

Pixellated Posterior Sampling of Point Spread Functions in Astronomical Images

CONNOR STONE,^{1,2,3,4} RONAN LEGIN,^{1,2,3} ALEXANDRE ADAM,^{1,2,3} NIKOLAY MALKIN,^{2,3,5} GABRIEL MISSAEL BARCO,^{1,2,3}
LAURENCE PERREAUL-LEVASSEUR,^{1,2,3,6,7,8} AND YASHAR HEZAVEH^{1,2,3,6,7,8}

¹*Ciela - Montreal Institute for Astrophysical Data Analysis and Machine Learning, Montréal, Canada*

²*Université de Montréal, Montréal, Canada*

³*Mila - Quebec Artificial Intelligence Institute, Montréal, Canada*

⁴*University of Toronto, Toronto, Canada*

⁵*University of Edinburgh, Edinburgh, Scotland*

⁶*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA*

⁷*Trottier Space Institute, Montréal, Canada*

⁸*Perimeter Institute, Waterloo, Canada*

Abstract

We introduce a novel framework for upsampled Point Spread Function (PSF) modeling using pixel-level Bayesian inference. Accurate PSF characterization is critical for precision measurements in many fields including: weak lensing, astrometry, and photometry. Our method defines the posterior distribution of the pixelized PSF model through the combination of an analytic Gaussian likelihood and a highly expressive generative diffusion model prior, trained on a library of HST ePSF templates. Compared to traditional methods (parametric Moffat, ePSF template-based, and regularized likelihood), we demonstrate that our PSF models achieve orders of magnitude higher likelihood and residuals consistent with noise, all while remaining visually realistic. Further, the method applies even for faint and heavily masked point sources, merely producing a broader posterior. By recovering a realistic, pixel-level posterior distribution, our technique enables the first meaningful propagation of detailed PSF morphological uncertainty in downstream analysis. An implementation of our posterior sampling procedure is available on GitHub.

Keywords: methods: statistical, techniques: image processing

1. INTRODUCTION

All astronomical observations are subject to image degradation arising from the redistribution of true sky flux due to the instrument’s physical characteristics. Distortions originating from imperfect optics, telescope jitter, scattered light, atmospheric turbulence (for ground based observatories), and diffraction are inescapable realities of real world observation (T. I. Liaudat et al. 2023). The composite effect of these distortions is formalized as the Point Spread Function (PSF): the observed image of an ideal point-like light source. Accurate characterization of the PSF remains a persistent challenge in observational astronomy (J. Krist 1993; J. E. Gunn et al. 1998; M. D. Perrin et al. 2012; R. G. Abraham & P. G. van Dokkum 2014; T. I. Liaudat et al. 2023; T. Schutt et al. 2025; J.-C. Cuillandre et al. 2025). Throughout this work, we refer to the effective PSF (ePSF), which encompasses not only the optical path PSF but also detector-level effects such

as pixel response and charge diffusion. Critically, the PSF is not static. For space-based observatories like the Hubble Space Telescope (HST), the ePSF exhibits significant spatiotemporal variability driven by mechanical and thermal variations over time, such as orbital focus changes or “breathing”. This variability, which also depends on position in the detector and wavelength, means that every observation is convolved with an essentially unique PSF, necessitating observation-specific modeling.

The accurate characterization of the PSF, and robust quantification of its morphological uncertainty, is critical for a broad spectrum of astronomical investigations. For example, precision shape measurements for weak gravitational lensing surveys are fundamentally limited by knowledge of the PSF (S. Farrens et al. 2022; T. I. Liaudat et al. 2023). An accurate and well-defined PSF model underpins the structural parameters derived from galaxy scaling relations (A. van der Wel et al. 2014) and

is essential for identifying fine-scale structures, such as the effect of dark matter subhalos in strong gravitational lensing analyses (S. Vegetti et al. 2010). Beyond structural analysis, the PSF is foundational for high-precision astrometry (J. Anderson & I. R. King 2000; P. Bennet et al. 2024) and accurate flux measurements (P. B. Stetson 1987; K. Kuijken 2008). In high-contrast imaging, a precise PSF model is required for robust subtraction of stellar flux to reveal faint companions (R. Galicher & J. Mazoyer 2024), while Active Galactic Nuclei (AGN) studies rely on it to deblend the central point source from its host galaxy (M.-Y. Zhuang & Y. Shen 2024). In all these applications, an accurate and precise PSF model is crucial to avoid bias in even fundamental measurements such as photometry; further the PSF model uncertainty is a source of systematic error if not accounted for correctly.

Given the recognized importance of the PSF, a diverse array of methodologies have been developed for its determination. These can be broadly separated into synthetic and empirical approaches. Synthetic methods utilize a first-principles approach, such as the analytic Airy disk pattern produced by a circular aperture (C. P. Society 1864; J. F. W. Herschel 1899). More sophisticated wavefront-propagation approaches compute the expected PSF from detailed instrument models (M. D. Perrin et al. 2012; L. Desdoigts et al. 2023). These may capture faint details that are otherwise challenging to extract, and can connect the components of the PSF with physical characteristics of the telescope. However, the complex, time-dependent nature of the PSF often necessitates its direct empirical inference from observed images.

The simplest empirical strategy involves analyzing isolated stars. This has traditionally been accomplished through parametric modeling, which fits an analytic function (e.g., a Moffat profile) to the star image (A. F. J. Moffat 1969; C. Y. Peng et al. 2002; C. J. Stone et al. 2023). While computationally efficient, this approach is realistically unable to capture the complex, high-frequency structures of a diffraction-limited PSF. More flexible basis methods, using Principal Component Analysis (PCA) (M. J. Jee et al. 2007), shapelets (A. Refregier 2003; S. Paulin-Henriksson et al. 2008), or starlets (K. Michalewicz et al. 2023), can trade off between detail and computational complexity. Other methods rely on stacking many stars to build a high-S/N ePSF template (J. Anderson & I. R. King 2006), but this yields a fixed point estimate that does not perfectly represent the unique PSF of a given observation.

Fundamentally, pixel-level forward modeling represents the most principled, data-driven methodology for

accurate PSF recovery. This approach models the PSF on a high-resolution pixel grid, offering the flexibility to capture arbitrary morphological features (E. Bertin 2011; M. Jarvis et al. 2021; Q. Liu et al. 2022). An upsampled pixel-level PSF model will have more free parameters than observations to constrain them, making this a severely ill-posed problem. Traditionally, this is overcome via regularization (E. Bertin 2011), or an implicit prior learned by a machine learning model (J. Herbel et al. 2018; D. Stark et al. 2018).

A Bayesian framework turns the ill-posed problem of recovering the upsampled PSF given low resolution noisy samples into the well-posed problem of determining the posterior distribution $P(M|D)$ over PSF models (M) given some data (D). Via Bayes' Theorem, this posterior is defined by a prior over PSF models $P(M)$ and a likelihood for the observed data $P(D|M)$. The likelihood may be formulated exactly under the assumption of independent Poisson-distributed pixel noise (often approximated as a Gaussian in the high count limit). The core challenge in Bayesian PSF inference is the specification of a realistic prior, $P(M)$. The complexity of PSF structures forced previous works to rely on simple, ad-hoc approximations for this prior. In fact, all traditional methods described above may be understood as priors imposed to restrict the solution space; though they are highly misspecified in this context.

In this work, we tackle the central challenge of representing a prior over PSF models explicitly. We replace the simple parametric, regularized, or implicit priors with a highly expressive, pixel-level prior $P(M)$ learned from high resolution sample images. To define our prior, we leverage score-based generative diffusion models (Y. Song et al. 2021), a framework seeing explosive growth in astrophysics (e.g., A. Adam et al. 2022; N. Mudur & D. P. Finkbeiner 2022; B. Remy et al. 2023; B. T. Feng et al. 2023; Z. Wu et al. 2024; R. Legin et al. 2025; N. Dia et al. 2025; A. Adam et al. 2025). This approach enables, for the first time, robust sampling of the full, pixel-level posterior distribution $P(M|D)$, yielding substantially improved PSF reconstructions and a principled propagation of morphological PSF uncertainty to downstream analysis. In Figure 1 we demonstrate our main comparison, our PSF posterior samples give visually realistic models and superior, noise-like, residuals. A final challenge, which we comment on but do not solve, is the generalization to a posterior $P(M(\vec{x})|D)$ of PSF models over the full detector frame rather than simply at the position of individual point sources.

In Section 2 we will present the Hubble-Space-Telescope (HST) data used for a concrete example of our analysis framework. In Section 3 we explain how

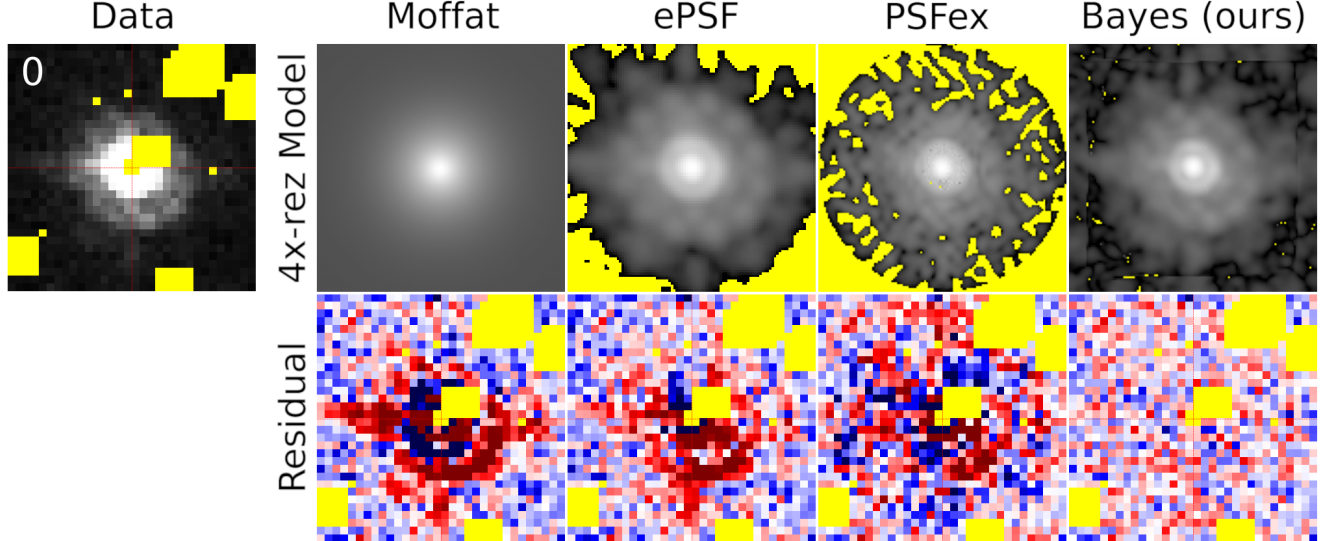


Figure 1. Methodological comparison. For a bright point source data cutout, we present three traditional modeling results alongside a single posterior sample from our procedure. The Moffat model includes only a smooth decline in brightness and leaves many structured residuals. The ePSF and PSFex models include structural details, resulting in unstructured, though still implausibly large residuals. Our Bayesian posterior sample gives detailed structure and pure noise-like residuals. Yellow pixels are masked for various reasons described in the text. A p-value test on the residuals (described in the text) gives 0.08 for our posterior sample, while for the other algorithms it confidently rejects the null hypothesis (6×10^{-17} , 2×10^{-22} , and 4×10^{-13} respectively).

the collected data forms a likelihood $P(D|M)$ which is shared over all methods considered here. In Section 4 we examine three widely used traditional methods for modeling PSFs. In Section 5 we present our fully Bayesian method for PSF reconstruction and demonstrate the robustness of the results. In Section 6 we compare the various methods and comment on the applicability of our work.

2. DATA

While the methodology developed in this work is applicable to any astronomical imaging dataset, we demonstrate its efficacy using Hubble Space Telescope (HST) Advanced Camera for Surveys / Wide Field Channel (ACS/WFC) F814W filter images⁹. To ensure the most accurate noise modeling and robust likelihood calculation for our fully Bayesian approach, we operate exclusively on the flat-field calibrated, unstacked *flt* frames. This is a deliberate choice not to use the common image stacking technique of Drizzle (A. S. Fruchter & R. N. Hook 2002). Since every observation is distorted by a unique PSF (e.g., via focus breathing over an HST or-

bit), Drizzle would be combining structurally different PSF models to make a stack. While it is common to use dithered images to improve sampling of the PSF to recover high resolution models, in such analyses care is taken to ensure the combined data share similar properties (detector position, orbital position, temperature, etc.). This is not the case for our *flt* frames and so stacking before modeling would be ill advised. The dataset comprises 108 individual *flt* frames, corresponding to multiple exposures across the selected fields¹⁰.

Point source identification was performed using the matched filtering algorithm: IRAF StarFinder (D. Tody 1986), as implemented within the *photutils* framework (L. Bradley et al. 2025). This routine identifies candidates by comparing a simple Gaussian kernel and then applying filters based on metrics of roundness and sharpness to isolate compact, unresolved sources. Many small extended objects and artifacts were selected in the initial IRAF Starfinder run alongside true point sources. We performed a visual inspection as a quality control step. This process reduced the initial 3,573 identified candidates to a final working sample of 1,355 high-quality point sources analyzed in this work.

⁹ This specific choice is motivated, though separate from, our other science goal: the high-fidelity analysis of galaxy-galaxy strong gravitational lenses from the Sloan Lens ACS (SLACS) Survey (A. S. Bolton et al. 2006). Our PSF models will be subsequently leveraged in R. Legin et al. (in prep.)

¹⁰ See R. Legin et al. (in prep.) for a discussion of the selected targets from the SLACS survey corresponding to the *flt* frames in question.

Each frame is accompanied by the requisite metadata, including a bad pixel mask identifying regions affected by detector artifacts, hot pixels, and cosmic rays. We implement a conservative masking strategy to ensure data quality. Specifically, any cosmic ray identified by the pipeline that spans more than one pixel is dilated by a single pixel to account for charge deposition in adjacent pixels, a phenomenon we observed in test analysis runs. All pixels exceeding 30,000 electrons are also masked, this is the empirical threshold for potential saturation in these data. The uncertainty associated with each pixel measurement is derived from the standard deviation provided in the *ERR* map. This is converted to the corresponding variance map (the square of the *ERR* map) for our Gaussian likelihood calculation under the assumption of independent pixel measurements.

For each of the vetted point sources, we extract a 32×32 pixel cutout. This cutout size corresponds to a 128×128 model image at four times upsampled resolution, which was set by the template ePSFs (see [Section 4.2](#)) and computational considerations for our diffusion model prior (see [Section 5.1](#)). Each cutout is treated as a distinct realization of the PSF. To support the subsequent Bayesian modeling, a detailed set of associated metadata is stored, including the matching variance and mask cutouts, sky background level, the World and detector pixel coordinates of the cutout center, and the image exposure time. Before analysis, the raw flux in units of electrons is converted to normalized flux units (electrons per second) by dividing the data cutout by the exposure time. Similarly, the variance map is normalized by dividing by the square of the exposure time. [Figure 2](#) illustrates an example fit frame and a collection of extracted star cutouts¹¹.

3. THE LIKELIHOOD

We adopt a standard Gaussian likelihood based on the assumption of independent, normally distributed noise in each pixel. The probability of observing the data vector D given a set of model parameters is defined as:

$$\mathcal{L}(D|M, x, y, f, s) = \mathcal{N}(D|f \cdot \mathbf{F}(M, x, y) + s, \Sigma), \quad (1)$$

where D is the vector of flux values corresponding to the unmasked pixels in the data cutout. M is a PSF model which may be evaluated at arbitrary pixel coordinates (it could be written as $M(i, j)$) even fractional pixel coordinates. For most models under consideration,

M is a four-times-upsampled grid of pixels which are converted to a smooth model via bilinear interpolation; the parametric Moffat model is intrinsically defined at any position. The set of nuisance parameters includes the relative sub-pixel shift vector (x, y) , the total source flux scalar f , and the uniform sky background s . Σ is an $N_{unmasked} \times N_{unmasked}$ diagonal covariance matrix, where the diagonal elements are the variance values derived from the HST *ERR* map for each unmasked pixel. By excluding masked pixels from D and Σ , the mask is implicitly incorporated into the likelihood calculation.

Having a diagonal covariance matrix is further reason not to use the Drizzle algorithm, as the stacking procedure produces highly covariant noise in the stacked pixels. While it is possible to model the covariance due to Drizzle, this would necessitate a dense Σ matrix as opposed to merely using the diagonal, making the technique less computationally efficient. Some pixel covariance may still be present in the *flt* images (e.g., due to charge diffusion), though it was small enough not to noticeably affect our results in [Section 5](#).

The core of the forward model is the operator $\mathbf{F}$, which maps the high-resolution PSF model M to the detector-scale observations. For each observed data pixel, $\mathbf{F}$ creates a 16×16 grid of points and evaluates the shifted model M at those locations. Averaging those 16^2 (256) samples gives the integral of flux in the data pixel contributed by the PSF model. This high resolution integration is motivated by the need to account for sub-pixel centering effects crucial for high-precision astrometry and photometry.

This likelihood formulation serves as the fundamental objective function for all subsequent analyses, including both traditional and fully Bayesian approaches. The form of our likelihood is highly standard in the field of astronomical image analysis. Further generalizations which include covariance between pixels do exist, such as the SLIC method ([R. Legin et al. 2023](#)), though we found no evidence that such techniques were necessary in our data.

4. TRADITIONAL METHODS

4.1. Parametric PSF Modeling

To establish a baseline for comparison, we first explore a standard parametric approach to PSF modeling. We utilize the Moffat profile ([A. F. J. Moffat 1969](#)), a common analytic form that is highly effective at characterizing the core structure of a PSF, although it lacks the complexity to capture high-order features such as diffraction spikes. The fitting is performed independently on each cutout using the general-purpose astronomical modeling code, *AstroPhot* ([C. J. Stone et al.](#)

¹¹ Note that star 10 in the [Figure 2](#) cutouts was later dropped due to a difficult to see interloper, therefore later figures index the stars without star 10.

Frame j9op01l8q, star positions

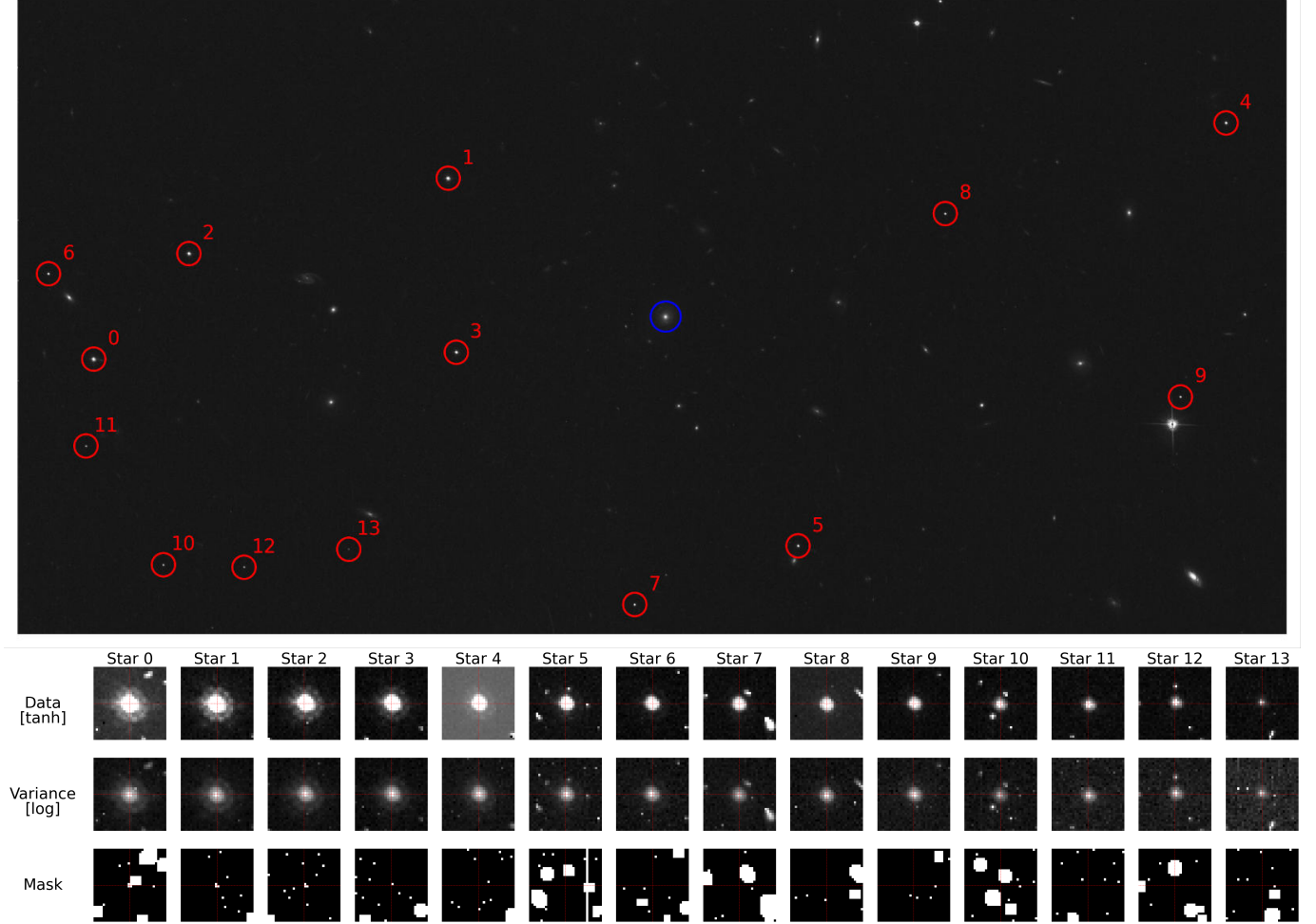


Figure 2. Example data from a single HST *flt* frame. Top: The full frame, with selected stars (red) and the science target (blue). Bottom: Cutouts of the data, variance, and mask for each star. These cutouts illustrate the core challenge: each star presents a unique, complex PSF morphology that must be modeled from noisy, incomplete (masked) data.

2023). In this implementation, the model consists of a single Moffat profile defined by its five parameters (centroid x and y , scale length, shape index, and amplitude), alongside a constant sky background as an additional nuisance parameter. The optimization is executed by minimizing the negative log-likelihood, which is computed for all unmasked pixels using the Gaussian formulation defined in Section 3.

The resulting fits are visually summarized in Figure 3. While the simple Moffat profile successfully models the general structure of the PSF, it leaves pronounced systematic residuals. For the brightest sources, these residuals frequently exceed 5σ significance, which effectively rules out the possibility that the remaining structure is attributable to Gaussian noise. This conclusion is quantitatively supported by the residual histograms in Figure 3. A Kolmogorov-Smirnov test is performed on the central region of the residuals and yields vanishingly

small p -values. The central region residuals correspond to the most pronounced deviations; they are selected by dividing the cutout into a 3×3 grid and taking the central segment.

Despite the significant systematic residuals, the formal uncertainties computed by *AstroPhot* for the Moffat parameters are exceedingly small. This result is not indicative of high accuracy but is a direct consequence of severe model misspecification. With only six parameters being constrained by approximately one thousand unmasked pixels, the model is highly overdetermined. Consequently, these small formal errors are misleading and do not provide a reliable measure of the true uncertainty associated with the real-world PSF.

4.2. ACS/WFC Focus-Diverse ePSF Models

Given the inherent limitations of the six-parameter Moffat model, a more flexible, pixelized representation

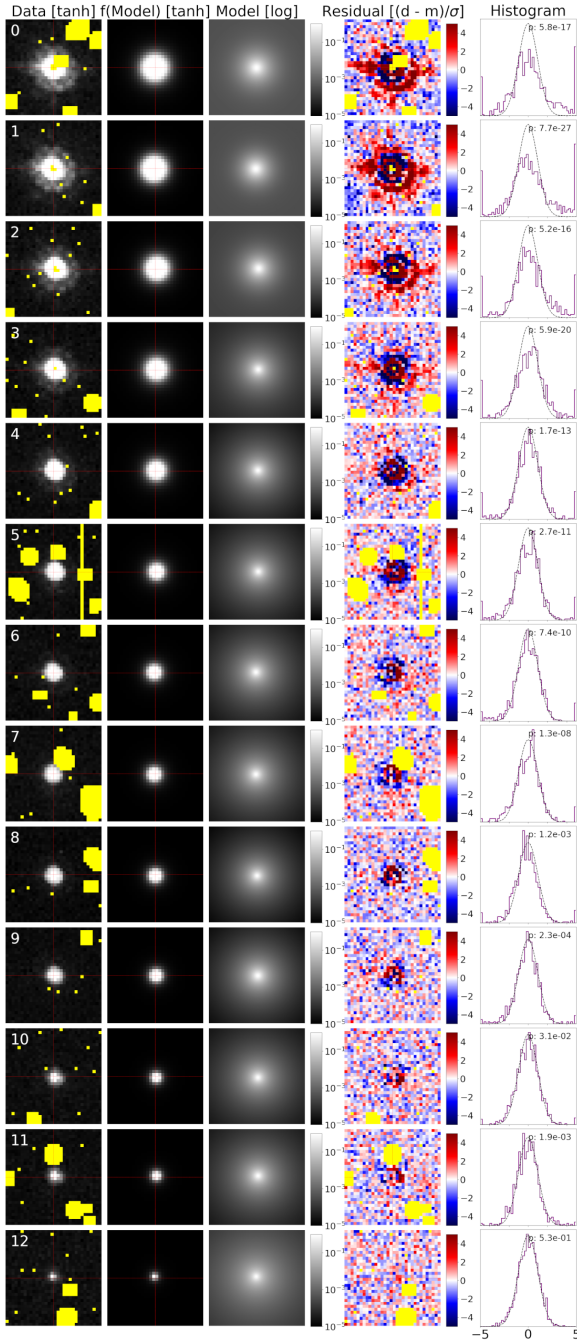


Figure 3. Traditional Method 1: Parametric Moffat Fit. From left to right the columns show the data, the best-fit Moffat model, the $4\times$ high-resolution model, the residuals, and a histogram of the residuals. Yellow pixels in the data and residuals are masked pixels (see text). The p-values in the top right of each histogram are from a Kolmogorov-Smirnov test on the central residuals against a Normal distribution (P. Virtanen et al. 2020). The key takeaway is that this method fails: the strong, structured patterns in the residuals (column 4) and the non-Gaussian residual histograms (column 5, with p-values ≈ 0) prove that a simple analytic function is a misspecified model for the true PSF.

is required to capture the detailed structure of the PSF. The effective PSF (ePSF) library developed by J. Anderson & I. R. King (2006) provides such a model collection. An “ePSF” model combines the ideal instrumental distortion with detector-level effects, such as the pixel response and charge diffusion, yielding the final projected PSF necessary for most science cases. The library was constructed by stacking thousands of high signal-to-noise stars and enforcing smoothness criteria across the spatial grid. Each template is a 101×101 pixel model defined at a $4\times$ supersampled resolution (10,201 free parameters per template). To account for spatial variation, the library provides a grid of 9×10 templates across the detector.

The PSF is subject to time-dependent variations, including focus changes driven by the Hubble Space Telescope’s orbital thermal cycles (A. Bellini et al. 2018). To account for these perturbations, subsequent work has provided methods to derive an image-specific, customized template set, facilitated by easy-to-use web tools (G. S. Anand et al. 2023). We adopt the G. S. Anand et al. (2023) template set specific to each frame for our analysis.

We then fit these fixed templates to our stellar cutouts. This is performed using Maximum Likelihood Estimation (MLE), minimizing the negative log-likelihood defined in Section 3. For each cutout, only the four nuisance parameters are allowed to vary: the sub-pixel shift (x,y), the total flux (f), and the sky background (s). The optimization converges rapidly given the limited number of free parameters.

As demonstrated in Figure 4, the ePSF templates represent the observed cutouts with significantly greater fidelity than the simple Moffat profile. The most prominent systematic residuals, such as diffraction spikes and ring structures, are successfully removed by the more detailed model. However, the residuals still contain numerous pixels exceeding 5σ significance, especially in the PSF cores. This indicates that the fixed templates, while broadly effective at modeling HST PSF structures, do not perfectly represent the true PSF for a specific observation.

Fundamentally, this approach provides only a point estimate of the PSF structure. Since the template itself is fixed, the fit yields only formal uncertainties on the four nuisance parameters, failing to propagate uncertainties in the PSF’s morphology. This leads to biased and overconfident results even for the nuisance parameters which would be used in downstream analysis (astrometry and photometry). This lack of model uncertainty coverage motivates our full Bayesian methodology.

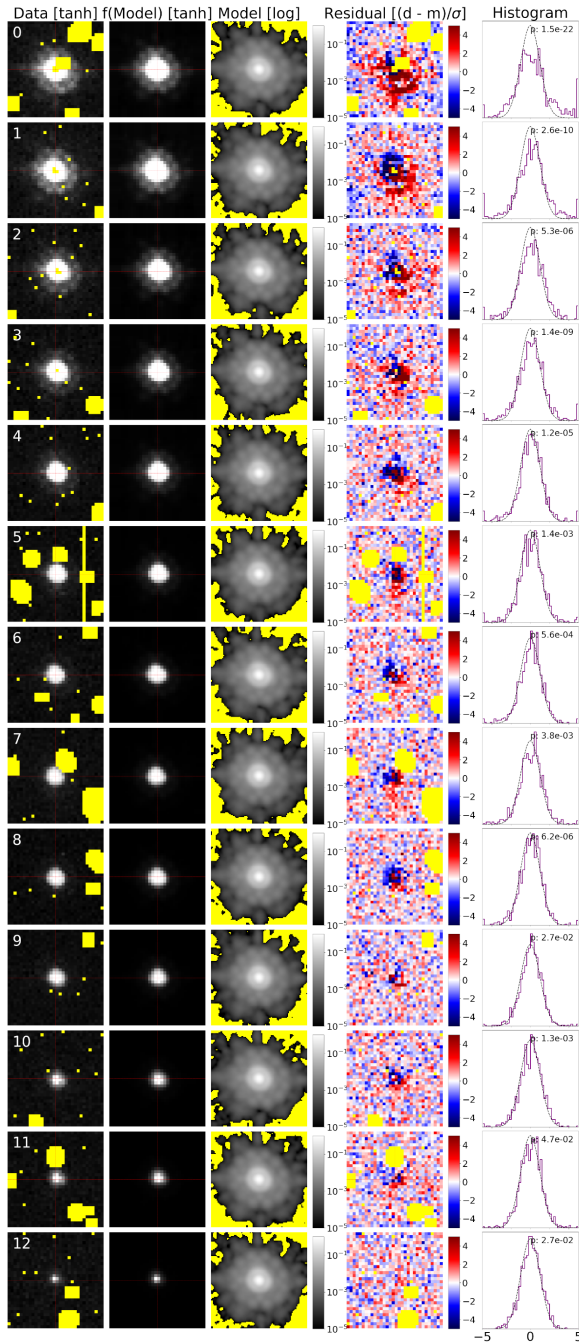


Figure 4. Traditional Method 2: Fixed ePSF Template. The format is identical to Figure 3, yellow pixels in column 3 indicate pixel values below 10^{-5} . This represents an improvement: the ePSF model (column 2) captures complex features like diffraction spikes, resulting in smaller residuals than the Moffat model. However, significant $> 5\sigma$ residuals remain, proving that a single, fixed “point estimate” template is still an imperfect match to the unique PSF of any given star.

4.3. Regularized Pixelated PSF Modeling

While template-based ePSFs offer improved accuracy, their reliance on external imaging campaigns means they are not perfectly tuned to the specific instrumental state of a given frame. We now investigate a method that performs pixelized optimization directly using the on-frame stellar observations. We adopt **PSFex** (E. Bertin 2011), a widely used algorithm designed to produce an empirical, spatially-varying PSF model by performing regularized pixel-level fitting to multiple point source cutouts simultaneously.

The PSFex workflow begins with source detection via **SourceExtractor** (E. Bertin & S. Arnouts 1996) to identify candidate stars. We use the default **SourceExtractor** configuration except to set an appropriate initial PSF FWHM, pixel size, and desired cutout size. PSFex then performs a χ^2 minimization to determine the pixel values in a supersampled model, constraining the model’s spatial variation across the detector with a low-order polynomial. Critically, because this is a severely ill-posed problem (fitting thousands of pixel values from a limited number of stars), PSFex employs Tikhonov regularization (A. N. Tikhonov 1963) to stabilize the solution. This regularization acts as an implicit, simple prior, penalizing pixel values that deviate from a smooth model. We use the default PSFex configuration, except to set a $4\times$ supersampled model resolution and a first-order polynomial for spatial variation¹².

For direct comparison, we use the final PSFex model as a fixed template and perform a Maximum Likelihood Estimation for the four nuisance parameters (position, flux, and sky), identical to the procedure in Section 4.2. The results are shown in Figure 5. While the overall fidelity appears comparable to the ePSF templates, the residuals for bright stars are often larger. Disconcertingly, in some cases (e.g., Star 4), the residuals are too small to be plausibly drawn from a Normal noise distribution, indicating that the flexible PSFex model has overfit the data. This overfitting is a direct consequence of the misspecified prior: Tikhonov regularization, which merely encourages smoothness, is a poor approximation for the true detailed structure of the HST PSF. Consequently, the fit does not yield the expected $\chi^2/\text{DoF} \approx 1$ and, like the other traditional methods, provides only a single point estimate without a rigorous quantification of morphological uncertainty.

¹² higher-order models were observed to overfit our data given the limited number of stars in the field

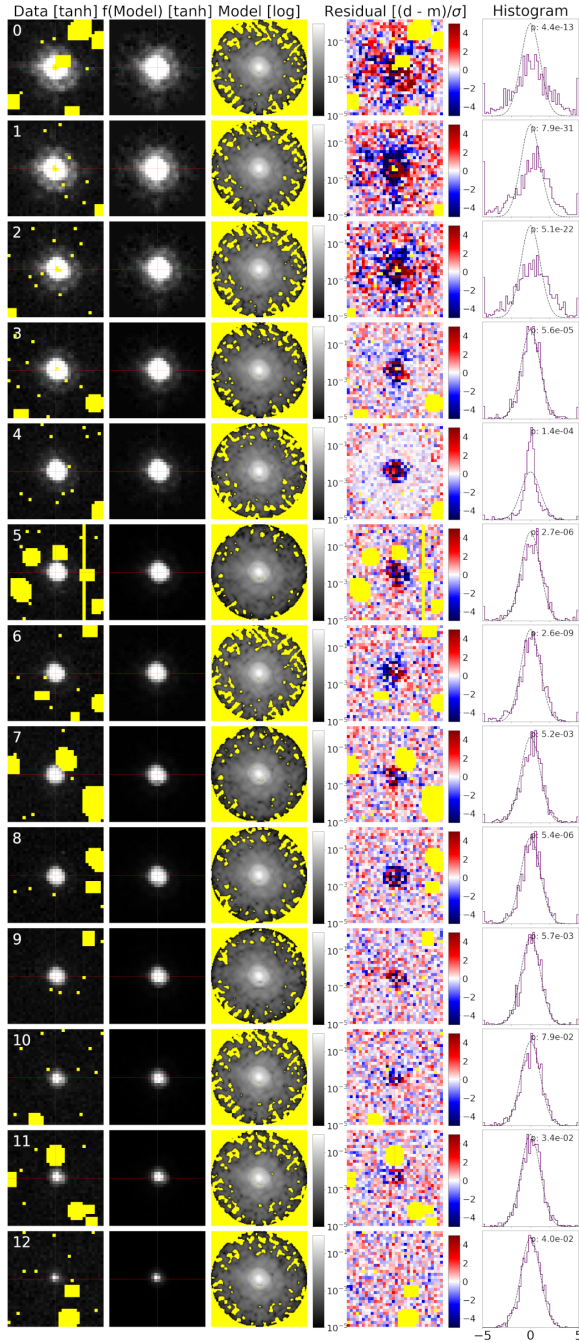


Figure 5. Traditional Method 3: Empirical PSFex Model. Format is identical to Figure 3. This method shows mixed failure: most stars have implausibly large residuals, while others (e.g., Star 4) have residuals that are too small, indicating the flexible model has overfit the data. This demonstrates that PSFex’s simple smoothness regularization is a misspecified prior, motivating our new Bayesian approach.

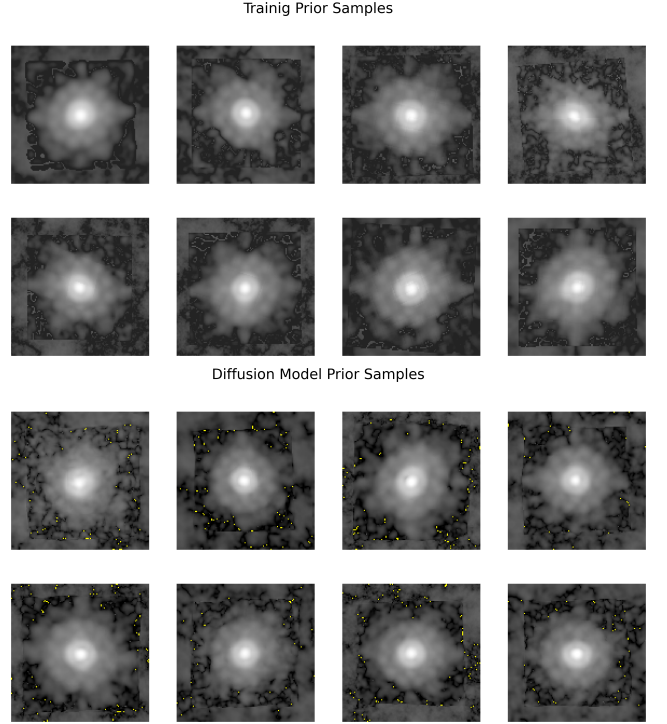


Figure 6. Top: Example prior samples based on the ePSF templates from the J. Anderson & I. R. King (2006) library, used for training. Bottom: New, unique samples drawn from our trained diffusion model prior. The prior samples (bottom) are highly realistic, capturing the complex morphology of the training data (top), but are not simple copies. This demonstrates our model has learned a flexible, generative prior (a manifold of realistic PSFs) rather than just memorizing the training set.

5. PIXEL LEVEL BAYESIAN PSF MODELING

5.1. Constructing a Prior PSF Model

To perform a Bayesian analysis, we must first define a robust prior distribution, $P(M)$, over the space of plausible high-resolution PSF models M . All traditional methods, as discussed in Section 4, can be understood as imposing a simple, and ultimately misspecified, prior. Here, we instead construct a highly expressive, non-linear prior by training a generative model to learn the complex morphological features of realistic PSFs. We leverage score-based generative diffusion models (Y. Song et al. 2021), which implicitly represent this complex prior through a learned “score function”, $\nabla_M \log(P(M))$. The application of diffusion models to problems in astronomy has seen recent explosive growth (e.g., B. Huang et al. 2018; K. Karchev et al. 2022; A. Adam et al. 2022; N. Mudur & D. P. Finkbeiner 2022; B. Remy et al. 2023; B. T. Feng et al. 2023; A. Bourdin et al. 2024; R. Legin et al. 2024; Z. Wu et al. 2024; R. Legin et al. 2025; E. Angeloudi et al. 2025; N.

Dia et al. 2025; A. Adam et al. 2025). This framework allows us to sample the posterior $P(M|D)$ by combining the log likelihood gradient, $\nabla_M \log(P(D|M))$, with this learned score function of the prior.

A key requirement for this approach is a representative training set. We use the 9,776 individual ePSF templates described in Section 4.2. Given we showed in Section 4.2 that the ePSF templates match individual cutouts imperfectly, training on them may seem an odd choice. However, our premise is that while a single template fails, the full ensemble of templates robustly spans the manifold of realistic PSF morphologies (including focus, spatial, and thermal variations). A further benefit of using score-based diffusion models is that they produce conservative priors with support outside their training samples; this allows for out of distribution posterior samples when the likelihood is highly informative (A. Adam et al. 2022; F. Rozet et al. 2024; S. Wagner-Carena et al. 2025; G. M. Barco et al. 2025). Our goal is to learn a prior that is broader than the training set. We achieve this in two ways. First, we apply extensive data augmentation to the training samples, including shifts, flips, scaling, rotations, skews, and blurring, to expand the effective prior space (each applying to 50% of the training samples). Second, as we will discuss, we intentionally stop the model’s training early to prevent it from “memorizing” the discrete templates.

Our diffusion model, implemented using a U-Net architecture (O. Ronneberger et al. 2015) from the `score_models` package¹³, has several key technical features. For numerical efficiency, the 101×101 templates are padded to a 128×128 resolution; the empty outskirts are filled with a Gaussian Random Field whose power-law spectrum approximately matches the observed PSF outskirts. To enforce the physical constraint of non-negativity, all training models are thresholded at a floor value of 5×10^{-5} relative to the peak flux. This value was chosen to be safely above the minimum noise level (1×10^{-5}) used in the diffusion model’s sampling process, again ensuring numerical stability.

Given the high dynamic range of the PSF (over 10^5), we found it numerically unstable to model the full pixel values directly. Instead, we compute an average template $\bar{M}$ and train the diffusion model to learn only the deviations from this mean. This dynamic-range-reducing transformation is applied purely for numerical stability and has no effect on the validity of our analysis (i.e., it applies without loss-of-generality).

Training duration proved to be a critical hyperparameter. Extended training caused the model to overfit, concentrating its score function on the training examples. This “spiky” prior would then resist solutions favored by the likelihood, giving large residuals. By empirically selecting an “early stopping” point, we obtained a model that balances prior expressiveness with flexibility. As shown in Figure 6, all the morphological features of the training data samples can be seen in our prior samples, meaning the overall structure of the PSF has been learned. That said, samples drawn from our prior have slight visual distinctions from the training examples, confirming that we have successfully defined a broader, more flexible prior manifold.

A minor, yet notable, boundary artifact is visible in the extended wings of the prior samples, producing a faint square structure. This feature is directly inherited from the padding process used to expand the 101×101 templates to the 128×128 input size for the U-Net. Critically, this structure occurs at fluxes approximately 10^{-4} relative to the PSF peak, rendering it irrelevant for standard photometric and shape analyses. For highly specialized investigations, the posterior samples can be safely cropped back to 101×101 pixels. Alternatively, sampling on extremely high signal-to-noise sources allows the dominant likelihood term to effectively overpower the prior in the low-flux regimes, naturally correcting the artifact, as demonstrated in A. Adam et al. (2022).

Several promising avenues exist for improving our diffusion model prior. Our current reliance on the J. Anderson & I. R. King (2006) ePSF templates could be improved by training the model on high-precision wavefront PSF models (M. D. Perrin et al. 2012; L. Desdoigts et al. 2023), which might provide a more physically motivated basis and extend to fainter features. Our prior model is also agnostic to the position, flux, or color/spectrum of the source being modeled. Since the PSF is known to be a function of these properties, we could further narrow the prior model distribution by incorporating them into the model, rather than producing a single broader model that covers all cases. These directions for improvement would be interesting directions for further progress, but are beyond the scope of this work.

5.2. Posterior Samples

Our diffusion model prior requires a linear forward model to enable efficient sampling (Y. Song et al. 2021; A. Adam et al. 2022). However, the full likelihood in Equation 1 is non-linear with respect to the combined sub-pixel shift (x, y) , flux f , and sky level s . We there-

¹³ https://github.com/AlexandreAdam/score_models

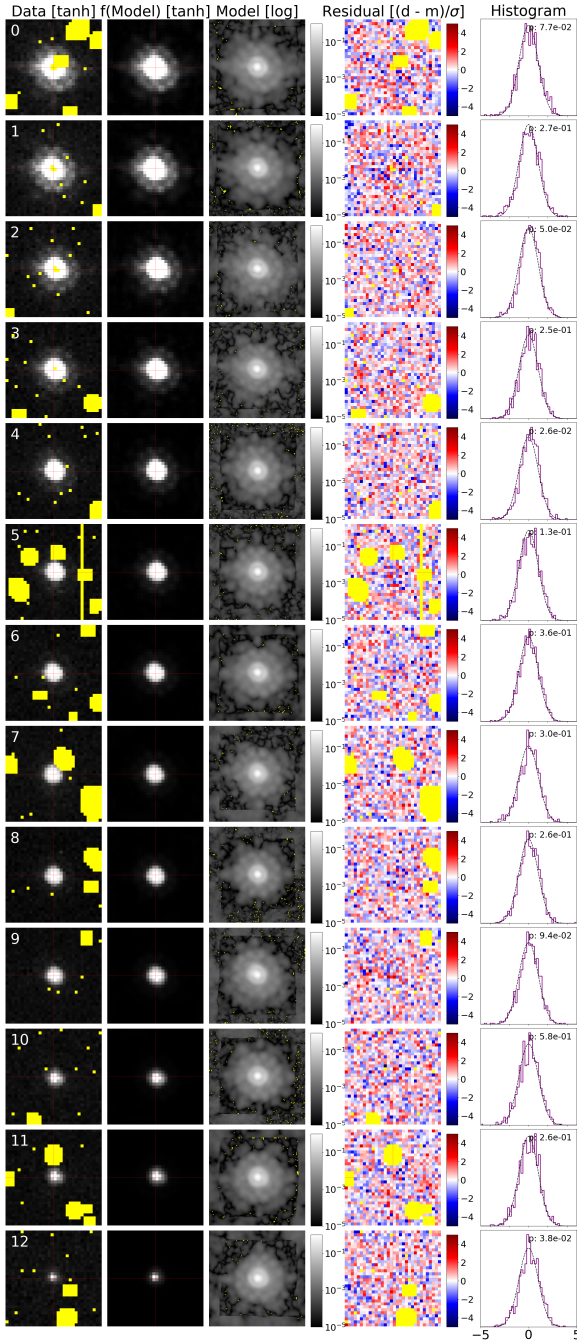


Figure 7. Pixel-level Bayesian posterior PSFs (our work). Format is identical to Figure 3. The model shown (column 3) is a single, random sample from the posterior rather than a point estimate of the PSF. The residuals (column 4) are unstructured and consistent with pure noise, and the residual histograms (column 5) are Normal (KS test p-values often ≈ 0.5). This demonstrates our framework produces a statistically sufficient model that fully explains the data.

fore cannot sample all parameters simultaneously. Instead, we adopt a pragmatic, iterative approach that alternates between optimization and sampling:

1. Optimize Nuisance Parameters: We hold the PSF model M fixed and find the Maximum Likelihood values for the nuisance parameters (x, y, f, s) .
2. Sample the PSF: We fix the nuisance parameters at their optimized values, and sample from the PSF posterior.

For the initial round we set the PSF model to $\bar{M}$ the average ePSF template sample. The forward model $\mathbf{F}$ is now a simple linear operation mapping M to the data in step 2. This linearity permits the use of the “convolved likelihood approximation” from A. Adam et al. (2022), a computationally efficient method for diffusion model posterior sampling. We cycle this two-step procedure twice to ensure all parameters have settled. We find this is sufficient, as any minor, remaining uncertainties in the nuisance parameters are effectively marginalized by the high-dimensional, pixel-level flexibility of the PSF posterior itself. This is because the PSF posterior sampling has some small flexibility on the position, flux, and sky level by controlling all pixels. In a final run, we produce 32 posterior samples to represent the posterior distribution for each cutout.

This sampling process involves numerically integrating a Stochastic Differential Equation (SDE) defined by the combined prior score function and the likelihood gradient. Conceptually, this procedure begins with pure noise and iteratively “de-noises” it; at each step, the sample is guided by both the likelihood (what the data requires) and the prior (what a realistic PSF looks like), ensuring the final sample is a valid posterior sample. For the integration, we employ the Heun SDE integrator from the `score_models` package. We use 4,096 integration steps and incorporate two Langevin corrector steps per iteration¹⁴, which helps reduce discretization errors from the SDE integration. An implementation and example of our posterior sampling procedure is available here: <https://github.com/Ciela-Institute/HSTPSF>.

Figure 7 presents a single, randomly chosen posterior sample for each cutout. The results are a clear improvement over the methods in Section 4. The residual images exhibit essentially no discernible structure, and the χ^2/dof is approximately one for all examples. This strongly indicates that our Bayesian framework has produced a statistically sufficient model, successfully ab-

¹⁴ During the initial two cycles, where we optimize the nuisance parameters, 2,048 SDE steps and no corrector steps are used.

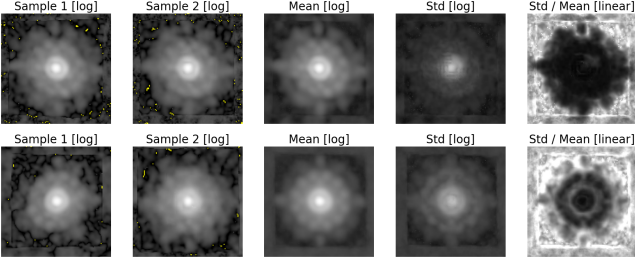


Figure 8. Variance of Posterior samples. Columns show two random posterior samples (1-2), the posterior mean (3), the standard deviation (4), and fractional standard deviation (5). These are based on 32 posterior samples per cutout. This figure demonstrates the power of our method: we now have a principled, pixel-level quantification of morphological uncertainty (which may be marginalized as in column 4). As expected, the bright star (top) is highly data-constrained (low variance), while the faint star (bottom) is less constrained (higher variance), producing a broader posterior.

absorbing all non-noise signal from the data into the posterior ensemble. While a faint residual is visible in cutout 1, this is not present in all posterior samples for that star, highlighting the importance of using the full posterior ensemble rather than a single realization.

Figure 8 visualizes the quantified uncertainty by showing the marginal variance map across the 32 posterior samples for two example sources. As expected, the high-SNR star (top) is more tightly constrained than the low-SNR star (bottom). For both, the bright central regions are data-constrained, while the faintest outskirts are almost entirely prior-driven. The posterior mean is shown for visualization, but we stress that the posterior mean is not a posterior sample and holds no special statistical meaning, as the true posterior is highly non-Gaussian.

5.3. Interpolating to an image position

Our Bayesian procedure yields posterior samples $\{M_k\}$ at the discrete locations of observed point sources. However, most downstream analyses require a PSF model at an arbitrary position $\vec{x}$ within the detector plane (e.g., at the centroid of an extended galaxy). This necessitates a method for interpolating the PSF field. Interpolation is a non-trivial challenge in PSF modeling. Standard methods, such as fitting a low-order polynomial across the detector (e.g., **PSFex**), are severely under-constrained given our typical source density of 2 to 20 stars per frame. A more rigorous, high-dimensional Gaussian Process (GP) is computationally prohibitive (C. E. Rasmussen 2003), as our PSF model exists in a $D = 16,384$ dimensional space. Furthermore, it is unlikely that our relatively small collection of 32 posterior samples per point source are sufficient to robustly estimate the high-dimensional covariance re-

quired for a meaningful GP model over the field. While a fully Bayesian approach would involve a prior conditioned on detector position allowing for PSF posteriors over the full detector field, this is beyond the scope of this work.

Given the intractability of a rigorous interpolation, we adopt a pragmatic, non-parametric weighted mean approach. An interpolated PSF model $M(\vec{x})$ is computed as a weighted average of posterior sample draws from all observed sources, $\{M_k\}$. The weight w_k for each source k is defined as:

$$w_k = \frac{1}{\sigma_k^2} e^{-0.5 \frac{\|\vec{x}_k - \vec{x}\|^2}{1024^2}}, \quad (2)$$

where $\|\vec{x}_k - \vec{x}\|$ is the spatial distance, σ_k^2 is the variance of that source’s posterior, and the weights are normalized to sum to one. The spatial weighting kernel is intentionally broad; this ensured that the interpolation is approximately evenly weighted with the variance.

To propagate the posterior uncertainty, we employ a Monte Carlo approach: we randomly select one posterior sample from each source’s ensemble, perform the weighted interpolation, and repeat this procedure multiple times. This generates a “pseudo-posterior” ensemble at the target position $\vec{x}$, capturing morphological variations.

While this weighted interpolation scheme is a practical compromise, we validate its performance using a leave-one-out cross-validation. In this test, we remove a star, interpolate a pseudo-posterior to its location using the remaining stars, and compare the result to the held-out data. The results of this validation, which demonstrate superior residuals to the methods in Section 4, are presented in Appendix B.

6. DISCUSSION AND CONCLUSIONS

We have successfully demonstrated a novel framework that produces pixel-level Bayesian posterior Point Spread Functions (PSF) in HST imagery using a Gaussian likelihood and generative diffusion model prior. This approach was benchmarked against three traditional PSF modeling techniques: simple parametric fitting (Moffat), fixed template libraries (ePSF), and empirical regularized likelihood methods (PSFex). In comparison, our method not only yields visually more realistic PSFs but, critically, achieves likelihood values orders of magnitude greater. Our PSF models give residuals statistically consistent with noise in most cases and adequately capture the detailed structure of the effective PSF. The production of a full posterior distribution for the PSF morphology enables the previously intractable marginalization over model uncertainty in subsequent

analyses. This is not merely a formal improvement; we demonstrate in [Appendix A](#) that traditional methods can introduce systematic flux biases of order 5%. Our procedure also yields unbiased measurements for significantly fainter stars than traditional methods, simply producing a broader posterior in accordance with the lower signal-to-noise. An implementation and example of our posterior sampling procedure is available here: <https://github.com/Ciela-Institute/HSTPSF>.

One limitation of the work is the requirement of PSF training samples for the generative model used as prior. In our case, previous works by [J. Anderson & I. R. King \(2006\)](#) extracted highly detailed upsampled PSF models which we could adopt for training. Another viable approach is to use analytically modeled PSFs with wavefront propagation ([L. Desdoigts et al. 2023](#)). In the absence of either type of pre-determined PSF models, another approach would be to learn the upsampled PSF model directly from the data using an iterative maximum likelihood algorithm such as described in [F. Rozet et al. \(2024\)](#); [G. M. Barco et al. \(2025\)](#).

The primary limitation identified in this work is the use of a non-parametric, ad-hoc weighted mean for interpolating the PSF posterior to arbitrary image locations. Our chosen weighting scheme (based on posterior variance and spatial proximity) is computationally efficient and practical. A leave-one-out cross-validation test in [Appendix B](#) demonstrates this interpolation scheme is superior even to the fitting results of traditional models, though it lacks the rigor of a full Bayesian formulation. Integrating a principled interpolation scheme, such as a high-dimensional Gaussian Process or a flexible polynomial expansion (like that used in PSFex), is an exciting direction for future work. The most formal approach would be a PSF prior conditioned on detector position; given a likelihood at the position of individual stars, this would produce posterior realizations of the full frame variable PSF model.

Despite the current limitations in interpolation, the availability of full, quantified PSF posterior distribu-

tions holds transformative implications across a number of astronomical disciplines. As discussed in the introduction, this work directly enhances the precision of measurements in fields including weak gravitational lensing, astrometry, high-contrast imaging, galaxy scaling relations, and photometry. By providing a statistically complete description of the PSF uncertainty, this method introduces a new paradigm of precision PSFs which push to the limits of information in a given point source cutout. We anticipate that this work will enable levels of astrophysical precision previously considered intractable.

ACKNOWLEDGMENTS

C.S. acknowledges the support of a NSERC Postdoctoral Fellowship and a CITA National Fellowship. This research was made possible by a generous donation by Eric and Wendy Schmidt via the Schmidt Sciences Foundation. Y.H. and L.P. acknowledge support from the Canada Research Chairs Program, the National Sciences and Engineering Council of Canada through grants RGPIN-2020- 05073 and 05102. N.M. acknowledges funding from CIFAR, Genentech, Samsung, and IBM. The work of A.A. and R.L. were partially funded by NSERC CGS D scholarships. R.L. acknowledges support from the Centre for Research in Astrophysics of Quebec and the hospitality of the Flatiron Institute. G.M.B. acknowledges support from the Fonds de recherche du Québec – Nature et technologies (FRQNT) under a Doctoral Research Scholarship (<https://doi.org/10.69777/368273>).

Software used: `astropy` ([Astropy Collaboration et al. 2022](#)), `AstroPhot` ([C. J. Stone et al. 2023](#)), `caskade` ([C. Stone et al. 2025](#)), `matplotlib` ([J. D. Hunter 2007](#)), `numpy` ([C. R. Harris et al. 2020](#)), `PyTorch` ([A. Paszke et al. 2019](#)), `PSFex` ([E. Bertin 2011](#)), `SourceExtractor` ([E. Bertin & S. Arnouts 1996](#)), `scipy` ([P. Virtanen et al. 2020](#)), `score_models` (Adam et al., in-prep.)

REFERENCES

- Abraham, R. G., & van Dokkum, P. G. 2014, *PASP*, 126, 55, doi: [10.1086/674875](https://doi.org/10.1086/674875)
- Adam, A., Coogan, A., Malkin, N., et al. 2022, arXiv e-prints, doi: [10.48550/arXiv.2211.03812](https://doi.org/10.48550/arXiv.2211.03812)
- Adam, A., Stone, C., Bottrell, C., et al. 2025, *AJ*, 169, 254, doi: [10.3847/1538-3881/adb039](https://doi.org/10.3847/1538-3881/adb039)
- Anand, G. S., Grogan, N. A., Anderson, J., Cohen, Y., & Bellini, A. 2023, The ACS/WFC Focus-Diverse ePSF Webtool,, Instrument Science Report ACS 2023-06, 12 pages
- Anderson, J., & King, I. R. 2000, *PASP*, 112, 1360, doi: [10.1086/316632](https://doi.org/10.1086/316632)
- Anderson, J., & King, I. R. 2006, PSFs, Photometry, and Astronomy for the ACS/WFC,, Instrument Science Report ACS 2006-01, 34 pages

- Angeloudi, E., Huertas-Company, M., Falcón-Barroso, J., et al. 2025, arXiv e-prints, arXiv:2509.25340, doi: [10.48550/arXiv.2509.25340](https://doi.org/10.48550/arXiv.2509.25340)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Barco, G. M., Adam, A., Stone, C., Hezaveh, Y., & Perreault-Levasseur, L. 2025, ApJ, 980, 108, doi: [10.3847/1538-4357/ad9b92](https://doi.org/10.3847/1538-4357/ad9b92)
- Bellini, A., Anderson, J., & Grogan, N. A. 2018, Focus-diverse, empirical PSF models for the ACS/WFC,, Instrument Science Report ACS 2018-8
- Bennet, P., Patel, E., Sohn, S. T., et al. 2024, ApJ, 971, 98, doi: [10.3847/1538-4357/ad5349](https://doi.org/10.3847/1538-4357/ad5349)
- Bertin, E. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435
- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: [10.1051/aas:1996164](https://doi.org/10.1051/aas:1996164)
- Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703, doi: [10.1086/498884](https://doi.org/10.1086/498884)
- Bourdin, A., Legin, R., Ho, M., et al. 2024, arXiv e-prints, arXiv:2408.00839, doi: [10.48550/arXiv.2408.00839](https://doi.org/10.48550/arXiv.2408.00839)
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2025, astropy/photutils: 2.2.0, 2.2.0 Zenodo, doi: [10.5281/zenodo.14889440](https://doi.org/10.5281/zenodo.14889440)
- Cuillandre, J.-C., Bertin, E., Bolzonella, M., et al. 2025, A&A, 697, A6, doi: [10.1051/0004-6361/202450803](https://doi.org/10.1051/0004-6361/202450803)
- Desdoigts, L., Pope, B. J. S., Dennis, J., & Tuthill, P. G. 2023, Journal of Astronomical Telescopes, Instruments, and Systems, 9, 028007, doi: [10.1117/1.JATIS.9.2.028007](https://doi.org/10.1117/1.JATIS.9.2.028007)
- Dia, N., Yantovski-Barth, M. J., Adam, A., et al. 2025, arXiv e-prints, arXiv:2501.02473, doi: [10.48550/arXiv.2501.02473](https://doi.org/10.48550/arXiv.2501.02473)
- Farrens, S., Guinot, A., Kilbinger, M., et al. 2022, A&A, 664, A141, doi: [10.1051/0004-6361/202243970](https://doi.org/10.1051/0004-6361/202243970)
- Feng, B. T., Smith, J., Rubinstein, M., et al. 2023, arXiv e-prints, arXiv:2304.11751, doi: [10.48550/arXiv.2304.11751](https://doi.org/10.48550/arXiv.2304.11751)
- Fruchter, A. S., & Hook, R. N. 2002, PASP, 114, 144, doi: [10.1086/338393](https://doi.org/10.1086/338393)
- Galicher, R., & Mazoyer, J. 2024, Comptes Rendus Physique, 24, 133, doi: [10.5802/crphys.133](https://doi.org/10.5802/crphys.133)
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040, doi: [10.1086/300645](https://doi.org/10.1086/300645)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Herbel, J., Kacprzak, T., Amara, A., Refregier, A., & Lucchi, A. 2018, JCAP, 2018, 054, doi: [10.1088/1475-7516/2018/07/054](https://doi.org/10.1088/1475-7516/2018/07/054)
- Herschel, J. F. W. 1899, Treatises on physical astronomy, light and sound contributed to the Encyclopaedia metropolitana (R. Griffin)
- Huang, B., Zhang, K., Lin, Y., Schölkopf, B., & Glymour, C. 2018, in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18 (New York, NY, USA: Association for Computing Machinery), 1551–1560, doi: [10.1145/3219819.3220104](https://doi.org/10.1145/3219819.3220104)
- Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Jarvis, M., Bernstein, G. M., Amon, A., et al. 2021, MNRAS, 501, 1282, doi: [10.1093/mnras/staa3679](https://doi.org/10.1093/mnras/staa3679)
- Jee, M. J., Blakeslee, J. P., Sirianni, M., et al. 2007, PASP, 119, 1403, doi: [10.1086/524849](https://doi.org/10.1086/524849)
- Karchev, K., Anau Montel, N., Coogan, A., & Weniger, C. 2022, arXiv e-prints, arXiv:2211.04365, doi: [10.48550/arXiv.2211.04365](https://doi.org/10.48550/arXiv.2211.04365)
- Krist, J. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 536
- Kuijken, K. 2008, A&A, 482, 1053, doi: [10.1051/0004-6361:20066601](https://doi.org/10.1051/0004-6361:20066601)
- Legin, R., Adam, A., Hezaveh, Y., & Perreault-Levasseur, L. 2023, ApJL, 949, L41, doi: [10.3847/2041-8213/acd645](https://doi.org/10.3847/2041-8213/acd645)
- Legin, R., Ho, M., Lemos, P., et al. 2024, MNRAS, 527, L173, doi: [10.1093/mnras/slad152](https://doi.org/10.1093/mnras/slad152)
- Legin, R., Isi, M., Wong, K. W. K., Hezaveh, Y., & Perreault-Levasseur, L. 2025, ApJL, 985, L46, doi: [10.3847/2041-8213/add681](https://doi.org/10.3847/2041-8213/add681)
- Legin, R., Stone, C., Adam, A., et al. in prep., TBD
- Liaudat, T. I., Starck, J.-L., & Kilbinger, M. 2023, Frontiers in Astronomy and Space Sciences, 10, 1158213, doi: [10.3389/fspas.2023.1158213](https://doi.org/10.3389/fspas.2023.1158213)
- Liu, Q., Abraham, R., Gilhuly, C., et al. 2022, ApJ, 925, 219, doi: [10.3847/1538-4357/ac32c6](https://doi.org/10.3847/1538-4357/ac32c6)
- Michalewicz, K., Millon, M., Dux, F., & Courbin, F. 2023, Journal of Open Source Software, 8, 5340, doi: [10.21105/joss.05340](https://doi.org/10.21105/joss.05340)
- Moffat, A. F. J. 1969, A&A, 3, 455
- Mudur, N., & Finkbeiner, D. P. 2022, arXiv e-prints, arXiv:2211.12444, doi: [10.48550/arXiv.2211.12444](https://doi.org/10.48550/arXiv.2211.12444)

- Paszke, A., Gross, S., Massa, F., et al. 2019, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc.), 8024–8035. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, 484, 67, doi: [10.1051/0004-6361:20079150](https://doi.org/10.1051/0004-6361:20079150)
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, 124, 266, doi: [10.1086/340952](https://doi.org/10.1086/340952)
- Perrin, M. D., Soummer, R., Elliott, E. M., Lallo, M. D., & Sivaramakrishnan, A. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8442, *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, ed. M. C. Clampin, G. G. Fazio, H. A. MacEwen, & J. M. Oschmann, Jr., 84423D, doi: [10.1117/12.925230](https://doi.org/10.1117/12.925230)
- Rasmussen, C. E. 2003, in *Summer school on machine learning* (Springer), 63–71
- Refregier, A. 2003, *MNRAS*, 338, 35, doi: [10.1046/j.1365-8711.2003.05901.x](https://doi.org/10.1046/j.1365-8711.2003.05901.x)
- Remy, B., Lanusse, F., Jeffrey, N., et al. 2023, *A&A*, 672, A51, doi: [10.1051/0004-6361/202243054](https://doi.org/10.1051/0004-6361/202243054)
- Ronneberger, O., Fischer, P., & Brox, T. 2015, *arXiv e-prints*, arXiv:1505.04597, doi: [10.48550/arXiv.1505.04597](https://doi.org/10.48550/arXiv.1505.04597)
- Rozet, F., Andry, G., Lanusse, F., & Louppe, G. 2024, in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=7v88Fh6iSM>
- Schutt, T., Jarvis, M., Roodman, A., et al. 2025, *The Open Journal of Astrophysics*, 8, 26, doi: [10.33232/001c.132299](https://doi.org/10.33232/001c.132299)
- Society, C. P. 1864, *Transactions of the Cambridge Philosophical Society*, Vol. 10 (University Press)
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., et al. 2021, in *International Conference on Learning Representations*. <https://openreview.net/forum?id=PXTIG12RRHS>
- Stark, D., Launet, B., Schawinski, K., et al. 2018, *MNRAS*, 477, 2513, doi: [10.1093/mnras/sty764](https://doi.org/10.1093/mnras/sty764)
- Stetson, P. B. 1987, *PASP*, 99, 191, doi: [10.1086/131977](https://doi.org/10.1086/131977)
- Stone, C., Adam, A., Coogan, A., Perreault-Levasseur, L., & Hezaveh, Y. 2025, *Journal of Open Source Software*, 10, 8786, doi: [10.21105/joss.08786](https://doi.org/10.21105/joss.08786)
- Stone, C. J., Courteau, S., Cuillandre, J.-C., et al. 2023, *MNRAS*, 525, 6377, doi: [10.1093/mnras/stad2477](https://doi.org/10.1093/mnras/stad2477)
- Tikhonov, A. N. 1963, *Sov Dok*, 4, 1035
- Tody, D. 1986, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 627, *Instrumentation in astronomy VI*, ed. D. L. Crawford, 733, doi: [10.1117/12.968154](https://doi.org/10.1117/12.968154)
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28, doi: [10.1088/0004-637X/788/1/28](https://doi.org/10.1088/0004-637X/788/1/28)
- Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, *MNRAS*, 408, 1969, doi: [10.1111/j.1365-2966.2010.16865.x](https://doi.org/10.1111/j.1365-2966.2010.16865.x)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Wagner-Carena, S., Akhmetzhanova, A., & Erickson, S. 2025, *arXiv e-prints*, arXiv:2510.05205, doi: [10.48550/arXiv.2510.05205](https://doi.org/10.48550/arXiv.2510.05205)
- Wu, Z., Sun, Y., Chen, Y., et al. 2024, *arXiv e-prints*, arXiv:2405.18782, doi: [10.48550/arXiv.2405.18782](https://doi.org/10.48550/arXiv.2405.18782)
- Zhuang, M.-Y., & Shen, Y. 2024, *ApJ*, 962, 139, doi: [10.3847/1538-4357/ad1183](https://doi.org/10.3847/1538-4357/ad1183)

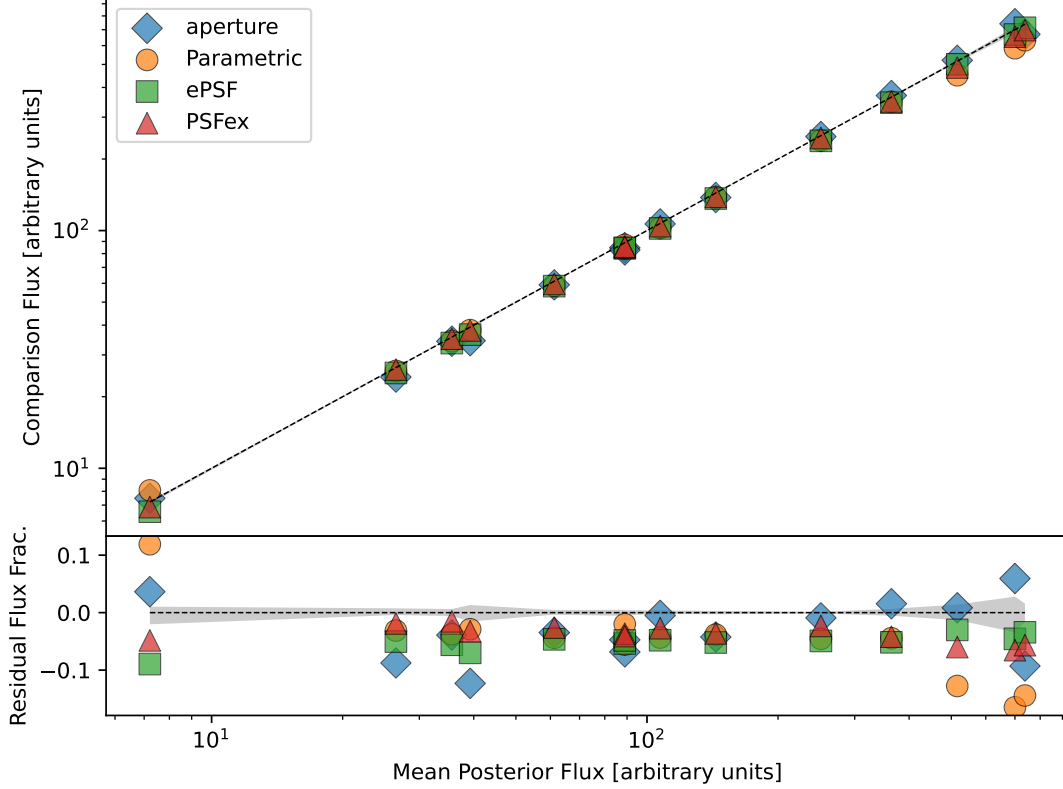


Figure 9. Biased photometry from misspecified priors. The posterior mean flux (x-axis) is compared against aperture, Moffat, ePSF template, and PSFex photometry (y-axis). A dashed line shows ideal agreement for reference. The lower subfigure shows the fractional residual flux ($\frac{F_{\text{estimate}} - F_{\text{posterior}}}{F_{\text{posterior}}}$), including a grey band for the 16-84 percentile of our posterior flux values. The takeaway is that model misspecification leads to biased flux measurements: the traditional methods (Aperture, Moffat, ePSF, PSFex) show a consistent $\sim 5\%$ flux bias. Our posterior (grey band) provides an accurate, unbiased flux measurement with a principled uncertainty estimate which propagates morphological uncertainty.

APPENDIX

A. FLUX MEASUREMENT COMPARISON

While flux estimation was not the primary goal of this analysis, it is instructive to compare the flux measurements derived from the various PSF fitting techniques. Figure 9 compares the single point-estimate fluxes from the traditional methods against the posterior mean and 16th-84th percentile range derived from our Bayesian posterior samples. It is immediately clear that all traditional methods produce fluxes that are biased relative to our posterior estimates. The parametric Moffat model is the most severely affected; its significant model misspecification (see Figure 3) translates directly into a large photometric bias. The PSFex and ePSF template models perform more consistently as a function of flux, but both still exhibit a systematic $\sim 5\%$ bias, underestimating the flux for most point sources. Interestingly, there appears to be no systematic trend with flux, potentially indicating that our pixel flux limit (30,000 electrons) is successfully deselecting any non-linearities in our data. We also include aperture flux values in the comparison, though these should be considered highly inaccurate as no correction was made for the masked pixels, which can be numerous, and would bias the estimates low.

Overall, these results provide a quantitative warning: even a seemingly simple, summary statistic like total flux is highly sensitive to the accuracy of the underlying PSF model. This demonstrates that great care must be taken in a model’s selection and validation to avoid propagating significant systematic biases into downstream scientific

measurements. Enticingly, this means there is likely opportunity for much higher precision in already available data, so long as models are sufficiently accurate.

B. PSF INTERPOLATION TEST

In [Section 5.3](#), we introduced a non-parametric, weighted-mean scheme to interpolate our PSF posterior to arbitrary detector locations. This appendix provides some validation for that pragmatic choice. We compare two distinct pixel-wise interpolation schemes:

1. **Weighted Mean:** The method from [Section 5.3](#), which weights each source’s PSF model by its posterior variance and spatial proximity (using a broad 1024-pixel Gaussian kernel).
2. **Linear Fit:** A scheme that fits a linear model in detector coordinates (a 2D plane) to the value of each pixel independently, weighting each source by its posterior variance.

To test these methods, we perform a leave-one-out cross-validation. For each star in our sample, we remove it, generate an interpolated PSF at its location using only the remaining stars, and then fit this interpolated model to the held-out star’s data (optimizing only for position, flux, and sky).

The results are presented in [Figure 10](#). As expected, the interpolated models perform worse than the direct posterior fits shown in [Figure 7](#). As seen with the other tests, the brightest point sources are the most difficult to model sufficiently accurately, while the faint star residuals are consistent with noise even for the interpolated models. However, they are visually superior to all three traditional methods shown in [Figures 3 - 5](#), with residuals that are far more consistent with noise. This confirms that our framework, even when interpolated, provides a more accurate PSF model.

Comparing the two interpolation schemes, the weighted mean (left panel) produces slightly lower residuals and, critically, suffers from fewer “catastrophic failures” where the interpolation results in non-physical negative fluxes (seen as yellow pixels in the right panel). For this reason, we adopt the more robust weighted-mean scheme for our work, while acknowledging that a more rigorous, fully Bayesian interpolation model remains a promising avenue for future research.

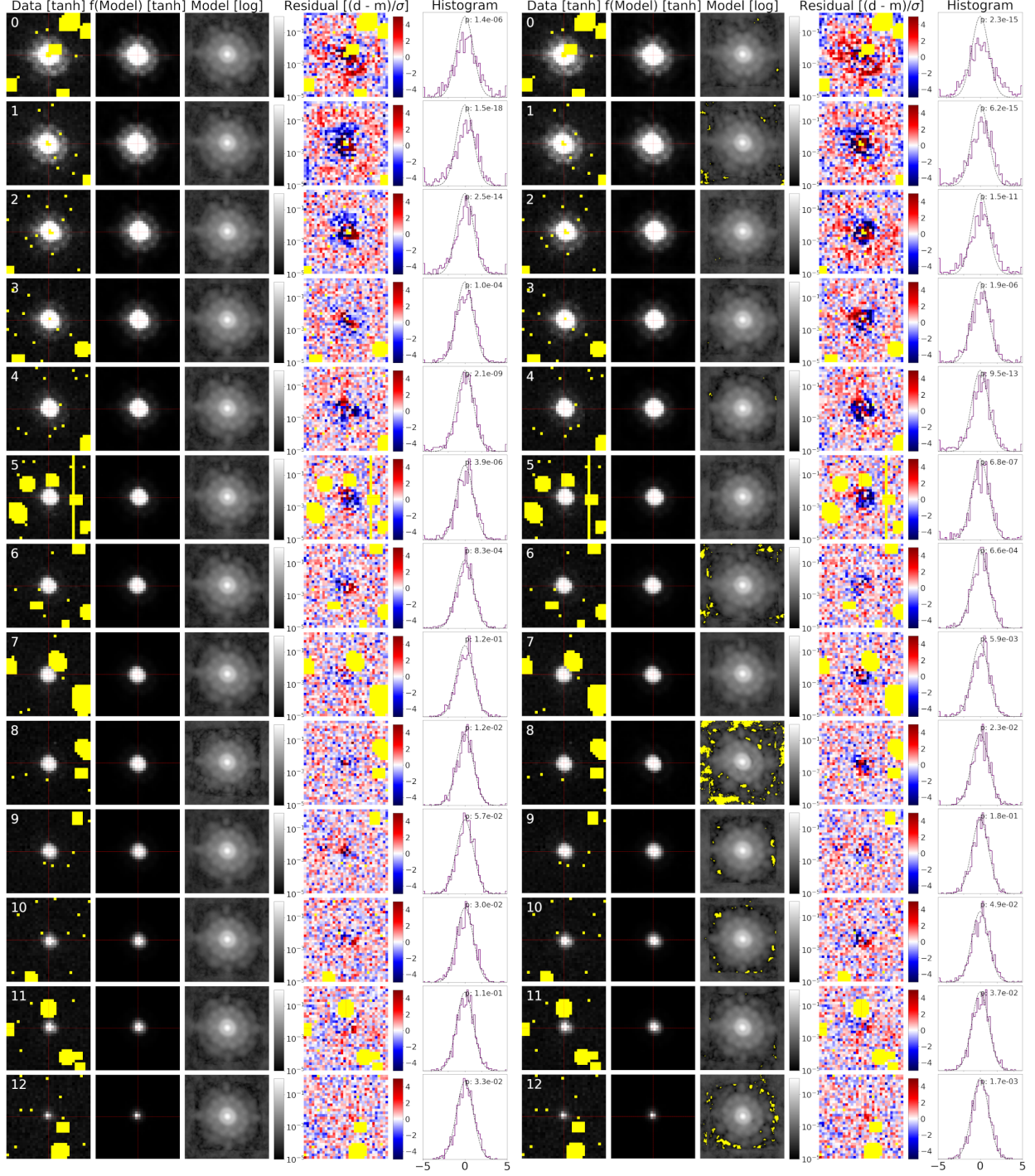


Figure 10. Cross-validation: interpolating the PSF to arbitrary positions. Format is identical to Figure 3. This figure shows a leave-one-out test: each star was removed, and a new PSF was interpolated to its location using only the other stars on the frame. This interpolated model was then fit with an MLE to the held-out star. The resulting residuals are visually superior to all traditional methods in Section 4. This validates our weighted-mean scheme (left) as a practical method for generating a PSF pseudo-posterior at any arbitrary position.