Draft version April 25, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

A Hierarchical PSF Reconstruction Method

Pedro Alonso,[1, 2] Jun Zhang,[1, 2] and Cong Liu[1, 2]

[1] *Department of Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*
[2] *Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai 200240, China*

ABSTRACT

Reconstruction of the point spread function (PSF) plays an important role in many areas of astronomy, including photometry, astrometry, galaxy morphology, and shear measurement. The atmospheric and instrumental effects are the two main contributors to the PSF, both of which may exhibit complex spatial features. Current PSF reconstruction schemes typically rely on individual exposures, and its ability of reproducing the complicated features of the PSF distribution is therefore limited by the number of stars. Interestingly, in conventional methods, after stacking the model residuals of the PSF ellipticities and (relative) sizes from a large number of exposures, one can often observe some stable and nontrivial spatial patterns on the entire focal plane, which could be quite detrimental to, e.g., weak lensing measurements. These PSF residual patterns are caused by instrumental effects as they consistently appear in different exposures. Taking this as an advantage, we propose a multi-layer PSF reconstruction method to remove such PSF residuals, the second and third layers of which make use of all available exposures together. We test our method on the i-band data of the second release of Hyper Suprime-Cam. Our method successfully eliminates most of the PSF residuals. Using the Fourier_Quad shear measurement method, we further test the performance of the resulting PSF fields on shear recovery using the field distortion effect. The PSF residuals have strong correlations with the shear residuals, and our new multi-layer PSF reconstruction method can remove most of such systematic errors related to PSF, leading to much smaller shear biases.

Keywords: techniques: image processing – instrumentation: detectors – telescopes – astrometry – (cosmology:) gravitational lensing

1. INTRODUCTION

The point spread function (PSF) measures the diffraction of light in optical systems, which makes point-like sources appear extended. In astronomical images, the PSF effect distorts the shape and size of all celestial objects. For ground-based telescopes, the properties of the PSF are mainly determined by the telescope optics (Jarvis et al. 2008) as well as the atmospheric turbulence (Roddier 1981; Fétick et al. 2018; Hébert et al. 2018; Xin et al. 2018). A minor contributor to the PSF is the pixelation effect, resulting from the finite pixel size of the CCD images (High et al. 2007; Zhang 2010; Kannawadi et al. 2021; Shen et al. 2022; Hirata et al. 2024).

PSF plays a particularly important role in weak lensing studies, since it convolves the lensed galaxies and distorts their shapes akin to the cosmic shear. This makes the PSF effect the most important source of systematics in shear measurements (Rhodes et al. 2007; Miller et al. 2013; Lu et al. 2017; Liu et al. 2023). To ensure the reliability of the measurements, precise modeling of the PSF is essential.

When modeling the PSF, it is convenient to describe it as a combination of time-variant and time-invariant features. Time-variant features vary from exposure to exposure and are produced by the atmospheric turbulence and some instrument-related issues such as misalignments in the optical components of the telescope, mechanical deformations caused by gravitational or thermal effects, tracking errors, or instrumental instabilities. On the other hand, time-invariant features are quite stable from exposure to exposure. Optical designs, mechanical elements of the telescope, as well as characteristics

Corresponding author: Jun Zhang
betajzhang@sjtu.edu.cn

of the imaging sensor, focal length, or aperture size can all induce time-invariant features on the PSF.

Multiple methods have been proposed to model the PSF effect for shear measurement (Kaiser et al. 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998; Kaiser 2000; Bernstein & Armstrong 2014; Zhang et al. 2015; Bernstein et al. 2016). With time-variant features complicating the PSF modeling, most of these methods rely solely on individual exposures to build the PSF model, thus being limited by the number of available stars. An effective way to deal with a limited number of stars is to assume a specific functional form for the distribution of the PSF, e.g., polynomial functions. Several methods that follow this approach have been successful in capturing most of the PSF features on the exposure level. Other methods, based on Principal Component Analysis (PCA) (Jarvis & Jain 2004) or machine learning (Herbel et al. 2018; Jia et al. 2020), have also shown impressive results in capturing the spatial features of the PSF. The optimal PSF reconstruction method generally depends on the particular shear measurement being considered, and the particular dataset.

In current methods, however, when stacking the PSF ellipticity residuals of the PSF models of many exposures, some complex features appear (Bosch et al. 2017; Jarvis et al. 2020; Zhang et al. 2022). These features are time-invariant, and therefore related to instrumentation. In this work, we present a novel PSF reconstruction method that performs three interpolations in a hierarchical manner, with the aim of modelling the high-frequency yet stable spatial features of the PSF distribution. Our method builds upon the PSF reconstruction method of Liu et al. 2024 (in preparation), constructing a model of the PSF power spectrum. We apply our method to the i-band of HSC pDR2 data, in which systematic PSF residuals have been previously reported. Our method successfully removes most of the systematic PSF residuals found in HSC DR2, significantly improving the PSF model.

The structure of this paper is as follows. §2 introduces the HSC dataset. The first layer of interpolation is explained in §3. The second and third layers of interpolation are described in §4, which also includes our main results. We conclude in §5.

2. HSC DATASET

The Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) project (Aihara et al. 2018; Miyazaki et al. 2018; Komiyama et al. 2018; Furusawa et al. 2018) is an optical multi-layer imaging survey that covers approximately 1400 deg^2 in five bands (g, r, i, z, y) in its *Wide* layer ($r \sim 26$). In addition, the survey includes two deeper layers, *Deep* and *UltraDeep*, covering 27 deg^2 ($r \sim 27$) and 3.5 deg^2 ($r \sim 28$), respectively. The project utilizes the Hyper Suprime-Cam, a wide field optical camera built on the 8.2m Subaru Telescope, and aims to address some of the most important problems in astrophysics and comsmology, with a focus on weak gravitational lensing, galaxy evolution, supernovae, and galactic structure. Their data is publicly available at their official website[1].

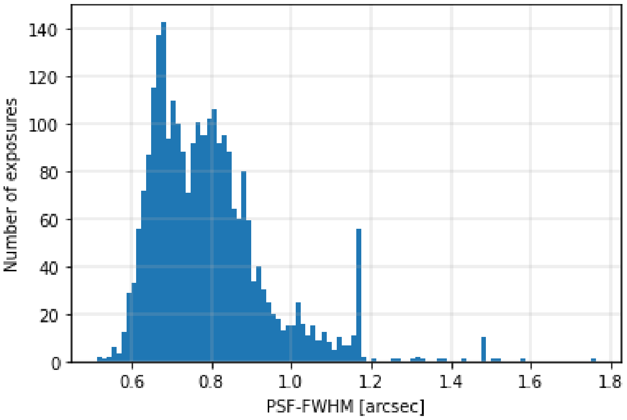

Figure 1. Distribution of the PSF FWHM of the i-band of HCS pDR2.

In this paper we use the i-band data of the second public data release of the HSC (HSC pDR2; Aihara et al. 2019). HSC pDR2 covers an area of 300 square degrees in the *Wide* layer, in all five bands. The data was collected over 174 nights of observation, from March of 2014 to January of 2018. HSC pDR2 includes significant improvements over the previous release (HSC pDR1), which include improved background subtraction, PSF modeling, and object detection procedures. However, we only use the background-removed CCD images, performing our own PSF reconstruction. The galaxies are selected from the official HSC catalog (Aihara et al. 2019). Fig. 1 shows the distribution of the PSF full width at half maximum (FWHM), in real space, for the i band of HSC pDR2.

3. PSF RECONSTRUCTION

Our method builds upon the PSF reconstruction method of Liu et al. 2024 (in preparation), which is part of the Fourier_Quad shear measurement pipeline (Zhang et al. 2022). Fourier_Quad utilizes the quadrupole moments of the power spectrum of galaxy images to measure the cosmic shear. In line with this approach, we

[1] https://hsc-release.mtk.nao.ac.jp/

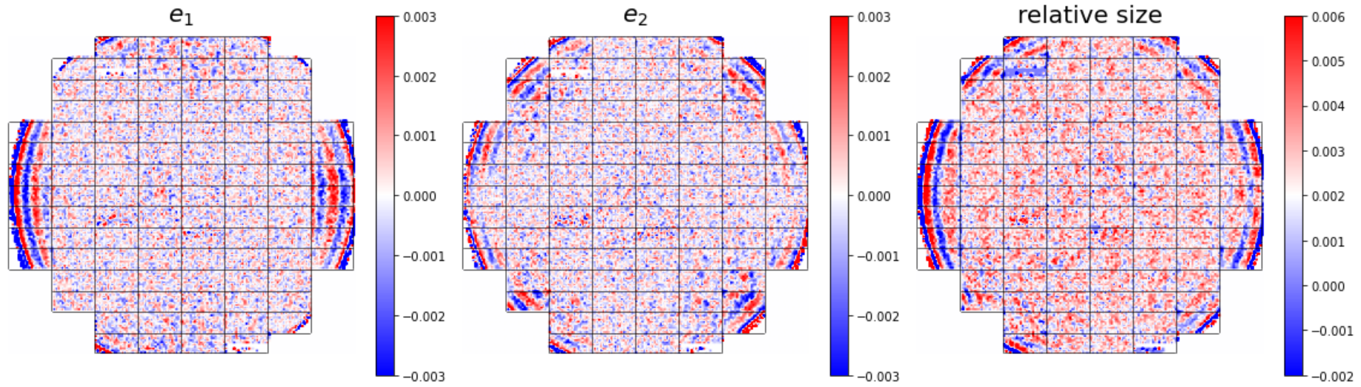

Figure 2. Stacked ellipticity (e_1, e_2) and relative size residuals from all exposures of the i-band of HCS pDR2.

perform PSF interpolation directly on the power spectrum of the stars, building a model of the power spectrum of the PSF (PPS hereafter).

3.1. *Star selection*

Several techniques have been developed for star-galaxy separation. Some methods rely on morphological features (Slater et al. 2020) or color information of the sources (Pollo et al. 2010). Moreover, modern approaches based on machine learning methods such as decision trees (Vasconcellos et al. 2011) or deep convolutional neural networks (Kim & Brunner 2017) have shown promising results in star-galaxy separation.

In this work, we follow the following procedures to select out bright stars for PSF reconstruction (Liu et al. 2024, in preparation):

1. For each exposure, we select sources with SNR $\geq$ 100 to form our initial set of star candidates. The rest of the steps only make use of their power spectra, which are all normalized so that the power $P(\vec{k}=0)$ is unity. Note that in Fourier space, point sources should have the most extended profiles. We therefore measure the area of each candidate (defined as the number of pixels above 0.02). The distribution of the area typically has a Gaussian shape. We throw away those candidates that are away from the peak of the distribution by more than 3σ to remove the outliers.

2. From the remaining star candidates, we form the first star model (power spectrum) pixel-by-pixel. Each pixel value is determined by sorting the corresponding pixel values from the power of all candidates, and taking the lower bound of the top 25%.

3. The similarity between the candidates and the star model can be quantified by defining a χ^2 as:

$$\chi^2 = 2\sum_{i=1}^{N}\left(I_i^n - I_i^{model}\right)^2 / \sum_{i=1}^{N}\left(I_i^n + I_i^{model}\right) \qquad (1)$$

in which I_i^n refers to the value of the $i^{\rm th}$ pixel in the power image of the $n^{\rm th}$ candidate, and $I_i^{\rm model}$ refers to the corresponding pixel value of the model. N is the total number of pixels involved, which are typically chosen to be those located in the middle part of the stamp. The distribution of the χ^2 forms a Gaussian-like function, and we remove those candidates whose χ^2 are more than 3σ away from the peak of the distribution.

4. We build the new star model, this time as a function of location, by interpolating the pixel values of all the remaining candidates with polynomial functions of order nine. Note that if there are not enough candidates left for the fitting, we simply stop processing the exposure further. We again use eq.(1) to define the χ^2 between each candidate and the star model at its location. From the distribution of χ^2, we again remove the candidates that are more than 3σ away from the peak. The surviving candidates are treated as stars for our PSF reconstruction.

3.2. *First layer of interpolation*

Our first interpolation is a 2D polynomial fitting of third order on the power spectrum of the stars on each CCD image, following the procedure of Liu et al. 2024 (in preparation). To ensure the reliability of the PSF model, we only include CCDs containing a minimum of 20 stars. Each star power spectrum is centered on a 48x48 stamp, and the interpolation is done pixel-by-pixel.

To evaluate the quality of the interpolation, we calculate the PPS residuals, represented as the ellipticity and relative size residuals at the positions of the stars. The ellipticity components, e_1 and e_2, and the size are defined based on the quadrupole moments, Q_{ij}, as:

$$
\begin{aligned}
e_1 =& \frac{Q_{20} - Q_{02}}{Q_{20} + Q_{02}} \\
e_2 =& \frac{2Q_{11}}{Q_{20} + Q_{02}} \qquad (2) \\
size =& \frac{Q_{20} + Q_{02}}{Q_{00}},
\end{aligned}
$$

where

$$Q_{ij} = \sum_{P(k)>P_0} P(k) k_x^i k_y^j. \qquad (3)$$

We only include pixels above $P_0 = 0.02 \cdot P(k=0)$ in the calculation.

Ellipticity residuals are calculated as the difference between the ellipticities of the original ($e_{1,true}$, $e_{2,true}$) and the predicted ($e_{1,pred}$, $e_{2,pred}$) stamps, while the relative size residuals are calculated as $(size_{true} - size_{pred})/size_{true}$. Fig. 2 shows the stacked ellipticity and relative size residuals including all exposures. While the residuals are very small for most part of the exposure, we find some remaining systematic residuals near the boundaries, which are not visible on single exposures but become prominent after stacking a large number of them. Note that similar PSF residual patterns are also observed with the official HSC pipeline (Bosch et al. 2017), as well as in the DES data (Jarvis et al. 2020) and the DECaLS data (Zhang et al. 2022). In the next layer of interpolation, we build a model for these systematic PSF residuals.

4. IMPROVED PSF RECONSTRUCTION

The systematic PPS residuals observed in fig. 2 are related to instrumental effects rather than the atmospheric turbulence. Given the strong spatial and temporal dependencies of the atmospheric turbulence, it is unlikely to produce any persistent features on the PSF residuals. In the following we present our second layer of interpolation, which builds a model for the systematic features of the PPS residuals. Unlike in the first interpolation, the features that we want to model are systematic, common to all exposures. Therefore, instead of performing an interpolation on individual chips, we collect the PPS residual stamps from all exposures and place them into a single exposure. We then perform a single interpolation on each CCD of that exposure. This procedure significantly increases the amount of available data for interpolation.

To collect the PPS residuals from all exposures and place them into a single exposure, we must first take into account the different PSF sizes in different exposures (fig. 1). We rescale the PPS residual stamps to a common size, with the re-scaling factor determined as the mean PSF size of the exposure. It is important to emphasize that the above procedure of rescaling the PSF stamps from different exposures and place them into a single exposure is only beneficial when modelling systematic features, common to all exposures. In our first interpolation, the main contributor to the PSF was the atmospheric turbulence, which strongly varies from exposure to exposure, hence the interpolation was performed on individual exposures.

4.1. *Rescaling of the PPS residuals*

We rescale each PPS residual to a reference PSF FWHM of 1 arcsec in real space. For each PPS residual stamp, we calculate the rescaling factor s as the mean PSF FWHM size, in real space, of all stars in the same CCD in unit of arcsec. To build the rescaled PPS residual stamp, we project the coordinates (u, v) of each pixel back to the original stamp, as $(u \cdot s, v \cdot s)$. In most cases, the projected coordinates $(u \cdot s, v \cdot s)$ lies within four pixels of the original stamp. The value of the (u, v) coordinate in the rescaled stamp is calculated as a weighted average of those four neighboring pixels in the original stamp, with the weights determined by the inverse of the pixel distances to $(u \cdot s, v \cdot s)$.

4.2. *Principal Component Analysis (PCA) of the PPS residuals*

Once the PPS residual stamps have been rescaled, we place them into a single exposure, and apply principal component analysis to the stamps on each CCD. Principal component analysis (PCA; Shlens 2014) is a widely used method for dimensionality reduction in data analysis. It transforms an N-dimensional system into a lower dimensional representation, characterized by Principal Components, or PCs. These PCs form the basis of the new space, representing the main features of the data. They are orthogonal to each other, which helps eliminate the correlation between variables, making PCA particularly useful for large datasets, where numerous correlated variables make data interpretation difficult. The main motivation for applying PCA to the PPS residuals is to capture the main features of the data, eliminating unnecessary information or noise. Additionally, dimensionality reduction drastically decreases the computational time, since instead of performing the interpolation pixel-by-pixel, now we perform an interpolation for each principal component.

We choose 100 as the number of principal components, resulting in each stamp being represented by the 100

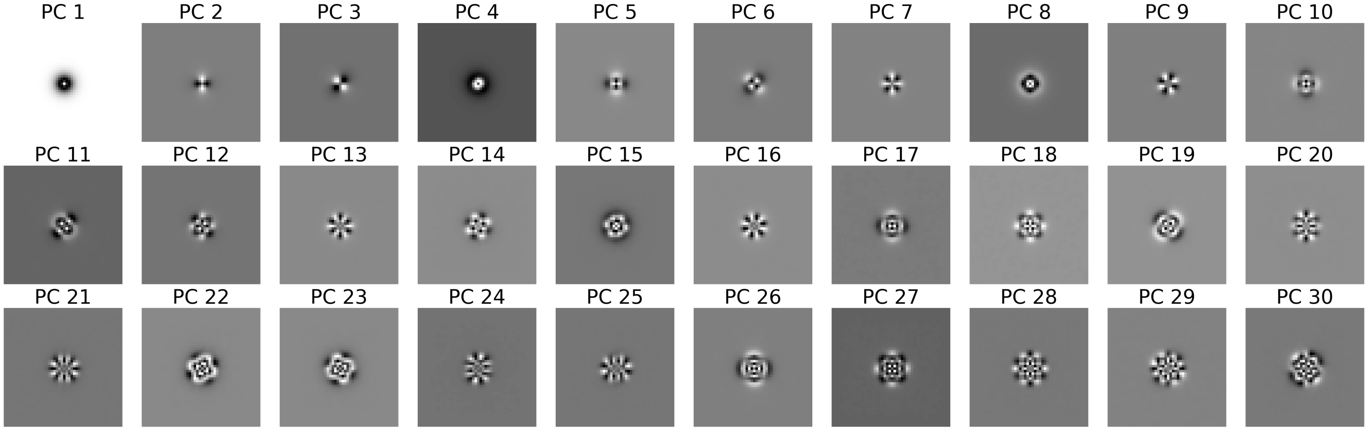

Figure 3. First 30 principal components of the CCD number 50, located in the central region of the exposure.

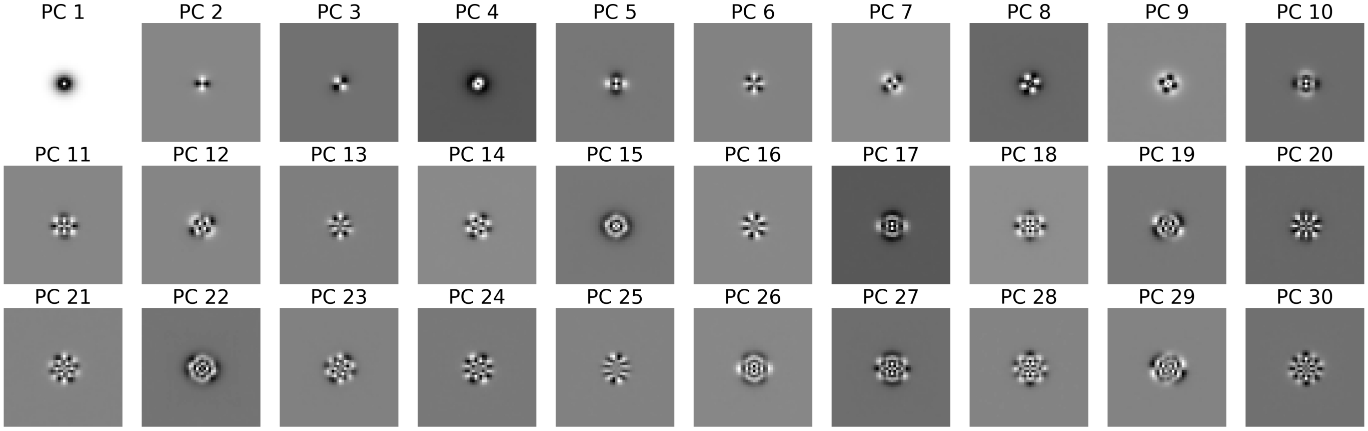

Figure 4. First 30 principal components of the CCD number 53, located in the boundary of the exposure.

coefficients to the PCs. Fig. 3 and fig. 4 show the first 30 principal components of the CCDs number 50 and 53, respectively (Aihara et al. 2018). The CCD number 50 is located in the central region of the exposure, whereas the CCD number 53 is located in the boundary. We observe that their first four principal components are almost identical, diverging from the fifth, showing the different features of the PPS residuals in the two regions.

4.3. *Reconstruction of the PPS residuals*

We build a model for each principal component coefficient at the CCD level by interpolating the PC coefficients of the PPS residuals within that CCD. We test two different interpolation methods, polynomial and random forest, and compare their performance.

4.3.1. *Polynomial*

The features of the PPS residuals are particularly complex near the boundaries of the exposure, thus a single interpolation on a CCD level is unlikely to fully capture them. Since the amount of data per CCD is now very large, we further divide each CCD into four equal parts (2×2 in the CCD plane), and fit a polynomial of order six to the PC coefficients of the PPS residuals within each part. The predicted PPS residual stamp (48x48) at the position of each star is constructed as the sum of the PCs, weighted by the predicted PC coefficients. Finally, we rescale the current PPS residuals model back to the original PSF size at each position. We follow the same procedure as §4.1, with the scaling factor being the inverse of s.

The improved PPS model is built by adding the PPS residuals model to the original PPS model from the first interpolation. Fig. 5 shows the new ellipticity and relative size residuals, calculated at the position of the stars, as the difference between the ellipticity components/relative size of the true star and the new PPS model. We observe a significant improvement, with the systematic PPS residuals vanishing almost completely.

4.3.2. *Random Forest*

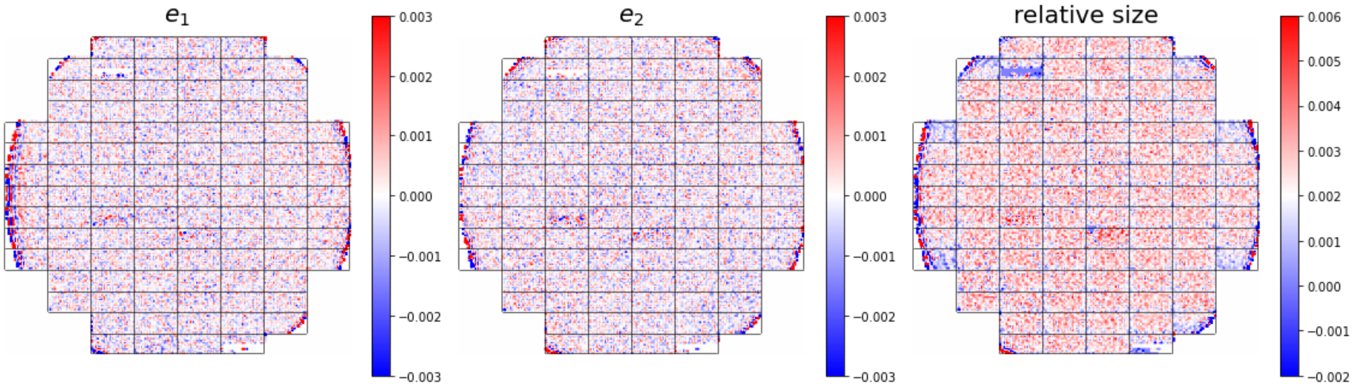

Figure 5. Stacked ellipticity (e_1, e_2) and relative size residuals of all exposures of the i-band of HSC pDR2, after the second interpolation, for the case of a polynomial fitting of order 6 as our second interpolation.

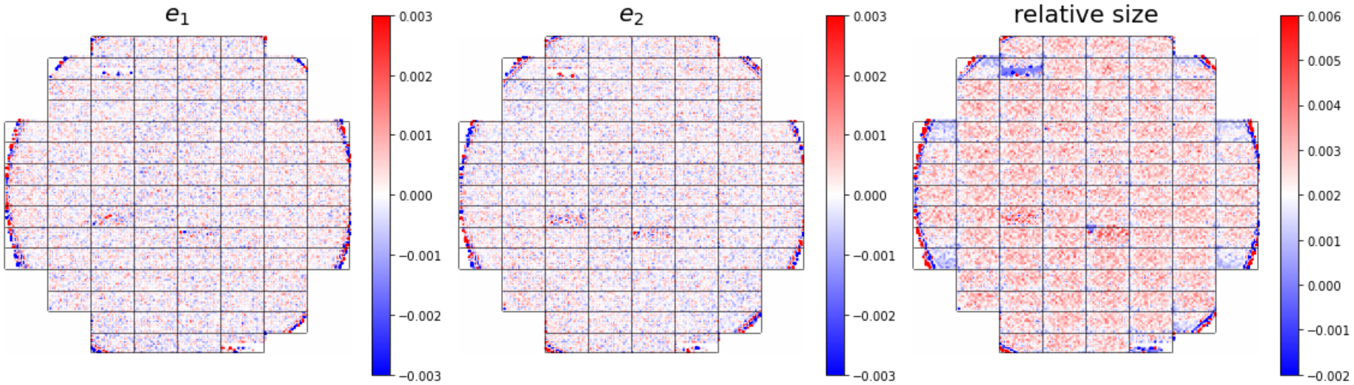

Figure 6. Stacked ellipticity (e_1, e_2) and relative size residuals of all exposures of the i-band of HSC pDR2, after the second interpolation, for the case of random forest as our second interpolation.

In this section we follow the same procedure as in §4.3.1, but using a machine learning algorithm called random forest to interpolate the principal component coefficients of the PPS residuals. Random Forest (Ho 1995; Breiman 2001) is a widely used machine learning algorithm, applicable to both classification and regression problems. It is an ensemble method that combines the predictions of multiple models, called decision trees, to infer a final prediction. Each decision tree is built on a randomly selected subset of the data, and the output of the random forest is the average output of all the decision trees. This significantly reduces the risk of overfitting.

Each decision tree is built through a series of binary splits of the data, starting from the root node. At first, two subnodes of the root node are created, based on a splitting condition. Following the same procedure, each of the subnodes is further divided into two new subnodes, based on new splitting conditions, and so on, building the tree. When a stopping condition is met, we stop splitting that node.

In our case, each decision tree performs a regression on a randomly selected subset of our data, where each datapoint is described by the coordinates (x, y) on the CCD, and the principal component coefficient value, z. Starting from the root node, we perform the first split of the data. The algorithm splits the data into two groups or subnodes (A and B), based on a condition—either on x or y—that better splits the data according to the values of z. It is important to highlight that the algorithm sees x and y as distinct features of the data, rather than as coordinates. As a result, each split is based on a single feature, either x or y. To determine the optimal splitting condition for each node, we use the mean squared error (MSE), a widely adopted metric for splitting in regression tasks. We define the MSE of each possible split as:

$$MSE = \sum_{(x,y)\in A} (z - z_{pred,A})^2 + \sum_{(x,y)\in B} (z - z_{pred,B})^2, \quad (4)$$

where $z_{pred,A}$ and $z_{pred,B}$ are the predicted PC coefficients associated to each group, and it is the same for

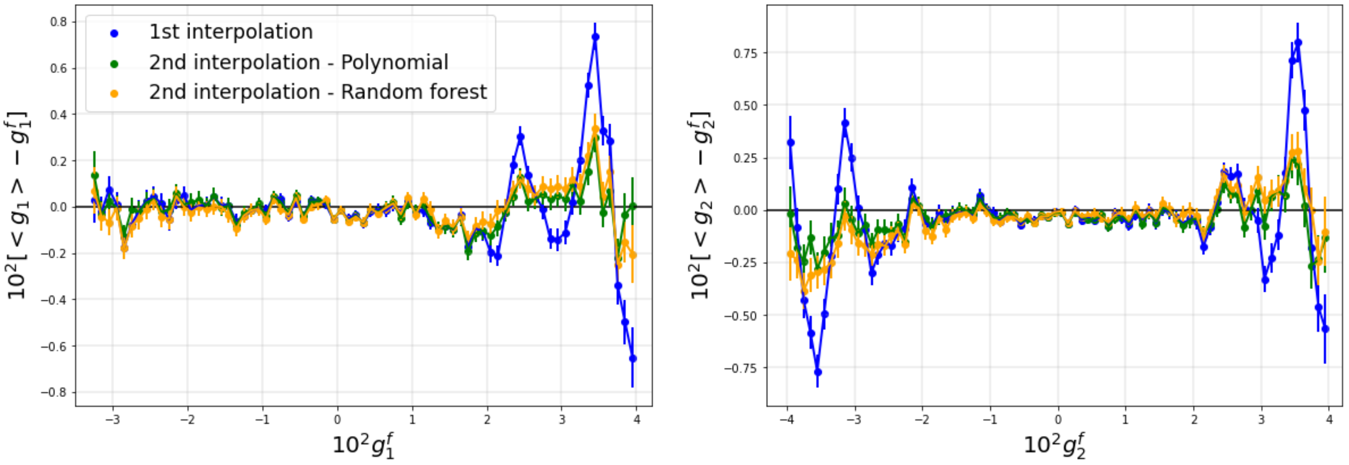

Figure 7. Field distortion test after first interpolation (blue curves), and after the second interpolation, for the case of polynomial of order 6 (green curves) and random forest (yellow curves). Results are shown for g_1 (left panel) and g_2 (right panel). Data points show $1-\sigma$ errorbars.

all elements in the group. The split with the lowest MSE value is the optimal split.

We continue splitting each subnode (A and B), following the same procedure as for the root node. We stop splitting a node when it contains less than 500 datapoints. This stopping criterion is found empirically, as further splitting the data results in overfitting.

A decision tree generates predictions for new coordinates (x, y) by following the splitting conditions for each node. Starting at the root node and moving down the tree, the new data reaches a leaf node, i.e., a node without subnodes. The predicted PC coefficient for the new coordinates is the value associated to that leaf node, z_{pred}.

As in §4.3.1, we divide each CCD into four parts (2×2 in the CCD plane), and build models for the PC coefficients of the PPS residual stamps within each part. We build 100 decision trees for each principal component coefficient, and determine the predicted PC coefficient at each CCD position, z_{pred}, as the average predictions of all decision trees. We build the PPS residual model as the weighted sum of the principal components, with the weights given by the PC coefficients predicted by random forest, z_{pred}. To obtain the final PPS residual model at each CCD position, we rescale the PPS residual model back to the PSF size of each exposure. Fig. 6 shows the ellipticity and relative size residuals of our improved PPS model, calculated as the sum of the PPS model from the first interpolation and the PPS residual model. As in the polynomial case, we obtain very small PPS residuals, removing almost completely the systematic PPS residuals of fig. 2. Although polynomial and random forest obtain comparable results, the PPS residuals of random forest are slightly smaller. However, this does not necessarily mean that random forest is going to be more accurate in making predictions at the position of galaxies. As a non-parametric algorithm, random forest could in principle capture more complex features on the data. However, they are also more prone to overfitting. To evaluate the predictions at galaxy positions of our new PPS models, we use the field distortion test.

4.4. *Field distortion test*

The field distortion (FD; Zhang et al. 2019) is an optical aberration characterized by the deviation from global rectilinear projection. It induces a distortion in the shape of galaxies in a similar way as the cosmic shear, which can be directly derived from astrometry parameters. Zhang et al. (2019) proposed a method to use the field distortion to evaluate the accuracy of shear recovery directly on real galaxies, by comparing the measured field distortion induced shear (FDS) and the true field distortion inferred from astrometry. This is known as the field distortion test. We refer to Zhang et al. (2019) for a derivation of the FDS equations.

We use the FD test to evaluate the performance of our PSF model on real galaxies, and compare the results of polynomial and random forest. We evaluate the true FDS, g_1(FD) and g_2(FD), against the FDS recovered from galaxies, g_1(gal) and g_2(gal).

Fig. 7 compares the FD test results for $\mathrm{SNR_F} \geq 4$ (Li & Zhang 2021), after the first and second interpolation. Note that the FDS signals are removed from the galaxies. The x-axis represents the true FD signal and the y-axis the difference between the recovered FD signal from astrometry and the true one. The results after the first interpolation clearly show the imprints of the PPS residuals of fig. 2, and suggest that all galaxies with a FD signal of $|g_{1,2}(\mathrm{FD})| > 0.02$ should be removed from the shear catalogs, as their PSF models are not reliable. We find a significant improvement after the second interpolation, particularly in the $|g_{1,2}(\mathrm{FD})| > 0.02$ range. This improvement allows the shear catalog to include a much larger number of galaxies, thus enhancing its statistical power.

Fig. 7 shows that although random forest obtained slightly smaller PPS residuals compared to the polynomial case, both methods perform equally well in predicting the PPS model at galaxy positions. This results demonstrate the efficacy of the FD test in evaluating the PSF reconstruction accuracy directly on galaxy positions.

4.5. *Further improvements*

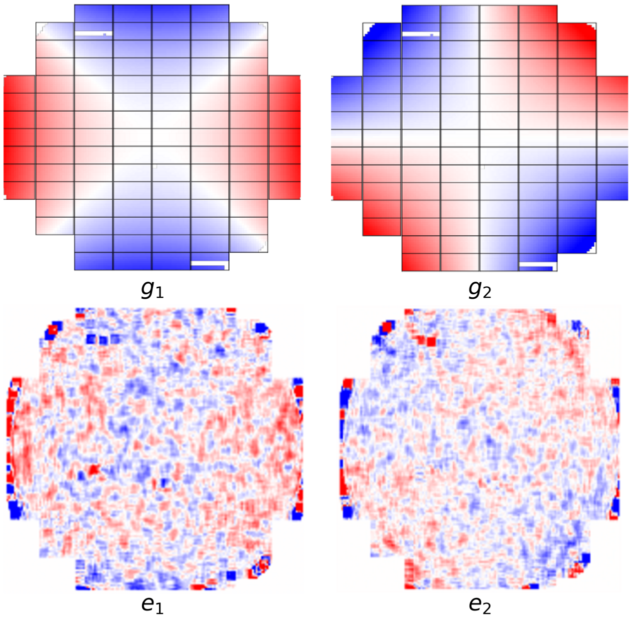

Figure 8. Field distortion signals (g_1, and g_2, top panels) and smoothed PPS residuals (e_1, and e_2, bottom panels).

After smoothing the PPS residuals shown fig.5, we find an interesting phenomenon: the PPS residuals show a very small but systematic slope, similar to that shown by the field distortion induced shear (FDS). Fig.8 displays the FDS (g_1 and g_2), and the ellipticity residuals (e_1 and e_2), which show a clear correlation. This correlation was not observed earlier mainly because this effect is too small and it was covered by the noisy PPS residuals signal.

We perform a third interpolation that aims to correct for this correlation. As we observe in fig.8, the systematic PPS residual is now global, thus we perform a single interpolation on the scale of the whole exposure.

As in the second layer of interpolation, we first rescale the PPS residual stamps—obtained after the second interpolation—to a reference PSF FWHM of 1 arcsec, in real space. We place them into a single exposure and apply principal component analysis (§4.2) to the PPS residual stamps in entire exposure. We use 100 principal components in our analysis. To find this number, we evaluate the quality of our PPS model for different number of PCs, using the field distortion test. Fig. 9 shows the correlation between the PPS residuals (e_1 and e_2) and the FD signal (g_1 and g_2) for different number of PCs (10, 20, 30, 50, and 100). From 10 to 50 PCs, we observe an clear improvement as the number of PCs increase, with the results stabilizing after 50 PCs, without a noticeable improvement. Based on these results, we conclude that a minimum of 50 PCs is necessary to capture all the features of the PPS residuals. However, it is important to note that these results might vary depending on the specific dataset, thus we recommend maintaining the number of PCs above 50, ideally 100, to ensure optimal performance.

Using all available data, each set of PC coefficients is fit to a polynomial of third order, building a model for the PC coefficient. These models are used to predict the PC coefficients at different positions on the exposure. Next, a model for the PPS residuals is built for each coordinate on the exposure as a weight sum of the PCs, weighted by the predicted PC coefficient at that position. Lastly, we rescale the PPS residual model back to its original PSF size, as in §4.3. This PPS residuals model is added to the PPS model after the second interpolation, building our final PPS model.

Fig. 10 compares the PPS residuals with the field distortion shears after the second and third interpolations. Although we observe a slight improvement after the third interpolation, this is not quite significant. We have explicitly tested various polynomial orders for our third interpolation, ranging from 2 to 6. The third order demonstrated the best performance. In addition, we have also tested our interpolation using random forest, but in this case it performs significantly worse than polynomial, likely due to overfitting.

We summarize all the steps of our PSF reconstruction scheme in a flowchart shown in fig. 11.

5. CONCLUSION

Point spread function (PSF) reconstruction is a crucial step towards accurate shear measurements, as it directly affects the observed galaxy morphology. Although current PSF reconstruction methods can capture most of

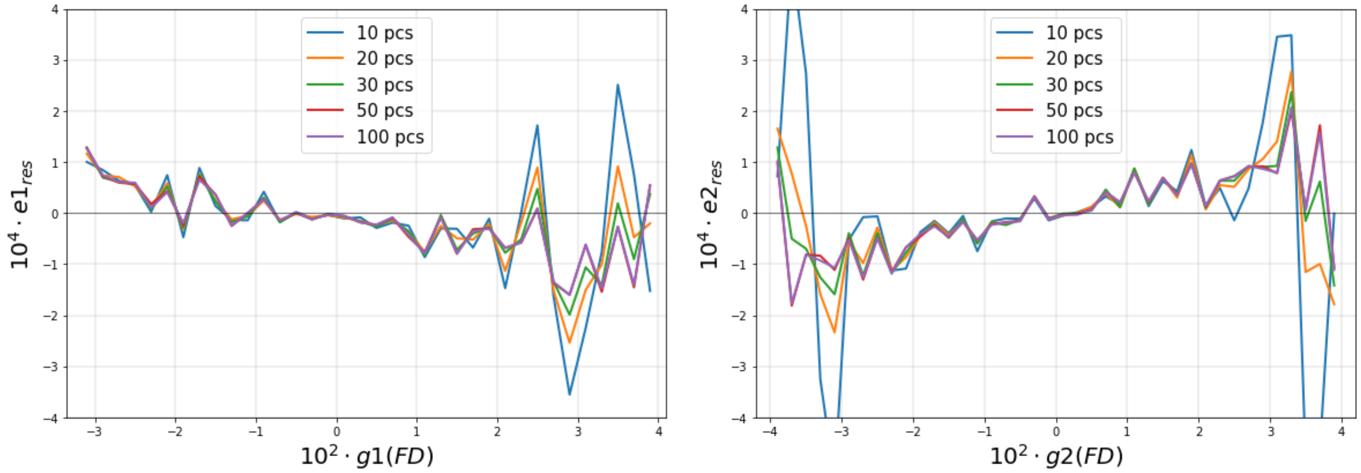

Figure 9. PPS residuals versus field distortion signal at the position of the stars. Different curves represent different number of PCs. The left panel shows the results for e_1 vs g_1 and the right panel shows the results for e_2 vs g_2.

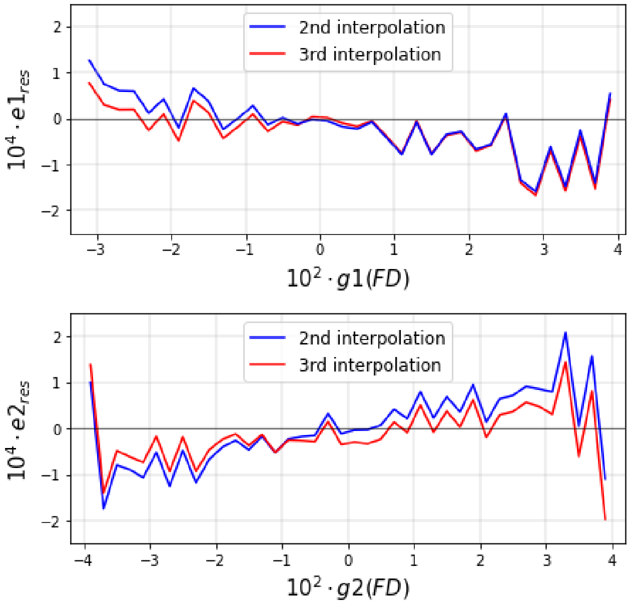

Figure 10. PPS residuals versus field distortion signal at the position of the stars. We compare the results after the second interpolation (blue curves) and after the third interpolation (red curves). The top panel shows the results for e_1 vs g_1 and the bottom panel shows the results for e_2 vs g_2.

the PSF features, there are often some PSF residuals remaining that limit the accuracy of, e.g., shear measurements. It is therefore essential to develop new techniques that can further improve current PSF reconstruction methods. In this work we introduce a novel PSF reconstruction method composed by three layers of interpolation, following a hierarchical scheme. Using data from the second data release of the Hyper Suprime-Cam (HSC DR2) as an example, our method significantly reduces the systematic PSF residuals, improving the PSF model.

The first layer of interpolation (§3.2) is a 2D polynomial fitting of third order to the power spectrum of the stars, pixel-by-pixel, on each CCD separately. This interpolation captures most of the spatial features of the PSF, building our initial PSF model.

The second layer of interpolation, described in §4, is the most important and novel part of this work. In this layer we model the systematic PSF residuals remaining after the first interpolation, using the data of all the exposures simultaneously. Note that to do so, we need to first uniform the image sizes of the PSF residuals. We apply PCA to the PSF residuals and build a model for each principal component coefficient. We then build the PSF residual model as a weighted sum of the principal components, with the weights being the predicted PC coefficients. To make corrections on the PSF model, we rescale the PSF residuals model back to the original PSF size of each exposure and add it to the original PSF model. Fig. 5 and fig. 6 show the PSF residuals of our improved PSF model, for polynomial and random forest interpolations, respectively. In both cases our method successfully removes most of the systematic features of the PSF residuals. In addition, we test our model directly on real galaxies, using the field distortion test (§4.4). We obtain significant improvements compared to the results of the first interpolation. Fig. 7 shows the field distortion test results for the cases of polynomial and random forest interpolations. Both interpolations have comparable performances, leading to very small shear biases.

In §4.5 we study a correlation between the PSF residuals and the field distortion signal, which is found after

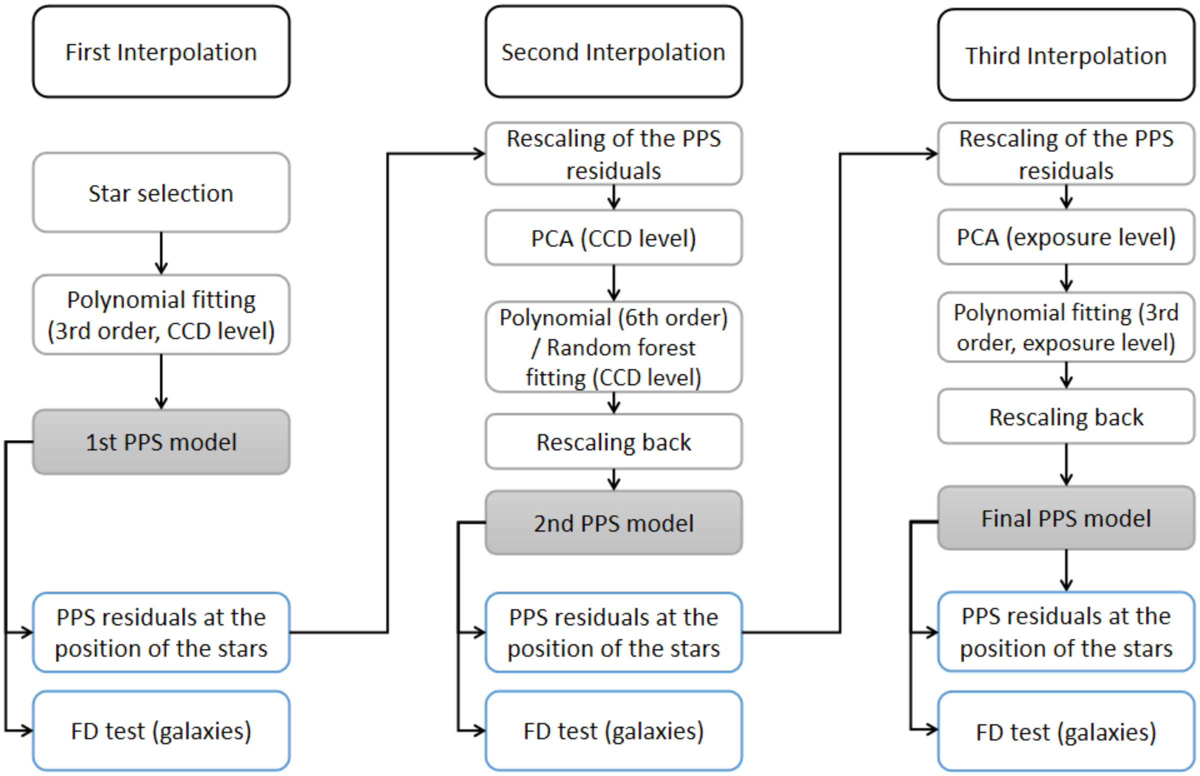
First Interpolation
Star selection
Polynomial fitting (3rd order, CCD level)
1st PPS model
PPS residuals at the position of the stars
FD test (galaxies)
Second Interpolation
Rescaling of the PPS residuals
PCA (CCD level)
Polynomial (6th order) / Random forest fitting (CCD level)
Rescaling back
2nd PPS model
PPS residuals at the position of the stars
FD test (galaxies)
Third Interpolation
Rescaling of the PPS residuals
PCA (exposure level)
Polynomial fitting (3rd order, exposure level)
Rescaling back
Final PPS model
PPS residuals at the position of the stars
FD test (galaxies)

Figure 11. Structure of our PSF reconstruction method, including three layers of interpolation.

smoothing the PPS residuals after the second interpolation (see fig.8). In this case, the systematic features are global, hence we perform a single interpolation on the exposure level, making use of all available exposures together. As in the second interpolation, we first rescale the PSF residuals to a common size. Then, we apply PCA to the entire data, and build a model for each PC coefficient. Our new PSF residuals model is built as a weighted sum of the principal components, with the weights given by the predicted PC coefficient at each position on the CCD. Lastly, we rescale the PSF residuals model back to the original PSF size of each exposure, and add it to the PSF model. Fig.10 presents the PPS residuals versus the field distortion shears, showing a small improvement after our third interpolation, although not quite significant. Nevertheless, this effect is minor and does not significantly impact the PSF model.

Overall, our model introduces a way to model the systematic PSF residuals, successfully removing most of the systematic PSF residuals in HSC DR2. In the current framework, our machine learning approach (random forest) performs similarly to polynomial, in the second interpolation. More exotic machine and deep learning algorithms may further improve the interpolation, reducing the PSF residuals.

In conclusion, our method aims to be a stepping stone towards building better PSF models, reducing the systematic biases induced by the PSF and helping produce more accurate and reliable shear catalogs, which are essential to the understanding of the distribution of dark matter, galaxy evolution, and to constrain cosmological parameters.

6. ACKNOWLEDGEMENTS

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan. The Subaru Telescope is honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

JZ is supported by the National Key Basic Research and Development Program of China (2023YFA1607800, 2023YFA1607802), the NSFC grants (11621303, 11890691, 12073017), and the science research grants from China Manned Space Project (No. CMS-CSST-2021-A01). The computations in this paper were run on the π 2.0 cluster supported by the Center of High Performance Computing at Shanghai Jiaotong University, and the Gravity supercomputer of the Astronomy Department, Shanghai Jiaotong University.

REFERENCES

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4, doi: 10.1093/pasj/psx066

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114, doi: 10.1093/pasj/psz103

Bernstein, G. M., & Armstrong, R. 2014, MNRAS, 438, 1880, doi: 10.1093/mnras/stt2326

Bernstein, G. M., Armstrong, R., Krawiec, C., & March, M. C. 2016, MNRAS, 459, 4467, doi: 10.1093/mnras/stw879

Bosch, J., Armstrong, R., Bickerton, S., et al. 2017, Publications of the Astronomical Society of Japan, 70, S5, doi: 10.1093/pasj/psx080

Breiman, L. 2001, Machine Learning, 45, 5. https://api.semanticscholar.org/CorpusID:89141

Fétick, R. J. L., Neichel, B., Mugnier, L. M., Montmerle-Bonnefois, A., & Fusco, T. 2018, MNRAS, 481, 5210, doi: 10.1093/mnras/sty2595

Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, S3, doi: 10.1093/pasj/psx079

Hébert, C.-A., Macintosh, B., & Burchat, P. R. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10700, Ground-based and Airborne Telescopes VII, ed. H. K. Marshall & J. Spyromilio, 107005E, doi: 10.1117/12.2314311

Herbel, J., Kacprzak, T., Amara, A., Refregier, A., & Lucchi, A. 2018, JCAP, 2018, 054, doi: 10.1088/1475-7516/2018/07/054

High, F. W., Rhodes, J., Massey, R., & Ellis, R. 2007, PASP, 119, 1295, doi: 10.1086/523112

Hirata, C. M., Yamamoto, M., Laliotis, K., et al. 2024, Monthly Notices of the Royal Astronomical Society, 528, 2533, doi: 10.1093/mnras/stae182

Ho, T. K. 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 1, 278–282 vol.1, doi: 10.1109/ICDAR.1995.598994

Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, ApJ, 504, 636, doi: 10.1086/306102

Jarvis, M., & Jain, B. 2004, Principal Component Analysis of PSF Variation in Weak Lensing Surveys. https://arxiv.org/abs/astro-ph/0412234

Jarvis, M., Schechter, P., & Jain, B. 2008, arXiv e-prints, arXiv:0810.0027, doi: 10.48550/arXiv.0810.0027

Jarvis, M., Bernstein, G. M., Amon, A., et al. 2020, Monthly Notices of the Royal Astronomical Society, 501, 1282, doi: 10.1093/mnras/staa3679

Jia, P., Wu, X., Yi, H., Cai, B., & Cai, D. 2020, AJ, 159, 183, doi: 10.3847/1538-3881/ab7b79

Kaiser, N. 2000, ApJ, 537, 555, doi: 10.1086/309041

Kaiser, N., Squires, G., & Broadhurst, T. 1995, ApJ, 449, 460, doi: 10.1086/176071

Kannawadi, A., Rosenberg, E., & Hoekstra, H. 2021, MNRAS, 502, 4048, doi: 10.1093/mnras/stab211

Kim, E. J., & Brunner, R. J. 2017, MNRAS, 464, 4463, doi: 10.1093/mnras/stw2672

Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, PASJ, 70, S2, doi: 10.1093/pasj/psx069

Li, H., & Zhang, J. 2021, ApJ, 911, 115, doi: 10.3847/1538-4357/abec6d

Liu, Q., Er, X., Wei, C., et al. 2023, Research in Astronomy and Astrophysics, 23, 075021, doi: 10.1088/1674-4527/acd589

Lu, T., Zhang, J., Dong, F., et al. 2017, AJ, 153, 197, doi: 10.3847/1538-3881/aa661e

Luppino, G. A., & Kaiser, N. 1997, ApJ, 475, 20, doi: 10.1086/303508

Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858, doi: 10.1093/mnras/sts454

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1, doi: 10.1093/pasj/psx063

Pollo, A., Rybka, P., & Takeuchi, T. T. 2010, A&A, 514, A3, doi: 10.1051/0004-6361/200913428

Rhodes, J. D., Massey, R. J., Albert, J., et al. 2007, ApJS, 172, 203, doi: 10.1086/516592

Roddier, F. 1981, Progess in Optics, 19, 281, doi: 10.1016/S0079-6638(08)70204-X

Shen, Z., Zhang, J., Li, H., et al. 2022, AJ, 164, 214, doi: 10.3847/1538-3881/ac8ff9

Shlens, J. 2014, arXiv e-prints, arXiv:1404.1100, doi: 10.48550/arXiv.1404.1100

Slater, C. T., Ivezić, Ž., & Lupton, R. H. 2020, AJ, 159, 65, doi: 10.3847/1538-3881/ab6166

Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, AJ, 141, 189, doi: 10.1088/0004-6256/141/6/189

Xin, B., Ivezić, Ž., Lupton, R. H., et al. 2018, AJ, 156, 222, doi: 10.3847/1538-3881/aae316

Zhang, J. 2010, MNRAS, 403, 673, doi: 10.1111/j.1365-2966.2009.16168.x

Zhang, J., Liu, C., Vaquero, P. A., et al. 2022, The Astronomical Journal, 164, 128, doi: 10.3847/1538-3881/ac84d8

Zhang, J., Luo, W., & Foucaud, S. 2015, JCAP, 2015, 024, doi: 10.1088/1475-7516/2015/01/024

Zhang, J., Dong, F., Li, H., et al. 2019, ApJ, 875, 48, doi: 10.3847/1538-4357/ab1080