The James Web Space Telescope

Inquiry to the measurement of the z index of galaxies by JWST compared to previous results

Jack Gale

A dissertation presented for the degree of
Bachelor of Physics (Honours)

School of Science
Royal Melbourne Institute of Technology
Australia
29 August 2025

Declaration of Originality

The work contained in this dissertation is original. The work is a review of all readily available literature on the topic of redshift and the methods used by HST and JWST to categorise these values. The review, summary and conclusion are all original work. All references were found via literature search by the author. All figures and tables which have been reproduced from these references have been cited appropriately. Where a specific wording was incorporated into the dissertation, this was indicated as a quote and cited appropriately.

Jack Gale 29/08/2025

Contents

1 Introduction

When observing objects in space, we see a window to the past. Since light travels at a finite speed, astronomic objects as we see them are never in their current state, but a state of when they emitted the light we observe. Thus, we spend a great deal of time observing distant objects for their distance in time, rather than their distance in space from us. Our observable universe is limited to the age of the universe and how far light has travelled in that time, with the most distant objects being the oldest[1]. Since Hubble and Lemaître, accepted theory is that of the expanding universe and the Big Bang cosmology[2]. The foundations of the theory lay in the observation of redshift; light waves from distant objects appear lengthened. Lemaître argued this was due to an expansion of space between the emitter and our observation of the light, thus explaining Hubble's constant, H_0. Redshift can be measured using equation 1, and is a fundamental measure of all observations. Redshift is thus a good indicator of distance, due to the dependence of the expanding space the transmitting redshifted light travels through.

$$z = \frac{\lambda_{obs}}{\lambda_{em}} - 1 \tag{1}$$

While the Cosmological Principle assumes a homogeneous and isotropic universe, which is backed up by most observations of the universe, [3] suggests that the our local galactic neighbourhood may influence measurements of fundamental cosmological properties, such as H_0. These properties influence all cosmological models, since we must account for the expansion of the universe when determining the distance implied by such redshift values. With this in mind, it is more important than ever to make true measurements of physical properties of cosmological observations, such as redshift. Our capabilities of observation have dramatically increased since Hubble and Lemaître, prominently with the advent of space-based telescopes. Space-based telescopes have the advantage of being above the Earth's atmosphere, which obscures and distorts light prior to capture in the lens. Certain wavelengths never reach the ground, as well as motion blur caused by the moving gases in the atmosphere. The two most prominent space telescopes of our time are the Hubble Space Telescope (HST) and the James Web Space Telescope (JWST).

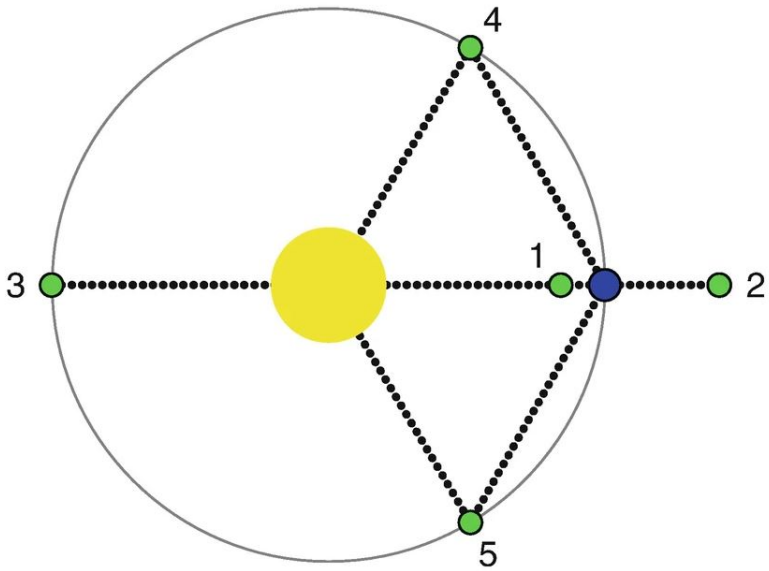

Figure 1: The 5 Lagrange points anchored by Earth's and the Sun's combined gravitational pull (from [4]).

HST was launched in 1990, while JWST was launched very recently, in 2021; both are still in active operation today. The HST is in Low Earth Orbit (LEO) at an altitude of 515 km, and has taken some 1.6 million observations over its lifetime for over 108 million astronomical objects. It has a primary mirror diameter of 2.4 m, with an ultraviolet-visible light resolution of 0.04 arcseconds per pixel, and an infrared light resolution of 0.13 arcseconds per pixel. The HST has a pointing accuracy of 0.007 arcseconds[5]. The HST's visual range is smaller than the JWST, with only small overlap. The HST is designed to see wavelengths of 90 to 2,500 nm, while the JWST from 600 to 28,500 nm. The JWST is at the L2 point, illustrated in figure 1, a gravitationally stationary point in solar orbit, in line with Earth's own solar orbit. These points can gravitationally bind small masses with respect to the orbits of the two large bodies[4]. The JWST has a primary mirror diameter of 6.5 m, with a resolution of 0.07 arcseconds[6].

1.1 Redshift Evaluation Methods

Observation of the universe is delicate business, which must involve models, approximations, and understanding of phenomena which may distort results. Measured redshift values may be affected by local Doppler effects from our reference frames movement (Earth's spin, orbit, and solar system trajectory) and from the emitter itself. Since redshift, or the "z" value, of systems is directly measurable it has widespread use; being dimensionless, the z-index can easily be used to measure time *or* distance. However, in order to apply distances or age values to a z-value, cosmological parameters, Doppler "peculiar velocities", and other influences must be considered. Further, the z-value is highly dependent on the method of observation - spectroscopic or photometric. With both methods, the filtering of foreground objects must also be undertaken in order to isolate a source, especially in the case of high-z objects[1].

Spectroscopic techniques of redshift measurement are perhaps the most direct and reliable, but are dependent on very specific sources. The spectroscopic technique takes the observed spectra of an object and compares it to similar known spectra. Objects which are particularly useful for such tasks are called "standard-candles". Type II supernovae have characteristic luminosity and can be used to compare redshift when observed, given their known and uniform appearance[7].

Photometric techniques have also been developed which can be used to measure the redshift of observed cosmological objects. This method instead relies on the observed flux ratios in different filter bands of observed objects. Photometric methods rely on many bands and narrow filters, which intrinsically limit its application to observations with long exposure times. These flux ratios are again compared against template spectra[1].

For the performance assessment we use some standard metrics. First, the difference between redshift values must be normalised, due to the differences growing in proportion to the distance. Equation 2 normalises the difference.

$$\Delta z_{norm} = \frac{z_{jwst} - z_{hst}}{1 + z_{hst}} \tag{2}$$

Then, the median absolute deviation (σNMAD) is defined in equation 3; this value represents the median of the absolute difference between two datasets value of the redshift. σNMAD is a good indicator which is not representative of the outliers of a study, but rather the datasets as a whole. For this study, z_{phot} and z_{spec} are replaced by z_{HST} and z_{JWST} respectively. The overall redshift bias is defined by the median of Δz_{norm}.

$$\sigma NMAD = 1.48 \times |\Delta z_{norm} - \tilde{\Delta z_{norm}}| \tag{3}$$

Equation 4 defines a quantitive measure for which redshift values should be treated as outliers[8].

$$|\Delta z| > 0.15(1 + z_{spec}) \tag{4}$$

This dissertation aims to evaluate the technologies behind the HST and the JWST and how they are leveraged to obtain redshift values for near and distant cosmological objects. The two space telescopes will be evaluated against each other using a common dataset to determine if there are significant differences in the measured values of redshift. In doing so, insight will be gained in how significantly previous results may have been biased due to the measured redshift value at the time.

2 COSMOS2020 and COSMOS2025

The Cosmic Evolution Survey (COSMOS) catalogue was used as a basis for the comparison between the HST and the JWST, owing to its large dataset and comparable results. COSMOS has been conducted multiple times, over 5 year intervals. As such, the COSMOS2020 catalogue was compiled using primarily the HST[9], while the COSMOS2025 was able to leverage the JWST[10]. COSMOS covers 2 square degrees of the sky, around JS2000 RA +150.11916667, DEC +2.20583333 (in the constellation Sextants). This area was chosen for its lack of near Milky Way objects. These catalogues are essential in the setting of references for redshift templates, which are used to calibrate and measure the redshift of new sources.

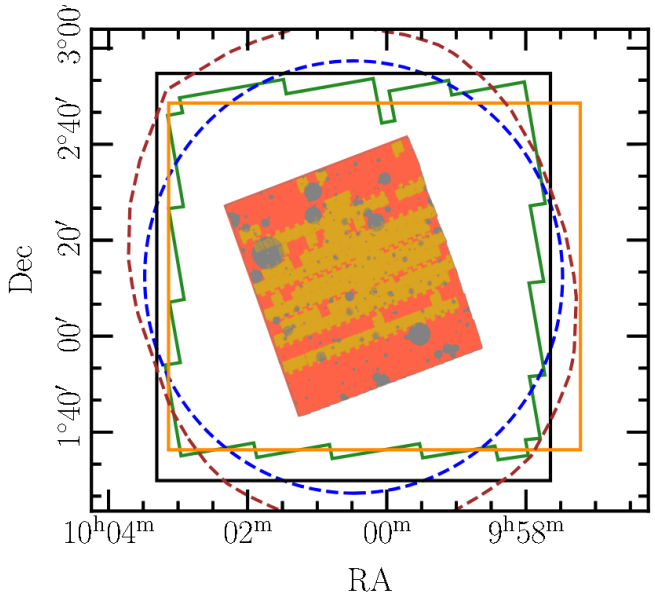

Figure 2: Survey footprint of the COSMOS2025 catalogue, with grey regions being bright star masks (from [10]).

The COSMOS2020 catalogue uses photometric techniques to derive z values, using LePhare[11] to fit the fluxes to z values[9]. The COSMOS2025 catalogue also used the LePhare fitting code, but had the fitting data expanded to include new multi-colour coverage from JWST NIRCam imaging. This expanded fitting allows for the colour-redshifted space to be used without the heightened risk of degeneration as faced by previous catalogues. Further, both catalogues have dust-fitting models applied which consider dust clouds which obscure our view of the COSMOS region[10]. Figure

2 illustrates the COSMOS2025 catalogue footprint and highlights the regions which have extra filters applied which counteract bright sources in the field.

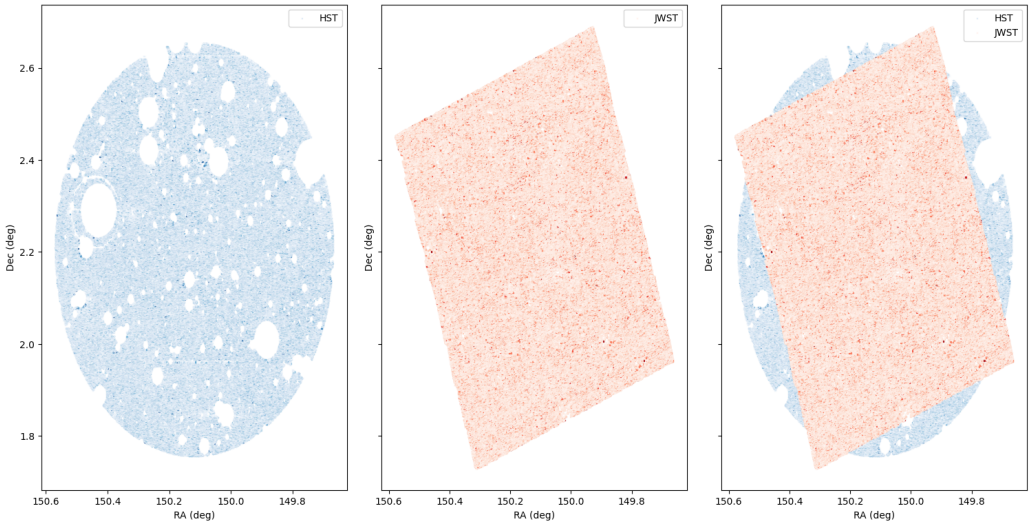

Figure 3: HST COSMOS2020 and JWST COSMOS2025 catalogues and their overlap.

The datasets which were used in this study are illustrated in figure 3. The HST based dataset utilised the COSMOS2020 catalogue, while the JWST based utilised the COSMOS2025 catalogue. The COSMOS2020 catalogue was retrieved using Python's Astroquery library. Astroquery is a set of tools for querying astronomical web forms and databases[12]. However, the COSMOS2025 catalogue was not available using the same service, and had to be downloaded and parsed from the data release[10]. This study required only the *photom_fits* and *lephare* libraries. With these datasets a comparable set of galaxies was available to gauge the z-indexing methods together.

Both datasets were sufficiently large, being in the millions of sources each. As such, they were parsed using multiprocessing techniques. Appendix A includes code snippets which highlight the multiprocessing used to extract and parse the datasets using python. The grey masked regions present in the COSMOS2025 catalogue are not present in the COSMOS2020 catalogue, which instead eliminates those regions altogether. Figure 3 illustrates the regions of the COSMOS2020 catalogue which are extracted due to nearby bright sources, which matches the COSMOS2025 grey masked regions. The two datasets used have sufficient overlap as shown by figure 3.

3 Difference Between Datasets

This study only considered galaxies, since they would be moving with sufficient z-values. From the datasets considered, only 784,016 sources from the COSMOS2025 catalogue were valid for assessment with the 3,611,107 sources from the COSMOS2020 catalogue. Each dataset was split up into multiple subsets, which were evaluated, cleaned, and parsed separately, before being recombined. The python multiprocessing toolbox allowed different processors to handle these tasks simultaneously. From this, 261,498 sources were matched with a tolerance of 0.3 arcseconds; if multiple matches were made, those with the smallest distance prevailed. Statistical analysis was not as computationally expensive as the initial parsing of the data, since the COSMOS2020 and COSMOS2025 catalogues had been matched and were reduced significantly in size. This analysis was completed on the entire matched dataset, such that median values were representative of the whole.

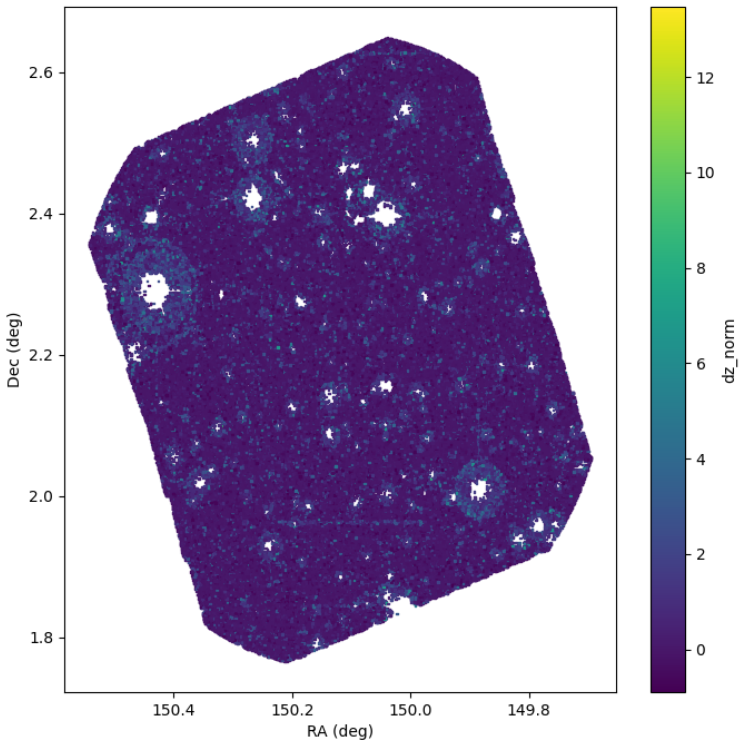

Figure 4: Normalized difference of redshift value between matched datasets.

Visualising equation 2, figure 4 highlights the areas which are affected by significant variations of the reported z-index between the two datasets. Across the range of sources, the normalised difference generally shows good agreement. However, significant disagreement can be seen in the regions close to the masked nearby sources.

Table 1: Statistical results of the difference between COSMOS2020 and COSMOS2025.

Measure	Value	Benchmark Value
Bias	9.2779e-3	0.05
$\sigma NMAD$	0.10378	0.05
Outliers	35.692%	15%

Such significant disagreement is reflected $\sigma NMAD$ (equation 3) and the number of outliers (equation 4) present in the sample. Table 1 highlights the results of the study against the benchmark values expectant of large catalogues. The bias of the comparison is much lower than the benchmark value by an order of magnitude. This suggests an overall well matched dataset, with little to no systematic errors. This indicates that the matching algorithm used was robust and that the calibration between the two datasets was conducted well. However, being a median, the bias doesn't indicate the presence of any outliers, nor their significance. While there are expected to be some outliers, due to the JWST's extended capabilities over the HST, the number of outliers is source for concern. With over a third of the sample falling out of the normalised outlier range, some systematic difference between the COSMOS2020 and COSMOS2025 catalogues is suggested. The masked regions illustrated in figure 2 and present in figure 4 are

candidate for the difference. This is further highlighted with the $\sigma NMAD$ value being over double the benchmark. While the benchmark is flexible, being even tighter for more focussed datasets (ones with exclusively low or high z sources), the 0.05 benchmark is representative for studies with a wide range of redshift values. Our calculated value being double suggests that there is not a strong 1:1 correlation between COSMOS2020 and COSMOS2025.

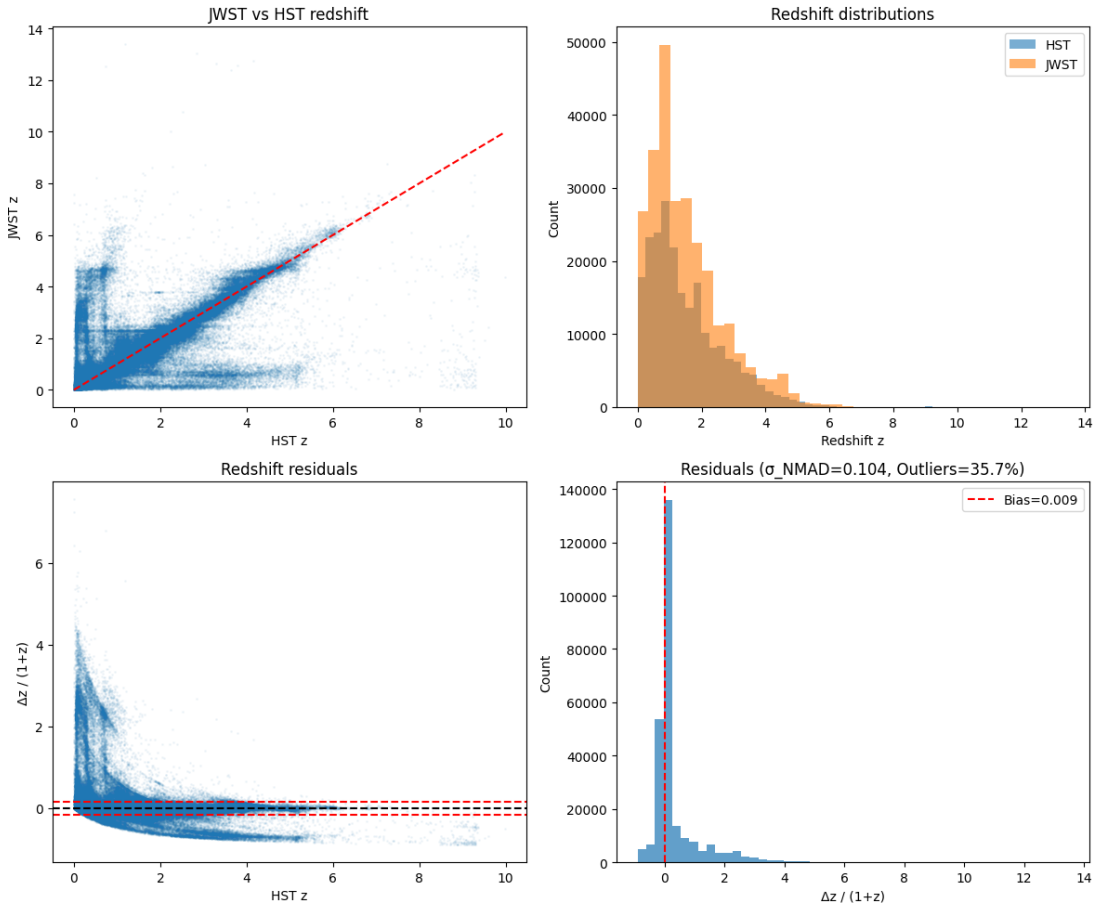

Figure 5: Redshift statistics of the matched datasets.

The difference between the two catalogues is not statistical noise, as proven by the low bias value. Figure 5 shows good correlation between COSMOS2020 and COSMOS2025, as illustrated by the strong tendency to lie on the "matched" line. The spread about this line also illustrates the $\sigma NMAD$ value, confirming the two datasets z value distribution failing to match with good tolerance. The new catalogue reevaluates many sources across both low and high z sources. Most sources evaluated were low z ($z < 3$), while there were still a significant amount of high z ($z > 3$) sources for evaluation. Figure 5 also illustrates the redshift residuals between the two datasets and can be used to probe the cause of the outliers. Many sources lie on the black line, which suggests agreement. The sources which lie between the red line are within the tolerance bounds, while those are representative of the outliers. There are many positive outliers at low z, with the JWST enhanced COSMOS2025 catalogue reporting much higher values for z. The imaging equipment has a much greater range than the HST, which may explain why low z galaxies, which may have previously been obscured and distorted by interstellar dust clouds, have been reevaluated has much higher values. Further, the negative outliers at high z suggest an extreme overestimation by the COSMOS2020 catalogue, likely caused by some generous overestimation of faint sources by the old catalogue and photometric model.

4 Conclusion

The investigation into the difference between the measured redshift values for galaxies in the COSMOS catalogues has yielded statistically significant results. 35.692% of identical catalogue objects have significant measured redshift values, with the spread, measured by the $\sigma NMAD$ function, is more than double the expected result. These results suggest that there are significant differences between the COSMOS2020 and COSMOS2025 catalogues. The cause of these differences can be attributed to new techniques being implemented for the photometric measurement of galactic redshift values, and an improvement in telescopic technology, namely the introduction of JWST data into the catalogue. While many differences in redshift for galaxies have been found, predominantly in the low and high redshift regions, the most concerning is the difference in redshift around regions affected by near-object filters (see figure 4).

Redshift catalogues are essential in current research with the z-index being an integral measure of many galactic properties, including H_0, a fundamental unit which drives expansion and cosmological models. These results illustrate how sensitive any model can be, and highlight that any model and results must be made with care and attention to the source and bias of their data. This report recommends that any models and assumptions made with previous COSMOS catalogues be remade with the current COSMOS2025 catalogue, owing to the increased capability of the JWST, with consideration of the data included in the masked regions.

References

[1] I. Appenzeller, *High-Redshift Galaxies: Light from the Early Universe*. Springer Berlin Heidelberg, 2009, ISBN: 9783540758242. DOI: 10.1007/978-3-540-75824-2.

[2] G. Lemaître, "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques," *Annales de la Société Scientifique de Bruxelles*, vol. 47, pp. 49–59, Jan. 1927.

[3] A. Pandya, K. Migkas, T. H. Reiprich, *et al.*, "Examining the local universe isotropy with galaxy cluster velocity dispersion scaling relations," *Astronomy & Astrophysics*, vol. 691, A355, Nov. 2024, ISSN: 1432-0746. DOI: 10.1051/0004-6361/202451755.

[4] D. Rouan, "Lagrangian points," in *Encyclopedia of Astrobiology*, M. Gargaud, W. M. Irvine, R. Amils, *et al.*, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 1356–1357, ISBN: 978-3-662-44185-5. DOI: 10.1007/978-3-662-44185-5_858.

[5] *Hubble by the Numbers - NASA Science — science.nasa.gov*, https://science.nasa.gov/mission/hubble/overview/hubble-by-the-numbers/, [Accessed 25-08-2025].

[6] S. T. S. Institue, *Web stats*, https://webbtelescope.org/contents/articles/webb-stats, [Accessed 25-08-2025], Dec. 2021.

[7] M. Dainotti, *Gamma-ray Burst Correlations* (2053-2563). IOP Publishing, 2019, ISBN: 978-0-7503-1575-3. DOI: 10.1088/2053-2563/aae15c.

[8] H. Hildebrandt, T. Erben, K. Kuijken, *et al.*, "Cfhtlens: Improving the quality of photometric redshifts with precision photometry*," *Monthly Notices of the Royal Astronomical Society*, vol. 421, no. 3, pp. 2355–2367, Apr. 2012, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.20468.x.

[9] J. R. Weaver, O. B. Kauffmann, O. Ilbert, *et al.*, "COSMOS2020: A Panchromatic View of the Universe to z∼10 from Two Complementary Catalogs," *apjs*, vol. 258, no. 1, 11, p. 11, Jan. 2022. DOI: 10.3847/1538-4365/ac3078. arXiv: 2110.13923 [astro-ph.GA].

[10] M. Shuntov, H. B. Akins, L. Paquereau, *et al.*, *Cosmos2025: The cosmos-web galaxy catalog of photometry, morphology, redshifts, and physical parameters from jwst, hst, and ground-based imaging*, 2025. arXiv: 2506.03243 [astro-ph.GA]. [Online]. Available: https://arxiv.org/abs/2506.03243.

[11] S. Arnouts, L. Moscardini, E. Vanzella, *et al.*, "Measuring the redshift evolution of clustering: the Hubble Deep Field South," *mnras*, vol. 329, no. 2, pp. 355–366, Jan. 2002. DOI: 10.1046/j.1365-8711.2002.04988.x. arXiv: astro-ph/0109453 [astro-ph].

[12] A. Ginsburg, B. M. Sipőcz, C. E. Brasseur, *et al.*, "astroquery: An Astronomical Web-querying Package in Python," *aj*, vol. 157, 98, p. 98, Mar. 2019. DOI: 10.3847/1538-3881/aafc33. arXiv: 1901.04520 [astro-ph.IM].

Appendices

A Python code snippets of the multiprocessing features used to parse large amounts of data

Listing 1: Retrieving the JWST and HST data using multiprocessing

```
manager = mp.Manager()
return_dict = manager.dict()

p1 = mp.Process(target=get_jwst_data, args=(0,return_dict))
p2 = mp.Process(target=get_hst_data, args=(1,return_dict))
processes = [p1, p2]

print("Starting data collection...")
for p in processes:
    p.start()
print("Joining...")
for p in processes:
    p.join()

jwst    = return_dict[0]
hst     = return_dict[1]
```

Listing 2: Retrieval and cleaning of the JWST data using multiprocessing

```
def get_jwst_data(i, return_dict):
    print("Retrieving JWST photom data...")
    with fits.open('COSMOSWeb_mastercatalog_v1_photom_primary.fits') as photom:
        #photom.info()
        ra_jwst = photom[1].data['ra']
        dec_jwst = photom[1].data['dec']
        no_warn = photom[1].data['warn_flag'] # should be 0
        no_faint = np.abs(photom[1].data['mag_model_f444w']) # should be <30
        star_mask = photom[1].data['flag_star_hsc'] # should be 0
    print("Retrieving JWST lephare data...")
    with fits.open('COSMOSWeb_mastercatalog_v1_lephare.fits') as lephare:
        #lephare.info()
        z_jwst = lephare[1].data['zfinal']
        is_galax = lephare[1].data['type'] # should be 0 for galaxy

    print("Building JWST table...")
    jwst = Table({'ra': ra_jwst, 'dec': dec_jwst, 'z_jwst': z_jwst, 'is_galax': is_galax, 'no_warn': no_warn, 'no_faint': no_faint, 'star_mask': star_mask})

    # Cleaning the data to only include robust galaxies
    print("Cleaing JWST data...")
    tot_rows = len(jwst)
    nprocessors = 6
    temp_manager = mp.Manager()
    temp_return_dict = temp_manager.dict()

```

```
    processes = []
    last_row_end = 0
    for j in np.arange(nprocessors):
                row_start = ceil(j * tot_rows / nprocessors)
                row_end   = floor((j + 1) * tot_rows / nprocessors
                    )
                if row_start != last_row_end:
                    row_start =last_row_end
                last_row_end = row_end
                print("Starting rows ", row_start, " to ", row_end
                    )
                p = mp.Process(target=clean_jwst_data, args=(j,
                    temp_return_dict, jwst[row_start:row_end]))
                p.start()
                processes.append(p)

    print("Ready to join clean JWST data...")
    for p in processes:
        p.join()

    print("JWST data built and ready!")

    return_dict[i] = jwst
```

Listing 3: Calculating statistics using multiprocessing

```
nprocessors = 8
manager = mp.Manager()
return_dict = manager.dict()

processes = []
last_row_end = 0
for j in np.arange(nprocessors):
            row_start = ceil(j * tot_rows / nprocessors)
            row_end   = floor((j + 1) * tot_rows / nprocessors)
            if row_start != last_row_end:
                row_start = last_row_end
            last_row_end = row_end
            print("Starting cols : [", row_start, ",", row_end, "]
                ")
            p = mp.Process(target=find_dif, args=(matches[
                row_start:row_end],j,return_dict))
            p.start()
            processes.append(p)

print("Ready to join...")
for p in processes:
    p.join()

dz_parts = []
print("Concatonating statistics...")
for j in np.arange(nprocessors):
    print("Processor ", j, "complete!")
    dz_parts.append(return_dict[j])
dz_norm = np.concatenate([np.asarray(part) for part in dz_parts])
```