D3.6.1: Impact of observational effects on LSST photo-z

Work Package 3.6

Project Acronym	LUSC-B
Project Title	UK Involvement in the Legacy Survey of Space and Time
Document Number	LUSC-B-34

Submission date	24/09/2022
Version	1
Status	Final
Author(s) inc. institutional affiliation	Qianjun Hang, Benjamin Joachimi (University College London)
Reviewer(s)	Shahab Joudaki (University of Waterloo), Markus M. Rau (Argonne National Laboratory)

Dissemination level	
Public	

Version History

Version	Date	Comments, Changes, Status	Authors, Contributors, Reviewers
0.1	11/10/2022	See deliverable review report	Shahab Joudaki, Markus M. Rau
1.0	18/11/2022	Final following Exec Group approval	Qianjun Hang, Benjamin Joachimi

Table of Contents

List of Figures

List of Tables

1. Executive Summary

We have investigated how different observational effects such as sky brightness, seeing, and number of exposures can affect the photometric redshift (photo-z) distribution for LSST. The observation strategy for Rubin is to cover a large survey region before building up the depth. During the first few years of observation, therefore, it is expected that the inhomogeneity in depth due to e.g. varying weather condition is large. Effectively, one can regard each pointing as a 'mini' survey with different observational systematics and limiting magnitudes. This could potentially be a problem for weak lensing analysis because the signal is sensitive to the mean redshift of the tomographic bin, especially for the precision required by LSST.
In this investigation, we focus on the 'gold' sample from the first (Y1) and fifth year (Y5) data release, for which the simulated observation conditions for the Rubin Observatory (OpSim) as well as the DC2 DM catalogue are used. We split the sample into tomographic binning for lens sample between 0.1 in $0.2 \leq z \leq 1.2$ according to Y1 and Y5 requirements respectively. One of the main aims is to check whether the spatially varying observing conditions introduce fluctuations to the mean and scatter of each tomographic bins that are larger than the Rubin requirement. The photo-z is estimated using a template-fitting algorithm, `BPZ_lite`. We find that for the sample with reduced photo-z outliers, the shifts in the mean redshift and the scatter in each tomographic bin is consistent with the random noise of the sample, and comparable to the Rubin requirement.
Additionally, we also looked at the impact of spatially varying observational conditions in each band on the cModel magnitude, the magnitude error, colour, and galaxy over-density.

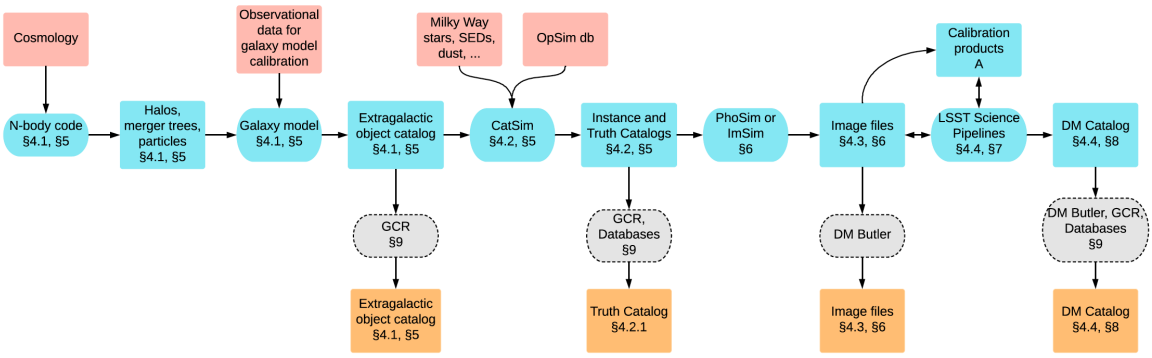


Figure 1: Workflow for the DC2 simulation, adopted from [1]. Opsim db, the Truth Catalogue, and the DM Catalogue are used here.

2. Introduction

In this note, we investigate how different observational effects such as sky brightness, seeing, and number of exposures can affect the photometric redshift (photo-z) distribution for LSST. The observation strategy for Rubin is to cover a large survey region before building up the depth. During the first few years of observation, therefore, it is expected that the inhomogeneity in depth due to e.g. varying weather condition is large. Effectively, one can regard each pointing as a 'mini' survey with different observational systemtaics and limiting magnitudes. The change in limiting magnitude can result in shifts in the estimated photometric redshift distribution, as shown in Fig.8 of [2] for the KiDS-1000 data as the r-band limiting magnitude is varied between $r_{\rm lim} = [25.9, 26.6]$. This could potentially be a problem for weak lensing analysis because the signal is sensitive to the mean redshift of the tomographic bin, especially for the precision required by LSST. [3] showed that the inferred cosmological parameter from cosmic shear analysis can shift up to $\sim 1\sigma$ for a KiDS-15000 like survey, which roughly accounts for the increased survey area of LSST (see their Fig.5).
In this note, we focus on the 'gold' sample from the first (Y1) and fifth year (Y5) data release, for which the simulated observation conditions for the Rubin Observatory (OpSim) as well as the DC2 DM catalogue [1] are used (see Fig. 1). The required photo-z quality is for Stage IV cosmological analysis is listed in LSST DESC Science Requirements Document [4]:

- For Y10 weak lensing analysis, the systematic uncertainty in the mean redshift of each source tomographic bin should not exceed $\langle z \rangle = 0.001(1+z)$; the scatter should not exceed $\sigma_z = 0.003(1+z)$;
- For Y1 weak lensing analysis, the similar requirements are $\langle z \rangle = 0.002(1+z)$ and $\sigma_z = 0.006(1+z)$;
- For Y10 LSS analysis, the systematic uncertainty in the mean redshift of each tomographic bin should not exceed $\langle z \rangle = 0.003(1+z)$; the scatter should not exceed $\sigma_z = 0.03(1+z)$;
- For Y1 LSS analysis, the similar requirements are $\langle z \rangle = 0.005(1+z)$ and $\sigma_z = 0.1(1+z)$.

We follow the assumed tomographic binning for lens sample in [4]: for Y10, we adopt 10 tomographic bins spaced by 0.1 in $0.2 \leq z \leq 1.2$, and for Y5, we adopt 5 bins spaced by 0.2 in the same redshift range. In all cases, we use the Y10 requirement as reference for the Y5 data. One of the main aim is to check whether the spatially varying observing conditions introduce fluctuations to the mean and scatter of each tomographic bins that are larger than the Rubin requirement. Additionally, we look at the impact of spatially varying observational conditions in each band on the cModel magnitude, the magnitude error, colour, and galaxy overdensity.
The the python codes and Jupyter notebooks used can be accessed in the Github repo: https://github.com/lsst-uk/DC2-observation-systematics/tree/main/.

3. Simulated Observations

3.1. maf maps

We generate observation property maps using `minion_1016`[1] which is the OpSim input for the DC2 DM catalogue. Notice that this version adopts a DC2-specific dither pattern. When accessing, `descDitheredRA`, `descDitheredDec`, and `descDitheredRotTelPos` should be used. Because this is a rather old version of OpSim, there are several known issues with this simulation, including:

- sky model is not up-to-date;
- weather conditions are too optimistic;
- all the SCP and galactic plane observations get executed in the first year (and often higher airmass).
- some observational effects are simulated in OpSim but not implemented in the DM catalogue. For example, the airmass is in the DM catalogue fixed to $X = 1.2$ across all bands.

One of the results of these issues is that the first year seeing conditions are much better compared to others (also noted in [5]).
Summary statistics are obtained from the OpSim database using the Rubin Observatory Metrics Analysis Framework (MAF). We look at the following observational conditions in each of the *ugrizy* band:

`airmass, dist2Moon, fiveSigmaDepth, filtSkyBrightness, FWHMeff, CoaddM5.`

For items that have a pointing history (e.g., all of the above except `CoaddM5`), we obtain the median, mean, maximum, and minimum values over the survey footprint with an observing period of 1 year (`nights < 365`) and 5 years (`nights < 1825`) respectively. We apply the 'healpix slicer' which pixellize the survey footprint into a healpix map with nside = 128.[2]. In practice, we use the median maf maps.
In Fig. 2 and 3, we show the spatial variation of the median values for Y1 and Y5 Rubin observational conditions over the DC2 WFD footprint (The 'missing' bit at the upper right corner of the footprint is the deep drilling field). The first row shows the case for all bands combined and the subsequent rows show the case for each of the *ugrizy* bands.

3.2. supreme maps

A set of survey property maps called the `supreme maps`[3] are available for the DC2 dr6 catalogue. These maps are generated directly from the coadd information, excluding any clipped regions during the the coaddition process. The maps are stored in the `HealSparse`[4] format, which compress the original high resolution healpix map with `nside_sparse`= 32768 to efficient storing size. We down-grade these maps to nside= 128 for comparison with the maf maps and for this analysis. Fig. 11 shows a comparison between the median airmass map from MAF and the coadd supreme map for 5-year observation. It can be seen that they are not exactly the same in terms of spatial variation as well as the specific values, but they are highly correlated in each band.

[1] The database can be accessed on NERSC:
`/global/cfs/cdirs/descssim/DC2/minion_1016_desc_dithered_v4_trimmed.db`
[2] The code for how these maps are generated is included: `/notebooks/DESC_DC2_minion_1016.ipynb` in the Github repo. The original codes are written by Lynne Jones.
[3] See https://confluence.slac.stanford.edu/display/LSSTDESC/DC2+Run2.2i+DR6+Survey+Property+Maps for a detailed description. To access on NERSC:
`/global/cfs/cdirs/lsst/shared/DC2-prod/Run2.2i/addons/supreme/dr6-wfd/`
[4] https://github.com/LSSTDESC/healsparse

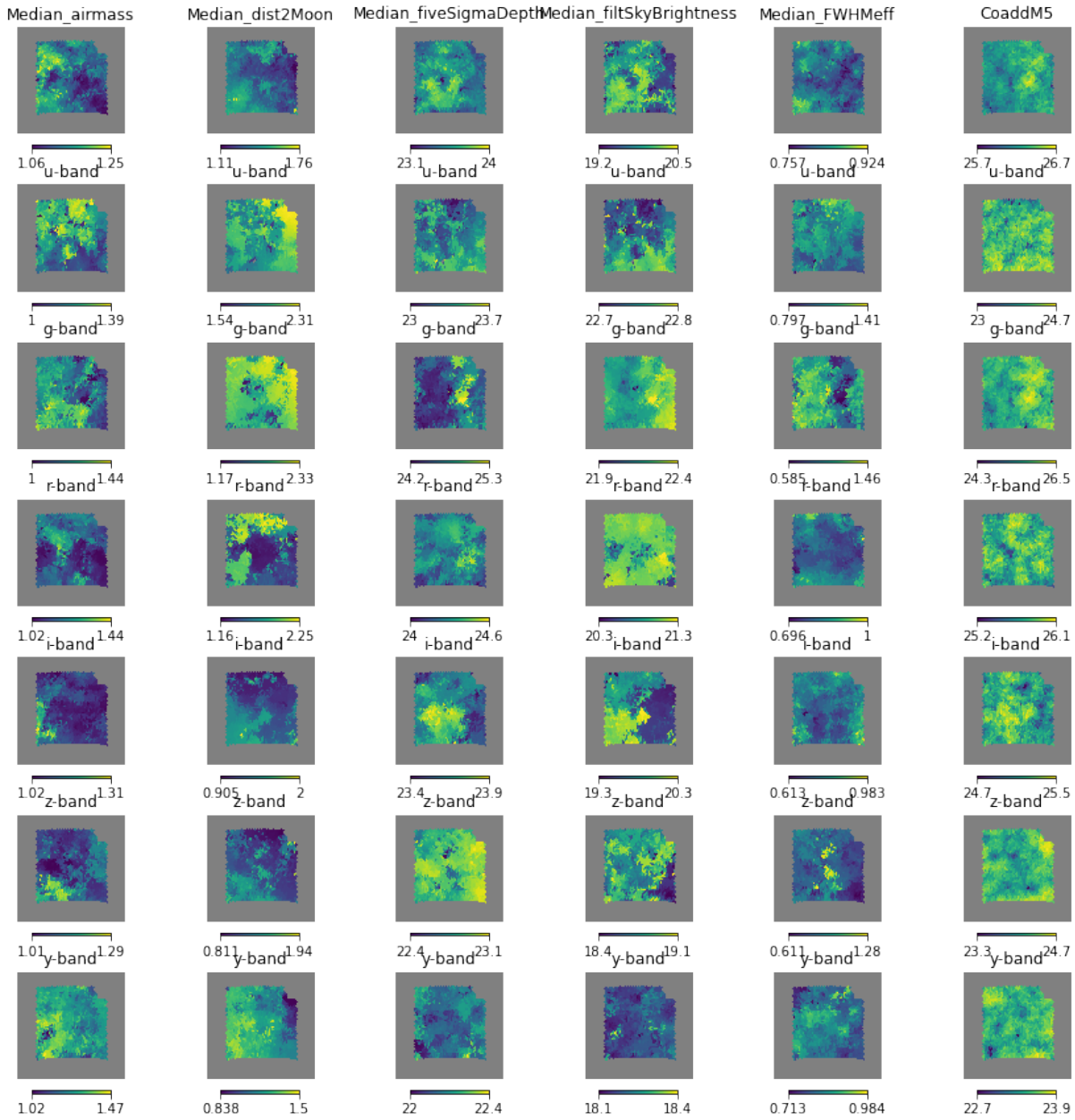


Figure 2: The median observational variables from `minion_1016` for 1-year observation in each photometry band.

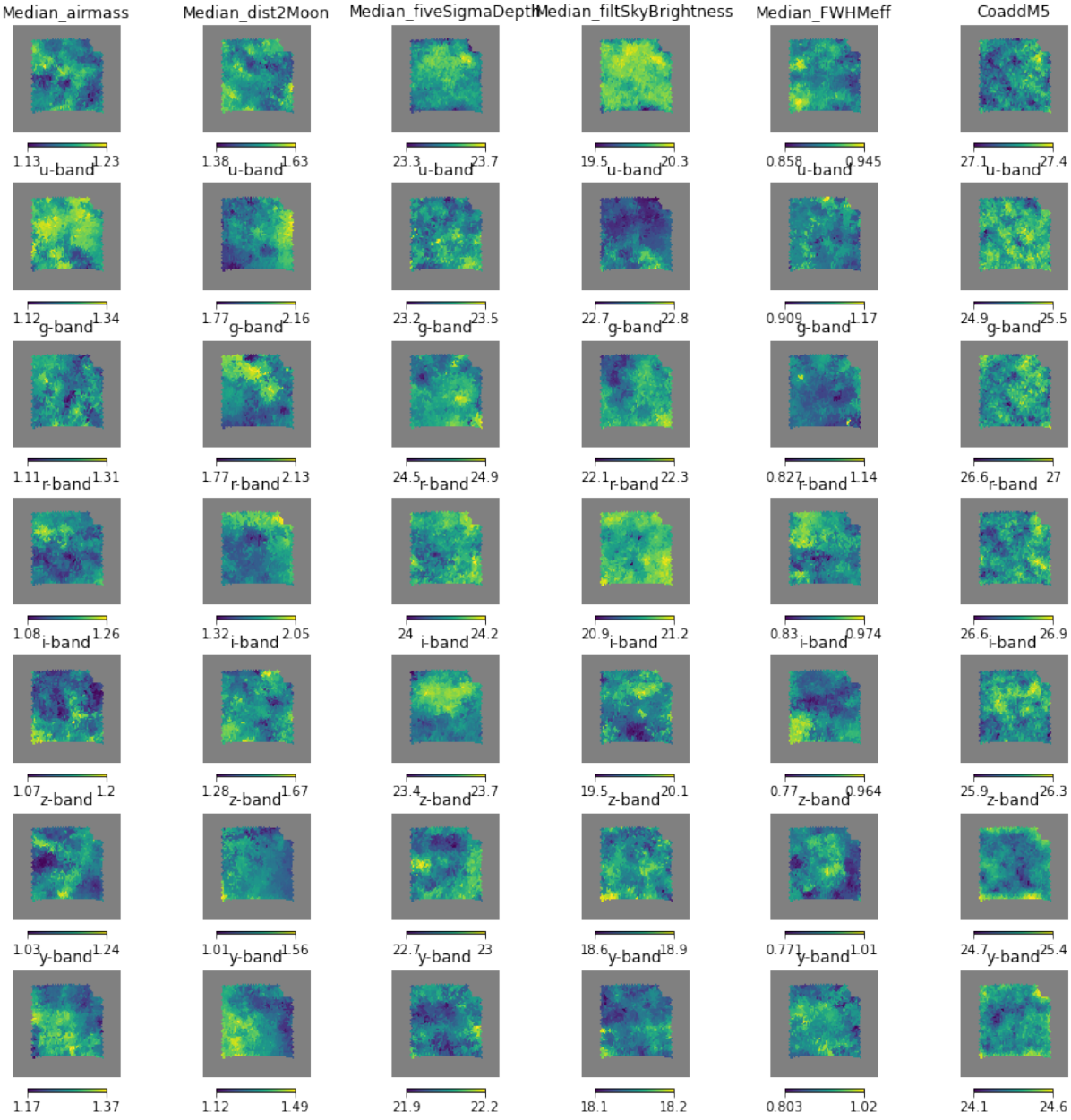


Figure 3: Same as Fig. 2 but for 5-year observation.

In principle, these maps should correspond more directly to the DM catalogue, because the DM catalogue is also derived from the coadds. The available maps include:

`airmass, bgmean, boresight_dist, coadd_mask, dcr_[e1,e2], exptime, maglim_aper, maglim_psf, nexp, psf_area, psf_[e1,e2], psf_size`

We make use of the `nexp_sum` and `psf_size_wmean` map.

3.3. Other systematic maps

We also include the Milky Way dust extinction map $E(B-V)$ over the DC2 footprint. This map offers an different angle to the systematics we consider here: firstly, it should not correlate with atmospheric systematics because they arise from different origin; secondly, in terms photo-z, the extinction introduces bias in the measured colour, whereas the other effects mainly changes the limiting magnitude. To compare the errors, we apply de-reddening[5] to the magnitudes, which implements a list of pre-determined $A_\lambda/E(B-V)$ values for each band, and looks up the $E(B-V)$ value from a dust map for each individual RA and DEC. The correction is applied by subtracting

$$\left[\frac{A_\lambda}{E(B-V)}\right] E(B-V) \tag{1}$$

from the measured magnitudes.

3.4. Correlations between different observation variables

Different observational variables are not completely independent. We compute the correlation matrix between different variables (see Fig. 4). The covariance matrix is given by:

$$\hat{C}_{ij} = \frac{1}{N_{\rm pix}} \sum_{k=1}^{N_{\rm pix}} (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j), \tag{2}$$

where $N_{\rm pix}$ is the total number of pixels within the footprint, x_i^k is the value of variable i (e.g. u-band median airmass) at pixel k, and $\bar{x}_i$ is the mean value of the variable averaged over the footprint. We observe some expected trend in the covariance. For example, `coaddM5` is strongly correlated with the number of exposures and `fiveSigmaDepth` in each band respectively, and is negatively correlated with the seeing angle.
This may be useful for generating the correlated observing conditions in the future, which can be used in forward modeling and applied as a degrader for the photo-z package.

3.5. LSST error model

The magnitude error as a function of magnitude and observing conditions for Rubin is detailed in [6]. For convenience, we quote the relevant equations below. The magnitude error in a single visit is given by

$$\sigma^2 = \sigma_{sys}^2 + \sigma_{rand}^2, \tag{3}$$

where σ_{sys} is the systematic error which is required to be < 0.005 mag for the calibration system and procedure. They estimate the random error using SDSS as a reference [7]:

$$\sigma_{rand}^2 = (0.04-\gamma)x + \gamma x^2, \tag{4}$$

where $x \equiv 10^{0.4(m-m_5)}$. m_5 is the 5σ depth for point source, and γ depends on the system throughput. The m_5 depends on a set of observing conditions in the following way:

$$m_5 = C_m + 0.50(m_{sky}-21) + 2.5\log_{10}(0.7/\theta_{eff}) + 1.25\log_{10}(t_{vis}/30) - k_m(X-1), \tag{5}$$

[5] https://github.com/LSSTDESC/DC2-analysis/blob/master/contributed/DC2_object_deredden.ipynb

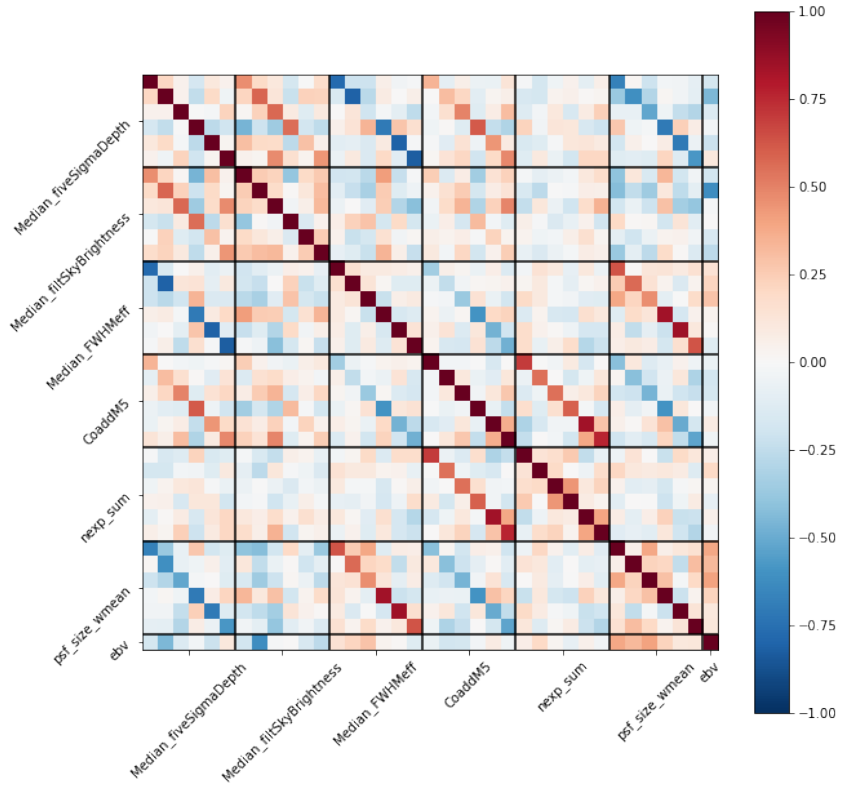


Figure 4: The correlation matrix between various median observational systematic maps for 5-year observation.

where C_m is a constant that depend on the overall throughput of the system, m_{sky} is the sky brightness in AB mag arcsec^{-2}, θ_{eff} is the seeing in arcsec, t_{vis} is the exposure time in seconds, k is the atmospheric extinction coefficient, and X is the airmass. The default values of the parameters in the above equation per band are given in Table 2 in [6]. Finally, the magnitude error for N-year observation is computed by σ/Nn_{vis}, where the mean number visit per year n_{vis} can be derived from Table 1 in [6].
Currently the module `LSSTErrorModel` in the `RAIL`[6] `degradation` package adopts this model to generate magnitude error using the default values in [6]. Notice that in the high SNR ratio limit, the error is implemented as a fraction of flux,

$$\sigma_m = 2.5 \log_{10}(1 + \sigma). \tag{6}$$

We have provided an additional module `ObsCond` which allows users to input these systematic maps (or the newer versions of OpSim) and compute the corresponding magnitude error depending on the assigned pixels.

4. DC2 DM catalogue

A detailed description of the DC2 DM catalogue can be found in [1]. The Extragalactic catalogue (`GCR`) is based on a N-body cosmological simulation with identified halo merger history and realistic galaxy models. `CatSim` provides observation conditions from the `OpSim` database, as well as foreground contamination such as Milky Way stars and dust. These effects are applied on simulated galaxy images via the `PhoSim` or `ImSim` pipeline. Finally, the simulated images are passed through the LSST Science Pipelines to generate the final data products. The the wide-fast-deep cadence (WFD) of DC2 covers about 300 deg^2 in RA $= [50, 70]$ and DEC $=$

[6] https://github.com/LSSTDESC/RAIL

$[-45, -27.5]$. The footprint is divided into 'tracts' of similar size. The Y1 and Y5 LSST mock data correspond to dr2[7] and dr6[8] respectively.

4.1. Sample selection

We select galaxies based on the criteria for the 'gold' sample; for Y10, this sample is defied to be $i < 25.3$. We work out the magnitude limit for the Y5 gold sample via two methods. One way defined in [4] is to use one magnitude shallower than the median 5-σ point source detection depth. Using the 5-year `CoaddM5` map, we find an i-band limit of $i = 25.1$. One can also take the SNR approach which roughly scales as $\sqrt{T}$, where T is the time of observation. The change in magnitude compared to Y10 is $2.5\log(\sqrt{10/T})$, where here $T = 5$ years. This gives a magnitude limit of $i = 24.9$. We thus adopt the mean value of $i < 25$ as the Y5 gold sample selection. Similarly for Y1, we adopt $i < 24$ as the cut for the gold sample.
In addition to the magnitude cut, we also apply a set of basic cuts adopted in weak lensing studies:

```
basic_cuts = [
GCRQuery('extendedness > 0'),      # Extended objects
GCRQuery((np.isfinite, 'mag_i')), # Select objects that have
                          # i-band magnitudes
GCRQuery('clean'), # The source has no flagged pixels
                 # (interpolated, saturated,
                 # edge, clipped...)
                 # and was not skipped by the deblender
GCRQuery('xy_flag == 0'), # Flag for bad centroid measurement
GCRQuery('ext_shapeHSM_HsmShapeRegauss_flag == 0'),
                  # Error code returned by shape
                  # measurement code
GCRQuery((np.isfinite, 'ext_shapeHSM_HsmShapeRegauss_sigma')),
                  # Shape measurement
                  # uncertainty should not be NaN
GCRQuery('mag_i_cModel < 25.3'), # Gold sample selection
]
```

4.2. Photo-z for dr6 and dr2

In addition, a photometric redshift catalogue is available for dr6 using `BPZPipe` (based on template fitting) and `FlexZPipe` (based on machine learning). Notice that the representative or training sample is only complete up to $i < 25.0$. For dr2, we select a 5% subsample to run the photo-z estimation using `BPZ_lite`. Notice that before running the estimator, all magnitudes need to be de-reddened. We notice also that the number of sample here may not be large enough to statistically show the level of fluctuation, after testing the effect of sample size by down-sampling the dr6 data. Two point estimates for a galaxy redshift is available: `photoz_mean` and `photoz_mode`. In practice we find that using `photoz_mean` gives a sightly smaller scatter compared to `photoz_mode`.
It should also be noted that the quality of the photo-z for the full gold sample is not yet at the Rubin requirement, as will be shown later. We can exclude the outliers by excluding non-detection, introducing a cut in `photoz_odds` > 0.9 (a measure of how compact the photo-z PDF is), as well as a cut in the template fitting χ^2: `photoz_mode_ml_red_chi2` < 2. We find that limiting to this higher quality photo-z sample does improve significantly the $\langle z \rangle$ and σ_z, as expected. However, the scatter introduced by different observational conditions is not changed significantly. We show an example case in Section 5.1.1.

[7] `dc2_object_run2.2i_dr2_wfd_with_addons`
[8] `dc2_object_run2.2i_dr6_with_addon`

Thus, here we focus on the *relative* shifts in $\langle z \rangle$ and σ_z in quantiles of given observing conditions, rather than the absolute values. In all cases, we use the middle quantile, q3, as the reference. It is beyond the scope of this note to investigate means to further reduce outlier rate in `BPZ_lite`, or to compare with other photo-z estimators unless they are given in the catalogue.
Notice also that the definition of $\langle \delta z \rangle = \langle z_{phot} - z_{true} \rangle$ is on the full distribution, and $\sigma^2_{\delta z} = \langle \delta z^2 \rangle - \langle \delta z \rangle^2$. Due to the non-Gaussian distribution of δz, it may be more desirable to use more robust measure, e.g., the interquartile range (IQR) as adopted in [8]. This makes a difference especially in the measure of $\sigma_{\delta z}$. Here we do not use IQR for computational convenience, but notice that this could increase $\sigma_{\delta z}$ by a factor of 1.5 or more compared to the IQR measure. The point here, however, is to look at the relative shifts due to different observing conditions, so the increase in the absolute values is not a main concern.

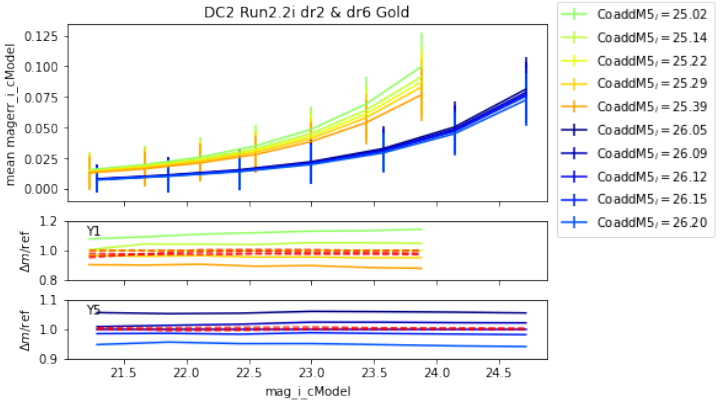
5. Impact of spatially varying observing conditions on photo-z

5.1. Variation in mean magnitude error, photo-z bias, and photo-z scatter

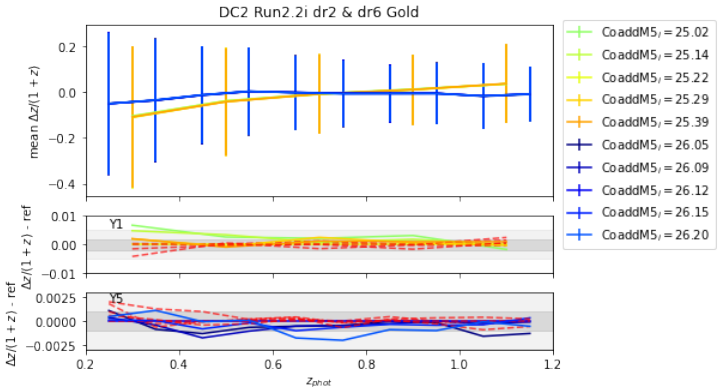
We show the variation in the i-band magnitude error σ_m, the photo-z bias $\langle (z_{phot} - z_{true})/(1 + z_{phot}) \rangle$, and the photo-z scatter $\sigma_{\delta z/(1+z)}$ from five quantiles of the i-band coadd 5σ depth based on the `CoaddM5` pixel values for both Y1 and Y5 in Fig. 5. This example is selected because i-band is the detection band, and the coadd depth is the dominant factor for the change in magnitude errors (and thus photo-z estimation). For Y1, the i-band coadd depth varies in range $m_5 = [25.02, 25.39]$, and for Y5, $m_5 = [26.05, 26.20]$ across the DC2 footprint. Fig. 5a shows the mean and scatter of the i-band magnitude error (`magerr_i_cModel`) as a function of the measured i-band magnitude (`mag_i_cModel`) in the top panel, and the bottom panel shows the ratio of the mean magnitude error and the reference line (the third quantile in each case). We see that the change in m_5 introduces a simple multiplicative factor to the magnitude errors, which is about 10% for Y5 and 20% for Y1 between the most distant quantiles (q1 and q5). This effect will be implemented in `ObsCond` in `RAIL`. Fig. 5b shows the mean and scatter of the photo-z bias - defined as the difference between z_{phot} and z_{true}, weighted by $1 + z_{phot}$ in a given photo-z bin in the top panel, and the deviation of the mean from the reference line is shown in the bottom panels. Notice that the absolute photo-z bias and scatter here is large due to outliers, especially at low redshift bins. The relative shifts between the different quantiles does not show a general trend; for Y5, the scatter is at the level of ± 0.002, and for Y1 it is ± 0.005, except at low redshift bins where the scatter can be ± 0.01. On the same plot, we also show a case in red dashed lines where the pixels are randomly chosen in each quantile, thus serves as a reference for random noise given the sample size. We see that the relatively larger scatter towards lower redshift is consistent with the random sample. This level of variation shown here is consistent with the Rubin requirements for LSS analysis (light grey band), but may not meet the requirements for weak lensing analysis (dark grey band; this is partially limited by the statistical uncertainty given the sample size). Fig. 5c shows the variation in photo-z scatter (same as the error bar in Fig. 5b) with the i-band depth. The relative shifts in scatter is about ± 0.005 for Y5, and ± 0.01 for Y1 except for a larger scatter at low redshift bins. For Y1, this seems to be consistent with the random noise (red dashed line), whereas for Y5, the scatter is slightly larger compared to the random sample. This level of fluctuation is negligible for LSS analysis (light grey band), but may be significant for the Y5 weak lensing analysis (dark grey band).

5.1.1. Effect of photo-z outliers

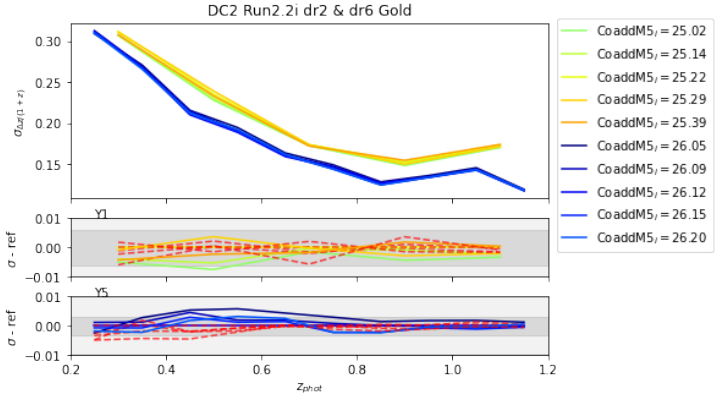
We apply a further selection on the photo-z quality here to make the sample more realistic. We use the Y5 data here because the cut in photo-z quality removes a significant proportion of the sample. First we remove objects that may exceed the observation limit: $u > 27.79$, $g > 29.04$, $r > 29.06$, $z > 27.98$, and $y > 27.05$. We also remove objects with very broad posterior or poor reduced χ^2, i.e., `photoz_odds` < 0.9 and `photoz_mode_ml_red_chi2` > 2. This removes 60% of objects, and the photo-z bias for the unbinned sample reaches $\langle (z_{phot} - z_{true})/(1 + z_{phot}) \rangle = 0.01$



(a) The i-band magnitude error as a function of the i-band cModel magnitude, split by five quantiles in the i-band coadd depth for Y1 and Y5 data.

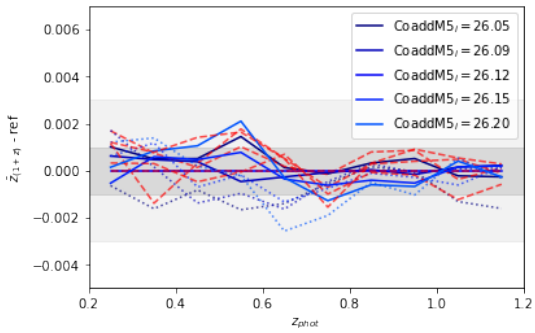


(b) Similar to Fig. 5a, but now showing the photo-z bias in each photo-z bin.

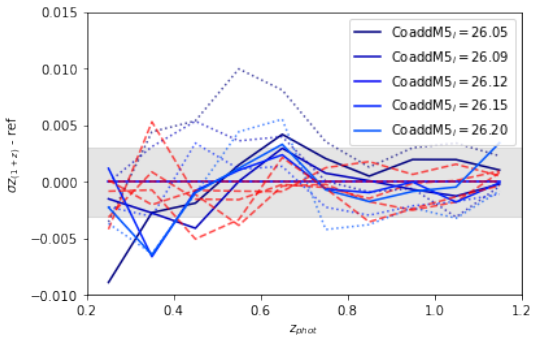


(c) Similar to Fig. 5a, but now showing the photo-z scatter in each photo-z bin.

Figure 5: Variation of magnitude error, photo-z bias, and photo-z scatter with i-band 5σ depth using the `coaddM5` maf map for Y1 and Y5 data in the DC2 simulation.



(a) The relative change in photo-z bias, defined as $\langle(z_{phot} - z_{true})/(1 + z_{phot})\rangle$, for the selected sample with less photo-z outliers (solid) and without selection (dotted). The darker grey band corresponds to the Rubin requirements for Y10 weak lensing, and the lighter grey band shows that for LSS analysis.



(b) The relative change in photo-z scatter, defined as the standard deviation of the photo-z bias, for the selected sample with less photo-z outliers (solid) and without selection (dotted). The darker grey band corresponds to the Rubin requirements for Y10 weak lensing.

Figure 6: Comparison of Y5 sub-samples with (solid) or without (dotted) cuts for photo-z outliers in terms of the relative change in photo-z bias and scatter, given five quantiles of the i-band coadd 5σ depth. The red dashed lines show the case where pixels are randomly selected in each quantile, indicating the level of random noise in this measurements.

with a scatter of 0.09. This still does not quite meet the Rubin requirement. More strict cuts have been applied but does not improve the bias or scatter further. It is a known issue with `BPZ` that the mean redshift is biased low, and may be corrected with spectroscopic calibration. It is beyond the scope of this note, however, to carry out the calibration or to investigate other photo-z estimators.

Fig. 6 shows the variation in photo-z bias and scatter with the above cuts to remove outliers (solid), to be compared with the sample without cuts (dotted). We further show in red dashed lines a case where the pixel selection is random in each 'quantile', in order to assess the level of statistical uncertainty. We see that for Y5, the difference is negligible for the photo-z bias, but improved for the photo-z scatter with reduced outliers. In both cases, excluding outliers results a level of fluctuation consistent with the randomised case, a level which is comparable to the Rubin requirements for Y10 weak lensing analysis. This means that a larger sample is needed to assess whether nor not the fluctuation is within the Rubin requirements.

5.1.2. Beyond i-band and other observational variables

In this section we look at the observing conditions in other bands. We take the example again in the coadd 5σ depth, and we include the results for other observing conditions in the Appendix. We look at three quantities using the Y5 data: the relative change in the mean magnitude error compared to the reference line (q3), the magnitude bias (defined as `mag_[...]_cModel - mag_[...]_truth`), and the relative change in the mean redshifts in 10 photo-z bins, defined as $(\langle z^{qi}_{true}\rangle - \langle z^{q3}_{true}\rangle)/(1 + \langle z^{q3}_{true}\rangle)$ for the i^{th} quantile. The scatter has been divided by a factor $\sqrt{N}$, where N is the number of galaxies in the bin.

Fig. 7 shows these quantities for five quantiles ('q1'-'q5') from the brightest 20%, to the faintest 20% `CoaddM5` pixels in each of the six LSST bands. In the top two rows, the magnitude measures are in the same band as the coadd map. We see that the trend in each band is similar. In the bottom row, the grey band marks ± 0.002, which gives a sense of the Rubin requirements. It is noticeable that in the middle row, the difference between measured magnitude and truth is as large as $0.5 - 1.0$ at the tails. The mean bias at the brighter end is surprisingly large, about -0.2 dex. This may be caused by the substantial tail towards the negative values, due

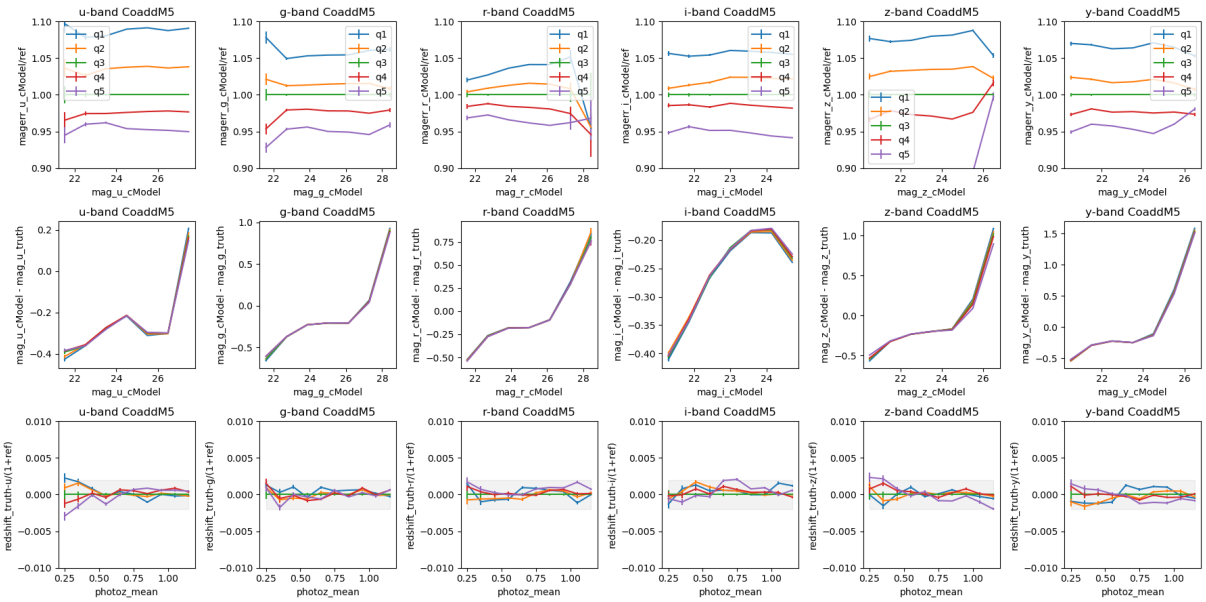


Figure 7: Variation of the mean magnitude error (cModel), the magnitude bias (measured magnitude minus the truth), and true redshift in bins of magnitude and photo-z, with three bins in `coaddM5` values, using the DC2 dr6 data. The different lines represent cases in 5 quantiles of the pixel values from the systematic maps. Different columns show results using systematics in different photometric bands.

to the observation magnitude limit. Finally looking at the shifts in the mean redshift from the photo-z bins, it seems that the level of variation consistent with the Rubin requirements, except that in the u-band, the lowest redshift bin is more affected. In Fig. 8 we show the same measurement for the Y1 subsample. Notice that the level of variation is now larger in all three quantities. However, due to the smaller sample, the measurement on mean redshift can be driven by statistical noise.

We also investigate whether the i-band depth has an impact on the measurements in other bands, due to the fact that i-band is used for detection. Fig. 9 shows this the response of other bands to the i-band coadd depth for Y5 data. We find that in terms of magnitude errors, most bands have a much smaller correlation with the i-band depth, at the level of $2-3\%$ at most. Here we also limit to a faint sample with $24.0 < i < 25.3$.

In Fig. 12 - 17 we show the same measurements for other observational conditions: `Median_fiveSigmaDepth`, `Median_filtSkyBrightness`, `Median_FWHMeff`, `nexp_sum`, `psf_size_wmean`, and `ebv`. We see that the trend is very similar for `coaddM5` and `nexp_sum`, giving the largest deviation in magnitude errors, and the lowest redshift bin can vary beyond 10% level with the u and i band systematics. For the other systematics considered here, the g band magnitude error is most stable against variation. For `Median_FWHMeff`, it seems that the i and z band are the most relevant for a shift in the mean redshift in the lowest bin. Fig. 18 - 20 show the response of other bands to the i-band median 5σ depth, median seeing, and total number of exposures respectively. We see again that the level of fluctuation in other bands is small in general.

5.2. Full distribution of magnitude, colour, and photo-z measures at faint end

The mean and standard deviation may not be a faithful measure if the full distribution is non-Gaussian. For completeness, we look at the full 2D distributions of a range of galaxy properties relevant for photo-z estimation with finer binning. We focus on the faint end of the i-band limit, $25 < i < 25.3$, for the Y5 data, and we take two subsamples: galaxies within the brightest and the faintest 10% pixels from the i-band `Median_fiveSigmaDepth` map. For the two samples, we look at the following 2D histograms:

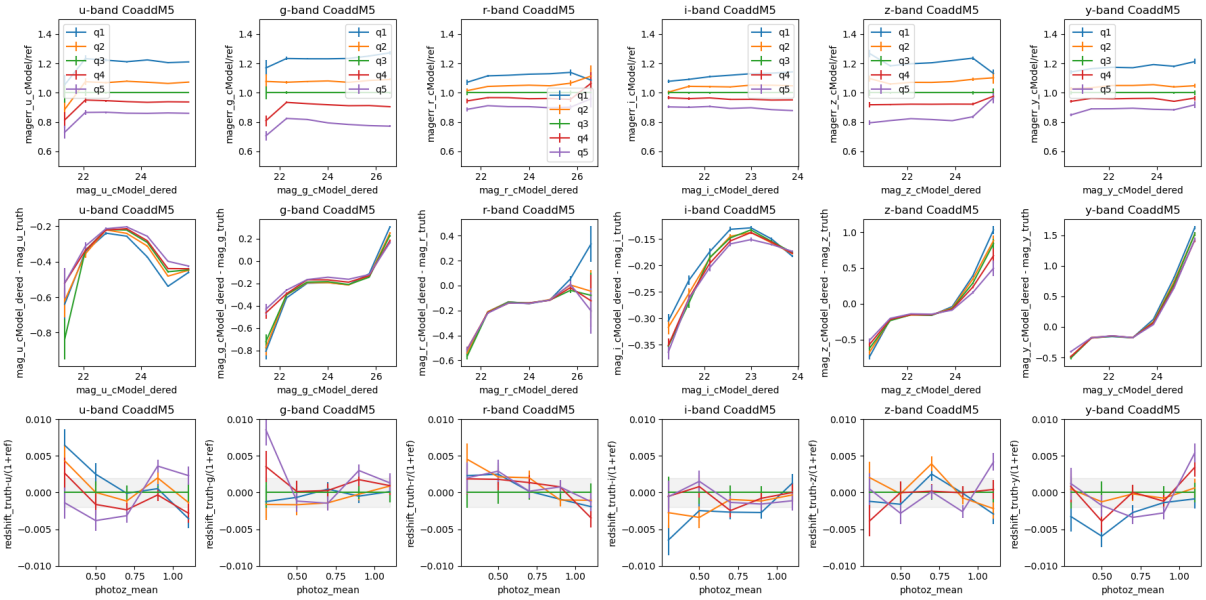


Figure 8: Same as Fig. 7, but for the 5% Y1 data with photo-z.

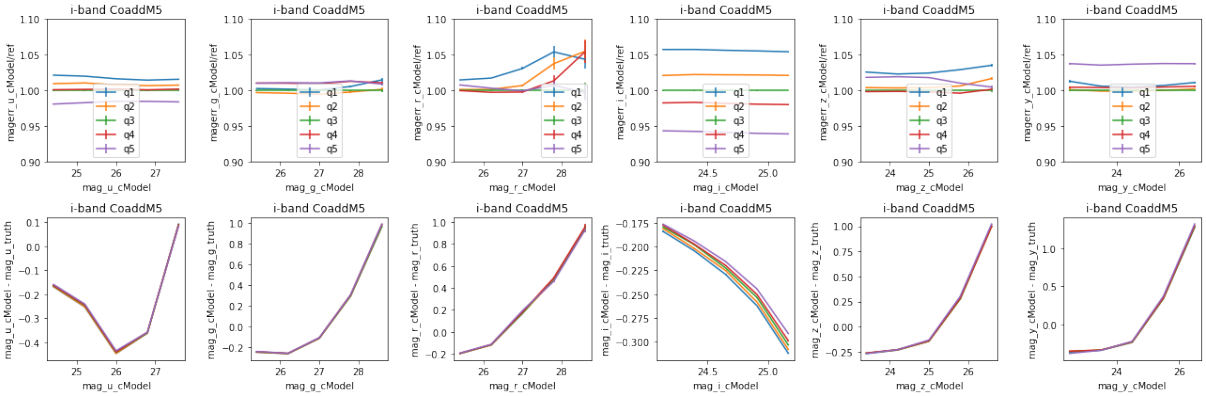


Figure 9: Same as Fig. 7 but for i-band coadd depth only.

- `magerr_[...]_cModel` vs `mag_[...]_cModel`;
- `mag_[...]_cModel` vs `mag_[...]_truth`;
- cModel colour vs truth;
- cModel colour vs `mag_[...]_cModel`;
- photo-z versus true redshifts, `photoz_odds` vs true; redshifts, i-band magnitude and magnitude error.

The supplementary figures are included in appendix Fig. 21 - 25. In all cases, the upper row shows measurements from the shallower 10% pixels ('bad' observing condition), whereas the lower row shows the deeper 10% pixels ('good' observing condition). We include brief comments below.
Fig. 21 shows measured magnitude error versus measured magnitude, and Fig. 22 shows the measured magnitudes in the six photometric bands versus the true magnitudes, thus reflecting the distribution of magnitude bias. These two figures directly correspond to the measurements in Fig. 18. As expected, the difference between the 'good' and the 'bad' samples is only subtle, except for the i-band magnitude error. The red lines in Fig. 21 show the magnitude error computed using Eq. 3 - 5 with the default parameter values. one can see how they underestimate the magnitude errors at the faint end, especially for the y and z bands. It is also noticeable in Fig. 22 that the measurements in y and z bands seem to be in general biased.
Fig. 23 and Fig. 24 show the comparison in terms of colour bias and colour-magnitude space. Both of these plots do not show significant variation regarding to the brighter and fainter sample.
Fig. 25 shows the 2D histogram of various combinations of photo-z quantities versus the true redshift and magnitude. The first column shows photo-z versus true redshift, and it is clear that beyond $z = 1.5$ the photo-z is not well estimated with `BPZ`, resulting in the big spike at this redshift. There is also a significant degeneracy at $z_{phot} = 0.5$. The rest of the columns show trends of `photoz_odds` with true redshifts, i-band magnitude and magnitude error. Across all panels, the 'good' and 'bad' samples do not show significant difference. However, one should again note that any effects in photo-z is at sub-percent level, which does not show up given the binning we adopt here.

5.3. Variations in galaxy number densities

As a sanity check, we also measure the correlation between galaxy over-density, $N/\bar{N}$, with various systematic maps. Again, we focus on the i-band here, and we split the sample into a few bins in i-band magnitudes in both Y1 and Y5 data. In this section, we show the Y1 results as an example in Fig. 10, and include the Y5 results in the Appendix. In each of these exercise, we select pixels from the corresponding survey condition map and compute the mean galaxy number per pixel, normalised by the mean number of galaxy counts using the full footprint.
Fig. 10 shows the variation with `Median_airmass`, `Median_dist2Moon`, `Median_fiveSigmaDepth`, `Median_filtSkyBrightness`, `Median_FWHMeff`, and `CoaddM5` in each row. In each column, we show samples in several bins in i-band cModel magnitude: $20 \leq i < 22$, $22 \leq i < 24$, $24 \leq i < 25$, and $25 \leq i < 25.3$, and the last column shows the full sample with $i < 25.3$. One can see that the largest trend comes from the faintest sample near the i band limit, specifically for the median 5σ depth, seeing, and the coadd 5σ depth, reaching $\pm 10\%$ variation in Y1 (5% in Y5).
The same plot for Y5 is shown in Fig. 26. Additionally, in Fig. 27, we also show the correlations with the `nexp_sum` and `ebv` maps. For the faint sample with $25 < i < 25.3$ in Y5, we also split the sample into four photo-z bins between $0 < z < 2$ to see if any trend exists at different redshifts (Fig. 28). However, the photo-z at this faint magnitude does not well separate the sample in true redshifts - in the lowest and highest bin, the true redshift distribution is quite broad. Here, the figure just serves as an demonstration for possible exercises one could conduct for further investigation, as density variation is not the main purpose of this note.

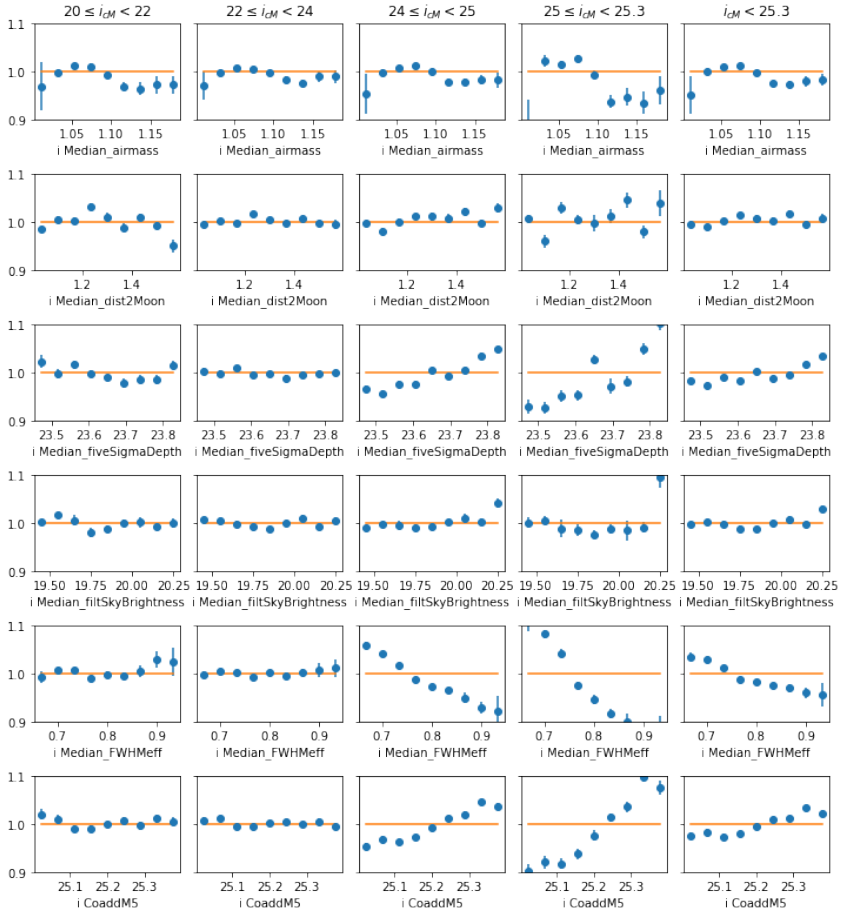


Figure 10: Mean galaxy over density as a function of various systematics, in bins of i-band magnitudes using the DC2 dr2 catalogue.

This exercise may be interesting for studying possible systematics in the clustering redshift method. While in principle, one should expect that the foreground systematics in the photometric survey to be uncorrelated with the LSS in spectroscopic surveys, it may contribute to spurious structures and give biased result. One could, for example, include weighted randoms to test the level of bias introduced by the inhomogeneous number fluctuation arising from observational systemtatics.

6. Possible directions for further investigations

There are several short-comings of the current analysis:

- The current photo-z quality does not quite meet the Rubin requirement even after removing outliers using a simple cut in `photoz_odds`. One could improve the quality by including more informed cuts e.g. in colour space, or explore other photo-z estimators.
- The number of Y1 sub-sample needs to be increase so that the variation in $\langle z \rangle$ and σ_z is not noise dominated.
- Currently we do not adopt the robust mean and robust scatter as in [8]. We also do not compute the statistical uncertainty on these quantities via, e.g. bootstrap.
- The DC2 simulation may not be realistic, thus the level of fluctuations may be underestimated.

Regarding to the last point, two possible avenues could be followed:

- Comparison with an existing survey with its survey condition maps. Currently, HSC is the best candidate with the advantage that the same image processing pipeline is adopted. The down-side is that the survey strategy there is depth before area, so that the expected systematics is probably smaller, and the depth is greater in HSC compared to LSST given the same observing period. It should also be noted that the HSC have multiple fields that are smaller than DC2, especially in DR1. The HSC DR1 survey property maps are accessible on NERSC:

  `/global/cfs/projectdirs/lsst/groups/LSS/HSC_data/decasu_maps/`

  These maps are put together by Javier Sanchez.
- Explore 'source injection', which may be used to study atmospheric / instrumental effect on the observed image / magnitudes in a controlled and systematic way.

References

[1] LSST Dark Energy Science Collaboration (LSST DESC) et al. The LSST DESC DC2 Simulated Sky Survey. *APJS*, 253(1):31, March 2021.

[2] B. Joachimi et al. KiDS-1000 methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis. *AAP*, 646:A129, February 2021.

[3] Sven Heydenreich et al. The effects of varying depth in cosmic shear surveys. *AAP*, 634:A104, February 2020.

[4] The LSST Dark Energy Science Collaboration, Rachel Mandelbaum, Tim Eifler, Renée Hložek, Thomas Collett, Eric Gawiser, Daniel Scolnic, David Alonso, Humna Awan, Rahul Biswas, Jonathan Blazek, Patricia Burchat, Nora Elisa Chisari, Ian Dell'Antonio, Seth Digel, Josh Frieman, Daniel A. Goldstein, Isobel Hook, Željko Ivezić, Steven M. Kahn, Sowmya Kamath, David Kirkby, Thomas Kitching, Elisabeth Krause, Pierre-François Leget, Philip J. Marshall, Joshua Meyers, Hironao Miyatake, Jeffrey A. Newman, Robert Nichol, Eli Rykoff, F. Javier Sanchez, Anže Slosar, Mark Sullivan, and M. A. Troxel. The lsst dark energy science collaboration (desc) science requirements document, 2021.

[5] B. Sánchez et al. Snia-cosmology analysis results from simulated lsst images: from difference imaging to constraints on dark energy, 2021.

[6] Željko Ivezić et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *APJ*, 873(2):111, March 2019.

[7] Branimir Sesar et al. Exploring the Variable Sky with the Sloan Digital Sky Survey. *AJ*, 134(6):2236–2251, December 2007.

[8] Melissa L. Graham, Andrew J. Connolly, Željko Ivezić, Samuel J. Schmidt, R. Lynne Jones, Mario Jurić, Scott F. Daniel, and Peter Yoachim. Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies. *AJ*, 155(1):1, January 2018.

Annex A Supplement for Section 2.2

Annex B Supplement for Section 4.1.2

Annex C Supplement for Section 4.2

Annex D Supplement for Section 4.3

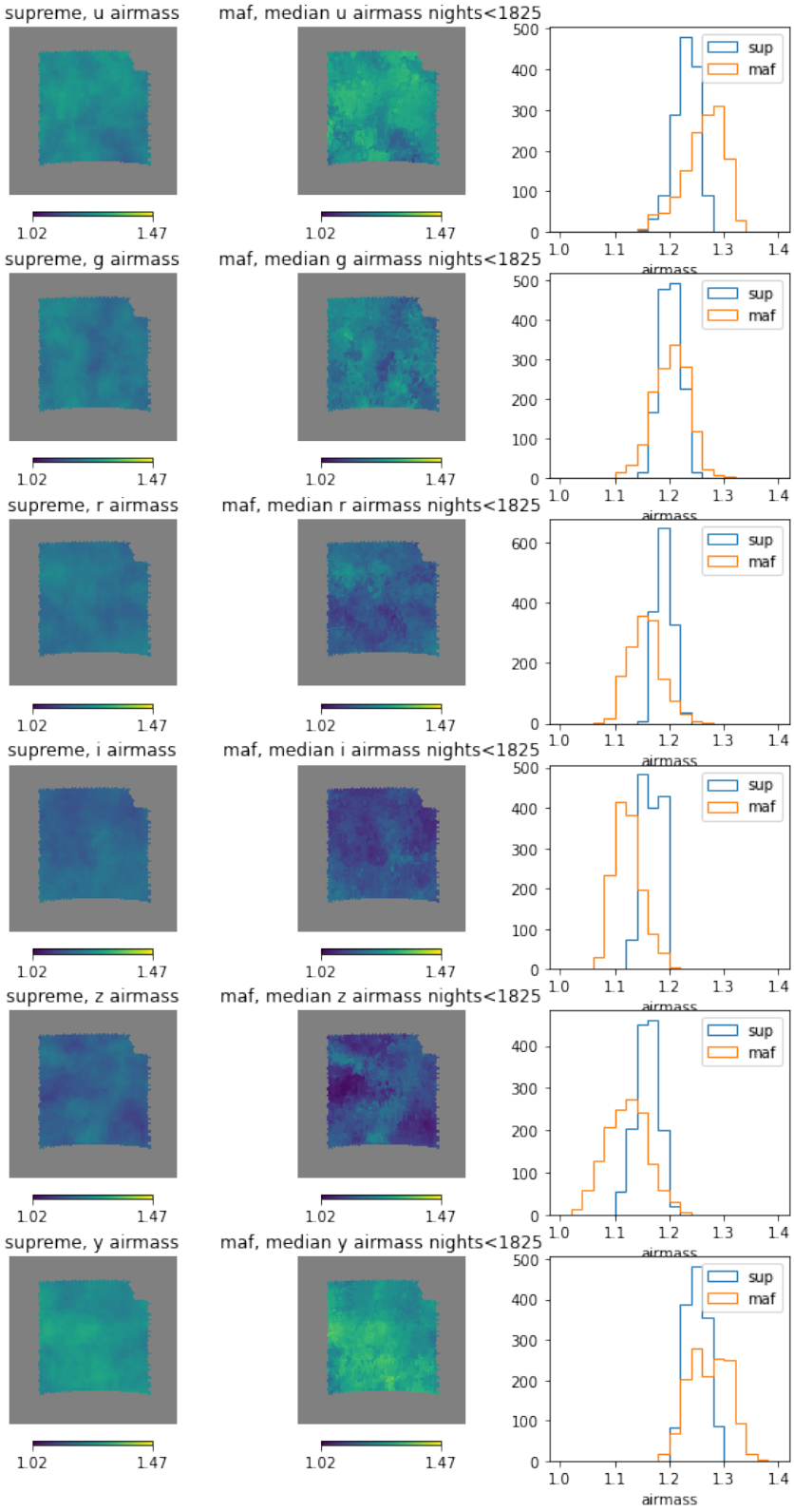


Figure 11: Comparison in terms of spatial variation and distribution between the supreme map and the MAF median airmass in each of the bands.

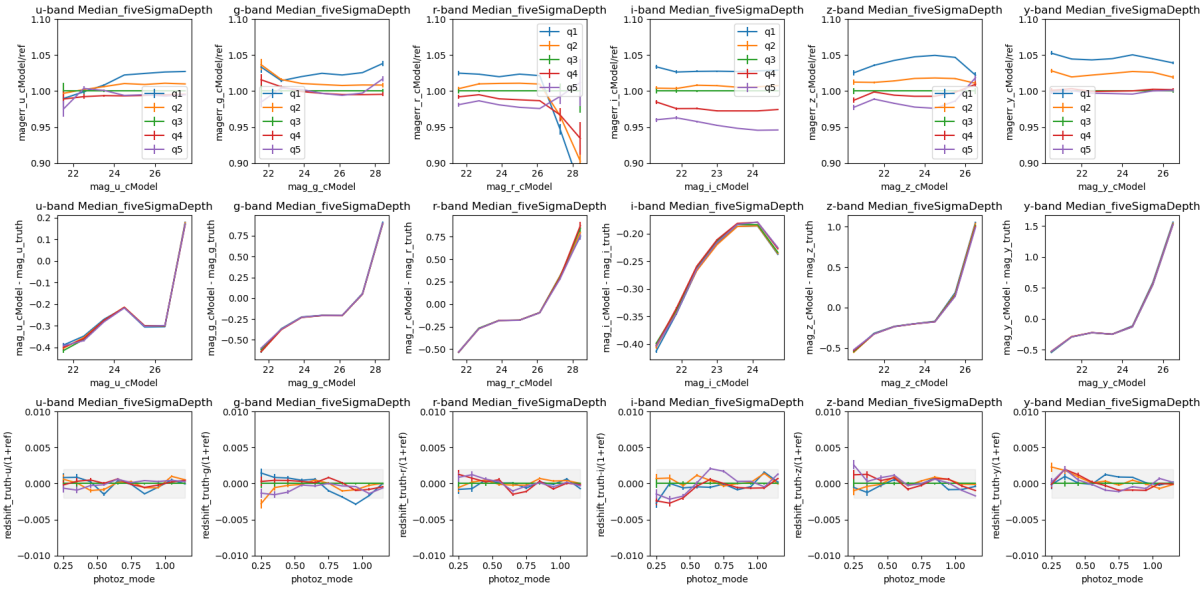


Figure 12: Same as Fig. 7 but for `Median_fiveSigmaDepth`.

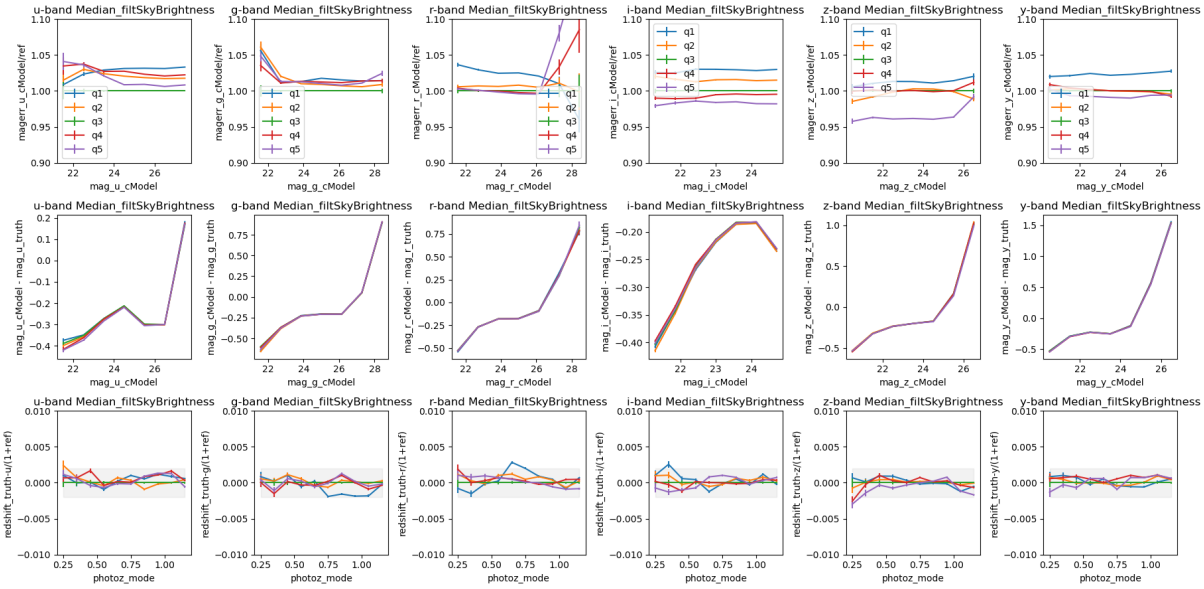


Figure 13: Same as Fig. 7 but for `Median_filtSkyBrightness`.

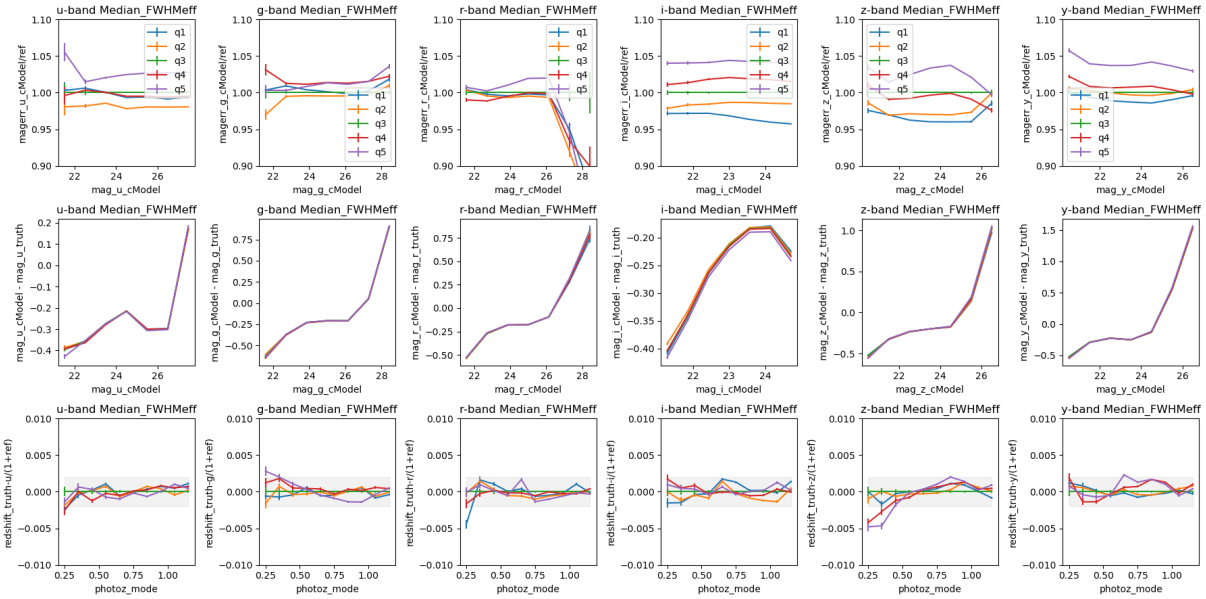


Figure 14: Same as Fig. 7 but for `Median_FWHMeff`. Notice that the trend is reversed, because a smaller seeing angle now means a better observing condition.

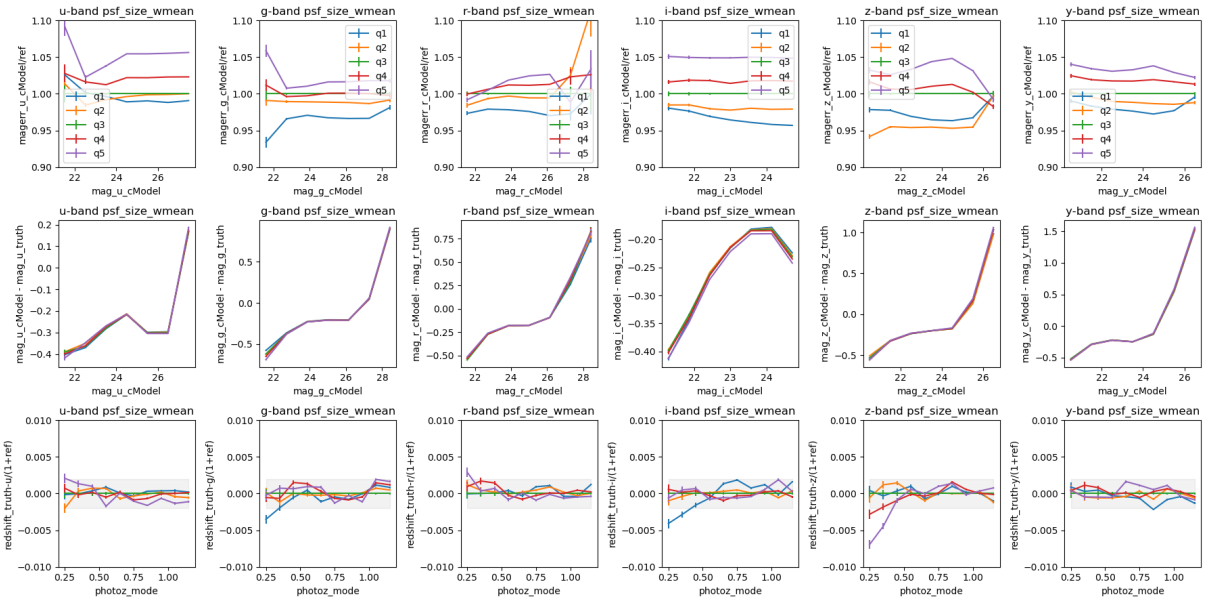


Figure 15: Same as Fig. 7 but for `psf_size_wmean`.

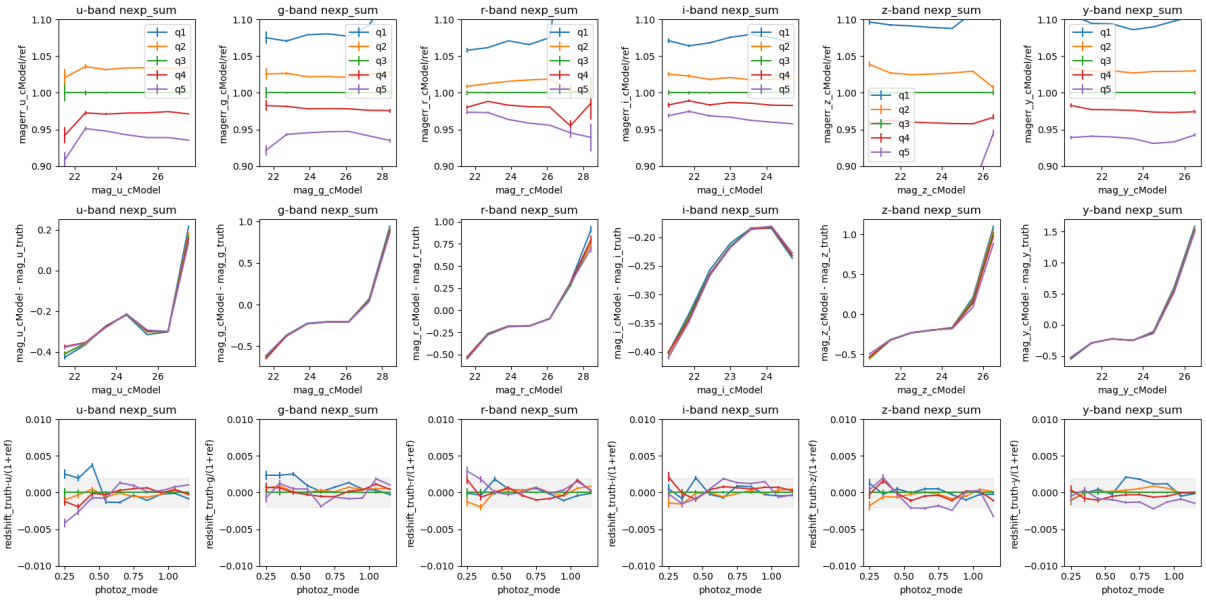


Figure 16: Same as Fig. 7 but for `nexp_sum`. This is very similar to `coaddM5`, as expected, because the number of exposure should correlated strongly with the depth.

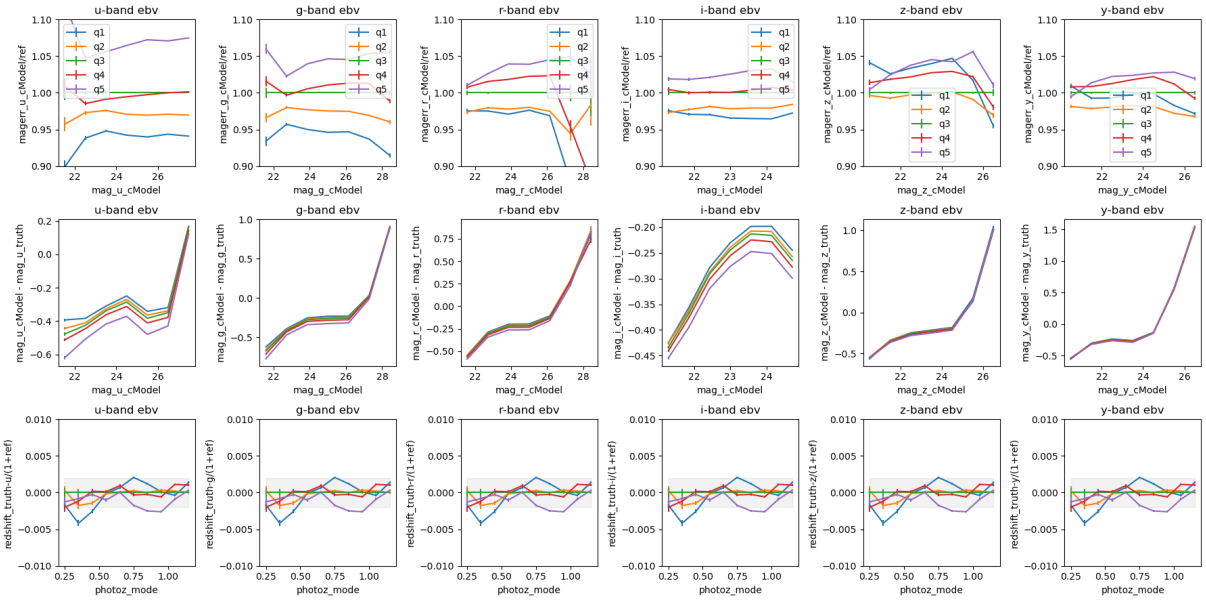


Figure 17: Same as Fig. 7 but for `ebv`. All measured magnitudes are de-reddened.

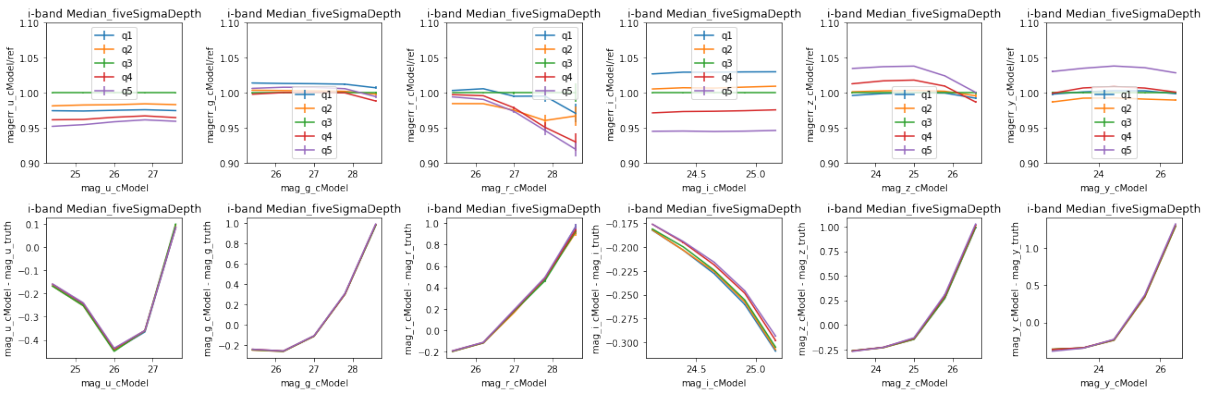


Figure 18: Same as Fig. 12 but for i-band systematics only.

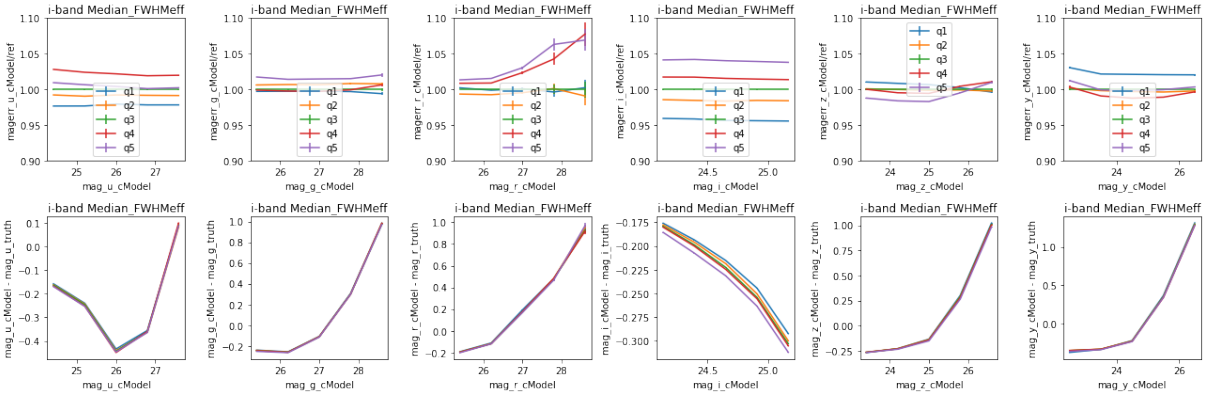


Figure 19: Same as Fig. 14 but for i-band systematics only.

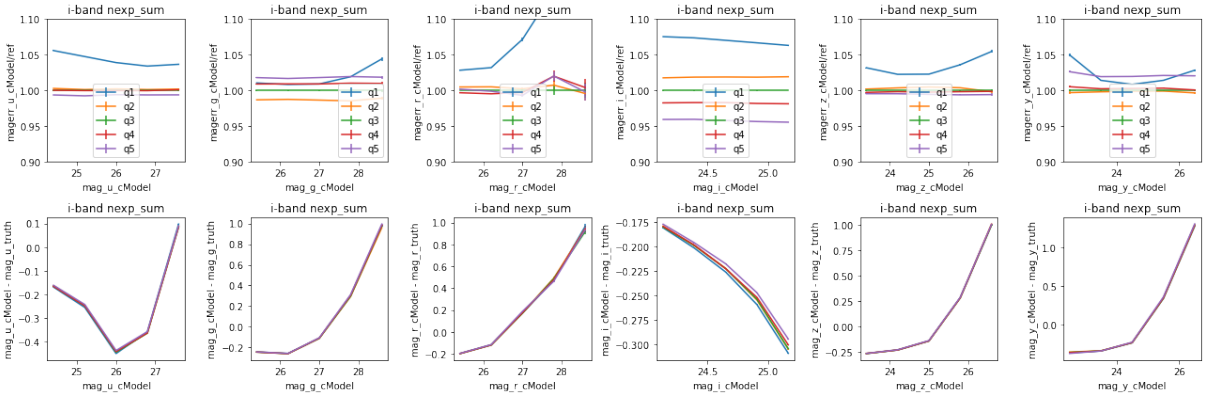


Figure 20: Same as Fig. 16 but for i-band systematics only.

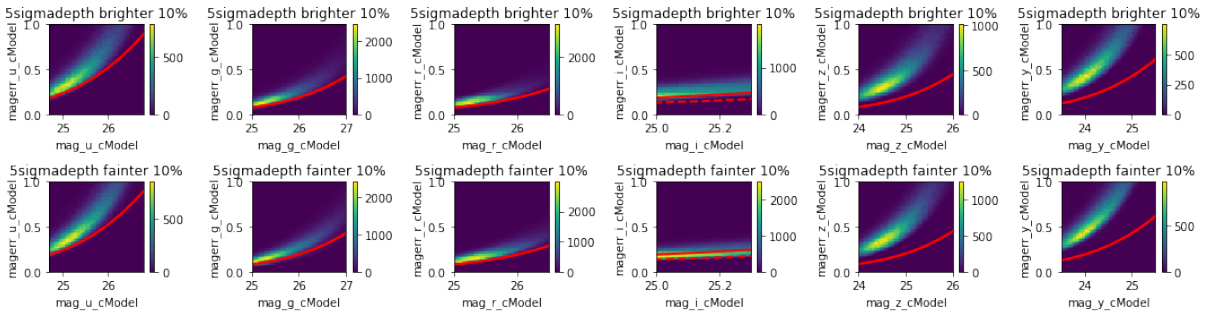


Figure 21: 2D histogram of measured magnitude error versus magnitudes, from the 10% brighter pixels (upper) and fainter pixels (lower) of the i-band `Median_fiveSigmaDepth` map.

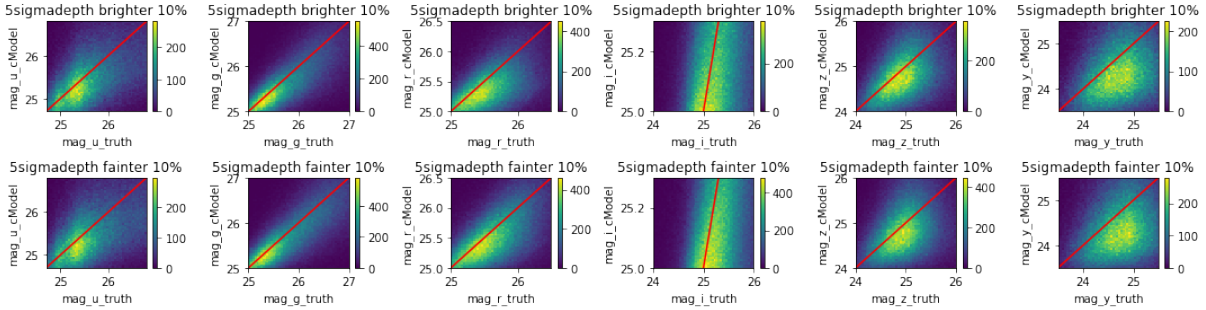


Figure 22: 2D histogram of measured magnitudes versus truth with the same sample selection as Fig. 21.

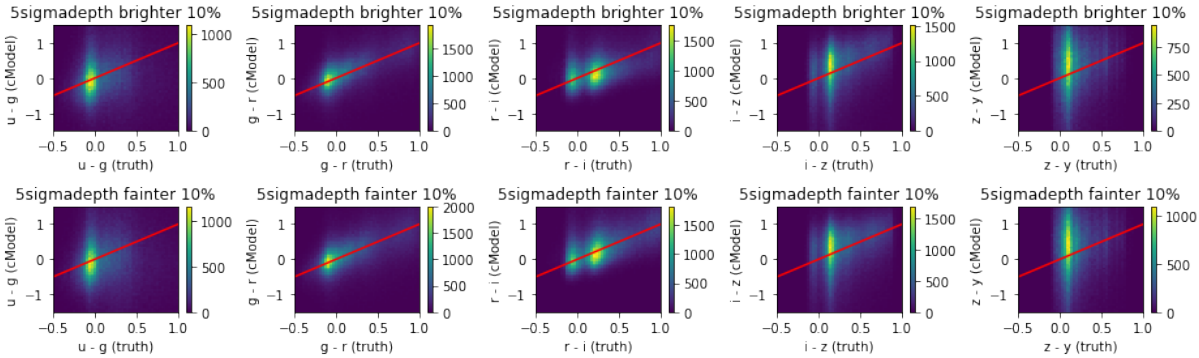


Figure 23: 2D histogram of measured colour versus truth with the same sample selection as Fig. 21.

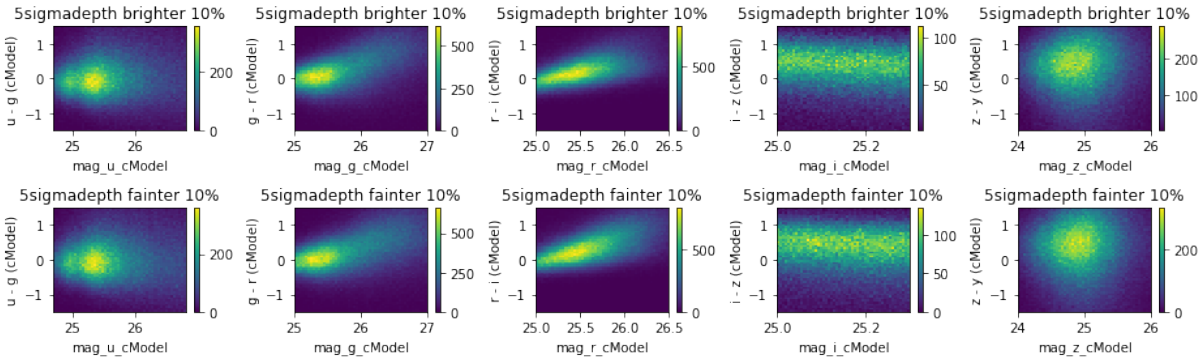


Figure 24: 2D histogram of measured colour error versus magnitudes with the same sample selection as Fig. 21.

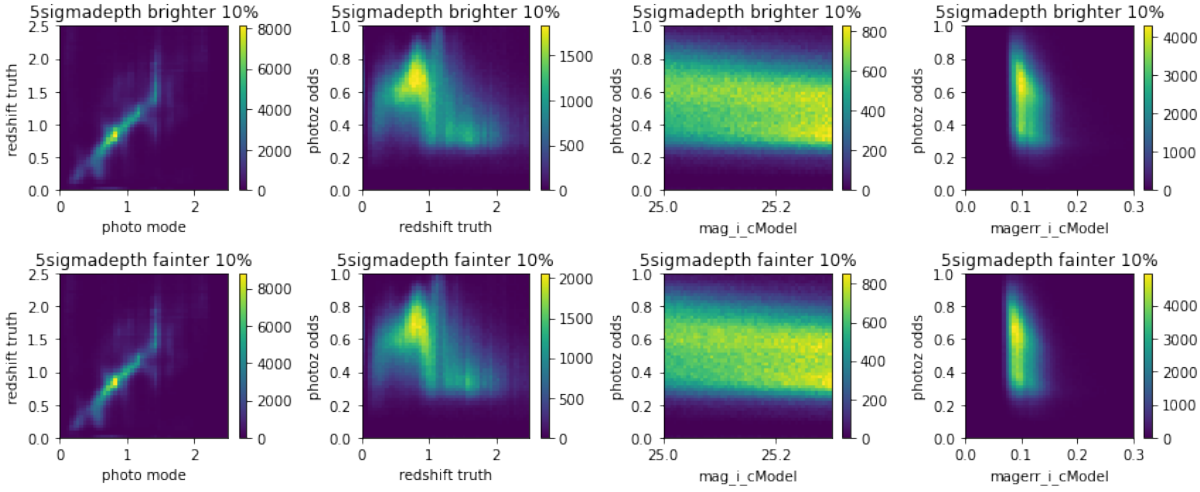


Figure 25: The 2D histogram of various redshift measures and magnitudes with the same sample selection from Fig. 21.

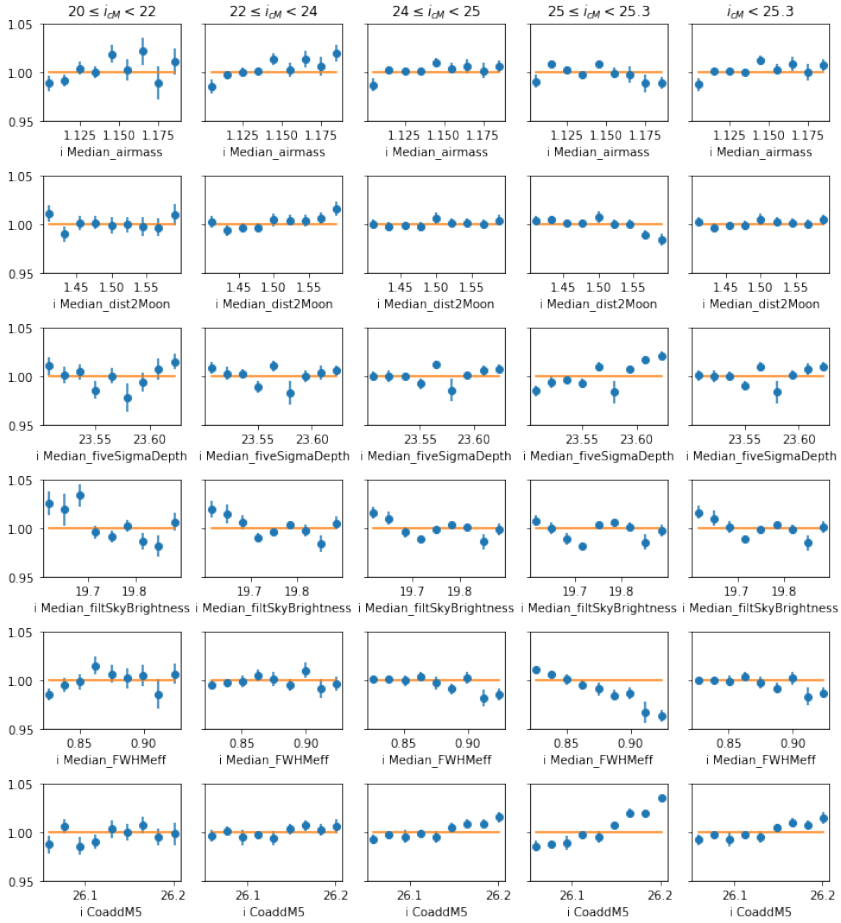


Figure 26: Same as Fig. 10 but for dr6.

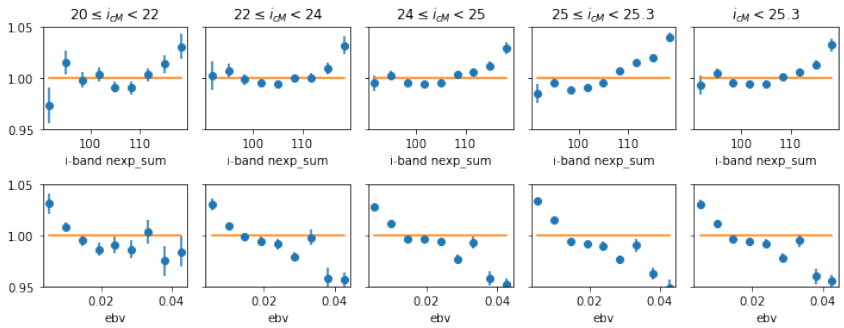


Figure 27: Same as Fig. 26 but for systematic maps `nexp_sum` and $E(B-V)$.

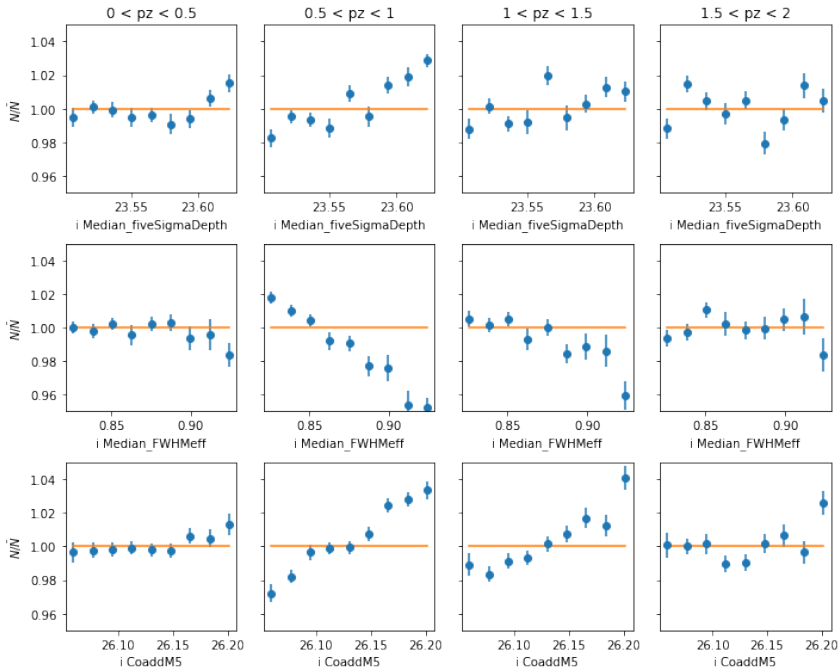


Figure 28: Mean galaxy over density as a function of various systematics, in bins of photo-z mode the DC2 dr6 catalogue for the faint samples with $25 < i < 25.3$.