Validation of PSF models for *HST* and other space-based observations

Bryan R. Gillis [●],[1][*] Tim Schrabback,[2] Ole Marggraf,[2] Rachel Mandelbaum [●],[3] Richard Massey [●],[4,5]
Jason Rhodes[6,7] and Andy Taylor[1]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory Edinburgh, Edinburgh EH9 3HJ, UK*
[2]*Argelander Institute for Astronomy, University of Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*
[3]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[4]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham DH1 3LE, UK*
[5]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[6]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena CA 91109, USA*
[7]*California Institute of Technology, 1200 E. California Blvd., CA 91125, USA*

ABSTRACT
Forthcoming space-based observations will require high-quality point spread function (PSF) models for weak gravitational lensing measurements. One approach to generating these models is using a wavefront model based on the known telescope optics. We present an empirical framework for validating such models to confirm that they match the actual PSF to within requirements by comparing the models to the observed light distributions of isolated stars. We apply this framework to TINY TIM, the standard tool for generating model PSFs for the *Hubble Space Telescope* (*HST*), testing its models against images taken by *HST*'s Advanced Camera for Surveys in the Wide Field Channel. We show that TINY TIM's models, in the default configuration, differ significantly from the observed PSFs, most notably in their sizes. We find that the quality of TINY TIM PSFs can be improved through fitting the full set of Zernike polynomial coefficients that characterize the optics, to the point where the practical significance of the difference between model and observed PSFs is negligible for most use cases, resulting in additive and multiplicative biases both of order $\sim 4 \times 10^{-4}$. We also show that most of this improvement can be retained through using an updated set of Zernike coefficients, which we provide.

Key words: gravitational lensing: weak – methods: data analysis.

1 INTRODUCTION

Optical observations are fundamentally limited by the point spread function (PSF) of the observing instrument and conditions. This describes the manner in which a point light source will be affected to spread the paths of photons, displaying not as a point but as an extended profile. Various effects contribute to this, including atmospheric diffraction, diffraction due to obscurations within the instrument, aberrations in the instruments optics, and various detector effects (Krist, Hook & Stoehr 2011). Accurate knowledge of the PSF is thus necessary for any pursuit that requires knowledge of the undistorted light distribution of an object.

For instance, weak gravitational lensing uses the shapes of objects before PSF distortion as an estimator for the gravitational shear field. Elliptical PSFs will contribute to the observed ellipticities of sources, and since in general the PSF spreads detected photons, this contributes to circularizing them. These effects, respectively, result in additive and multiplicative biases in the ellipticities of sources, and so it is necessary to have a proper characterization of the PSF in order to make unbiased measurements of source shapes (Kaiser, Squires & Broadhurst 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998; Bartelmann & Schneider 1999; Kaiser 2000).

For ground-based observations, the PSF is dominated by atmospheric conditions that vary on short time-scales, and so its profile in any given observation must be determined empirically, using point source objects such as stars to sample it and interpolating appropriately (Jarvis & Jain 2004; Rowe 2010; Hamana et al. 2013; Lu et al. 2017). For space-based observations, however, it is possible to design a system so that the PSF is relatively stable, determined only by instrumental effects such as diffraction due to obscurations, optical aberrations, and polishing errors. It is therefore theoretically possible to instead model the PSF based on an understanding of the instrument's optics (Krist et al. 2011). This is particularly useful for measurements that require high precision such as the upcoming *Euclid* mission (Laureijs et al. 2011), as for an empirical model, the number of observed stars might not be sufficient to overcome noise, particularly in the wings of the PSF. Additionally, the *Euclid* mission is designed to have as stable a PSF as possible, making it an ideal candidate for using a model PSF.

Before such PSF models can be used for measurements, it will be necessary to validate them through comparisons with observations, to ensure that the models actually are faithful recreations of the instrumental PSF. It is not immediately obvious how best to perform such validation, and it is necessary to weight the chances and impacts of false positives and false negatives for any validation procedure. In this paper, we discuss the decisions that must be made, and we present a validation framework which is targeted at ensuring a PSF model is suitable for weak lensing measurements.

[*] E-mail: brg@roe.ac.uk

We demonstrate this framework by applying it to TINY TIM (Krist 1993; Krist et al. 2011) model PSFs and comparing them to observations taken with the *Hubble Space Telescope* (hereafter '*HST*'). TINY TIM has a long history of use in processing *HST* images, for purposes such as e.g. measuring the sizes of star clusters (Whitmore et al. 1999) and distant galaxies (van Dokkum et al. 2008), measuring the width of a protostar jet (Burrows et al. 1996), strong gravitational lensing observations (Surpi & Blandford 2003), and weak gravitational lensing observations (Leauthaud et al. 2007).

The literature is sparse, however, when it comes to validation tests of TINY TIM. Of note, van der Wel et al. (2012) tested Tiny-Tim-based PSF models against a sample of 46 isolated stars and found a 4 per cent discrepancy in the amount of light outside of an 0.2 arcsec circular aperture. Imperfections in the model were also noted in Krist et al. (2011), where they note 'For subtraction in the wings, usually an observed reference PSF provides superior results than using a model, as the match to the fine scale structure is better.' Hoffmann & Anderson (2017) also compared TINY TIM to an empirical PSF, and found that the empirical model outperformed TINY TIM by a factor of 2 in their quality-of-fit metric. Insight can also be gained from comparisons of observed PSFs with models generated in other manners, such as Jee et al. (2007)'s analysis of PSF models generated with a principle-component analysis technique.

Further testing of TINY TIM is thus warranted to see if these results can be replicated with a larger sample and to test further properties of the PSF beyond its light profile that might affect lensing measurements and other applications which rely on TINY TIM models.

In Section 2, we discuss the data and models we use to test our framework. In Section 3, we present our proposed testing framework for the validation of PSF models. In Section 4, we present and discuss our control tests. We discuss our findings in Section 5 and conclude in Section 6.

2 DATA AND MODELS

2.1 Data

For our tests, we use data from the *HST* observed with the Advanced Camera for Surveys ('ACS', Avila et al. 2016) in the wide-field channel using the F606W filter. We use a set of 217 pointings observed between the dates 2009 July 9 and 2011 July 16, selected for high-stellar density. These are generally observations of globular clusters such as NGC 104, but also include images of the nearby dwarf galaxy NGC 185. Each pointing comprises a set of two 4096×2048 pixel images, with an approximate pixel scale of 0.05 arcsec pixel^{-1}, for an observed area of ~ 5.8 arcmin^2 for each image. The pointings cover regions of varying stellar number density and have varying exposure, resulting in a greatly variable number of usable stars per image (from ~ 10 to ~ 200, with significant bimodality in the distribution).

The full reduction pipeline used for our data is detailed in Schrabback et al. (2018), but in short, the images are reduced primarily with the standard ACS calibration pipeline CALACS. The images are corrected for the effects of charge-transfer inefficiency through the procedure presented in Massey et al. (2014). Each of the images is then processed by the MultiDrizzle tool (Koekemoer et al. 2003) to correct for cosmic rays, but the final 'drizzle' step (which stacks exposures together) is not performed, as this alters the effective PSF of the stacked image and complicates analysis of it. Since we work only with unstacked images from this point forwards, we must be careful to ONLY perform operations that are valid on unstacked images, which means e.g. avoiding interpolation.

2.2 TINY TIM PSF models

For this work, we test the PSF models generated by TINY TIM (Krist 1993; Krist et al. 2011), the standard tool for generating model PSFs for the *HST*. TINY TIM calculates the PSF through a wavefront model, using the known obscuration pattern of the *HST* and assuming that all obscuration happens in a single plane (see section 3.1 of Krist et al. 2011 for justification of this assumption). The wavefront is represented by an $n \times n$ array, where n is sufficiently large that the wavefront is Nyquist sampled, and the wavefront is propagated to the focus with a fast Fourier transform. The square of the magnitude of the electromagnetic field is then taken to produce the model PSF. TINY TIM also includes prescriptions to model aberrations due to defocus and polishing errors, and can also include a prescription for jitter if desired.

The effective PSF is also affected by the diffusion of charge between neighbouring pixels. This step can be modelled by TINY TIM if desired, but it is often more useful for the user to request a subsampled model PSF. Such a model PSF can be shifted so that it represents the proper placement of the point source relative to the centre of the nearest pixel. The subsampled model PSF can then be rebinned and convolved with the 3×3 charge diffusion kernel provided by Tiny Tim to generate the final model PSF, which is what we do in our analysis.

We test TINY TIM with a use-case similar to the approach of e.g. Rhodes et al. (2007), who fit focus offset value input to TINY TIM for each image by finding the value which resulted in PSF models being generated which best matched the observed stars. We use the downloadable TINY TIM tool, version 7.5,[1] to generate model PSFs. The tool's first executable, `tiny1`, asks the user a series of questions which allow it to determine the PSF to be generated, and it stores the necessary information in a parameter file. This file also contains values which the tool does not ask the user about but which can be modified, such as the assumed coma and astigmatism. As we expect users will not do this unless it is shown to be necessary, we will start by testing the scenario in which the only adjusted parameters are those which the tool asks about. These parameters and our selections for them are as follows:

(i) Camera: 15 (ACS – Wide Field Channel).
(ii) Detector: Chosen per image based on the 'CCDCHIP' header keyword.
(iii) Position on detector: Chosen per star, taking the nearest of a fixed 32×16 grid of positions. This constrains the number of PSFs which need to be generated, and was found to not noticeably impact our results per our analysis on control fields (see Section 4).
(iv) Filter passband: F606W.
(v) Spectrum: 1; 15 (Use the K7V spectrum, which is chosen to represent a typical star in the sample. We discuss the impact of this in Section 4.)
(vi) PSF diameter: 2.0 arcsec.
(vii) Focus-secondary mirror despace: Fit per image. As *HST* 'breathes' due to its passing in and out of the Earth's shadow, the position of the focus relative to the secondary mirror changes over time and cannot be perfectly predicted for any given observation (Krist 2003; Anderson & King 2006). We thus have to fit the best focus position by simulating multiple sets of PSFs for each image.

The chosen Position on Detector value is used by TINY TIM to determine field dependent aberrations and charge diffusion, which

---

[1] http://tinytim.stsci.edu/static/tinytim-7.5.tar.gz

are used in the generation of the PSF. If the user desires, it is possible to edit the generated parameter file to change the Zernike coefficients that determine aberrations, as we test in Section 5.2. The values in this parameter file correspond to the values at the centre of the field, and TINY TIM adds these to position-dependent offsets for each coefficient to determine the value to use for each generated PSF.

A finely subsampled model PSF is then generated through running `tiny2` with the generated parameter file. For the ACS, it is also necessary to apply distortion due to the fact that it is installed off-axis. This is done through the `tiny3` command, which also allows the user to determine their desired subsampling factor for the final model PSF. We choose a subsampling factor of 8×, based on our analysis on control fields (see Section 4). This factor is large enough to allow us to shift the model PSF to adequately match the proper subpixel centre for any given star. We apply this shift for each star, based on the relative positions of the best-fitting centre of the star and the subsampled model PSF, and then rebin the model PSF to the same pixel scale as the detector. Finally, we apply the charge-diffusion kernel provided for each subsampled model PSF to generate the final model PSF for each star. At this stage, we compare the subpixel centre for the star and rebinned model PSF. If the difference is greater than the subsampled pixel scale, we return to the subsampled model and shift it an additional amount corresponding to this offset, and then rebin and convolve with the charge-diffusion kernel again. We repeat this process until the centre converges.

3 PSF VALIDATION FRAMEWORK

3.1 Designing tests

For a given PSF model, ideally its predictions of the PSF from a point source will differ from its observed light profile only in noise and the effects of environment (i.e. other nearby sources on the sky). We can therefore propose various statistical tests which we expect that an ideal model will typically pass. The most straightforward such test is a χ^2 test on the fluxes of a set of pixels, comparing the predicted flux from the model to the observed flux surrounding an isolated point source, using a proper characterization of the noise and taking into account the number of parameters of the PSF model which must be fitted to the data. Such a test has the following issues, however:

(i) It tests the statistical significance of a departure from a perfect model, but a statistically significant failure of the test does not mean that the model is unusable. Depending on what the model is used for, its imperfections may have negligible or even no effect on the resulting measurements.

(ii) Truly isolated point sources are difficult to identify. Stars are effectively point sources, and the brighter stars can be reasonably distinguished from galaxies through a selection in size–magnitude space, but many apparent stars are in fact unresolved binary star systems (e.g. Lada 2006 finds ∼31 per cent of stellar systems host more than one star), and this binary nature will result in the light profile being larger and more elliptical than that of a point source.

(iii) This test requires that the background behind each point source be uniformly zero. If the background is not perfectly subtracted, this will result in the χ^2 statistic for the comparison being spuriously large.

We discuss these issues in the following subsections.

3.1.1 Relevant quantities

As the primary motivation of this work is to develop a validation framework for PSF models to be used for gravitational lensing, we desire our tests to be failed if and only if imperfections in the PSF would lead to significant errors in the measured gravitational lensing signals. To determine what tests might be necessary, let us briefly provide an overview of how lensing signals are measured. Lensing signals are determined from the measured shear estimates derived from galaxies light distributions. If galaxies are assumed to be elliptical and randomly oriented, then their measured ellipticities[2] will provide unbiased estimates of the amount of shear in the absence of pixelization and the influence of a PSF. The influence of the PSF will affect the apparent ellipticities of observed galaxies, and so it must be properly characterized for an accurate measurement of a lensing signal. See e.g. Mandelbaum (2015) for further details on the theory and methods underlying weak gravitational lensing.

The two most common approaches to shear-measurement algorithms are moments-based and model-fitting approaches. Moments-based approaches primarily use measurements of the weighted quadrupole moments of observed galaxies and PSFs[3] to calculate the most likely ellipticity of the undistorted galaxy image. Model-fitting approaches use model profiles for the undistorted galaxies, convolve these with the PSFs, and test the convolved models against the observed galaxies to find the best-fitting models, and the ellipticities of these models are then used. Often, an approach called 'Metacalibration' (Sheldon & Huff 2017) is applied, testing the impact of perturbations to the data on the shear estimates and using this to calibrate the shear estimates. This still assumes a perfect PSF model in the most straightforward implementations though, and as such will not remove the sensitivity to imperfections in the model.

For the purposes of testing PSFs, the manner in which moments-based shear-measurement algorithms work is particularly illuminating. These methods are primarily sensitive to the quadrupole moments of the PSFs, and so PSF models that have imperfections in these moments will have direct effects on the estimates of galaxies' undistorted ellipticities. The PSF's dipole moments are also relevant, as they will affect the determination of its centre and thus the calculations of its quadrupole moments. As the PSF conserves flux, its monopole moment is only useful for normalizing other moments.

As used in gravitational lensing analysis, the normalized weighted multipole moments are defined as

$$M_x = \frac{1}{m_0} \iint x \, w(x, y) \, I(x, y) \, \mathrm{d}^2 A,$$

$$M_y = \frac{1}{m_0} \iint y \, w(x, y) \, I(x, y) \, \mathrm{d}^2 A,$$

$$M_{xx} = \frac{1}{m_0} \iint x^2 \, w(x, y) \, I(x, y) \, \mathrm{d}^2 A,$$

$$M_{yy} = \frac{1}{m_0} \iint y^2 \, w(x, y) \, I(x, y) \, \mathrm{d}^2 A,$$

$$M_{xy} = \frac{1}{m_0} \iint xy \, w(x, y) \, I(x, y) \, \mathrm{d}^2 A, \qquad (1)$$

---

[2]For ellipticity, we use the definition $|e| = (1 - r)/(1 + r)$, where r is the axial ratio of the galaxy.

[3]Methods typically also use higher order moments for corrections to their estimates. It is not necessary to take this into account here, as the quadrupole moments are the most significant factors going into the estimates, and they sufficiently constrain the PSF fitting.

where

$$m_0 = \iint w(x, y)\, I(x, y)\, \mathrm{d}^2 A,$$

$w(x, y)$ is a weight function that quickly approaches zero at some finite radius, $\mathrm{d}^2 A$ represents an integral over the full relevant area, and the central position $(x, y) = (0, 0)$ is generally chosen so that M_x and M_y will be zero for some weight function. In the case of a pixelized image, the integral can be replaced with a summation. The two ellipticity components of an object can be calculated through one of many algorithms using these moments. For example, one of the simpler algorithms is

$$\hat{\epsilon}_1 = \frac{M_{xx} - M_{yy}}{M_{xx} + M_{yy} + 2\sqrt{|\mathsf{M}|}},$$

$$\hat{\epsilon}_2 = \frac{2 M_{xy}}{M_{xx} + M_{yy} + 2\sqrt{|\mathsf{M}|}}, \tag{2}$$

(Seitz & Schneider 1997) where

$$|\mathsf{M}| = M_{xx} M_{yy} - M_{xy}^2. \tag{3}$$

Note that this particular algorithm is not commonly used, as the reliance on $|\mathsf{M}|$, which includes the products of moments, gives the algorithm high sensitivity to noise. More commonly, algorithms instead use simply $M_{xx} + M_{yy}$ in the denominator and apply other corrections that do not rely on the products of moments (e.g. Kaiser et al. 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998). However, the mathematical simplicity of this algorithm is more beneficial for our purpose here of estimating the impact of model PSF inaccuracies on shear estimates.

In the simplified scenario of unweighted moments, the normalized quadrupole moments of the undistorted galaxy can be calculated from the difference of the normalized moments of the observed galaxy and PSF:

$$\mathsf{M}^{(u)} = \mathsf{M}^{(o)} - \mathsf{M}^{(p)}, \tag{4}$$

where $\mathsf{M}^{(u)}$ represents the quadrupole moments of the undistorted galaxy, $\mathsf{M}^{(o)}$ those of the observed galaxy, and $\mathsf{M}^{(p)}$ those of the PSF. Combined with equation (2), we can identify and define the most relevant moments-based quantities for the PSF as

$$M_+ = M_{xx} - M_{yy},$$
$$M_\times = 2 M_{xy},$$
$$M_* = M_{xx} + M_{yy} + 2\sqrt{|\mathsf{M}|}. \tag{5}$$

As the dipole moments of the PSF are used to determine its centroid, it is important to test them as well, so we will additionally include M_x and M_y in our analysis.

From work by Paulin-Henriksson et al. (2008) and Rowe (2010), we know in particular that errors in the PSF size will lead to multiplicative biases in shear estimates, and errors in the PSF shape will lead to additive biases. We can identify here M_* as related to the size of the PSF, and M_+ and $M_\times$ as related to the shape of the PSF, which implies errors in the former will cause multiplicative biases, and errors in the latter will cause additive biases. This is confirmed by our own calculations in Appendix B.

Note that of these five terms, all except M_* are independent to changes to the background by a flat offset in the limit of zero pixel size, due to the cancellation of terms in the calculation. Even outside this limit, this will tend to be the case on average due to the uniform random positioning of stars relative to the centres of the nearest pixels. For the quadrupole terms (but not the dipole terms), this is also true for changes to the background by a linear gradient. This means that even if the observation has a non-zero background in the region surrounding a star used for comparison with the model, all quantities except for M_* will be relatively insensitive to this effect. However, as M_* is sensitive to this, a measured difference in it between the model and observed stars may be due either to imperfect flat-fielding or background subtraction, or to the model itself being imperfect. Worse, it is possible for the effects of imperfect flat-fielding or background subtraction to cancel out the effects of model imperfections, falsely reassuring us that the model is adequate. Therefore, either we must ensure that any imperfections in subtracting off the sky background are unbiased with respect to the locations of the stars we select, or else we must test a different quantity in its place.

For our work here, we choose the latter approach, developing an alternative quantity which we will use in the place of M_*. The key requirement on this quantity is that it should in general be positively correlated with changes in M_* due to changes in the central flux distribution but be insensitive to the addition of a flat light distribution. As M_* can be seen as representative of the distribution's size, we look for an alternate size measurement which is background-independent.

To accomplish this goal, we generate a new size estimator M_s through the following algorithm:

(i) Bin all pixels in the image by distance from the centroid, in bins of size 1 pixel.

(ii) For each bin i except the innermost, determine $W_i = \overline{I_{<i}} - \overline{I_i}$, where $\overline{I_i}$ is the mean flux of all pixels in this bin, and $\overline{I_{<i}}$ is the mean flux of all pixels interior to it. For the innermost bin, $W_0 = 0$. These W are designed to be insensitive to the background level – a flat background added to the image will have no effect on the calculated W_i.[4] Due to the symmetry of this procedure, they will also be insensitive to any linear gradient in the background.

(iii) Determine the size measure $m_s' = \sum_i (d_i w(d_i) W_i) / \sum_i (w(d_i) W_i)$, where d_i is the distance from the centre of the image to the middle of each bin and $w(r)$ is the weight function used for other moments (which is limited to being circularly symmetric for this method).

(iv) Repeat the above for an image in which the flux of the central pixel is 1 and all other pixels are zero, to determine m_s^0. Determine the final size estimate $M_s = (m_s' - m_s^0)^2$. The subtraction here is performed to regularise M_s so that it converges to zero for objects of the minimum possible size, and it is squared so that it has equivalent units to M_*.

This size estimator is analysed in detail in Appendix A. Notably, we find that in the circumstances we intend to use it for, this estimator has the additional advantage that it is relatively less sensitive to noise than other commonly used size estimators.

We now have the parameters M_+, $M_\times$, M_s, M_x, and M_y, all of which we expect to be the same between an ideal model PSF and a selection of isolated stars, differing only due to noise. We will refer to this set of parameters collectively as M_k.

---

[4]This form was inspired by the formula for the signal from weak gravitational lensing around a mass distribution, the result of which is degenerate with the addition of a constant-value mass sheet to the field of view.

3.1.2 Combining weight functions

The parameters we developed in the previous section are sensitive to the choice of weight function used. If the weight function is strongly weighted towards the core of the PSF, these parameters will be relatively insensitive to imperfections outside the core, which might pose issues if galaxy ellipticities are ultimately measured with a weight function which is less weighted towards the core (with similar issues in the opposite scenario). Many shear-measurement algorithms use adaptive weight functions, which makes it impossible to know which weight function to use here. What we can do is to bracket the possible weight functions, using one weight function which is weighted to the core of the PSF, and another which is weighted to the wings. For our tests, we use the following weight functions:

$$
\text{Core}: w_c(R) = \begin{cases} \exp\left(-R^2/\left(2\sigma_c^2\right)\right) & R \le R_{\max}, \\ 0 & R > R_{\max}, \end{cases}
$$

$$
\text{Wings}: w_w(R) = \begin{cases} 1 & R \le R_{\max}, \\ 0 & R > R_{\max}, \end{cases} \tag{6}
$$

with $\sigma_c^2 = 0.15$ square arcseconds and $R_{\max} = 0\farcs5$. Note that the number of pixels within an annulus scales linearly with R, so for the wings weight function, the total influence of all pixels at radius R will also scale with R, up to the cut-off radius $R_{\max}$.

An ideal PSF model will have all five of our tested M_k parameters being on average the same for the models as for the star images for both the core and wings weight functions. However, the fact that these weight functions sample overlapping regions means that they are not independent of each other, and so we must take into account the resulting covariances of each M parameter for the two weight functions in our final statistic for the model. We can do this by defining and using the quantities '$Q_k^{(+)}$' and '$Q_k^{(-)}$', which are linear combinations of the M_k parameters as measured with the two weight functions:

$$
Q_k^{(+)} = \frac{1}{2\sqrt{2}}\left[M_k^{(c)}\left(1 + \frac{\sigma\left(M_k^{(w)}\right)}{\sigma\left(M_k^{(c)}\right)}\right) + M_k^{(w)}\left(1 + \frac{\sigma\left(M_k^{(c)}\right)}{\sigma\left(M_k^{(w)}\right)}\right)\right],
$$

$$
Q_k^{(-)} = \frac{1}{2\sqrt{2}}\left[M_k^{(c)}\left(1 + \frac{\sigma\left(M_k^{(w)}\right)}{\sigma\left(M_k^{(c)}\right)}\right) - M_k^{(w)}\left(1 + \frac{\sigma\left(M_k^{(c)}\right)}{\sigma\left(M_k^{(w)}\right)}\right)\right],
$$

$$\tag{7}$$

where $M_k^{(c)}$ is the M_k parameter using the core weighting function, $M_k^{(w)}$ is the M_k parameter using the wings weighting function, and $\sigma(M_k^{(c)})$ and $\sigma(M_k^{(w)})$ are the standard deviations of these quantities. The covariance of these two quantities can be shown to be zero, and in the scenario where $\sigma(M_k^{(c)}) = \sigma(M_k^{(w)})$, we will have the property $(Q_k^{(+)})^2 + (Q_k^{(-)})^2 = (M_k^{(c)})^2 + (M_k^{(w)})^2$, implying the $Q_k^{(\pm)}$ values will be of similar magnitude to the $M_k^{(c/w)}$ values.

Using these linear combinations has the additional benefit that it allows us to address the fact that the PSF centres are determined as the positions which make the dipoles zero for a certain weight function. If we were using a single weight function, it would be impossible to detect any differences between the dipole moments of the model and observed point sources, as these differences would be removed in the centring step. However, when we are using two weight functions, only the dipoles using one of them will be set to zero, and we can get information from the dipoles that use the other weight function.

If the weight function is flat near the centre (as is true of both tophat and Gaussian weight functions), then changes in the dipole moments will be linearly correlated with changes in the centre to a first-order approximation. Thus, if we shift the centre so that, for instance, $M_x^{(c)}$ and $M_y^{(c)}$ are made equal to zero, $M_x^{(w)}$ and $M_y^{(w)}$ will be reduced by

an amount proportional to the original $M_x^{(c)}$ and $M_y^{(c)}$. The final $M_x^{(w)}$ and $M_y^{(w)}$ values will then be analogous the differences between the values for each weight function. Looking at the linear combinations we presented above, we see that $Q_{x/y}^{(-)}$ also uses a difference between the moments from the two weight functions. We can therefore still use $Q_{x/y}^{(-)}$ for our analysis. This gives as a final set of eight values which provide a linearly independent basis: $Q_x^{(-)}$, $Q_y^{(-)}$, $Q_+^{(+)}$, $Q_+^{(-)}$, $Q_\times^{(+)}$, $Q_\times^{(-)}$, $Q_s^{(+)}$, and $Q_s^{(-)}$. In the following sections, we will use the shorthand 'Q_k' to refer to this particular set of parameters.

3.1.3 Quality of fit parameters

As many PSF models have one or more free parameters (the focus-secondary-mirror offset in the case of TINY TIM) which must be fit for each image, it is useful to determine a single value which can serve as a quality-of-fit metric. The focus parameter can then be fit by minimizing this value. A natural value to use for this purpose is the χ^2 value for the Q_k parameters, but this would require us to calculate a theoretical estimate of the errors for each Q_k value. This is not a trivial matter, as the determination of the centre of each star interacts with the calculation of the moments in a complicated manner. We therefore wish to use empirical estimates of the error in each parameter, calculating it from the scatter of the differences between the star and model values of that parameter for each image.

This, however, raises the issue that the resulting parameter from a χ^2-like calculation will not follow a standard χ^2 distribution, since the information in the scatter is reused both in the error calculation and in the χ^2 calculation. Additionally, a χ^2 value may not be optimal for fitting purposes: If a fit is imperfect, it indicates the statistical significance of a failure, rather than the practical significance of it. In our case, if we cannot fit perfectly, we would like to find the closest possibility, which may not be the same as that which minimizes the statistical significance of the failure.[5] We can calculate a value more indicative of the practical significance through the following process. We start by calculating Z_k^2 values through

$$
Z_x^{2(-)} = \frac{(0.71\,\text{px}^{-2})^2}{N}\sum_{i=0}^{N}\left(Q_{x,\text{star},i}^{(-)} - Q_{x,\text{model},i}^{(-)}\right)^4,
$$

$$
Z_y^{2(-)} = \frac{(0.71\,\text{px}^{-2})^2}{N}\sum_{i=0}^{N}\left(Q_{y,\text{star},i}^{(-)} - Q_{y,\text{model},i}^{(-)}\right)^4,
$$

$$
Z_+^{2(\pm)} = \frac{(0.14\,\text{px}^{-2})^2}{N}\sum_{i=0}^{N}\left(Q_{+,\text{star},i}^{(\pm)} - Q_{+,\text{model},i}^{(\pm)}\right)^2,
$$

$$
Z_\times^{2(\pm)} = \frac{(0.14\,\text{px}^{-2})^2}{N}\sum_{i=0}^{N}\left(Q_{\times,\text{star},i}^{(\pm)} - Q_{\times,\text{model},i}^{(\pm)}\right)^2,
$$

$$
Z_s^{2(\pm)} = \frac{(0.43\,\text{px}^{-2})^2}{N}\sum_{i=0}^{N}\left(Q_{s,\text{star},i}^{(\pm)} - Q_{s,\text{model},i}^{(\pm)}\right)^2. \tag{8}
$$

---

[5]Here, a notable difference between minimizing the statistical and practical significances of fitting failures arises due to our decomposition into the $Q_k^{(+)}$ and $Q_k^{(-)}$ values. As $Q_k^{(i)}$ represents a difference between highly correlated values, the statistical error on it is much smaller than the sum, $Q_k^{(+)}$. As such, if we were to take a χ^2-like value, weighting the differences from zero by their errors, a similar practical difference [e.g. (0.1,0.1) versus (0.1, −0.1)] will correspond to a much larger statistical significance if it corresponds to a non-zero $Q_k^{(-)}$ value than if it corresponds to a non-zero $Q_k^{(+)}$ value.

and a final fitting statistic:

$$X^2 = \sum_k Z_k^2, \tag{9}$$

summing over the set of eight Z_k^2 values. (Here, we use 'Z_k^2' as a shorthand for this set of eight parameters.) The formulae here are justified in Appendix B, but in summary:

(i) $Z_x^{2(-)}$ and $Z_y^{2(-)}$ approximate the worst-case scenario for the mean square of contributions to shear bias from centroiding issues.

(ii) $Z_+^{2(\pm)}$ approximates the mean square contribution of PSF shape inaccuracies to the first additive component of shear bias (c_1).

(iii) $Z_\times^{2(\pm)}$ approximates the mean square contribution of PSF shape inaccuracies to the second additive component of shear bias (c_2).

(iv) $Z_s^{2(\pm)}$ approximates the mean square contribution of PSF size inaccuracies to the both multiplicative components of shear bias ($m_1 + m_2$).

(v) X^2 approximates the mean square of total contribution to shear bias.

We can minimize X^2 to fit the optimal PSF model, but due to the presence of noise in the images, if we wish to judge whether or not a given image passes a test of the PSF model, we will need a baseline for its expected value and variance in ideal circumstances. We accomplish this by applying our testing procedure to a set of control images, as detailed in Section 4.

The choice of the value X^2 to minimize is admittedly arbitrary, and an argument could be made that it is better to weight the contributions to additive and multiplicative biases differently, given the different magnitudes of and requirements on them. However, it is first necessary to see if one or the other might be more problematic before choosing such weights, and so we use equal weights in our analysis here, and invite others to let our results inform how they might choose to weight these components.

It is also worth comparing to requirements that might be imposed on the PSF. For instance, *HST* images are planned to be used for validation of the processing pipeline for the *Euclid* mission (Laureijs et al. 2011). This mission imposes the following shear bias requirements on its full pipeline:

$$m_1, m_2 \leq 2 \times 10^{-3},$$
$$c_1, c_2 \leq 5 \times 10^{-5}. \tag{10}$$

The requirements on *HST* images for validation of the *Euclid* pipeline are likely to be much less strict, but as these requirements have not yet been calculated, we will use the above values as example requirements for our work here. This will allow us to judge, for instance, whether the multiplicative or additive requirements on shear bias are more difficult for PSF models to meet.

These shear bias requirements will correspond to requirements on the various Z_k^2 values. However, in order to test against these requirements, we must first account for the fact that noise in observations translates to non-zero typical Z_k^2 values (which are in effect scaled mean square deviates) even for an ideal PSF model. If we assume that the Z_k^2 values we calculate for the control fields in the ideal scenario, which we will here label as $Z_{k,\text{Ideal}}^2$, then we can consider the Z_k^2 values for observations to be the sum of this, a contribution from excess variance (which might, for instance, be caused by the model not fully capturing the spatial dependence of the PSF) and a contribution from a difference in the mean of the measured moments between the modelled and observed PSFs. That is,

$$Z_k^2 = Z_{k,\text{Ideal}}^2 + Z_{k,\text{ExcessVariance}}^2 + Z_{k,\text{MeanDeviate}}^2. \tag{11}$$

It is only the latter two terms which we want to test against requirements on the PSF, which we can do by imposing the requirements on the difference between the measured Z_k^2 values and those measured for the control images, $Z_{k,\text{Ideal}}^2$.

Additionally, if we wish to ensure that we meet requirements to a given threshold of certainty (for instance, 95 per cent), we must require that this difference is below the required value by at least $1.645\sigma\left(Z_k^2\right)$, where the factor 1.645 is the z score that corresponds to a one-sided p-value of 0.95, and $\sigma\left(Z_k^2\right)$ is the standard deviation of this Z_k^2 value, as measured from tests on control images. This gives us the requirements:

$$Z_+^{2(\pm)} - Z_{+,\text{Ideal}}^{2(\pm)} \lesssim 2.5 \times 10^{-9} - 1.645\sigma\left(Z_+^{2(\pm)}\right),$$
$$Z_\times^{2(\pm)} - Z_{\times,\text{Ideal}}^{2(\pm)} \lesssim 2.5 \times 10^{-9} - 1.645\sigma\left(Z_\times^{2(\pm)}\right),$$
$$Z_s^{2(\pm)} - Z_{s,\text{Ideal}}^{2(\pm)} \lesssim 4 \times 10^{-6} - 1.645\sigma\left(Z_s^{2(\pm)}\right). \tag{12}$$

Here, we assume that centroiding is handled carefully so that even if issues are present with the x and y moments of the PSF, they do not impact shear estimation. Note that these requirements are approximate due to the assumptions we made in our calculations of the weighting factors for the Z_k^2, which assume a typical galaxy size of $1.5\times$ the size of the PSF, and the specifics of the validation procedure used might require more or less stringent requirements.

3.2 Implementing tests

We implement our tests on each image independently. For each image, we start by using the SEXTRACTOR utility (Bertin & Arnouts 1996) to identify all candidate objects in the image, using the image's exposure time to set the proper zero-point for it in the configuration. To form our star sample, we then impose the following cuts on the generated catalogue:

(i) CLASS_STAR $\geq$0.99
(ii) 22.4 $\leq$ MAG_AUTO $\leq$25.4 (approximately signal-to-noise ratio 20–200)
(iii) FLUX_AUTO/FLUXERR_AUTO $\geq$50

This ensures that all objects in the sample are likely stars and are detected with enough significance to be useful for our tests. The lower limit on the magnitude is used to prune stars that are possibly bright enough to show non-linear effects or saturate their central pixels. For each star in the sample, we then determine the distance to the nearest other object (star or otherwise) in the catalogue. We remove any stars from the sample for which the nearest neighbour is within $2R_{\max}$ (1″) of it or the edge of the image is within $R_{\max}$ of it to form a sample of isolated stars. It is still possible that some of these stars might neighbour a faint object which was not identified by SEXTRACTOR or that they might in fact be a binary system, so it will be necessary later to check the sample for outliers.

For each of the isolated stars in our sample, we first determine the ideal centre position (x_c, y_c) for it as the position for which $M_x^c = M_y^c = 0$ (where the c superscript represents calculation using the core weight function). To do this, we start with the centre provided by SEXTRACTOR and calculate M_x^c and M_y^c. We then shift the centre positions by these values:

$$x_{c,\text{new}} = x_{c,\text{old}} + M_x^{(c)},$$
$$y_{c,\text{new}} = y_{c,\text{old}} + M_y^{(c)}. \tag{13}$$

This process is iterated until the convergence of both x_c and y_c. Using this centre, we then calculate the moments and Q_k values for the star, as detailed in Sections 3.1.1 and 3.1.2.

We now proceed to determine the best-fitting focus value for each image by fitting for the focus value that minimizes X^2 (as defined in equation 9) for the image. The calculation of X^2 for each image is detailed in Section 3.1.3, except for the determination of outliers. To do this, we first calculate all $M_k^{(c)}$ and $M_k^{(w)}$ values (i.e. M_+ etc. calculated with each of the core and wings weight functions) except for $M_x^{(c)}$ and $M_y^{(c)}$ (which will be zero due to the centre fitting) for each star, assuming a focus offset value of -1 µm, which is at the centre of the observed range of fit values, -8 to $+6$ µm. For each of these parameters, we then calculate the mean and standard deviation of the differences between the measured values for the stars and their corresponding models.

Since we expect that there will be some contamination of our sample due to objects misclassified as stars, binary star systems, and other blends, we apply Chauvenet's criterion to each parameter, marking as outliers for this parameter stars where:

$$\frac{\left|(M_{k,\text{star},i} - M_{k,\text{model},i}) - \overline{(M_{k,\text{star}} - M_{k,\text{model}})}\right|}{\sigma(M_{k,\text{star}} - M_{k,\text{model}})} > D_{\max}(N), \quad (14)$$

where $D_{\max}(N)$ is chosen such that for N realizations of a normal distribution, there is a less than 50 per cent chance that any value will be this far or farther from the mean, and N is the initial number of stars. This process is iterated, updating the standard deviation and mean (but not N) until the sample has converged. A star will be considered an outlier if any of its Q^c or Q^w parameters is marked as an outlier.[6]

We use the samples of non-outlier stars to calculate the X^2 value for each image, and fit for the focus that provides the minimum such value. Since we are only fitting a single parameter, we apply a simple brute-force procedure, sampling focus values between -6 and 6 µm at intervals of 1 µm. We take the best of these values, and then use a downhill simplex method to determine the best focus to within 0.1 µm. This gives us the best-fitting focus value, and we use the statistics this value provides for our analysis.

4 CONTROL TESTS

It is important to check our testing procedure against control images for various reasons. It allows us to validate that it works and has no apparent bugs, it informs us of the ideal fitting statistic X^2 for various scenarios, and it allows us to perform convergence tests.

4.1 Control image design

Our control images are generated to match the properties which are shared by all our test images:

(i) Dimensions: 4096×2048 pixels
(ii) Pixel scale: 0.05 arcsec px^{-1}
(iii) Gain: $2.0e^{(-)}$/ADU
(iv) Instrumental zero-point[7]: 26.50

We also make the following choices to make the image representative of one of the exposures in our sample, using the values from one of the exposures of NGC-0104, labelled as `jb6v09shq` in the archive:

(i) Exposure time: 1298 sec
(ii) Pixel noise: Approximated by Gaussian distribution with $\sigma = 51e^{(-)}$ px^{-1}. This is a conservative overestimate of the noise to account for the contributions of the wings of bright stars in addition to the image background and read noise.
(iii) Chip: 1

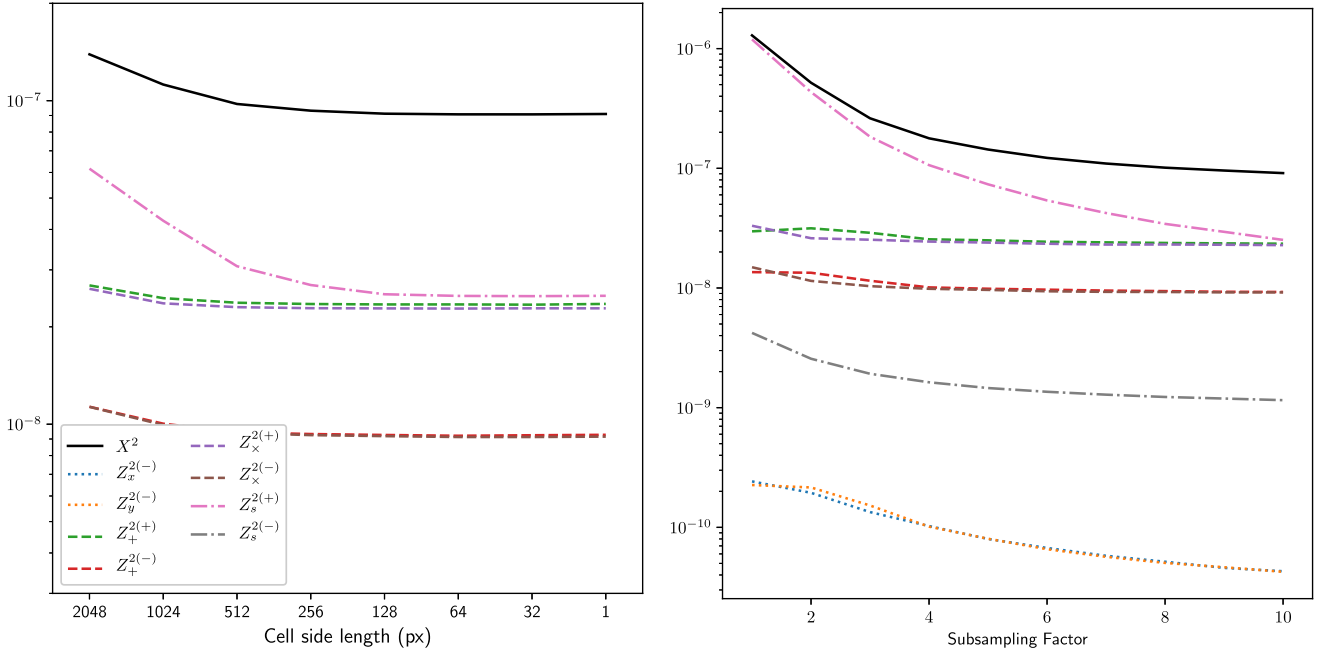We generate a set of 100 images, with focus offset values distributed uniformly between -6.0 and 4.0, which covers the typical range of values expected in observations. For each image, we generate 1000 mock stars, with positions drawn from a uniform random distribution on the image and magnitudes drawn from a uniform random distribution between 22.4 and 25.4, and with appropriate shot noise applied to the images. TINY TIM is called to generate a PSF model for each star (using its actual position, rather than the nearest grid point as done elsewhere in our analysis), subsampled at a factor of $10\times$, which is the maximum factor allowed by Tiny Tim. Unlike in our testing procedure, where we rebin models through simple summation after shifting, here we use the GalSim toolkit (Rowe et al. 2015) to interpolate and integrate the models. This is much slower, but it allows for more precise determinations of the rebinned PSF models. We will later test the accuracy of the faster simple summation approach for various subsampling factors in Section 4.2. By default, each star is generated using the K7V spectrum in TINY TIM.

We generate a set of variant control images, each accounting for different effects which would be difficult or time-consuming to account for in our testing procedure, to estimate the impact these effects would have on our testing procedure. We limit our analysis to representative cases of different effects, using the following image variants, with random seeding used to maintain the positions of stars between variants:

(i) Base: All default options, as described above.
(ii) Binaries: For a randomly-selected 30 per cent of stars, an additional mock star is added to the image at a random position within a circle of radius 1 px around the star's position. The additional star's magnitude is drawn from the same distribution as the other stars.
(iii) Wide binaries: As Binaries, except in all cases where a star is selected to be a binary, the additional mock stars are added to the image within circles of radius 2 px around the stars' positions instead of circles of radius 1 px.
(iv) 1D guiding error: To simulate the effects of guiding error (see Lucas et al. 2018, section 5.2.3), each model PSF is convolved with a tophat profile of length 0.2 px in the x-dimension and length 0 in the y-dimension.
(v) 2D guiding error: Differential velocity aberration causes both an elongation of images similar to that of guiding error and a scale change (Pirzkal et al. 2001). The latter is small enough to be negligible, but the former is of concern. We simulate this in the extreme case where this effect is orthogonal to the elongation caused by guiding error by convolving each PSF model with a tophat profile of length 0.2 px in both the x- and y-dimensions.
(vi) Galaxy background: To test the impact of possibly unresolved blends, we add to the image a background consisting of a randomly generated galaxy field, with size and magnitude distributions designed to approximate what would be observed in the F606W filter. Galaxies with magnitude $\lesssim 28$ are rendered.
(vii) Varying spectral type: Rather than using the K7V spectrum for all stars, stars are generated with spectral types drawn from a random distribution. This distribution is a simple model that allows all types but is strongly weighted towards redder types, with $P(i) \propto i^3$,

Figure 1. Convergence tests for the PSF gridding scheme (left) and subsampling factor (right). In each plot, the vertical axis shows the mean quality of fit statistics for a test of control fields using a given set-up, where the solid black line shows X^2, the sum of the other statistics. In the left plot, the horizontal axis shows the side-length in pixels of each cell in the PSF gridding scheme, and in the right plot, it shows the subsampling factor used. Note that in the left plot, many lines are not shown as scaling the plot to show them would result in making it difficult to judge the slope of the X^2 line, which is the most important result here.

where i is the index of the spectral type as provided to TINY TIM. For reference, the default K7V spectral type is index 15 out of 17 total.

(viii) Full: This variant includes the combination of effects from the binaries, 2D guiding error, galaxy background, and varying spectral-type variants.

Two additional possible complicating factors are jitter and imperfect correction for charge transfer inefficiency ('CTI', see Massey et al. 2014). Jitter results in a blurring of scale $\lesssim 7 \times 10^{-3\prime\prime}$ (Clapp 2009), or 0.14 px, which is subdominant to the guiding error and differential velocity aberration effects and not worth testing separately. CTI manifests as a blurring of images in the readout dimension. Aside from the fact that CTI is a non-linear transformation, this is similar to the convolution we use for our Guiding Error variants, and so it is not necessary to use separate variants for imperfect CTI correction.

4.2 Convergence tests

In Fig. 1, we plot the results of our convergence tests. In the left plot, we test the gridding scheme for PSF models, where models are generated only for a fixed number of points at the centres of grid cells, and these models are used for all positions within their cells. This can greatly save time in testing if multiple stars reside in the same cell (even on different images), as fewer PSF models will need to be generated, but we must assess the impact of this approach to ensure that it does not significantly impact the results of our testing framework. We thus test the framework on our set of control images, using the known focus offset value for each and a set of cell sizes from 2048 px² (only two cells per image) to 1 px². For all cases, we use the maximum possible subsampling factor of 10 to ensure that no additional errors are introduced from a lack of convergence for that variable.

We see from the plot that if the gridding scheme is too coarse, X^2 can be increased by up to a factor of ~ 1.5. A cell size of 512 px²

provides a reasonable balance of time saved versus accuracy, but for our purposes we decide to use a size of 128 px². By this point, the accuracy has definitely converged, and it will likely still have converged if accuracy is improved (e.g. by increased exposure time). This corresponds to 512 cells per image, which means few stars on the same image will share a cell. It still allows for significant time savings if models are reused for different images, though.

In the right-hand panel of Fig. 1, we test the subsampling factor used for the generated PSF models before they are shifted and rebinned to match the positions of stars. In this case, the effect on X^2 is more drastic, increasing it by up to an order of magnitude when no subsampling is used. Here, a subsampling factor of 4 provides a good balance between time and accuracy. We choose to use a subsampling factor of 8 for the same reasoning as before; it has converged by this point, and likely still will have if the accuracy is improved.

To ensure high accuracy for our tests, we choose a gridding scheme with cell sizes of 128 px² and a subsampling factor of 8.

4.3 Expected fitting statistics

In order to determine the expected fitting statistic X^2 for an ideal scenario, we apply our testing framework to each control field, using the converged grid scheme and subsampling factor determined in Section 4.2. We test both using the known focus offset value and fitting for the best value, and we plot the resulting X^2 and Z_k^2 values in Table 1.

Let us first compare the fitting statistics from using the known focus offset versus fitting it, to understand the impact of the fitting procedure on the results. In the case where we use the known focus offset, the size-related parameter $Z_s^{2(+)}$ is the largest contributor to X^2 in all scenarios, but particularly when unresolved binary stars are included. When the focus offset is instead fit, the magnitude of $Z_s^{2(+)}$ tends to decrease slightly, while other statistics slightly rise, resulting

Table 1. The fitting statistics resulting from applying our testing procedure to each control image variant. The details of the variants are listed in Section 4.1, and the fitting statistics are defined in Section 3.1. The top table lists the fitting statistics resulting from a test in which the known focus offset for each control image is used, the middle table lists the statistics when the focus offset is fit for each image, and the bottom table lists the statistics when all possible optical parameters are fit for each image.

Type	X^2	$Z_x^{2(-)}$	$Z_y^{2(-)}$	$Z_+^{2(+)}$	$Z_+^{2(-)}$	$Z_\times^{2(+)}$	$Z_\times^{2(-)}$	$Z_s^{2(+)}$	$Z_s^{2(-)}$
				Known focus offset					
Base	1.0×10^{-7}	5.1×10^{-11}	5.1×10^{-11}	2.4×10^{-8}	9.5×10^{-9}	2.3×10^{-8}	9.2×10^{-9}	3.5×10^{-8}	1.2×10^{-9}
Binaries	1.6×10^{-7}	3.7×10^{-11}	3.9×10^{-11}	2.0×10^{-8}	8.6×10^{-9}	1.9×10^{-8}	8.6×10^{-9}	9.9×10^{-8}	1.5×10^{-9}
Wide binaries	1.7×10^{-7}	4.2×10^{-11}	3.9×10^{-11}	2.3×10^{-8}	9.9×10^{-9}	2.2×10^{-8}	9.7×10^{-9}	1.1×10^{-7}	2.0×10^{-9}
1D guiding error	1.0×10^{-7}	5.1×10^{-11}	5.1×10^{-11}	2.4×10^{-8}	9.5×10^{-9}	2.3×10^{-8}	9.2×10^{-9}	3.5×10^{-8}	1.2×10^{-9}
2D guiding error	1.0×10^{-7}	5.1×10^{-11}	5.0×10^{-11}	2.4×10^{-8}	9.6×10^{-9}	2.3×10^{-8}	9.5×10^{-9}	3.5×10^{-8}	1.2×10^{-9}
Galaxy Background	1.0×10^{-7}	5.7×10^{-11}	5.8×10^{-11}	2.4×10^{-8}	9.6×10^{-9}	2.3×10^{-8}	9.2×10^{-9}	3.4×10^{-8}	1.2×10^{-9}
Varying spec. type	1.0×10^{-7}	5.1×10^{-11}	5.0×10^{-11}	2.4×10^{-8}	9.6×10^{-9}	2.3×10^{-8}	9.4×10^{-9}	3.5×10^{-8}	1.3×10^{-9}
Full	1.6×10^{-7}	4.2×10^{-11}	4.4×10^{-11}	2.0×10^{-8}	8.9×10^{-9}	2.0×10^{-8}	8.6×10^{-9}	1.1×10^{-7}	1.6×10^{-9}
				Fit focus offset					
Base	1.1×10^{-7}	9.9×10^{-11}	1.0×10^{-10}	2.6×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	9.6×10^{-9}	3.9×10^{-8}	1.3×10^{-9}
Binaries	1.6×10^{-7}	6.9×10^{-11}	7.1×10^{-11}	2.2×10^{-8}	9.9×10^{-9}	2.0×10^{-8}	8.9×10^{-9}	9.5×10^{-8}	1.7×10^{-9}
Wide binaries	1.8×10^{-7}	7.0×10^{-11}	7.0×10^{-11}	2.4×10^{-8}	1.1×10^{-8}	2.3×10^{-8}	1.0×10^{-8}	1.1×10^{-7}	2.4×10^{-9}
1D guiding error	1.1×10^{-7}	1.0×10^{-10}	1.1×10^{-10}	2.5×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	9.5×10^{-9}	3.9×10^{-8}	1.3×10^{-9}
2D guiding error	1.1×10^{-7}	1.0×10^{-10}	1.0×10^{-10}	2.5×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	9.8×10^{-9}	3.9×10^{-8}	1.3×10^{-9}
Galaxy Background	1.1×10^{-7}	1.1×10^{-10}	1.1×10^{-10}	2.6×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	9.4×10^{-9}	3.8×10^{-8}	1.3×10^{-9}
Varying spec. Type	1.1×10^{-7}	9.3×10^{-11}	1.0×10^{-10}	2.5×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	9.6×10^{-9}	3.9×10^{-8}	1.3×10^{-9}
Full	1.6×10^{-7}	7.7×10^{-11}	8.9×10^{-11}	2.3×10^{-8}	1.1×10^{-8}	2.1×10^{-8}	9.1×10^{-9}	9.6×10^{-8}	1.9×10^{-9}
				Fit all parameters					
Base	1.1×10^{-7}	9.3×10^{-11}	7.8×10^{-11}	2.6×10^{-8}	1.1×10^{-8}	2.4×10^{-8}	1.0×10^{-8}	3.6×10^{-8}	1.4×10^{-9}
Full	1.5×10^{-7}	8.0×10^{-11}	7.7×10^{-11}	2.2×10^{-8}	1.0×10^{-8}	2.1×10^{-8}	9.2×10^{-9}	8.9×10^{-8}	2.5×10^{-9}

in only a slight decrease in the total X^2. This is likely due to the fact that modifying the focus offset has the most significant impact on the size of the PSF model, and since this is already the largest contributor to X^2 – even when the observed size is incorrect (due to noise and possibly other factors) – X^2 can be improved by fitting a different focus offset. We will keep this effect in mind in our analysis.

In Fig. 2, we plot the fitted focus offset for each control image against the actual value for each control variant. We see that when unresolved binary stars are not included, the fitted value is very close to the actual value, but when binaries are included, the fitted values become more extreme, biased to either side of ~ -3. This is due to the minimum PSF size resulting from this value, and binaries biasing the fit to prefer larger model PSFs. Since larger PSFs result from focus offsets further from ~ -3, the fitted values are thus biased away from it, resulting in a bifurcation in the fitted values.

As the fit values are what we will obtain with observed images, we focus the remainder of our discussion on the values in that table. Looking first at the base image variants, which include no extra complicating factors, we see that the quantities that impact additive shear bias, $Z_+^{2(\pm)}$ and $Z_\times^{2(\pm)}$ are all significantly above the threshold to determine if the model PSFs meet the example requirements we have imposed, while those that impact multiplicative bias, $Z_s^{2(\pm)}$, are well below the threshold. The standard deviations of these values (not shown in the table) are as follows:

$$\sigma\left(Z_+^{2(+)}\right) = 3.2 \times 10^{-9},$$
$$\sigma\left(Z_+^{2(-)}\right) = 1.6 \times 10^{-9},$$
$$\sigma\left(Z_\times^{2(+)}\right) = 3.2 \times 10^{-9},$$
$$\sigma\left(Z_\times^{2(-)}\right) = 1.6 \times 10^{-9},$$
$$\sigma\left(Z_s^{2(+)}\right) = 7.2 \times 10^{-9},$$
$$\sigma\left(Z_s^{2(-)}\right) = 2.2 \times 10^{-10}. \tag{15}$$
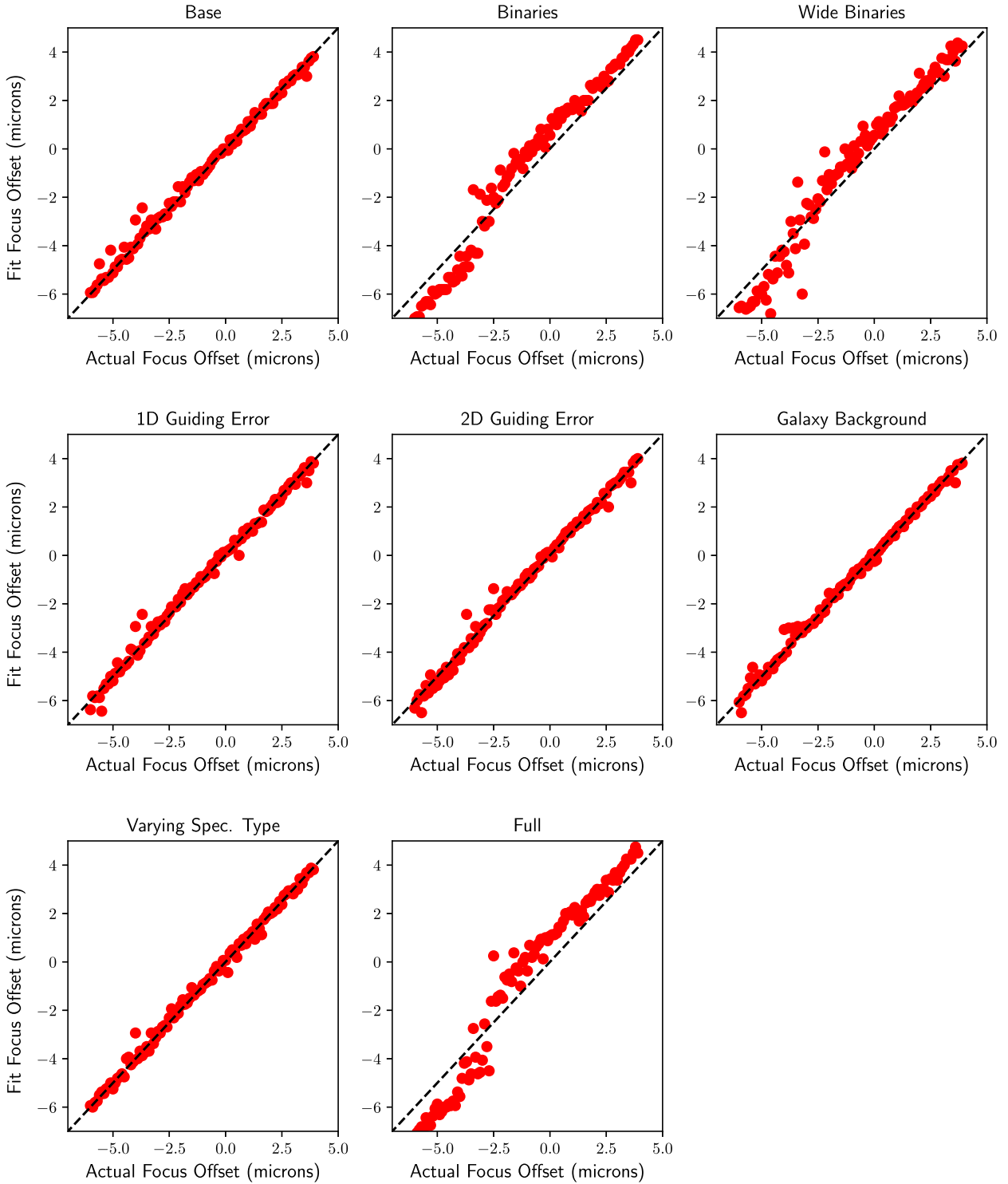
Unfortunately, the standard deviations for the factors that contribute to additive bias on shear measurements are all sufficiently large they result in the threshold to confidently state that a PSF model meets this requirement being negative. This means that in practice we will not be able to confidently make this claim, although we may be able to confidently claim that a PSF model fails to meet this requirement. For the case of multiplicative bias, the standard deviation and ideal values of the contributing factors are sufficiently low that we can say that a model PSF confidently meets our multiplicative bias requirements if

$$Z_s^{2(+)} \lesssim 4 \times 10^{-6}, \text{ and}$$
$$Z_s^{2(-)} \lesssim 4 \times 10^{-6}. \tag{16}$$

If it is in fact necessary to ensure the PSF model meets the example requirements for additive bias that we are using, testing only the brightest stars in an image will make this test more feasible mathematically, as it will reduce the impact of noise on the Z_k^2 values, but this will come at the expense of worse sampling of the PSF across the image. Taking deeper exposures can additionally help mitigate this problem. However, this applies only to the ideal scenario, which assumes that various other complicating factors are properly handled. It also assumes that there is no loss of data due to bright stars becoming oversaturated and useless, or if there is, it is counteracted by fainter stars becoming usable.

The remaining rows in Table 1 show the representative impact of these effects, which we will now discuss.

The most significant complicating factor is the presence of unresolved binaries in the sample of stars. Although our testing procedure includes both a detection step (with cuts on SEXTRACTOR's CLASS_STAR parameter and object size) and an outlier-rejection step, the presence of binary stars still increases the fitting statistic significantly, to $\sim 1.6 - 1.8 \times 10^{-7}$. This error cannot be reduced through longer exposures, but instead will require binary stars to

Figure 2. The fit focus offset plotted against the actual focus offset for each control image. Each panel shows a different control variant, including different complicating factors. The dashed line indicates an ideal fit, where all fitted focus offsets match the input values.

be identified and excluded from the sample, for instance through analyses of their spectra as measured by supplementary observations. Interestingly, there is only a small difference in the fitting statistic when the width of binaries is increased, which is likely due to wider binaries being more likely to be identified in the outlier-rejection process and excluded from the sample.

All other factors tested (the presence of a background of resolved and unresolved galaxies, guiding error, and varying spectral type) have no significant impact on the fitting statistic.

For our analysis in this paper, we take the approach of ignoring these effects in our testing procedure and comparing the resulting fitting statistics to those for the Full variant.

5 RESULTS AND ANALYSIS

Having confirmed that our testing procedure works on control images and obtained data on the expected fitting statistics for an ideal PSF model in ideal and more realistic situations, we now move on to testing TINY TIM PSFs on the *HST* observations we introduced in Section 2.1. In Section 5.1, we present the results of testing TINY TIM with its default configurations, and we test fitting more advanced configurations in Section 5.2. We compare our results against di Nino et al. (2008) and Niemi & Lallo (2010)'s model for *HST*'s focus offset in Section 5.3, and discuss our results in Section 5.4.

5.1 Testing results

We can judge the quality of model PSFs by looking at the X^2 and Z_k^2 statistics of our tests, as defined in Section 3.1.3. For an ideal model in ideal circumstances, we would expect these values to be clustered around those found for the Base control images, as seen in the first row of Table 1. As multiple factors which we have not accounted for can complicate the analysis, even an ideal model will likely have larger quality-of-fit statistics. The final row of Table 1 presents a reasonable estimate of the upper bound, using representative values for various possible complicating factors, and so comparisons against these values will present a more conservative test of the PSF models.

We plot histograms of our fitting statistics for all tested images after fitting the best focus offset for each in Fig. 3, and compare them to the expected values from the Base and Full control images, shown as the dashed and dotted black lines, respectively. The corresponding grey lines show these values plus twice the standard deviation among all control fields, and so represents the approximate maximum values we would expect to see if the PSF models are ideal. If we look just at the total statistic, X^2, the values for the tested images are clustered significantly above the expected value from the Full control, which implies either an issue with the model or that our control underestimates the impact of complicating factors by nearly an order of magnitude.

This is similarly the case when we look at the Z_k^2 values that together compose X^2. Of these values, all are on average significantly above the expectations for the Full control. In the worst case, for $Z_y^{2(-)}$, the resulting statistics are on average nearly 1.5 orders of magnitude higher than the expectation from the control images.

As the $Z_+^{2(+/-)}$ and $Z_\times^{2(+/-)}$ values generally greatly exceed those values for the control images, we can state that our example additive bias requirement ($c < 5 \times 10^{-5}$) will not be met by this PSF model. The multiplicative bias requirement will be met however, as although the $Z_s^{2(+/-)}$ values generally exceed the control values, it is not by nearly enough to violate our requirement of $m < 2 \times 10^{-3}$.

We can gain further insight into the issue here by looking at the relationships between the quality of fit parameter X^2 and other parameters. We show this in Fig. 4, where we plot the relationships between X^2 and the best-fitting focus value, the chip with which each image was observed, the observation date, and the number of stars in each image. There is no apparent relationship between X^2 and the observation date, suggesting that there is no time-dependent effect in this period affecting the PSF (at least to our level of sensitivity). Similarly, if our exclusion of objects with nearby neighbours were not sufficient to deal with crowding, we would expect to see a trend between the number of stars in each image and X^2, and we see none. This suggests that our exclusion is indeed sufficient, at least for detected objects (and for undetected objects, the control tests outlined in Section 4.3 show that this effect is minimal).

However, we do see notable oddities when we compare X^2 to the best-fitting focus offset value and the chip with which each image was observed. We see that the focus offset values are clustered between 0 and 6 μm, where we expect from the *HST*'s breathing that they will span the range of −6 to 6 μm. The chip with which each image was observed also seems to play a role – images observed with chip 2 systematically have lower best-fitting focus offset values by roughly a micron, and larger X^2.

One might consider that this is in part due to the fact that the fitting algorithm prioritizes fitting the size of the PSF model, as it is the largest contributor to X^2. Since we cannot efficiently remove all unresolved binary stars from the sample, these bias the model to a focus offset value which provides a larger size. It seems that the fitting algorithm generally finds the best solution to this at larger focus offset values. However, this explanation is not fully consistent with our tests on control fields, as although the fitting does prefer more extreme focus offset values, as seen in Fig. 2, it is split between high and low values, with the mean bias actually being to a too-low focus offset. One potential explanation to this is that the symmetry observed in the control fields only holds for small deviations between the model and observed PSFs, and large deviations result in this symmetry breaking and appearing as we observe here.
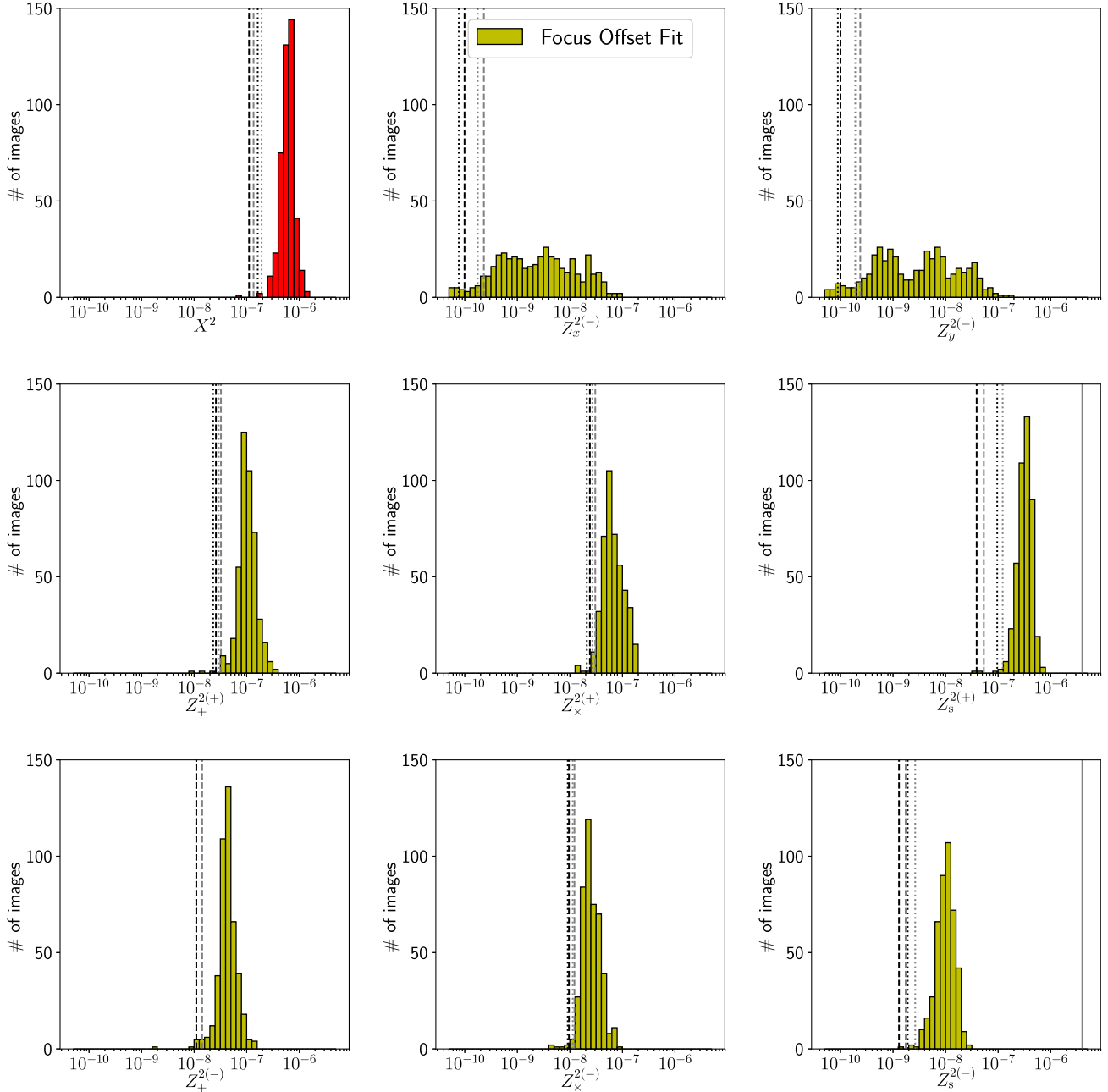
Further analysis of the data shows that in the control fields, the model PSFs fail to match the sizes of the observed PSFs, but differences between the model and observed sizes are tightly clustered, while in the actual fields, the model PSFs match the sizes of the observed PSFs on average, but there is very large scatter in this relationship. This suggests that a possible explanation for this discrepancy might be that there is more spatial variation in the PSFs than is accounted for in the model. Inspection of the data shows that while there is a statistically significant correlation between position and size difference for many images, the nature of this correlation is not consistent between images. This could be due to temperature variations and gradients distorting the image plane in ways that are not accounted for by the TINY TIM model.

The difference between the two chips might be considered to be due to the fact that there is a vertical offset between them. However, TINY TIM already does implement a correction for this effect. Even if it didn't, a focus offset difference between the two chips of just one micron, as we tend to see here, would correspond to a height difference of hundreds of microns in the chips (Cox & Niemi 2011), which is implausibly large. This effect was also noticed by Cox & Niemi (2011), who attributed it to most likely being due to differences in spherical aberration and charge diffusion between the two chips. As we will discuss further in Section 5.2, we find this to be an insufficient explanation.

It seems that while the impact of unresolved binaries or some similar effect is responsible for some of the difference between our expected and measured X^2, this cannot explain all of it. Many of the fitting statistics are significantly worse in the real images than in our control tests, and this cannot be explained by any of the possible complicating factors we tested. The most likely conclusion at this point is that the issue lies with the model PSFs. In the following section, we will test options for improving the models.

5.2 Advanced configuration of TINY TIM

As we found in the previous section, the TINY TIM PSF models we tested seem to fail to adequately characterize the observed PSFs, even when accounting for complicating factors that might result in apparently poor fits. However, the models we tested were limited to

Figure 3. Distributions of the quality of fit parameter X^2 across all tested images, as the contributions of each tested moment, presented as Z_k^2 (see equation 8). The expected values for an ideal fit for each parameter are shown with the dashed black lines, and the expected values allowing for complicating factors are shown with the dotted black lines. The grey lines show these values plus twice their respective standard deviations among all control fields. The solid grey line, for the $Z_s^{2(+/-)}$ panels, indicates the threshold for these values such that the PSF model will meet our example requirements on multiplicative bias $m < 2 \times 10^{-3}$. Our example requirements on additive bias are failed if the $Z_+^{2(+/-)}$ or $Z_\times^{2(+/-)}$ values exceed the control values.

the default configuration for TINY TIM. It is also possible to modify various parameters of the model, notably the Zernike polynomial coefficients which characterize the optics, but these options are not normally presented to the user. Given the discrepancies we have found between the models and reality, it is reasonable to consider that the default values for these parameters might be incorrect, at least for the observations we tested.

To test this hypothesis, we repeat our testing procedure, this time fitting each image not only for the focus offset (which in fact corresponds to the fourth Zernike polynomial's coefficient) but also

for the coefficients of all polynomials Z_2 through Z_{21}. Only the focus offset is expected to vary from image to image, so ideally the other fitted parameters will be consistent across all images, or else vary only with the date of observation (which would imply some factor changed the optics over time).

In addition to these, the possibility was also raised in private correspondence with TINY TIM's author, John Krist, that the charge diffusion kernel determined by TINY TIM is only an approximation and might not be accurate enough. As the possible variations to the kernel are infinite, we will test only one modification which

Figure 4. Left: The quality of fit parameter X^2 for each image plotted against the best-fitting focus value, with points coloured according to which of the two detector chips the image was taken with. Right: The quality of fit parameter χ^2 for each image plotted against the date of observation, with points coloured according to the number of non-outlier isolated stars in the image.

will roughly characterize whether too much or too little intensity is diffused. We introduce a parameter c, and the eight non-central pixels of the kernel are then multiplied by c. The central pixel is then rescaled so that the kernel sums to 1. A value of $c > 1$ then corresponds to more diffusion, and a value $c < 1$ corresponds to less diffusion. If there is an indication that the default value $c = 1$ is not appropriate, further investigation can be undertaken to determine a better model for charge diffusion.

Finally, as we discussed in Section 4.3, it is possible to include the effects of guiding error and similar effects in the PSF model. This can be roughly approximated by convolving the PSF with a rectangular tophat profile, which will have free parameters representing its height, width, and orientation. However, this has the drawback that including this effect will necessitate a much greater amount of computer time: In addition to adding three more parameters which must be fit, this convolution would require the use of an interpolation and integration approach to rebinning PSF models, rather than simple summation. As this can result in up to double the total time required, and since the effect of guiding error is dwarfed by the effect of unresolved binary stars in the sample, we choose not to include this effect in our analysis.

We perform our fitting procedure using a two-stage approach. First, the focus offset is fit using a brute-force followed by steepest-descent algorithm, while all other parameters are held constant. Secondly, a steepest-descent algorithm is used to fit all parameters. This is done to help avoid the fitting algorithm getting trapped in a local minimum. Since the focus offset is the most significant factor, fitting it first ensures that this approach will provide results at least as good as the focus-only approach, and will not get trapped in a local minimum far from the global minimum.
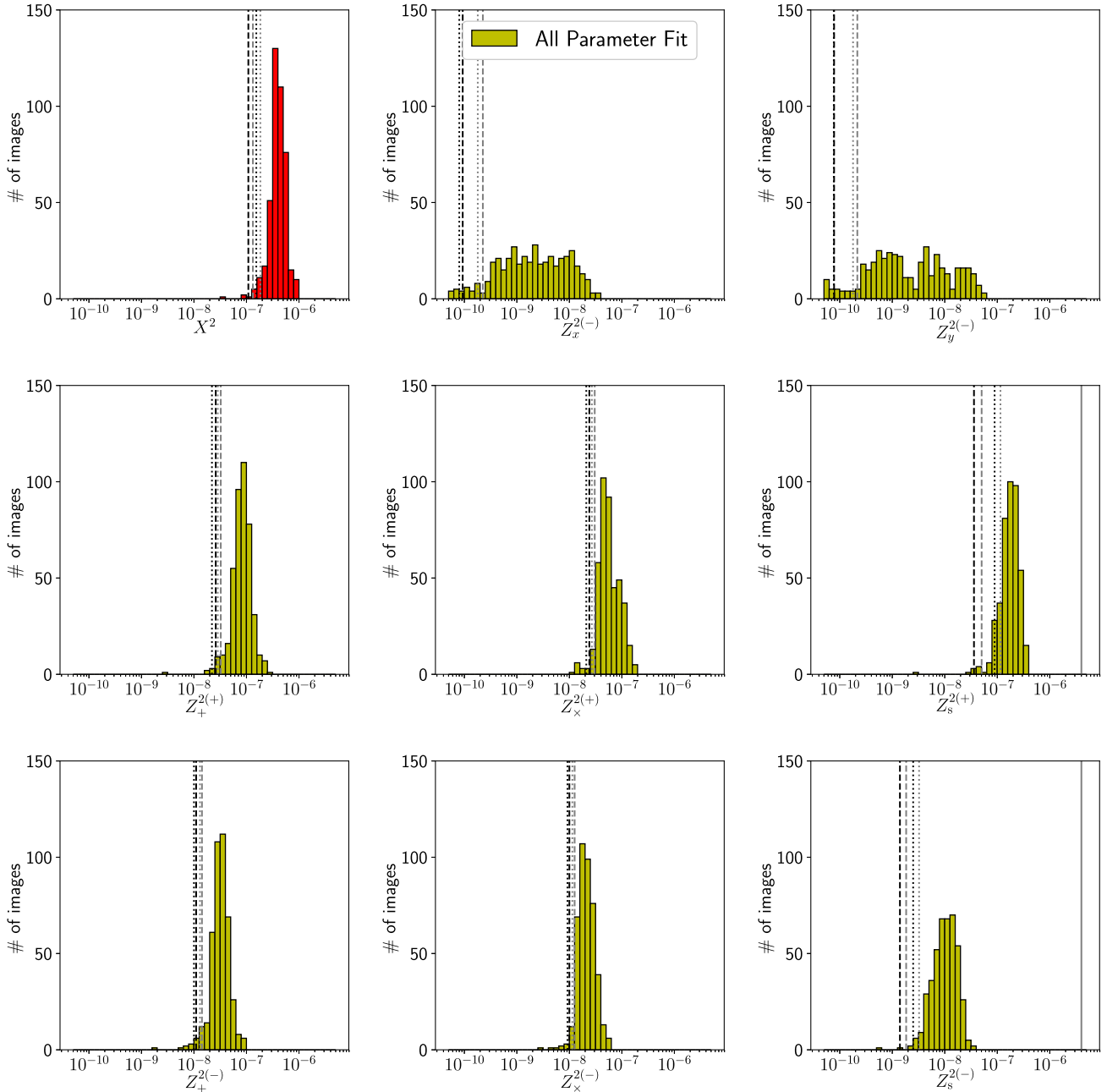
In Fig. 5, we plot the resulting fitting statistics when all optical parameters are fit for each image. This shows a small but noticeable improvement over the case where only the focus offset is fitted, and the overall statistic, X^2, lies closer to the mean value found when testing on representative control images. The fact that we still do not

reach the expected quality of fit from the control tests might be due to our control images not properly accounting for some physical effect which is present in reality, or it might be due to a flaw in Tiny Tim's modelling.

The comparison to our example requirements remains the same even with this fit; the PSF model meets our requirements on multiplicative bias, but fails to meet our requirements on additive bias, although it is somewhat closer to meeting these requirements when all optical parameters are fit to each image. This is well beyond the accuracy limit to which Tiny Tim was originally tested, though, and is unlikely to be large enough to have any practical impact on shear measurements using *HST* data. It is only an issue here, since we are testing the PSF model against the requirements for the *Euclid* mission, which are much stricter given the much greater volume of data involved.

The left-hand panel of Fig. 6 illustrates the relationship between the quality of fit and the fitted focus offset value now that all optical parameters are being fit. We no longer see a significant difference between the X^2 for the two chips, but a difference in the best-fitting focus offset remains. This suggests that the issues we saw when only the focus offset was fit were possibly due to the use of other improper optical parameters being used. From the right-hand panel, which shows the relationship between the quality of fit, the date of observation, and the number of stars per image, we see no significant change from before aside from the previously-noted fact that the quality of fit has generally improved.

In Fig. 7, we plot the best-fitting values for each of these fitted parameters against the fitted focus offset, with Tiny Tim's default value for each shown for comparison with a solid black line. From this plot, we see that certain parameters are consistently fit to different values from the defaults used by Tiny Tim, and in some cases there is also a correlation between these fitted values and the fitted focus offset. Of particular note is the X coma parameter, which differs very significantly from its default value: Its mean fitted value is 0.0159, compared to the default value of 0.003.
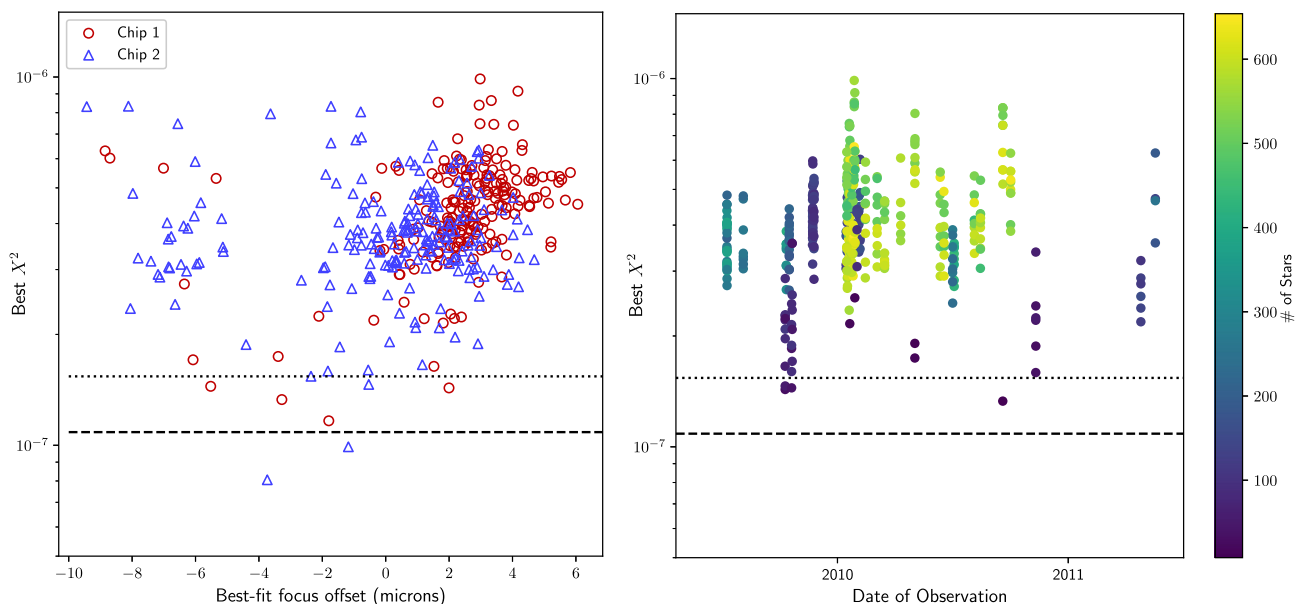
Figure 5. As Fig. 3, except all optical parameters were fit for each image.

It is possible that some or all of the cases where we see a correlation between these values and the focus offset might be due simply to a degeneracy in the models rather than a correlation between the actual values, but an actual correlation is possible. This could occur if the heating and cooling pattern of *HST* has more complicated effects than simply changing the focus offset. This could be caused by, for instance, certain parts of the telescope shading other parts, resulting in deformation due to different portions of it being heated by different amounts. This could result in the optical parameters for the PSF model being altered at the same time that the focus offset is altered, appearing as a linear correlation between them.

If our hypothesis is correct that the default optical parameters used by TINY TIM are incorrect, then if we fit only the focus offset and use either the mean fitted values for the other parameters or the predictions from a linear regression of them against the focus offset, we should see significant improvement in the fitting statistics as compared to using the default parameters. If we do not see this, then it would imply that the improvement in fitting statistics here is likely due to fitting to noise, degeneracies between the focus offset and other parameters, and complicating factors such as unresolved binary stars.

In Fig. 8, we show the resulting fitting statistics from fitting the focus offset with either the mean best-fitting optical parameters (solid bars) or a best-fitting linear relationship between these parameters and the focus offset (transparent bars). Comparing this plot to Fig. 3, we see that either approach provides a notable improvement over using the default parameters, but not quite as much as fitting all optical parameters to each image. This also fails to bring the X^2

Figure 6. As Fig. 4, except all optical parameters were fit for each image.

statistic below the expected value predicted from our representative control tests. This implies that the discrepancy between the models and reality cannot be explained entirely by the optical parameters having shifted since they were originally fit.

As using a linear relationship between the fitted focus offset and other parameters showed similarly little benefit, the discrepancy also cannot be adequately explained by these parameters varying due to the heating and cooling cycle of the instrument, like the focus offset does, unless they vary with a different periodicity or phase. We tested this hypothesis through Fourier analysis (see Appendix C), but this failed to uncover any evidence of any parameters varying with a different periodicity or phase from the focus.

5.3 Comparison to focus model

A model to predict the focus offset of *HST* was developed by di Nino et al. (2008) and Niemi & Lallo (2010). As reported in Cox & Niemi (2011), from tests comparing the model's predictions to measurements of the focus offset made through the phase-retrieval process described in Krist & Burrows (1995), this model predicts the focus offset to within 1 μm ∼50 per cent of the time, and to within 2 μm ∼80 per cent of the time.

In Fig. 9, we show a comparison of our fitted focus offset values to those predicted from the model (taken from the Annual Summary files published online, for the time of the exposure). We can see that the fitted and model values correlate well for large values of the focus, but the relationship worsens for lower values. Part of this is due to the known effect in our fitting procedure that focus offset values are biased away from ∼−3 μm. However, this effect is not sufficiently large to explain the cluster of images for which our fitted focus is ∼−7, while the model focus is ∼−2.

We confirm the finding of Cox & Niemi (2011) that there is a significant difference between the two chips in comparison to the model, of approximately 0.5 μm. However, Cox & Niemi (2011) attributed this to possible differences in charge diffusion and spherical aberration between the two chips, but our results here fit for these parameters individually for each chip, and this difference
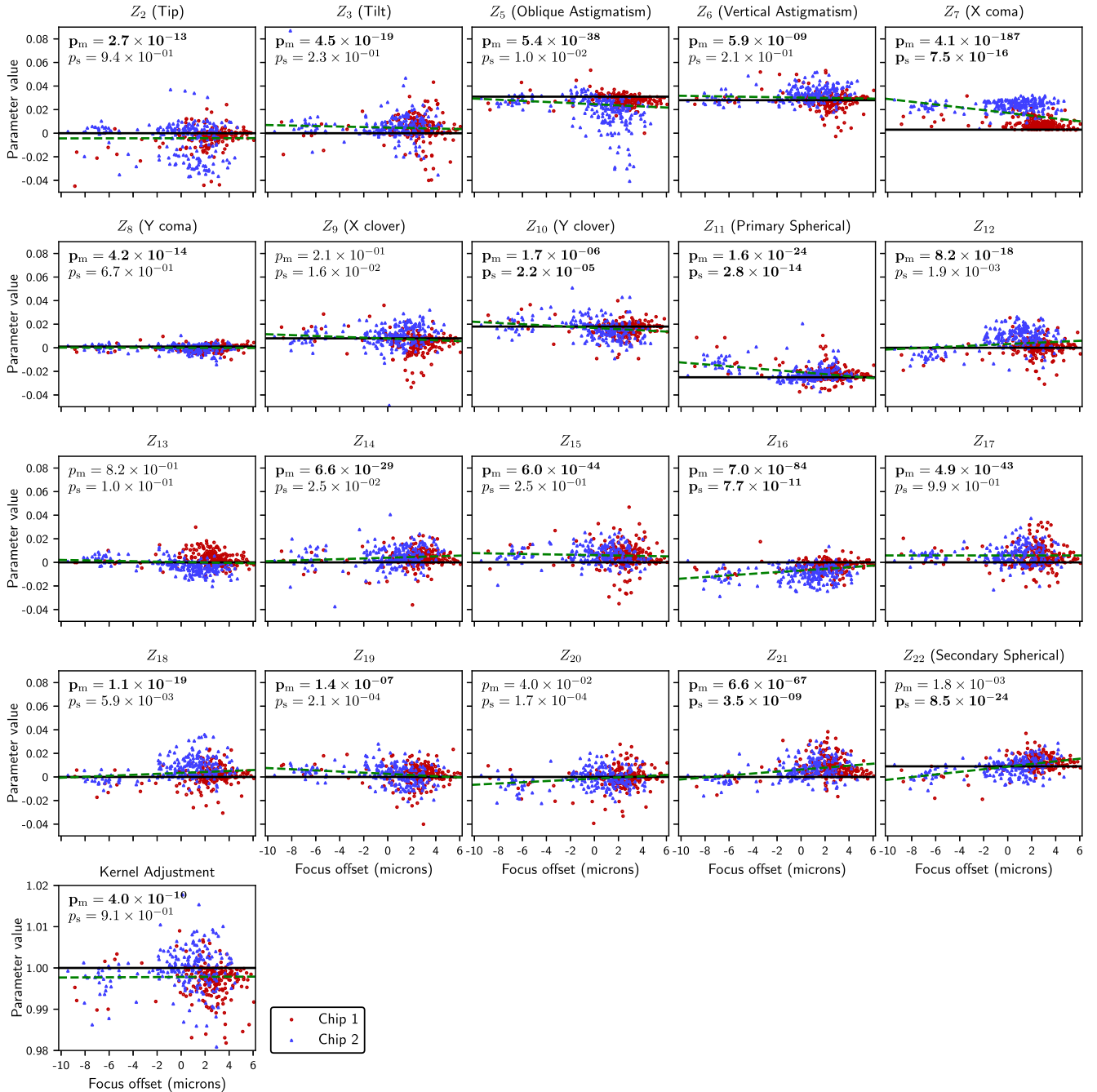
remains. This suggests that this is not the full explanation for this effect, and other possibilities will have to be investigated.

Since the model's predictions were validated against measurements to be generally accurate within 2 μm, it is worth considering why our measurements differ by more than this threshold. The key difference between our fitting procedure and the phase-retrieval process used for validation of the model (see Krist & Burrows 1995) is that the latter uses only out-of-focus observations to fit the focus offset, while we use in-focus observations. Inaccuracy in the extrapolation from out-of-focus to in-focus measurements, for instance due to unknown non-linearity in the motion of the secondary mirror, could result in these measurements differing. Another notable difference is that the measurements used for validation of the model were all made from images of a star at the centre of the detector, while our measurements use stars from various positions on the detectors. If the model fails to properly account for spatial variation of the PSF (see our discussion of this possibility in Section 5.1), this could also account for the difference.

5.4 Discussion

In Section 5.1, we showed that our PSF validation framework indicates a significant discrepancy between the model PSFs generated by the TINY TIM tool and the PSFs of observed stars when using the default parameters and only fitting the focus offset. The magnitude of this discrepancy is too large to be explained by the various effects we expect to see in observations but which were not accounted for in the TINY TIM models (see Section 4.1), the most notable of which is the presence of unresolved binary stars in the sample. One possible explanation of this that we highlight is that the model might not fully capture that spatially varying nature of the PSF, as we see indications that this effect might vary between images, perhaps due to time-dependent temperature gradients distorting the image plane.

We showed in Section 5.2 that if the user takes advantage of the advanced configuration options for TINY TIM and fits all possible optical parameters for each image, the fit can be significantly improved. The improved models unfortunately still fail to match
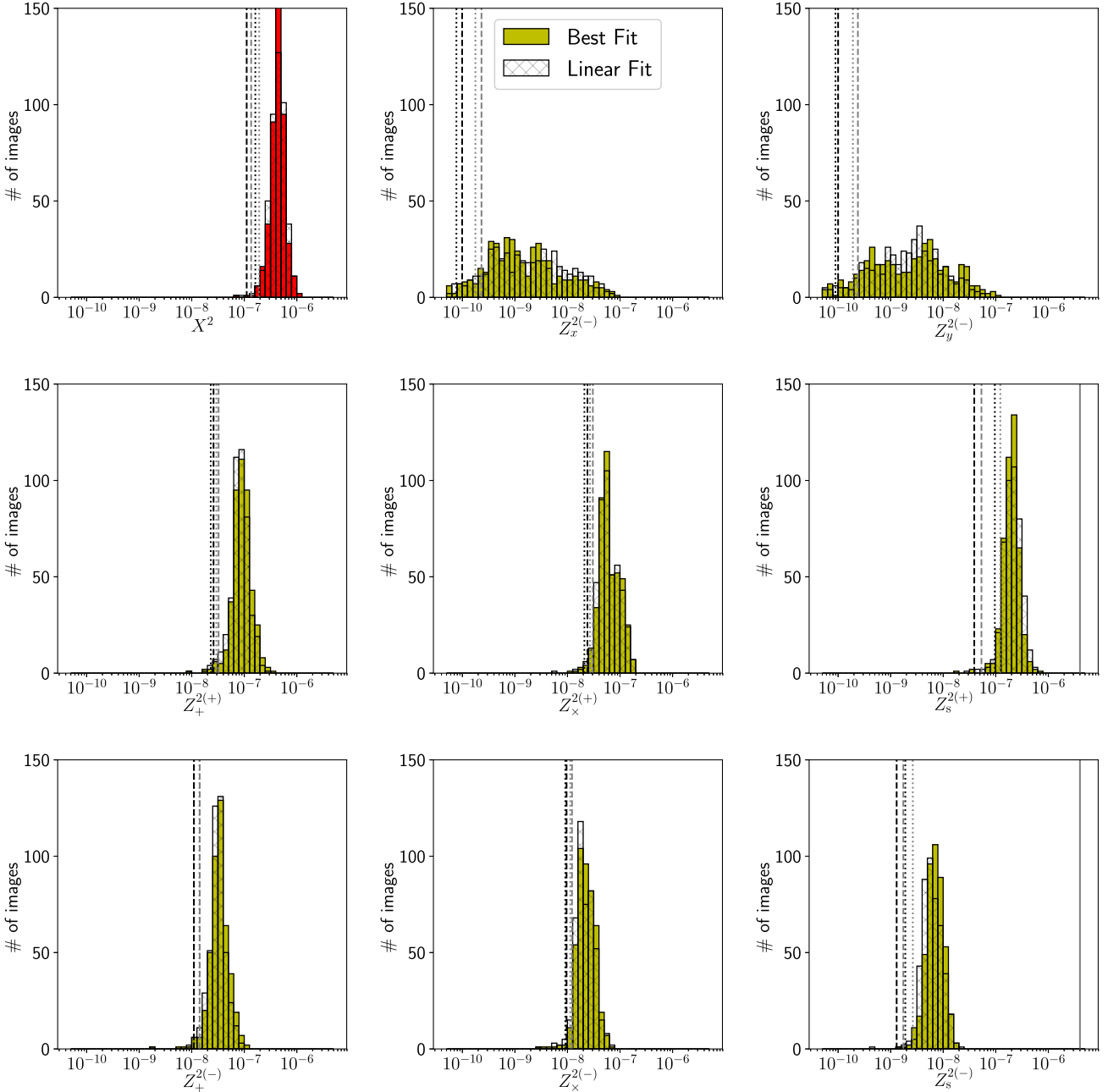
Figure 7. The best-fitting optical parameters over all tested images, plus the adjustment c to the charge diffusion kernel, plotted against the fitted focus offset value. The solid black line indicates the default value used for each by TINY TIM, and the dashed green line is the best linear fit to the data. The blue triangular points correspond to data from Chip 1, and red circular points to data from Chip 2. The p-values stated on each panel are the two-tailed Gaussian probabilities that the mean fitted value is consistent with the default value and that the slope is consistent with zero. When the mean is separated from the default value by at least one standard deviation, or the slope is separated from zero by at least one standard deviation, the respective p-value is emboldened.

the observed PSFs, but the remaining difference is small enough that it is unlikely to cause issue for most use cases on *HST* data. It does fail to meet the more stringent requirements for the *Euclid* mission on additive bias, but it as yet uncertain whether this requirement will in fact need to be met in order for *HST* images to be suitable for validation of the *Euclid* pipeline.

Unfortunately, performing a fit on all possible optical parameters is significantly more time-consuming than fitting the focus offset

alone, and may not be possible with images which only contain a few stars. We investigated two possible ways to improve the fit without having to fit all parameters, by fitting just the focus offset and using either the mean values from the all-parameters fit or a linear relationship between them and the focus offset. Both of these methods provided notable improvement over the focus-offset-only fit. They did not provide quite as significant a benefit as the all-parameter fit, but they require orders of magnitude less
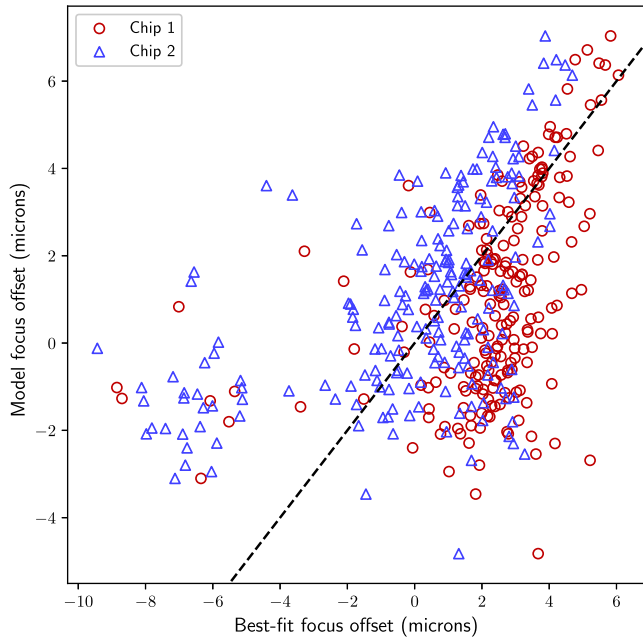
Figure 8. As Fig. 3, except showing cases where the mean best-fitting optical parameters (aside from the focus offset) are used for the model PSFs (solid bars) and where a linear fit between the optical parameters and the focus offset is used (transparent bars). The focus offset is still fit for each image.

computer time to perform, and are thus recommended. In the case of the use of the best-fitting parameters, we calculate that the remaining inaccuracies in the PSF model will cause approximate multiplicative shear-measurement bias of $m_1 \approx m_2 \approx 4.6 \times 10^{-4}$ and additive bias of $c_1 \approx 3.6 \times 10^{-4}$ and $c_2 \approx 3.0 \times 10^{-4}$ (see the calculations in Appendix B for further details on how this is determined).

In order to aid others, we have provided a PYTHON wrapper script in Appendix D to generate TINY TIM PSFs using either the best-fitting optical parameters or a linear fit with the focus offset.

In Section 3.1.3, we had to make an arbitrary choice about what single parameter minimize in our fitting procedure, and we decided to use X^2, which is an equally weighted sum of the Z_k^2 values

which contribute to additive and multiplicative bias, as we did not beforehand know which would be more of an issue. Now that we have our results, we can see that while the $Z_s^{2(+)}$ value was the most significant contributor to X^2 and had the greatest differences between the model and observed PSFs, it still ended up falling well within the range that would confidently meet the *Euclid* mission's requirements on multiplicative bias. On the other hand, while the $Z_+^{2(\pm)}$ and $Z_\times^{2(\pm)}$ values did not contribute as much to X^2 nor differ as much between the model and observed PSFs, they greatly exceeded the range of values that would be consistent with the requirements on additive bias for the *Euclid* mission. This suggests that for future work, it would be more useful to weight the $Z_+^{2(\pm)}$ and $Z_\times^{2(\pm)}$ values more than the $Z_s^{2(+)}$ value, so that the contribution to additive bias might be

Figure 9. The focus offset, as predicted by the model from di Nino et al. (2008) and Niemi & Lallo (2010), plotted against our fitted focus offset values from our all-parameter fit. The dashed line indicates an ideal match between the model and our fitted values. Recall that our fitted values are biased away from a focus offset value of ∼−3, so the actual focus offset values are likely closer to ∼−3 μm than plotted here.

reduced, while the contribution to multiplicative bias is still within requirements.

In Section 5.3, we compared our fitted focus offset values to those predicted by the model from di Nino et al. (2008) and Niemi & Lallo (2010). For positive focus offsets, we found a reasonable correspondence between our fitted values and the model's values, and we confirmed the offset they found between the two chips of the detector. However, for negative values of the focus offset, we fitted significantly lower values than were predicted by the model. This suggests that there is some issue either with the TINY TIM PSFs models for negative focus offsets or with the model's predicted values in this regime. Further investigation into this issue is thus warranted.

6 SUMMARY AND CONCLUSIONS

In this paper, we have presented a framework for the validation of models for space-based PSFs, with a particular focus on the requirements for weak gravitational lensing measurements. As lensing measurements are most directly affected by the dipole and quadrupole moments and size of the PSF, we focused on testing these aspects. We developed a framework which tests these parameters, providing indications when a model fails and allowing it to be determined if a failure will be significant for weak gravitational lensing measurements.

As an example of how this framework can be applied in practice, we used it to assess the quality of model PSFs generated by the TINY TIM tool. To form a basis for analysis, we generated control images on which we first applied the framework. Mock stars were placed on these control images, using PSFs generated from the TINY TIM tool with a known configuration for each image. We then applied the testing framework to each image and compiled the results,

determining the typical statistics we expect from this framework in ideal and more realistic usage scenarios.

We then applied the framework to test model PSFs generated with TINY TIM against stars observed in *HST* ACS images. In the first test, we used the default configuration for TINY TIM, only fitting the focus-secondary-mirror despace for each image. Here, we found that there was a significant difference between the models and observations, most notably with the measured sizes of the model PSFs differing from those of the observed PSFs with significantly more scatter than would be predicted from noise alone. This finding of a significant difference between the model and observations is consistent with the findings of van der Wel et al. (2012), who found a similar effect on a smaller sample of stars observed with the F160W filter, and with the findings of Rhodes et al. (2007).

We then tested more advanced usage of TINY TIM, to see if it is possible for the end-user to generate improved models in some manner. To do this, we tried additionally fitting all Zernike polynomials for the optics up to Z_{22} for each image. In this case, the quality of the fit improved significantly. Although the mismatch between the model and observed PSFs is still statistically significant, the practical significance is minimal for the purposes of weak lensing. Most of this improvement can be retained through the use of the best-fitting optical parameters from this fit and only fitting the focus offset, or else by using a linear relationship between these parameters and the focus offset, with a significantly lower computer time cost.

We thus conclude that our testing framework is a valuable tool for assessing the quality of model PSFs for the purposes of weak gravitational lensing. Additionally, through testing the framework on the TINY TIM tool, we have been able to identify deficiencies with it in its standard configuration and are able to provide a script which will generate improved models, which we supply in Appendix D.

DATA AVAILABILITY

The data underlying this article were accessed from the Barbara A. Mikulski Archive for Space Telescopes (http://archive.stsci.edu/), and may be obtained from the MAST archive at https://dx.doi.org/10.17909/t9-5nb9-8543. The derived data generated in this research will be shared on reasonable request to the corresponding author.

REFERENCES

Anderson J., King I. R., 2006, PSFs, Photometry, and Astronomy for the ACS/WFC, Instrument Science Report ACS 2006-01.
Avila R., Grogin N., Anderson J. et al., 2016, ACS Instrument Handbook Version 15.0. STScI, Baltimore
Bartelmann M., Schneider P., 1999, A&A, 345, 17
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Burrows C. J. et al., 1996, ApJ, 473, 437
Clapp B. R., 2009, J. Astronaut. Sci., 57, 419
Cox C., Niemi S.-M., 2011, Technical Report, Evaluation of a Temperature-Based HST Focus Model. STScI, Baltimore
di Nino D., Makidon R. B., Lallo M., Sahu K. C., Sirianni M., Casertano S., 2008, Technical Report, HST Focus Variations with Temperature. STScI, Baltimore
Hamana T., Miyazaki S., Okura Y., Okamura T., Futamase T., 2013, PASJ, 65, 104
Hoekstra H., Franx M., Kuijken K., Squires G., 1998, ApJ, 504, 636
Hoffmann S. L., Anderson J., 2017, A Study of PSF Models for ACS/WFC. STScI, Baltimore
Jarvis M., Jain B., 2004, preprint (arXiv:astro-ph/0412234)
Jee M. J., Blakeslee J. P., Sirianni M., Martel A. R., White R. L., Ford H. C., 2007, PASP, 119, 1403
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Koekemoer A. M., Fruchter A. S., Hook R. N., Hack W., 2003, in Arribas S., Koekemoer A., Whitmore B., eds, HST Calibration Workshop: Hubble after the Installation of the ACS and the NICMOS Cooling System, MultiDrizzle: An Integrated Pyraf Script for Registering, Cleaning and Combining Images. STScI, Baltimore, p. 337
Krist J., 1993, in Hanisch R. J., Brissenden R. J. V., Barnes J., eds, ASP Conf. Ser. Vol. 52, Astronomical Data Analysis Software and Systems II, TINY TIM: an HST PSF Simulator. Astron. Soc. Pac., San Francisco, p. 536
Krist J., 2003, ACS WFC & HRC Field Dependent PSF Variations due to Optical and Charge Diffusion Effects, Instrument Science Report ACS 2003-06
Krist J. E., Burrows C. J., 1995, Appl. Opt., 34, 4951
Krist J. E., Hook R. N., Stoehr F., 2011, in Kahan M. A., ed., Proc. SPIE Vol. 8127, Optical Modeling and Performance Predictions V. SPIE, Bellingham, p. 81270J–1
Lada C. J., 2006, ApJ, 640, L63
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leauthaud A. et al., 2007, ApJS, 172, 219
Lu T., Zhang J., Dong F., Li Y., Liu D., Fu L., Li G., Fan Z., 2017, AJ, 153, 197
Lucas R. A. et al., 2018, ACS Data Handbook. STScI, Baltimore
Luppino G. A., Kaiser N., 1997, ApJ, 475, 20
Mandelbaum R., 2015, in Cappellari M., Courteau S., eds, Proc. IAU Symp. 311, Galaxy Masses as Constraints of Formation Models – Galaxy Halo Masses from Weak Gravitational Lensing. IAU, Oxford, p. 86
Massey R. et al., 2014, MNRAS, 439, 887
Niemi S.-M., Lallo M., 2010, Technical Report, Phase Retrieval to Monitor HST Focus: II. Results Post-Servicing Mission 4. STScI, Baltimore
Paulin-Henriksson S., Amara A., Voigt L., Refregier A., Bridle S. L., 2008, A&A, 484, 67
Pirzkal N. et al., 2001, A&A, 375, 351
Rhodes J. D. et al., 2007, ApJS, 172, 203
Rowe B., 2010, MNRAS, 404, 350
Rowe B. T. P. et al., 2015, Astron. Comput., 10, 121
Schrabback T. et al., 2018, MNRAS, 474, 2635
Seitz C., Schneider P., 1997, A&A, 318, 687
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Surpi G., Blandford R. D., 2003, ApJ, 584, 100
van der Wel A. et al., 2012, ApJS, 203, 24
van Dokkum P. G. et al., 2008, ApJ, 677, L5
Whitmore B. C., Zhang Q., Leitherer C., Fall S. M., Schweizer F., Miller B. W., 1999, AJ, 118, 1551

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Figure A1. Comparison of our background-independent size estimator, M_s, with moments-based size estimators through tests on an Airy profile of varying size λ/d, using a weighting function favouring the core of the profile (left) and the wings of the profile (right).

Figure A2. As Fig. A1, except with the Q_s size estimate on the horizontal axis and the squares of the other two estimates on the vertical axis, and without displaying versions containing mock noise.

Figure C1. The best-fitting optical parameters, plus the adjustment c to the charge diffusion kernel, for each image plotted against the date of observation.

Figure C2. The amplitudes of the Fourier modes of all optical parameters when they were all fit to each image.

Appendix D. Generating TINY TIM Psfs with our fitted parameters.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SIZE ESTIMATOR TESTS

In Section 3.1.1, we explained our need for a background-independent size estimator. In this appendix, we present tests and analysis of this estimator.

In Fig. A1, we show comparisons of M_s with two moments-based size estimators: an estimator not including a determinant term, calculated as $M_{xx} + M_{yy}$, and an estimator including a determinant term, calculated as $M_{xx} + M_{yy} + 2\sqrt{|\mathbf{M}|}$. Both of these estimators have been used for shear-measurement; the former by e.g. Kaiser et al. (1995), and the latter by e.g. Seitz & Schneider (1997). The comparison illustrates that M_s, like the other two estimators, is monotonic increasing with the size of the measured profile. It is less sensitive when the size is smaller than one pixel, but it is much less prone to oscillations when the wings weighting function is used.

The comparison also illustrates the response of each estimator to mock noise, applied with the same prescription used in Section 2.2, using the same total flux for each profile. This shows that M_s has significantly different noise properties from the other estimators.

Notably, it has a higher signal-to-noise ratio when the wings weighting function is used.

It is also useful to directly compare the moment combinations used for ellipticity estimation to the corresponding M_s values. This allows us to convert observed differences in M_s between models and observed stars to the equivalent changes in these moment combinations, which will allow estimates of how significantly weak lensing measurements would be affected. We show this in Fig. A2, plotting these moment combinations against M_s.

APPENDIX B: RELATING FITTING STATISTICS TO SHEAR BIAS

Using our definition of the relevant moments for shear estimation in equation (5), as well as equations (2) and (4), we can express the calculation of the ellipticity parameters of the undistorted galaxy as

$$
\hat{e}_1 = \frac{M_+^{(u)}}{M_*^{(u)}} = \frac{M_+^{(o)} - M_+^{(p)}}{M_*^{(o)} - M_*^{(p)}},
$$
$$
\hat{e}_2 = \frac{M_+^{(u)}}{M_*^{(u)}} = \frac{M_\times^{(o)} - M_\times^{(p)}}{M_*^{(o)} - M_*^{(p)}}. \tag{B1}
$$

Let us now consider how these are affected by perturbations in $\mathbf{M}^{(p)}$, as might be caused by an imperfect PSF model being used.

Starting with M_+:

$$
\hat{e}_1(1 + m_1) + c_1 = \frac{M_+^{(o)} - \left(M_+^{(p)} + \delta M_+^{(p)}\right)}{M_*^{(o)} - M_*^{(p)}}
$$
$$
= \frac{M_+^{(o)} - M_+^{(p)}}{M_*^{(o)} - M_*^{(p)}} - \frac{\delta M_+^{(p)}}{M_*^{(o)} - M_*^{(p)}}, \tag{B2}
$$

while $\hat{e}_2$ is unchanged. This implies that a perturbation in the M_+ of the PSF model will result in an additive bias on $\hat{e}_1$, with the relationship

$$
c_1 = -\frac{1}{M_*^{(u)}} \delta M_+^{(p)}. \tag{B3}
$$

Similarly, we can calculate

$$
c_2 = -\frac{1}{M_*^{(u)}} \delta M_\times^{(p)}. \tag{B4}
$$

For M_*:

$$
\hat{e}_1(1 + m_1) + c_1 = \frac{M_+^{(o)} - M_+^{(p)}}{M_*^{(o)} - \left(M_*^{(p)} + \delta M_*^{(p)}\right)}
$$
$$
\approx \frac{M_+^{(o)} - M_+^{(p)}}{M_*^{(o)} - M_*^{(p)}} \left(1 + \frac{\delta M_*^{(p)}}{M_*^{(o)} - M_*^{(p)}}\right), \tag{B5}
$$

and similarly for $\hat{e}_2$, so

$$
m_1 = m_2 = \frac{1}{M_*^{(u)}} \delta M_*^{(p)}. \tag{B6}
$$

M_x and M_y do not directly enter the equations for estimating ellipticity, but are used instead to determine the centres of objects. For a moments-based method, this should generally not impact the ellipticity estimate as long as the PSF is centred using the same weight function as the galaxy. For model-fitting methods, however, we cannot rule out that this might have an impact. To estimate the maximum magnitude of this impact, we will consider the worst-case scenario for a moments-based method, where one of our two extreme weighting functions is used to centre the PSF, and the other is used to measure its moments.

If the difference in moments between the two weight functions is $\delta(M_x^c - M_x^w)$ when measured near the centre, and this is small relative to the scale of the weight functions, then $\delta(M_x^c - M_x^w)$ will approximate the difference in centre positions for the two weight functions. It can be shown from equation (1) that when measured offset from the centre by a small distance d_x, M_{xx} will increase by d_x^2. The situation is similar for M_{yy}. This implies that a difference in these moments between the weight function can possibly result in changes to the second-order moments, contributing both additive and multiplicative bias of magnitudes

$$
|m| \approx \frac{4}{M_*^{(u)}} \left[\left(\delta\left(M_x^c - M_x^w\right)\right)^2 + \left(\delta\left(M_y^c - M_y^w\right)\right)^2\right],
$$
$$
|c| \approx \frac{1}{M_*^{(u)}} \left[\left(\delta\left(M_x^c - M_x^w\right)\right)^2 + \left(\delta\left(M_y^c - M_y^w\right)\right)^2\right]. \tag{B7}
$$

The factor 4 in the m bias is due to a factor of 2 from M_{xx} and M_{yy} each showing up twice in the definition of M_* and another factor of 2 since any change in the size affects both m_1 and m_2.

Since we wish to combine all statistics together for a single quality-of-fit parameter, we must decide how to weight additive versus multiplicative bias. For the sake of simplicity, we decide to weight them both equally for this work. We thus have the following scaling relationships between shear bias $b = m_1 + m_2 + c_1 + c_2$ and each parameter:

$$
\frac{\delta b}{\delta M_x^{(p)2}} = \frac{5}{M_*^{(u)}},
$$
$$
\frac{\delta b}{\delta M_y^{(p)2}} = \frac{5}{M_*^{(u)}},
$$
$$
\frac{\delta b}{\delta M_+^{(p)}} = \frac{1}{M_*^{(u)}},
$$
$$
\frac{\delta b}{\delta M_\times^{(p)}} = \frac{1}{M_*^{(u)}},
$$
$$
\frac{\delta b}{\delta M_*^{(p)}} = \frac{2}{M_*^{(u)}}, \tag{B8}
$$

with the proviso that the bias relationships for $M_x^{(p)}$ and $M_x^{(p)}$ are estimates of a worst-case scenario and only relevant for the difference in measurements between weight functions.

We can use these expressions if we wish to relate a perturbation in $\mathbf{M}^{(p)}$ to the amount of bias which will result if we know the $M_*^{(u)}$ of the undistorted galaxy. In order to determine estimates representative of the worst-case scenario, we assume that the smallest possible galaxies are being measured, with their size comparable to the size of the PSF. Here, this corresponds to a typical size of $M_*^{(u)} \sim 5\,\mathrm{px}^2$ when measured with the core weighting function, and $M_*^{(u)} \sim 9\,\mathrm{px}^2$ with the wings weighting function.

For our fitting procedure, we decided to linearize the relevant parameters for each weight function through equation (7), as well as to use an alternative, background-independent size estimator in place of M_*. For the size estimator, we can see from Fig. A1 that M_s has a nearly linear relationship with M_* for an Airy profile, with a slope of

$$
\frac{\delta M_*^{(p)}}{\delta M_s^{(p)}} \approx 1.5. \tag{B9}
$$

We can thus calculate

$$
\frac{\delta b}{\delta M_s^{(p)}} = \frac{\delta b}{\delta M_*^{(p)}} \frac{\delta M_*^{(p)}}{\delta M_s^{(p)}} \approx \frac{3}{M_*^{(u)}}. \tag{B10}
$$

Since we decided to linearize the $M^{c/w}$ values for each weight function into linear combinations $Q^{(\pm)}$, we cannot directly apply equations (B8) and (B10) without undoing the benefits of linearization. Instead, since in any case we cannot calculate exact bias projections without knowing what weight functions will be used, we decide to use representative values instead. Making use of the fact that in the scenario where $\sigma(M_k^{(c)}) = \sigma(M_k^{(w)})$, we will have the property $(Q_k^{(+)})^2 + (Q_k^{(-)})^2 = (M_k^{c})^2 + (M_k^{(w)})^2$, we can assume that the $Q_k^{(\pm)}$ values will be of similar magnitude to the $M_k^{c/w}$ values, and thus

$$\frac{\delta b}{\delta Q_k} \approx \frac{\delta b}{\delta M_k}. \tag{B11}$$

Finally, since we are using a combination of the core and wings weighting function, we must use a representative size $M_*^{(u)}$ intermediate the two size calculations, and so we choose to use the average value $M_*^{(u)} = 7\,\mathrm{px}^2$.

This gives us the ultimate set of representative bias scaling relationships,

$$\frac{\delta b}{\delta (Q_x^{(-)})^2} \approx \frac{5}{7\,\mathrm{px}^2} \approx 0.71\,\mathrm{px}^{-2},$$

$$\frac{\delta b}{\delta (Q_y^{(-)})^2} \approx \frac{5}{7\,\mathrm{px}^2} \approx 0.71\,\mathrm{px}^{-2},$$

$$\frac{\delta b}{\delta Q_+^{(\pm)}} \approx \frac{1}{7\,\mathrm{px}^2} \approx 0.14\,\mathrm{px}^{-2},$$

$$\frac{\delta b}{\delta Q_\times^{(\pm)}} \approx \frac{1}{7\,\mathrm{px}^2} \approx 0.14\,\mathrm{px}^{-2},$$

$$\frac{\delta b}{\delta Q_s^{(\pm)}} \approx \frac{3}{7\,\mathrm{px}^2} \approx 0.43\,\mathrm{px}^{-2}, \tag{B12}$$

which we can use to calculate Z^2 values for each statistic, each of which will be representative of the square of the shear bias contributed by each parameter:

$$Z_x^{2(-)} = \frac{(0.71\,\mathrm{px}^{-2})^2}{N} \sum_{i=0}^{N} \left(Q_{x,\mathrm{star},i}^{(-)} - Q_{x,\mathrm{model},i}^{(-)}\right)^4,$$

$$Z_y^{2(-)} = \frac{(0.71\,\mathrm{px}^{-2})^2}{N} \sum_{i=0}^{N} \left(Q_{x,\mathrm{star},i}^{(-)} - Q_{x,\mathrm{model},i}^{(-)}\right)^4,$$

$$Z_+^{2(\pm)} = \frac{(0.14\,\mathrm{px}^{-2})^2}{N} \sum_{i=0}^{N} \left(Q_{+,\mathrm{star},i}^{(\pm)} - Q_{+,\mathrm{model},i}^{(\pm)}\right)^2,$$

$$Z_\times^{2(\pm)} = \frac{(0.14\,\mathrm{px}^{-2})^2}{N} \sum_{i=0}^{N} \left(Q_{\times,\mathrm{star},i}^{(\pm)} - Q_{\times,\mathrm{model},i}^{(\pm)}\right)^2,$$

$$Z_s^{2(\pm)} = \frac{(0.43\,\mathrm{px}^{-2})^2}{N} \sum_{i=0}^{N} \left(Q_{s,\mathrm{star},i}^{(\pm)} - Q_{s,\mathrm{model},i}^{(\pm)}\right)^2. \tag{B13}$$

Under the assumptions that these quantities are generally independent and that each is as likely to cause a positive or negative bias, we can treat the sum of these values,

$$X^2 = \sum_k Z_k^2, \tag{B14}$$

as representative of the square of the total shear bias contributed by inaccuracies in the PSF. This provides us with our desired goal of a single representative quantity which we can minimize in the fitting procedure.

APPENDIX C: TIME DEPENDENCE OF BEST-FITTING OPTICAL PARAMETERS

In Section 5.2, we showed that fitting all possible optical parameters to each image provided a significantly better match between the model PSFs and observed stars than only fitting the focus offset. When the mean best-fitting values for each parameter were used, or when linear relationships between these values and the focus offset were used, the improvement over the focus-offset-only case was not nearly as significant. In this appendix, we illustrate other tests we performed on these best-fitting values to test if there is evidence for any time-dependent behaviour.

In Fig. C1, we plot the best-fitting optical parameters and our adjustment to the charge diffusion kernel against the date of observation and show the best-fitting linear relationship. If any parameter varies gradually or discontinuously with time, we would expect to a significantly non-zero slope in the best-fitting linear relationship here. However, the slope is nearly zero for all parameters except Z_{21}. With this many parameters tested, it is not unusual that one might have a significantly non-zero slope due to noise alone, and so overall we see no evidence here to indicate a significant gradual or discontinuous relationship between any parameters and the date of observation.

Another possibility is that one or more of the optical parameters might vary periodically with time. To test this, we calculate Fourier modes for each parameter through

$$A_f = \left| \sum_j (v_j - \langle v_j \rangle) \exp 2\pi i f t_j \right|, \tag{C1}$$

where summation is performed over all images j, v_j is the value of a given optical parameter (or the modification to the charge-diffusion kernel) fit for image j, $\langle v_j \rangle$ is the expected value of v_j from a linear fit to the focus offset, f is the frequency, and t_j is the time of observation for image j. The subtraction of $\langle v_j \rangle$ here is done to reduce or remove the impact of degeneracies with the focus offset on the calculated amplitudes here. Note that due to sparse sampling, the calculated amplitudes for different frequencies are not necessarily independent.

We plot the resulting amplitudes in Fig. C2. We would expect any periodic behaviour to manifest as spikes in these plots. Parameters such as Z_7 (the 'X coma') and Z_{17} show coherent variations in the amplitude across a large range of frequencies, which likely arises simply from coincidental noise in the time-domain values, and does not have any physical meaning.

The dotted lines in this plot correspond to integer multiples of *HST*'s orbital frequency. Any spikes near these values would correspond to variations out of phase with changes to the focus, but we see no such spikes. There may however be variations in phase with the focus, but we cannot disentangle these from the effects of degeneracies between the parameters.

Overall, there is not enough evidence to confirm the presence of periodic behaviour in any of the optical parameters, but the data are too noisy to entirely rule it out.

APPENDIX D: GENERATING TINY TIM PSFS WITH OUR FITTED PARAMETERS

In order to aid others in easily generating TINY TIM PSFs using the best-fitting optical parameters we have determined in this paper, we list both the mean values and the parameters for the linear fit in Table D1, and we provide PYTHON code that uses them to call TINY

TIM to generate subsampled PSFs. This code is hosted online at https://bitbucket.org/brgillis/tinytim_psfs and in the electronic version of this document.

Table D1. The mean values and slope and intercept of a linear fit against focus offset (in µm) for the optical parameters used by TINY TIM, determined from our all-parameter fits.

Name	Mean	Intercept	Slope
Z_2 (Tip)	−0.0042	−0.0043	0.0000
Z_3 (Tilt)	0.0046	0.0048	−0.0002
0 degree astigmatism	0.0241	0.0246	−0.0005
45 degree astigmatism	0.0300	0.0302	−0.0002
X coma	0.0159	0.0172	−0.0012
Y coma	0.0000	0.0000	0.0000
X clover	0.0074	0.0079	−0.0004
Y clover	0.0163	0.0169	−0.0005
Spherical third	−0.0217	−0.0207	−0.0008
Z_{12}	0.0037	0.0002	0.0005
Z_{13}	0.0001	0.0003	−0.0002
Z_{14}	0.0043	0.0039	0.0003
Z_{15}	0.0059	0.0061	−0.0002
Z_{16}	−0.0061	−0.0069	0.0007
Z_{17}	0.0059	0.0059	0.0000
Z_{18}	0.0039	0.0034	−0.0004
Z_{19}	0.0020	0.0026	−0.0005
Z_{20}	−0.0008	−0.0014	0.0005
Z_{21}	0.0072	0.0062	0.0008
Spherical fifth	0.0101	0.0088	0.0011
Kernel adjustment	0.9978	0.9978	0.0000

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