Referee report

The paper "Characterization and Photometric Performance of the Hyper Suprime-Cam Software Pipeline" (Song Hung et al.) describes the photometric performance of the HSC pipeline hscPipe by injecting synthetic stars and galaxies into real HSC images. This paper is a sort of the supplement of hscPipe, specific to evaluate the performance of hscPipe for SSP science. Therefore, this report should be appeared in the hscPipe technical paper or in the appendix of scientific papers with hscPipe applied to SSP. Although the results are described in detail, the discussion seems to be poor in terms of new achievements or ideas of software technology. It lacks the discussion on the versatility for general purposes. In order to make this paper more useful for general users of SynPipe and worth publishing as a full paper of PASJ, the referee proposes the revise of the paper as mentioned below.

Major comments:

1. The section of "Discussion" is missing.

First, summarize the main results, and then discuss on what follows.
(1) Details of new achievement and idea of software technology should be explained.
(2) The discussion on what are the superior points of SynPipe to existing pipelines should be added, by comparing with existing pipelines.
(3) The errors originated from hscPipe and SynPipe should be discussed, by comparing with pure statistical noise (Poisson noise of targets and sky, readout noise, etc.).
(4) Show the strong and weak points of SynPipe. It would be very useful for users of SynPipe to study stars or galaxies with HSC. Is this SynPipe applicable to other imaging data, e.g., of SuprimeCam, MegaCam, or specific to HSC data?

2. Interpreter language

The authors stress the usage of Python. However, such shell script languages are generally much slower than compile languages. Especially HSC images are so large that the analysis should be cost effective. In fact, in Section 4.3, the paper demonstrates the expensive CPU cost. General users of SynPipe cannot afford 108 cores (in p.15). There seem no merits in Python-based.

On the other hand, the bulk of the calculations in GalSim are carried out in C++ to save the CPU cost, while the user interface is in Python. I wonder if most parts of SynPipe can be compiled with C or C++ to save the CPU cost and to make the analyses much faster. Alternatively, show why this paper uses shell script language (Python) in all process instead of compile languages? In this regards, "the overall efficiency of SynPipe is similar to the real data reduction process" is stressed in p.12. Please show how efficient it is, regardless of the slow shell script language.

3. The definitions are missing.

(1) magnitude

The magnitude, e.g., i~25 mag, is useless for this kind of statistical performance test. It has different meaning between those on the images with e.g., 10-minute and 20-minute exposures.

The difference of limiting magnitudes and S/Ns among g, r, i, z can be known from the observed or implemented photon number of objects and sky. (Although "5 sigma at 26.4 mag of SSP Wide layer" is written, there is no description or definition of the sigma). The authors should show the conversion factor of magnitude to photon number for e.g., 25 mag.

Sky background noise, dark noise, and readout noise of the images used in the paper should be listed. Therefore, the readers will know the pure statistical Poisson or random errors of the model stars and galaxies and differentiate the errors originated in the analyses of SyncPipe and hscPipe.

(2) S/N

What are the Flux_PSF and Flux_ErrPSF used for S/N? No definition is described in the paper, or give the reference.

(3) usePixelWeight=TRUE or FALSE (text in p. 33 and Fig. 17 in p.32)

What is "the pre-pixel variance information is used". The details should be noted.

4. cModel photometry

The good performance of cModel photometry is a natural consequence of this kind of pipelines, because the paper injects the model galaxies of GalSim in the real image, while hscPipe is also uses GalSim model galaxies for cModel photometry. Therefore, the present paper just confirms the "internal" consistency.

In general, software packages like hscPipe are optimized to achieve the maximum performance to reproduce the parameters of the model stars or galaxies, which are internally generated. Therefore, such an internal consistency is a minimum requirement of his sort of pipelines. In fact, the construction scheme of model galaxies are not always the same. One of the example is the value of the central pixel of galaxy images.

The values of the central pixel of Sersic profile should be carefully defined, because the central pixel value is not the value of Sersic profile at r=0, where the value is very large for large Sersic index. The pixel value should be the average over the central pixel. Another example is the Gauss profile at outer edge where the slope is so steep that the value of pixel center is not the average of the pixel value.

The values are defined in many ways (e.g., artdata in IRAF, GALFIT, NoiseChisel by Akhlaghi and Ichikawa 2015). The authors should show the performance of the present pipeline with the model galaxies defined with other packages.

5. Fig.4 and 9

From the input photon number of model stars, sky background, readout noise, (and dark), the S/N in terms of random statistical errors can be calculated. The S/N location of pure noises should be depicted in (a) i-band and (b) g-band, so that readers can identify the difference between S/Ns obtained in hscPipe and pure statistical noises.

6. Pure statistical errors in figures.

As noted above, pure statistical random errors (combined with Poisson noises of stars/galaxies and sky background, dark and readout noise) should be drown also in Figs. 5, 6, 7, 10, 11, 12, and 15 as a function of magnitude, so that readers can differentiate magnitude errors obtained in hscPipe from the pure statistical noise.

7. Precision and accuracy of PSF magnitude in p. 18

Show the pure statistical errors along with the performance of hscPipe. If the errors obtained with hscPipe and SynPipe are much larger than pure statistical errors, discuss the origin of the difference.

8. p. 7-8

"Unlike most publically available photometric pipeline, it involves high-level reduction process"

What is the "high-level reduction process"?

9. Purity and completeness.

This paper uses the stars and galaxies detected with hscPipe with good photometric quality. Those with bad quality, non-detection, false detection, and misclassification (star or galaxy), which are not shown in the figures in the present paper, should be studied in terms of the purity and completeness. Show the purity and completeness in new figures for stars and galaxies in g (the highest fraction of misclassification) and i (highest success rate) as a function of magnitude.

Minor comments:

1. p.7

"and the astrometry is accurate to ~10 and 40 max internally and externally"

What is the unit of 10 and 40? Is it "mas", not "max"?

2. p.10

"SyncPipe generates a rectangular cutout image of the PSF model"

Why is a rectangular cutout used? Stars are circular and galaxies are more or less circular, so that models should be in circular cutout. The slope at the outer edge is very steep, so that model fit with rectangular cutout sometimes give wrong results for parameters, e.g., r_e (effective radius) of galaxies with large Sersic index and small r_e.

3. p. 11

"SynPipe adds appropriate noise to the image"

What is meant by "appropriate"? Poisson noise? Show it more clearly.

4. p. 13 Fig 3. caption

(lower panel) from left to right are the Sersic...

-> (lower panel) from left to right are the input Sersic...

Alternatively, the abscissas label of the figures of lower panel should be "Input Sersic Index, Input Axis Ratio... " etc.

5. Definition of highly blended objects

Fig. 4 (p. 14) Highly blended stars

-> Highly blended stars with b> 0.05

Fig. 5 (p. 16) Highly blended objects

-> Highly blended objects with b> 0.05

6. p. 22

"Fig 8 displays the precision of PSF colors...."

The stars at reddest and bluest i-z and z-y colors are not successfully recovered by hscPipe. On the other hand, Fig. 13 for model galaxies shows no such tendency. Any comments are needed.

7. p.22

"Also, for blended stars .... However, such biases appear to be less ...."

Why blended stars show a bias bluer in g-i and redder in i-y? The bias seems to be very large. Why is it "less sever"? In the first paragraph of this subsection (5.1.3) "PSF photometry is the most appropriate way to measure colors for point sources" The large bias for blended stars seems to be very sever. Any comments are needed.

8. p.29
"while overestimating the (i-y) color at the very red end."

No figure for (i-y) is depicted. Is it (i-z)?

9. Astrometric Calibration, Fig. 14

The unit of arcsec is meaningless without the pixel scale. In the text and caption of Fig. 14, the pixel scale should be added, or use the pixel number instead of arcsec.

10. the definition of highly blended (p.33)

In figs. 5, 10 and in summary, "the highly blended" is defined as those with b> 0.05. But the first paragraph in p. 33 defines it with 0.1 for galaxies. It would be confusing. In addition, Fig. 16 depicts b=0.1 for both stars and galaxies with vertical dashed lines. To avoid the confusion, it should be at b=0.05 (log b= -0.3).

11. Tables 1, 2, 3, and 4
"based on the statistics of the difference between output ..."
"is described using the statistical uncertainties of ..."

What is meant by "statistics" and "statistical"? In Fig. 5 and 10, for example, the running-median is used. If "statistics" is median, give more clear definition. "Statistics" is too indefinite.