



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

UNIVERSITAT AUTÒNOMA DE BARCELONA
DEPARTAMENT DE FÍSICA



data mining and precision simulations of Cosmic Surveys

By

SANTIAGO SERRANO ELORDUY

ADVISORS:

ENRIQUE GAZTAÑAGA
FRANCISCO JAVIER CASTANDER

TUTOR:

ENRIQUE FERNANDEZ

Cosmology and Fundamental Physics Department
Institute of Space Science, ICE-CSIC/IEEC

A thesis submitted for the degree of
Doctor of Philosophy in Physics

Barcelona, September 2021

ABSTRACT

The current generation of cosmology surveys requires larger and more accurate observations than ever before. While Edwin Hubble only used 46 galaxies measured from Vesto Slipher almost 100 years ago, present studies require analyzing millions - sometimes billions - of galaxies. Not only the data volume is now many orders of magnitude larger, but the accuracy of each measurement had to increase accordingly. As this growth has been increasing exponentially, recent surveys had to suffer a paradigm shift in the way the data were analyzed.

The practice of a single person analysing the data in a single computer is not valid anymore. A bridge between scientists and engineers had to be built to deal with the complex data processing. Opportunely, the era of Big Data arrived in time, providing the necessary tools to overcome this technological barrier.

In the same way, simulations had to step up, with subtle instrumental features not accounted for before, more realistic universe models from recent observations and faster algorithms that allow producing large sets of precise simulations where tight requirements could be validated.

In this thesis we explore different state of the art cosmological surveys, ground and space based, exploring the spectrum of light from the UV to the infrared. From broadband to spectroscopy. A similar approach has been taken in all cases; an end-to-end simulation where new algorithms can be designed, optimized and verified in this new era of precision cosmology. We will detail the new algorithms and technical solutions in various areas of a cosmic survey that allowed us to achieve the target scientific results.

TABLE OF CONTENTS

	Page
1 Introduction	1
1.1 Observational Cosmology	1
1.2 Earth and Space based observatories	3
1.3 Astronomical Images	5
1.4 Motivation	7
2 Guiding telescopes	10
2.1 Auto-Guiding concept	10
2.2 Guider simulations	11
2.2.1 Stellar catalogue	11
2.2.2 Short exposure PSF	11
2.2.3 Pixel simulation	12
2.2.4 Open shutter effect	13
2.3 Survey Guiders	14
2.3.1 Dark Energy Survey	14
2.3.2 Physics of the Accelerating Universe Survey	19
2.3.3 Dark Energy Spectroscopy Instrument and Survey	26
2.4 Enabled scientific results	38
2.5 Summary	39
3 Data Mining a Cosmic Survey	40
3.1 The PAU Survey and Data Management	40
3.2 Simulations	42
3.2.1 Cosmology & Mock Galaxy catalogue with MICE	42
3.2.2 Stellar catalogue	44
3.2.3 Instrument, telescope and sky models	44
3.2.4 Survey Strategy Simulator	50
3.2.5 Pixel Simulations	53
3.3 Observing at the WHT	60
3.3.1 Survey strategy definition	61

3.3.2	Night planning and Survey monitor	64
3.3.3	Adaptive exposure time with SN estimator	68
3.4	Reduction of narrow band images	69
3.4.1	Data overview	71
3.4.2	Instrumental detrending	72
3.4.3	Astrometric calibration	82
3.4.4	Photometric calibration	85
3.5	Photometry of narrow band images	99
3.5.1	Forced Photometry	100
3.5.2	Flux measurement and error estimate	104
3.5.3	Survey mask	108
3.5.4	Quality assurance & validation	110
3.5.5	Known limitations & further work	117
3.6	Operation and data orchestration	119
3.6.1	PIC data center	120
3.6.2	Operation	122
3.6.3	Data flow and data orchestration	122
3.7	Enabled scientific results	128
3.7.1	Photometric Redshift Estimation	128
3.7.2	Intrinsic alignments and clustering	129
3.8	Summary	130
4	Space Based Simulations	132
4.1	The Euclid Mission	132
4.1.1	The VIS instrument	134
4.1.2	The NISP instrument	135
4.1.3	The Ground Segment	136
4.1.4	OU-SIM	137
4.2	The True Universe	138
4.2.1	Cosmology & Mock Galaxy catalogue with Flagship	139
4.2.2	Stellar catalogue	141
4.2.3	SED library	144
4.2.4	Galaxy profile generators	149
4.3	The mission database	154
4.3.1	The reference survey	155
4.3.2	Models and Parameters	157
4.4	Instrument simulators	158
4.4.1	VIS Simulator	159
4.4.2	NISP-P Simulator	162

4.4.3	NISP-S Simulator	166
4.4.4	EXT Ground Based Simulator	168
4.5	Operation and production of simulations	169
4.5.1	Infrastructure and Science Data Centers	171
4.5.2	The Sim Planner	172
4.5.3	Scientific Challenges	172
4.6	Summary	175
5	Synopsis and final conclusions	177
A	Synthetic spectra samples	181
B	Data base schema	188
	Bibliography	195

INTRODUCTION

1.1 Observational Cosmology

Humanity has been constantly trying to understand the origin and rules of the Universe we live in. It was not until the beginning of the XX century that our knowledge expanded from our own Galaxy to further distant galaxies. The discovery of a much larger Cosmos was only possible thanks to major theoretical and technological advances. Since then, our understanding of the nature and origin of the Universe has been constantly growing, together with larger and more accurate observations.

Distances and Redshift One of the most important challenges in Observational Cosmology is the ability to accurately measure radial distances. At closer distances, like nearby stars, we can use its parallax motion at different locations of earth in its transit through the solar system to determine its average distance. However at the enormous cosmic distances between galaxies, this approach is not anymore valid and alternative techniques had to be discovered. In 1908 Henrietta Swan Leavitt discovered a particular behaviour of Cepheids, a star that pulsates its brightness in a very well predictable manner, allowing astronomers to define a reference object that reached much further than the parallaxes technique. With these standard candles across the Universe, Edwin Hubble managed to determine the distance of further away galaxies. Another key event happened in 1912, when Vesto Slipher estimated the radial velocity of distant galaxies by measuring the doppler shifts from the emission lines in the spectrum. Most galaxies showed positive radial velocity from their redshifted spectrum. It was then in 1927 when the astronomer Georges Lemaître found the relation between velocity and distance, leaving Edwin Hubble two years later to define the Hubble Law with his name, and propose that this relation was explained by the expansion of the Universe, at a rate of 500 km/s/Mpc . From now on, we would estimate

the distance of a galaxy based on his spectral redshift. Much later in the 90s, when trying to calibrate Hubble's constant with more accurate supernova standard candles, it was found that the expansion was not happening at a constant rate, but rather accelerating.

Since the first velocity measurements that carried out Slipher more than 100 years ago, numerous cosmic surveys have continued to estimate galaxy redshifts. Most cosmological studies based in observations require large statistical analysis that reduce the sampling variance of observing a limited number of galaxies. For this reason, surveys have been trying to map larger and larger areas in the sky, providing three-dimensional maps where the theoretical models can be validated. Thanks to the rapidly evolving technology, current surveys can map very large areas in the sky with thousands of square degrees down the very faint galaxies, which result into maps that reach further away, in some cases close to the limits of the observable universe.

Past surveys At the beginning of the 80s, the CfA Redshift Survey (Tonry, Davis, 1979) tried for the first time to map the large scale structure of the universe by measuring 2,200 galaxies one by one, extending the survey at the end of the decade to approximately 15,000 galaxies, down to a B magnitude of 14.5. Later in the 90s, the 2dF Galaxy Redshift Survey (Colless et al., 2001) managed to observe $\sim 250\,000$ galaxies on the Anglo-Australian Telescope. Thanks to an instrument capable of observing 400 objects simultaneously, the 2dF reached a limiting magnitude of 19.5 with a median redshift of 0.11 over 1500 deg^2 of sky. Another step up was reached by the Sloan Digital Sky Survey (SDSS) (York et al., 2000), with almost a million galaxies observed between 2000 and 2005, over 8000 deg^2 in its five optical bandpasses.

Current surveys In this thesis, we will focus on various surveys in the decade of 2010, which significantly increased the volume and complexity of data thanks to large and dedicated instruments. The first to start in 2012 was the Dark Energy Survey¹ (The Dark Energy Survey Collaboration, 2005), reaching a magnitude 24 in the *i* band in its 5000 deg^2 . Observations finished on 2019, having observed more than 220 million galaxies, it holds the largest cosmology map to date. On 2015 the Physics of the Accelerating Universe Survey² (PAU) started, and with a camera (Padilla et al., 2019) designed to optimize the galaxy redshifts, it delivered the most accurate distances with imaging data. The last survey that started recently in 2021, is the Dark Energy Spectroscopy Instrument³ (DESI Collaboration, 2016), which plans to observe 35 million spectra from very distant galaxies and quasars, which will also hold the largest spectroscopic sample, with incredibly accurate 3D positions.

Future Surveys In the near future, very large and ambitious cosmological surveys will begin. In this thesis we will also describe work for the Euclid Mission (Laureijs, 2011), a space observa-

¹<https://www.darkenergysurvey.org/>

²<https://pausurvey.org/>

³<https://www.desi.lbl.gov/>

tory that will map 15.000 deg² of sky, observing more than a billion galaxies with exquisite detail. Also expected to start in 2023, the Vera Rubin Observatory (LSST collaboration, 2019) with its 8-meter class dedicated telescope will start a massive survey from Earth, reaching depths around magnitude 27 in the i-band, and will become the deepest and largest survey with its 37 billion sources when it concludes in 2033.

In the past 4 decades, each cosmic survey has increased the volume from the previous set by an order of magnitude. This exponential rise requires a proportional increase in the size and complexity of the data processing required, as it will be presented in this study.

1.2 Earth and Space based observatories

Contrary to other sciences where one can interact or modify the subject to study, in astronomy we are limited to observe the light (and recently also gravitational waves) that arrives to Earth from distant sources. For this purpose we have been building observatories across the whole planet to observe our night sky. With a thick atmosphere that blurs out our images, best observatories are located as high as possible in spots with clear and dark skies such as Hawaii, Chile or the Canary Islands.

Pros and Cons Since the 60s, we have been launching small observatories to space, removing the atmosphere and all the bad effects it comes with it. It allows to observe in bands where the atmosphere absorbs all light such as Gamma Rays, X-Rays or a good amount of the Infrared spectrum. The image quality is superior, with no variations from the turbulent airflows or clouds that prevent any observation. And finally we can locate our space observatory against the sun, allowing 24 hours of continuous observation. However, sending observatories to space has several disadvantages. Probably the most important is the cost, which can be orders of many magnitude higher than the equivalent telescope on earth. The level of qualification required to send even the simplest piece of equipment makes the preparation much slower than ground-based equivalents, as mistakes or upgrades are most of the cases not possible. Both scenarios having pros and cons, observatories are still being constructed on Earth and in Space.

Telescopes In cosmic surveys, we collect the light with optical reflector telescopes, which are mainly composed by a large mirror in its back, and depending on the design, other mirrors or glass elements to redirect and focus the light beam towards the instrument that detects the photons.

Due to the need of observing more and fainter galaxies, telescopes have been increasing its collecting area, from the 24-inch (0.6m) refractor that Slipher used in 1912 to measure the first redshifts, to the 8m telescope that is being built for the Vera Rubin Observatory, with 200 times more surface. Current telescopes are gigantic structures weighting hundreds of tons, capable of pointing anywhere in the sky with a precision around one arcsecond (less than one thousandth

of your fingertip with your arm extended). As exposures are typically long, it is important that telescopes can track with precision for long periods of time. Even a telescope can blindly follow accurately a fixed position of the sky compensating Earth's movement, the excellent image quality and long exposure times necessary in cosmic surveys require very minor corrections to fully stabilize the image and avoid the telescope to drift. In Chapter 2 we describe three autoguiding systems that required to be designed specifically to meet the image quality requirements of three cosmology surveys.

When computers were not available, most telescopes installed a simpler equatorial mount. These mounts had its polar axis aligned with the one of Earth, allowing to track objects by fixing the altitude and moving the azimuth at a constant rate to compensate for Earth rotation. DES and DESI were installed on telescopes with such configuration. Once the technology allowed to track the two axis independently with a more complex movement, mounts switched to an altitude-azimuth configuration, making the structure lighter and more symmetric, allowing larger mirrors such as the 4.2m William Herschel Telescope, the 8.4m Vera C. Rubin Observatory or the 10.4m Gran Telescopio de Canarias.

Instruments The last key element in the overall system is the instrument, which performs the measurement over the incident light collected by the telescope. In large cosmological surveys we can mount two main types of instruments: imagers and spectrographs. In the first case we will obtain images of the sky, passing the incoming beam through a filter. If we observe with a single filter, we will obtain information of the size, shape, position and brightness of the various sources in the band we observed. However it is common to observe at various bands to obtain color information for each source. This is particularly relevant when trying to estimate the redshift of sources, where each band will add a point in the spectrum. Redshifts obtained from images are called photometric redshifts (or photo-z for short). The more bands we have, the more accurate we will determine the photo-zs. On the other side we have spectrographs, where light of each source is dispersed across our detector, allowing very accurate redshift determination as emission lines can be identified. Spectrographs can create a slit where the source is, leaving only the light of the source pass before it is dispersed, or putting a fiber at the position of the source, or dispersing all light in a slit-less configuration, which will produce contamination between overlapping spectra. The disadvantage of spectroscopy against photometry is that we will have to select the objects we want to obtain the spectra, resulting in a less complete, shallower and biased (from our selection) mapping of the universe. From the surveys in this study, DES mounts an optical imager with five bands, DESI is a spectrograph with 5000 fibers and PAU sits in between being an imager with 40 narrow bands. The Euclid spacecraft mounts two instruments, a visible imager and a near infrared that does both imaging and slit-less spectroscopy. All these instruments are tailored to do Cosmology and have been optimized to measure either efficiently or accurately galaxy redshifts.

1.3 Astronomical Images

After observing the sky with our instruments (either in the telescope or remotely in a spacecraft), we end up with a set raw images that contain the information we need to extract, but yet are not ready to do science with them. These are called single-epoch images as they have been observed at a single period in time (contrary to coadd images that combine multiple exposures a various times). Images in astronomy are typically stored in FITS files (which stands for Flexible Image Transport System), a format defined by astronomers that contains both the imaging information (data) and its associated metadata (header) such as exposure time, filter, telescope coordinates, time of observation and much more. The FITS files are then packets of Header Data Units or HDUs. Nowadays instruments are built by large mosaics of detectors, and therefore are FITS files contains multiple HDUs, one per detector or amplifier.

Elements Images contain the sources we want to do science with them, which in cosmic surveys are mainly galaxies and quasars. The other main type of objects that we will find in our observed images are stars, which even they are not our main target, they will be essential in the calibration process, as we will see later. There are other types of astronomical objects that will appear such as asteroids, nebulae and even earth satellites and planes, but these less common and not of our interest for cosmology. Lastly we have cosmic rays, which are high-energy particles travelling at nearly the speed of light which will interact with our sensor causing dots and traces in our image that will complicate the measurements we are interested to perform.

In imager instruments, the image will contain a picture of the sky with the actual shape and distribution of the objects in it. In spectroscopy, the look of the image will depend on how the instrument captures the light for each source (with slits, slit-less or fibers as explained previously), but in one way or another, we will have the dispersed light into the image where the pixel values contain the spectrum of each source as a function of wavelength.

Behind all astronomical objects we detected in our image, we will find an almost flat level of flux that we call background. Depending on the band we observe or the location of our telescope (space or ground) one source of background will be more important than other but the most relevant are Zodiacal Light, atmosphere airglow, scattered light from the nearby stars and the moon. As these sources of illumination have a very large scale, we assume to be constant across our field of view, although as we will see later in this study, there are cases where this is not true and the analysis is significantly more complicated.

Unfortunately the process of taking images produces noise, limiting our capability to observe fainter objects. This is called readout noise and it is produced by the electronics that acquire the image. Other important sources of noise are the shot noise of both objects and background, associated to the particle nature of light. Photosensitive devices produce noise themselves which we call dark current, although in cooled down devices this is often negligible.

And finally our own instrument plus telescope system will alter the true information of light,

by distorting, reducing or simply not responding homogeneously to the incident light. We call these effects instrumental signatures and most of the calibration process aims to correct these artificial trends.

A key element that determines the quality of the whole imaging system is the Point Spread Function (PSF), which describes the distribution of a point source at the focal plane. Stellar objects are assumed to be point like sources due to their relative small size and far distance. In ground based telescopes the PSF shape typically follows a Moffat profile and it is driven by the atmosphere turbulence (even it is also affected by the optical system of the telescope and instrument). The width of the profile is usually measured in Full Width Half Max (FWHM) and under good conditions we reach a PSF FWHM below 1 arcsecond. In space the PSF tends to be more complex as it is not blurred out by any atmosphere, and the pure optical + detector shape is remaining, following a less analytical profile distribution.

Calibration Once all images and calibration images are obtained, we can begin the processing and calibration of the images which we call data reduction. The instrumental de-trending is a long and complex processing, specially if the survey requires high precision measurements, like cosmic surveys nowadays. A very common process during this stage of calibration is to correct the response of the light with a flat-field calibration image. These are images that are produced with a uniform flat light illuminating our instrument. As the instrument will perturb the light similarly to our reference flat field, we can divide our images with the flat field to flatten the response of our observation.

Another important step is the astrometric calibration, where we correct the optical distortions and telescope pointing inaccuracies. This is made using the stars available in the image, serving as reference points in our field of view. Comparing these reference positions with a well calibrated reference catalogue like Gaia (Gaia Collaboration, 2018), we can reposition all sources in our image.

A final step is to correct the fluxes of the objects, known as photometric calibration. The intensity of the sources in our image will vary from one exposure to another depending on multiple factors such as atmosphere variations, change in response of the detector or degradation of transmission in the optical system. Furthermore we need our fluxes at a given standard of flux so we can compare our measurements with observations from other surveys. The photometric calibration process is very specific to each instrument and to the available of reference sources in the field we observed. Generally we use stars as the reference flux due to its simpler shape and abundance across the sky. Then we compare these to photometrically calibrated stars which allow us to compute the scaling factor that needs to be applied to all fluxes in our image to transform our arbitrary fluxes into a given standard magnitude system. Magnitude standards can be defined such that a given reference body like Vega star has 0 magnitude in all bands. However it is more common in cosmology to define magnitudes in the AB system, which does not refer to a particular source but to a calibrated spectra in absolute units.

After all instrumental, astrometric and photometric calibrations have been applied to our observed images, we can carry out all measurements necessary for the scientific analysis.

1.4 Motivation

In this thesis we aim to explain the process we followed in the different areas of various cosmic surveys, detailing how new technologies enable the accuracy required in such large and complex projects.

Large surveys and the big data era As previously stated, current cosmological studies have been increasing their volume and complexity of data up to a point that traditional analysis methods were not valid anymore. In the past, the data obtained by an astronomer could be processed in the telescope itself or in a personal computer at the research institute. Today, the massive volume and complex analysis require to operate the processing in a data center, or even in a grid of computing centers. In 2000 when SDSS started their observations, the volume of data produced in a single week was larger than all previous data collected in history of astronomy. The whole SDSS data collected in years will be negligible compared to the 60 PB of raw observations that will be produced by the Vera Rubin Observatory.

Fortunately, the increased volume of observations came with new and more powerful computing technologies that enabled its necessary analysis. The era of big data arrived in time providing the data management tools we need, and here we will describe those that resulted more effective in the problems we encountered in our way.

New techniques The data reduction techniques used in previous surveys cannot be simply applied, and new scalable algorithms had to be designed that meet the more strict needs of today's studies. Higher level languages such as Python enable fast and versatile programming not possible with older programming languages like Fortran or C. Even standard astronomy libraries such as IRAF (Tody, 1986) that have been used for decades are becoming obsolete with astronomical Python libraries like Astropy (Astropy Collaboration, Astropy Contributors, 2018) or Photutils (Bradley et al., 2020). The combination of parallel processing in High Throughput Computing data centers with these new advanced libraries changed the paradigm in which data reduction pipeline are being built.

Large volumes of data enable data mining techniques extracting further information imperceptible to previous generation algorithms. Machine learning algorithms are being trained in astronomical data, providing ever more subtle instrumental detrending corrections and extracting further information, not possible to train when the data sets were smaller.

The technical interface gap The data management of a cosmic survey has become a hybrid task, requiring both a computer and an astronomical background. In the past astronomers used

to carry out all tasks, taking care of both design and operation of the data reduction, all the way til the final scientific analysis. Nowadays this has become such a technical task that support from computer scientists and engineers is mandatory.

This thesis aims to highlight the importance of close interaction between development of scientific codes and the operations team in a practice called DevOps. It has been proved essential the bridges between both teams, such that engineers understand the scientific goal and scientists understand the technical solutions.

In Chapter 2 we will explain the implementation of three autoguiding systems, a key algorithm in the operation of these instruments made for three cosmic surveys.

In Chapter 3 we will detail the entire data flow of a cosmic project, from the survey strategy and observation at the telescope, to the whole data reduction process using advanced data mining techniques.

In Chapter 4 we will summarize a precise simulation pipeline for a space mission aiming to map the geometry of the Universe.

In Chapter 5 we will summarize the work done and review the solutions proposed.

GUIDING TELESCOPES

We begin explaining work in the process of developing and commissioning instruments, the early steps in the process of building cosmic surveys. Precisely in the system that stabilizes the telescope pointing inaccuracies: the auto guider. Similarly to other areas in the project, these systems had to be improved from the existing ones as the image quality requirements had to step up from the previous surveys.

In this chapter we describe various auto-guiding systems designed for three different cosmology instruments, mounted on 4-meter class telescopes. The different systems have specific requirements based on the target science each survey needs to achieve.

2.1 Auto-Guiding concept

One of the most difficult tasks for all telescopes is the ability to follow the sky precisely. This is even more challenging on telescopes with such a large primary mirror as the ones where we mounted our instruments, weighing hundreds of tons. Typical astronomical observations span from a few seconds to hours of exposure time. Therefore it is critical that these huge constructions point to the exact same location in the sky throughout the whole exposure. Even though these telescopes can blindly follow the sky with a very low tracking error, after a few minutes they all need pointing corrections to prevent drifting from the nominal position. Therefore most telescopes have built-in auto-guiding systems that provide closed-loop feedback to the tracking system to ensure the observed sky stays in place for any exposure time. However, new cosmological surveys have very strict guiding requirements, and due to its wide field of view, they cannot reuse the telescope default guiding system and custom guiders need to be designed and built.

The approach followed by most auto guiding systems to ensure a precise telescope tracking is to have two cameras: one dedicated to the long scientific exposure and a second one with very

short exposures dedicated to guiding. On the guiding camera we first need to pick up one or more guiding stars, continuously measure the guide star motion and send back to the telescope the measured motion to correct for possible telescope tracking errors. The signal that comes from drifting and not from atmosphere turbulence will be corrected and compensated by the telescope, ensuring the image exposure stays in place during the necessary time.

2.2 Guider simulations

Sky time is precious and it is important that any new system is tested as much as possible prior to being mounted on a telescope to reduce commissioning time. Most of the sub-systems of an instrument can be tested in the laboratory where these instruments are built, such as the detector readout, the cooling system, the filter interchanger, etc. But the guiding system requires realistic simulations of the sky to be tested during the design and integration phase as it requires stars in the field of view, which cannot be reproduced easily in an optical bench. For this purpose we have designed realistic guide image simulations that reproduce the data that we will observe once the instrument is mounted.

2.2.1 Stellar catalogue

In order to monitor and correct the telescope tracking errors during the whole science exposure, we have to continuously read the guide detectors at a fast rate. Therefore guide exposures are very short (around 1 second) and only stars are expected to be above the detection threshold in this kind of images. A star catalogue needs to be produced to feed the image simulation with stellar sources. Having a realistic distribution of magnitudes as a function of galactic latitude is critical as we need to ensure in the design phase that we will always be able to find stars with enough signal to noise to guide.

The signal to noise will depend on many factors, such as telescope optics, band-pass, detector noise, sky brightness, etc. and those will be detailed in later sections, but the range of magnitudes of interest is approximately $14 < r_{AB} < 19$. Under this range, and due to the fact that during the design phase of these projects the Gaia catalogue was not available - RP and BP integrated bands were released in April 2018 with the Data Release 2 (Gaia Collaboration, 2018) -, we decided to use Sloan Digital Sky Survey (SDSS) stars (SDSS Collaboration, 2015) as they cover the bands and magnitude range of interest.

2.2.2 Short exposure PSF

Typical Point Spread Function (PSF) of astronomical images from ground based telescopes are composed by atmosphere blurring, telescope motion - mostly due to guiding errors -, instrument diffraction, optics diffusion and pixel sampling. To simulate a realistic PSF in guiding images, we need to bypass the first two components: with enough exposure time, the atmosphere's turbulent

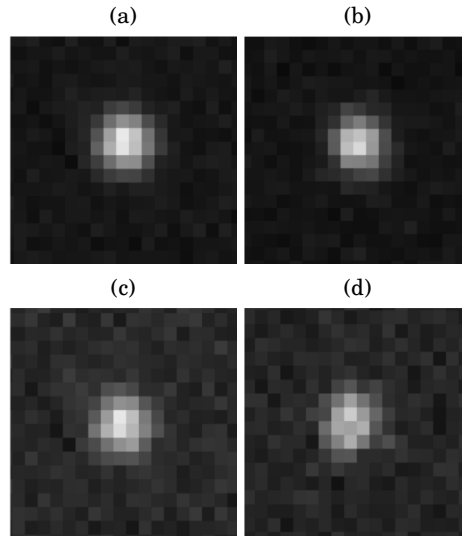


Figure 2.1: Comparison between long and short exposure profiles. (a) 1 second exposure with Gaussian profile, (b) 1 second exposure with the custom short exposure profile, (c) 0.5 second exposure with Gaussian profile, (d) 0.5 second exposure with the custom short exposure profile

motion spreads the photons in a Kolmogorov distribution, which resembles a normal distribution. Guiding exposure times are often shorter than a second, resulting in a non smooth profile. In a simplified approach, we reproduce this effect by stacking a sequence of delta functions with centers chosen from a Gaussian distribution. The difference between a long exposure analytical profile and the short exposure profile can be seen in Figure 2.1. On short exposures, the PSF exhibits a characteristic speckle pattern while for longer exposures, it smooths out the speckles delivering a smoother and circular PSF. This is particularly critical in the guiding algorithm as the centroiding accuracy is noticeably reduced under this irregular profile.

2.2.3 Pixel simulation

Putting all pieces of the simulation together brings us to the pixel simulation core. Initially and with the objective of having full control of the pixel values, we created GUIDESIM a simple image guider simulation tool with a reduced number of instrumental effects.

All projects under study have reused the same detectors dedicated for science as the guide detector, which consist of Charged Couple Devices (CCD), the most common detector nowadays for optical imaging. The simulator emulates the read amplifier gain and a Gaussian read noise. A zeropoint is defined to determine the conversion between electrons per second and apparent magnitude. Sources from the sky are limited to stars, sky background and cosmic rays, the main components in these short exposure images. No galaxies are included in this simulation. The PSF follows the algorithm to render short exposure PSFs defined in §2.2.2.

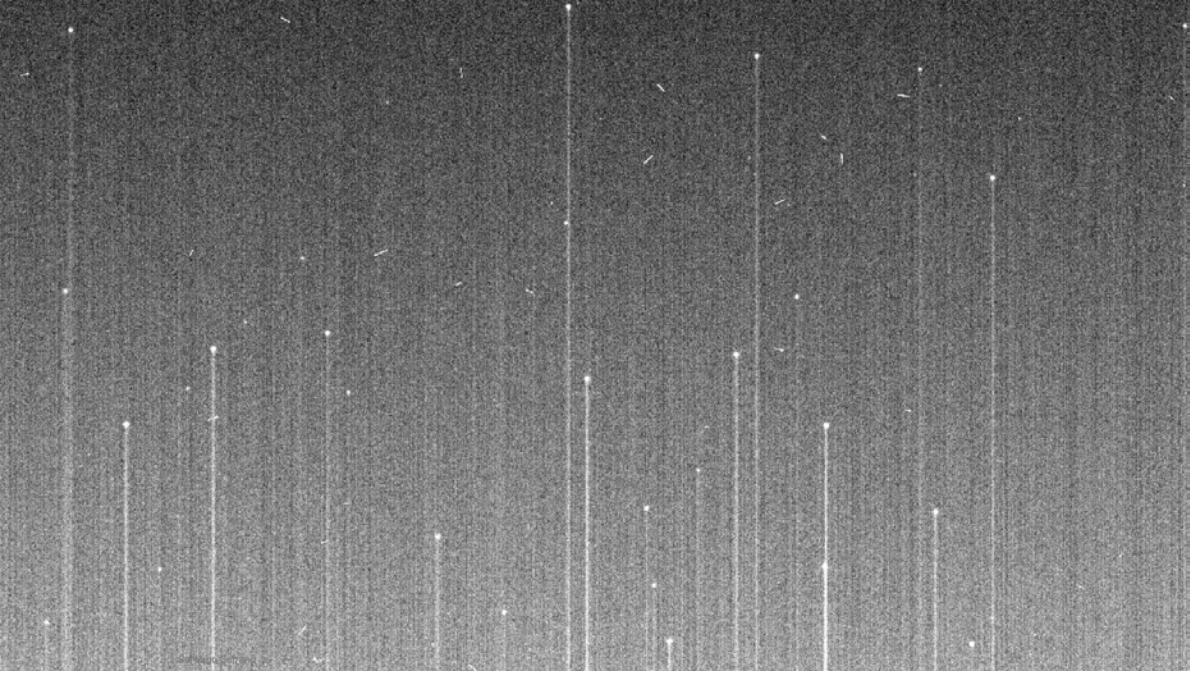


Figure 2.2: Guide image simulation with the effect of open shutter, caused by the incoming light during the readout process.

Additionally we used SKYMAKER¹ (Bertin, 2009), as it provides more complex optical PSF, simulating the pupil from the primary mirror, obscuration from spider arms, diffusion from optical elements and even its own implementation of the atmosphere turbulence on short exposures. This also allowed an independent validation of our own tool.

2.2.4 Open shutter effect

The detectors dedicated for guiding require fast and continuous readout while the main exposure is integrating light. This implies that the shutter cannot be closed to prevent light hitting the guide detectors while these are in the readout process. Therefore a trail of light appears in the read direction as light from sources continue to emit while CCD charges are moving towards the readout register. This is a very specific feature that we implemented on both GuideSIM and SkyMaker pixel simulators, and will affect the centroiding performance, as well as the background subtraction, as it will contain a significant gradient. The only additional parameter required is the read speed at which the detector shifts the rows.

¹<https://www.astromatic.net/software/skymaker>

2.3 Survey Guiders

In this study we have designed, implemented and tested the auto-guiding systems for three leading cosmology projects: DES, PAU and DESI. A brief description of each survey, the host telescopes, the peculiarities of each algorithm and the performance on sky is described in the corresponding subsections.

2.3.1 Dark Energy Survey

The Dark Energy Survey² (DES) (The Dark Energy Survey Collaboration, 2005) is an ambitious international project aiming to map hundreds of millions of galaxies in research of the unknown dark energy that causes the universe to accelerate its expansion. The survey began in 2013 and after six years, in 2019, the observations were completed covering a total of 5000 deg² of the southern sky. The survey has selected carefully the area in the sky convenient for the cosmological measurements, outside the galactic and ecliptic planes, where contamination from local sources is higher.

DES studies the effects of dark energy with four main probes: supernovae, revealing the expansion of the universe with precise distances. Galaxy clusters and its relation with the predicted halo mass. Baryon Acoustic Oscillations from galaxy clustering at distances and number of galaxies never tested before providing information of the cosmic expansion rate with unprecedented precision. And finally Weak Lensing, a gravitational effect from distant galaxies when they travel through matter in the way towards us. This final effect requires measuring very precisely the ellipticity of galaxies and the auto-guiding requirements will be driven by the accuracy required to measure this extremely subtle effect.

2.3.1.1 The SMARTS 1-meter telescope

During the design phase, and prior to the camera to be completed and installed in its host telescope, we tested our guiding system prototype with an engineering model of the detector in the SMARTS 1-meter telescope, at the Cerro Tololo Interamerican Observatory (CTIO). This is a relatively small equatorial telescope sitting next to its big brother, the Victor M. Blanco telescope. Constructed by the Yale University to search for Near Earth Objects, it resulted in a very useful host to perform early tests. In early 2004 the telescope installed its own auto-guiding system, where we could compare performances and set a performance benchmark.

2.3.1.2 The Victor M. Blanco 4-Meter telescope

The majestic Victor M. Blanco telescope is a 4-meter aperture telescope located also at the Cerro Tololo Interamerican Observatory, constructed in 1976 and the largest of its size in the

²www.darkenergysurvey.org

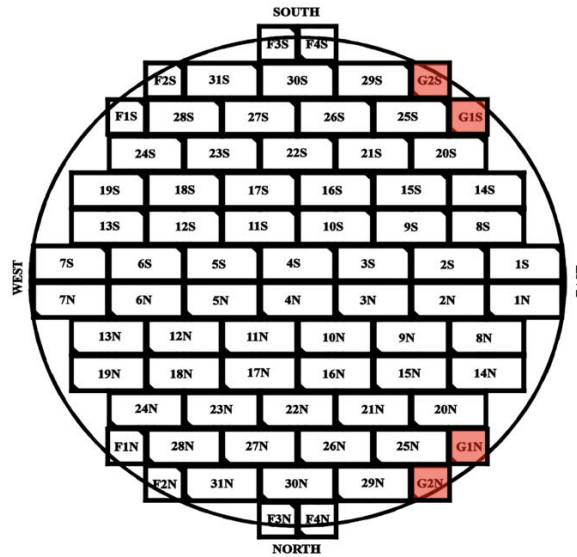


Figure 2.3: The focal plane layout of DECam with all 74 CCD detectors. 62 are dedicated for science, 8 for focusing and 4 for guiding (highlighted in red).

southern hemisphere until 1998. Constructed in an equatorial mount, this is one of the largest telescopes in this kind of mount still operating. Due to its design nature it delivers an outstanding good tracking accuracy in elevation, better than in azimuth. Thanks to the Ritchey-Chrétien configuration, it projects a wide field of view, necessary for the DES survey where covering large areas of the sky in each exposure is necessary.

2.3.1.3 The Dark Energy Camera

The Dark Energy Camera (DECam) (Flaugher et al., 2015) is a high-performance wide field imager, designed specifically for the Dark Energy Survey. It is composed by an optical corrector of five elements, a set of broad band filters (ugrizY) and 74 fully depleted CCD detectors inside a cooled vacuum Dewar. The camera is mounted in the prime focus of the Victor M. Blanco telescope, yielding a field of view of 3 square degrees. With its very sensitive array of 570 Mpixel, it can map large areas of the sky in a very short amount of time.

From the total set of detectors, there are four dedicated to guiding (as shown in Figure 2.3), around the edges of the 62 imaging ones in the partially-vignetted area of the focal plane. Similarly, there are eight detectors located at slightly different optical path lengths than the fiducial focal plane aimed for focusing. Even that the size of both guiding and focusing detectors is half that of the science ones (2k x 2k vs 4k x 2k), the CCDs are identical for all applications.

2.3.1.4 Algorithm

The front-end electronics allow an independent readout of the guide detectors, while the imaging array is integrating light. In this way we can continuously monitor the guide stars, independently from the length of the science exposure.

The DECam guider system is designed to support closed-loop telescope tracking to stabilize the exposures to an accuracy of a few tenths of an arcsecond. The guider uses information from four $2k \times 2k$ CCDs located at the edge of the DECam focal plane. During a science exposure the guide CCDs are operated in region-of-interest mode, which read only a subset of the CCD (typically 50×50 pixels) instead of the whole detector, dramatically reducing the 10 second readout time to few milliseconds. These ROI are read out continuously with a typical exposure time of 600 ms. The postage-stamp images of the guide stars are rapidly analyzed by the guider software, which looks for shifts in the position of the stars detected. Every second a position error signal is sent to the telescope control system to close the tracking loop.

The guide algorithm is the core of the guider software. It determines the change in guide star position and delivers the error signal to the Telescope Control System (TCS). Following this objective the guide algorithm handles three main tasks: Acquisition - Find and set the best reference star for guiding (once), Tracking - Follow the reference star with the best signal to noise, and Compute offset - Determine and deliver the offset to the TCS.

For each DECam exposure, the guider executes the acquisition task once during the initialization stage. For the rest of the exposure the guider operates in a loop continuously running the tracking and the compute offset tasks. Acquisition is carried out during the initialization phase. The goal is to select the stars best suited for guiding from the stars available in the full frame images from the 4 guide sensors. The SExtractor software (Bertin, Arnouts, 1996) is used to identify stars. It has proven to be a fast, precise and reliable tool for this task. On average, SExtractor provides more accurate stellar positions for stars that have a higher signal-to-noise (S/N). Before selecting the highest signal-to-noise stars, the guide algorithm applies a filter to remove the following from the list of detected objects: stars with one or more saturated pixels, double stars, stars close to or truncated by a chip boundary, stars with neighbors that are bright and close enough to significantly bias its flux, galaxies, cosmic rays, and detector defects.

In addition, the algorithm searches for isolated stars within a user-defined range, to avoid possible confusion with the guide star during the tracking stage. Once a guide star is selected for each of the available guide CCDs its position is taken as the zero-point location in the detector for this exposure.

After the acquisition phase when guide stars have been set, the system is reconfigured to readout only a Region of Interest (ROI) around each guide star. As previously said, this dramatically reduces the readout time of the guide CCDs from 10 seconds to read the whole detector to few milliseconds just for the ROI. This is achieved by skipping the rows and columns that do not overlap with the ROI at a much faster readout speed. The ROI size is a configuration

parameter. The default area is 50×50 pixels, which at the scale of $0.27''/\text{pixel}$ of DECam, corresponds to a box of $13.5''$ wide, enough to cover large PSFs plus the telescope tracking errors. A signal is sent to the online system to start the exposure and to readout the four guide sensors in a continuous loop. Each set of postage stamp images is analyzed by the guider software and the correction signal is sent to the TCS.

Tracking corrections are sent at a rate of about 1 Hz. The system is designed to minimize jitter between updates. During the tracking phase, the guider attempts to find the reference guide star selected during the acquisition phase, typically a fraction of a pixel away from its original position. For this task we use PyGuide³, a simpler image analysis tool that provides fast and accurate centroids at a constant rate. The critical task during this step is to ensure the correct match with the sources found in the postage stamp images to the initial set of guide stars. For the guider to achieve the required performance, it is very important to avoid losing or mismatching the reference guide stars. For this purpose the guider algorithm employs a user adjustable search range surrounding the zero point position set during the acquisition phase. The search range is large enough to have the true guide star inside (larger than the maximum tracking error) and small enough to cover just few stars from the several available in the ROI image. Ideally only the reference guide star will be within the search range as non-isolated stars are discarded at the Acquisition stage.

In case there is more than one detection match, the algorithm chooses the star with most similar flux to the reference guide star flux. The guide star flux is updated after every iteration to safeguard against changing sky conditions during the exposure. In general, only the reference star is located in the search range but in case we find more than one star, this method is very efficient in identifying the guide star.

With the guide stars detected in the postage stamp images, computing the offset from a single CCD is as simple as measuring the distance from the current position of the guide star found and the previous zero point. It is slightly more complicated when multiple guide stars have to be considered. A simple way to combine the information from the multiple guide stars would be to average the offsets. However, since it is known that the position accuracy is higher in high S/N stars, the offset is averaged by weighting the offset with the corresponding guide star's S/N. By using the weighted mean, the more reliable stars have more importance in the final offset. In addition, the guider algorithm uses information about the agreement/discordance among the tracking error information given by the different CCDs. This means that if three CCDs provide a very similar signal whereas the fourth suggests a different correction, the algorithm discards the outlier and does not include this information when calculating the correction offset. This selection is accomplished by measuring the Euclidean distance between the offsets from all guide stars. If the discordance exceeds a threshold, the highest discordant guide star is discarded. The threshold is adjusted for every iteration depending on the correction signal differences found

³<http://staff.washington.edu/rowen/PyGuide/Manual.html>

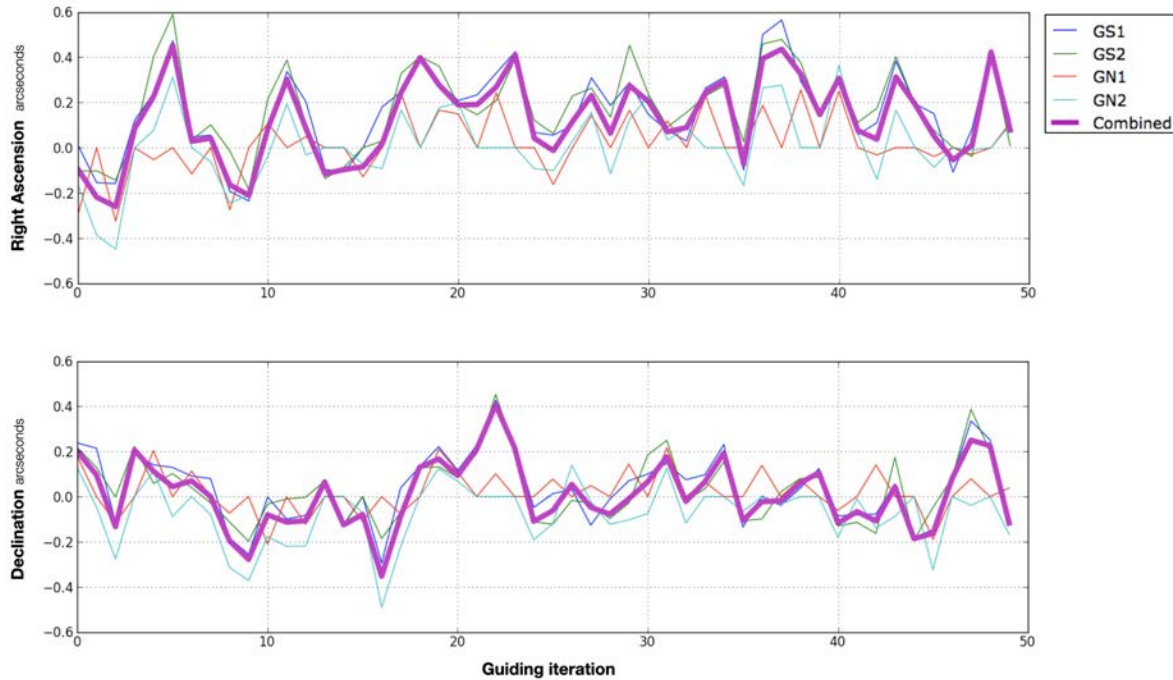


Figure 2.4: This example shows how the multiple guide signals follow a random pattern around the center (driven by the minor telescope tracking errors) with a standard deviation of $0.2''$, similar to the pixel size. In this particular example, guide CCD GN1 shows a worse performance due to a lower S/N guide star.

among the available guide CCDs. In other words, more discordant signal offsets will have a more tolerant threshold than very accurate signals. Having an adaptive threshold is important as different atmospheric conditions may lead to very different discard thresholds. During the main guiding loop, the algorithm removes a guide CCD from further processing if it has to be discarded for more than five consecutive iterations. It will also be disabled if the guide star cannot be found for the same number of consecutive times. The number of failed iterations can be adjusted by the observer.

2.3.1.5 Performance on sky

The combined correction offsets as well as the results from individual guider CCDs for a typical DECam exposure are shown in Figure 2.4. The integrated view of the correction signals in Figure 2.5 shows the RA/Dec corrections peaking at the target coordinates which is another indication of the excellent tracking performance of the Blanco telescope.

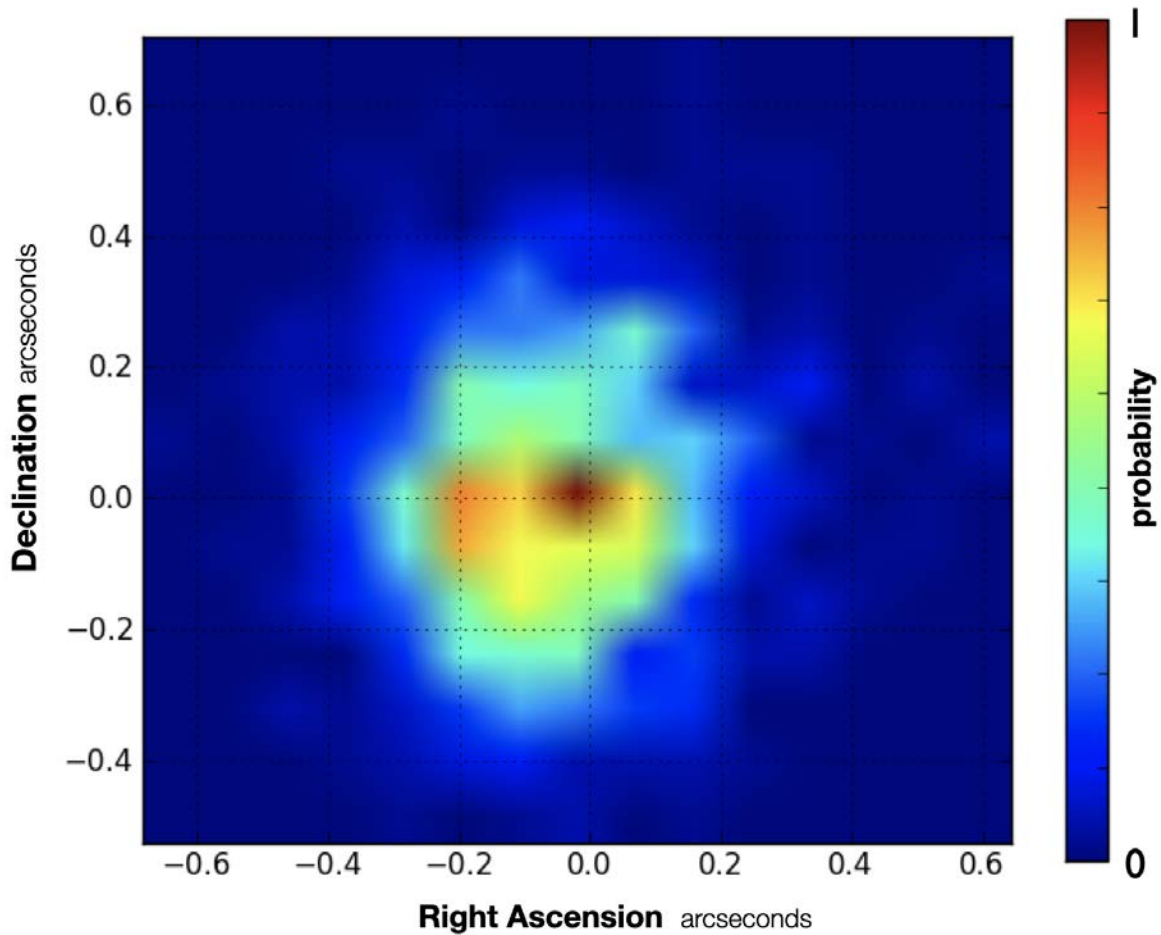


Figure 2.5: The X/Y corrections of the combined signal at the Blanco telescope with DECam guiding system.

2.3.2 Physics of the Accelerating Universe Survey

The Physics of the Accelerating Universe (PAU) Survey is an innovative project that will help to better understand the yet unknown nature of dark energy and dark matter of the universe. With a very peculiar camera specifically designed to optimize photometric redshifts, the survey will provide unprecedented redshift accuracy for millions of galaxies.

2.3.2.1 The William Herschel 4.2-Meter Telescope

The William Herschel Telescope (WHT) is a large 4.2 meter telescope located at the observatory of El Roque de los Muchachos, in the Spanish Canary island of La Palma. When it was built, it was the third largest telescope in the world, and it is currently the second largest telescope in Europe, after the Gran Telescopio de Canarias. It has an Altazimuth design and can mount instruments on its prime focus, in the Cassegrain focus and on the two Nasmyth foci. The PAU Camera is

mounted on its prime focus, where the field of view is the largest. Due to its alt-azimuth design, it requires a de-rotator to compensate for the field rotation. Even though it has a moving mass of 186 tons, it delivers an outstanding pointing accuracy of 2 arcseconds RMS and a tracking accuracy of 1 arcsecond over 10 minutes without autoguiding. The rotator tracks very accurately its field rotation and therefore there is no need to correct for rotation for the applications of the PAU survey.

2.3.2.2 The Physics of the Accelerating Universe Camera

PAUCam is an 18-detector imager camera (Padilla et al., 2019) with a very particular set of narrowband filters specifically designed to optimize the determination of photometric redshifts. It contains the typical six broad band optical filters ugrizY, plus an additional 40 narrow band set covering from 450 nm to 850 nm in steps of 10 nm. Due to the large number of filters, these cannot cover the whole detector mosaic, and are distributed in filter trays with different filters sitting on top of each detector. To reduce detector's dark current, the focal plane is cooled down inside the cryogenic and vacuum dewar with liquid nitrogen in parking position and with cyrocoolers while observing in the prime focus. Only the eight central detectors are covered with narrow band filters as the external ones are highly vignetted and present stronger optical distortions. The surrounding detectors are therefore used for calibrations and guiding purposes. A schema of the focal plane in the WHT can be seen in Figure 2.6.

2.3.2.3 Algorithm

As described in the WHT section, the telescope delivers an excellent tracking performance, of 1" over 10 minutes of pointing. As the PAU survey exposures are relatively short, with a maximum integration time of 180s, guiding is not strictly mandatory. However the PAU Camera is also used by community observers that may expose longer, therefore we need to implement a guiding system that improve the open loop performance.

When the guiding system is enabled, two of the outermost corner detectors of the focal plane (G12 and G17) are devoted to guiding, leaving the science exposures to be done with the remaining 16 CCDs, including the most illuminated detectors in the focal plane.

When the shutter opens at the beginning of the exposure, the guiding detectors continuously take images with exposure times between 1 and 5 seconds, depending on the sky conditions and stellar density. The guider software then analyzes the images from these two guiding detectors and, using the stars available in the field, identifies the telescope tracking error and sends the correction signal to the TCS interface.

To minimize the latency of the correction signal sent to the TCS, only a subset of the detector is read out. This subsection is called Region Of Interest (ROI), which is typically adjusted to 1000×2000 pixel subset in the most illuminated side of the detector. In this area, the image is quite distorted and partly vignetted.

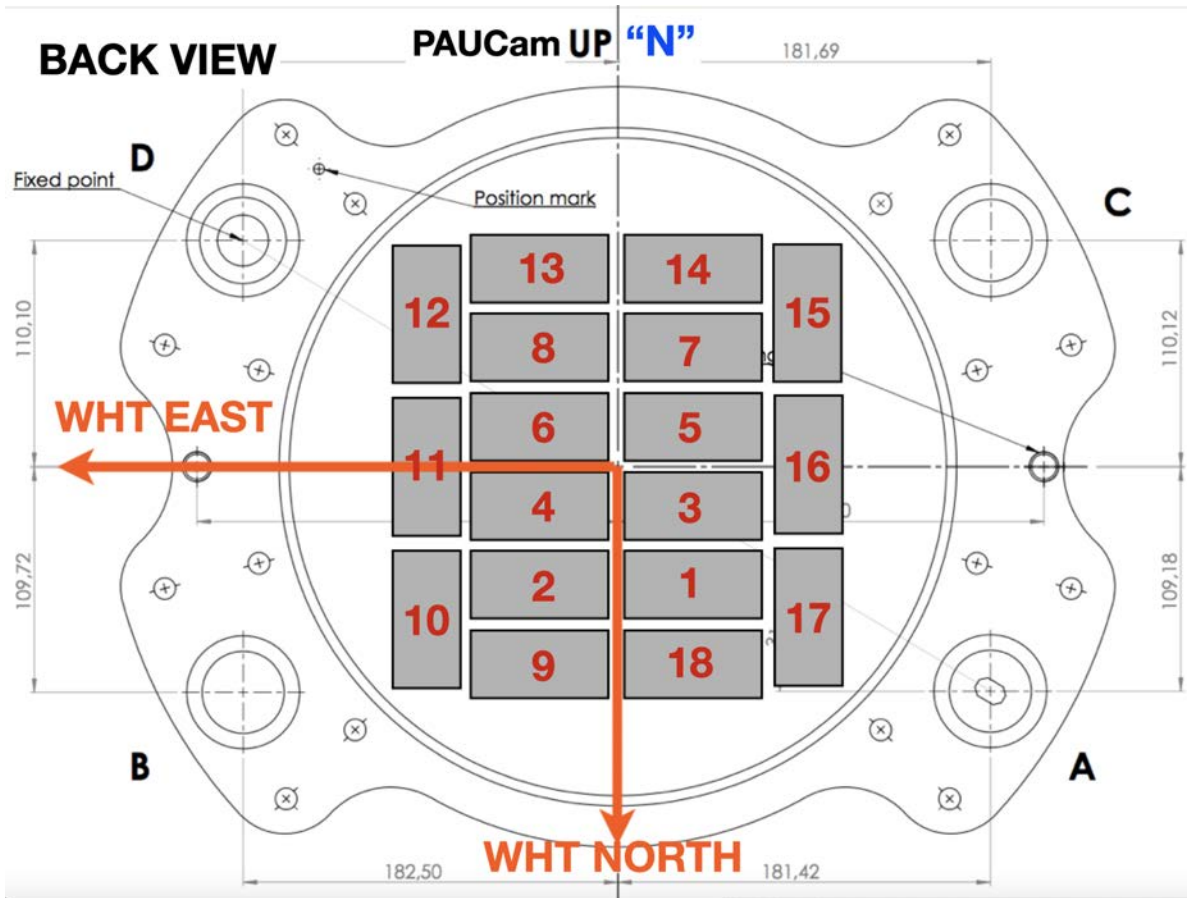


Figure 2.6: PAUCam focal plane in WHT sky. Detectors 1 to 8 hold the narrow band filters while the external ones have broad band filters for calibration verifications. Detectors 12 and 17 can be used for the integrated auto-guiding system.

In the first place the guiding algorithm will quickly reduce the image, subtracting the bias signal and correcting the gain. The image is then analyzed by SExtractor, providing a list of source detections. These detections are filtered to ensure that stars are neither blended nor very close to each other, nor in the border, unaffected by cosmic rays, etc. The remaining list of candidate guide stars are selected and will be searched in the next iterations when the new guiding images arrive. Depending on the stellar density of the field, 3 to 15 guide stars are selected on each guide detector. Stars that are not found in new iterations will be removed from the list, assuming that might be moving objects, badly selected cosmic rays in the first place or too low signal to noise stars that went below the detection threshold. Having multiple sources as guide stars provides more accurate guiding, even under this area with more distortion of the focal plane. The individual tracking errors found in each star are combined using a weighted average by each star S/N ratio.

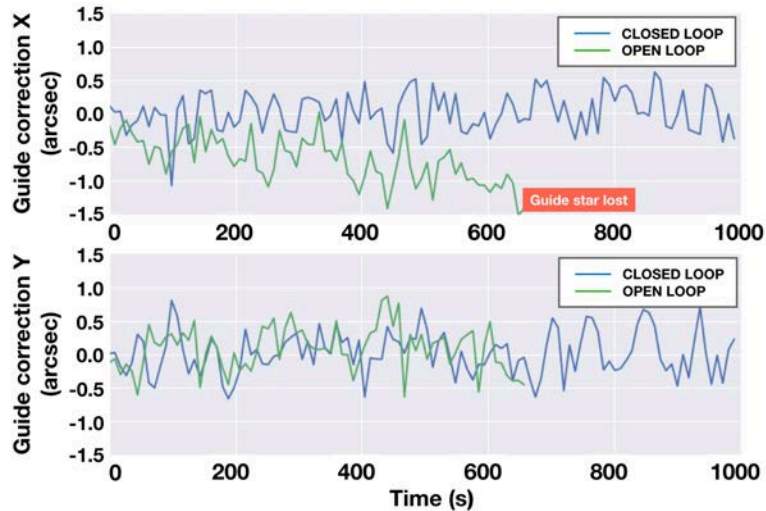


Figure 2.7: Guiding Correction on X and Y as a function of time for open and closed loop configuration.

2.3.2.4 Performance on sky

After integration, readout and analysis, the error signal is sent to the TCSi every 6 to 8 seconds. This correction rate is enough to ensure a stable telescope tracking within sub arcsecond accuracy for an unlimited exposure time.

Figure 2.7 demonstrates how the guider corrects the telescope drift when guiding in closed loop. The figure shows the guiding correction computed in X and Y for the case where the signal is sent to the TCSi (closed loop) or when the correction is not applied. One can see that the tracking of the telescope can track the stars even without auto-guiding up to around 4 minutes but the stars are completely lost when longer exposures are made.

2.3.2.5 Auto-Focus system

In order to set the camera in focus, we have designed a specific readout and analysis system that provides accurate measurements of the focus position at each detector. To speed up the acquisition of the image at different focus positions, we produce a single exposure image, where charges are shifted between the different steps, producing an image where the same source (star) is exposed several times, each at a different focus position. Figure 2.8 shows an example of such an image, where the same star is seen in different places of the CCD separated by a fixed number of pixels (a jump). The first focus is identified because the distance to the second one is doubled (a double jump).

Once the image is read and built, the control system automatically triggers the analysis of this ‘stacked’ focus exposure. In the first place the raw image is quickly reduced applying a gain and overscan correction. Over the reduced focus image, we run SExtractor to obtain a

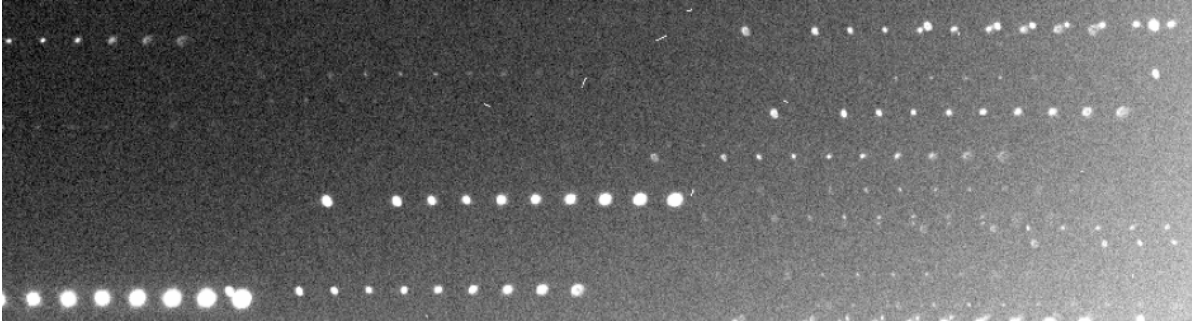


Figure 2.8: Focusing image resulting from a focus search iteration. For every star, several images are taken in the same file.

list of detected sources. As we need to identify the pattern of one double readout jump, plus N jumps, we create an R-Tree of all detected stars, which allows us to perform a very fast spatial matching, keeping the sequences that match all the jumps. The measured FWHM of each star in all the focus sequences identified for each CCD detector are median combined to produce a master curve. This curve is fitted to a 6th-degree polynomial that allows to automatically detect the focus position that minimizes the FWHM of the PSF. Figure 2.9 shows an example of a focus sequence analysis, automatically performed by the PAUCam Control System. The top plot shows the FWHM as a function of the focus position for all analyzed stars. Some of the stars have low signal or they are in a more distorted area of the focal plane. The red curve is the mean of all measurements. The middle plot shows the average per CCD. One can see small variations because of the tolerances in the CCD fabrication and installation process. The bottom plot shows the results with the fitted polynomial and the preferred focus value.

The results are presented in the GUI application to allow the observer to perform any needed focus correction.

2.3.2.6 Tip & Tilt analysis

To ensure that the focal plane was installed flat in the prime focus of the WHT, we performed a tip-tilt analysis during the commissioning, and in the later installations for the next operation runs. To carry out this analysis, a sequence of defocused images is taken, around the default focus position at 82.3mm . Figure 2.10 shows the same portion of the image taken in the tip-tilt tests at four different focus positions. The numbers show the measured FWHM of the detected stars.

At each focus position the FWHM of the stars in each CCD is measured, providing an average PSF FWHM per CCD. Similarly to the focusing algorithm, the position of the focus (where the FWHM minimizes) is estimated in every CCD sequence (see an example for CCD3 in Figure 2.10) providing 18 focus points in the focal plane.

When the focal plane is completely flat to the optical path at the prime focus, the individual detector focus should lie at the same focus position. In case the camera does not sit completely

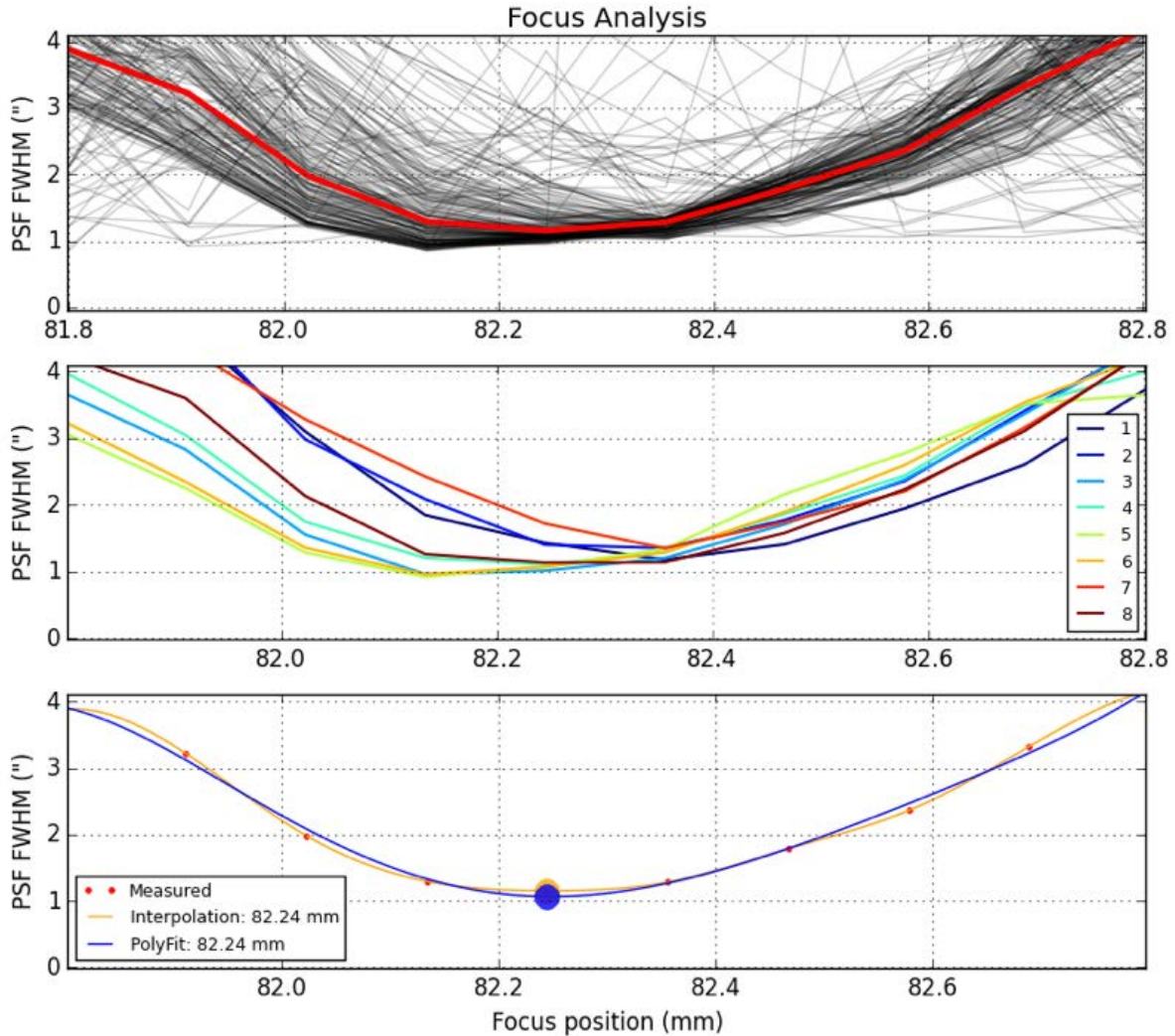


Figure 2.9: Focus analysis: Top, with all stars, middle with the average results per CCD and bottom with the fitted result and the preferred focus value.

flat in the telescope mount, the three of the four mounting points (one is fixed) are adjustable to correct for the possible tip and tilt miss-positioning.

To estimate the amount of tip and tilt, a 3D plane is fit to the focus positions (z) at the CCD positions (x,y). Following the mechanical drawings the position of the Kinematic Mounts are identified and the tip-tilt code returns the offset that needs to be applied for a flat focal plane (see Figure 2.11). Due to the fact that there is a significant amount of distortion in the 10 outer detectors that may alter the result, only the 8 central detectors were used in the 3D fit.

Thanks to an accurate design, verification and installation of the camera, all the analysis resulted into a very flat focal plane from the very first installation during commissioning, to the later installations of PAUCam in the next observing runs. The corrections resulting from

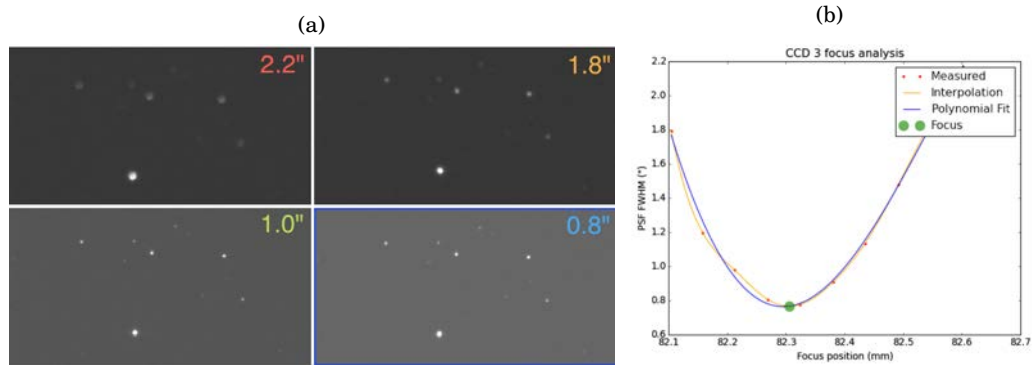


Figure 2.10: Tip & Tilt processing. (a) The same portion of the image taken in the tip-tilt tests at four different focus positions. The numbers show the measured FWHM of the detected stars., (b) Position of the best focus point for CCD3 during the tip-tilt analysis.

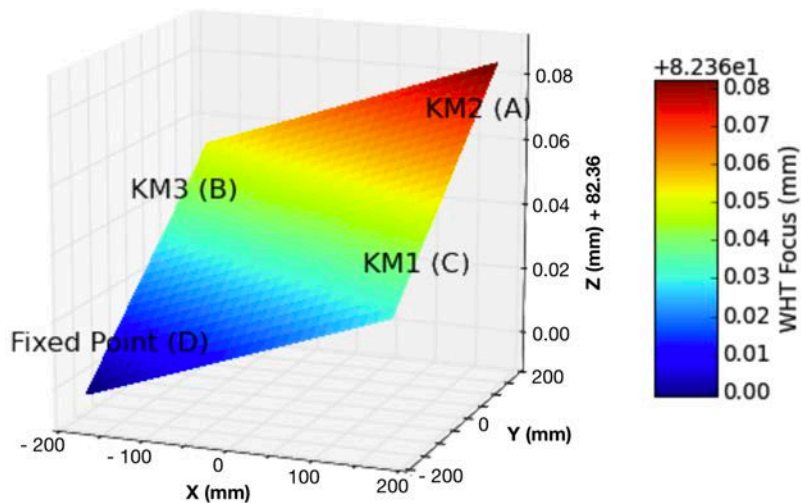


Figure 2.11: Tip & Tilt result. Offset to be applied in every adjustable mount to have a completely flat focal plane. All analysis resulted into a very flat installation that needed no mount adjustment.

this analysis where smaller than the resolution of the three mechanical mounts ($< 10 \mu\text{m}$) and therefore no adjustment was applied after any installation.

2.3.3 Dark Energy Spectroscopy Instrument and Survey

The Dark Energy Spectroscopy Instrument (DESI) is a very ambitious camera made to carry out a large cosmology survey measuring the effect of dark energy on the expansion of the universe. The most accurate measurement of radial distance is done with spectroscopic redshifts, and DESI can measure 5000 sources in each exposure, much faster than any other instrument built up to date. This allows one to study large areas of the sky, necessary for any cosmological survey. Over its 5-year mission, DESI will obtain spectra for approximately 30 million galaxies over one-third of the night sky, mapping a huge volume of space stretching over 11 billion years into the past.

The instrument contains 5000 robotic positioners that can adjust the optical fibers to its exact target location. These send the light through 50 meters of fiber down to 10 spectrographs that extract the spectra of each source. In order to align the instrument with the sources on the sky, there are 10 independent detectors called Guiding, Focusing & Alignment (GFA) Cameras. These devices correct the telescope positioning and tracking to point exactly where it is expected for the light of the 5000 sources to enter exactly in the fiber hole, each with its 1.5 arcsecond diameter core. The requirements of alignment and guiding are therefore much restrictive for DESI than in the other two surveys.

2.3.3.1 The Mayall 4-Meter Telescope

DESI is installed at the Mayall reflector Telescope located at the Kitt Peak National Observatory, in Arizona. This 4-meter telescope was built in 1970 and it is a twin of the Blanco Telescope in Chile that hosts DECam, with the same characteristic equatorial mount. Therefore, a similar pointing and tracking performance is expected with respect to the Chilean telescope.

2.3.3.2 Algorithm

In DES and PAU, the guiding system could grab the guide stars wherever they were and guide at that position. If the telescope was not pointing at the exact target location (within sub-arcsecond accuracy) the post-processing astrometry calculation would correct for that. However in DESI we need to ensure we are at the exact target location before starting the observation and we remain at this position throughout the whole exposure within a small tolerance. For that purpose, we use six out of the ten GFA cameras to make sure we are in position and perform the auto-guiding procedure. The astrometric solution is analyzed by a tool called PlateMaker that uses the full GFA image readout and determines the necessary offset to be in position. This relatively large offset is sent to the telescope and a second round of PlateMaker and full GFA readout is done. Now being much closer to the target position, PlateMaker decides which guide stars to use, sends

its pixel positions and decides at which sub-pixel position they need to be. The Region of Interest for the six guider GFAs are configured and only the postage stamps are sent to the guider. Up to four ROIs per GFA are allowed resulting into a maximum of 24 guide stars to be tracked by the guider.

The guider then analyzes the postage stamps, searches for the guide star in the expected location by PlateMaker and computes the error \vec{e}_{star} to the target location such as:

$$(2.1) \quad \vec{e}_{star} = (x, y)_{target} - (x, y)_{centroid}$$

As the pixel array of the GFA detectors is rotated across the focal plane, we need to use the astrometric World Coordinate System (WCS) information for each camera to transform the offsets into a common frame that is sent to the TCS. The offsets are then transformed to equatorial coordinates and sent directly to the telescope for correction. To combine the guide errors we take into account the signal to noise of each guide star in a weighted average such as:

$$(2.2) \quad \vec{e}_{combined} = \frac{\sum \vec{e}_{WCSstar} s/n_{star}}{\sum s/n_{star}}$$

where $\vec{e}_{WCSstar}$ are the individual guide star errors transformed to the common frame and s/n_{star} are the estimated signal-to-noise ratios. The telescope applies a proportional feedback f_{TCS} on the error signal sent by the guider to avoid tracking instabilities, typically set to 30%. There are other complex options that the telescope can apply to deal with the guide correction such as a PID loop or a Kalman filter than are not within the scope of this study and will not be discussed here.

As previously said, the target position of the guide star defined by PlateMaker may be pixels away from the current position. In the process of moving the guide stars to its target locations, we set the status of the guider to ALIGN status. In this mode, the guider searches the guide stars of new incoming guide images at its expected location, and not at the target position. The expected location will be:

$$(2.3) \quad (x, y)_{expected} = (x, y)_{centroid} + (\vec{e}_{combined} \times f_{TCS})$$

as we need to take into account the rate at which the telescope will be moving towards the target location.

We define a tolerance radius and number of guide iterations inside the target location to transition between ALIGN status and GUIDE status. Once in GUIDE status, the expected location will always be the target location as we should be moving around this position.

Similarly to other guiders, we define a number of iterations where a guide star is not found at its expected location before disabling that particular ROI.

In addition to the default **PlateMaker mode**, we implemented a simpler mode called **Self Selection**, where the guider selects by itself the guide stars from the full readout image, simplifying the interfaces and allowing to debug the guider by itself. Summarizing, we have two guide modes that perform the following steps:

- **Mode 1 - PlateMaker** (default)
 1. **ALIGN**: Slowly move the current positions to the target positions defined by PlateMaker.
 2. **STABILIZE**: Wait until the current positions are around the target positions for a number of iterations.
 3. **TRACKING**: Maintain guide stars at the target positions constantly sending the measured errors to the telescope.

- **Mode 2 - Self selection**
 1. **SET STARS**: Analyze the full readout image and select the best guide stars. Set the star positions and configure ROIs.
 2. **TRACKING**: Maintain guide stars at the original position constantly sending the measured errors to the telescope.

Self Selection mode is simpler and convenient for debugging the system as it does not require PlateMaker, but cannot be used in operation as it does not ensure the telescope is pointing at its target location and therefore the light from the sources may not be aligned with the fibers.

2.3.3.3 Performance on sky

Before installing the final DESI camera, two prototype instruments were tested at the Mayall telescope. Here we describe the goals and results of each of the three installations.

ProtoDESI

In summer 2016, the ProtoDESI (Fagrellius et al., 2018) experiment was built to demonstrate that robotic positioners could be aligned to receive the light of a source in its optical fiber. In protoDESI we had a single detector dedicated to guiding. Even with this simple setup, it was a great challenge as several key steps had to be tested for the first time. It was the first time we interacted with PlateMaker, and guided in catalogue mode, where stars were pre-selected by an external software. It took some time to debug both codes, but we finally managed to put the guide star in the middle of a 50x50 pixel stamp and the guider started guiding directly with it. Furthermore, we tested the alignment process, where the guider moved its guide stars from an original position to its target destination, bit by bit. An example of the alignment process and guiding with PlateMaker and Guider code can be seen in Figure 2.13, where the original guide star was more than 2 arcseconds away, and the guider moved it back to its center location in

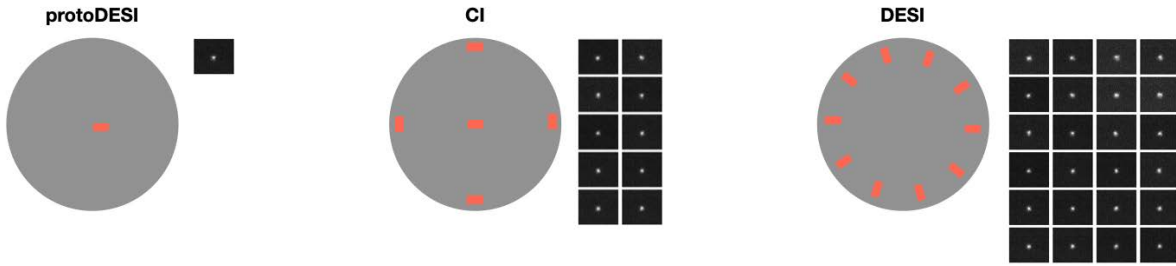


Figure 2.12: The three different commissioning scenarios. In first place we had ProtoDESI, a single GFA experiment for preliminary validation of the system and the robotic positioners. In a second step, and with the final DESI corrector, there was the Commissioning Instrument, with 5 GFA-like cameras. And lastly the DESI instrument with its 10 GFAs, where 6 of them are devoted for guiding.

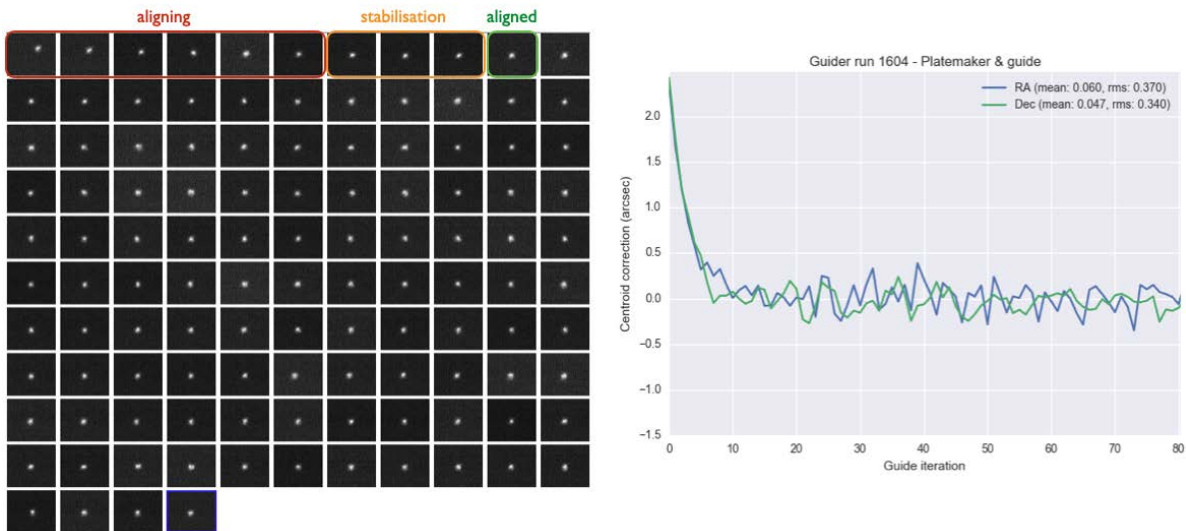


Figure 2.13: An alignment and guiding test in ProtoDESI where the guide star was originally ~ 2 arcseconds away from its target location defined by PlateMaker and the guider positioned it and continued guiding further with it.

6 guide iterations. After the alignment and stabilization, the guider kept the guide star at its center location for more than 80 iterations.

Further tests were done to ensure the guider was capable of recovering sudden jumps throughout the tracking. During a guiding process, we manually shifted the telescope few arcseconds and watched the guider correcting the jump at the opposite direction and drift it back to the center. The guider successfully recovered jumps up to 2" in both Right Ascension and Declination axes. Figure 2.14 show an example test where the telescope was shifted 1" and the guider corrected and continued guiding at the original location.

A final stress test was made to see for how long the guiding system could keep tracking.



Figure 2.14: A guiding test in ProtoDESI where we manually moved the telescope tracking by 1" and the guider picked up the jump and recovered back to its center to continue the guiding process.

Typically DESI exposures are 20 minutes long, therefore we need to ensure we can guide at least during this period of time. Even with ProtoDESI's single detector and single guide star, we guided for more than an hour until we manually aborted the exposure, proving that the guiding system was stable.

The Commissioning Instrument

The Commissioning Instrument (CI) was installed at the DESI focal plane behind the final DESI Corrector in 2019. The number of sensors dedicated to guiding increased to 5 (as seen in Figure 2.12). Additionally the CI installed 22 fiber-illuminated fiducials, a set of fibers fixed in the focal plane that when illuminated from the back, it can be seen by a small camera that points towards the focal plane, allowing to model optical distortion and mechanical deformations. This is essential as those models allow to place the robotic positioners at the exact target location.

With an increased number of detectors and distributed along the final corrector, we would test guiding multiple stars at different detectors, and with rotated arrays. This implied using a calibrated set of WCS for the five cameras, a major difference compared to previous systems. Furthermore, as multiple stars were selected by catalogue from PlateMaker, it could be that inconsistent positions were found as different stars would indicate different directions of guiding.

We successfully integrated the multi-detector guiding with WCS in CI. In the first attempt, the performance was not yet reaching the requirement goal as two main issues were raised during the CI run. In first place, we found the telescope to oscillate tracking after some minutes of guiding. That was identified to be an issue of the Proportional-Integral-Derivative (PID) loop. This effect can be seen in Figure 2.15 as the large oscillations. The second issue was that guide stars seemed to separate apart into inconsistent directions as the exposure was progressing. This

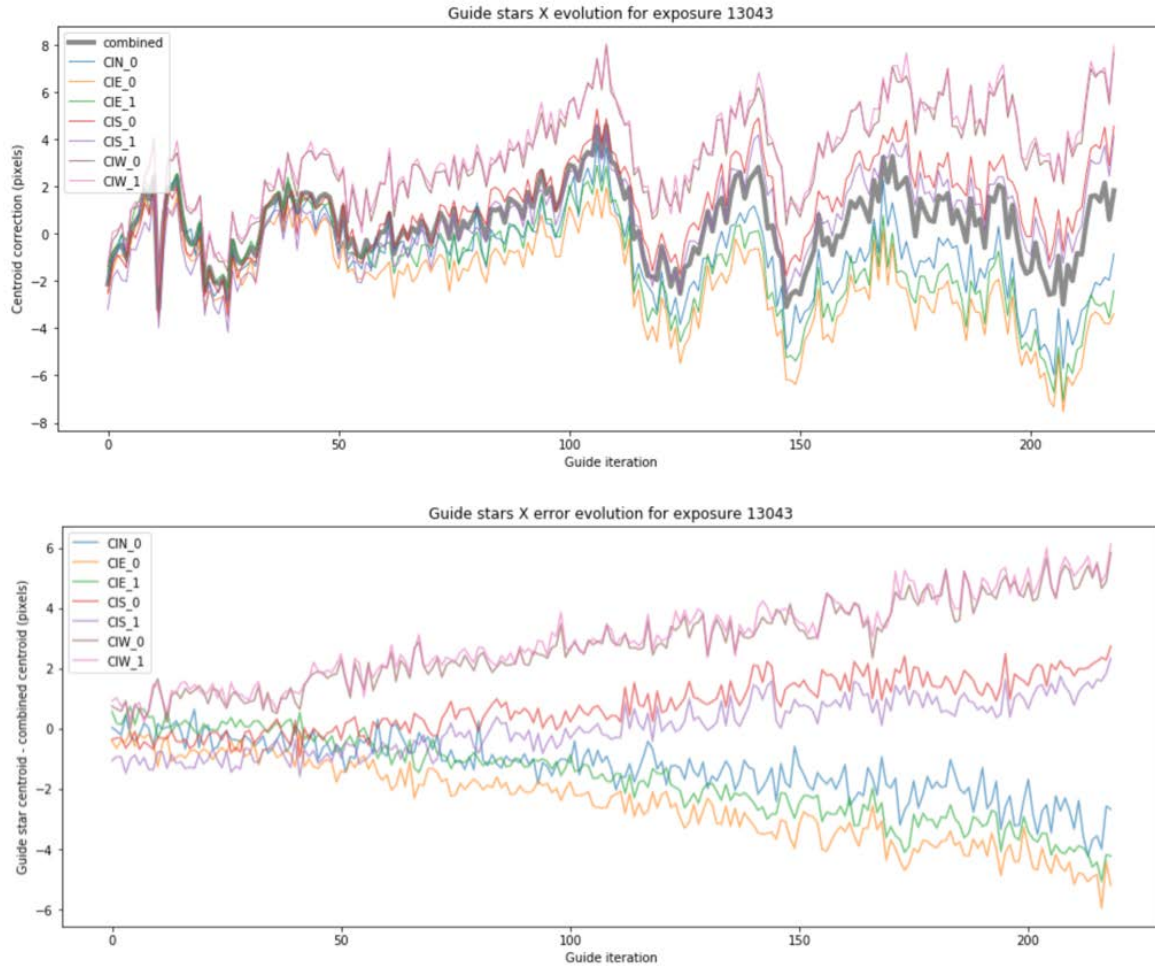


Figure 2.15: A guiding sequence in the Commissioning Instrument test, where the unstable tracking and field rotation is clearly visible.

was identified to be related to field rotation. Being an equatorial mount, field rotation was not expected to be present, however due to precession and nutation, the axis of Earth precesses on a timescale of ~ 26.000 years. As the telescope was built almost 50 years ago, there is a very subtle polar mis-alignment that implies focal plane rotational errors of 0.002 deg. In most cases this effect is negligible, but due to the large field-of-view of DESI and tight requirement on tracking to ensure light enters into fibers in such large 8 deg^2 focal plane, the effect had to be corrected. Even if equatorial telescopes do not have a (de)rotator to compensate for field rotation, the complex mount of the DESI barrel contains a hexapod that is capable of performing a slight rotation to compensate for it.

With the guiding data, composed mostly from guide-star centroid measurement sequences from the various cameras, we designed a software that could estimate the field rotation throughout a sequence. In all previous guiders we assumed the guiding errors are just a translation in the

2D tangential plane, and this is simply corrected by averaging the offsets between the measured guide star centroid and its original position in each axis. However adding rotation significantly complicates the calculations as individual guide star centroids cannot be treated independently and a global solution that involves rotation and translation is required.

The solution we found is to implement a Kabsch algorithm (Kabsch, 1976) based in a Singular Value Decomposition (SVD) where the rotation matrix R between two datasets of points could be computed in a global solution. The first dataset P_{target} is the target positions (where PlateMaker specified to the guider that guide stars need to be to stay in place) and the second dataset $P_{centroid}$ is the measured centroids at a given iteration. Both datasets need to be in focal plane coordinates and therefore we need to transform from the local detector positions using the WCS of each GFA. First we need to re-center both datasets to the center such that only a rotation transformation at the origin can convert one dataset to the next. We build the covariance matrix as the product of the target vector at origin with the centroid vector at the origin transposed:

$$(2.4) \quad H = (P_{target} - \bar{P}_{target})(P_{centroid} - \bar{P}_{centroid})^T$$

The SVD factorization allows us to decompose H as:

$$(2.5) \quad H = U\Sigma V^T$$

where U and V are real orthogonal matrices representing the spatial rotation while Σ is a rectangular diagonal matrix representing the scaling. From the rotation components we can compute the rotation matrix R as:

$$(2.6) \quad R = VU^T$$

Additionally this solution provides an alternative translation t taking into account the rotation previously solved. Using the centroids from the previous datasets we can solve:

$$(2.7) \quad RP_{target} + t = P_{centroid}$$

$$(2.8) \quad R\bar{P}_{target} + t = \bar{P}_{centroid}$$

$$(2.9) \quad t = \bar{P}_{centroid} - R\bar{P}_{target}$$

In an offline test with the centroid data from the observations, we run the Kabsch algorithm and we verify that there is a clear field rotation trend on some of the guide sequences. The 2D guide signal (translation) is also different when the rotation is taken into account, but the effect is not dominant. An example of a sequence with a strong and clear field rotation can be seen in Figure 2.16.

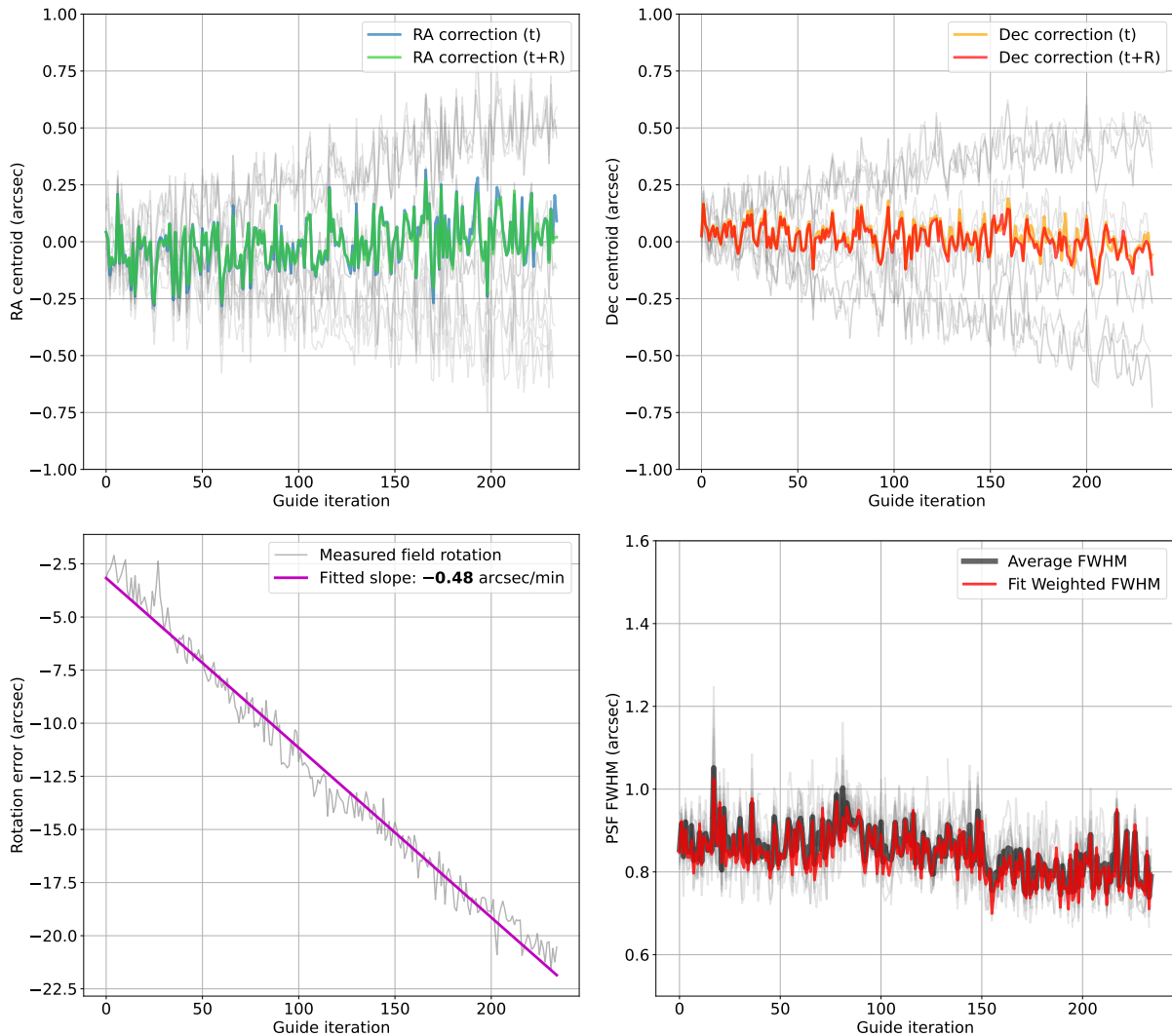


Figure 2.16: Offline analysis of sequence 43860 with the Kabsch algorithm to measure the field rotation throughout the guiding sequence. In the two upper plots we have the correction centroids for Right Ascension and Declination. The grey lines correspond to the individual guide star centroids, that diverge one from each other as the exposure goes on due to field rotation. In each coordinate axis we plot the translation correction only (t) as computed by the guider at the Mayall and the offline computation taking into account Rotation and Translation (t+R). The difference becomes more evident once the rotation error is higher at the end of the exposure. The lower plots show the measured rotation and its fitted slope, estimating a field rotation rate of -0.48 arcsec/min and the measured FWHM which seems stable throughout the exposure.

However the field rotation rate significantly varies at different observations. As expected due to the telescope polar misalignment, it has a relation with Declination and Hour Angle (HA). We studied the option to implement the Kabsch algorithm into the online guider code to estimate and correct in real time the field rotation. However due to the noisy field rotation measurements of the individual guide iterations, the slope could only be determined reliably after a significant amount of iterations. Furthermore the field rotation rate could be modelled and adjusted at the hexapod prior to each exposure. Therefore we had to properly estimate the errors of the rotation rate slopes to accurately create a model that reproduces the rate depending on the Declination and HA of the exposure. To estimate the slope of the measured rotation as a function of time to derive the rotation rate, we first seek to fit the n data points (x_i, y_i) to the linear model given here:

$$(2.10) \quad \hat{y} = \hat{m}x + \hat{b}$$

Then we compute the following sums of squares where:

$$(2.11) \quad SS_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$(2.12) \quad SS_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

The Least Squares Estimator of the Slope is \hat{m} such as:

$$(2.13) \quad \hat{m} = \frac{SS_{xy}}{SS_{xx}}$$

and the Least Squares Estimator of the Intercept is \hat{b} where:

$$(2.14) \quad \hat{b} = \bar{y} - \hat{m}\bar{x}$$

Taking into account that there are n data points and $p = 2$ regression parameters, leaving $n - 2$ degrees of freedom therefore the standard deviation of $y(x)$ is $s_{y,x}$ where:

$$(2.15) \quad s_{y,x}^2 = \left(\frac{1}{n-2} \right) \sum_{i=1}^n (y_i - \hat{y})^2$$

and finally, the standard deviation of the slope \hat{m} is s_m where:

$$(2.16) \quad s_m^2 = \frac{s_{y,x}^2}{SS_{xx}}$$

Analyzing 2300 guiding sequences and rejecting those with a high slope error, we produce the correction model that is shown in Figure 2.17. The modelled field rotation residual was added to the hexapod rotation rate to compensate for this effect and the average value was significantly reduced and centered at 0, as seen in Figure 2.18. The model is still noisy and incomplete for

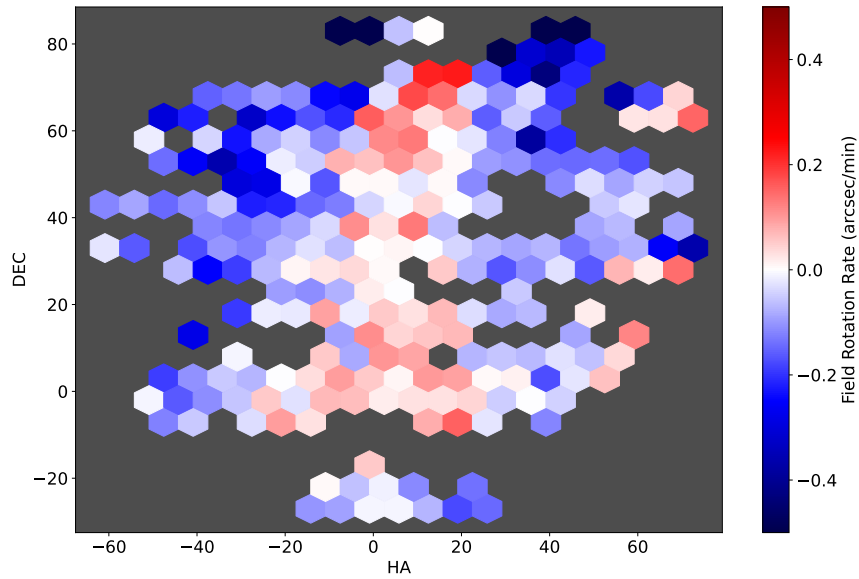


Figure 2.17: A model of the field rotation rate residual measured by the guider as a function of telescope Hour Angle and Declination. This residual is due to telescope polar misalignment and is expected to be corrected by DESI's hexapod.

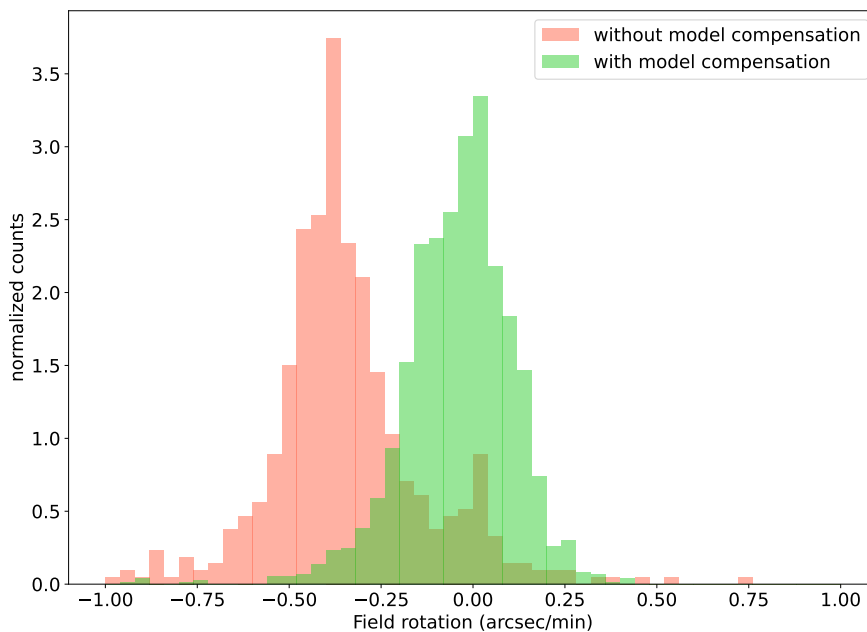


Figure 2.18: The distribution of the measured field rotation at the guide exposures without any correction (prior to March 2020) in red and with the compensation added to the hexapod rotation from the sky model produced from the Kabsch algorithm (after January 2021) in green.

some sky areas and further observations at more extreme HA and Dec are expected to improve its completeness.

DESI

Finally, with the complete instrument mounted at the Mayall's prime focus, we had to verify that the challenging requirements on tracking and positioning were met. Those related to the guider performance are summarized into the following requirements:

- **Key Performance Parameters E2E function:** *For zenith angles $\leq 60^\circ$, Mayall guiding shall be demonstrated to be 200 mas RMS or better.*

From the output guide signal we can compute the Guide 1D & 2D RMS. This is computed as the RMS of the output guide corrections in RA and Dec and combined. This metric involves the overall system performance, from the ability of the guider to measure and track the guide stars to the performance of the telescope in reacting to the corrections sent to the TCS. With this metric we can validate the main guiding requirement in DESI established by the KPP E2E function of 200 mas RMS. Measuring the average RMS values of the last 500 exposures during commissioning, we measure an average value of 129 mas, and 90% of the time below 200 mas. A representative histogram of the Guide RMS can be found in figure 2.19. With an equatorial mount, a better performance on Declination is expected with just 76 mas RMS. The data used in this metric is still commissioning data and further improvements are planned that could upgrade the quality of the observations, reducing even more the RMS measured in this test.

- **Focal Plane Array Req. 2001:** *The FPS shall deliver a tracking error signal better than 0.03 arcseconds.*

This requirement points solely to the ability of the guider to measure the true tracking error of the telescope. This was tested in simulations and the RMS performance of the centroiding algorithm used (PhotUtils (Bradley et al., 2020)) easily reached the accuracy needed. However real observations have more complex systematics that may degrade the performance from the simulated test. A realistic metric we accounted to validate this requirement is to compute the standard deviation of the individual guide star centroids. This removes the possible constant offsets between guide star centroids and provides an independent measurement of pure centroiding accuracy and consistency. Performing the centroid STD test over the same subset as the previous metric, with the last 500 exposure sequences from the commissioning, results into a standard deviation of just 18 mas, once again with 90% of exposures below the 30 mas limit. An illustration of the distribution of centroid accuracy can be found in Figure 2.20.

The previous tests validate the readiness of the guiding system to operate for the DESI survey, ensuring that the tracking is stable enough to maintain the 5000 sources aligned with all robotic

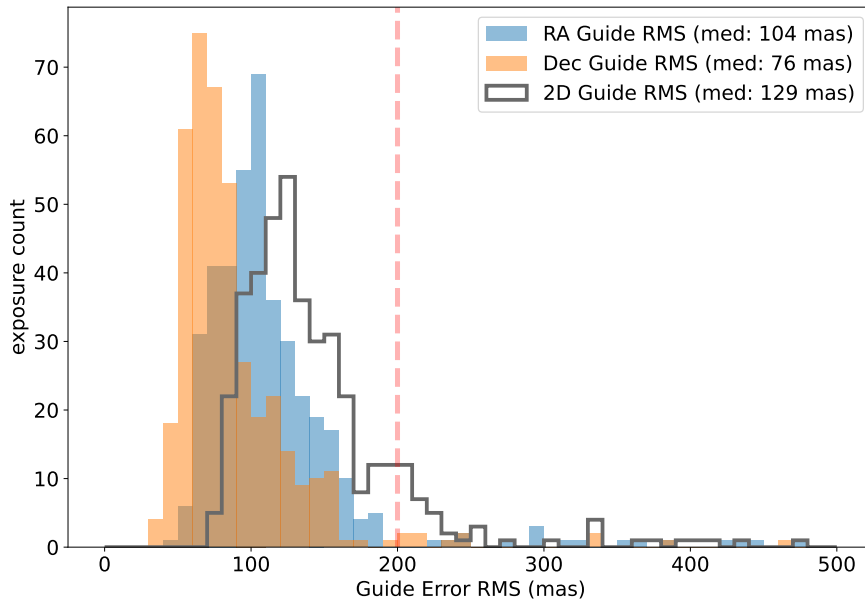


Figure 2.19: The distribution of the measured Guide signal RMS for the two axes and the combined one. The average value reaches the requirement limit specified by the DESI survey. The sample set shown in this sample corresponds to the last 500 exposures during commissioning in early 2021.

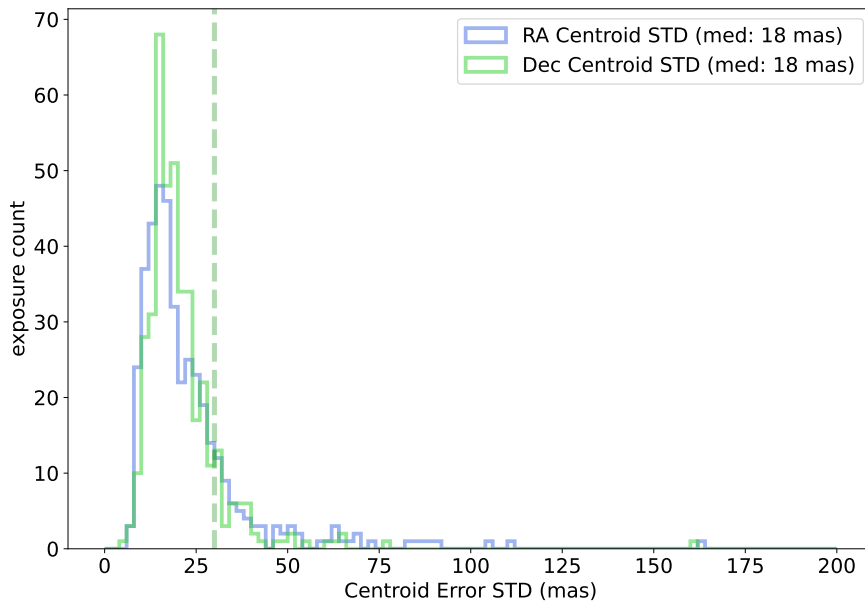


Figure 2.20: The distribution of the guide star centroids standard deviation, a metric that validates the accuracy of the tracking error measurements. The average value reaches the requirement limit specified by the DESI survey. The sample set shown in this sample corresponds to the last 500 exposures during commissioning in early 2021.

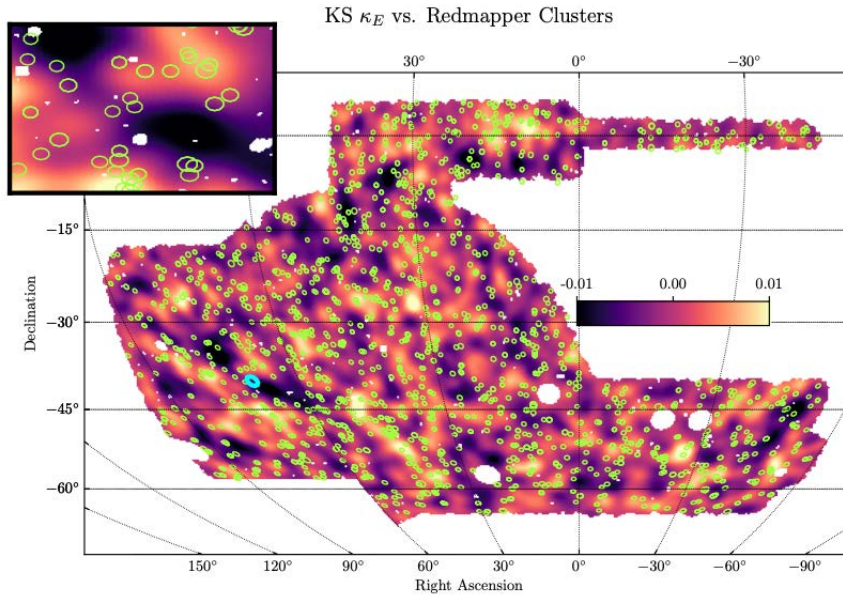


Figure 2.21: Mass map from DES Year 3 derived from the weak lensing measurements with clusters in the redshift range $0.3 < z < 0.5$ (green circles) superimposed.

fibers throughout the entire exposure.

2.4 Enabled scientific results

The accurate tracking achieved with the auto-guiding systems previously described, combined with the rest of the instrument systems being optimized and an accurate data reduction enabled outstanding scientific results in multiple areas. DESI barely started their scientific observations and therefore no major discoveries are yet being found. However the Dark Energy Survey completed their observations in 2019, providing since the first year of data remarkable cosmological results. It has provided the largest galaxy redshift catalog to date (Abbott et al., 2021), which improved our knowledge of the current cosmological parameters, from 700 million sources. Additionally, the exceptional image quality allowed to estimate weak lensing measurements over 1500 deg^2 which resulted into the largest and most detailed dark-matter map from our Universe (Gatti et al., 2021), as can be seen in Figure 2.21.

The PAU Survey results will be highlighted the Chapter 3 where the whole project and data management is described. However PAUCam was offered to the community to perform their own observations at the WHT, and it offered also noticeable non-cosmological scientific results such as in Martínez-Lombilla et al. (2019), enabling deep images as the one shown in Figure 2.22.

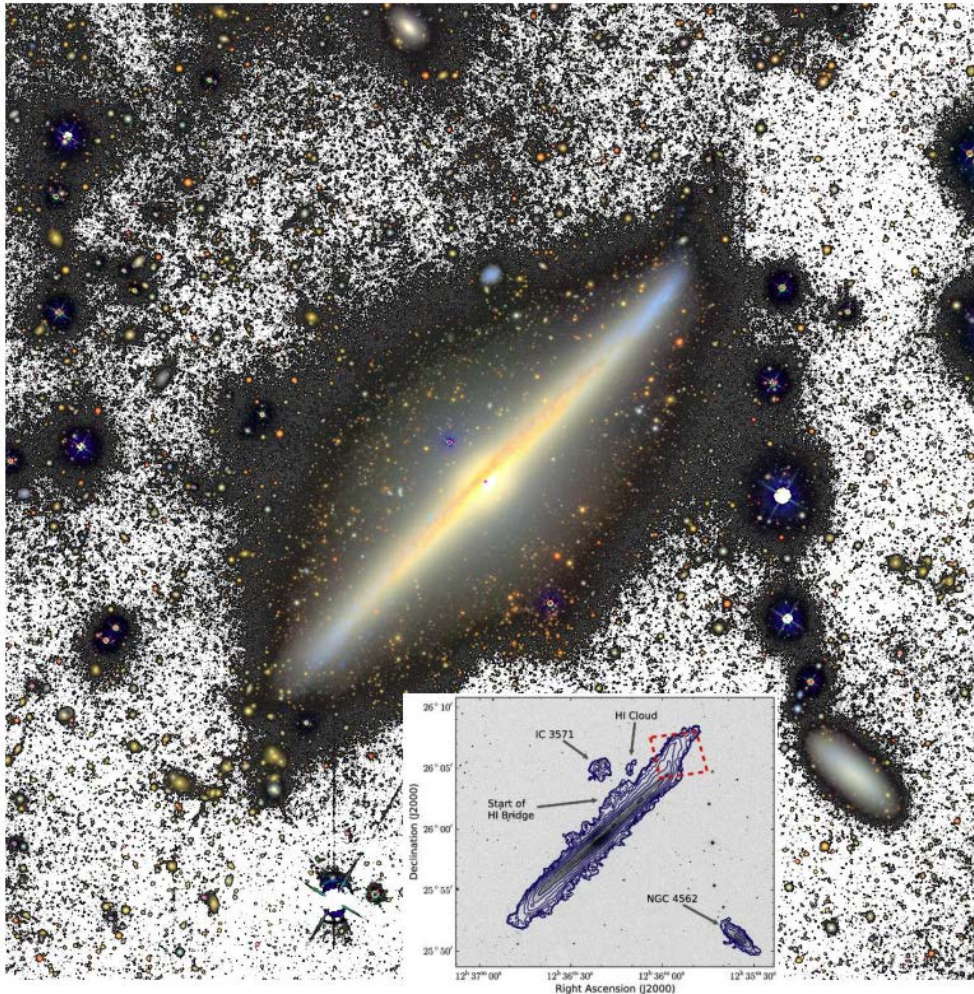


Figure 2.22: Deep observations of NGC 4565 with PAUCam in filters g , r and i in a combined 4h/filter coadd image reaching a depth of $30\text{mag}/\text{arcsec}^2$.

2.5 Summary

In this chapter we have presented three auto-guider systems based on a multi-star ROI concept, in three different implementations for the DES, PAU and DESI surveys. We explained the specific pixel simulations we designed for this purpose, which resulted key in the development and validation of all three guiders. The tracking requirements were always reached, even in the most demanding scenario for the DESI spectrograph, reaching tracking error below 150 mas RMS. The guider data also allowed us to optimize the subtle field rotation present in the Mayall Telescope. With similar techniques, we could implement a Focusing and Tip-Tilt algorithms that supported the commissioning and observations of PAUCam.

The accurate tracking resulting from the implemented systems enabled key scientific results.

DATA MINING A COSMIC SURVEY

In this chapter we describe all the complex data analysis on a large scale cosmology Ground-Based survey, particularly to the Physics of the Accelerating Survey. From its early pixel simulations, to calibration of raw images all the way up to science ready catalogues. We also describe efficient observing tactics in the telescope and big-data operations at the computing center.

3.1 The PAU Survey and Data Management

The Physics of the Accelerating Universe Survey (PAU) is an ambitious international survey aiming to better understand the yet unknown physics behind the evidence of an accelerating expansion of the universe. The approach in which PAU seeks to discover further the origin of the mysterious dark energy and dark matter is by measuring better than anyone the most difficult metric in cosmology: the radial distance of the galaxies. The project presents an innovative technique with 40 narrow band filters, allowing to obtain a low resolution spectra of all objects in the sky where its cosmological redshift can be inferred with unprecedented accuracy. This method lies between the regular photometric redshifts obtained from just 5 or 6 broad band filters (such as DES, SDSS,...) and the high resolution spectroscopy surveys (BOSS, DESI,...). The first ones have the advantage of sampling faster the sky but its accuracy is at least one order of magnitude worse than PAU. On the other side, the spectroscopic surveys have a very high redshift precision but with the disadvantage of being slower and the need of pre-selecting the sources at which the spectroscopy is made. Therefore PAU sits on a unique spot where no selection is made and high precision redshifts are obtained. This is at the expense of observing with many more filters the sky, which limits the area that can be surveyed.

The goal of the survey is to obtain photometric redshifts for galaxies down to $i_{AB} \sim 22.5$ in the redshift interval $0.1 < z < 0.9$, with a precision of $\sigma_z < 0.003(1+z)$ over an area of $\sim 100 \text{ deg}^2$.

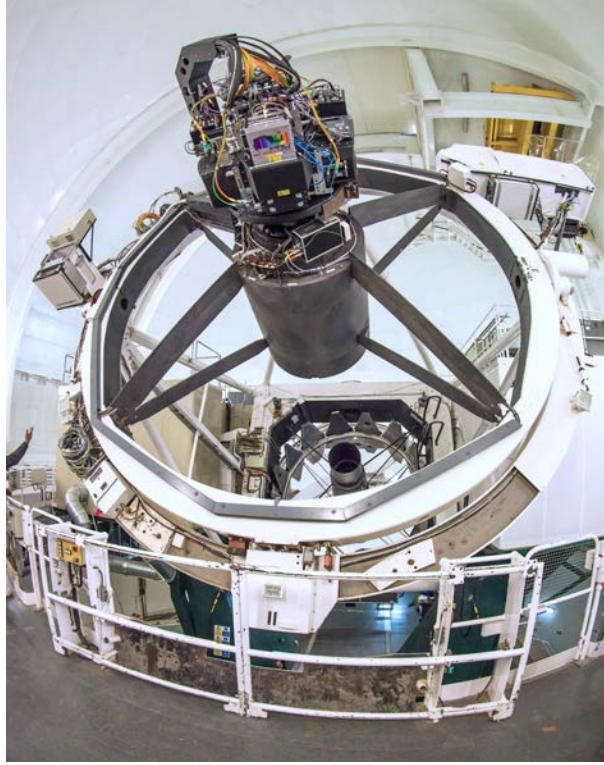


Figure 3.1: The PAU Camera mounted at the William Herschel Telescope.

In addition, PAU will yield a low resolution spectra for millions of stars and galaxies, leaving a unique legacy value enabling a wide range of astrophysical studies aside from the cosmological ones.

As explained in Chapter 2, a new large field of view instrument had to be build, named PAUCam. This camera has been installed at the majestic William Herschel Telescope with its massive 4.2m diameter mirror, as seem in Figure 3.1. With the current WHT Prime Focus corrector, the camera covers ~ 1 degree diameter Field of View (FoV), of which, only the inner ~ 40 arcmin diameter are unvignetted. The focal plane consists of a mosaic of 18 $2k \times 4k$ Hamamatsu fully depleted CCDs, with high quantum efficiency up to $1\mu\text{m}$ in wavelength. To maximize the detector coverage within the FoV, filters are placed in front of the CCDs inside the camera cryostat (made out of carbon fiber) using a challenging movable tray system. The camera uses a set of 40 narrow band filters ranging from ~ 4500 to ~ 8500 Angstroms complemented with six standard broad-band filters, ugrizY.

Carrying out such survey implies large volumes of data and complex analysis. The camera produces ~ 350 GB of raw data per night of observation and the processed data will multiply its size to enable the further processing analysis. The unique band set and instrument configuration requires new algorithms specifically tailored for this dataset. In the end, the survey will map millions of stars and galaxies, with its 40 band measurements and many more parameters each,

with multiple observations at different times. Furthermore the constantly changing weather conditions and other unpredictable situations that happen in the observatory such as Calima from the Sahara's desert, frozen camera window due to condensation, closed petals at the primary mirror and many more, make the data processing of PAU a major challenge. Manual intervention or visual inspection of all the data is not anymore an option. Therefore we need to use advanced data mining techniques that ensure a reliable processing, ensuring that we take the most of the massive data set of PAU and that statistical properties of the survey are not biased.

To deal with such complex data set and build the necessary infrastructure and software that can handle the data mining of it, we created the PAU Data Management (PAUdm) team. Its responsibilities start from the transfer of data to the Port d'Informació Científica where all processing takes place, to the calibration, analysis and distribution of the final data products.

3.2 Simulations

In such a large and complex project, it is critical to ensure that the scientific goals can be reached prior to the construction of the camera and development of the survey. For this reason we built an end-to-end pixel simulation pipeline from the cosmological dark matter simulations to the particular instrumental effects of PAUCam. These simulations provided such valuable information that have also been used after the survey has started with real and more accurate models to provide answers on complex systematic effects.

3.2.1 Cosmology & Mock Galaxy catalogue with MICE

PAU aims to study the large scale structure of the universe. Therefore it is critical to have a realistic galaxy catalogue as input to our simulations that accurately reproduce the cosmology and physics that we are aiming to study in this survey. In particular we need to pin down the properties of the dark-energy that drives the observed accelerated expansion of the universe. For such a challenging task of reproducing the universe large scale structure with high precision, dark matter N-Body simulations have become a standard method to mock cosmological observations.

The Marenstrum Instituto de Ciencias del Espacio (MICE) Team ran a massive N-Body simulation (Fosalba et al., 2015a) at the Marenstrum Supercomputer in Barcelona, with more than 70 billion dark-matter particles in a box of $3h^{-1}$ Gpc aside. This simulation samples from the largest linear scales down to the highly non-linear regime of structure formation where gravity drives dark-matter. The N-Body simulation produces comoving outputs at different times as well as a lightcone where structures evolves with distance up to a redshift of 1.4. Following an innovative technique called *Onion Universe* (Fosalba et al., 2008), the MICE Team could derive accurate gravitational lensing parameters for each galaxy in the simulated lightcone, essential to validate the science goals of PAU. A sample of the N-Body simulation can be seen in 3.2.

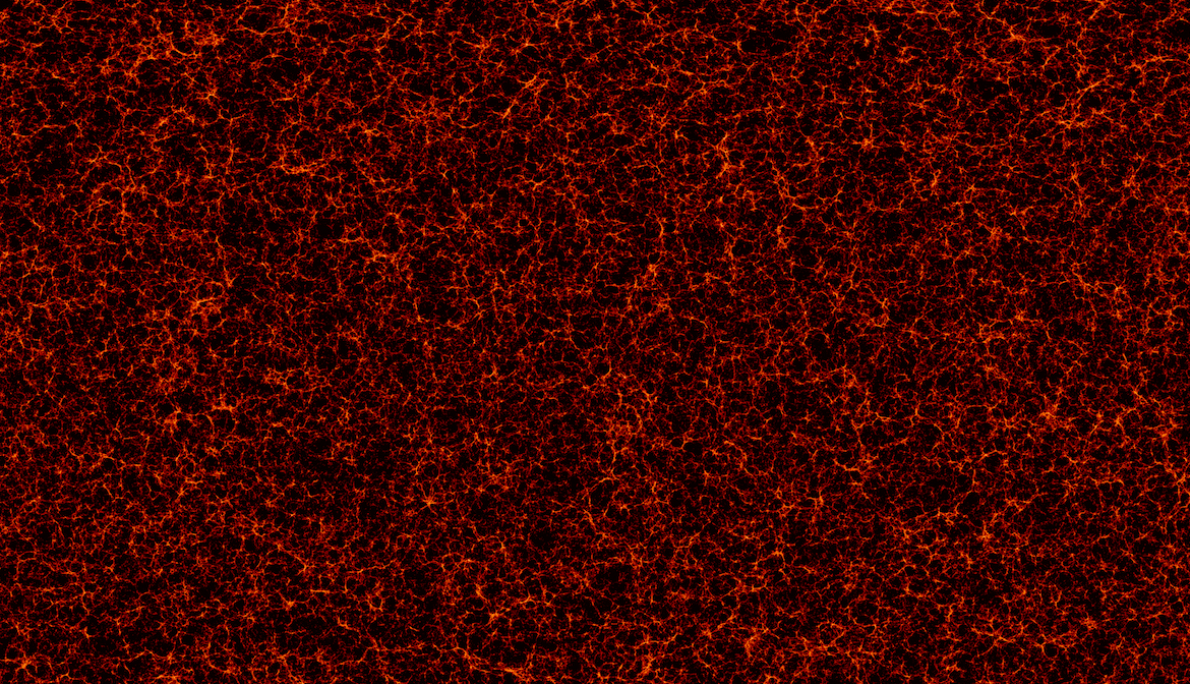


Figure 3.2: A snapshot of the MICE dark matter simulation. The image above illustrates the dark matter particles of a third of the sky in a single redshift shell. The structure and filaments of the dark matter particles can be seen and from these we build the halos and galaxy catalogues that enter into the PAU pixel simulation.

As an intermediate step towards the galaxy catalogues, we need to build the dark matter halo catalogues (Crocco et al., 2015) using a Friend-of-Friends algorithm both for the comoving outputs and lightcone. At a final step, the halos are populated with galaxies following an halo occupation distribution (HOD) (Carretero et al., 2015). The resulting mock galaxy catalogues reproduce the observed local galaxy luminosity function and the colour-magnitude distribution as observed by the Sloan Digital Sky Survey (SDSS). It also reproduces the observed galaxy clustering properties as a function of luminosity and colour which is key for the scientific analysis of the PAU Survey.

In order to render these galaxies into images, we need additional parameters that model the spectra and shape of each galaxy. We assign to each galaxy a SED template and an extinction curve such that with a reference magnitude and its redshift we can reconstruct the incident spectra. To compute the flux at each band, we convolve the spectra with the passbands in PAU, obtaining a flux that can be fed into the pixel simulator. Additionally, the catalogue includes its bulge and disk lengths, the position angle and its axis ratio allowing the pixel simulation pipeline to render its light profile accurately.

3.2.2 Stellar catalogue

Even when the PAU survey aims to study galaxies, stars are crucial throughout the whole data processing. Stars are the main source for photometric, astrometric and PSF calibrations. On the negative side, stars add confusion as they blend with the galaxies behind and complicate the photometry and separation of their flux. For these reasons amongst others, it is necessary to include a representative sample of stars in the image simulation pipeline. Real catalogues cannot be used as they do not provide complete information (specially for narrow bands) or are not faint enough for the depth of the survey (i.e. Gaia will only reach $g_{AB} < 21$ while PAU must reach stars at a fainter $i_{AB} < 22.5$ or more. Therefore we need to obtain our stellar sample from a Milky Way model, rather than from real observations. A convenient option was the Besançon model (Robin, Creze, 1986) as it provides a realistic sample of stars at any location in the sky, giving us the flexibility to test different stellar densities at the various PAU fields. To obtain the fluxes of the stars at each narrow band filter, we assign an SED template from the Pickles library (Pickles, 1998) based on the stellar properties defined by the Besançon model for each star. Once band fluxes are computed, the catalogue has all properties to be rendered in the image simulations.

3.2.3 Instrument, telescope and sky models

The models described in this section provide the information that allow us to transform the star and galaxy photons into counts at the pixel image, with all the effects that alter its true value and shape due to atmosphere, telescope or instrument features.

3.2.3.1 Pre-commissioning models

In this section we describe the models assumed by the pixel simulation prior to install the camera at the WHT. These were very useful simulations used to verify that the survey could meet its scientific requirements and provided a practical dataset to develop the data management tools for the data handling and calibration.

- **Sky Transmission:** This model represents the transmission of the atmosphere. It belongs to a model obtained at the Apache Point Observatory at airmass 1. The filter curve of this model can be seen Figure 3.3.
- **Sky Emission Lines:** Calibrated intensities of emission sky lines from the UVES Sky Emission Atlas (Hanuschik, 2003). This is the source of Airglow emission that contributes to the 65% of background in dark nights.
- **Zodiacal light spectrum:** Zodiacal light spectrum in $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1} \text{arcsec}^{-2}$ with a 25% contribution of background light in dark nights.
- **Solar light spectrum:** Spectrum of the sun to be used as moonlight spectrum on grey and bright nights.

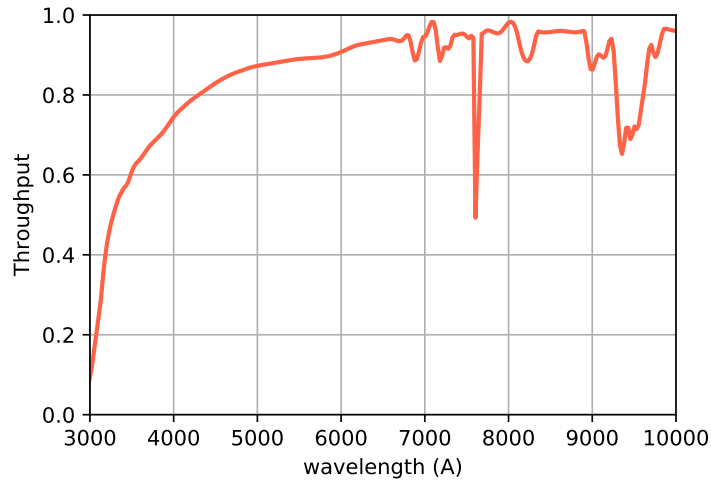


Figure 3.3: A model of the sky transmission used in the pixel simulation pipeline of PAU.

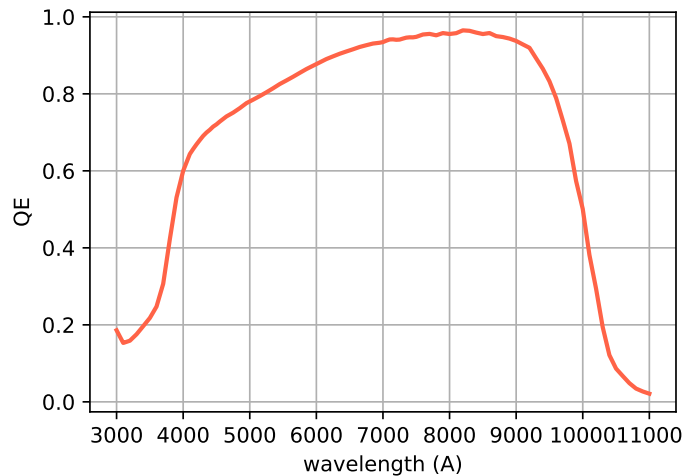


Figure 3.4: A model of the Hamamatsu CCD detectors quantum efficiency, measured in our optical laboratories.

- **Hamamatsu CCD Quantum Efficiency:** Quantum efficiency of the Hamamatsu CCD detectors. Measured in our optical lab in August 2013 and averaged across all 21 detectors received. The good infrared behaviour of these detectors can be seen in Figure 3.4.
- **Milky Way Extinction maps:** Even we initially started using the SFD dust reddening maps (Schlegel et al., 1998), we switched to the more resolution and accurate Planck Thermal dust maps (Planck Collaboration, 2014). This model is applied to all extragalactic sources simulated in the pipeline. The stellar sources from the Besançon simulation already contain a more complex 3D model extinction and therefore this map is not used for the

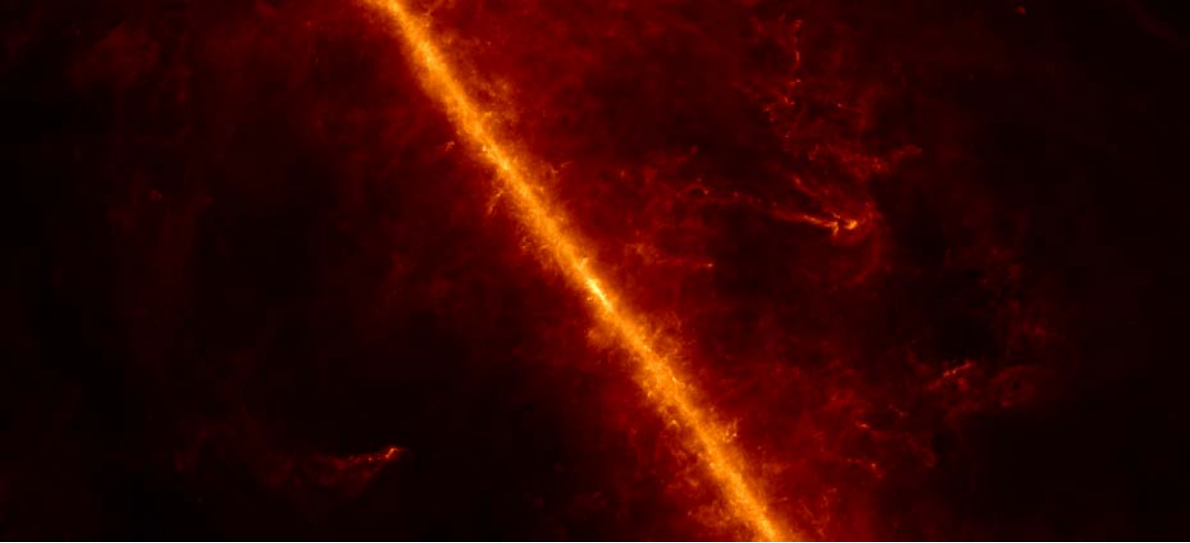


Figure 3.5: A close up of the thermal dust map in equatorial coordinates from the Planck collaboration released in 2014, pointing at the galactic center where the extinction is the highest.

stellar sources. A sample of the dust map is seen in Figure 3.5.

- **PAU Filter System:** The PAU filter set. This is composed by 6 broad band filters ugrizY and the 40 narrow band set. These were characterized at our optical laboratory at room temperature and at different integration angles from 2.9 to 10.1 degrees, representative of the incident angles during observation. Figure 3.6 show the 46 transmission set of PAU including atmosphere, telescope, filter and detector.
- **WHT Throughput:** The optical transmission of the William Herschel Telescope, including corrector lenses and mirror reflexivity. Evaluated by Tibor Agocs (ING) March 2010.
- **WHT Vignetting:** Transmission due to optical vignetting from the center of the focal plane towards the edge. The transmission at 20 arcminutes from the focal plane center is expected to be at 90% and quickly decay to 10% at 30 arcminutes, leaving effectively the 40 arcminute diameter for science observations. The effect of vignetting in the focal plane of PAU can be seen in Figure 3.7.
- **Focal Plane Layout:** The position and orientation of each detector in the focal plane, including gaps, readout directions and the detector name. This model also provides information of the filter tray layouts, specifying which filter stays on top of each detector for all 6 broad band filter trays and 6 narrow band filter trays.
- **General detector parameters:** For all Hamamatsu detectors, we measured a common full well capacity (of 210.000 e^-), an ADC saturation level at 18 bits, the plate scale of 0.263"/pixel, a pixel size of 15 μ m, a size of vertical and horizontal pre and over scan.

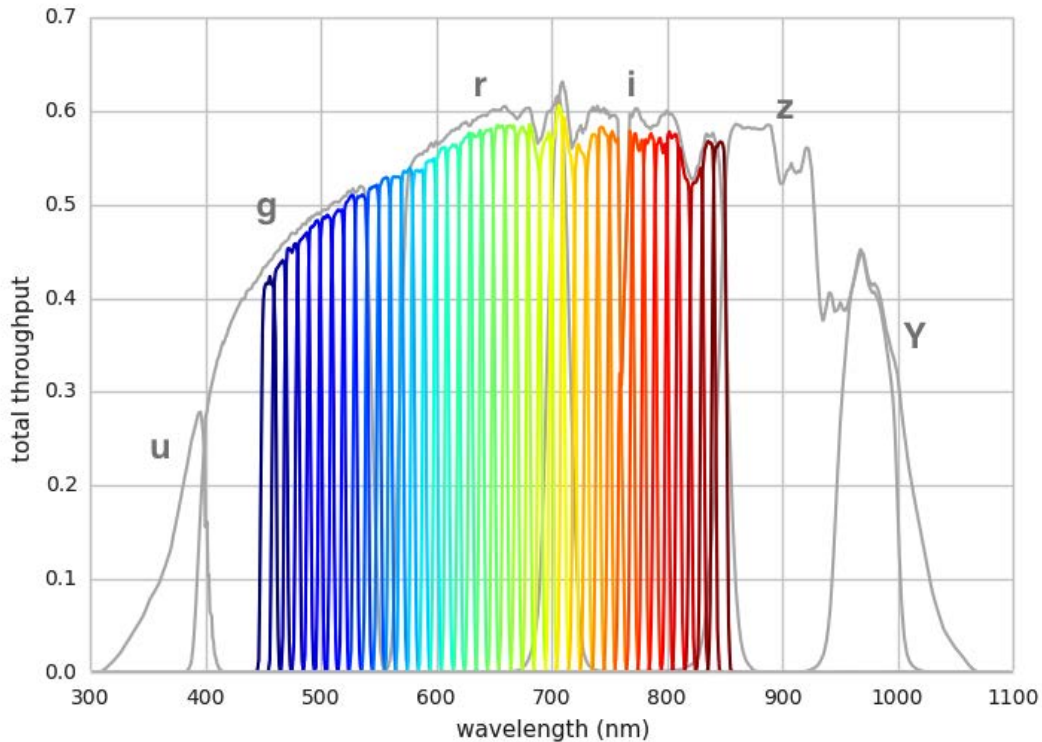


Figure 3.6: The entire PAU filter throughput set including all transmission effects in the chain: atmosphere, telescope, filter and detector.

- **Particular detector parameters:** For each science detector mounted in the focal plane array of PAU, we modelled the gain, bias and readnoise of each of the 4 amplifiers.

3.2.3.2 On-sky models

Soon after the camera started operating, we had large sets of data, with a level of accuracy not accessible before the commissioning of the instrument. There were also instrumental effects that we did not account for originally, such as the scattered light. Below we describe the models that were taken from on-site and on-sky images.

- **QE maps:** From the dome flat field calibration exposures, we can infer the Quantum Efficiency maps that define how pixel do no respond uniformly to light. These calibration exposures include the vignetting curve that results into a non-uniform illumination, not due to Quantum Efficiency, but to optical limitations. We can easily remove this component by subtracting a low frequency component from the image, leaving only the pixel-to-pixel small scale variations due to defects and variable response of the pixels. A sample QE model is shown in Figure 3.8.

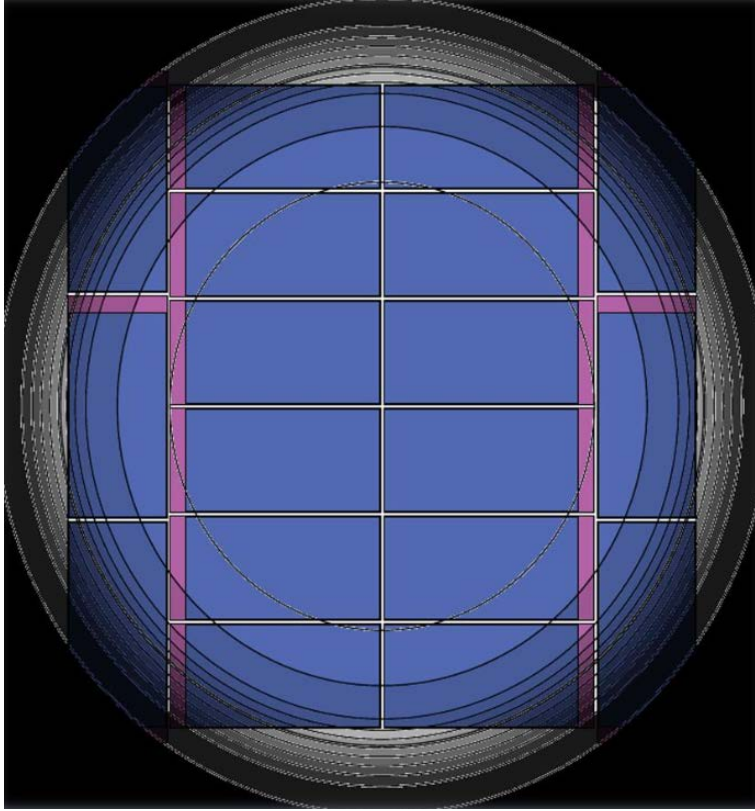


Figure 3.7: The vignetting at the WHT prime focus with the PAU focal mosaic at scale. The 8 central detectors have enough light to perform scientific studies with the narrow band filters while the outer detectors are left with broad band filters for calibrations.

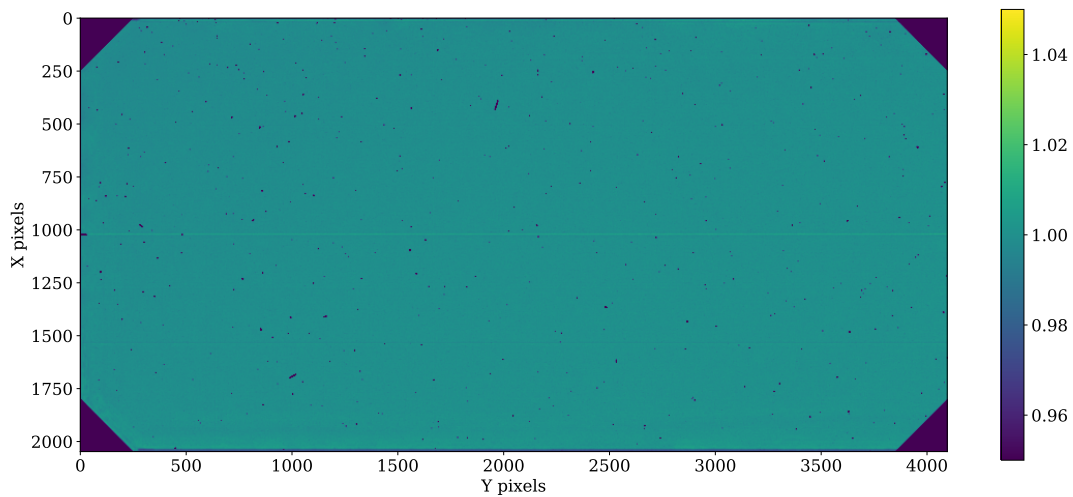


Figure 3.8: A quantum efficiency model to be used in simulations for a single detector-filter pair obtained from the processing of dome sky flats. This example corresponds to CCD 8 and filter NB695.

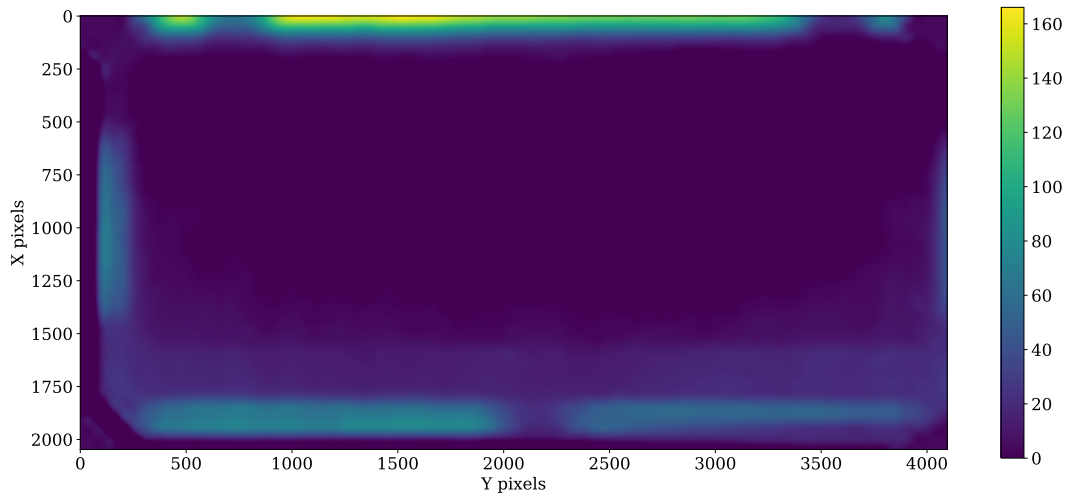


Figure 3.9: A model representing the scattered light measured from real observations. To generate this model, 100 sky exposures from different observations were combined in a median average, remaining only the constant background and illumination contribution.

- **Scattered light maps:** The scattered light in the PAU Camera has been one of the major complications during its processing. An effect caused by the reflections at the edges of the filter piece and the filter tray resulted into a non-uniform background at the edges of each detector image. Custom algorithms have been built to deal with this effect and simulations from this model were critical to develop such processing code. This model is built from sky flats, as a result of a combining multiple sky exposures from the same band. With a median average, we can filter out objects that vary from one exposure to the next, remaining only the constant features, such the background scattered light. A sample of the scattered light model used in simulations is shown in Figure 3.9
- **Illumination maps:** The illumination of the focal plane is not uniform due to optical distortions and vignetting, but neither the theoretical model nor the dome flats properly represent the real illumination pattern when observing the sky. To generate the illumination model, we need to compute again the sky flats and subtract the illumination curve from the background. This process has some complications due to the scattered light that sits in front of the illumination, but rather as an additive effect with respect to the multiplicative effect of the illumination. After some processing we can build a more accurate illumination model to use in simulations (and in the reduction of the images as well) shown in Figure 3.10.

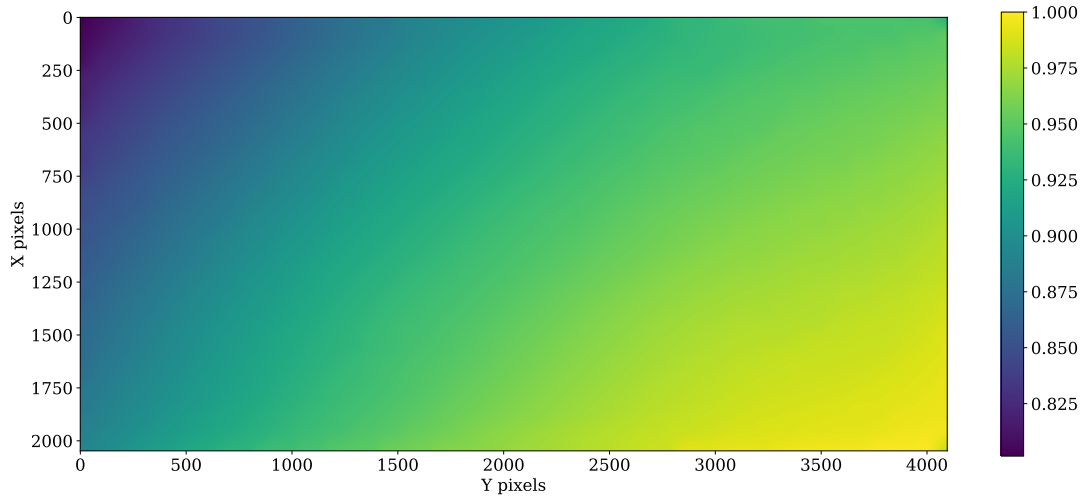


Figure 3.10: A model representing the true illumination at the focal plane of PAUCam. It better represents the optical response than the dome flat field and it is used in the pixel simulations and in the data reduction of the images to compensate for vignetting.

3.2.4 Survey Strategy Simulator

The observing time at the WHT is very precious and we need to make sure we make the most out of it. Furthermore PAUCam is a complex instrument and its focal plane layout with different filters for each detector complicate further the scheduling of targets. We built a tool *ObsSim* to simulate how PAU observes the sky, to optimize the telescope time and test different dither patterns.

3.2.4.1 ObsSim Planner

The Observation Simulator (ObsSim) planner is a tool for planning observations and defining the pointings of PAUCam required to cover an area in the sky. Typically we want to cover a rectangle in the sky of tenths of square degrees with the entire filter set of 40 narrow bands. If each filter would cover the entire focal plane, one could establish a dither pattern that shifts the pointing by several arcseconds to remove the detector gaps. However each of the 8-central detectors contain a different narrow band, which could be seen as 8 independent cameras shifted by few arcminutes, each observing with a different filter. The disadvantage of this configuration is that it is inefficient to observe small fields with all the 40 NB set as there is a constant boundary of 1 detector (in the long direction) in the Right Ascension axis and 3 detectors (in the short direction) in the Declination axis. This can be seen in Figure 3.11 and a relation of coverage efficiency vs total area for the case of having 1, 2, 3 and 4 fields in the sky.

After the definition of the field boundaries and the dither pattern, the ObsSim Planner returns the list of target pointings to be observed in all trays to fully cover the field. This was not the final

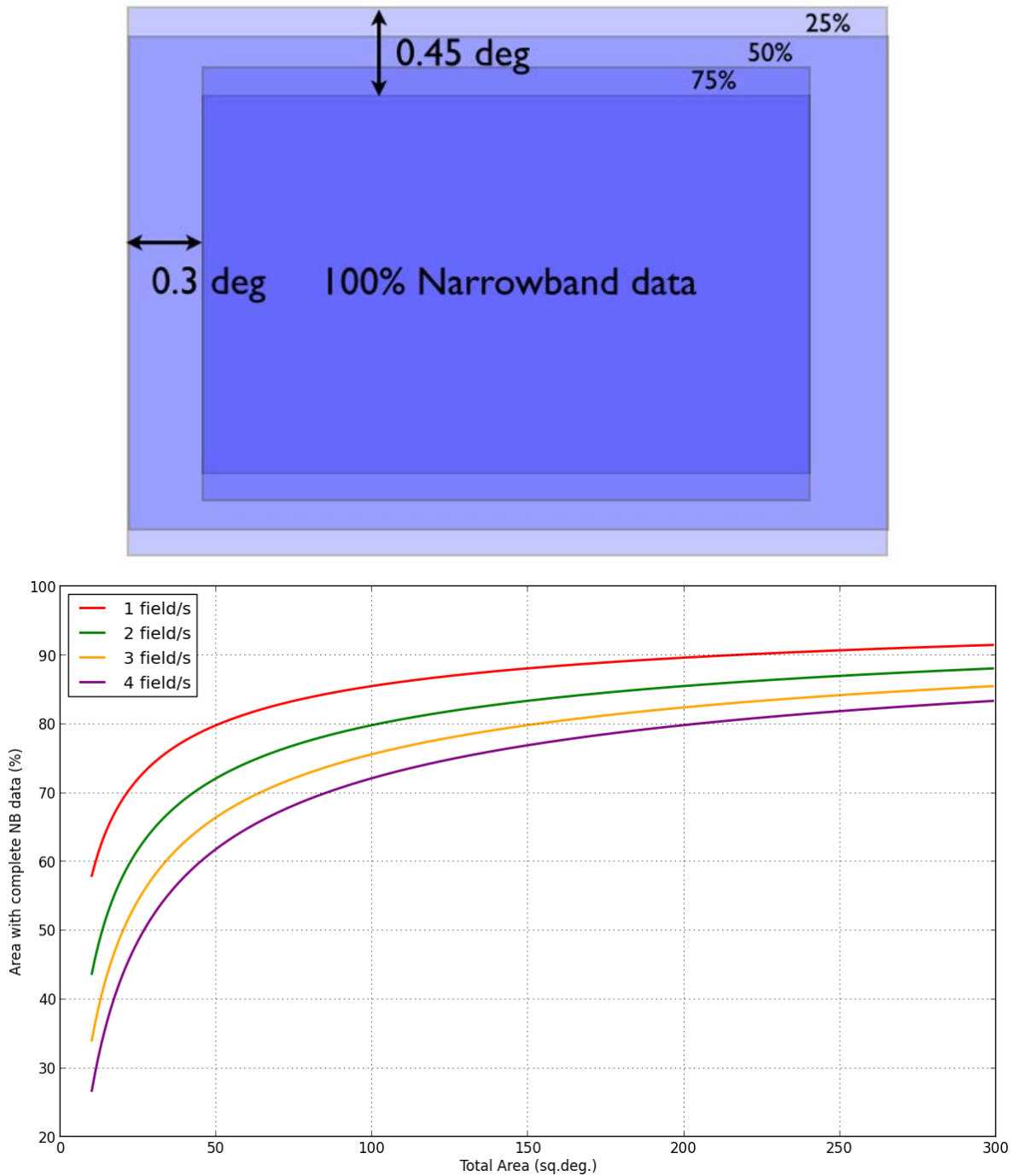


Figure 3.11: Top plot: A schema illustrating the percentage of narrow band coverage due to the particular disposition of filters in the focal plane. There is a constant margin of 0.3 degrees and 0.45 degrees in the RA and Dec directions respectively. This results in smaller fields being less efficient in terms of 100% of narrow band coverage with respect to total exposure time. Bottom plot: the relation between total survey area and the 100% narrow band coverage efficiency for different scenarios with 1, 2, 3 and 4 fields.

tool to define the Survey Strategy of PAU, but resulted useful prior to the commissioning to define a realistic scenario for the pixel simulation pipeline. The more accurate study and definition of the final survey strategy is described in §3.3.

3.2.4.2 ObsSim Scheduler

Once the targets were defined by the ObsSim planner for one or more fields, we had to schedule specific targets during a mock night of observation. For this purpose, we built a tool that reproduces the sky conditions and decides which observations suits best at each time. In the first place, the tool analyzed which fields remain visible at the telescope location and between astronomical twilights for that particular night. With the available fields visible at the beginning of the night, the scheduler selects the field that stays at a higher altitude at the current time to minimize airmass. Then it analyzes the background taking into account all contributions to it such as airglow, zodiacal light and moonlight. If the current field is too close to the moon, the background levels will rise and the field will be discarded automatically by the algorithm and switched to another at lower airmass if available. If the background levels are within the specified limits, the Scheduler will analyze which filter tray has its minimum distance from its dark night background conditions. Once the tray with best conditions is set, it will begin scheduling targets at specific times in the night. The background of each tray is only analyzed once per night to minimize the number of filter changes as the process of moving trays takes time and performing this operation too many times could deteriorate the mechanical parts of the camera. Once there are no more observations to perform for the current tray in the current field, other filter trays analyzed. If the field goes beyond the visibility limit (typically set at 2 airmasses), the scheduler will try to search for new fields to observe.

Similarly to the ObsSim Planner, this algorithm did not schedule the real observations at the WHT, as some observing strategy decisions were left to the astronomer. Furthermore this tool did not integrate possible issues in the telescope such as rotator limit, severe weather conditions, etc. However this system provided a quite realistic sequence of observation and resulted very useful to analyze the efficiency of the observing. The schedules was able to study a real period of nights and define the percentage of time where observations could be done, because there were fields visible and with dark enough background conditions. With this metric of night efficiency we could study different field configurations and establish at which periods of the year we could observe with the highest efficiency so no WHT time was lost. Figure 3.12 show an example study of a single field observation in the period of 31 days. The field was visible through all the month but the moon rise and got too close to the field, exceeding the background levels and preventing the observation during those days.

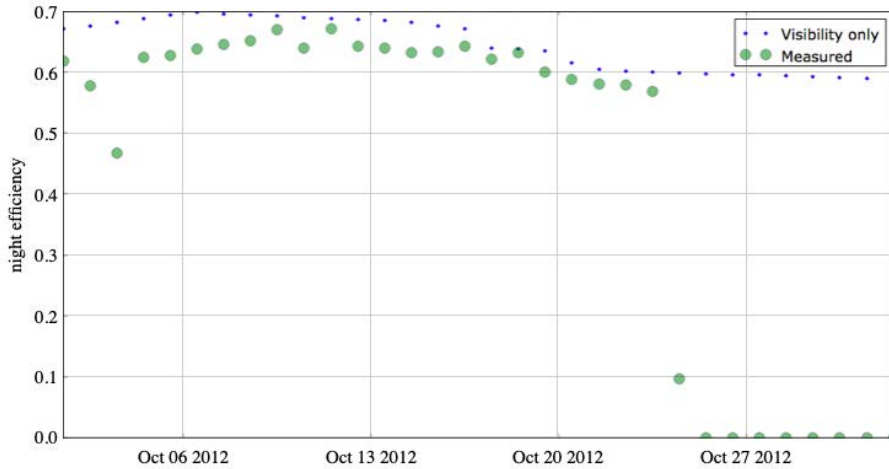


Figure 3.12: A sample run of the ObsSim scheduler during the period of a month. The single field in this study was visible all the time but due to an accurate background simulation we could verify that was not suited for observation.

3.2.5 Pixel Simulations

This section describes the final step of the chain, where all ingredients described so far are put together to build the pixel simulations. We differentiate two scenarios that will be described below: the Image simulation with all simulated data and the Image Emulation, where we try to replicate real observations.

3.2.5.1 Image simulation

At this stage we have the mock galaxy catalog with the cosmology defined by MICE, a realistic stellar catalogue as a function of galactic latitude and longitude, all the instrumental, telescope and atmosphere models to reproduce how PAUCam sees the objects and a survey plan that mimics the observation patterns. The only missing piece is a tool that with all the models and parameters from the catalogue, generates the pixel images. Our core rendering tool is SkyMaker (Bertin, 2009), a software from the Astromatic package that produce pixel images of ground-based telescopes.

In first place we build the catalogues for SkyMaker in pixel positions. Unfortunately the World Coordinate System (WCS) is not yet defined (they will be defined in this simulation) so we need to make use of the telescope pointing, the plate scale and mechanical positions of each detector. The fluxes at each band were computed in the generation of the catalogues, taking into account the total system transmission. At this stage we just need to pick up the corresponding band of the filter we are simulating and set it in the format of the SkyMaker catalogues. Then we configure SkyMaker with all the parameters it can simulate such as exposure time, seeing FWHM, background magnitude, magnitude zeropoint, gain, readnoise, full well capacity, M1 and

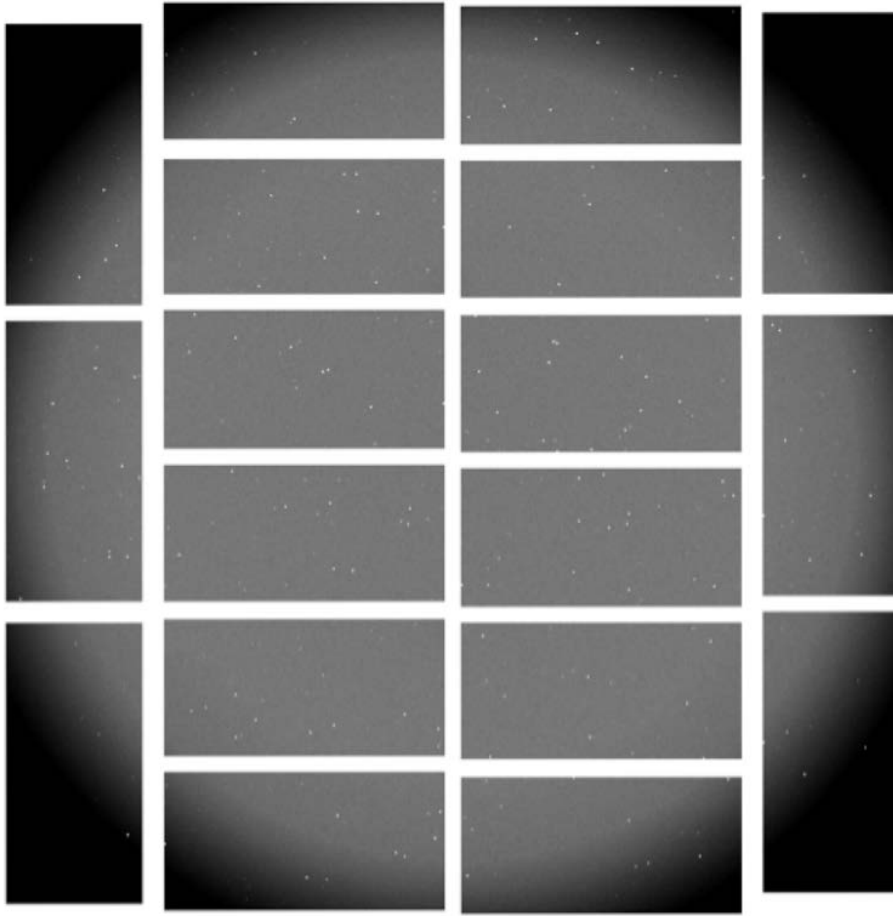


Figure 3.13: A simulated exposure of PAUCam with the pixel simulation pipeline. This simulation includes a cosmological galaxy mock, a realistic stellar catalogue, modelled background and environmental conditions and all known models from the atmosphere, telescope and instrument.

M2 mirror diameter, number of telescope spider arms, etc. SkyMaker returns then an image with most of the effects and a realistic PSF from the mirror optics. Effects like pixel blooming are also computed from the value of the detector full well. Optical vignetting and a quantum efficiency map were not available in SkyMaker we branched the code to include these missing effects. As a post-processing, we divide the image into 4 sections and apply the corresponding gain. All the metadata from the simulation and models is included in the FITS header in the same format as the camera will deliver its raw images. The WCS for each of the 72 amplifiers is computed and added as well to the FITS header. A sample PAUCam mosaic image simulated with this pipeline is shown in Figure 3.13.

Additionally, the simulator produces calibration frames such as bias and flat field images, with the same models used to produce the scientific sky images. Typically for an observation set, 10 bias images and 5 flat field images with different noise realizations are produced.

These simulations have been crucial to develop a first data reduction pipeline before the commissioning of the instrument. Thanks to this, as soon as PAUCam started to produce real sky images, we could analyze the data and return a rapid feedback to solve issues found in the images. This was also very valuable to prepare all infrastructure such as archive, transfer system, data base, etc. as these simulated sets of whole nights had the exact characteristics and format as the raw images from the instrument. Furthermore we develop scientific challenges on these simulated frames, where astrometric and photometric performances were evaluated and optimized.

3.2.5.2 Image emulation

Real sky data was more complex and had additional instrumental effects that the pixel simulation pipeline did not include, and therefore significant work has been done to improve the quality of the PAU data since the first light of the instrument. Part of the development for those new features found in the observed data that was not available in the simulations could be done with the actual real data and with the information from calibration sources or reference catalogues. However for other situations, it was necessary to replicate those new effects and reconstruct the real data in simulations. We call this process image emulation and has several fundamental differences with respect to the pure Image Simulation. On one side, we cannot use mock stars or galaxies as these do not match the real sources and one-to-one comparison tests cannot be done. Fortunately we have reference surveys that are deeper than PAU where we can obtain sources for certain areas in the sky. The Canadian French Hawaii Lensing Survey (CFHTLenS) which overlaps with most PAU fields, have a magnitude limit 2 magnitude deeper which is sufficient for the simulation purpose of our analysis. Additionally, we make use of real models to include more complex effects where an analytical model is not available or it is too complex to represent with accuracy.

In this process we build a stellar catalogue that is a combination of Gaia for the bright end and CFHTLenS on the fainter end. We could not obtain the stellar sample from a single catalogue as Gaia reaches g_{AB} of 21 and CFHTLenS does not contain the brighter stars. Therefore we establish the transition point at magnitude 17, which ends in a smooth distribution as seen in Figure 3.14.

Due to its wide spread use in the astronomy and cosmology community and continuous development for new features, we switched from SkyMaker (used in the main pixel simulation pipeline) to Galsim as our main rendering core. Our main programming language is Python and Galsim also integrates better to any script or notebook enabling more complex and custom configurations.

To render the stars, we use a simple Moffat profile such as:

$$(3.1) \quad I(R) \sim (1 + (r/r_0)^2)^{-\beta}$$

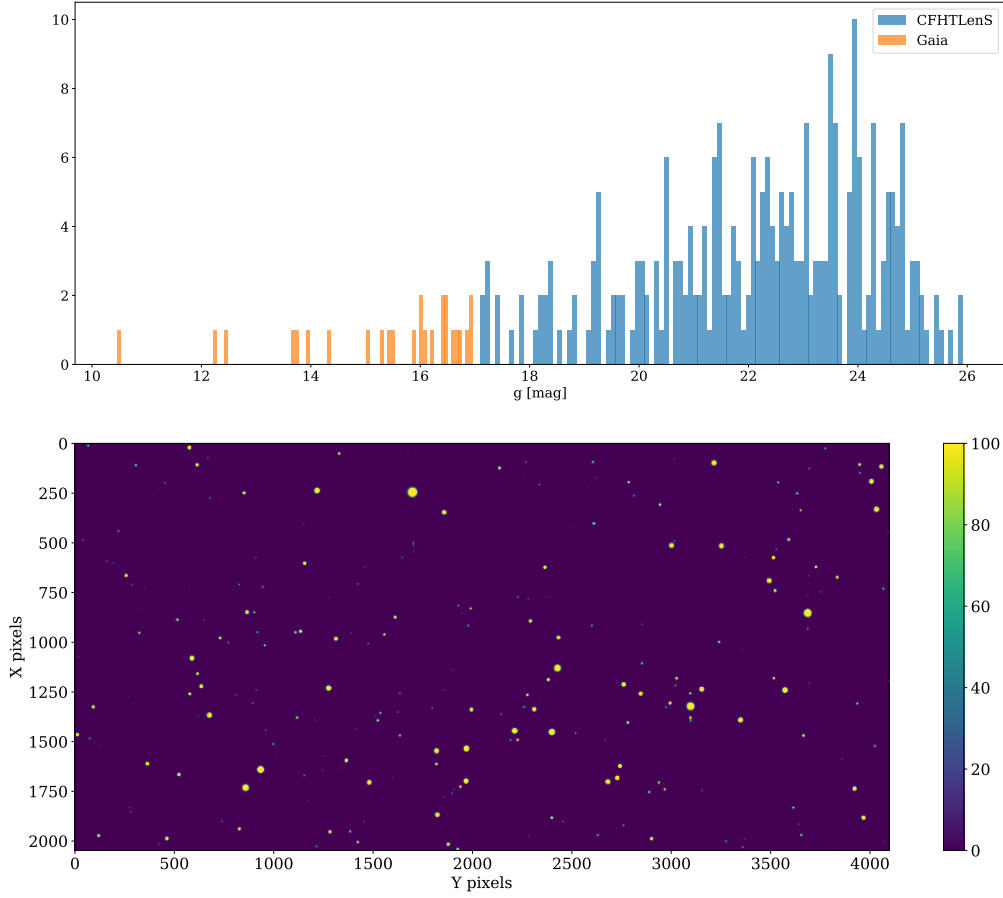


Figure 3.14: Emulation of real stars from Gaia and CFHTLenS. A combination of both catalogues is required to cover the entire magnitude range. The distribution above shows the selected stars for a single PAU detector. The image below contains the stellar profiles rendered in an empty image.

with a beta parameter of 4.75 and the measured PSF FWHM from the image we want to emulate. As the model is defined as a function of the scale radius r_0 , we convert from the measured FWHM such as:

$$(3.2) \quad r_0 = \frac{\text{FWHM}}{2\sqrt{2^{(1/\beta)} - 1}}$$

To render galaxies we convolve the same Moffat profile as for the stars with a single Sérsic profile as:

$$(3.3) \quad I(r) \sim e^{-(r/r_0)^{1/n}}$$

where r_0 is the scale radius and its is directly provided by the CFHTLenS catalogue, in a parameter were they computed the PSF deconvolved galaxy size. The Sérsic index n of the profile

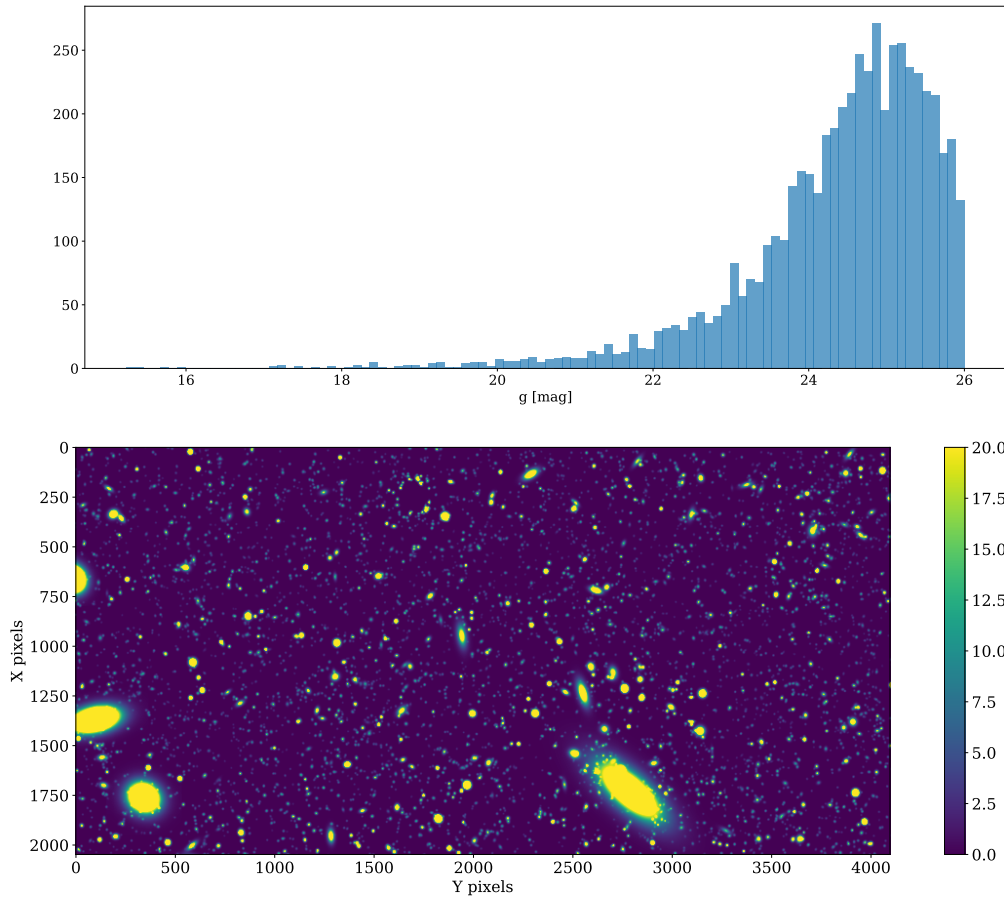
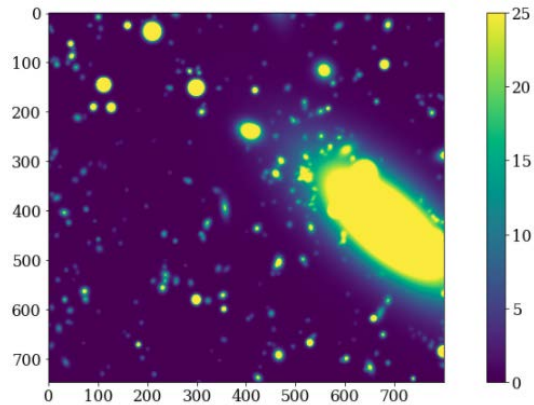


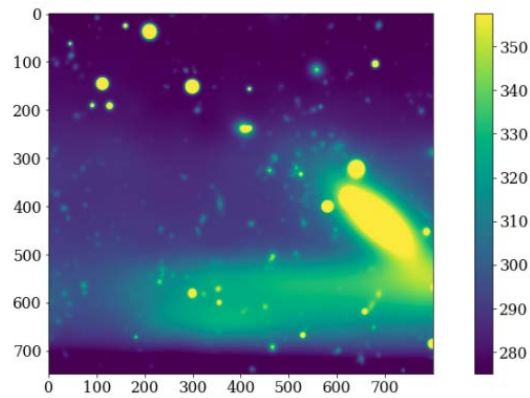
Figure 3.15: Emulation of real galaxies from CFHTLenS. The distribution above shows the selected galaxies for a single PAU detector. The image below contains the galaxy profiles convolved with a simple PSF rendered in an empty image.

is estimated from the Bulge-to-total ratio in a lookup table, derived double Sérsic profile fits internal to CFHTLenS. With this information we can render the galaxies in the detector, as shown in Figure 3.15.

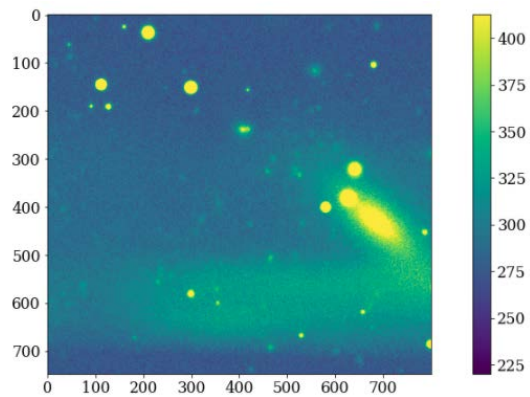
At this stage we can include the instrumental and atmospheric effects as measured from the observations. First of all we combine both star and galaxy noiseless images together:



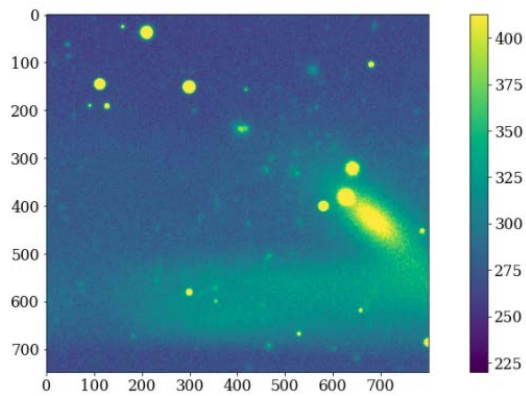
Then we include the additive components of the background and scattered light. The background is a flat offset added to all pixels and corresponds to the median background measured in the reference image. The scattered light model is obtained with sky flats as explained in §3.2.3.2. Both are additive effects and therefore just raise the level of the simulated sources:



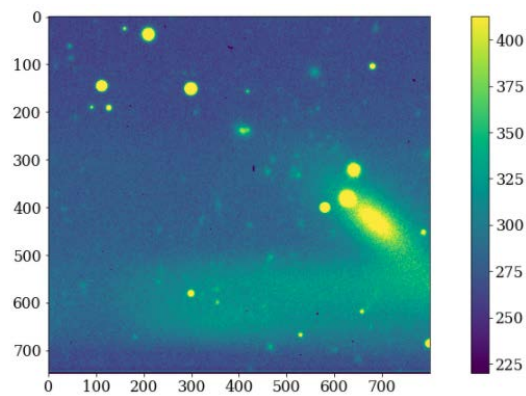
Now that all source emitting components are included in the image, we can apply the shot noise drawing samples from a Poisson distribution:



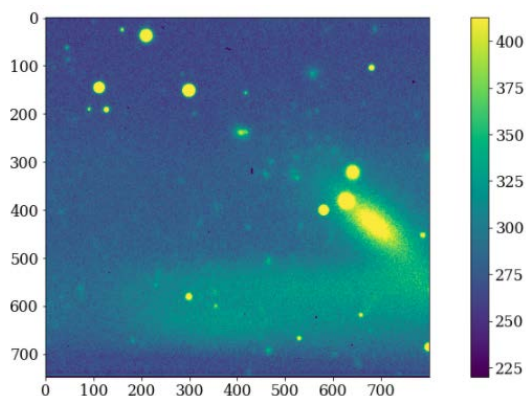
We now enter into the optics of the telescope and we multiply the optical vignetting and illumination pattern, modelled from real images:



After the optics we multiply again by the variations of quantum efficiency of the detector, as modelled from real flat field calibration frames:



Finally, we add the electronic readout noise that we modelled as a Gaussian noise image from noise statistics measured in the overscan regions, where only this component is present:



With this process, we emulated a real observations with instrumental data and can perform further analysis and tests with the true information of what we included in each step. This process has been particularly interesting to deal with the scattered light, developing algorithms

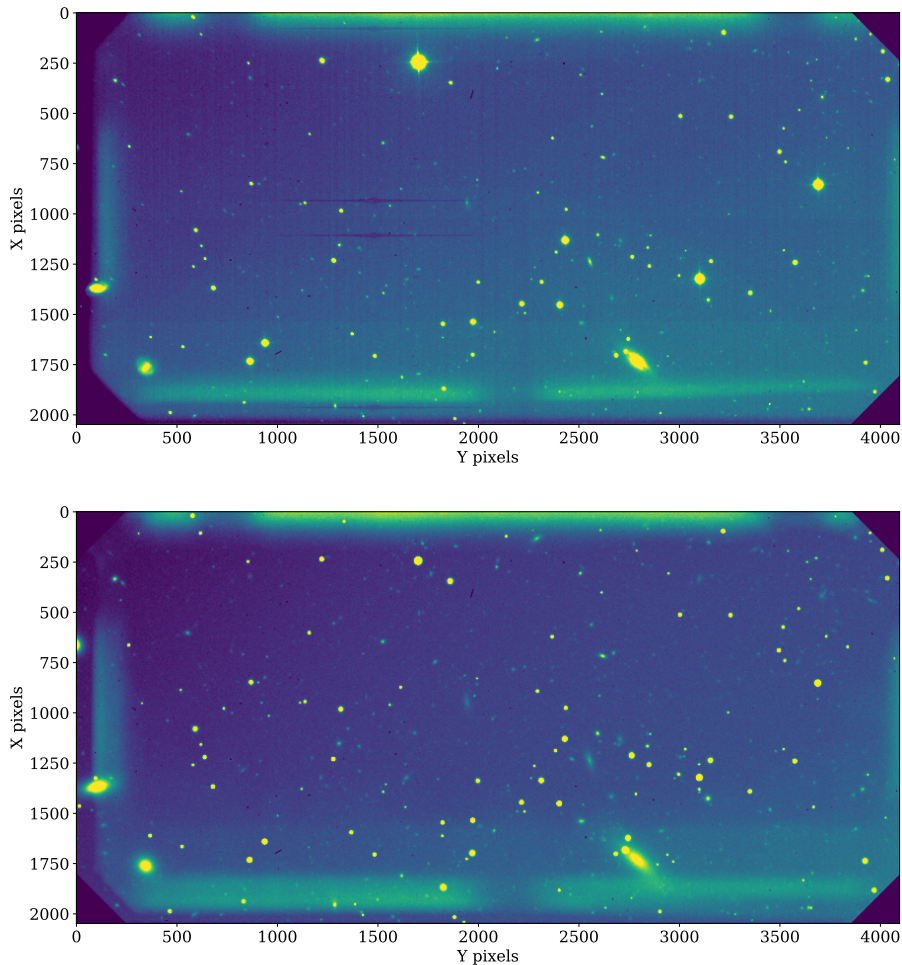


Figure 3.16: Result of image emulation in PAU. The top image shows the reference image we are aiming to reproduce with catalogues and instrument models. The bottom image the final emulated image. The images are statistically equivalent, useful for some particular studies. Few remaining effects yet differ and can be easily spotted such electronic cross-talk, readout patterns and a simpler PSF used in the emulated image.

that can separate its additive component from the other elements in the image. This will be further explained in the data reduction §3.4.2.4. An example of the reference true image and the emulated image in this process can be seen in Figure 3.16. There are still some missing effects visible in this example, such as the cross-talk, but for the scope of the study in this emulation, the inclusion of it was not necessary.

3.3 Observing at the WHT

In large and long-term observing surveys it is fundamental to come up with a strategy that makes the best use of the available time at the telescope, maximizing the scientific goals of the

survey. Taking data at the telescope is a complex and uncertain process where many different situations appear, from telescope and instrument technical issues, to quickly changing weather, changing sky brightness due to the moon and other more unpredictable circumstances that may affect the observation. Having a clear and robust plan is key to minimize the loss of time and data, and adapt to the circumstances quickly.

In the following sections we describe the strategy we followed in the PAU Survey as well as the tools designed for the astronomers we had at the control room to operate the survey.

3.3.1 Survey strategy definition

The power of PAU lies in its capability to deliver low-resolution spectra with no selection, resulting into the highest precision photometric redshifts to date. Combining this strength with the gravitational lensing measurements from other surveys that have exquisite optical quality images enables the capability to make outstanding cosmological studies. For this reason we aim to observe fields that have already been observed by other cosmology surveys with complementary measurements. Furthermore we can make use of their broad band photometry, avoiding spending time observing in those bands.

The PAUS collaboration decided to choose three fields as their primary targets: CFHTLenS W1, KIDS G09 (also named W2 in PAU) and CFHTLenS W3. This implied that our observing time had to be scheduled between November and May, if full nights are scheduled. If half nights can be scheduled, then this range can be increased. In addition to the three previous fields, we consider an additional field as back up if we cannot observe any of the primary fields in the time scheduled: CFHTLenS W4. We also observed in the COSMOS field with more dithers and used it as our main calibration field due to the extended number of bands and spectra available. The target fields of the PAU survey can be found in 3.17.

As shown in Figure 3.11, targeting 3 fields instead of 4 with a total area of $100 (deg)^2$ could provide an additional 4 percent efficiency gain just by avoiding the edge effect due by the filter tray layout. Therefore our priority is to avoid time during Summer to focus on the other 3 primary fields.

Even though most mosaic cameras perform dither patterns around the target location to cover the detector gaps, our camera setup with different filters on top of each detector would not need such pattern. However dithering provides advantages on removing transient events (such as cosmic rays, satellite trails, airplane lights, etc). Also it allows averaging the position were a source lies in the detector, avoiding having particular sources on bad areas such as the CCD corner or the CCD edges where more scattered light is present. Too many dither exposures over the same area is inefficient due to a constant readout time of 20 seconds. The baseline for the main fields in PAU was then set to 3 dithers and 5 for the COSMOS calibration field.

The fields are covered as a sequence of telescope pointings, each pointing containing the corresponding number of dithers. The pointings start from a particular location and follow a

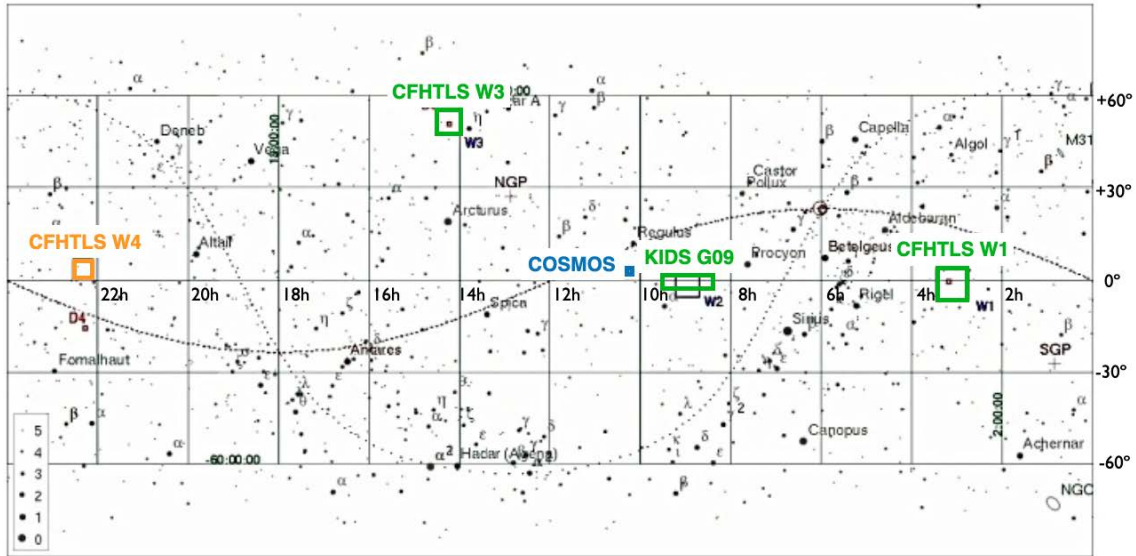


Figure 3.17: The 5 PAU fields in the sky. CFHTLS W4 was left as a backup field in case the allocated time was around Summer and COSMOS field was the main calibration field for development and validation. Image credit: CFHTLS

spiral pattern around the initial location. This is particularly interesting as there are areas within a field with particular interest, such as stripes of calibrated spectra. An example of a field sequence pattern in W1 can be seen in Figure 3.18.

The default exposure times are a compromise to cover a large area and reach deep magnitudes. The goal is to reach magnitudes $m_{AB} \sim 23$ at a $S/N \sim 5$. The distribution of times between filter trays should be chosen as to produce the best photo- z performance. This was hard to quantify, as different galaxy types and different redshifts prefer different distributions. A compromise is to choose a distribution of exposure times that gets similar S/N for all trays. Initially we set an exposure time that delivered a S/N too low in the bluer bands. These exposure times were maintained for the calibration fields on which we started the observations. Therefore we balanced the exposure time for the main fields to ensure a flat spectrum with the resulting exposure times listed below:

Tray name	Exposure Time (s)	AB	S/N
FT_NB455_525	3*90	23.75	2.2
FT_NB535_605	3*90	23.5	2.7
FT_NB615_685	3*100	23.3	3.8
FT_NB695_765	3*100	23.1	3.7
FT_NB775_845	3*120	22.9	3.5

To allow for easy updates, queries and analysis of the survey progress, we created a database that run at the observatory were all the target information was stored. The database contained a

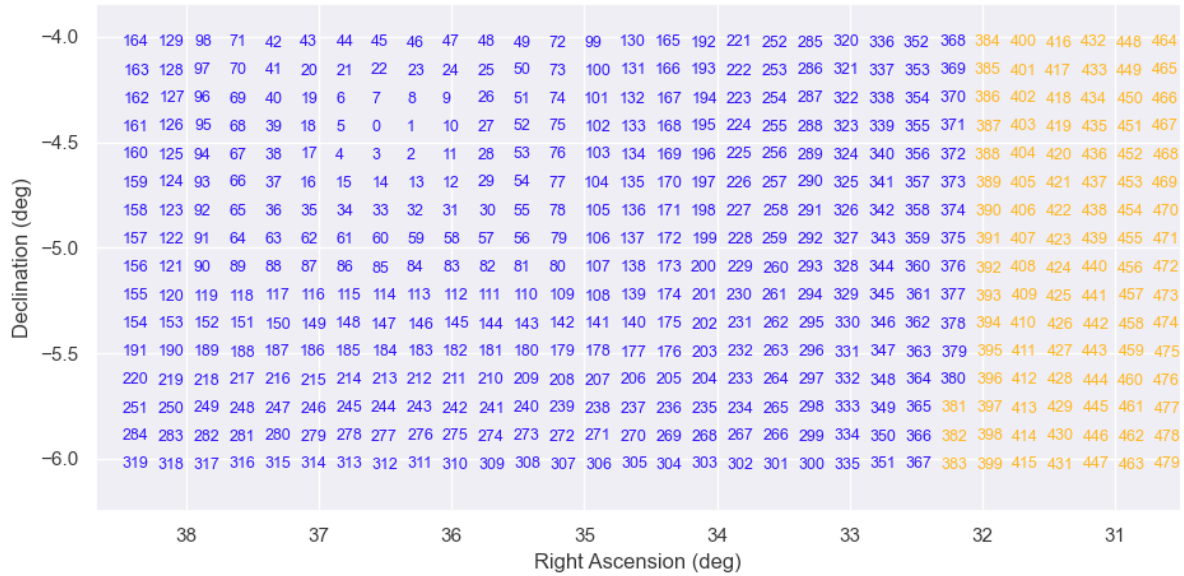


Figure 3.18: An example of pointing sequence in the W1 field for tray FT_NB615_NB685. It starts at 0 from a particular location and starts a spiral around the initial position. Blue numbers indicate pointings already observed and orange numbers are pointings to be completed.

single table with the following columns:

- **target_id** The target id. Targets are unique in this table and represent a combination of pointing, dither and filter tray.
- **field** The field at which the target belongs to. It can be W1, W2, W3, W4 or COSMOS.
- **pointing_id** The pointing id is the spiral sequence of sky locations inside a field. Dithers steps are grouped inside a pointing id.
- **dither_step** The repeated observations inside a pointing, with small offset to vary the position of sources at the detector.
- **filter_tray** The filter tray of the exposure. There are 6 broad band trays with a single large filter and 5 narrow band trays with different filters each covering a single detector.
- **ra** The Right Ascension coordinate of the target.
- **dec** The Declination coordinate of the target.
- **pa** The Position Angle of the target. All PAU survey targets are aligned with the North up, at 0 degree position angle.
- **exp_time** The exposure time of the target. Initially it is set to the default values described previously but if the exposure time is increased, it will update its value.

- **status** The status of the target. Targets to be observed are in status *SCHEDULED* and targets which have been completed are set to *OBSERVED*.
- **exp_num** The exposure number that has observed this target. Initially is set to None and once the target is observed, PAUCam update this element with the corresponding exposure number.
- **relative_sn** The relative Signal to Noise of the target. This value tracks the image quality of the observation to ensure a certain signal to noise is met. Further details about the adaptive exposure time and Signal to Noise estimates are explained in §3.3.3.

Once the targets to be observed in each field are defined, the corresponding entries in the database are filled. The PAU survey scheduled a total of 55346 targets to cover all 5 fields, with all dithers in all filter trays.

3.3.2 Night planning and Survey monitor

Here we describe a typical night of observation for the PAU survey and the rules that apply for an optimal scheduling of targets.

During the afternoon prior to the observation, bias and dome flat fields on all the filter trays planned to be used are taken. During twilight, when the sky is dark enough to open the shutter (approximately 30 minutes after the sunset) we observe available spectrophotometric standard stars in Oke (1990) that have a well-characterised spectral energy distribution and typically a smooth spectrum. These blue stars serve as an absolute calibration points for our narrow band set and are observed in a programmed sequence at the center of each detector with a narrow band filter. The standard stars that we have observed most are shown in Figure 3.19.

Once the sun is beyond 18° below the horizon, passed the astronomical twilight, the science observations can begin. At this stage the observer must decide the field to observe and which filter tray to set. The selected field must be the one at lower airmass (higher altitude) for the next hour. In any case starting airmass must be less than 1.7 or higher than 35° in altitude. At any given time, there is normally only one primary field observable, and therefore the decision on the field is normally simple. The only situation where a lower field could be chosen is when the moon is too close to the field that stays higher, and the background would be too bright to observe. The Moon illumination fraction, the Moon altitude and the Moon distance to the field is key in the decision of the filter tray to set.

Moon fraction	Moon alt	NB775_845	NB695_765	NB615_685	NB535_605	NB455_525
>50%	>10					
>30% <50%						
>15% <30%						
<15%	>0					
	<0					

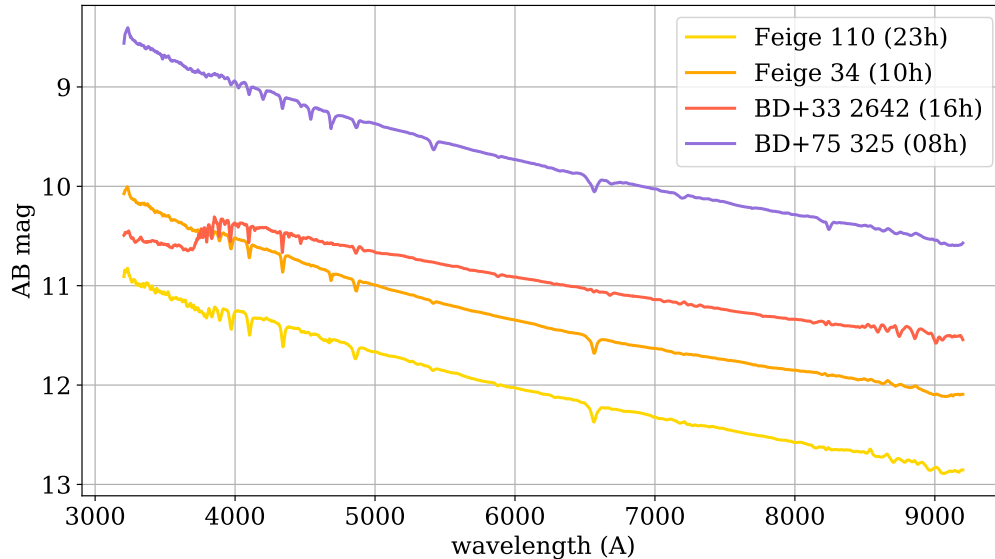


Figure 3.19: The four standard stars most observed with PAU from the Oke 1990 set. At a variety of Right Ascension hours, these are available at high altitude throughout the whole year, both at evening and morning twilight.

Additional unquantifiable considerations are to optimize the quality and homogeneity of the data. As bluer bands suffer from a larger seeing and higher extinction, we tend to observe the bluer trays at higher altitudes as well as in situations with better seeing conditions. Also we try to have an overall view of the observing run in terms of bright and dark time available to distribute the observing time between the trays as to reach at the end a uniform distribution.

The PAU Survey Strategy tool provides an elevation plot of the PAU fields visible in the current night, with sunset, sunrise and twilight times, as well as moon altitude and distance to each of the fields. This plot stays in the observer screen all night and provides very useful information in the decision tree explained previously. This figure is called *Night Plan* and can be seen in Figure 3.20.

With all the previous information we slew the telescope at the target location and perform a focus sequence there, to ensure the camera is at its optimal focal distance. At this stage PAUCam can start observing the field with a particular filter set. The Survey Strategy tool allows us to query the targets database and export the next pointings to observe. It will produce a YAML file with all pointing information in a format that can be inserted into the PAU Camera Control System (PAUCCS). PAUCam will automatically load the targets one by one and command to the telescope the next target position, wait until the telescope is properly tracking and begin the observation. It also updates the Survey Strategy database once completed with the exposure number of the target observation. In this way, the system is almost automatic for the whole

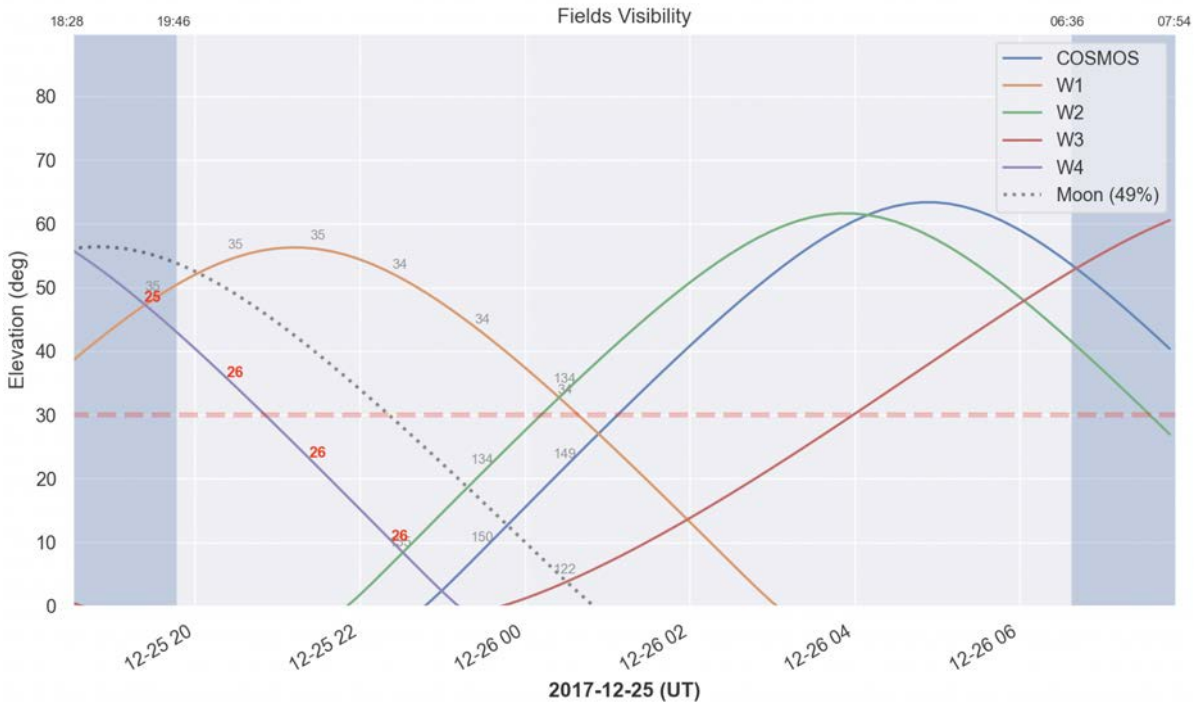


Figure 3.20: A screenshot of the *Night Plan* produced by the PAU Survey Strategy tool with the elevation of the visible fields, moon elevation, fraction and distance and relevant times throughout the night. In example during the Christmas night of 2019, all fields are available, but W4 is already too low and too close to the Moon at the beginning of the night. The observer will start with W1 field probably with the reddest tray due to the proximity to the Moon and its advanced fraction. At midnight it will switch to W2 with any tray as the Moon will set soon. At past 4am, it can slew to COSMOS until the end of the night. W3 will not be observed either as it is higher than COSMOS passed the twilight.

survey strategy sequence, leaving the astronomer only to monitor the progress of the survey and take decisions on what to observe next. The camera also performs automatic quality analysis over the observations in real-time, providing information to the observer on the levels of background, PSF quality and size, number of sources detected and read noise of the electronics.

The observer can monitor the progress of each field and overall survey status with the *Survey Pointing Status* plot. Here we can monitor the pointings completed for each field and filter tray and the number of dithers at each pointing. With this view one can quickly identify if a filter tray is far behind in the pointing sequence or if there are missing dithers to complete for a particular tray. Typically, sequences are observed in runs that last approximately an hour, exporting targets of the same filter tray and dither for 30 or so consecutive pointings. We separate in time the consecutive dithers to avoid having the same conditions on all 3 overlapping observations. A sample of the *Survey Pointing Status* can be seen in Figure 3.21.

After exporting sequences of targets throughout the night, we arrive to the morning twilight



Figure 3.21: A screenshot of the *Survey Pointing Status* plot produced by the PAU Survey Strategy tool. The figure displays the pointing status for each field and filter tray, indicating the number of dithers completed with the color of the points. This sample shows the current status of the fields after observations from 20A. The main fields W1, W2 and W3 extend to much higher pointing values up to 700, but here we only show the first 100 for clarity. This tool is very useful to monitor the progress of each field during observation and schedule sequences of targets to be observed.

were we will observe again spectrophotometric standards as in the evening. Once the sun begins to brighten the sky, 20 minutes before sunrise we close the dome and initiate the automatic transfer of the observed data to the PIC data center.

Even though the observations are quickly processed in the camera system, we rely on the full PAUdm data reduction pipeline (explained in Chapter 3.4) to obtain accurate quality measurements of the images observed. The data processing pipeline automatically triggers as soon as the data is transferred to the data center, enabling a rapid feedback in less than eight hours. The report is available before the observations for the next day start, so exposures that do not meet a certain quality for the survey are re-observed. The Survey Strategy tool allows to ingest the automatic report produced by the PAUdm over the previous night and set the bad quality observations back to SCHEDULED status. Those will appear as holes in the *Survey Pointing Status* plot. These are the metrics that mark an observation as invalid and require re-observing:

- median seeing larger than 1.6"
- transparency lower than 50% (or 0.4 magnitudes) as defined with the measured zero point
- fewer than 5 stars with $i < 21$ for calibrating the exposure
- other quality flags output by the nightly pipeline such as PSF model failure or astrometric solution errors

3.3.3 Adaptive exposure time with SN estimator

After few semesters of observation, we realized that some exposures that met the quality requirements, degrade the overall signal to noise when they are combined within the same pointing. For this reason we adapted the Survey Strategy tool to analyze the current weather conditions and dynamically adjust the exposure time such that the combined SN of the various dithers reach a certain threshold. This also allowed us to observe in very bad seeing conditions increasing from the previous limit of 1.5" seeing up to 2", with an increase of exposure time. In that way we could make use of times were weather was not good, by progressing slower but not throwing away the low quality data. Additionally we analyzed the previous data and estimated the SN in each pointing, allowing an additional dither for those pointings that did not reach a certain SN. This is why in Figure 3.21 there are some pointings with 4 dithers in the main fields, where we originally targeted 3.

The signal to noise of a specific observation is estimated as:

$$(3.4) \quad \text{SN}_{\text{measured}} = \frac{f_{\text{ref}}}{\sqrt{f_{\text{ref}} + f_{\text{sky}}A + \sigma_{\text{read}}^2 A}} \times \frac{\text{SN}_{\text{goal}}}{\text{SN}_{\text{ref}}}$$

where f_{ref} is a reference flux of an object expected to meet the target SN_{goal} , f_{sky} is the measured background in counts/pixel, σ_{read}^2 is the read noise variance and A is the area of an aperture that measures specific flux fraction of the reference source. This aperture will depend on the size of the objects and the measured sky seeing. Additionally we include SN_{ref} , a reference SN measured with the current exposure times at the reference object.

With the previous formula we have a relative SN of a single dither. However we observe on average 3 times the same area in the sky in each band, therefore we can compute the combined observed SN as:

$$(3.5) \quad \text{SN}_{\text{observed}} = \sqrt{\sum \text{SN}_{\text{measured},i}^2}$$

Therefore, when the relative SN is 1 or above, means that the goal SN has been reached. As the relative SN is expected to be completed once all dither exposures have been observed, we can estimate the remaining SN to reach the goal per exposure as:

$$(3.6) \quad \text{SN}_{\text{remaining/exp}} = \sqrt{\frac{1 - \text{SN}_{\text{observed},i}^2}{N_{\text{remaining}}}}$$

If we construct the SN formula with flux rates where f_{ref} is the flux rate times the exposure time as $r_{\text{source}}t_{\text{exp}}$ and the background f_{sky} as $r_{\text{sky}}t_{\text{exp}}$ we obtain:

$$(3.7) \quad \text{SN} = \frac{r_{source} t_{exp}}{\sqrt{r_{source} t_{exp} + r_{sky} t_{exp} A + \sigma_{read}^2 A}}$$

which we can reorder as:

$$(3.8) \quad (r_{source}^2 / \text{SN}^2) t_{exp}^2 + (-r_{source} - r_{sky} A) t_{exp} + (\sigma_{read}^2 A) = 0$$

Such quadratic equation $at^2 + bt + c = 0$ can have two solutions as:

$$(3.9) \quad t = \frac{-b \pm \sqrt{b^2 - 4ac}}{(2 * a)}$$

Applying the solution to our SN formula, we can obtain the predicted exposure time that reaches the remaining SN as:

$$(3.10) \quad t_{predicted} = \frac{r_{source} + r_{sky} A + \sqrt{(-r_{source} - r_{sky} A)^2 - 4 \frac{r_{source}^2 \sigma_{read}^2 A}{\text{SN}_{remaining/exp}^2}}}{2 \frac{r_{source}^2}{\text{SN}_{remaining/exp}^2}}$$

As this formula could end up in too short or too long exposure times, we establish a minimum for each tray and a maximum of 180s. The maximum limit is constrained by the tracking accuracy of the WHT, as by default PAU Survey images do not enable its autoguiding system. The PAU observer can enable the adaptive exposure time by just entering into the system the average seeing and sky background measured by the quick reduction real-time analysis of PAUCam for the current filter tray. After this procedure was put in place, combined observations had a much more homogeneous SN, with the corresponding benefits in the final scientific catalogue.

The Survey Strategy tool can plot the relative SN of all the observed pointings for each field and filter tray combinations, in a plot similar to the *Survey Pointing Status*, but with the SN estimated from the formulae previously stated. The tool can also identify which pointings with no pending dither observations do not reach the SN goal (or a fraction of it). We set by default a threshold of 0.7 to generate further dithers with the adequate exposure time to reach it. An example of the SN sequences and the evaluation of the quality in each can be seen in Figure 3.22.

3.4 Reduction of narrow band images

In this section we describe the PAUdm Nightly pipeline, a set of algorithm that take care of all image processing from the Raw images to clean and calibrated detector frames, where specific algorithms for photometric calibration and scatterlight correction had to be designed.



Figure 3.22: Example of the SN analysis of the PAU Survey Strategy tool. Top figure: the combined relative SN estimated for the observed dithers at the various fields and filter trays. Redder points indicate higher SN with respect to its goal. Depending on the weather conditions at different semesters, the SN could vary from field to field. In this example W2 and W4 show a higher SN than the rest. Once the adaptive exposure time was enabled, the variations between fields were reduced. Bottom figure: The analysis of which pointings do not meet their corresponding SN. In this example there are 5 pointings in the bluer tray of COSMOS that did not meet the 70% goal. The COSMOS field was completed after this technique was implemented and therefore we did not add additional dithers. In W2 and W3 fields multiple pointings did not reach the limit and a 4th dither was planned and observed, as can be seen in Figure 3.21

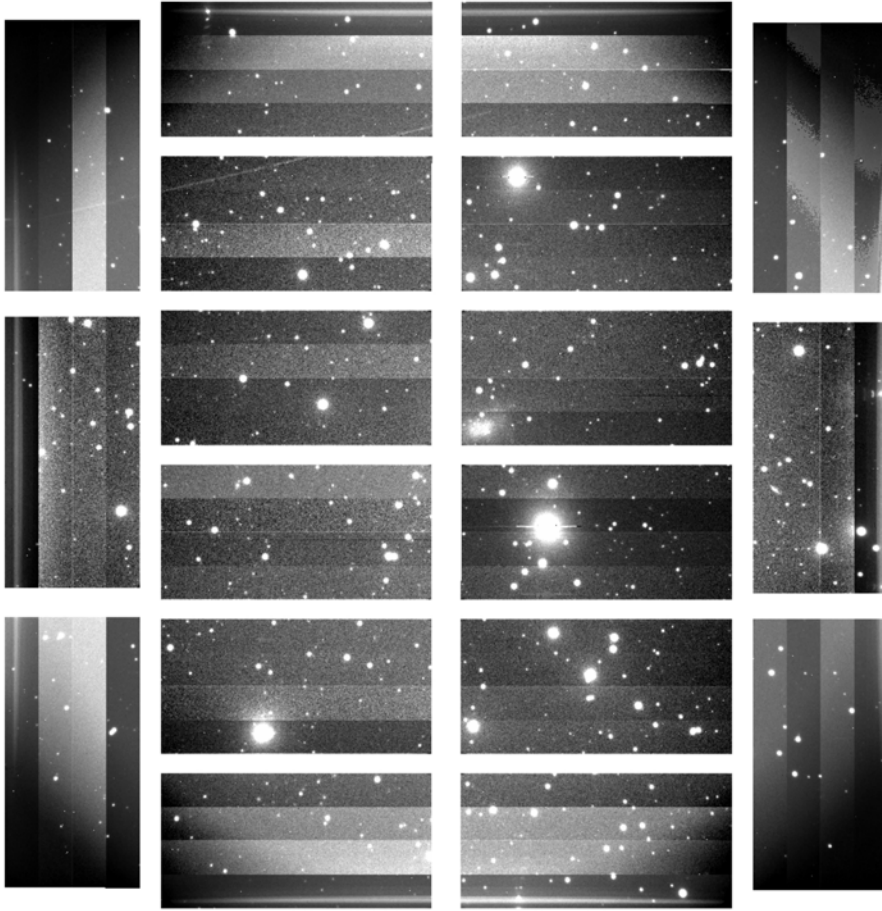


Figure 3.23: This figure shows a sky exposure of the full 18-detector PAUCam mosaic. As this a raw image, all instrumental signatures are present, and the 72 amplifiers can be identified as well as the vignetting from the WHT prime focus corrector.

3.4.1 Data overview

3.4.1.1 PAUCam Raw data

PAUCam is an 18-detector imager camera (Padilla et al., 2019) that produces multi-extension fits (MEF) files containing the pixel imaging data for the various types of frames, with associated metadata information in the header of each extension. Separate header-data units (HDU) are created for each amplifier, resulting into a 72 extension MEF file of about 670 MB. During the afternoon, bias and dome flats are observed for calibration purposes. In the twilight, when the sky is too bright for science observations, a high altitude standard star is observed with each detector-filter set for photometric calibration and total system throughput calibrations. Once the sky is dark enough, regular field observations are taken.

The PAUCam raw data is organized in observation sets, that group exposures from a single night. Typically for an observing night, the calibration frames such as bias and dome flat-fields

are stored in one observation set while the rest of the science sky exposure are stored in a separate one. Observation sets allow also to separate observations between the PAU survey and community observations in shared nights.

The three main types of raw exposures are Bias, Flats (dome) and Science. There are also other types of exposures such as stacked focus, PTC, test, etc. but those will not enter into the data reduction process.

3.4.1.2 PAUdm reduced data

After the image processing that will be described in the following chapters, 5 more types of images will be created; Master Bias, Master Flats, Reduced Science, Reduced Weight and Reduced Mask. Reduced mosaic exposures are separated into eight FITS files, one per detector on the most central and illuminated CCDs. Additionally, PSF models and astrometry catalogues will be stored in the archive, together with other associated data that will be stored in the PAU database such as image zeropoints and single-epoch detections.

3.4.2 Instrumental detrending

Raw images contain a significant amount of instrumental signatures that need to be neutralized or masked before carrying out any photometric measurement. This section describes the correction flow that take place in all raw science exposures from PAUCam.

3.4.2.1 Gain calibration and PTC analysis

During the assembly of the PAUCam instrument, we performed the first Photon Transfer Curve (PTC) analysis described in Janesick (2001) to estimate the gain for each of the four amplifiers in all detectors. Once in the mountain, with PAUCam installed in the Prime Focus of the William Herschel Telescope under a stable installation, we repeated the test as the environment conditions were different (low temperature, pressure, and humidity) and with the interaction with other devices and a different grounding. We decided to repeat this test on each observing run since 2015.

In order to process the PTC we need to obtain dome flats with a broad band filter, usually g or r, and obtain pairs of images with scaled exposure times, from one to 20-30 seconds, from zero exposure until saturation.

As the PAU Survey allocated narrow bands only on the eight central CCDs (the unvignetted area) we have only analyzed these detectors in the scope of this study. An analysis of the gain for the full focal plane (the 18 detector set) is described in Padilla et al. (2019).

Each $2K \times 4K$ Hamamatsu CCD of PAUCam has four outputs, and each one must be analysed independently, so the eight central CCD have 32 regions to analyse.

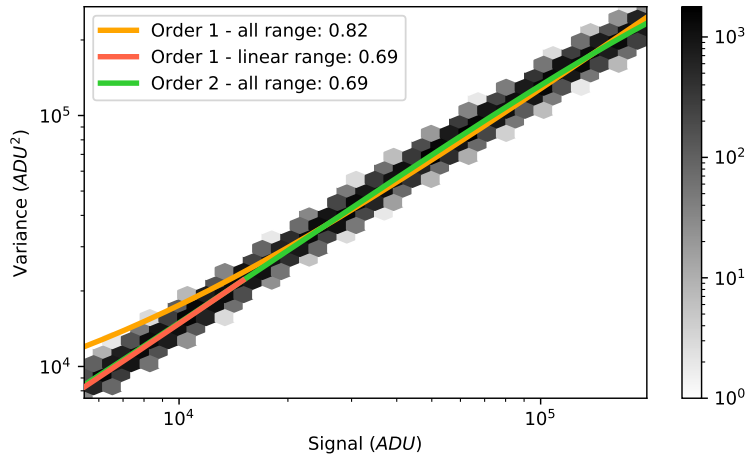


Figure 3.24: The relation between the mean value and its variance used to infer the gain as described in the PTC process for the first amplifier of CCD 1 in a test during the 19A period. The gain is measured as the inverse of the slope in the linear regime, where first component of a parabolic fit matches the slope of a linear fit in the linear regime of the detector.

To compute the PTC we subtract overscan median value for each image and average the pairs of images with the same exposure time to remove possible patterns and reduce the noise. A random choice of small squares of 100×100 *pixels* in the subtracted and averaged images are used to determine the variance (in ADU^2), and the mean signal (in ADU). The PTC represents the variance as function of the signal as showed in figure 3.24. The gain can be evaluated as a linear fit between the variance and the signal. The slope in the fit is the inverse of the gain, it in units of e^-/ADU . As the detectors do not respond linearly on the very high end of the range as shown in Figure 3.24, we have measured the slope for the range where the linear fit matches the non-linear fit for the first order component and computed the gain over the linear behaviour of the detector only.

Since the first installation of PAUCam in the Prime Focus of the WHT in 2015 we ran 16 PTC tests. Figure 3.25 show the values of the gain for each CCD/Amplifier.

3.4.2.2 Overscan, Gain, Bias and Flat correction

In first place a row by row overscan correction is performed. To take into account low frequency oscillations in the vertical readout, a 8-row Gaussian filtering is applied, allowing a more accurate subtraction of the electronic bias. The gains for the 72 amplifiers in the mosaic that were measured using a Photon Transfer Curve analysis are applied to each amplifier array, and ADUs are transformed to electrons in this early stage of the process. To compensate for readout patterns that are present in all exposures, a master bias frame is produced, combining around 10 individual zero-exposure bias frames. As we have identified that the first 2 or 3 exposure

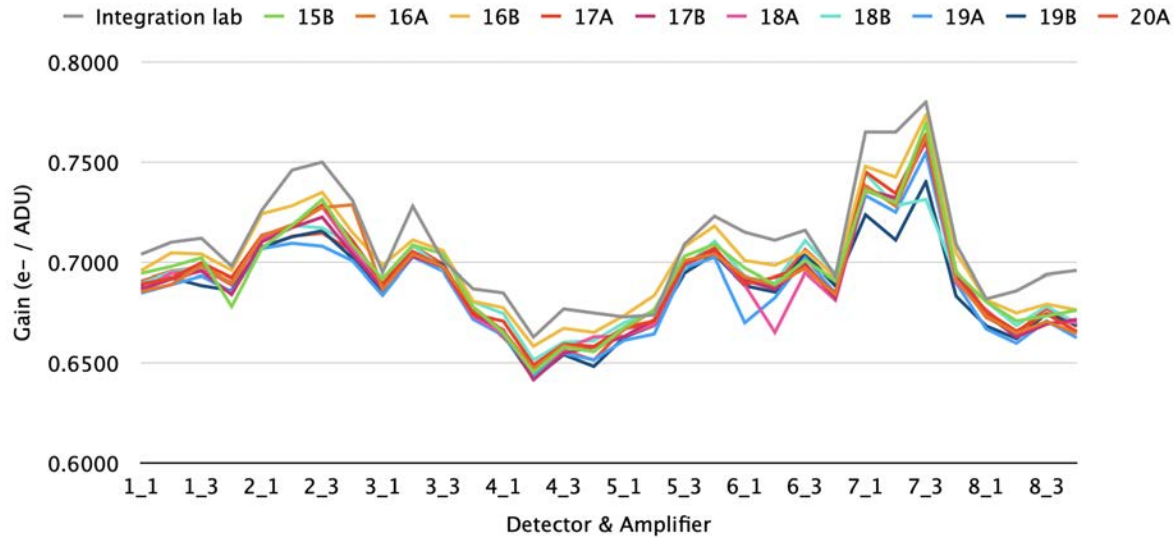


Figure 3.25: The gain for the eight central CCD and for each amplifier of its obtained in the various test obtained from 2015 until 2020. This figure highlights the importance of continuous gain calibration as the value varies with time and is very sensitive to different installations such as the one in the integration lab.

after a full readout system restart include some residual patterns, the individual bias frames are analyzed and those with high levels of noise are removed from the median average. Finally, a master flat is produced from the individual dome-flat exposures that are taken every afternoon, previous to the night observations. Scattered light residuals found in the raw flat images needed to be removed, as this signature in the flat images is an additive component of the light, while the flat field needs to contain only multiplicative factors of the main optical path. The process of removing the scatterlight will be described in the scatterlight correction section, together with the correction in the scientific sky images. Once a clean master flat has been produced, it is divided to the science images.

3.4.2.3 Crosstalk correction

As explained in §3.4.2.2, the incident photon flux in the CCDs is converted into photo-electrons which are then stored in the individual CCD pixel potential wells. During read-out the charge in each pixel is passed sequentially (clocked) to an on-chip amplifier that converts the charge into a voltage and amplifies this voltage before sending it to an ADC. To reduce read-out time, PAUCam CCD's have four output amplifiers, one for each read-out region consisting of 4096 rows and 512 columns.

Since the four amplifiers are read in parallel it is likely that current is induced through magnetic fields between the different channels read at the same time, effect which is commonly known as crosstalk. Given the way the charges are moved, those simultaneous pixels are located



Figure 3.26: On the right of this crop image an 8.5 magnitude star saturates the detector producing bleeding and a visible crosstalk signal on the mirrored amplifier in the left side of this image.

in the same row of the CCDs and in columns that are mirrored of each other with respect to the borders of the four read-out regions.

Too many electrons in a given pixel cause electrons to overflow to the nearby pixel wells in the same column. This effect is known as saturation bleeding and cause this elongated shape around very bright stars. If the pixel signal is above 18-bit (current depth of PAUCam ADCs), it will saturate at $N^{sat} = 2^{18} - 1$ (or 262,143) ADUs. In Figure 3.26 one can see ghost images parallel to the elongated star image due to the crosstalk effect. Even we can only see the ghost image from the saturated pixels, the crosstalk effect happens at all levels but is only visible when the signal is strong enough to stand out from the noise in the background.

To remove the crosstalk signal we need to estimate the induction ratios $r_{xy_{ij}}$ between each pair of amplifiers. Therefore, each amplifier would be composed by its direct integration signal I^{int} plus the signal of all the remaining amplifiers in the mosaic being readout at the same time scaled by the induction ratio.

$$(3.11) \quad I_{xy}^{tot} = I_{xy}^{int} + \sum_i^{18} \sum_j^4 I_{ij}^{int} r_{xy_{ij}}$$

for a given set of detector x and amplifier y .

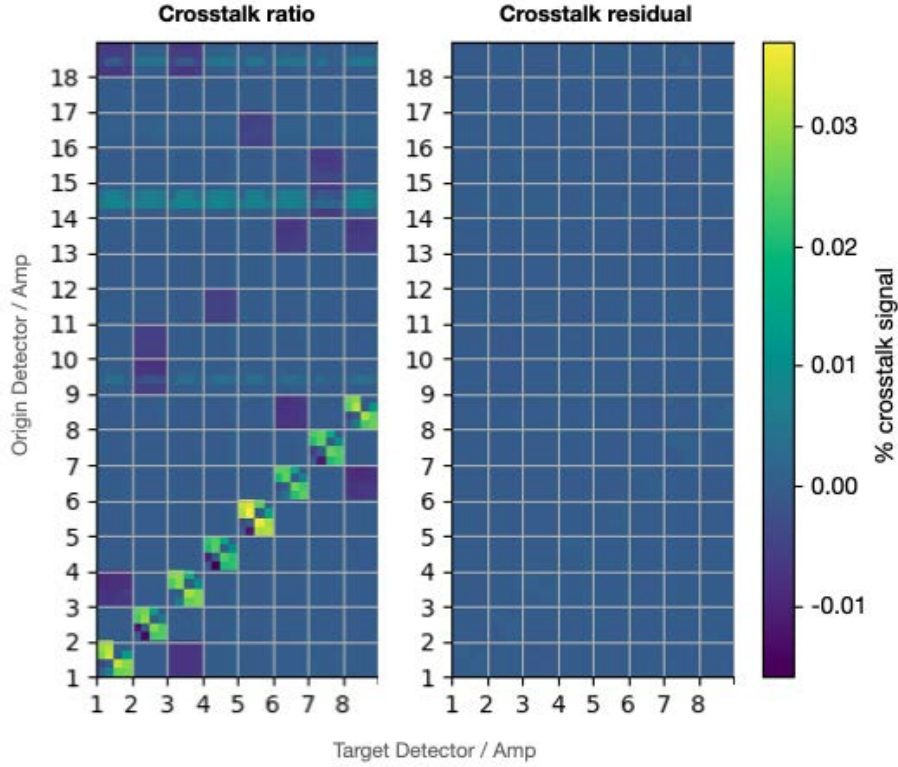


Figure 3.27: Left: The crosstalk induction ratios used to apply the correction. Right: The residual crosstalk ratios measured on a different dataset after crosstalk correction.

Even crosstalk analysis were made at the facility lab in Barcelona, the crosstalk ratios may vary once mounted in the telescope so we had to estimate them with sky data when the instrument is mounted. In order to estimate the ratios, we into the ghost positions of the saturated pixels, looking to the change in flux with respect to the background. If the mirrored position contains sources or is not flat enough, we discard the measurement.

$$(3.12) \quad r_{xy_ij} = \frac{f_{ij} - bg_{ij}}{N^{sat}} \quad (\text{where } x_y \text{ is saturated})$$

We compute the ratios for the 8 central detectors (where narroband filters are located) against the remaining 18 detectors and 4 amplifiers, resulting into a total of 2272 pairs. To remove the crosstalk signal, each amplifier needs to subtract the flux of the other 71 amplifiers by the corresponding ratio. Figure 3.28 illustrate how the ghost produced by the bright star on the right is now removed after applying the calibrated crosstalk ratios.

To verify the correct implementation of the crosstalk calibration, we have measured the same crosstalk ratios over a dataset that has been already crosstalk corrected. Figure 3.27 illustrate the matrix of crosstalk ratios and how the values go towards 0 once the calibration has been applied.

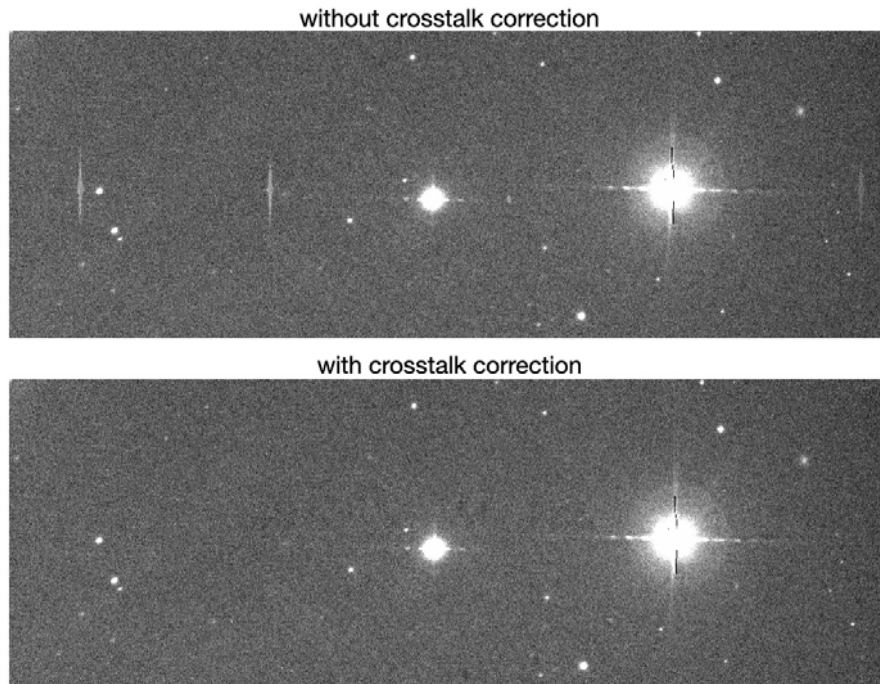


Figure 3.28: Sample of images with and without crosstalk correction enabled. The 3 ghost signals from the saturated star mirrored on the other amplifiers are completely removed after applying the corresponding ratios and subtraction.

In this figure it can also be seen how the intra-detector crosstalk ratios are significantly higher than the inter-detector ones, as expected due to further distance of wiring in the electronics.

3.4.2.4 Scatterlight correction

The numerous bands required for PAU required to engineer an efficient design to include all 46 filters. The final solution was to distribute multiple small filters, each covering a single detector in filter trays, such that with 5 trays and 8 narrow bands in each covering the central most illuminated detectors, we could cover the entire narrow band range. This design caused a side effect in the quality of the images, as significant reflections in the lateral edge of the filter and the filter trays was bouncing creating a localized scattered light in the edge of each detector image. This is clearly visible in the top Figures 3.31 and 3.32. This effect was soon identified and the camera was reopened in mid 2016 to redesign the filter trays and minimize the problem. After the camera intervention, the background caused by scattered light was reduced by a factor of 4 as it can be seen in Figure 3.29. The scattered light was not fully eliminated and specific processing had to be developed to correct the areas affected by this issue, without compromising the light from the sources we need to measure.

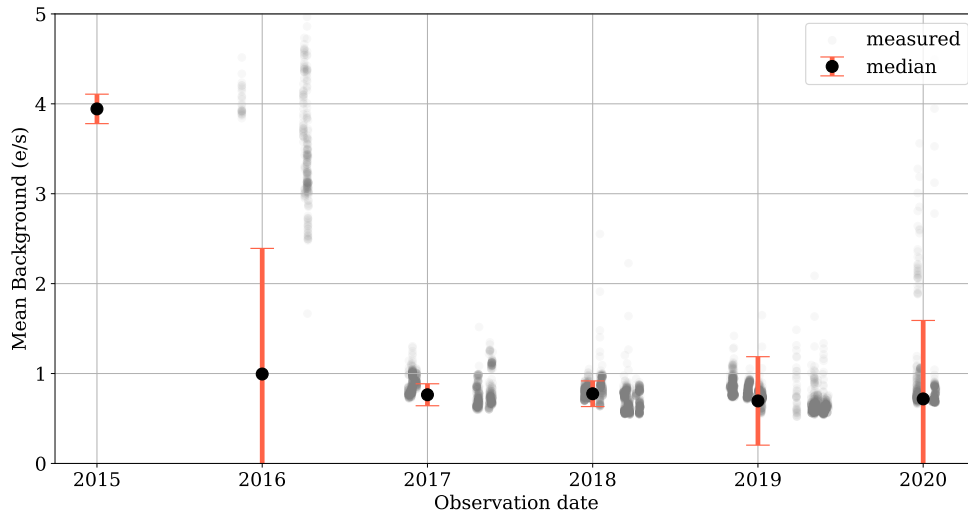


Figure 3.29: The evolution of the mean background in the reduced PAU images. With the original filter tray design, a large substantial light was reflected and at the edges of the filter pieces and filter tray, causing severe levels of scattered light arriving to the science images. An intervention was carried out in mid 2016 and the mean background was highly reduced from $\sim 4e^-/s$ to less than $1e^-/s$, as it can be seen in this plot in 16B observations and beyond.

Flat-fields Flat fields can be divided into two main frequency components across the focal plane: a low pass band due to vignetting and a high pass pixel due to pixel variations (i.e. dust, dead and hot pixels). Fortunately the scatterlight is in between with a mid-size frequency variation, so we can easily isolate and correct for it. In first place we use the broad-band flat fields made of a single large filter - that does not contain scattered light - to construct a vignette image, mostly produced by the prime focus corrector optics. With a low pass filtering we could isolate the vignette component and dismiss the high frequency variations. Dividing the narrow-band flats with the vignette profile flattens the image and leaves the SL as the lowest component in the image, ensuring that with a high pass filtering would leave the SL only component that can be then subtracted from the flat.

Science images In the case of science images, the process is more complicated as the large extended sources (i.e. large galaxies or nearby objects) may have a similar spatial frequency as the scattered light. In the cases where the target sources are distant and small galaxies, a low pass filtering with sigma-clipping isolate the scattered light, unaffected the photometry of the small sources. This is similar to background subtraction techniques, with the difference that the codes used take into account the preferable direction of the scatterlight across the edges of the detector-filter system.

To preserve the flux from extended sources, sky flats was the most effective solution. To

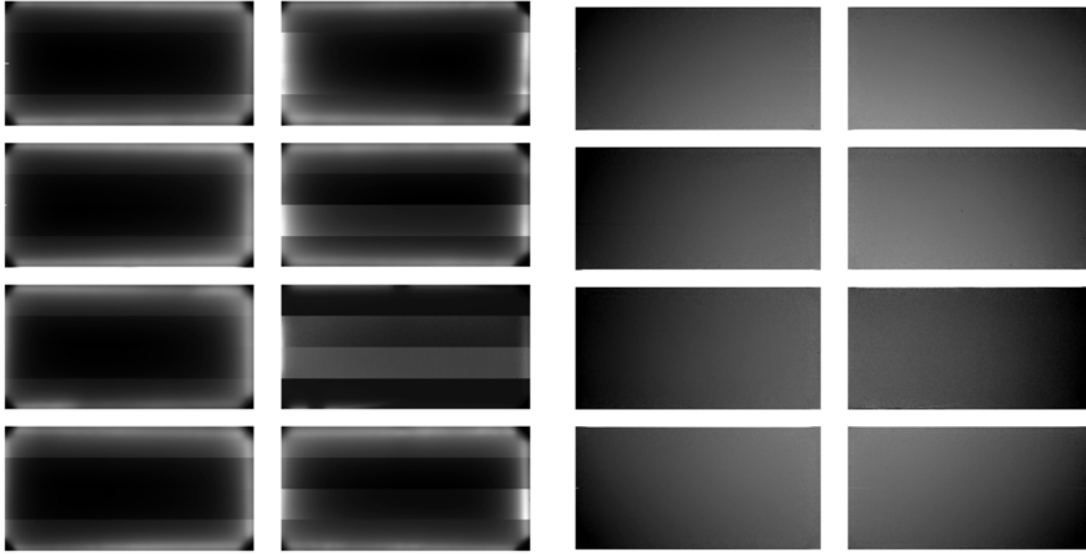


Figure 3.30: This figure shows how the scattered light in the flat fields was removed with the proposed algorithm, preserving both the low frequency component of the vignetting and the high frequency of the pixel-to-pixel variations.

produce the sky flats, a large set of images is combined following a median average stack. We identified that around 50 images is enough to provide an accurate modelling of the background, including scatterlight. A combination of multiplicative factors (residuals from the dome flat) and additive components (mostly scatterlight) are present in the sky flats and therefore need to be separated to subtract and divide the images to detrend the effects properly. The complication of this method is that requires multiple epochs to detrend an image, and the images stacked need to have similar background levels, which might not always be possible with shorter observations. However when possible this method provides an illumination correction more accurate than simply with the dome flats and is applied as enough dither data from the same tray is available.

3.4.2.5 Cosmic Ray detection, rejection and masking

The fully depleted detectors are also very sensitive to cosmic rays impacting the silicon sensitive area and leaving a trace that will impact the photometry if overlaps with the target source. The typical exposure times for the PAU survey last between 2 and 3 minutes which leave a low density of cosmic in the exposure. However the camera is capable of performing long exposures thanks to its integrated auto-guiding system, and in these cases having a proper cosmic ray identification will be critical. In any case, we perform a cosmic ray detection, rejection and masking, following a Laplacian filtering algorithm known as L.A.Cosmic (van Dokkum, 2001). The main idea is to take advantage of the specific sharp profile of the cosmic ray hits, due to the fact that the particle is blurred by the atmosphere PSF as the rest of the photons in the image and can be highlighted with a Laplacian filtering. For the noise model described in the algorithm, we provide

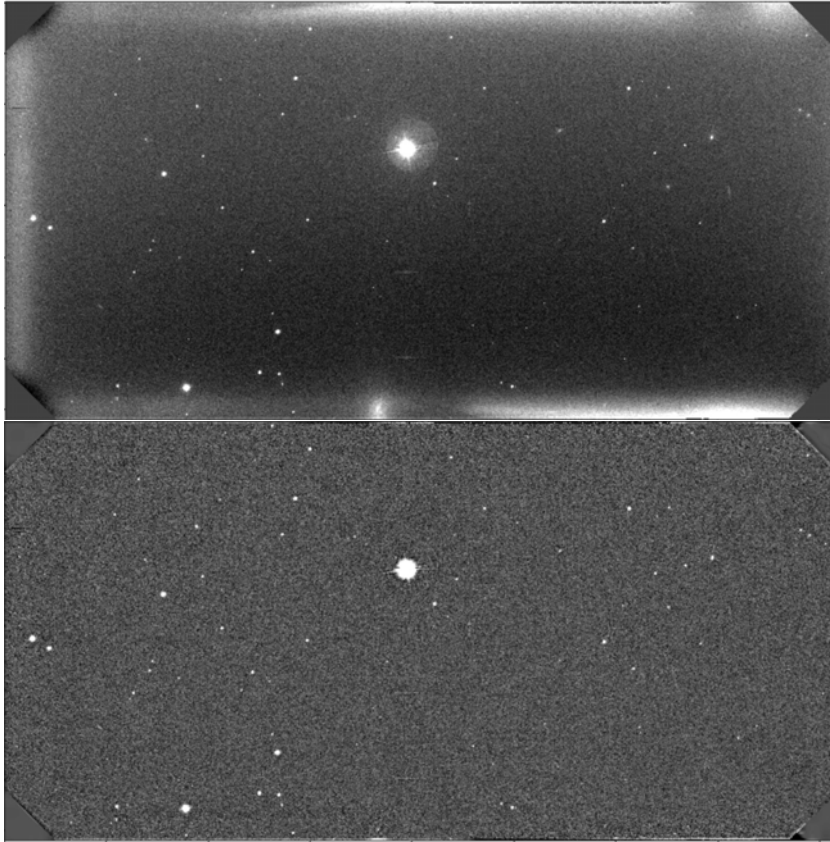


Figure 3.31: This figure shows how the scattered light in the scientific frames was removed with the proposed algorithm, preserving the flux from the stars and small galaxies, but affecting the larger extended sources and haloes.

the measurement of the readout noise taken in the overscan regions of the raw image.

3.4.2.6 Image Masking

In order to keep track of the history of each pixel, we attach a mask to each exposure image, of the same size as the original. The pixels in the mask are mapped into a bitmap where each bit corresponds to a certain issue. Bits 1 to 5 are left for SExtractor flags, and will specify when a detected source is crowded, merged, under a halo, truncated or deblended. Bits 6 to 12 will be set at the pixel level in the mask image to specify for crosstalk corrected, detector cosmetics (dead or hot pixels), saturated pixel, rejected cosmic ray, highly vignetted area, close to edge and very high distortion. Other bits are left for later photometry such as scatterlight affected background, very high extinction, discordant measurements, astrometry issues and noisy background.

To classify the cosmetics in the detector, a flattening process over the master flats is applied and pixels with less than 40% of the normalized flux will be added to the mask and flagged as bad pixels.



Figure 3.32: This figure shows how sky flats can correct for scatterlight and preserve light from extended sources (M101 in the pictured image).

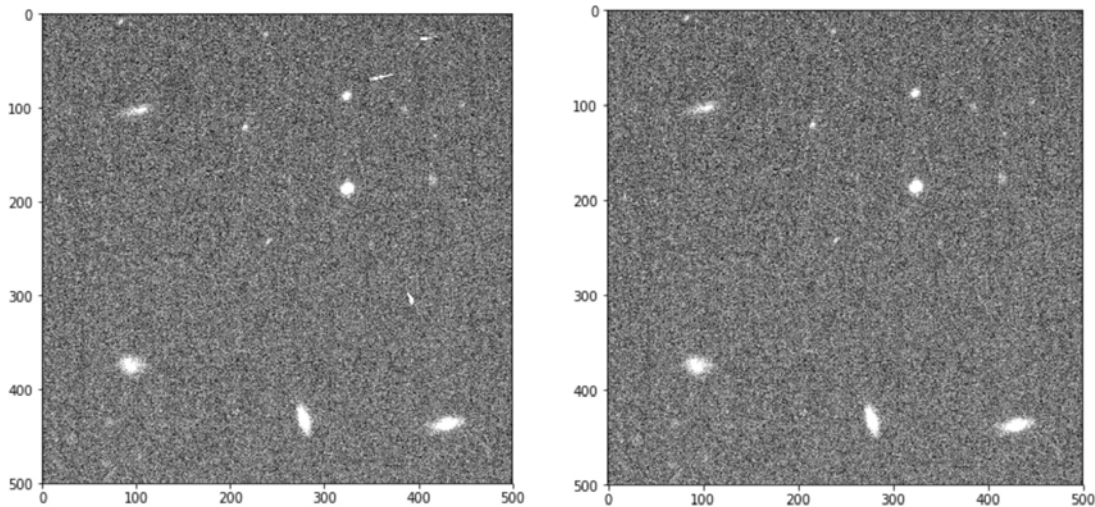


Figure 3.33: This figure shows how the Laplacian filter algorithm can successfully identify, mask and interpolate the Cosmic Ray hits in a single-epoch image.

Attached to the science frames it is also provided the weight map, being the master flat field with the elements in the mask reduced to a very low value. This will be later used to flag objects in the aperture photometry process and in the survey mask, described in §3.5.1 and §3.5.3 respectively.

3.4.3 Astrometric calibration

Once the image has been cleaned and detrended for all known instrumental effects described before, we can proceed to correct for telescope pointing inaccuracies and optical distortions. There are three key elements in this process; the reference catalogue, the source extraction and the astrometry matching and correction.

3.4.3.1 Gaia reference catalogue

As in 2018, Gaia has recently released the Data Release 2 (DR2) of their public dataset (Gaia Collaboration, 2018). This dataset provides a more complete catalogue, including proper motions, which can be critical for the astrometric solution. However as the observation period of PAU and Gaia overlap in time, the proper motions did not result into significant improvements in the astrometry accuracy.

3.4.3.2 WCS calibration

In order to get a first extraction of the sources in the image, we use SExtractor (Bertin, Arnouts, 1996) in a configuration mode specific for astrometry, with extended measurement of windowed positions and a vignette matrix for each source.

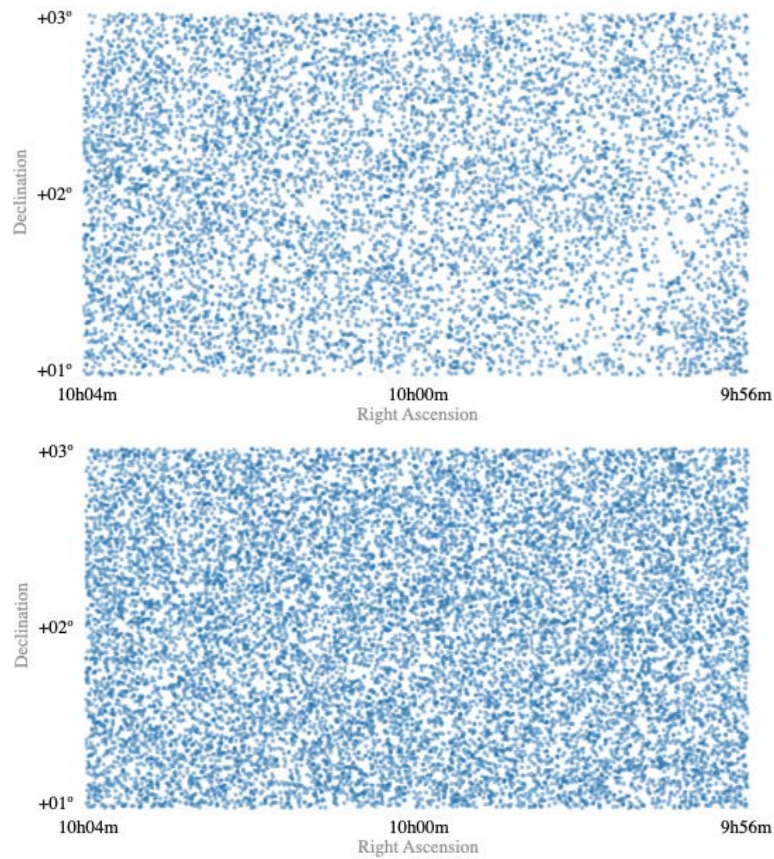


Figure 3.34: This figure shows the stars in the Cosmos field from Gaia in DR1 (top) and a more uniform and dense coverage in DR2 (bottom) enabling a better astrometric reference for calibration.

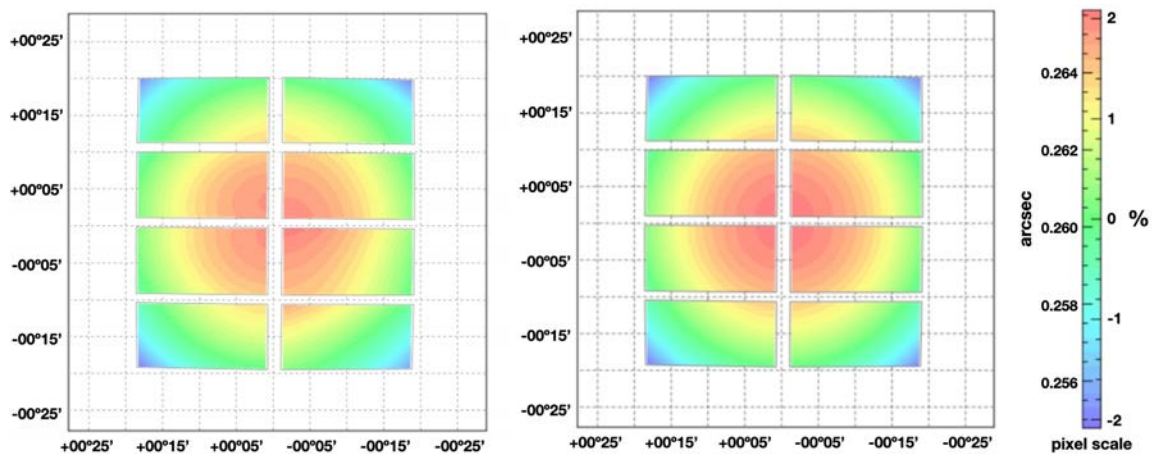


Figure 3.35: This figure shows the distortion map found by scamp with in a single-epoch exposure (left) and with 100 overlapping exposures (right).

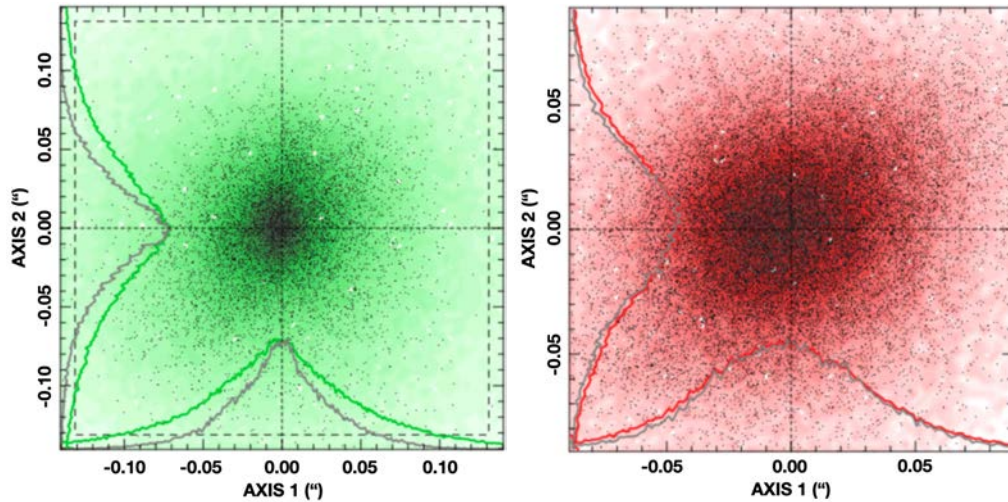


Figure 3.36: This figure shows astrometric residuals after the astrometry solution between internal overlapping sources (left) and to the reference catalogue (right).

The raw mosaic exposures come with a base World Coordinate System (WCS) (Calabretta, Greisen, 2002) that approaches the plate scale of PAUCam at the WHT. However this default WCS is not enough to calibrate single-epoch exposures and we need to compute more accurate WCS solution for the focal plane geometry. Furthermore, the solution for each detector position, rotation and scale to be defined independently as the base WCS comes from the mechanical layouts and not from precise measurements of the built focal plane. For this purpose we use SCAMP (Bertin, 2006) in a particular configuration, allowing for high order polynomial solution. We feed to SCAMP all overlapping exposures in cosmos for each filter tray to compute a precise solution with enough stars across the focal plane that allows to determine the position, rotation and scale of the entire mosaic of detectors. Figure 3.35 show the measured errors in both axis coordinates when calibrating the WCS with 100 overlapping exposures with respect to the reference Gaia catalogue and internally to PAU. From this validation analysis we estimate an absolute astrometric accuracy of ~ 40 milli-arcseconds (mas) RMS and an internal consistency of ~ 30 mas RMS.

For the last and most important element of the astrometry correction, we use again SCAMP as the main software to compute the astrometric solutions. It provides an updated WCS header for the FITS file that includes the offset and distortion correction under a 3 degree polynomial fit. Even SCAMP works best with overlapping exposures, we perform the astrometry solution on a single-epoch basis, as it already delivers 50 mas RMS of astrometric accuracy, below from what it is required to perform forced photometry (< 100 mas). However, a first calibration of the focal plane was performed with 100 overlapping exposures, that serves as a first guess reference for the single-epoch corrections. Even there are different filters over each CCD in the detector array, the global mosaic solution performs well and no chromatic distortions were remained from

SCAMP's solutions.

3.4.4 Photometric calibration

3.4.4.1 Concept

In a photometric night, the extinction correlates linearly with airmass. Traditionally the calibration of astronomical images has been performed taking observations of photometric standards (eg., Landolt) observed at different values of the airmass to compute the extinction coefficient. After the extinction model of the night was computed, it could be applied to the rest of the science exposures.

This approach works well for photometric data. For instance, SDSS with its dedicated telescope managed to observe only under good photometric nights. While DES observed in all conditions, bad weather nights were left for surveys that did not require accurate absolute photometry, such as their supernova (SN) survey. In the PAU Survey, with limited observation time, we had to design a method that could work under any observable condition, allowing to cross-calibrate with reference data when auto-calibration was not possible. That would take the most from the available observing time in the over-subscribed William Herschel Telescope.

Additionally, the fluxes measured with PAU cannot be reconstructed as a linear combination of the fluxes measured with the SDSS broad bands, as the narrow band set contains more spectral information. Without a method that allows to directly calibrate PAU from broad band data, a more sophisticated method had to be designed that allow to transfer the information from the reference SDSS calibration. Fortunately the PAUS fields fully overlap with well calibrated broad band data from the SDSS.

The approach we present here is to infer stellar templates from the SDSS broad band data and compute synthetic narrow band photometry from the stellar templates. These synthetic photometry is then compared to the PAU measurements of the same stars to obtain a zero point (ZP) for each star. Even though the stellar templates inferred from the SDSS broad bands for some particular stars could be wrong, a proper statistical analysis delivers accurate photometric zeropoint for each detector in the narrow band data when several stars can be used.

3.4.4.2 SDSS data

The Sloan Digital Sky Survey (SDSS) is a large survey observing most of the northern sky with imaging and spectroscopy. The survey has been conducted for many years in a dedicated 2.5m telescope at Apache Point Observatory in New Mexico. It is known to have very accurate and uniform photometric calibration down to 1% thanks to their continuous drift-scanning observing method and an exquisite absolute Übercalibration (Padmanabhan et al., 2008) and therefore it serves as a good reference for PAU. Ideally we would use SDSS spectra to calibrate the PAU narrow band images. However, there are not enough SDSS stellar spectra to ensure

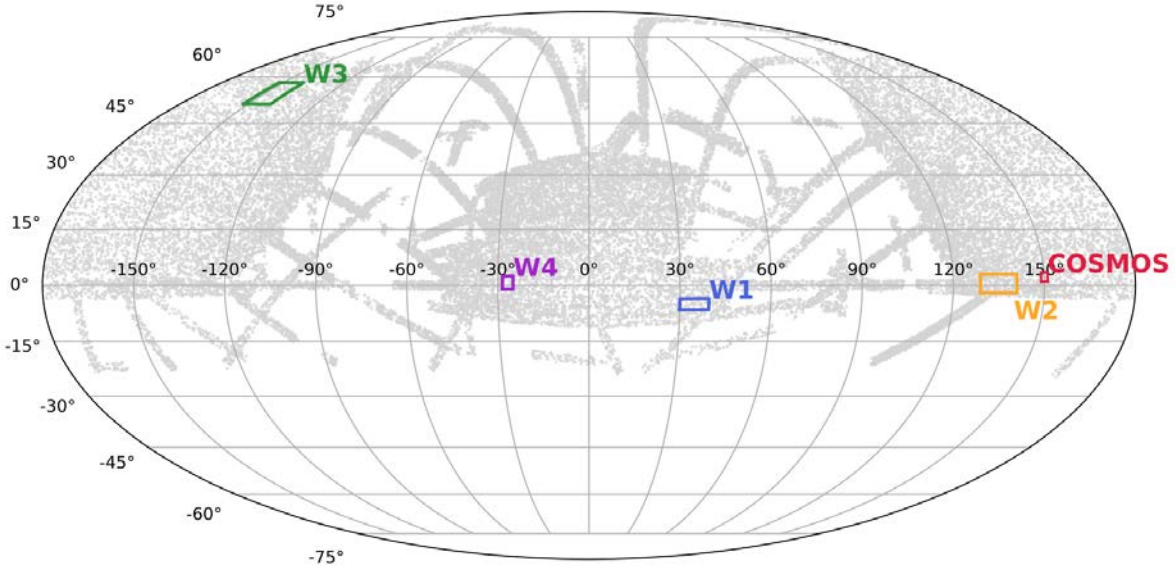


Figure 3.37: The SDSS DR12 star sample used as reference for the photometric calibration of PAU. This figure only displays around 40.000 stars brighter than $i_{AB} 10$, from the total of 260 million set. As seen from the figure, all PAU Survey plots are within the SDSS footprint and have enough stars to perform the calibration analysis.

calibrating every single PAU exposure image in all the fields with them. Therefore we use the stellar photometric sample, composed by ~ 260 million stars covering 14.500 deg^2 of the north sky. The overlap between the PAU Survey fields and the SDSS DR12 stars is shown in Figure 3.37.

We downloaded a local copy of the Star view from SDSS which essentially contains the photometric parameters (no redshifts or spectroscopic parameters) for all primary point-like objects, including quasars. From the long list of parameters available, we use only the PSF magnitudes (psfMag) in their ugriz broad band set. PSF measurements provide the highest SN on point-like objects such as the stars we are aiming to use. Additionally there is a *clean* parameter flag that allows us to filter possible photometric problems on their side.

This sample is selected based on star-galaxy separation techniques mostly based in morphology profiles. Therefore, as stated by SDSS, this sample contains quasars and we need to make sure these very distant galaxies do not enter into our calibration process. The most effective procedure is to apply a color cut that clearly differentiates the QSOs from the main stellar population. Based on the work of Richards et al. (2001), we propose the following cut to secure the rejection of most QSOs, without entering significantly in the stellar locus:

$$(3.13) \quad (g - r) > 1.25 - 1.12(u - g)$$

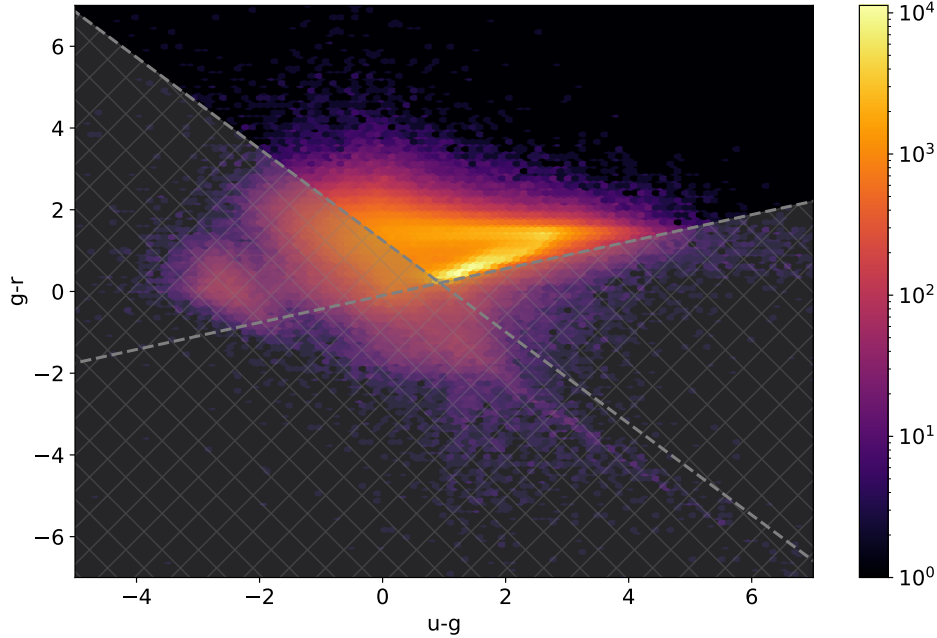


Figure 3.38: The color-color diagram of the Star catalogue in SDSS DR12. This catalogue is made of point-like objects and therefore QSO are also included. The proposed cut (marked with grey crosses) indicates the most likely QSO sources that we reject from the sample. This eliminates 25% of the sources.

$$(3.14) \quad (g-r) > -0.10 + 0.33(u-g)$$

As a result, 25% of the sample is rejected, leaving a cleaner star set for the calibration process. In Figure 3.38 the color-color diagram of the SDSS Star sample is shown, with the QSO cut we applied.

3.4.4.3 Stellar libraries

A stellar library is required in the proposed calibration process of PAU to transform the 5 points in the spectrum of the SDSS broad band set, to a high-resolution spectra that then can be downsampled to the low-resolution 40 narrow band set. Consequently, we need a realistic stellar sample that properly represents the whole stellar locus, providing good matches to all color combinations from the ugriz set.

Initially we chose the Pickles library (Pickles, 1998), a set of 131 flux-calibrated spectra encompassing all normal spectral types and luminosity classes at solar abundance, and metal-weak and metal-rich F-K dwarf and G-K giant components. The spectral coverage extends from 1150 to 10620 Å for all sources (and beyond for half of them) which perfectly overlaps with the

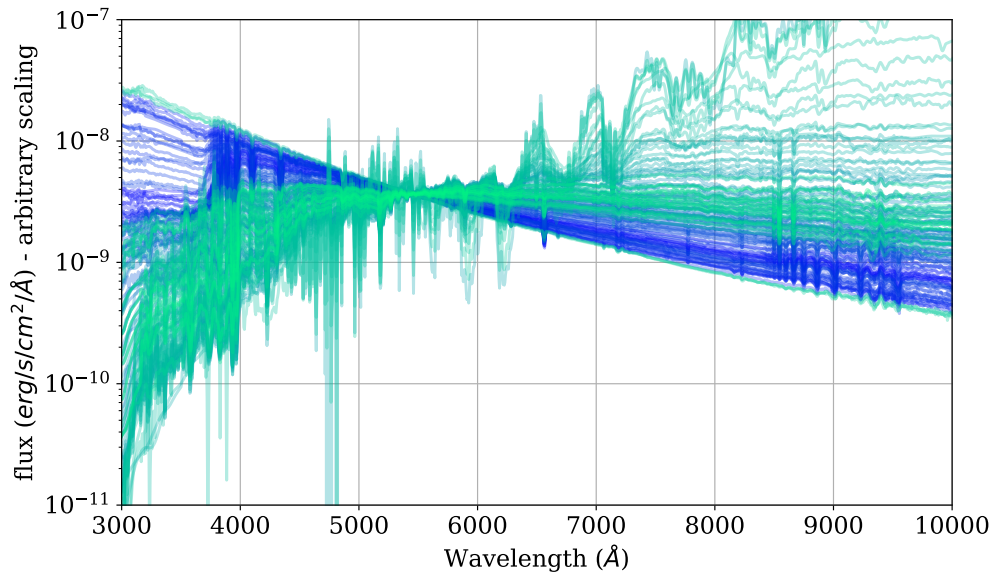


Figure 3.39: The 131 stellar template set from Pickles (1998) in the wavelength range of interest from 3000 Å to 1 μm

PAU narrow band data set (4500 to 8500 Å). The set of Stellar Pickles Templates can be seen in Figure 3.39

Lately, new spectral libraries have been released, such as the The X-shooter Spectral Library (XSL) (Gonneau et al., 2020). This library presents advantages with respect to the Pickles previously used. As shown in Figure 3.40 the XSL library covers a wider variety of spectral types, sampling more accurately the the colour-colour plot with its 666 templates. Moreover the spectrum of XSL was obtained as a single observation through the whole wavelength range, while the Pickles had to merge optical and near-infrared of distinct samples of stars, which could result in some inconsistencies.

3.4.4.4 Synthetic template fluxes

Now that we have chosen a stellar template library, we need to compute the synthetic fluxes that result from the combination of each stellar template and the transmission of all narrow band filters. As this is a computationally intensive task to be done for every single star and template we observe in our field, we pre-compute the synthetic fluxes from all templates with all narrow band filter transmissions. For every calibration star we will have its pre-computed 40 fluxes all stellar template (131 in the case of Pickles and 666 in the case of XSL) and will only have to scale the flux from the a reference flux in SDSS. Therefore we also integrate SDSS filters to get their synthetic flux so the scaling is direct. /fjccI find this paragraph unclear.

Additionally, the Milky Way extinction needs to be taken into account as it modifies the

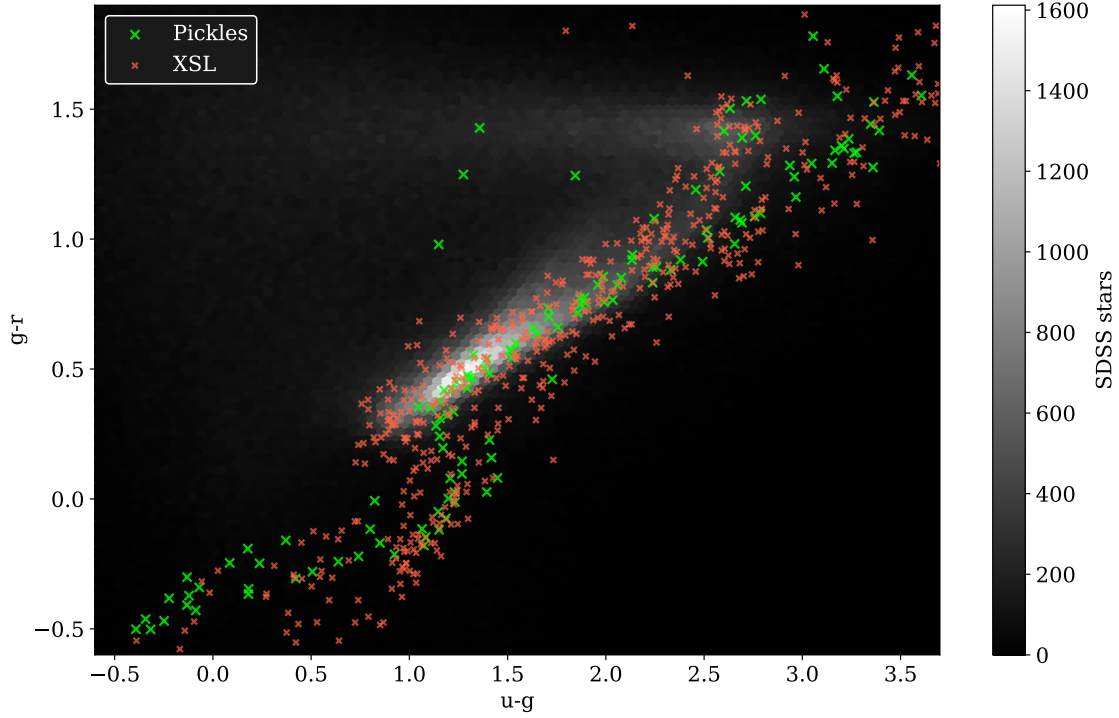


Figure 3.40: The colour-colour plot of the SDSS stars (grey distribution) with the Pickles (green) and XSL (red) template libraries. An average Galactic extinction was added to the templates to allow a direct comparison with the extinguished magnitudes of SDSS.

color of the spectrum altering the shape of the integrated fluxes. Consequently we also compute the synthetic fluxes at a range of extinctions following an O’Donnell (1994) extinction model (O’Donnell, 1994) with an R_v of 3.1.

As the template units are defined in f_λ with units of $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$ we can compute the integrated average flux density of the template at the specific passband such as:

$$(3.15) \quad F(t, R_i, A_v) = \frac{\int f_{\lambda_t} R_i(\lambda) 10^{-0.4 * A(\lambda)} \lambda^2 d\lambda}{c \int R_i(\lambda) d\lambda}$$

where t is the stellar template, R_i is the filter transmission of reference and A is the extinction in magnitudes from the O’Donnell94 model.

For the XSL set of 666 stellar templates, 40 PAU narrow bands + 5 SDSS broad bands and 10 extinction points, we end up with ~ 300 thousand pre-computed template synthetic fluxes.

3.4.4.5 Fitting SDSS fluxes to the templates

The next step is to fit the fluxes of a particular star in SDSS to all the stellar templates. We load the star PSF fluxes and its associated errors σ_s for the 5 ugriz bands. From the sky coordinates of

the star we can obtain its extinction from a reddening dust map and pre-computed the synthetic template fluxes at that particular extinction.

Now with all synthetic template fluxes f_t at the correct extinction, we define the χ^2 fit from the SDSS fluxes f_s to the arbitrary template fluxes as:

$$(3.16) \quad \chi^2 = \sum \frac{(f_s - \alpha f_t)^2}{\sigma_s^2}$$

with the scaling factor α that minimizes the χ^2 :

$$(3.17) \quad \alpha = \frac{\sum f_s f_t \frac{1}{\sigma_s^2}}{\sum f_t^2 \frac{1}{\sigma_s^2}}$$

and its associated error σ_α :

$$(3.18) \quad \sigma_\alpha = \frac{1}{\sqrt{\sum \frac{f_t}{\sigma_s^2}}}$$

To summarize, the fitting process described above returns for the combination of each star and all the stellar templates a scaling factor α , its corresponding error σ_α and the χ^2 fit.

3.4.4.6 Star and Image Zeropoints

Once the star (with its broad band SDSS flux) is fitted to all templates we can compute what is the star flux at each PAU narrow band. To estimate what is the most likely spectrum of that star, we select the template with the best (lowest) χ^2 fit. We can estimate the narrow band flux of that particular star $\langle f_{star} \rangle$ by multiply the scaling factor α to the pre-computed narrow band fluxes:

$$(3.19) \quad \langle f_{star} \rangle = \alpha_{best} f_{best}$$

where f_{best} will be the flux model with the best χ^2 at the narrow band we want to estimate.

However selecting the "best" χ^2 presents a statistical problem. If matches between SDSS stars and the template set would be perfect, the "best" solution would deliver the most accurate solution. However, there are no perfect matches as a star could lie anywhere in the color space and templates have discrete positions. Therefore it is statistically better to take into account all the templates weighted by their goodness of fit. Otherwise the "best" template can be arbitrarily selected by a random process. Therefore it is a safer approach to enter in the computation of the scaling factor all the template models and weight the average by the probability of each χ^2 fit.

The probability of a template fit can be defined as:

$$(3.20) \quad P_t = \frac{e^{-\chi_t^2/2}}{\sum_i e^{-\chi_i^2/2}}$$

Thus, the estimate of the expected flux at a given band is:

$$(3.21) \quad \langle f_{star} \rangle = \sum_t P_t \alpha_t f_t$$

and the estimated flux error:

$$(3.22) \quad \sigma_{\langle f_{star} \rangle}^2 = \sum_t P_t \sigma_{\alpha_t}^2 f_t$$

Now the estimated flux takes into account all possible template fits and is more robust to uncertain situations where not a clear best fit is defined. Each star measured in a PAU narrow band image $f_{measured}$ that matches SDSS provides a calibration zeropoint that can be defined as:

$$(3.23) \quad ZP_{star} = \frac{\langle f_{star} \rangle}{f_{measured}}$$

$$(3.24) \quad \sigma_{ZP_{star}} = \frac{\sigma_{\langle f_{star} \rangle}}{f_{measured}}$$

It is possible that the fitting process detailed previously fails due to limited information in the 5 SDSS band or outliers in the observed SDSS or PAU data. Fortunately, there are multiple SDSS stars in each PAU image and we can reject outliers by averaging multiple ZP_{star} .

The combined image zeropoint is simply the mean average:

$$(3.25) \quad ZP_{image} = \sum ZP_{star} / N$$

The image zeropoint error did not represent properly the error when propagating the error from the individual star zeropoints and a bootstrap procedure provided a more accurate measurement of the image zeropoint uncertainty. It is defined as the difference in the q-th quantile of the following values:

$$(3.26) \quad \sigma_{ZP_{image}} = 0.5 (q_{ZP_{star}} [0.84] - q_{ZP_{star}} [0.16])$$

where $q_{ZP_{star}}$ denotes the quantile associated with the distribution of ZP_{star} .

3.4.4.7 Source extraction and integration

The photometric calibration code is implemented inside the nightly pipeline. It has two main steps: the star photometry and the zeropoint calibration previously explained. For the first step, we run Source Extractor over the instrumentally detrended and astrometrically calibrated images. Our target calibration stars are moderately bright, comprising magnitudes between 14 and 19, that typically deliver a SNR higher than 20. For such bright stars we do not need to optimize the aperture with complex and PSF dependent methods that could be sensitive to the observation

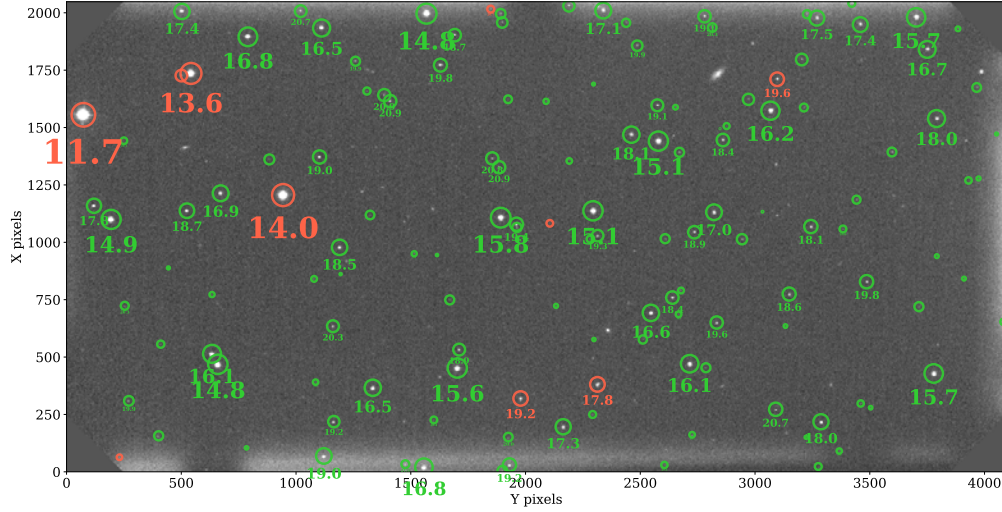


Figure 3.41: A NB665 image with stars that matched between PAU and SDSS. The number in the image corresponds to the i_{AB} magnitude measured by SDSS. The distribution starts from very bright stars of 12 magnitude all the way to magnitude 21. Green circles indicates "clean" SDSS photometry while red ones indicate that photometry is marked as "not clean". Very bright stars are therefore not used in the calculation of the photometric zeropoints.

conditions or optical distortions in the focal plane. Instead we aim for a constant large aperture ($\sim 4''$ radius) that gathers all light from the star independently from its PSF, ensuring more than 99% of flux in the worst conditions tolerated by the survey. We tested various configurations of aperture sizes, background modelling and scatter light correction, and the method described previously resulted to be the most reliable across the different observing conditions. Once the photometry is processed, we perform a spatial matching with SDSS DR12 (SDSS Collaboration, 2015), as only those stars will be of interest for the photometric calibration of the narrow band images. An example of a narrow band image of PAU with the calibration stars from SDSS can be found in Figure 3.41.

In the second step where the photometric calibration is computed, we provide as input the PAU star photometry in the image to be calibrated that matches with SDSS, the associated SDSS broad band photometry and the pre-computed template fluxes. With it the calibration code will return a star zeropoint ZP_{star} , zeropoint error $\sigma_{ZP_{star}}$ and best template χ^2_{best} fit for each star in the image, as well as the combined Image Zeropoint ZP_{image} and its bootstrapped zeropoint error $\sigma_{ZP_{image}}$. All values are stored in the PAUdm database for later processing and analysis. As PAUdm works with fluxes in all its measurements, the Zeropoints are all defined as a multiplicative ratio rather than in magnitudes which is commonly done in astronomy.

3.4.4.8 Validation with SDSS spectra

The photometric calibration is a key step in the reduction of the PAU data, and therefore we need to validate that the technique proposed previously is actually correcting the pixel data and at which level. Unfortunately there are no other surveys with the same broad band system to directly compare the fluxes. The alternative method is to check the calibration performance by comparing our calibrated fluxes with synthetic narrow band photometry from observed spectra that overlap with our images.

SDSS has large programs observing optical spectra and therefore there is significant overlap with their spectroscopy. In their Baryon Oscillation Spectroscopic Survey (BOSS) galaxies, quasars, and stellar spectra were collected, measuring a total of 1.5 million redshifts. Additionally, there is the Sloan Extension for Galactic Understanding and Exploration (SEGUE) measured $\sim 350,000$ spectra and their derived stellar parameters. Because it is simpler to model and to obtain the photometry, we begin the validation with stars from both programs.

Synthetic narrow band photometry In order to compare a spectra with PAU narrow band photometry, we need to generate synthetic photometry from the hi-resolution spectra. This provides a high quality reference to compare with any passband, specially valuable with the less common PAU narrow bands. This comparison will only be possible with objects that are observed by any spectroscopic survey and by PAU.

The first step in the process of generating the synthetic bands is to retrieve and homogenize the spectral data. In our case we have converted all fluxes to a more common f_λ with units of $erg/cm^2/s/\text{\AA}$. Generally, the data set contains the wavelengths where the spectra is sampled, the fluxes, the noise (or inverse variance) and a mask. Optionally SDSS also includes a measurement of the sky, that allows to identify possible contamination in strong emission or absorption lines. On a second stage, we interpolate the bandpass response $R(\lambda)$ to the sampling of the spectral data, as these two do not necessarily are in the same space. Later we mask both the spectral fluxes and the passband with the flagged measurements from the spectra mask. At this point we can compute the integrated average flux density of the source at the specific passband in $erg/cm^2/s/Hz$ such as:

$$(3.27) \quad \langle F_\nu \rangle = \int \frac{f_\lambda R(\lambda) \lambda^2}{c} d\lambda$$

and the its associated integrated response:

$$(3.28) \quad R_i = \int R(\lambda) d\lambda$$

Finally we can compute the synthetic magnitudes in the AB system with the following transformation:

$$(3.29) \quad m_{syn} = -2.5 \log \frac{\langle F_v \rangle}{R_i} - 48.6$$

It is also important for the statistical analysis to estimate the error of each synthetic band. As the flux in the spectra has been weighted by the response of the transmission, we must weight the noise in the spectra by the relative transmission throughout the entire passband.

$$(3.30) \quad \sigma_{f_v}^2 = \int \frac{R(\lambda)^2 \sigma_\lambda^2 \lambda^2}{c^2 R_i^2} d\lambda$$

where σ_λ is the noise in the high-resolution spectra. We can approximate the magnitude error such as:

$$(3.31) \quad \sigma_{m_{syn}} \approx 1.0857 \frac{\sigma_{f_v}^2}{\langle F_v \rangle / R_i}$$

Following the previous procedure we compute the photometry over all VIPERS spectra and all SDSS spectra that overlap with PAU over the 40 PAU narrow band set and the SDSS and CFHT broad band systems. We have flagged all measurements where the overlap between the systems response and the unmasked spectra is below 70%.

At this stage we can compute the synthetic narrow band photometry on the PAU filters with 170 SDSS spectra, serving as reference to $\sim 25,000$ PAU flux measurements.

Re-calibration of spectra To account for remaining aperture or PSF effects in the measured spectra we use total broad band (BB) photometry in the corresponding reference survey (SDSS or VIPERS) to re-calibrate each individual spectra. To do this we first estimate synthetic broad bands from the spectra, $F_S(BB)$, as shown in previous section. We then use the BB measured flux $F_O(BB)$ to find a multiplicative zero point, ZP , which is in general different for each BB:

$$(3.32) \quad ZP(BB) = \frac{F_O(BB)}{F_S(BB)}$$

We use ZP to re-scale each individual spectra. In the cases where we have 2 (or 3) BB measurements fully within the spectra wavelength coverage we combine them using a fit to a linear (or quadratic) function $ZP = ZP(\lambda)$, where λ is the mean of the bandpass response $R(\lambda)$. Each synthetic NB λ_{NB} from the spectra is re-scale by $ZP_{NB} = ZP(\lambda_{NB})$. Fig.3.43 shows the histogram of values of ZP_{NB} for all 40 NBs in 25644 independent measurements of 170 different SDSS calibration stars. The mean re-calibration is only 2% with a 4% scatter. Fig.3.44 show two typical example SDSS spectra.

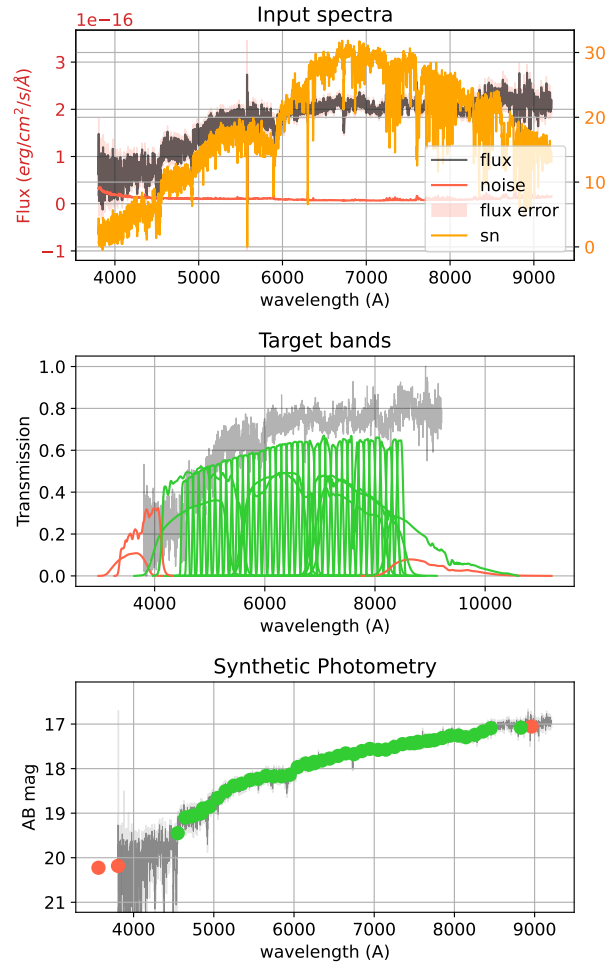


Figure 3.42: A sample $z=0.14$ galaxy spectra from SDSS used as input for synthetic photometry. The target bands are the 40 narrow band set from PAU plus the two broad band systems from SDSS and CFHT. Finally the computed synthetic photometry from the hi-resolution spectra. The bands without enough unmasked samples from the spectra are marked in red. More samples of synthetic spectra from SDSS and VIPERS can be found in Appendix A

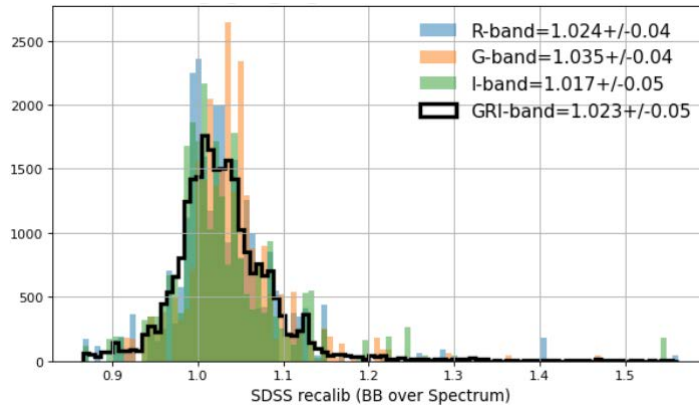


Figure 3.43: Color histograms show the values of the SDSS zero point spectra re-calibration ZP in Eq.3.32 for different BB as shown in the color labels. The black histogram combines G,R and I with a quadratic function of λ to find ZP for each synthetic NB. The labels show the mean and σ_{68} statistics.

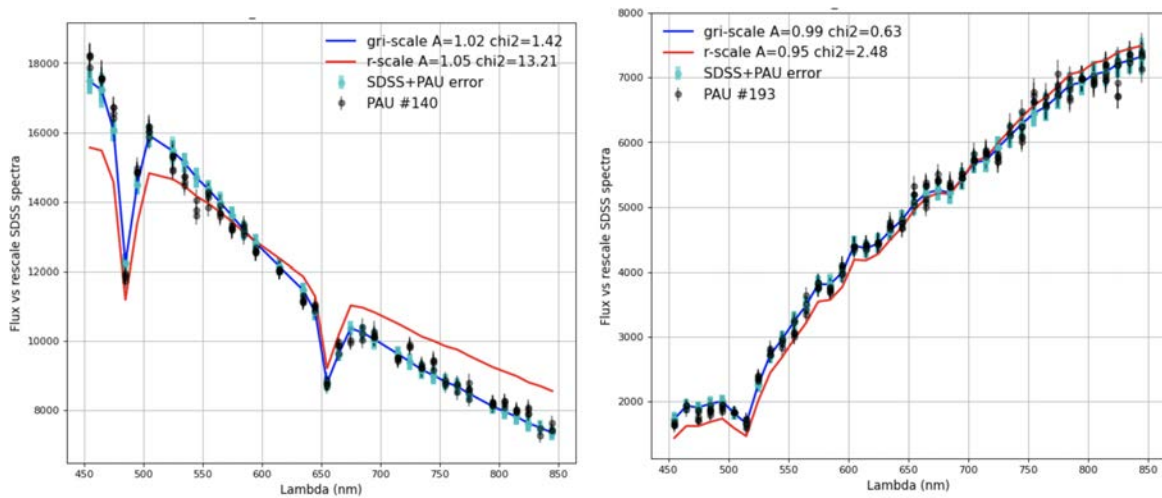


Figure 3.44: Example SDSS star calibration spectra for the validation study in PAU (points with errorbars). Red lines correspond to the raw SDSS synthetic narrow band photometry re-scale by a constant value ZP in Eq.3.32 to match the SDSS broad bands photometry of the same star in the R band. The blue lines uses a parabolic fit to the three broad bands (G , R and I) which fully overlap with PAUS. In some cases (as shown in left panel) this corrects for a color term in the spectra continuum. The spectra is also shifted by a global aperture calibration A given in Eq.3.33 (values shown in the labels), to account for possible differences in the aperture used in each observation. The "chi2" label shows the normalized χ^2 as compared to PAUS data.

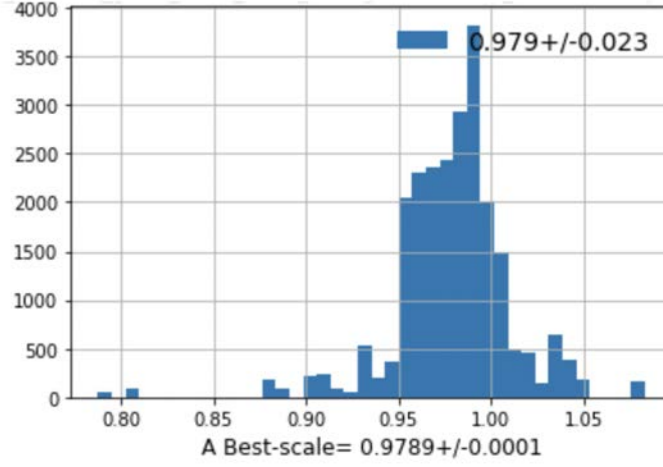


Figure 3.45: Aperture corrections A relative to PAUS data for $N = 194$ independent SDSS star calibration spectra and 42420 independent PAUS measurements of run id #955 in COSMOS. There is a very good overall agreement. The label show the mean and σ_{68} statistics.

Aperture corrections: global amplitude Once the SDSS spectra are re-calibrated with Eq.3.32, we also perform aperture correction of the amplitude of each individual spectra to the PAUS measurements. This is a fit to a linear constant A

$$(3.33) \quad A(i) = \frac{\sum f_{PAUS}(i) f_{SDSS}(i)}{\sum f_{SDSS}^2(i)}$$

between PAUS raw fluxes f_{PAUS} and SDSS re-scaled synthetic spectral f_{SDSS} (including the spectral recalibration). The sum is over all individual PAUS measurements in a given spectra (i) and it uses inverse variance weighting $w = 1/\sigma^2$, where σ is the join error added in quadrature. Typically there are 200 PAUS independent measurements (40 NB times 5 exposures) for each SDSS spectra. Fig.3.44 show two typical example SDSS spectra.

Fig.3.45 shows the distribution of values of A for $N = 194$ different SDSS star calibration spectra and 42420 independent measurements for PAUS run #955 in COSMOS. We find a mean value and scatter of $A = 0.98 \pm 0.02$, which indicates that PAUS data is in very good agreement overall with the SDSS calibration within 2% overall scatter.

Color terms We now check for any residual differences as a function of NB wavelength λ . Figure 3.46 shows the mean and scatter zero point difference for each NB λ :

$$(3.34) \quad ZP(\lambda) = \frac{\sum f_{PAUS}(\lambda) f_{SDSS}(\lambda)}{\sum f_{SDSS}^2(\lambda)}$$

between PAUS raw fluxes f_{PAUS} and SDSS re-scaled synthetic spectral f_{SDSS} (including the aperture correction A in Eq.3.33). The sum is over all individual PAUS measurements (42420 in

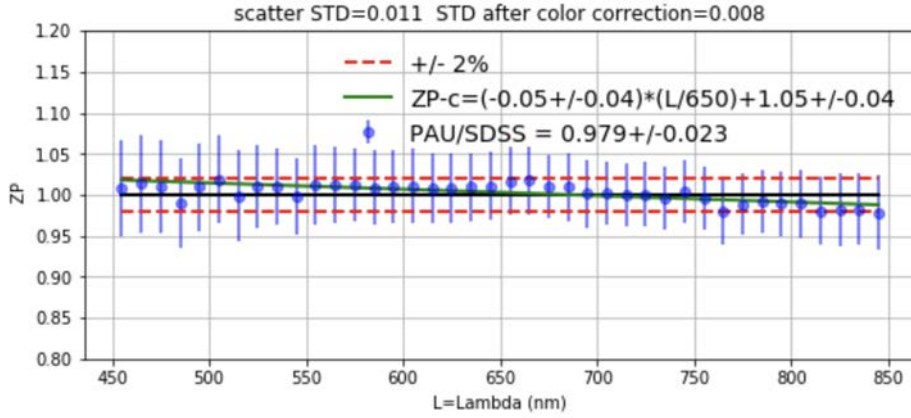


Figure 3.46: A validation study of the photometric calibration in PAU using synthetic narrow band photometry from SDSS spectra. The result demonstrates that the PAU absolute flux is above the 97% to the flux from the spectra and the band to band photometry is accurate down to 0.9%.

total) and uses inverse variance weighting $w = 1/\sigma^2$, where σ is the join error added in quadrature. There is a very good agreement overall as a function of λ . The typical scatter between the 40 NB around unity is only 1.1%. We fit for a linear color term and find:

$$(3.35) \quad ZP(\lambda) = 1.05 \pm 0.04 - (0.05 \pm 0.04) \left(\frac{\lambda}{650nm} \right)$$

which is consistent with unity within errors. The scatter between the 40 bands after correcting for this residual slope is just 0.8%.

3.4.4.9 Validation with ESO Spectrophotometric standards

As explained in §3.3.2, we observe various spectrophotometric standard stars every night during the morning and evening twilights. These are observed directly with each narrow band filter and are meant to provide an absolute calibration value. With magnitudes around 10-12 as seen in Figure 3.19, we observe them with very short exposure times, $t_{exp} \sim 10s$. Due to this much shorter exposure than in our science exposures, and because most of these do not overlap with the SDSS footprint, we cannot perform a direct comparison with the photometric calibration method explained in this section. Even so, we can evaluate the average ZP corrections with these spectrophotometric standards and compare them to the values obtain in each observed frame with our direct calibration method.

Initially with the detrended and astrometrically calibrated images we perform aperture photometry at the location of the standard star increasing the radius of the aperture until the flux asymptotically converges. At that radius, we obtain the observed flux. In order to compare it with the expected flux from the spectrophotometric standard, we need to correct for the atmospheric

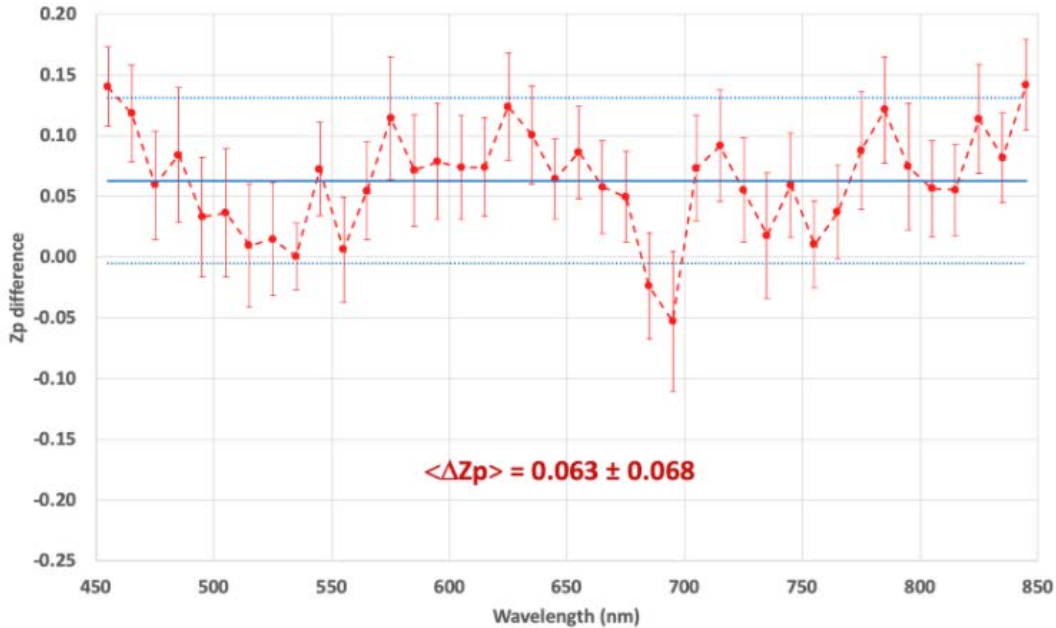


Figure 3.47: The difference average zeropoints from the main processing pipeline (PAUdm) to an independent method using spectrophotometric standards.

extinction. However, we cannot compute the atmospheric extinction coefficient as we only observe at one value of the airmass per twilight. Therefore, we rely on the values extracted from a model for the observatory (King, 1985). Once the PAU fluxes are extinction corrected, we compare them to the synthetic magnitudes calculated from the standard stars spectra and compute the zeropoints that make the observed extinction corrected fluxes agree with the synthetic ones. After analyzing 6498 images in all 40 narrow bands, we determine the median ZP at each band, removing the outliers with a Median Absolute Deviation (MAD) measure. The results shown in Figure 3.47 indicate that the average ZP in the different bands is equivalent to the one measured by PAUdm on the base procedure with templates with an offset of $0.063\Delta ZP \pm 0.068$. Even the individual standard star measurements present a noisy result due to its limited statistics, a fully independent procedure delivers a consistent solution to the PAU photometric calibration.

3.5 Photometry of narrow band images

In this section we explain how photometry of galaxies (and stars) we designed for the PAU Science and deliver scientific catalogues that optimize the redshift accuracy. The tools developed for this photometry are part of the Multi-Epoch and Multi-Band Analysis (MEMBA) pipeline of PAUdm.

3.5.1 Forced Photometry

The most important aspect that maximizes photometric redshift accuracy is preserving the colors of its measured bands. The narrow band images of PAU deliver low SN at the target magnitude of i_{AB} 22.5 and therefore we need to perform the photometry based on external reference catalogues. With the information on the reference catalogues, we can compute Forced Photometry over the PAU images, measuring the same amount of light in each band. This would not be possible with the PAU images themselves for the faintest objects as shape could not be properly estimated on sources with such low SN. From the reference catalogue we establish a consistent location, shape and scale of each object and define an aperture that preserves the flux fraction at all wavelengths. The only constraint for this technique is to have good astrometry accuracy as position of apertures over the sources will be defined blindly (no object detection or centroiding in the PAU images). As we defined in §3.4.3, we have a consistent astrometry at the sub-pixel level, enough for the purpose of Forced Photometry. As PAU images have different seeing and PSF sizes, it is also important to model the PSF such that apertures are scaled accordingly for a constant flux fraction. This process is described in §3.5.1.3.

3.5.1.1 Reference catalogues

Performing a forced photometry technique requires a reference catalogue that overlaps with the images observed by PAU. There are two main aspects that the reference catalogues must have: the necessary parameters to perform the forced photometry accurately and the complementary cosmological information that, in combination with the outstanding redshift accuracy of PAU, delivers a unique scientific spot. Moreover, the reference catalogue must be complete down to the magnitude limit of PAU such that the final combined catalogue has no target selection that could bias the cosmological measurements.

The selected fields targeted to run the survey are The Canada France Hawaii Lensing Survey (CFHTLenS) fields W1, W3 and W4, the Kilo-Degree Survey (KIDS) field G09 and COSMOS. In the CFHTLenS catalogue (Heymans et al., 2012) we have combined state of the art weak lensing estimates with THELI (Erben et al., 2013), shear measurement with lensfit (Miller et al., 2013), and photometric redshift measurement with PSF-matched photometry (Hildebrandt et al., 2012). In the case of the KIDS we use its latest release DR4 (Kuijken et al., 2019), also with outstanding cosmological lensing measurements (Kuijken et al., 2015). And finally, COSMOS where we built a merged catalogue from Laigle 2015 (Laigle et al., 2016) and the Zurich Structure & Morphology Catalog¹ for the accurate shape information. COSMOS is our main calibration sample and with so many multi-wavelength observations, it provided interesting >40 band photo-z tests.

Furthermore, all these catalogues contain the information to perform the forced photometry measurements such as sky coordinates, a star-galaxy classification, a reference i_{AB} magnitude,

¹Zurich COSMOS catalogue https://irsa.ipac.caltech.edu/data/COSMOS/gator_docs/cosmos_morph_zurich_colDescriptions.html

the scale of the source deconvolved from its observed PSF, the axis ratio to estimate its intrinsic ellipticity, the position angle and its sersic index or a parameter that allows to model it.

3.5.1.2 Background modelling

Accurate background subtraction is a key step to achieve precise photometry. The flux of the sources we need to measure stay on top of a floor of counts produced by the brightness of the sky, plus residuals of electronic bias and scattered light. The latter is particularly important due to the configuration of the filter trays in PAUCam, causing significantly more light to be reflected and scattered in the edges of the narrow band filter glass. An additional complication of the scattered light is that it can produce non-homogeneous background, much harder to estimate and subtract. Underestimating the background will have a stronger impact on faint sources with few counts, as it will add a bias that scales with the size of the aperture.

For all reasons stated above, a careful background estimation needed to be implemented. Amongst the different options, we designed an annulus around each source where we wanted to perform forced photometry. This provides an accurate estimate that takes into account large scale variations of the scattered light. The annulus had to be placed at a close distance to pick up smaller scale variations, but not too close to introduce flux from the source itself. Taking into account the typical size of sources we aim to measure, we set a fix limit for the inner annulus at 30 pixels from the center of the target source. To get enough pixel statistics the outer annulus was set at 45 pixels. All bad pixels that fall into the background annulus. A sample of an annulus with real pixels can be seen in Figure 3.48. Additionally, to avoid blending sources to affect the estimate, we perform a sigma clipping in the remaining pixels of the annulus, leaving only background free pixels. The median of the available pixels provides our estimate of the background in each source. This average value is multiplied to the area of the aperture and subtracted from the measurement in the main photometry process. The standard deviation and number of pixels used in the estimate is kept as it will be used in the flux error estimate.

An alternative Neural Network method to estimate and subtract the background of PAUCam narrow band images proposed in Cabayol et al. seems to deliver very accurate modelling of the background. It is not yet implemented in production and is planned to be integrated in the main pipeline for future releases.

3.5.1.3 Aperture scaling

We correct here for the impact of an elliptical PSF on a elliptical aperture for a galaxy. For simplicity, we present the results in circular coordinates, but this result is trivially extended to elliptical apertures by just using elliptical coordinates. The goal is to measure the fluxes at an elliptical aperture that corresponds to the same fraction of the total light after taking into account the effects of the PSF at the time and sky location when the image was taken.

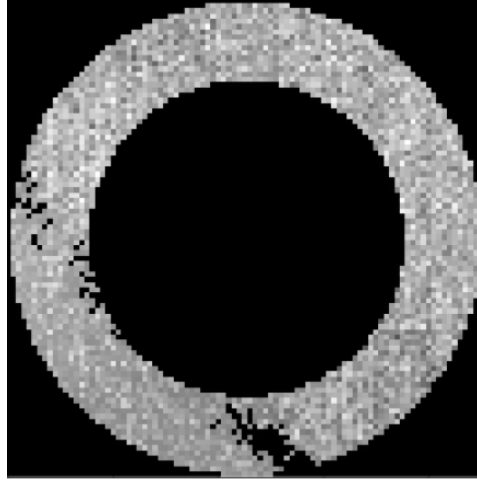


Figure 3.48: The annulus used to estimate the background of a particular source. It can be seen how pixels present in the image mask are discarded from the annulus and will not enter into the sigma clipping statistics.

Sersic Profiles and Aperture flux We will assume Sersic circular profile of slope n and scale r_0 for the surface brightness distribution:

$$(3.36) \quad I(r) = I(0) \exp[-(r/r_0)^{1/n}]$$

The total luminosity up to an aperture radius $r = A$ is:

$$(3.37) \quad L(A) = 2\pi \int_0^{r=A} dr r I(r)$$

For a Sersic profile, we can relate r_0 to the effective radius r_{50} , defined as the aperture which contains half the total light ($L(r_{50}) = L(\infty)/2$): $r_0 = [(0.86n - 0.142) \ln 10]^{-n} r_{50}$.

Convolved Profiles The observed surface brightness profile $I_o(r)$ will result from the convolution of the intrinsic $I(r)$ and the PSF kernel W_{PSF} . This is a 2D convolution, so even when the image is originally symmetric, $I(\vec{x}) = I(x)$, the convolved image I_o might not be symmetric:

$$(3.38) \quad I_o(\vec{r}) = \int_{Image} d\vec{x} I(x) W_{PSF}(\vec{r} - \vec{x})$$

For a circular PSF: $W_{PSF}(\vec{r} - \vec{x}) = W_{PSF}(|\vec{r} - \vec{x}|)$ we have:

$$(3.39) \quad I_o(r) = \int_0^\infty dx x I(x) \int_0^{2\pi} d\theta W_{PSF}(\sqrt{r^2 + x^2 - 2xr \cos \theta})$$

Moffat PSF We will use a radial Moffat PSF profile:

$$(3.40) \quad W_{PSF}(r) = \frac{\beta - 1}{\pi \alpha} \left[1 + (r/\alpha)^2 \right]^{-\beta}$$

which has $FWHM = 2\alpha \sqrt{2^{1/\beta} - 1}$ For a Gaussian PSF: $FWHM = 2.35482 \sigma$.

Effect of seeing on Aperture To obtain the same flux with different PSF we need to change the aperture to $r = A_o$ by solving:

$$(3.41) \quad L(A) = L_o(A_o) = 2\pi \int_0^{r=A_o} dr r I_o(r)$$

Stellar Apertures For a Gaussian $I_o(r)$ of width σ :

$$(3.42) \quad L(A) = \sqrt{2\pi} \sigma \left[1 - e^{-0.5A^2/\sigma^2} \right] = \sqrt{2\pi} \sigma \left[1 - e^{-4A_{FW}^2 \ln 2} \right]$$

where $A_{FW} \equiv \frac{A}{2\sigma\sqrt{2\ln 2}}$ in units of the FWHM. The fraction of light is then: $\frac{L(A)}{L(\infty)} = 1 - e^{-4A_{FW}^2 \ln 2}$ which for $A_{FW} = 1/2$ gives 50% of light (FWHM). For a Moffat profile the fraction is a bit smaller.

Implementation Computing the convolution integrals for the 2D Sersic profile with a Moffat PSF model for each measurement (120 per galaxy on average) is prohibitive as this operation may take several seconds when it is not hardly optimized. We take advantage of the very optimal implementation of the Half Light Radius calculation in Galsim (Rowe et al., 2015). Even Galsim claims this is only accurate down to a few percent, with these smooth profiles we achieve precision better than 1% in most cases. This is precise enough as the accuracy in our PSF modelling and the reference effective radius will be less accurate and will determine the final precision we can reach in estimating the aperture.

The sersic profile is generated with the effective radius of the reference catalogue and the sersic index estimated. In case the sersic index is not available in the reference catalogue, we assign an index of 4 for galaxies for elliptical types and 1 to galaxies classified as spiral.

The PSF is modelled as a moffat profile with a beta index of 4.75 and the FWHM measured in PSFEx for each detector image. The function `calculateHLR` is applied over the convolved object and it allows to compute either the HLR or a specific flux fraction. Typically we run the photometry over a variety of flux fractions, from 0.5 to 0.9. On average, the signal to noise is maximized around 0.625, even the optimal SN depends on each source. A example of how aperture size relates with flux fraction can be seen in Figure 3.49.

As the `r50` is defined as the effective radius on the major axis, we apply the same procedure for the minor axis, multiplying the `r50` by the axis ratio defined in the reference catalogue.

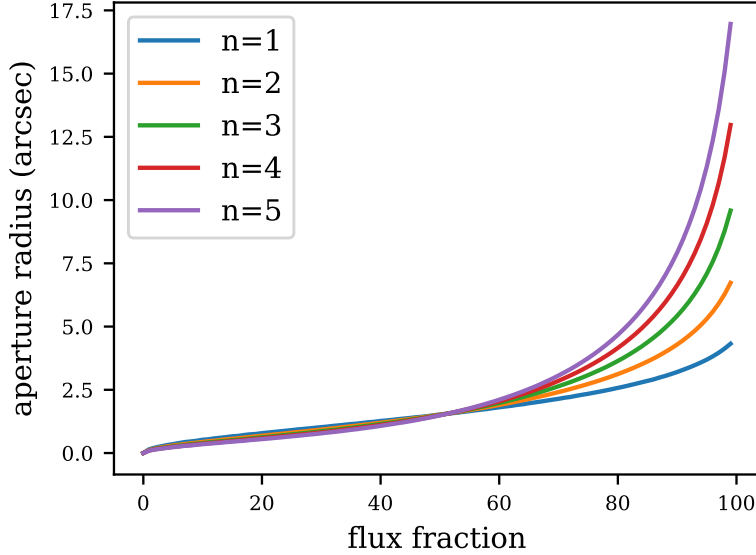


Figure 3.49: The aperture size radius as a function of flux fraction at various Sersic indexed. The sample source is a galaxy with an effective radius r_{50} of 1.5" and a Moffat PSF with a FWHM of 0.7" and a beta parameter of 4.75.

3.5.2 Flux measurement and error estimate

With the objective of measuring very low SN sources, we perform a forced photometry method where we blindly set the aperture in the image, without any centroiding or source detection process. As the broad-band reference catalogue is significantly deeper than the narrow-band PAU images, we can rely on good and complete positions and shapes from the reference catalogue. At this stage the narrow band images successfully passed all astrometry quality cuts and therefore, we can use the calibrated WCS solution to determine the pixel positions with sub-pixel precision of the source in the image.

After computing both major and minor aperture radius from the previous step, we can perform aperture photometry with elliptical apertures. For that we make use of PhotUtils library (Bradley et al., 2020), an affiliated package of Astropy (Astropy Collaboration, Astropy Contributors, 2018). The aperture flux F_{raw} is measured on the reduced image, still containing the component of the background. If the aperture contains flagged pixels in the mask, the flux measurement will be flagged containing the bit mask combination of all flags.

At the same location we estimate the background with an annulus as described in §3.5.1.2. The flagged pixels in the annulus do not enter into the combined flagging of the source, they are simply ignored in the statistics of the background calculation. The background modelling function returns the median background flux \bar{B} , standard deviation σ_{bg} and number of valid samples in the annulus N_{bg} .

The background subtracted flux of the source f_{src} is therefore

$$(3.43) \quad f_{src} = f_{tot} - \bar{B} * N_{ap}$$

where N_{ap} is the area of the aperture in pixels. Note the area is the exact fractional overlap of the aperture and each pixel.

The error in the flux is derived from the noise variance of a source in an aperture. As we apply the measurement on a single-epoch image, we assume there are no pixel-to-pixel correlations as it would appear on a remapped image.

We identify three components that contribute to the the flux uncertainty in this measurement. The source Poisson error, the noise in the background and the uncertainty in the estimation of the background.

The Poisson error of the source in the detector image is:

$$(3.44) \quad \sigma_{src} = \sqrt{g * F_{src}}$$

where g is the amplifier gain in the detector readout system and F_{src} is the raw detector counts of the source only. However the image is already in electrons, but has been divided by the exposure time t_{exp} to measure flux rate (e^-/s). Thus the shot noise error at the image units can be expressed as:

$$(3.45) \quad \sigma_{src} = \frac{\sqrt{f_{src} t_{exp}}}{t_{exp}} = \sqrt{\frac{f_{src}}{t_{exp}}}$$

Adding the independent background noise and background estimate error in the annulus we end up with a total combined error of:

$$(3.46) \quad \sigma_{tot} = \sqrt{\frac{f_{src}}{t_{exp}} + \left(N_{ap} + k \frac{N_{ap}^2}{N_{bg}} \right) \sigma_{bg}^2}$$

where k is estimated to be $\pi/2$ for a median average as we used in our background estimation method.

3.5.2.1 Flux co-addition

So far all measurements made happened at the single-epoch level. However the PAU survey observes each area in the sky multiple times for every filter. On average we perform 3 passes in the main fields (W1, W2 and W3) and 5 times in our calibration field (COSMOS - Laigle et al. (2016)). Observing the same area multiple times has some advantages, such as covering the

gaps between detectors, increasing the signal to noise, rejecting outliers, reducing the density of cosmic rays and increasing the dynamic range of the sources, as shorter exposure times will allow brighter stars not to saturate. This is done at the expense of an increased volume of data and a slower observing rate due to a constant readout time of 20s.

Most surveys combine their multiple layers at the image level, which is convenient for cosmic ray rejection (a median average almost completely removes all cosmic hits) but the stacking requires to resample the images, causing correlated noise which is complicated to model. Due to our objective of measuring very low signal to noise galaxies, we decided to stack the measurements at the catalogue level, performing all image measurements on individual exposures at their original pixel sampling. The combined measurements will be called coadded fluxes.

Before combining the single-epoch aperture measurements, they need to be corrected to a standard system so all fluxes are consistent. Light rate from the same source may vary due to particular observing conditions like variations in the atmospheric extinction on non-photometric nights, different telescope elevation resulting into different observing airmass or any other effect that varies the transmission with time. For this purpose we have calibrated each image and assigned a multiplicative factor ZP and it's corresponding calibration error σ_{zp} .

The calibrated single-epoch flux is simply defined as:

$$(3.47) \quad f_{calib} = f_{src} * ZP$$

and its calibrated error σ_{calib} , assuming non-linear error propagation with independent and not negligible variances, we derive:

$$(3.48) \quad \sigma_{calib} = \sqrt{\sigma_{src}^2 \sigma_{zp}^2 + \sigma_{src}^2 ZP^2 + f_{src}^2 \sigma_{zp}^2}$$

Now that we have a calibrated flux and its associated error for the individual measurements, we can proceed to combine all overlapping measurements into a coadd flux and error using an inverse-variance weighting method:

$$(3.49) \quad f_{coadd} = \frac{\sum f_{calib_i} / \sigma_{calib_i}^2}{\sum 1 / \sigma_{calib_i}^2}$$

where only non-flagged sources will enter into the statistics of each combined measurement.

Assuming that the overlapping measurements are independent we estimate the coadd error as:

$$(3.50) \quad \sigma_{coadd}^2 = \frac{1}{\sum 1 / \sigma_{calib_i}^2}$$

where N is the number of unflagged measurements to be combined.

Additionally we compute the reduced chi-square χ^2 as a measurement of consistency for the multiple measures:

$$(3.51) \quad \chi_{coadd}^2 = \sum_i^N \frac{(f_{calib_i} - f_{coadd})^2}{\sigma_{calib_i}^2} / (N - 1)$$

All three forced photometry coadd parameters are stored in the database for further processing and quality analysis.

3.5.2.2 Flagging

Throughout the whole processing of an image from its original raw state, we identify any possible cause that may affect the confidence of its value. To track each possible cause of problems we use flags both in the nightly processing and in the MEMBA pipeline. In the Nightly Pipeline which primary goal is to calibrate the images we track the flags at the pixel level. For that purpose we have created a mask image that is related with every single science image and contains the flag value in each pixel. In order to track all possible flag combinations in a single value, we have mapped each flag condition to a bit in the value of the pixel allowing for 16 different flags in a 16-bit depth image. In PAUdm we have defined the following image-type flags:

- **Cosmetics:** pixels not responding correctly to light, either hot ones that deliver constant high values or dead pixels that do not react to light inputs. Dust or imperfections in the detector mosaic or filter may appear here too.
- **Saturated:** pixels with so much flux that reached the ADC limit (18-bit in the case of PAUCam)
- **Cosmic Rays:** pixels identified as cosmic rays in the Laplacian filtering algorithm. Even pixel values will be interpolated from the neighbouring ones and cosmic rays may seem to have disappeared, the mask will keep track and the pixel value will not be use for science.
- **Vignetted:** areas in the focal plane with low transmission due to optical vignetting. The default value is set to 40%.
- **Crosstalk:** pixels contaminated by a strong signal of crosstalk from a related amplifier or detector (i.e. with mirrored saturated pixels).

On the contrary, in MEMBA pipeline we perform photometry of sources and flagging will take place at the catalogue level for each source measurement. The flags in the image that overlap the aperture will be propagated to the measurement flag and additionally we have defined the following catalogue-type flags:

- Edge: source too close to the edge or partially out of the image array.
- Distortion: source in an area with strong optical distortion or with an elongated PSF such that flux ratio in the aperture scaling may be inaccurate.
- Scattered light: source with intense and spatially dependent scattered light that could compromise the background subtraction.

In other areas of the processing where we make use of SExtractor (i.e. in the source extraction and photometry for the astrometry calibration) we make use of the standard SExtractor flags described in its software manual.

3.5.3 Survey mask

The particular layout of the narrow band filters in the camera trays and the fact that CCD detectors are separated leaving gaps in the focal plane results into a non homogeneous coverage of the sky for each pass band. Additionally there are telescope pointing errors that result into a more in-homogeneous sky coverage. In large cosmology surveys that intend to identify statistical correlations of galaxy positions and densities, it is mandatory to accurately identify how the survey has tiled the sky with its thousands of exposures.

For this purpose we have build a survey mask with two levels of information. First we generate the exposure mask, where we define for each filter how many times we observed each area in the sky for an entire field with a resolution of 5 arc-seconds. The mask is built taking into account variations in the system response from the flat field and flagged pixels. This will create a complex mask that introduces effects like vignetting, bright stars that saturate or corners in the detector not visible due to mechanical pieces in the optical path.

The second level is the bands mask and is build with the combination of all 40 exposure masks for each band. It represents the same area in the sky as the exposure masks but contains the number of bands available in each location of the sky with one or more effective units of exposure. A example of how both levels of mask can be seen in Figure 3.50.

The survey masks are associated with MEMBA runs, as the resulting coadd catalogue will be built with a set of images and this same set is the one that will enter into building the masks. As we are performing forced photometry from an external catalogue, the selection of sources will not depend on PAU observations and we will need to access to the survey mask provided by the external survey. However the final survey mask must be created as an intersect between the external mask and the PAU survey mask.

We used SWARP (Bertin et al., 2002), an Astromatic tool to remap and build image stacks. We feed SWARP with the science images and their corresponding flat field maps as the weight map. We obtain the exposure mask as the resulting combined weight map provided by SWARP in "MAP_WEIGHT" mode. As a by-product we obtain the science stacked image even this is not used as part of the main science processing.

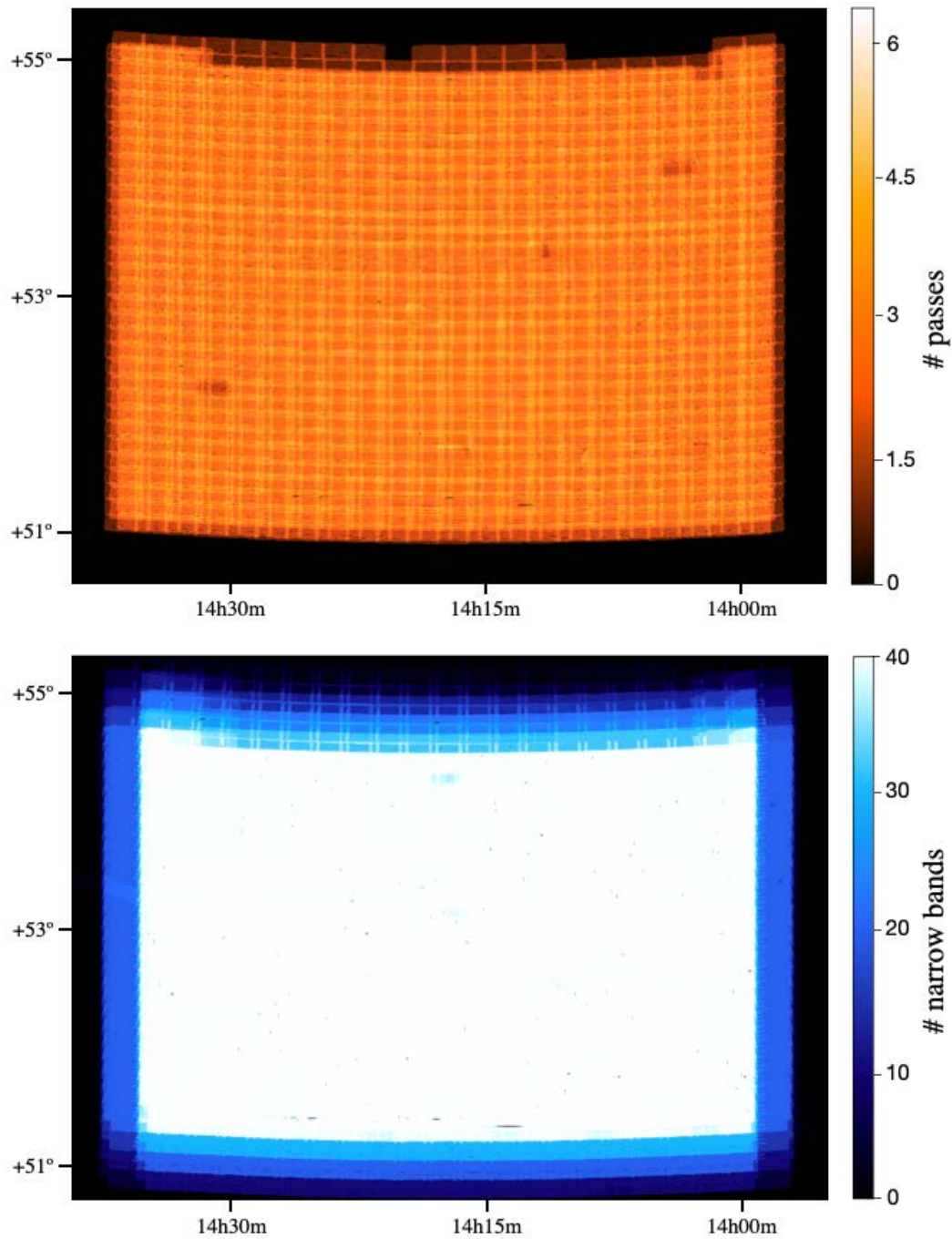


Figure 3.50: The PAU Survey mask in the 25 deg² W3 field. Top: The exposure mask of narrow band NB655 indicating the number of overlapping observations of this filter on the sky. Bottom: The bands mask indicating the number of narrow bands with at least one observation in each part of the sky.

3.5.4 Quality assurance & validation

One of the major challenges in PAUdm is the volume of data to be processed and analyzed. Tuning an algorithm to process a small dataset is simple as one can manually verify the correct behaviour of the processing and its output result. However with such large and complex dataset we need to build automatic control systems that verify that a particular code or configuration worked for the entire volume of data and raise an alarm or discard the data that did not meet some specific requirement.

3.5.4.1 Quality controls

At the time of writing this publication, the PAUdm system has processed more than 7 million images of PAU, and the number is increasing with additional reprocessing of data. The complexity and volume of this set requires an automated data quality control system to assure that the data products meet the expected requirements. Even the most imaginative and cautious developer will miss the variety of circumstances that data from an observatory can contain. From closed petals of the main mirror, to dust from the Sahara desert, to vapor condensation in the entry window of the camera in extreme weather conditions... these are some of the unpredictable conditions that we must catch to reject bad quality exposures and request observations to be repeated.

With this particular aim, we build a quality control system associated to processing jobs that verify the most relevant parameters that can be measured to reach a certain level of quality. In the nightly pipeline we defined the following quality control tests:

- Readnoise: check that electronic readnoise is under specification. This is measured in the overscan region of each amplifier. The default limit is set to $20e^-$.
- Flat field level: check that the flat field image is illuminated in the correct range of values. Too bright illumination could result into saturation and too faint illumination would increase the noise of the master flat field. The default range is set between 1000 and 120.000 ADUs.
- Saturation: check that the science images do not contain too many saturated pixels. A certain amount of saturated pixels is expected due to bright stars in the field but too many pixels is indicative of an issue in the exposure time, electronics or target selected. The default limit is 0.1% of saturated pixels.
- Cosmic rays: check that the cosmic ray detection algorithm does not classify too many cosmic ray pixels. Issues in the electronics or very noisy images may affect the sensitive CR detection algorithm and end up with too many pixels being classified as cosmics. The default limit is 1% of pixels.
- Astrometry: check that the contrast and χ^2 to the reference catalogue is good enough to ensure SCAMP found a reliable solution. This is critical as we rely on the single-epoch

astrometry and images with high extinction may end up with too few stars to deliver a solution. The default limits are greater than 3.0 in the XY contrast and below 50 for the reference χ^2 .

- Seeing: check that the average image PSF FWHM measured by PSFEx is below a certain value. Large PSFs reduce the signal to noise and limit the target sources of interest. The default limit is 1.8".
- Calibration stars: check that there are enough stars matched with SDSS to be used for the photometric calibration. The default limit is set to 5 stars.
- Zeropoint error: check that the estimated error in the photometric zeropoint is constrained. An unusually high ZP may be due to a non uniform response across the detector. The default limit is set to 0.2.

The quality controls processed in each job are piled up and propagated to the parent jobs so quality issues in large processing sets with many dependencies can be tracked easily.

3.5.4.2 Nightly report

Periods of observation typically last for 1 to 2 weeks. During this time it is very important to provide feedback to the astronomers in the observatory in the shortest time. This was one of the key constraints in the design phase of PAU data management and we achieved to process the whole night data set before the next night starts (~8 hours).

The nightly report is a web-based application that provides feedback to the astronomers on the quality of the data from the previous night so that observers can reschedule observations that did not meet a certain quality. It has also been critical to identify issues in the camera or telescope that were fixed with minimal delay for the remaining observing run. The nightly report has default quality limits necessary for the survey and generates a report file that can be ingested directly to the PAUCam control system for re-scheduling targets. The app also displays statistics and evolution plots for each night, with options to adjust the metrics to be analyzed and its time span, such as the ones in Figure 3.51. A total of 35 parameters can be displayed to help PAU astronomers understand the atmospheric, weather and instrument behaviour from any previous night.

In addition to the quality checks described in §3.5.4.1, the nightly report displays the processing status of the main blocks in the nightly pipeline: detrending, astrometry, PSF modelling and photometric calibration. The status in each block can be used as quality cut (i.e. repeating all observations where PSF modelling failed in any detector). An example of the nightly report status summary can be found in Figure 3.52.

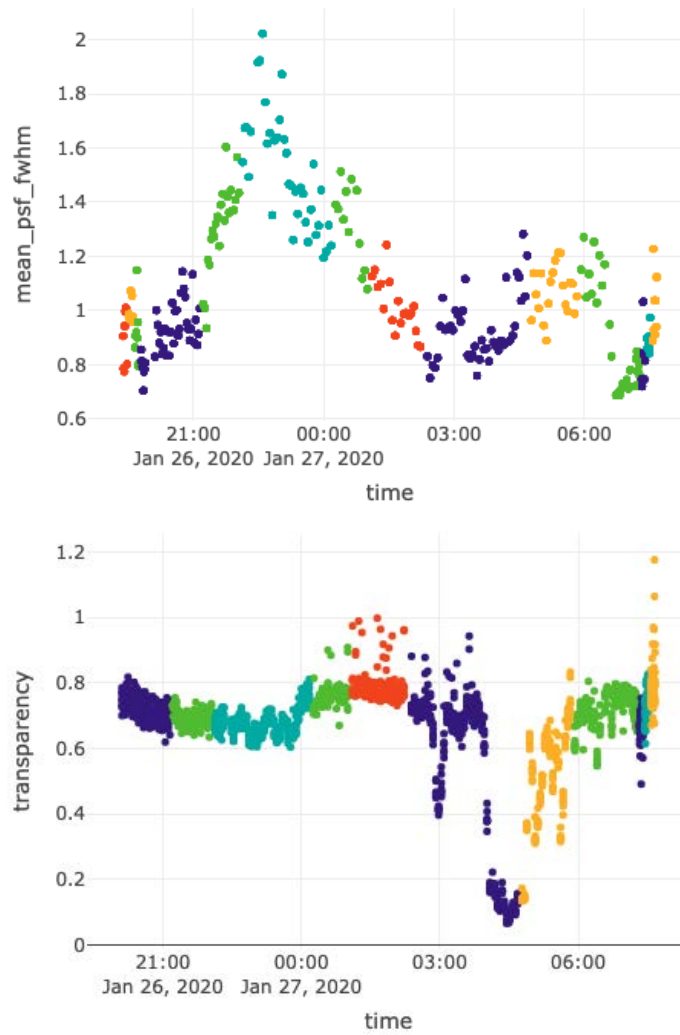


Figure 3.51: Some of the quality control plots that the nightly report produces as a result of the nightly pipeline processing. This data corresponds to one of the last nights of the PAU Survey before the COVID-19 pandemic break. Top: Evolution of the atmospheric seeing and how it got quite high before midnight, stabilizing low before the end of the night. Bottom: Evolution of the transparency, where some clouds entered at around 4am, reducing the absolute transparency to almost 10%.



Figure 3.52: The selection tool from the nightly report showing the status of the main 4 sections in the nightly pipeline. It allows to set limits specified by the survey to re-observe the targets that did not meet specific quality requirements.

3.5.4.3 Forced aperture inspector

With such a large narrow band set and with the additional overlapping exposures, each object depends on the correct reduction and calibration of hundreds of images. There are many things that can go wrong and even all quality tests passed in the processing of an image, the final low-resolution spectra for each source. It is critical to identify the sources of issues that may end up in outliers in PAU spectra as the photo-z code could translate that into a physical feature of the galaxy, resulting in a catastrophic redshift determination.

A very reliable reference to validate PAU measurements is high-resolution spectra from external surveys. The selection of the calibration fields was in fact driven by the overlap with datasets of carefully calibrated spectra. Our main calibration field COSMOS contains ~17.000 sources with good reference spectra. Additionally our main fields contain other datasets such as VIPERS (Le Fèvre et al., 2013) in W1 or DEEP2 (Newman et al., 2013) in W3, complementing the already extended set in COSMOS. From the calibrated spectra, we can infer the synthetic narrow band and perform a direct comparison to the PAU measurements [point to section of syth spectra or explain here]. Galaxy sources with spectra provide accurate spectroscopic redshift

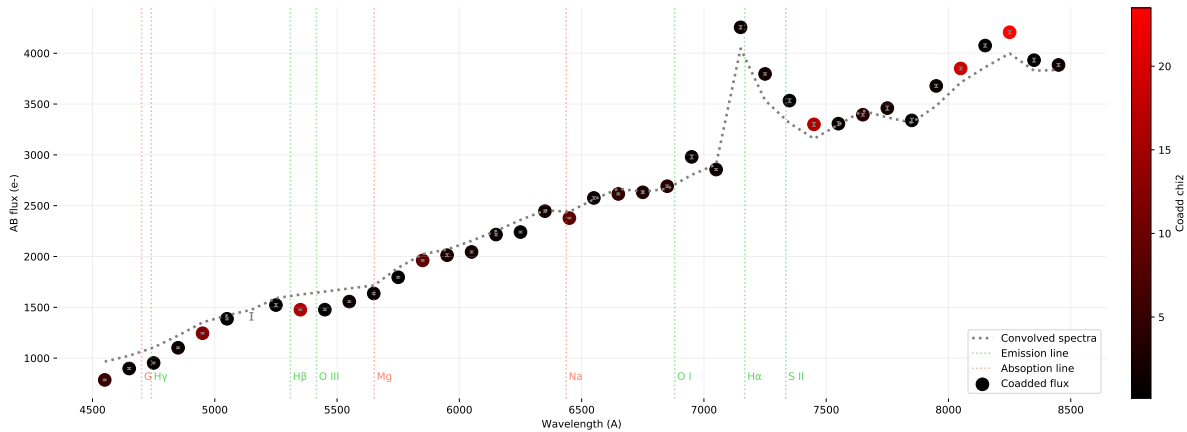


Figure 3.53: A bright galaxy at redshift 0.1 from the COSMOS field. The aperture inspector displays the coadd measurements from MEMBA and overlays a synthetic narrow band photometry from the corresponding SDSS hi-resolution spectra (when available). With the emission and absorption lines depicted, the redshift solution can be hinted, specially on emission-line galaxies such as the example above. The redder coadds indicate discrepancy between the single-epoch fluxes, suggesting possible issues in the combined measurement.

estimations which allow us to know where the expected emission and absorption lines are, providing additional confidence on a particular spectra shape.

We built another quality control web application where PAU measurements for a particular source are displayed together with the synthetic measurements from a reference spectra. When the redshift is available, we also display the emission and absorption lines in the expected wavelength position. This allows us to visually inspect PAU data and easily identify outliers or discrepancies with the spectra reference. The app user can click on a particular PAU band and it will show the measurements that contribute to that band. Clicking again on a single measurement will display the portion of that particular image with the aperture done in MEMBA and all quality parameters associated to the image or observing conditions. Very rapidly we can identify issues from the final PAU spectra to the original image that contributed to every measurement.

As this resulted a very powerful tool and many scientist used it, we added a reporting system where people inspecting the sources could specify the issue that caused a particular outlier. The list of possible issues grew up as we learned more and more about the PAU data, and we ended up with a list of 18 possible issues, such as scattered light, blended source, crosstalk, astrometry issue and more. Even this is a subjective test that required some training, we could extract valid statistics and correct for multiple systematics that originally caused trouble.

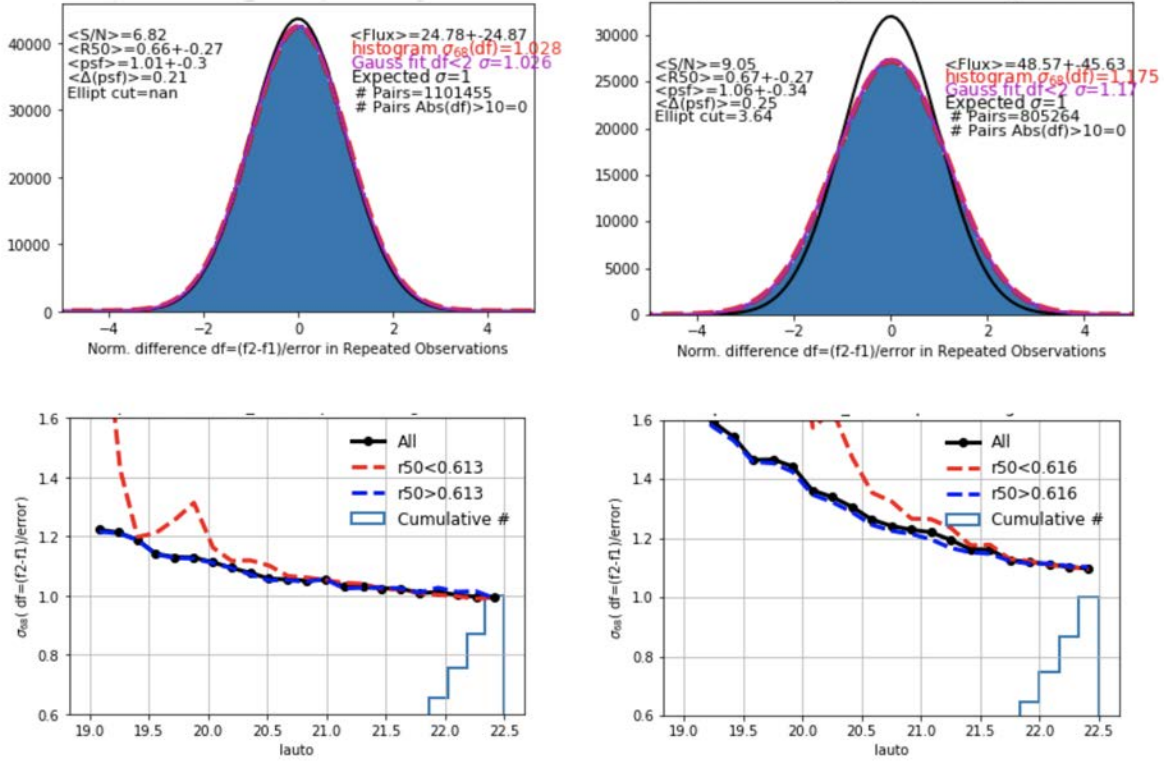


Figure 3.54: **TOP panels:** Duplicates for W3 galaxies (run 941) with NB=575nm (left) and NB=755nm (right). Histograms of values of $df \equiv (flux_2 - flux_1)/error$. There are $\approx 1.1E6$ pairs with mean S/N of 6.82 and mean flux of 24.78 for NB=575nm. **BOTTOM panels:** σ_{68} as a function of total broad band magnitude I_{auto} . Continuous line correspond to all duplicates. Dashed red line shows values with smaller half light radius $r_{50} < 0.616$ arcsec. The histogram below the line shows the relative number of duplicates which increases sharply at the faint end and dominate the statistics for a full population.

3.5.4.4 Duplicate observations test

Repeated exposures over the PAUS fields have been used to validate the final MEMBA photometry. We have between 3-10 independent fluxes for the same object in each NB band. We use this catalogs to build a sample of over a million pairs of duplicate (repeated) measurements of the same object. There are 45 pairs of measurements for the same object and about 40,000 separate objects, before cuts and masking. We identified duplicates as objects with same reference ID (which means they have same position in MEMBA force aperture). We select pairs with $S/N > 3$. The goal is to test if the errors in fluxes produced by MEMBA are consistent with repeated measurements for the same object as a function of different properties of the object and observation.

The top panels of Fig.3.54 show two examples of histogram in the values of normalized flux differences $df \equiv (flux_1 - flux_2)/error$ of duplicates. The error is just added in quadrature from the MEMBA error of $flux_1$ and $flux_2$. In general, the filled histograms of normalized differences

follow a normal distribution (shown by dashed red lines) but the width is typically a bit larger, $\sigma_{68} \simeq 1.028$ and $\sigma_{68} \simeq 1.175$ in these two cases, than unity (black line) which is what we would expect if MEMBA errors were perfectly accurate.

The bottom panels of Fig.3.54 show the width of the normalized duplicate distributions, σ_{68} , as a function of the total broad i-band of the reference catalog (W3 in CFHTLS) which we called I_{auto} . We can see that there is a strong dependence with I_{auto} which indicates that MEMBA errors are correct at the faint end but are underestimated at the bright end. Similar results are found for other PAUS fields.

We have seen similar tendencies in the photometry of duplicates for other broad band surveys. This happens for both galaxies or stars and using aperture photometry or also SExtractor total magnitudes (e.g. in the PAUS zero point calibration with SDSS stars). We interpret these results as a small bias in zero point calibration values, which have spatial variations which affect more strongly the brighter fluxes. This is not accounted for by error propagation of zero points to flux in Eq.3.48. A zero point bias of only 1% can produce up to 20% miss estimation in the error fluxes of bright galaxies, which have larger fluxes and smaller relative errors.

In Fig.3.55 we show σ_{68} for all other NB wavelengths λ in PAUS. The background green color intensity variations represents the relative density of pairs as a function of λ and σ_{68} . The variance is typically larger than $\sigma_{68} = 1$ which indicates that errors are underestimated by up to 20% in some NB. We find similar tendencies for all wavelengths. The yellow arrows point to NB575 and NB755 which correspond to the two extreme cases shown in more detailed in Fig.3.54.

3.5.4.5 Comparison with SDSS and VIPERS spectra

We use spectroscopic data from SDSS and VIPERS to predict the MEMBA flux measurements. We then compare the predictions with the actual measurements to test MEMBA and the previous data calibration steps.

Statistical validation analysis We perform a test similar to that presented in Fig.3.43-3.46 but using galaxies with SDSS and VIPERS spectra instead of just bright SDSS stars. The result for SDSS and VIPERS is shown in Fig.3.56. We can see larger errors than for the case of calibration stars because this is for fainter galaxies. But we can see that MEMBA works very well compare to SDSS. There is a small color term which is most likely related to the difference in the aperture of SDSS fibers and PAUS forced photometry en MEMBA. This term was not present for stars, because they are point objects but is expected for galaxies because they are extended and the spectra varies as a function of position in the galaxy. Similar results are found for VIPERS (bottom panel) which has smaller spectral overlap with PAUS. Note that VIPERS spectra seem to be extinction corrected while PAUS and SDSS are not. We have use the mean values in the VIPERS/SDSS mean spectra calibration to correct VIPERS from this before we compare to PAUS. Fig.3.57 shows two typical SED examples for galaxies at $z = 0.4$ and $z = 0.5$ with $i_{mag} \simeq 20$.

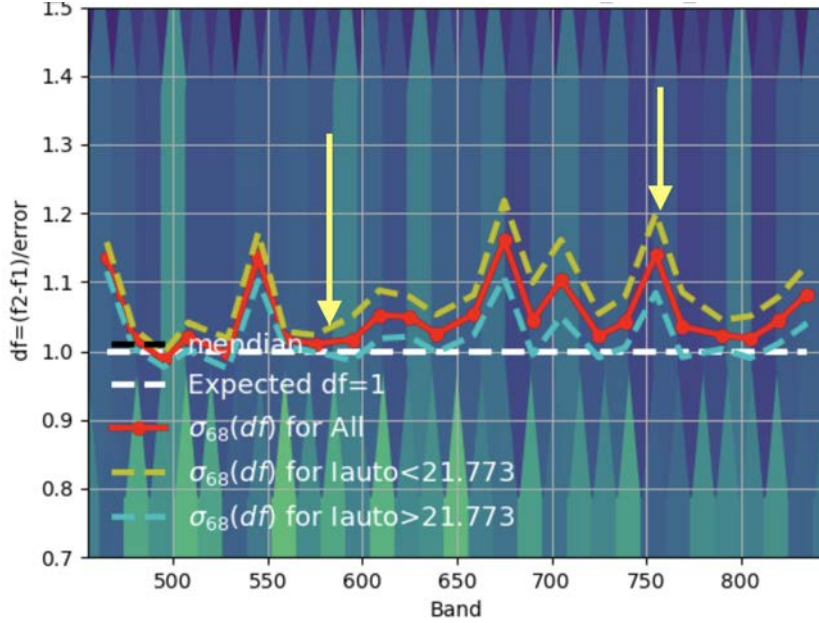


Figure 3.55: Normalized error (σ_{68} in Fig.3.54) in duplicate differences $df \equiv (flux_2 - flux_1) / error$ as a function of NB wavelength. Red continuous line corresponds to the value for all duplicates. Dashed yellow (blue) line shows the values for half of the sample with brighter (fainter) broad band reference I_{auto} magnitudes.

3.5.5 Known limitations & further work

Since the first light of PAUCam we have been constantly improving both Nightly and MEMBA algorithms, steadily better understanding the systematics and the instrument behaviour. Even we reached outstanding photometry and photo-z accuracy, there are still steps in the overall process that we know it can be improved. Recently we tested algorithms that seem to improve the overall survey performance. However in this paper we refer only to those algorithms that were implemented in the main pipeline and that delivered scientific results in already published papers. We plan to improve and present further algorithms with the corresponding upgrade in scientific results once these are stabilized and validated.

On the flux calibration step in the nightly pipeline, we know the current dome flat fields do not reproduce the sky illumination accurately. Therefore the detector response after flat-fielding calibration is not homogeneous. A possible solution is processing sky flats and leave the dome flats for small scale pixel variations and not to correct the large-scale vignetting and illumination patterns. This has the difficulty of the scattered light as an additive component which complicates the sky flat processing.

The background modelling implemented in MEMBA is a simple but reliable method. Due to the complex varying background in PAU images resulting from the flat field and scattered light residuals, a more complex background estimator that understands trends of the background can

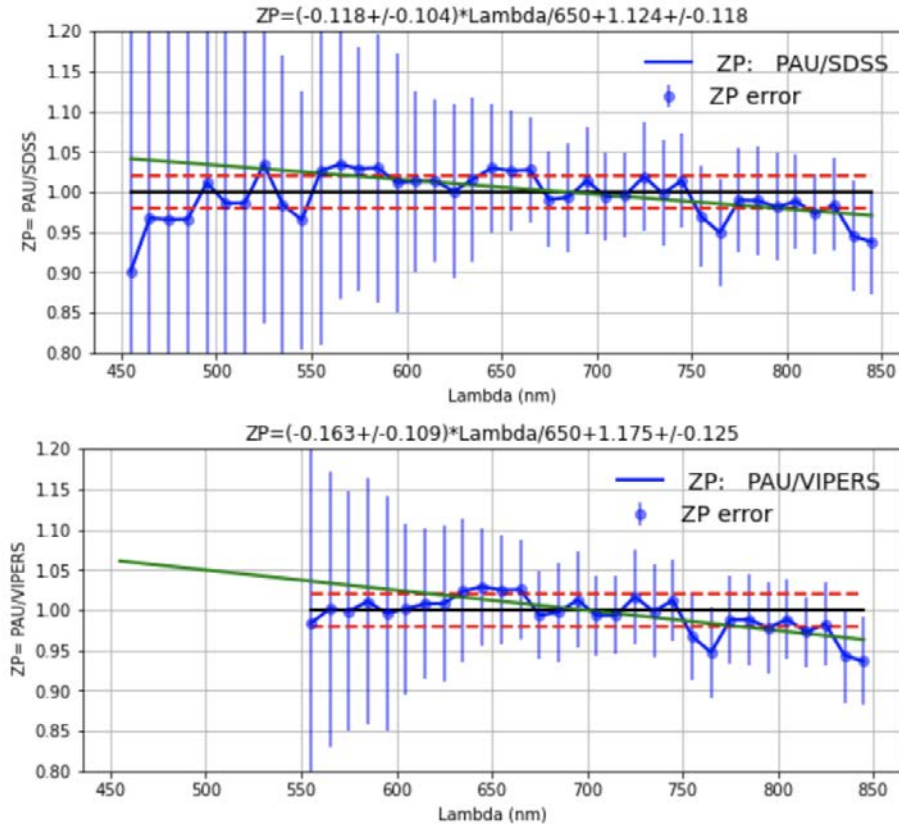


Figure 3.56: A validation study of the photometric calibration in PAU using synthetic narrow band photometry from SDSS (top) and VIPERS (bottom) galaxy spectra. This should be compared to Fig.3.46 which uses brighter stars.

provide significant benefits. This is specially important in low SN sources where small residuals of the background can bias the measurement. A machine learning technique has been studied and published in Cabayol-Garcia et al. (2020) and will soon enter into the main processing of PAUdm. There are also prototypes that provide flux estimates from neural network algorithms, improving the accuracy and increasing the signal to noise of the measurements.

The areas in the focal plane with most distortion are currently flagged. There is a possibility to increase the survey efficiency by accurately modelling the PSF at the edges of the focal plane and adapt the apertures to include a larger area of the unvignetted mosaic.

Modelling the PSF and computing the astrometric solution is done at the single-epoch level, independently for each exposure, for simplicity and because it delivers good enough precision. However more stable and accurate solutions can be obtained by computing astrometry and PSF models with multiple overlapping exposures, at the expense of complicating the nightly processing adding dependency between single-epoch reductions. The astrometric precision would only improve marginally at the sub-pixel level which could have some benefits for particular

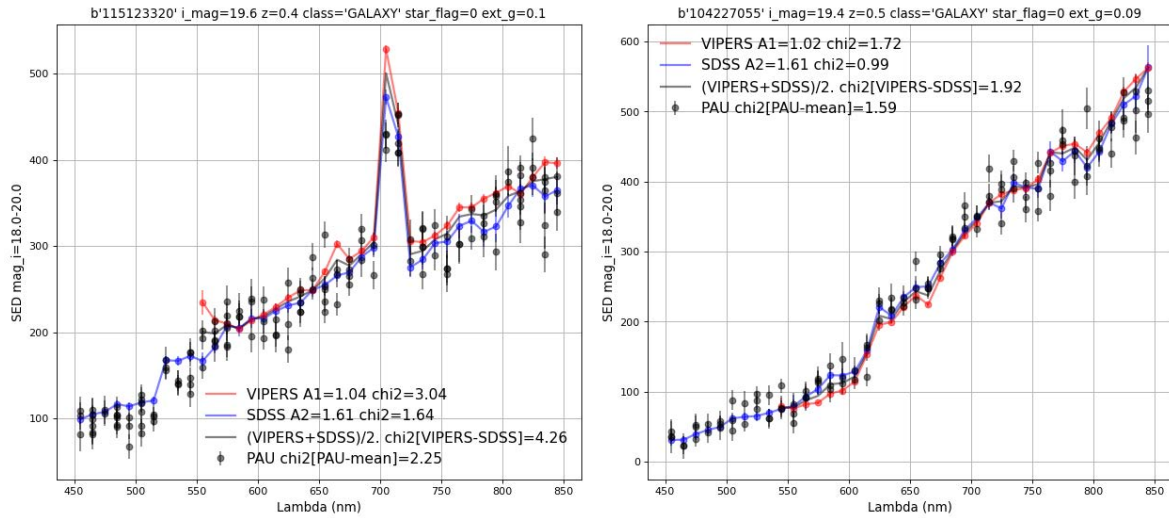


Figure 3.57: Example spectra of validation study of the photometric calibration in PAU (points with errorbars) with synthetic narrow band photometry from SDSS (blue) and VIPERS (red) galaxy spectra. SDSS (VIPERS) spectra have been multiply by A_1 (A_2), as shown in the labels, to account for possible differences in the aperture used in each observation. The "chi2" label shows the normalized χ^2 as compared to PAUS data.

applications. But the PSF modelling could provide a stable and varying model with very potential applications, contrary to the simple model we currently use.

Similarly to the previous point, the photometric calibration delivers ZP independently for each detector image. There are algorithms such as the Übercalibration that computes a global solution from the overlapping measurements between images, homogenizing the calibration and ensuring a flat response across wide area in the sky. However, as we calibrate against SDSS stars and those have been globally calibrated in a similar process, we would expect this improvement to be minor for PAU.

3.6 Operation and data orchestration

The PAU data management system has been designed to operate in the infrastructure at the Port d'Informació Científica (PIC). In this section we describe the infrastructure of operation, the technical aspects of the main scientific pipelines and the tools required to operate the pipeline under the available infrastructure. The solution presented below was not the original design as both the infrastructure and the project needs changed since the beginning of the operation. A more technical and infrastructure-oriented description of PAUdm is described in Tonello et al. (2019)

3.6.1 PIC data center

Port d'Informació Científica (PIC) is a High Throughput Computing (HTC) data center, serving as a Tier-1 facility of the Large Hadron Collider Computing Grid (WLCG) that also provides data services to Astrophysical and Cosmology surveys such as PAU, DES, DESI and Euclid. Additionally, large cosmology simulations that were computed in Super Computers are now stored at PIC such as MICE or Euclid Flagship enabling its scientific exploitation.

3.6.1.1 Infrastructure description

As an HTC center, the computer cluster of 8800 cores does not act as a super computer, but as independent processing elements capable of processing and moving large volumes of data. In PAU we have only a share of the total computing power, with a maximum of approximately 300 cores in parallel. How rapid the cores start to run our jobs depend on the load of the data center, mostly driven by LHC activity. The computing is orchestrated with a HTCondor framework, allowing to submit and monitor the processing jobs.

With respect to the storage capacity, PIC currently holds a massive 33PB in their tape system and 11PB of disk with 60 storage servers. PAU has a subset of the total archive with 400 TB of tape and a dedicated 200 TB disk server.

The network recently upgraded to 200Gbps connection, which enables faster data transfer from the observatory and sharing data with international collaborators. Additionally there are other useful services such as an online Jupyter Notebook server, allowing to quickly test and develop algorithms with direct access to the storage system.

3.6.1.2 PAU database

With such a large and complex dataset where millions of galaxies, measurements and images are related, it has been key to setup a relational database that tracks all the information and metadata of the survey. We have chosen a PostgreSQL database running on a powerful 12-core 96GB server in a twin configuration for reliability and performance. The database can be accessed by the pipelines via an object relational mapper (ORM) for better integration and reliability under the pipeline environment. It is also accessed by the different web applications such as the nightly report and the forced aperture inspector described in previous sections. As PAU is a large collaboration, we make available via the PAUdm website a dynamic view of the database to browse the schema and perform simple queries. Additionally, for development and validation purposes, it can be accessed via Python notebooks under a read-only role, dumping queries directly into dataframes with all the flexibility and potential that these objects provide.

The database model was designed to allow for reprocessing of data at any level, tracked by the production table. Careful constraints were set on each table to ensure unique entries under each production set. We defined 4 main pipelines: The Pixel Simulation, where we produce

survey and pixel image simulation for development and assessment of performance. The Nightly pipeline, already detailed in previous sections where the main image reduction and calibration happens. The nightly can process input productions from the pixel simulation or real observations from PAUCam. The MEMBA pipeline where we perform forced photometry over the nightly images. And finally Photo-z where we estimate photometric redshifts from MEMBA photometry. Each pipeline can be processed independently, allowing to process multiple times a given input production (i.e. different aperture photometry in MEMBA under various configurations with the same set of image reduction made by the Nightly pipeline). In addition to the data-related tables, we have a survey strategy database, synchronized cloned from the one at the observatory, containing all information about fields, exposure status and observation progress. These tables are used throughout the night to schedule targets that need to be observed by PAUCam. It is also necessary in the regular PAUdm processing to select the exposures associated to a survey field and that have been classified as valid exposures. Finally they have all data operation tables with information about job configuration, status and their dependencies. The use of these are described in §3.6.1.4. Additionally the whole database schema with the most relevant columns in each table can be found in Appendix B.

3.6.1.3 PAU archive

The PAU camera produces ~300GB of raw data per observing night. These are mostly FITS files that contain the exposure images and additional metadata in its header. The data is processed in the main pipelines where more sub-products are generated, multiplying the volumes of RAW data. The actual size of the RAW archive (until 20A observations) is 42 TB and it is safely archived as a two-copy tape and a third copy in disk. The processed data is significantly larger due to the increase bit depth and the various sub-products generated per exposure (science, mask, weights, PSF models...). Reduced data has a single copy in disk, except for published releases where we include an additional copy in tape. The raw archive tree organizes the data in observation sets, following the schema in the PAUCam temporary archive at the observatory. The reduced tree adds an additional layer to account for multiple reprocessing of the same data, that we call productions. The distribution of data archived in PAU can be seen in Figure 3.58.

Even most of the access to the archive system is made by the nodes in the computing farm, we wanted to make available both raw and reduced data to the PAU Collaboration. For that purpose we setup a WebDAV server that allows web access to the entire archive in a user friendly format.

3.6.1.4 Browthrower

The high volume of data and its complex analysis requires processing the pipelines in a data center with enough computing power such as PIC. With its High-Throughput Computing data center configuration we had to split each pipeline into smaller tasks than can be processed independently with limited consumption of memory and CPU time. Consequently, a pipeline may

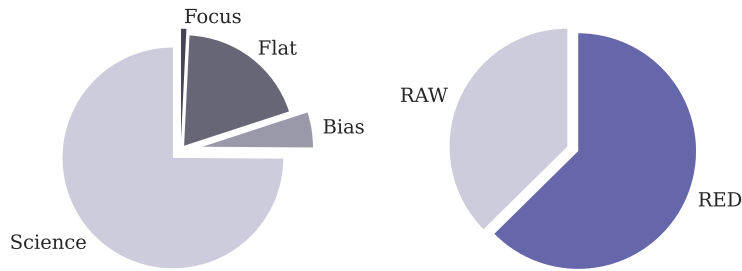


Figure 3.58: Distribution of storage for Raw and Reduced data until 20A. The Raw data is directly produced by PAUCam and sums a total of 42.87 TB and is stored in double tape copy plus an additional copy in disk for quick access. The Reduced data composed of reduced science, weights and masks, plus master bias and flat frames and surveys masks raising to a total of 71.52 TB.

result into hundreds or even thousands of jobs, with its configuration and inter-dependencies, that need to be launched, monitored and operated.

For that purpose we designed a job orchestration tool named Brownthrower (BT) that give us the flexibility to operate the pipelines at PIC. With BT we can create jobs, add dependencies between them so certain jobs do not begin processing until others are complete, share configuration between jobs and monitor the status of a pipeline and its sub jobs. To process these jobs we submit to the computing farm via HTCondor a set of pilot jobs that will be continuously grabbing free jobs to be processed (in status Queued and without pending dependencies) until all jobs are being processed.

3.6.2 Operation

With all the available infrastructure described previously, we can carry out the complex data management operations. Thanks to a close link between the operations team and the development of the astronomy pipelines, we have been evolving how we operate the data processing along the various years of the survey, which significantly changed since the beginning with simulations.

3.6.3 Data flow and data orchestration

The data flow in PAUdm is divided between main PAUdm archive, where large files are stored and the database, where metadata and information that requires complex selections is uploaded to the database. These balance is orchestrated by the different pipelines. In first place, the raw data is transferred from the observatory in La Palma to the archive set in Barcelona at PIC (as described in §3.6.3.1. Immediately, the exposure metadata is registered in the database. Next the Nightly Pipeline begins its image calibration and archives the clean images, its PSF models and the astrometric solutions in WCS. Photometric measurements and its calibration ZPs are uploaded to the database. Once enough sky area has been processed by the Nightly, MEMBA can begin the galaxy photometry and reingests its coadd catalogue once all measurements have been

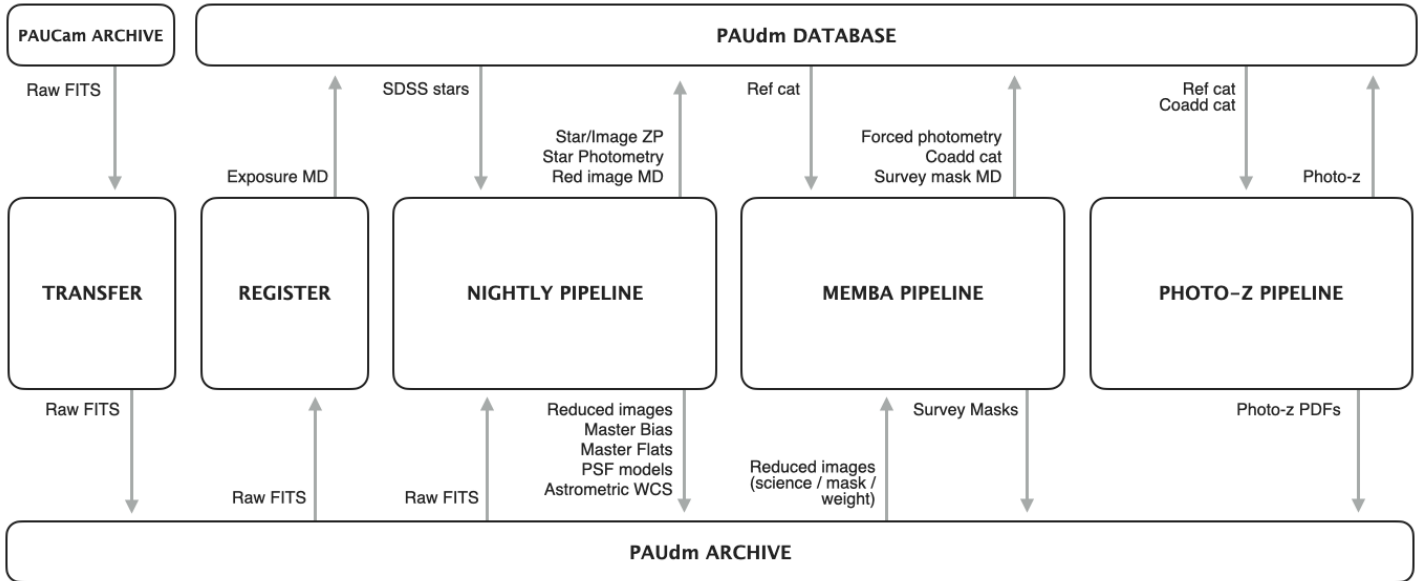


Figure 3.59: A simplified schema of the PAU data flow, where interactions between pipeline, storage and database are shown.

done. MEMBA is also in charge of producing survey masks and therefore these large files are properly stored in the archive. Finally, the Photo-z pipeline obtains MEMBA's measurements and computes the photo-z for each galaxy. The photo-z values and estimated errors enter the database while the large probability distribution files are archived in the storage. A summary of the data flow is depicted in Figure 3.59.

3.6.3.1 Transfer and register

The PAU camera produces large volumes of data, mostly from exposures stored Multi Extension FITS files of 600 MB each. A typical night of observation ~ 300 GB of data are generated. The camera control system contains a local storage that has capacity for approximately 5 nights of observation. Therefore with such a short-term storage, we need to perform the transfer of the data from La Palma to Barcelona in the shortest time possible and in a nightly basis as long as the bandwidth connection is not broken (thing that happened sometimes at the Observatorio del Roque de los Muchachos).

Once a block of observation is complete, a folder in the local archive of PAU is generated with an structure equivalent to the main archive at PIC. It contains all FITS files and an additional YAML file with information and checksums of the list of images in that particular Observation Set. At the end of the night, the observer triggers to archive the observation sets and these are moved atomically from the PAUCam local archive to a PAUdm server located at the WHT. Every morning, an automatic job at PIC triggers a *bcp* tool that synchronizes the archive at La Palma with the one at PIC. Once the transfer and checksum verification are complete, register

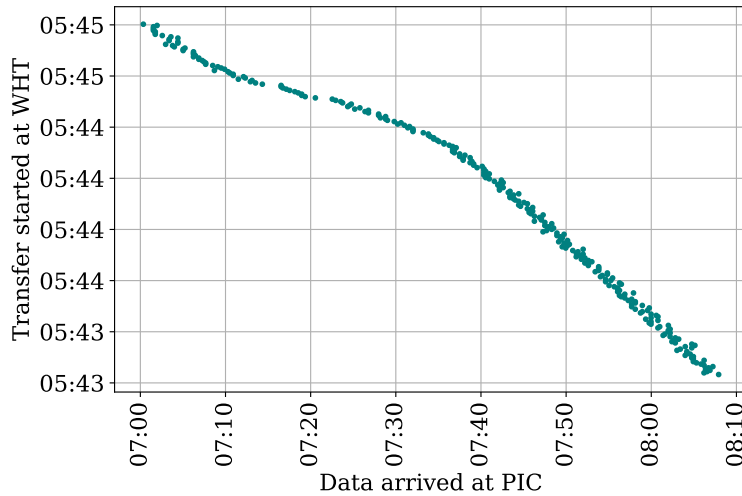


Figure 3.60: An example of a data transfer for one night of observation from 17A with 281 exposures. This plot indicates the time where the transmission started at the WHT server for each FITS file against the time where the whole transmission was completed and safely stored at PIC. The average transfer time for each file is less than 2 hours and the total transfer time was completed in about 2.5 hours

jobs are scheduled in the batch system to introduce the metadata of the raw exposures into the corresponding tables in the PAU database. An example of data transfer times for a night dataset is shown in Figure 3.60.

3.6.3.2 Nightly Pipeline

This is the pipeline in charge of the image calibration. It begins with a set of raw exposures, including flat and bias calibration images. It ends up with the science exposures astrometrically and photometrically calibrated, ready to perform the flux measurements.

The Nightly pipeline has two main steps. First we have the production of master bias and master flats and secondly the single-epoch reduction of science images. During observation periods, we process the pipeline in batches of observation sets. Typically an observation set is a single night but it allows to have more than one observation set per night in cases where observations come from different surveys, as PAUCam is also available in the WHT as a community instrument. When we operate in observing mode, the nightly pipeline tree starts with the master bias, then with the master flats of each filter tray and associated to each tray the corresponding sky images. This means that the image calibration won't start until the jobs of master bias and master flat are successfully completed. The pipeline tree can be seen in Figure 3.61.

There is a second mode to operate the nightly pipeline meant to process entire fields from multiple nights. This is used when we improve the nightly code and want to reprocess a given subset. First we process all masters from the nights we we can have exposures and at a second

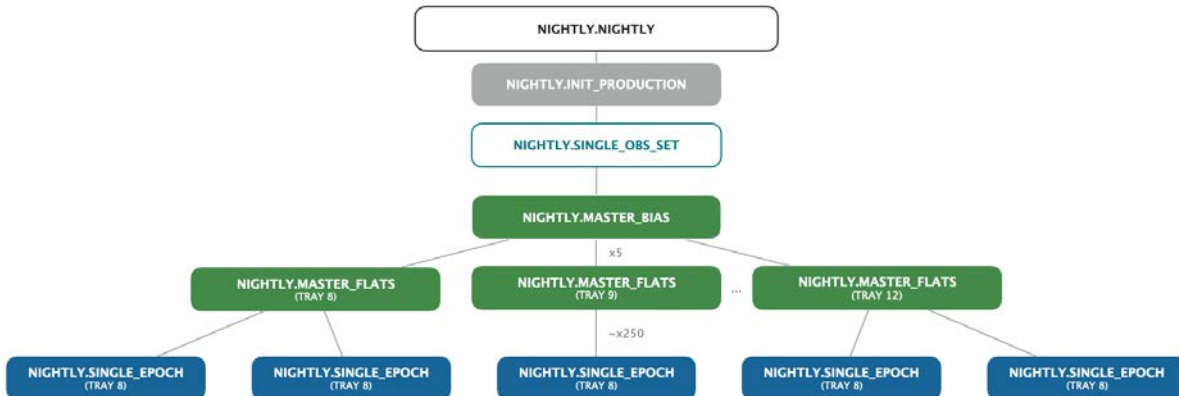


Figure 3.61: The dependency chart of the Nightly pipeline. Empty boxes define jobs creating sub jobs while filled boxes refer to jobs processing data at the computing farm. The dependency at the filter level starts with the master flat fields and the single epochs are related with its own masters.

stage we analyze all science images in parallel.

The pipeline significantly evolved since the beginning as we had to deal with heterogeneous data sets such as very cloudy skies, observation sets without calibration images, saturated flats, etc. A major effort had to be done to automatically detect any possible situation and either correct for it or classify the faulty data as invalid.

To allow for the 8-hour rapid feedback during an observation period, we launch 50 BT pilots to the computing farm that process all images in time. Reprocessing entire fields involve a much greater set of images and therefore we increase the number of pilots to 100. Master bias and master flat jobs are processed in 10 minutes while single-epoch exposure reduction can take up to 20 minutes per job.

The nightly pilots require intense i/o access to the archive system to retrieve raw data and ingest reduced images and for this reason we do not allow more than 100 nightly pilots to run in parallel.

3.6.3.3 MEMBA Pipeline

The Multi-Epoch and Multi-Band Analysis pipeline is intended to perform the photometry over the reduced images across the different bands and overlapping exposures from different epochs. It can be divided into three main steps: the forced photometry, the coadded catalogue and the production of survey masks.

Each forced photometry job takes care of running the photometry of a single detector image. It will load the reference catalogue, the corresponding reduced image and mask and will upload the measurements to the database once completed.

The coadd jobs are divided into areas in the sky. We select the areas by healpix pixels of N side 128. This is an efficient way to divide the sky and retrieve all overlapping measurements from

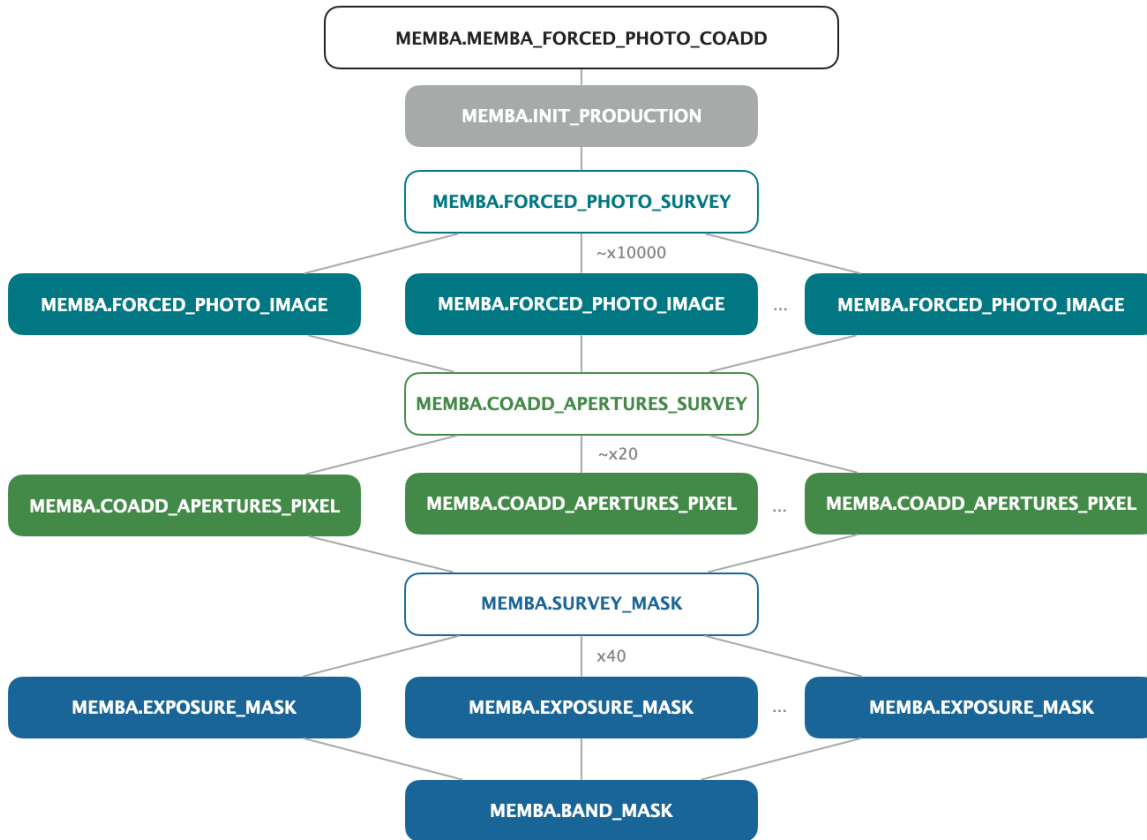


Figure 3.62: The dependency chart of the MEMBA pipeline. Empty boxes define jobs creating sub jobs while filled boxes refer to jobs processing data at the computing farm. The parallelization takes place at three levels: forced photometry, coadds and survey masks.

the database as the healpix pixel ID is stored as a parameters under the forced aperture table. Limiting the area of a coadd job homogenize the number of processed in each job, constraining the CPU and memory of each process.

The dependency of jobs is set between the forced photometry tasks and the coadds, as the later require the photometry to be complete at all bands and layer prior to the combination. On the contrary mask jobs can be processed independently as these do not require inputs from the forced photometry or coadd tasks. The pipeline tree can be seen in Figure 3.62.

As these do not have such intense access to the archive and interact mostly with the database, we can increase the number of parallel pilots up to 200. The CPU time of a MEMBA run is dominated by the forced photometry jobs as it can hold up to 30.000 images. Each job last approximately 15 minutes, resulting in a total process time of few days.

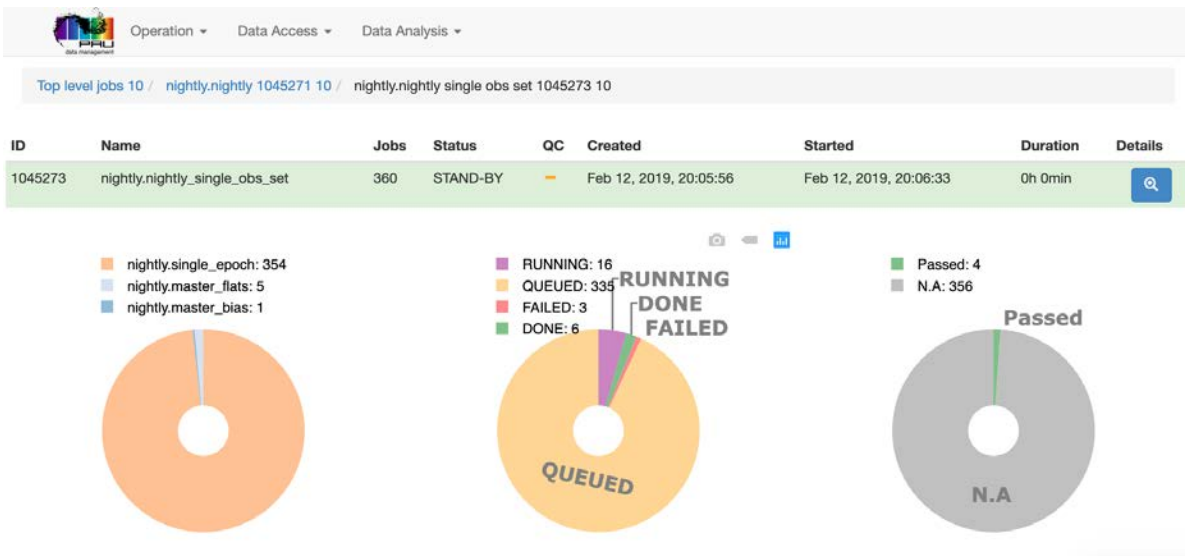


Figure 3.63: The PAUdm Job Monitor, a web-based tool to monitor the operation of the PAUdm pipelines.

3.6.3.4 Monitoring operations

Even the pipelines perform large analysis automatically, they need to be triggered manually (except the Register+Nightly during observing runs to reduce the feedback). Even when the code is stable and has been debugged thoroughly, situations require human intervention when something happens. It can be a processing pilot that went down, an issue with corrupted data or just some very weird image that crashes the algorithm. When dealing with pipelines that contain thousands or even tenths of thousands of job per run, an efficient manager and monitor needs to be set.

For this purpose we designed a PAUdm Job Monitor, a web-based tool that tracks the status of the the processing jobs. The tools allows the operator to view the manually generated top-level jobs and navigate through their sub-jobs, displaying the overall status and detailed configuration of each. In case an error occurred in a particular job, the web app displays its log and the traceback that caused the failure. Even all this would be possible with the Brownthrower Manager in a terminal window, the Job Monitor allows a much quicker and user-friendly way. A screenshow of the Job Monitor is shown in Figure 3.63.

To debug new or updated PAUdm code, a combination of the Job monitor and a test worker node provides an efficient procedure to test and fix until the code is stable and ready to be deployed. The test node allows to run a single job interactively under the same production environment without the need of submitting pilots to the farm and with the ability to see the logging in real time.

3.6.3.5 Catalog distribution with CosmoHub

Once the final scientific catalogues have been produced, we need to distribute them to all scientists in the collaboration to higher level studies are performed, out of the main data management system. The old way of making a large FITS or csv table and sharing it it not efficient due to the large volume and complexity of the data. Furthermore not all scientific studies require the entire data set so a more efficient system had to be built.

Initially, we created a web-based application that allow to select and filter the columns available in a particular table from the PAUdm PostgreSQL database. This solution became popular in the PAU collaboration that was extended to other cosmology surveys and simulations with large sets of data such as DES, MICE and Euclid. At some point the PostgreSQL reached its capabilities to deal with such large datasets and a new technical solution had to be found. With the emerging Big Data solutions, we switched to a Apache Hive which provided a much more scalable and powerful tool than the original database.

This was the beginning of CosmoHub (Tallada et al., 2020), a widely used data distribution and real-time analysis platform for Cosmological datasets. One of the biggest advantages is that very large datasets (with billion entries) can be analyzed in seconds, enabling 1D and 2D histograms that allow to analyze the set and possible subsets from complex queries. These subsets, once verified that contain the correct information the scientist is looking for, can be requested for custom file generation. Once the catalogue is produced, an email notifies the user tha the download of it can be done. All internal and public PAU coadd+photo-z catalogues are available in CosmoHub.

Additionally we provided an alternate solution for scientific analysis with Jupyter Notebooks that run at PIC. This is a recent service that PIC offered, and provides the advantage of performing custom analysis codes in an environment with connection to the PAU database and PAU archive. With just few commands and a shared read-only database account, any scientist can setup a python notebook with the final catalogue of PAU being loaded into a Pandas dataframe. The possibilities of it are endless.

3.7 Enabled scientific results

All previous work, from the planning and observations at the telescope til the final calibrations in photometry, allowed us to deliver accurate galaxy catalogues that enabled multiple scientific results that have been published in peer-reviewed journals. Here we highlight some of the representative results that could only be achieved with well calibrated PAU data.

3.7.1 Photometric Redshift Estimation

As previously said, the PAU camera has been designed to maximize the performance of photometric redshift estimates with its narrow band set installed in a wide field imager. While the

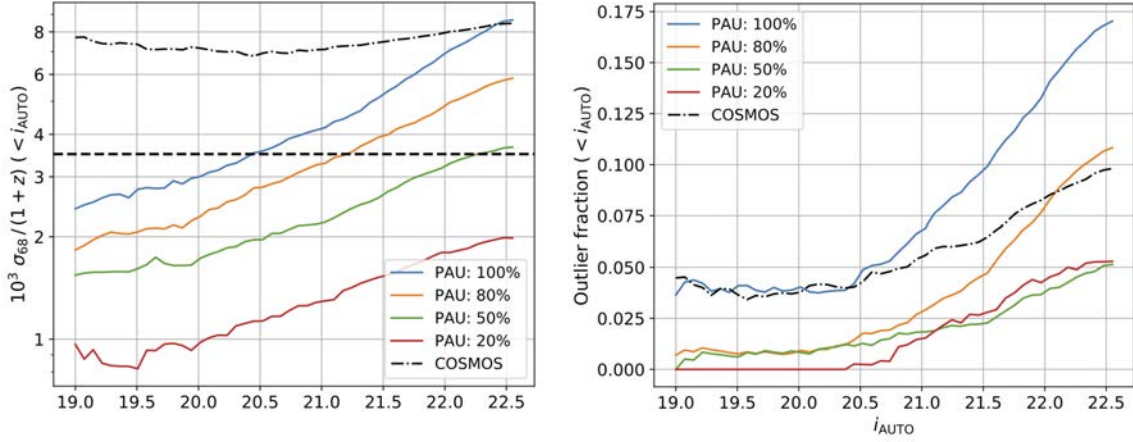


Figure 3.64: The $\sigma_{68}/(1+z)$ (left) and outlier fraction (right) for different quality cuts as a function of the cumulative magnitude bins. The solid lines show the results when 100, 80, 50 and 20 percent of the sample remain after a quality cut. The dashed line shows the COSMOS results without any quality cuts, using the public COSMOS2015 catalogue.

redshift precision that can be achieved with broad band photometry is around $\sigma_{68}/(1+z) \simeq 0.05$, comparisons with spectroscopic redshifts from zCOSMOS in Eriksen et al. (2019) show that PAU reaches a precision $\sigma_{68}/(1+z) = 0.0037$ down to $i_{AB} < 22.5$ when selecting the best 50% of the sources. The performance across the reference i magnitude is shown in Figure 3.64.

These results are significantly better than any previous state-of-the-art photo- z survey, matching the expected performance made with simulations of $\sigma_{68}/(1+z) \simeq 0.0035$ for $i_{AB} < 22.5$. The higher sampling resolution of the wavelengths in PAU is proved to be more effective than the higher signal-to-noise and wavelength coverage of the reference zCOSMOS (Lilly et al., 2009).

In combination with the 26 COSMOS broad band set, Alarcon et al. (2021) demonstrated that PAU Survey photo- z accuracy could improve the one in Laigle et al. (2016) by a factor of $1.66\times$ and $3\times$ at magnitude $i_{AB} \sim 18$ and $i_{AB} \sim 23$ respectively.

Recent developments in the photo- z code and data analysis allow improve the early demonstration presented in Eriksen et al. (2019), extend it to the whole survey area. Such precision over the current 50 deg^2 of sky will provide highly valuable cosmological information. The great accuracy of PAU enables to see the large scale structure with great detail as seen in Figure 3.65.

3.7.2 Intrinsic alignments and clustering

Another remarkable result where the PAU survey stands out is the ability to measure the intrinsic alignments of galaxies (Heavens et al., 2000), an effect caused by tidal forces applied from the surrounding large scale structure, causing the galaxies to align with the neighbouring galaxies and matter. This can significantly bias the weak lensing signal, which is looking for correlated shape distortions due to gravitation only. Understanding how galaxies correlate and

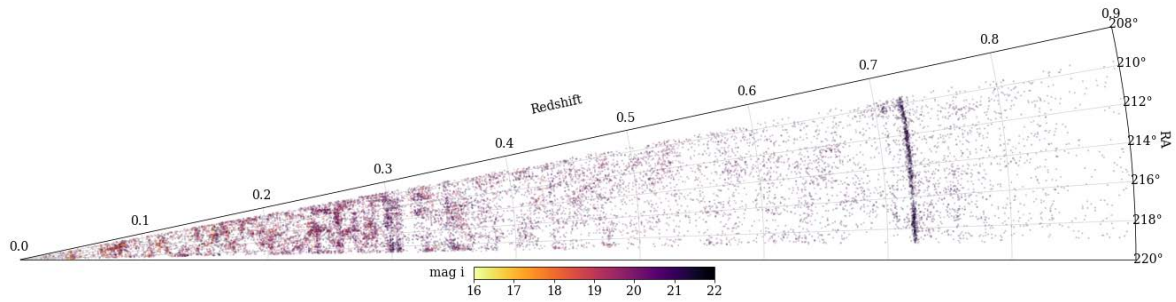


Figure 3.65: The large scale structure measured by PAU for a slice of the W3 field, representing the 10% of the galaxies in this same field. The apparent wall at 0.72 is expected to be an artifact of the photo-z estimator, yet to be corrected.

align is essential for any surveys aiming to perform weak lensing measurements, such as Euclid (Laureijs, 2011).

The first measurements of intrinsic alignments made only over the W3 field of PAU (shown in Figure 3.66 against the the Galaxy and Mass Assembly (GAMA) survey) show that, even for the red sample we were not able to detect any signal with enough significance, we find null signals at high precision in the blue sample. Even it was believed that bluer galaxies would have less alignment with the large scale structure, it could be confirmed from the analysis of this study.

3.8 Summary

In this chapter we have explained the entire data flow of the PAU Survey, starting from the telescope all the way to the scientific results.

First we presented the pixel simulation pipeline, that was fed with galaxies from a cosmological simulation, real stars from a milky way model and all the available instrument and telescope models. We developed an survey strategy simulator that replicated observation sets and sky conditions. Those were finally rendered in pixels under two different approaches: the pixel simulation and the pixel emulation, based on sky data.

Next we explained the work at the William Herschel Telescope, with observing procedures our particular survey strategy. Thanks to an optimized strategy and adaptive exposure time based on real-time conditions, we could homogenize the signal-to-noise across the different sky areas and bands.

Later and with the data now at the data center, we detailed the data reduction process in the Nightly pipeline, where instrumental detrending, astrometry and photometric calibration takes place. Specific algorithms had to be designed due to the particular filter set and tight photometric requirements of the survey. An innovative photometric calibration based in stellar templates was implemented and spectrophotometric standards and good SDSS reference spectra allowed us to validate the whole process.

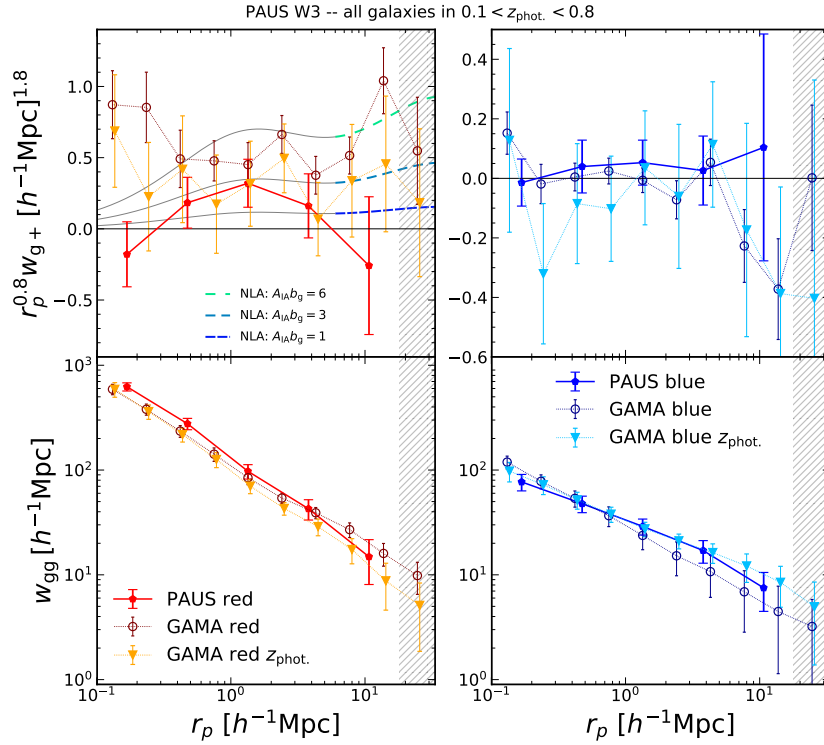


Figure 3.66: The projected position-shear (*top*) and position-position (*bottom*) correlations measured for red (*left*) and blue (*right*) galaxies in PAUS (filled pentagons) and GAMA (open circles/filled triangles).

With calibrated and reduced single-epoch images, we could perform the photometric measurements in the MEMBA pipeline. We describe the accurate flux measurements and background subtraction, putting emphasis on the error propagation necessary for the statistical analysis of the photo- z estimation and other scientific studies. We also validated the narrow band photometry against calibrated spectra from VIPERS and SDSS.

All previous algorithms and pipelines had to run under our computing data center and thus we described the technical aspects necessary to implement and operate such pipelines.

Finally we highlighted two examples of the scientific results enabled from the whole PAUdm data processing described in this chapter.

SPACE BASED SIMULATIONS

In this chapter we describe a complex simulation pipeline implemented for a Space-Based Survey: The Euclid Mission (Laureijs, 2011). We will give an overview of the project, its instruments and how the ground segment is organized. All the details of the simulation pipeline are explained including the input star and galaxy catalogues that we called True Universe, the organization and coordination of the complex parameter modelling and all the instrument simulators and common tools necessary to replicate a multi-channel space survey with the accurate precision it requires. Finally we will summarize the Scientific Challenges were this simulation pipeline was run under the distributed system at data centers around the world.

4.1 The Euclid Mission

Euclid is a medium-class ESA mission aiming to study the nature of the mysterious dark energy and dark matter, that fill up the universe causing its accelerated expansion. Euclid will observe a large area of the extragalactic sky, aiming to map with unprecedented accuracy billions of galaxies up to redshift 2. With its two main probes, Weak Gravitational Lensing and Galaxy Clustering, Euclid will be able track the imprints of dark energy and gravity in the period era over which dark energy played a significant role in accelerating the expansion of the Universe.

The satellite (shown in Figure 4.1) is composed by the Service Module (SVM) and the Payload Module (PLM). The SVM includes the sunshield, the star trackers and gyros, the thrusters, the micro-motions and slews control systems, with hydrazine and cold gas tanks, the Attitude and Orbital Control System (AOCS), the solar panel and electric power system, the thermal regulation system and the downlink communication system. The communications in the X and K band provide 55 Mbit/s data rate during its 4 hour of daily transmission. To store all the observed data prior to be transmitted, the SVM has a memory of 2.6 Tbit. The AOCS will meet the high

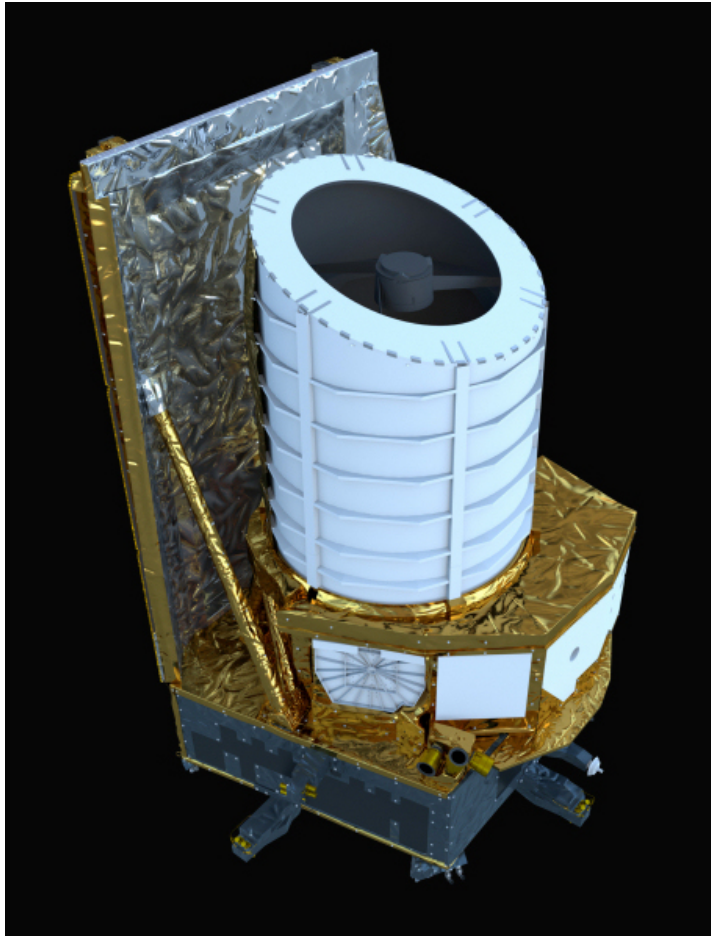


Figure 4.1: Artist's impression of the Euclid spacecraft. Credit: ESA/ATG medialab

quality imaging requirements with a pointing dispersion smaller than 35 mas in each exposure, providing an extremely stable attitude through the whole integration time.

In addition the PLM comprises the telescope, the PLM thermal control system, the Fine Guidance Sensor (FGS), the VIS and NISP instruments (delivered to Airbus Defence and Space division by the Euclid Consortium and detailed in §4.1.1 and §4.1.2 respectively) and the detectors (delivered by ESA). The telescope follows a Korsch design with a primary mirror of 1.20m diameter providing an optically corrected and unvignetted field of view of $0.79 \times 1.16 \text{ deg}^2$. The incoming light beam is split with a dichroic plate reflecting the bluer part that enters into the optical VIS instrument and transmitting the near-infrared light towards the NISP instrument, allowing to observe with both instruments at the same time. The optical design can be seen in Figure 4.2.

The launch is planned in 2022 by a Soyuz ST 2-1b launch vehicle departing from Europe's Spaceport in French Guiana. The mission will carry out a survey that it is expected to be completed in 6 years. The spacecraft will fly to the Second Sun-Earth Lagrangian point (L2) in approximately 30 days where it will perform a large amplitude halo orbit. The main wide survey

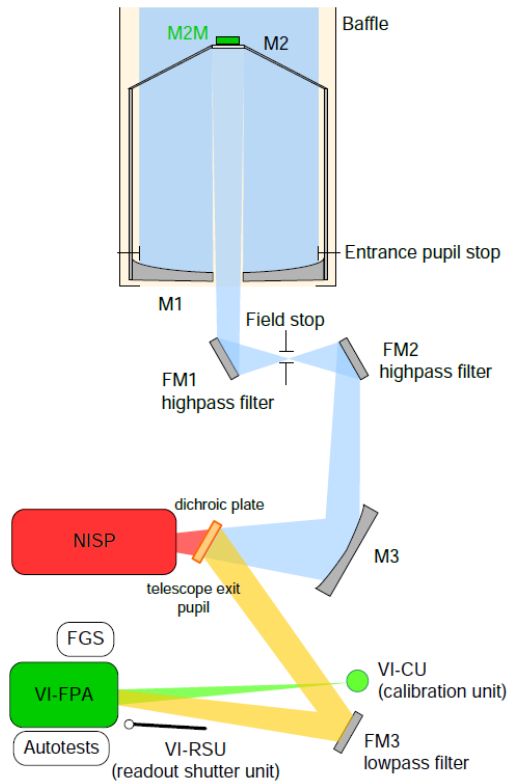


Figure 4.2: Overview of the PLM sub-systems. Credit: Airbus Defence and Space.

will cover the dark sky, away from the Milky Way galaxy plane and the ecliptic plane where zodiacal light is too bright for Near-Infrared observations, covering an area of 15.000 deg^2 . In this area the VIS instrument will measure extended sources of $VIS_{AB} = 24.5$ with a signal to noise of 10σ . On the infrared bands, the point-like sources will have a signal to noise of 5 at a magnitude of $Y, J, H_{AB} = 24$. A deeper survey will be observed with an area of $\sim 40 \text{ deg}^2$ reaching the same signal to noise 2 magnitude deeper than the wide survey.

4.1.1 The VIS instrument

The VIS instrument is an optical imager with a mosaic of 36 (6×6), 4096×4132 pixel CCD detectors provided by e2v technologies, specially optimized for this mission. It contains a single broad band from 550 to 900 nm (equivalent to a combined g+r+i band) shaped by the reflection of the dichroic plate and the quantum efficiency of the CCD detectors. The VIS focal plane covers a field of view of 0.57 deg^2 with a pixel scale of $0.1''/\text{pixel}$. The VIS instrument can be seen in a 3D model in Figure 4.3.

The exquisite image quality of the VIS instrument will allow to measure the shape of galaxies with enough accuracy to estimate the gravitational lensing effect caused by the large scale

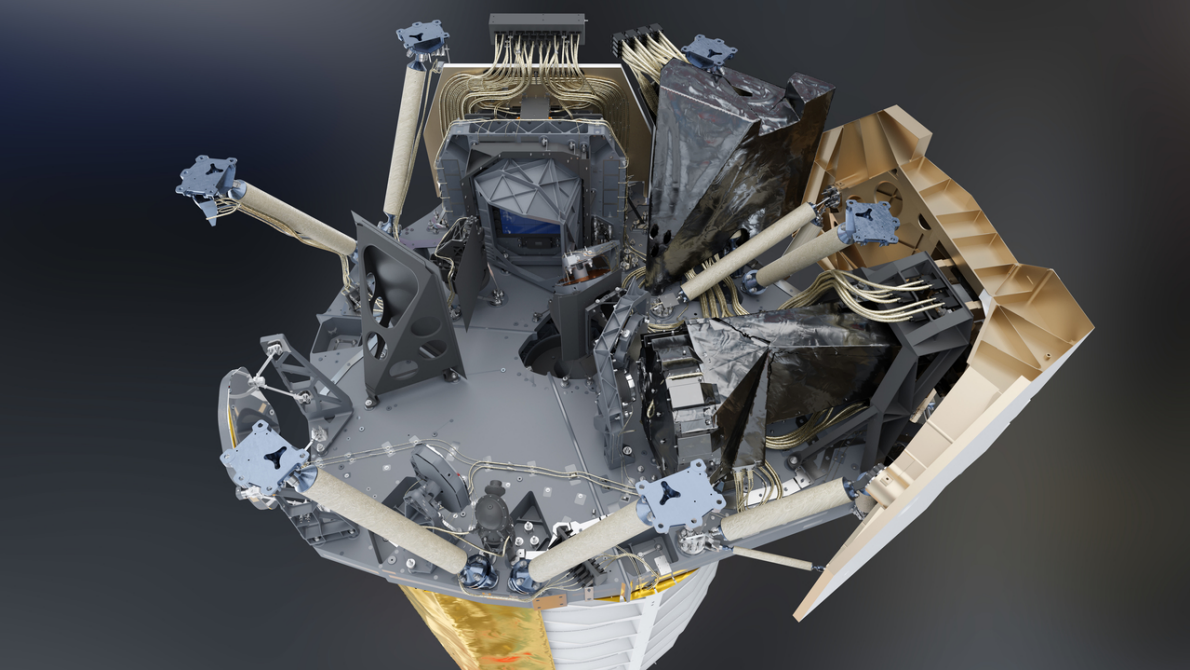


Figure 4.3: A 3D model of the Euclid payload with its two instruments. The VIS images is located on top while the near-infrared spectrometer and photometer (NISP) can be seen to the right. This plate also contains 3 fold mirrors that redirect the light to its corresponding focal plane, the dichroic which splits the beam into its blue and red side of the spectrum and the calibration lamps for obtaining in-flight flat fields. Image credit: Airbus Defence and Space - Toulouse

structures of the universe on distant background galaxies.

4.1.2 The NISP instrument

The Near Infrared Spectrometer and Photometer (NISP) instrument aims to provide broad band photometry in three bands (NISP-P) as well as near infrared low resolution spectra (NISP-S). In combination with the VIS band and Ground Based optical photometry, the 3 NISP broad bands will provide accurate photometric redshifts for billions of galaxies. The low-resolution spectra are obtained with gratings that disperse all the light without any slit, a technique called slit-less spectroscopy. The advantage of this technique is that the design is simpler and more robust than a slit mechanism and requires no target selection. The disadvantage is that there is strong contamination of spectra stepping one into each other. With the NISP-S low-resolution spectra, we can obtain spectroscopic redshifts enabling precise measures of the distribution and clustering of galaxies.

The NISP focal plane contains 16 (4x4) 2040x2040 pixel infrared detectors from Teledyne, with a field of view of 0.53 deg^2 and a pixel scale of $0.3''/\text{pixel}$. The photometric channel is equipped with 3 broad band filters (J, Y and H) that span the range of 900 to 2000 nm with transmissions

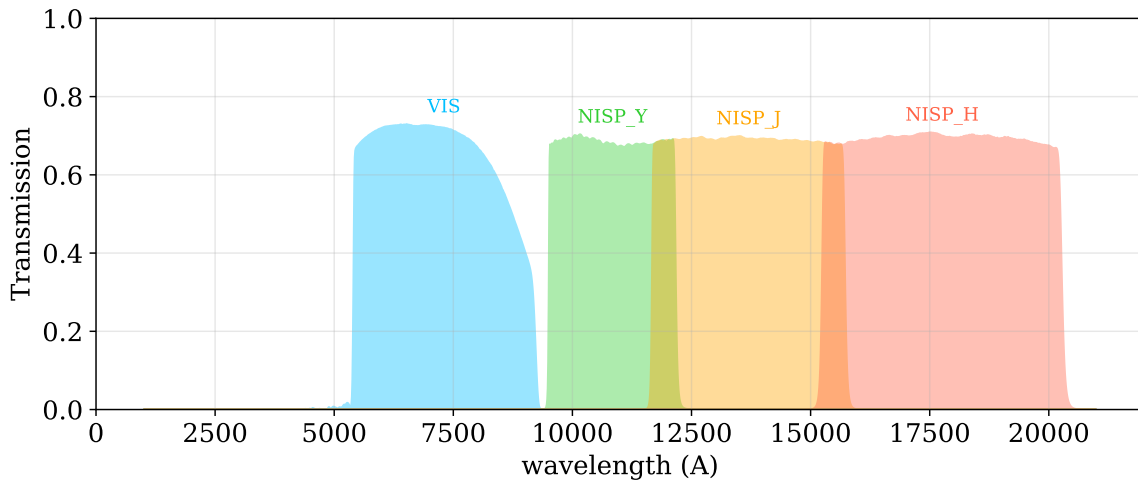


Figure 4.4: The optical transmissions of the VIS and 3 NISP passbands.

as shown in Figure 4.4. The spectroscopic channel has 4 gratings providing a spectral resolution of $R=380$ for a $0.5''$ source. Three gratings cover the redder side of the spectrum from 1250nm to 1850nm and the remaining one is the "blue" grating covering a range from 920 to 1250nm. The three red gratings cover the same wavelength range but have a rotated dispersion orientation to facilitate the de-contamination of overlapping slit-less spectra.

4.1.3 The Ground Segment

The ground segment is divided into two main components: the operations ground segment (OGS) and the science ground segment (SGS). The first one is in charge of operating the spacecraft, analyzing the telemetry and performing the down-link transmissions of data, all coordinated at the Mission Operations Center (MOC) located at the ESA's European Space Operations Centre in Darmstadt, Germany. The SGS is then composed by the Science Operations Center who manages the survey planning and performs quick look of the data, while the Euclid Consortium SGS is in charge of the whole data processing down to the production of the scientific results.

In order to achieve its goals the SGS has 9 national Scientific Data Centers (SDC) in a distributed common environment. Additionally, the SGS has distributed the main pipeline into 10 Organization Units (OU) each responsible to define, design and validate an area of the processing. This is the list of OUs and their corresponding tasks:

- **OU-VIS:** Instrumental detrending, astrometric and photometric calibration of the VIS imaging.
- **OU-NIR:** Instrumental detrending, astrometric and photometric calibration of the NISP-P imaging.

- **OU-SIR:** Processing, calibration and extraction of the 2D and 1D spectra from the NISP-S channel.
- **OU-EXT:** Ingestion, instrumental detrending, astrometric and photometric calibration of the Ground Based External survey data.
- **OU-MER:** Merging and photometry of Euclid and External data.
- **OU-SPE:** Estimate spectroscopic redshifts and spectral properties from the calibrated NISP-S clean spectra.
- **OU-SHE:** Measure shear and weak lensing from the clean VIS galaxies.
- **OU-PHZ:** Estimate photometric redshifts from the combined MER photometry.
- **OU-LE3:** Scientific processing and analysis from the Level 3 data.
- **OU-SIM:** Image simulations of VIS, NISP and EXT channels.

A schematic view of the SGS can be seen in Figure 4.5.

4.1.4 OU-SIM

Our work is focused on the design, organization and implementation of OU-SIM, an organization unit aiming to provide precise simulations to the whole Science Ground Segment. The simulations produced have two main goals: serve as test data to implement the SGS pipeline and prepare the infrastructure before the actual Euclid observations arrive. And secondly to validate the restrictive requirements that each OU has in terms of performance and data quality. These requirements derive from the top level ones that ensure Euclid can reach its scientific goals. This is particularly important in any space mission due to its elevated cost and the impossibility to fix anything once the spacecraft is launched. Additionally OU-SIM simulations have also been useful to determine the overall mission performance in the end-to-end Science Performance Verification (SPV) tests. These tests are performed several times throughout the Phase C and D of the mission preparation and require the latest knowledge of the instrument models and science inputs. As expected in such a complex mission, technical issues appear from time to time and OU-SIM simulations could reproduce the particular problem and the impact of the issue could be evaluated allowing further decisions to be taken.

The tight requirements in this mission required extremely precise simulations, with exquisite detail and all known instrumental and environmental models to make sure nothing was not tested prior to the launch. Even though OU-SIM only produced unprocessed level 1 simulations, we had to establish interfaces with all OUs, as each one of them places different simulation requirements that we had to introduce in our pipeline. Also, interfaces had to be built with the cosmological simulations science working group as our main supplier of galaxy catalogues, as

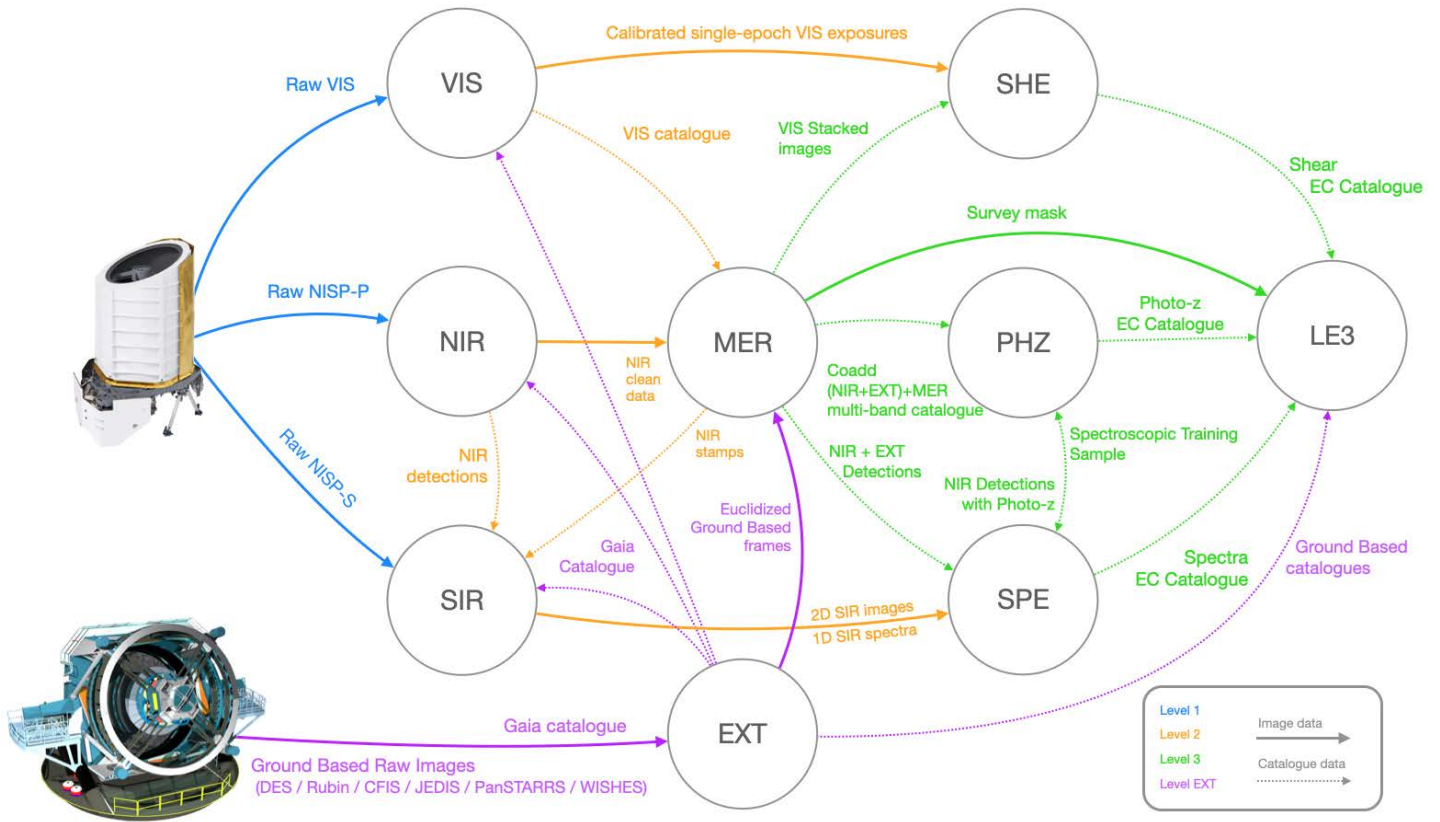


Figure 4.5: A schematic view of the SGS with the relations and data flow between the Organization Units. Before real observations start, OU-SIM provides simulated data for the 3 Euclid channels and the EXT data.

well as with the instrument teams who provided the models that we had to implement in the pixel simulation.

Thanks to the expertise in building the pixel simulation for PAU as explained in Chapter 3, we could have a more clear view of the needs for this greater simulation that were defined early in the development of OU-SIM.

4.2 The True Universe

Star and Galaxy sources are the main object of study for Euclid. It is therefore critical to have a realistic sample where the scientific performance of the mission can be evaluated. Being in a simulation we can define its true values (such as position, brightness, shape, etc) and therefore we call the input sources to the simulation *True Universe (TU)*. Comparing the true values from the catalogue with the ones measured in the pixel simulation through the SGS pipeline we can evaluate the performance at various levels and validate the compliance of the requirements at the different stages of the mission.

In this section we describe the galaxy and stellar TU catalogues used in the simulation, as well as the common spectra and thumbnail libraries that allow us to transform their parameters into fluxes of the pixel simulation.

4.2.1 Cosmology & Mock Galaxy catalogue with Flagship

Euclid aims to study the nature of dark energy by modelling the cosmological parameters from the observations of billions of galaxies. Therefore, we need to construct a galaxy catalogue that follows a given cosmology with great accuracy such that we can evaluate with end-to-end simulations how well Euclid will recover the input cosmology.

4.2.1.1 Dark matter simulation and halo catalogue

For this purpose the Flagship Simulation was created. This is the largest n-body simulation of the universe ever made, with 12600^3 particles over a simulation box of $3780 h^{-1}\text{Mpc}$ with an agreed cosmology close to the Planck 2015 values. The 2 trillion dark-matter particle simulation was run with the PKDGRAV3 (Potter et al., 2017) code on the Piz Daint supercomputer at the Swiss National Supercomputer Center (CSCS). The processing produced in real time an all sky particle light cone down to redshift $z = 2.3$.

The dark matter halos were computed from the particle lightcone using the ROCKSTAR halo finder, an adaptive hierarchical refinement of a friends-of-friends method, able to track substructure and relations from parent dark matter halos to the smallest subhalos.

Following the technique described in (Fosalba et al., 2015b), all-sky lensing convergence maps were built, which enable weak lensing effects to be included in the final galaxy catalogue.

4.2.1.2 Mock galaxy catalogue

As a final step of the chain, the mock was built similarly to the one explained in §3.2.1, implementing an improved Halo Occupation Distribution technique. This resulted into a mock galaxy catalogue with 2 billion galaxies covering a volume equivalent to the one Euclid will map. Particular care was put to model the H-alpha emission of the galaxies (Pozzetti et al., 2016), a line that is particularly important for NISP to determine the spectroscopic redshifts. Additionally, the size and shape distributions of the galaxies were updated and improved from the reference in MICE, following the distributions from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS), an HST survey with accurate morphology samples.

The final mock catalogue contains the following information that allows the pipeline to fully render its flux at pixel level:

- **ID:** A unique ID of the source and the ID of the parent dark matter halo.
- **Positional coordinates:** The source location in equatorial coordinates (RA, Dec) and the lensed positions due to the gravitational distortion.

- **True Universe kind:** An identifier of the object kind from the various true universe inputs such as flagship central or satellite, QSO, high-z and strong lensing, as different sources are rendered differently.
- **Redshift:** The true and observed redshift of the galaxy, including peculiar velocities.
- **Reference magnitude:** An absolute and apparent redshift of the galaxy at a given band. This is used to scale the spectrum at a given brightness and the band depends on the object type as we cannot define a common band where all galaxies emit.
- **SED template:** A reference SED template from a given library and the template for the intrinsic extinction law. This is used to reconstruct the spectrum (detailed in §4.2.3). Additionally, the color excess $E(B-V)$ for the extinction law needs to be provided.
- **Emission line fluxes:** The fluxes of various emission lines key for the spectroscopic redshift determination. The actual set of lines is H_α , H_β , OII, OIII, NII and SII.
- **Morphological parameters:** In order to render the light profile of the galaxy we need the Bulge and Disk half light radius, the ratio of the flux in the bulge component to the total flux (often written B/T), the sersic index of the bulge profile, the axis ratio defining its intrinsic ellipticity, the inclination angle of the disk and the position angle with respect to the north. How these parameters are implemented to reconstruct the galaxy shape are further described in §4.2.4.
- **Lensing parameters:** Based on the flagship lensing and convergence maps, we extract the Gamma 1 and 2 shear values that allow us to distort the shape. We also extract the convergence Kappa value which determines the shape and flux magnification of the galaxy.
- **MW extinction:** The optical extinction of the Milky Way in the V band. This is derived from the Planck Thermal dust maps at the observed (lensed) coordinates of the galaxy.

With the previous list of parameters we can include any galaxy in the catalogue in any of the simulation channels, both photometric and spectroscopic with all its complex instrumental effects. Additionally we compute the True "error free" fluxes for each galaxy in all pass bands of interest, from Euclid and the Ground Based External Surveys.

4.2.1.3 Primordial Universe sources

The Flagship galaxy mock aims to cover the scientific needs of the primary science in Euclid. However there are other fields where Euclid will provide countless new sources, far beyond of those available to date. One of the legacy science that will benefit the most from the unique wide and deep Euclid survey is the primordial universe. These are very distant galaxies where Euclid is expected to observe and the Flagship dark matter run does not contain, as its lightcone

reaches "only" a redshift of 2.3. The Primordial Universe Science Working Group provided a sample of high redshift galaxies and Quasars at a redshift greater than 6, as described in (Euclid Collaboration et al., 2019). Providing these sources in the same format and with the same set of parameters as the flagship galaxies we could easily introduce them into the True Universe set as additional sources to be simulated.

Similarly, the Strong Lensing Science Working Group provided a set of galaxies with projected lens mass density greater than the critical density, delivering arcs, Einstein rings and multiple images. Euclid is expected to measure more of these particular sources than any previous survey, with tenths of thousands of new objects. Having these in the TU set allows to train algorithms to train and classify them much more efficiently.

4.2.1.4 Visualizing the mock

These massive data sets as the Flagship galaxy catalogue are most often treated as computer tables that no one ever has the chance to see. Even those that are curious to visually observe it, struggle to render the mock, as common plotting libraries are not designed to deal with so much data. For scientific and outreach purposes we implemented a small pipeline based in Splotch (Dolag et al., 2008), a ray-tracer software to visualize cosmological data. In our pipeline we also programmed camera movements enabling video sequences flying around the dark matter boxes. In the example shown in Figure 4.6 we rendered a slice of the lightcone, from redshift 0 to its edge at 2.3. We rendered the central galaxies with a green map while satellites are shown in redder tones.

4.2.2 Stellar catalogue

Even though stars are not the primary scientific objective in Euclid, they are key in the whole analysis and calibration of the data. They are the reference for all astrometric and photometric calibrations, as well as for the extremely precise PSF modelling, required for all lensing measurements. On the negative side, they provide confusion to the classification of sources and produce numerous complications such as blending, bleeding, ghost reflections, scattered light and more. It is then critical to have a realistic and accurate distribution of the stars at each location of the sky, with representative density and stellar types.

With such a wide sky coverage and depth as that of Euclid, aiming to observe thousands of square degrees down to optical and near-infrared AB magnitudes below 24, we need a synthetic model that accurately represents the distribution of stars in the Milky Way. Similarly to the stellar simulation in PAU described in §3.2.2, we chose the Besançon model (Robin, Creze, 1986) as our main stellar population synthesis model of the Galaxy. The model provides a realistic representation of populations in the disk and halo, and even it also provides photometry at various bands. We only request a single reference measure in the 2MASS H band. The model provides the surface temperature, gravity and metallicity for each star that we use to derive the

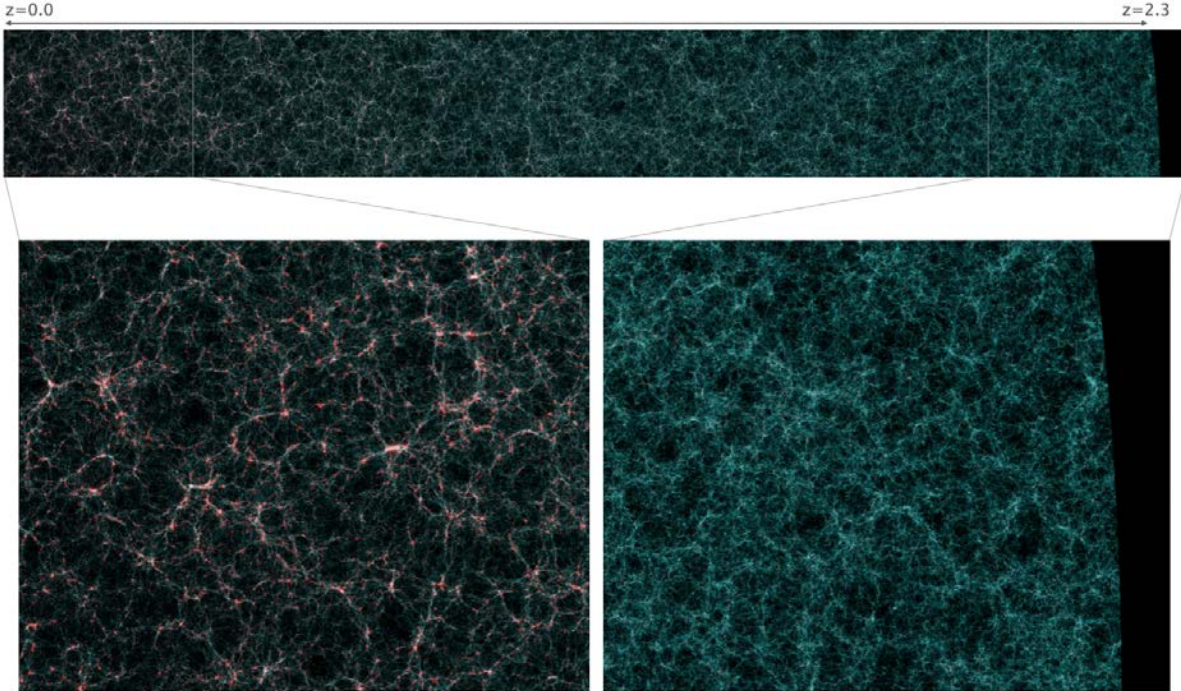


Figure 4.6: Top: False colour image showing a slice (0.3% of the total volume) of the full light-cone flagship mock. Central galaxies are coloured in green and satellites in red, where filaments from the large scale structure are seen in great detail. Bottom left shows a zoom in section with the local universe ($z=0$). Bottom right the zoomed in fraction of the furthest galaxies ($z=2.3$)

stellar template from the Basel 2.2 stellar library (Lastennet et al., 2002) (further details of the library and how we reconstruct spectra in §4.2.3. We ran an all-sky Besançon model simulation down to $H < 26$, skipping the Galaxy plane as Euclid will not observe there and it holds too many stars at those magnitudes. Even though the NISP instrument has a defined magnitude limit of $H < 24$, we simulated a deeper catalogue to include the lower SN sources into the pixel-level images, as these are also relevant in altering the photometry of other sources of interest, like the main galaxy sample. This resulted into a massive final stellar catalogue containing 3 billion stars with the following set of parameters:

- **ID:** A unique True Universe id, not overlapping with the galaxy sources.
- **Sky coordinates:** The equatorial coordinates of the star in RA and Dec.
- **Distance:** The radial distance of the source in Kpc. This is used to obtain the optical extinction using 3D dust model maps (Schultheis et al., 2014).
- **Reference magnitude:** The apparent magnitude in the Vega system for the 2MASS H band. This is used to scale the template spectra to its apparent brightness.
- **Extinction:** The Milky Way optical extinction in the V band.

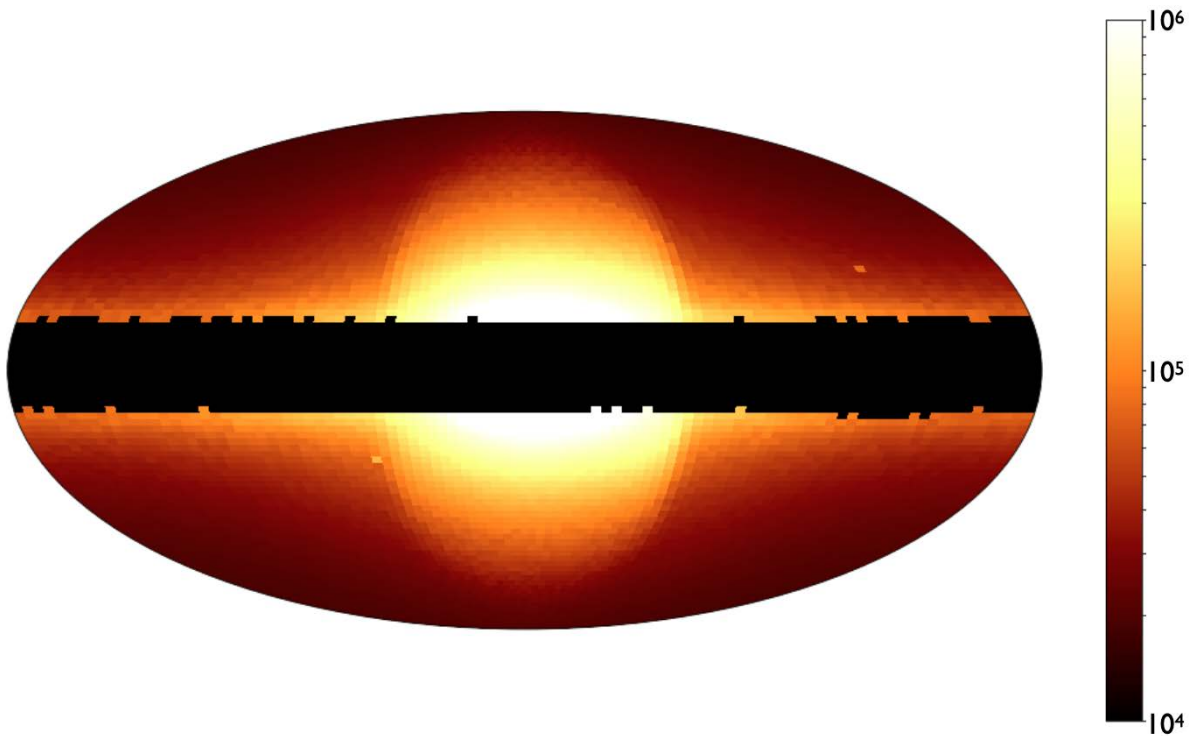


Figure 4.7: An all-sky (without $\pm 10 \text{ deg}^2$ from the galactic plane) of the Besançon simulation we computed for Euclid. It is complete down to $H < 26$, summing up a total of 3 billion stars.

Just as in the galaxy catalogue, true fluxes in all Euclid and external bands are also computed. The 2D count of the Besançon simulation run can be found in Figure 4.7.

The wide survey was designed using the reference zodiacal light model and straylight levels from the brightest stars near the galactic plane. Therefore the brightest end of the stellar catalogue had to be replaced with real stars from an all-sky catalogue. We chose the Tycho2 stellar catalogue (Høg et al., 2000) as we could use the Pickles procedure described in (Pickles, Depagne, 2010) to match stellar types for each source in the catalogue and then to our reference Basel library. This catalogue contains the 2.5 million brightest stars in the sky down to V magnitude ~ 11 .

One of the most difficult tasks of identifying the very distant quasars is being able to differentiate them from the point-like brown dwarf stars that have a comparable very red spectrum which delivers similar NIR colours. With the objective of a more complex QSO identification, we needed to include MLT type stars. The M-class stars were already included in the Besançon sample, however the redder LT were not, so we added new templates to the Basel Library and included the brown dwarfs following recipes described in (Euclid Collaboration et al., 2019).

In the first prototypes of the SGS pipeline, we used the True Stellar Catalogue as the astrometric reference for the calibration algorithms. However this ideal catalogue will not be available during real observations as Gaia will be the actual reference used. With the performances de-

scribed in the Gaia DR2 (Gaia Collaboration, 2018), we could replicate a realization of a Gaia catalogue based in our True Universe sample, with realistic astrometric and photometric errors based on the magnitude and type of star. Each star in the catalogue will then incorporate position, position errors, magnitude and magnitude errors on the G, BP and RP bands, down to $G \sim 21$. With that, a much more realistic Gaia-like astrometric reference is provided to the VIS, NIR and EXT calibration pipelines.

4.2.3 SED library

The OU-SIM pipeline is capable of delivering image simulations from the 3 Euclid channels and from various Earth telescopes and their corresponding instruments. As it will be explained in §4.4, VIS, NISP-P, NISP-S and EXT pixel simulators are implemented as separate independent codes. Having all different codes implementing common routines is risky and a potential source of bugs. For this reason we build a common SED library named SIMSPECTRA, in charge of reconstructing the observed spectra from the parameters in the catalogue and the corresponding template library. The incident spectra are reconstructed every time a given source needs to be rendered. Even though this seems inefficient, precomputing the spectra for all objects is prohibitive with so many objects, as catalogues of spectra are massive and rise the amount of memory and storage needed to process them. Reconstructing the spectra is a quick process (when optimized) and enables low-memory jobs that can process in parallel under distributed architectures.

Here we describe the implementation of the Galaxy and Stellar in the SIMSPECTRA common SED library.

4.2.3.1 Galaxy spectra reconstruction

Our base galactic template library is composed of COSMOS SED templates as described in Bruzual, Charlot (2003) and Polletta et al. (2007). These set of templates are assigned to each galaxy in the production of the mock, as well as the intrinsic dust extinction law from Prevot et al. (1984) and Calzetti et al. (2000). The galaxy template set used for these simulations is shown in Figure 4.8.

Here we describe the process of reconstructing the observed spectra from the parametrized information in the catalogue. Before anything, we homogenize all wavelength dependant elements into a common sampling to enable direct operation with the arrays. From the needs in the Euclid simulation pipeline, we need to ensure the spectra cover the wavelength range of interest and with enough resolution. Thus we establish a range between 600 and 24000 Å, in steps of 1 Å. With a spectral resolution of $R=380$ in NISP-S we should be able to sample properly the dispersed spectra.

First we initialize the spectrum with a linear combination of the galaxy templates. The way this library is constructed, allows for intermediate combination of spectra, enabling a wider

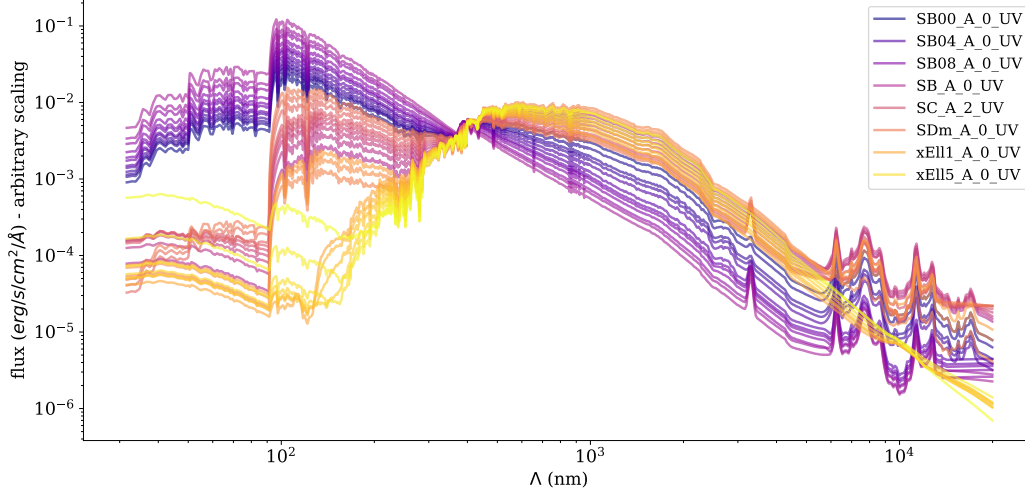


Figure 4.8: The 31 COSMOS template set from Bruzual 2003 and Polletta 2007

variety of colours and shapes, which give more realism to the simulation. We compute the combined template such as:

$$(4.1) \quad f_{combined}(\lambda) = (1 - i_{frac})f_{template}(\lambda)[i] + i_{frac}f_{template}(\lambda)[i - 1]$$

where i_{frac} is the fraction of the SED number above its integer. Next we load the dust extinction and scale it with its color excess E_{B-V} and an average visibility value $R_v \approx 5$ such as:

$$(4.2) \quad f_{ext}(\lambda) = f_{combined}(\lambda)10^{-0.4R_v E_{B-V}}$$

The next step is to shift the spectrum to the observed redshift z_{obs} . This effect is reproduced by scaling the sampled wavelengths by $(1 + z_{obs})$ and dividing the fluxes by the same factor $(1 + z_{obs})$. With the redshifted spectrum we can scale the template to its absolute flux. Our reference value to scale the spectrum is the absolute magnitude M_r . The scaling cannot be computed directly with it as we aim to scale the observed spectra, which requires an apparent magnitude. This process requires computing the luminosity distance d_L and the K_{Corr} correction.

The luminosity distance d_L is often given in the mock, but we can compute it with an Astropy FlatLambdaCDM cosmology function, setting the Hubble constant at $z = 0$ H_0 and Omega matter ω_m (density of non-relativistic matter at $z=0$) used in the dark matter simulation. In the case of the Flagship run, $\omega_m=0.319$ and $H_0=100 \text{ h km s}^{-1} \text{ Mpc}^{-1}$.

As our reference flux is not Bolometric and lies within a particular passband, we need to compute the K_{Corr} correction that transforms the absolute magnitude to its apparent one at the current redshift. By definition:

$$(4.3) \quad K_{Corr} = 2.5 \log_{10} \frac{f_{restframe}}{f_{observed}}$$

we obtain both $f_{restframe}$ and the redshifted $f_{observed}$ in the reference passband as:

$$(4.4) \quad f_v = \frac{\int \frac{f_\lambda R(\lambda) \lambda^2}{c} d\lambda}{\int R(\lambda) d\lambda}$$

where $R(\lambda)$ is the system response of the reference magnitude passband. With the K correction and the luminosity distance we can now compute the apparent magnitude:

$$(4.5) \quad m_r = M_r + (5 * \log_{10}(d_L) - 1) + K_{Corr}$$

The scaled observed flux in $\text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}$ at the reference passband from the AB definition is then:

$$(4.6) \quad f_{v,r} = 10^{-0.4m_r + 48.6}$$

and the scaled continuum spectrum is therefore:

$$(4.7) \quad f_{continuum}(\lambda) = f_{ext}(\lambda) \frac{f_{v,r}}{f_{v,observed}}$$

At this stage we can add the emission lines to the scaled template. As emission lines are already scaled and modelled with extinction, we compute them after applying the dust to the scaled continuum. The central frequency of each line is taken from Vacuum or Air, depending on whether the channel we simulate is Ground or Space based. The table below shows the set of emission lines and their corresponding central wavelength used for both air and vacuum:

Line	Vacuum (Å)	Air (Å)	ratio to primary
Ha	6564.614	6562.801	
Hb	4862.721	4861.363	
OII	3727.090	3726.030	
OII*	3729.880	3728.820	1:1
OIII	5008.239	5006.843	
OIII*	4960.295	4958.911	1:3
NII	6585.270	6583.450	
NII*	6549.860	6548.050	1:3
SII	6718.290	6716.440	
SII*	6732.680	6730.820	1:1

Secondary lines are indicated with * and the corresponding flux ratio to the primary is indicated, as no secondary fluxes are explicitly indicated in the TU catalogue. We estimate the

width of the emission line σ_{el} in Km/s following a Tully-Fisher relation (Tully et al., 1975), such as:

$$(4.8) \quad \log_{10}(\sigma_{el}) = (-0.10 + 0.01 * z_{obs}) * (M_r - 3.0) - 0.05 * z_{obs}$$

with the observed redshift z_{obs} (including peculiar velocities) and M_r the absolute magnitude in the r band. In any case we limit the approximation to a maximum velocity of 50 Km/s. We generate all 10 emission lines as Gaussian profiles with the integrated flux as defined in the catalogue, at the corresponding central wavelength with a sigma of σ_{el} . The reconstructed emission line set needs to be shifted in wavelength and scaled down in flux with redshift just as we did for the continuum spectra, with the $(1+z)$ factor. Then we compute the emission line contribution only $f_{el}(\lambda)$ spectrum as:

$$(4.9) \quad f_{el}(\lambda) = \sum_i^{10} \frac{f_{el_i}}{\sigma_{el} \sqrt{2\pi}} e^{-\frac{(\lambda - \mu_i)^2}{2\sigma_{el}^2}} \frac{1}{(1+z)}$$

where μ_i and f_{el_i} are the central wavelengths and TU line fluxes respectively for the ten emission lines in the corresponding medium.

The resulting Gaussians at the redder wavelengths can now be added to the calibrated continuum component of the spectra:

$$(4.10) \quad f_{galaxy}(\lambda) = f_{continuum}(\lambda) + f_{el}(\lambda)$$

With all the elements from the source galaxy, we introduce the effect of lensing magnification with:

$$(4.11) \quad \mu = \frac{1}{(1 - \kappa)^2 - |\gamma|^2}$$

$$(4.12) \quad f_{magnified}(\lambda) = f_{galaxy}(\lambda) * \mu$$

where κ is the convergence and γ the shear vector obtained from the flagship lensing maps.

Finally, the only remaining step is to apply the Milky Way extinction to the magnified galaxy spectra, at the specific location in the sky. From the Planck dust maps (Planck Collaboration, 2014) we obtain a precise color excess $E(B-V)$ and following an O'Donnell law (O'Donnell, 1994) with R_v of 3.1 we extinct the spectrum to obtain a final incident spectra as:

$$(4.13) \quad f_{incident}(\lambda) = \frac{f_{magnified}(\lambda) + f_{el}(\lambda)}{10^{0.4A_v(\lambda)}}$$

where $A_v(\lambda)$ is the O'Donnell law scaled by $E(B-V)$ and R_v extinction terms.

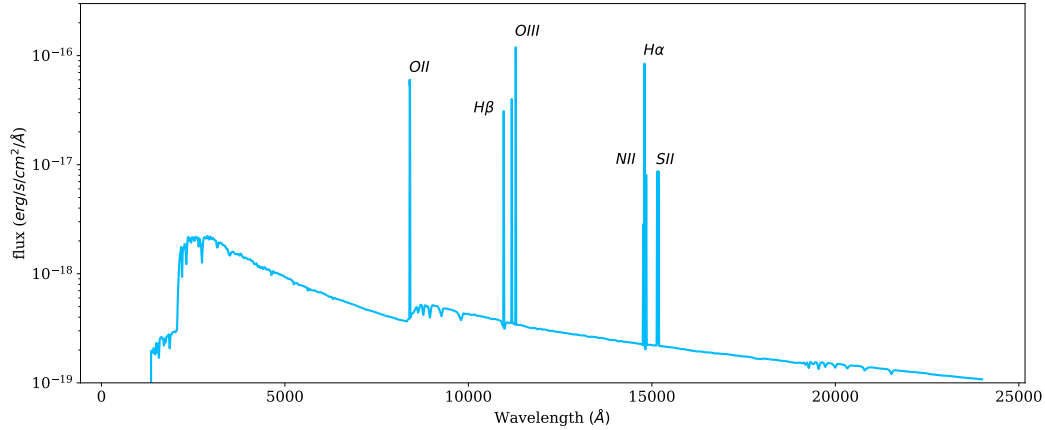


Figure 4.9: A reconstructed galaxy spectrum with parameter information from the True Universe catalogue.

An example of a reconstructed spectrum from catalogue parameters is shown in Figure 4.9.

Once we have a final spectrum, we can compute the true fluxes in all Euclid Bands (VIS, NISP-Y,J,H), Ground based surveys (DES [griz], Rubin [ugrizy], CFIS [ur], PanSTARRS [iz], JEDIS [g] and HSC [z] explained in §4.4.4) and Gaia[G,BP,RP]. These integrated fluxes are computed following the same procedure explained in §3.4.4.8. The integrated passband fluxes are not used by most instrument simulators, as these require the full chromatic information to precisely include the instrumental effects (such as wavelength dependent PSF or Quantum Efficiency). Only the simpler Ground-Based simulator uses the precomputed band fluxes, avoiding the use of spectra in the process of simulating pixels of a specific band. However this process is computationally expensive with the various integrals required in the method and computing the exact spectra and fluxes in all bands for the entire flagship catalogue is prohibitive. Instead we compute the exact values in a matrix sampling various points in the parameter space and perform a multi-space interpolation to obtain the integrated fluxes in all bands. Comparing the exact fluxes from the full spectra reconstruction process against the approximated interpolations show that the fast method has an accuracy of 0.005 mag RMS, enough for all SGS validation applications.

4.2.3.2 Stellar spectra reconstruction

As described in §4.2.2, our stellar library is the Basel 2.2 (Lastennet et al., 2002) template set. We managed to assign an SED from both the Besançon model simulation and the Tycho2 catalogue to a particular stellar template. The reconstruction of the stellar spectra is much simpler than the case of galaxies as there is no redshift, K correction or lensing magnification in the reconstruction process. There are just two steps to calibrate the template to the incident calibrated spectrum for each star. First we need to scale the flux to our reference H 2MASS band. With the already

available apparent magnitude in the H band we can simply compute the template flux $f_{v,template}$ in the observed band as defined in Equation 4.4 and the 2MASS H band flux $f_{v,H}$ as Equation 4.6 and scale the star spectrum as:

$$(4.14) \quad f_{star}(\lambda) = f_{template}(\lambda) \frac{f_{v,H}}{f_{v,template}}$$

The only remaining step is to correct the Galactic extinction. The Tycho2 stars are not extinguished corrected and therefore we do not apply further correction. Additionally, these very bright stars are expected to be nearby and few dust will affect their spectra. On the other side, Besançon stars are further apart and the reference magnitude does not include extinction. With the radial distance information provided by the model, we can apply a 3D dust model Schultheis et al. (2014) and obtain an accurate reddening factor. For those model stars, we apply a procedure similarly to the galaxy spectra as:

$$(4.15) \quad f_{star_{ext}}(\lambda) = \frac{f_{star}(\lambda)}{10^{0.4A_v(\lambda)}}$$

with the typical relative visibility value R_v of 3.1 in the case of the Milky Way.

Even though the wavelength range of the stellar templates cover the whole spectrum of interest, the reddest part from 12.000Å onwards has a reduced sampling in the Basel Library with 50Å intervals. This low-resolution characteristic of the templates prevents the NISP-S channel to properly sample the spectra in the simulation with the consequent calibration issues derived from the undersampled spectra. The wavelength calibration in NISP-S has to be accurate to better than half a pixel, which translates to approximately 6 Angstrom. Therefore, we would need stellar spectra sampled at 2 to 3 angstrom intervals. For this reason we replaced the feature-less low sampled redder part of the spectrum with real hi-resolution measurements following the continuum as defined in the template. In this way we do not alter the colors of the spectra but we allow accurate wavelength calibrations in the simulated NISP-S channel.

The stellar library was also extended with 365 MLT star templates that allow the additional Brown Dwarf stars to be assigned an SED, as the Basel 2.2 do not contain those particular stars.

4.2.4 Galaxy profile generators

4.2.4.1 Overview

Similarly to what is done for the spectra library, we also need a common tool to transform the morphological parameters from the catalogue to pixelized galaxy light profiles that can be rendered in the simulated image. We call this library SIMTHUMBNAILS and is capable of rendering the light profiles at the exact pixel scale for each instrument, using the WCS information of the target image to be rendered. Originally, the OU-SIM pipeline precomputed the galaxy light profile thumbnails at a high resolution to enable rendering the stamps in all channels. This approach presented two main problems. The first one is that it is never rendered

at the exact sub-pixel location and therefore some degradation in the profile had to be taken to shift it to the correct location, or a slight shift in position had to be assumed. This was particularly bad for the VIS simulations, where great care has to be taken in the shape of each galaxy to enable weak lensing measurement with the required accuracy. The second issue was to store, transfer and load such large data sets of stamps, with too few resolution for VIS and too much for the Ground Based simulations. The SIMTHUMBNAILS enabled a much more accurate light profile rendering and lighter data handling, at the expense of rendering the galaxy profile every time it is needed. The core of the library is based on Galsim (Rowe et al., 2015), which allows plenty of flexibility to render different profile configurations.

4.2.4.2 Current model

The latest model we have for the standard Flagship galaxies is a two-sersic profile tracing the bulge and disk components.

First we define the bulge as a Sérsic profile:

$$(4.16) \quad I_{bulge}(r) \sim e^{-b(r/r_e)^{1/n}}$$

where r_e is the half light radius and n the Sérsic index provided by the Flagship catalogue. The b parameter is calculated to give the right half-light radius (not necessary in case of providing the scale radius instead of the r_{50}). We also specify the relative flux to the model as the bulge fraction. This will set the correct relative flux to the disk component once both are added together. The previous version of this component had a constant Sérsic index set to 4, but HST-CANDELS data (Grogin et al., 2011) showed that a wider range of Sérsic indexes better represented the observations.

The current bulge model $I_{bulge}(r)$ is then transformed to the corresponding minor-to-major axis ratio with a shear function, with the bulge axis ratio defined in the catalogue.

Next we initialize the disk component with an exponential inclined model. As elliptical galaxies do not contain a disk component, this is only processed for spiral galaxies with a bulge fraction below 1. The exponential inclined model is a particular Sérsic profile of index 1, characterized by three properties: the inclination angle (where 0 degrees refers to a face-on galaxy and 90 degrees would indicate it is positioned edge-on), its scale radius (or effective radius), and the scale height of the disk. Instead of explicitly defining the scale of the disk, we set a relative ratio $h_{s/r}$ from the scale radius set to 0.1. The inclination angle can be computed from the axis ratio b/a such as:

$$(4.17) \quad \theta_{inclination} = \arcsin \sqrt{\frac{1 - \frac{b^2}{a^2}}{1 - h_{s/r}^2}}$$

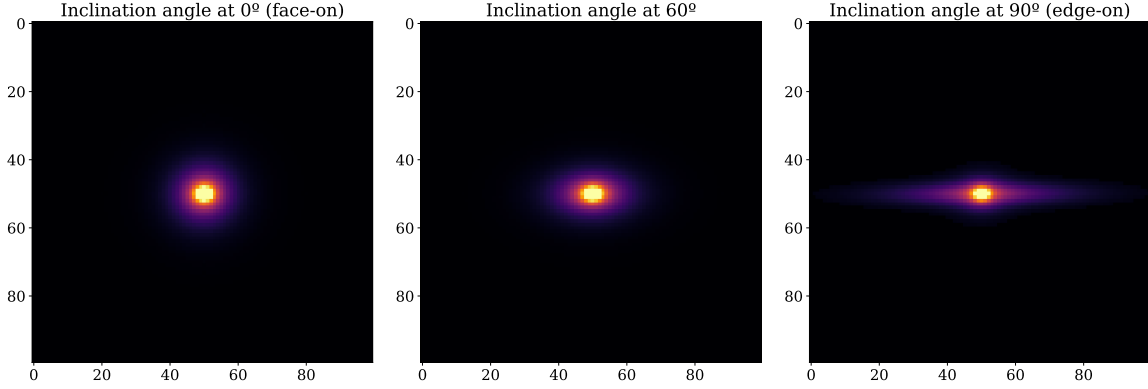


Figure 4.10: A sample of three spiral galaxies at different inclination angles (0, 60 and 90 degrees from left to right). The inclined exponential model better represents the three-dimensional shape of the disk, with a more realistic thickness than the previous Sérsic profile at index 1 used in the past.

The advantage of this profile is that replicates the thickness of the disk, that Sérsic profiles do not properly represent when the galaxy is edge-on. The relative flux of this profile is set to the remaining flux from the bulge fraction. A sample of the current model at various inclination angles can be seen in Figure 4.10

Both bulge and disk models are added and rotated to the corresponding position angle on the sky (with a convention of North up, increasing towards East).

4.2.4.3 Gravitational lensing and optical distortions

With both models into a single object we can apply the gravitational lensing effects, which distorts the shape and size of the galaxy.

In Galsim we apply a lens function with the reduced shear g and μ magnification such as:

$$(4.18) \quad g = \frac{\gamma}{1 - \kappa}$$

$$(4.19) \quad \mu = \frac{1}{(1 - \kappa)^2 - (\gamma_1^2 + \gamma_2^2)}$$

where κ is the convergence and γ are the shear parameters provided by the lensing maps in the flagship mock. In this process we shear and magnify the current bulge and disk profiles. Galsim can only compute a reduced shear < 1 even so this is not a limitation as we are mostly interested in the weak lensing regime which falls away from this constraint. Gravitational lensing

preserves surface brightness and the increase in scale due to magnification is compensated by the magnification applied in the incident spectra (explained in §4.2.3).

From the WCS we can infer a 2x2 local Jacobian matrix that represents the optical distortions of the field in a particular position of the focal plane, where we are aiming to render the galaxy. From the Jacobian matrix we derive the shear parameters that distort the galaxy within its size. This is particularly interesting in VIS simulations where the final shape of the galaxy is critical to estimate the weak lensing signal.

4.2.4.4 Flux scaling and truncation

The source flux now scales the profiles such that the integrated flux up to infinity contains all the flux. However, these profiles need to be truncated to limit the size of the stamp that will be added to the pixel image. Galsim allows to define the parameter *folding threshold* to estimate the size of the stamp such that a given amount of flux is preserved inside. One cannot simply scale the flux of the stamp to the total flux as this would increase the true surface brightness. We set a default *folding threshold* value of 99.5% of flux. In most cases, where the Sérsic index is low, the stamp contains almost all the flux as it is concentrated in the center. On high- n Sérsic profiles where the light decays slower, the flux loss can be closer to 0.5%, but in such extended profiles it will be very hard to recover the total light with photometric measurements.

4.2.4.5 PSF convolution

With the final profile at the given flux we can now convolve it with the Point Spread Function (PSF). As performing a full integral in real space of the distorted models with all pixel windows would be too slow, Galsim allows to perform the convolution in Fourier space with an FFT, which results in an accurate and fast computation. It is in the drawing step where the convolution and models are being computed. In the rendering step we provide the local WCS at the galaxy position which defines the pixel scale and projection at the focal plane. By default the profile is centered in the middle of the central pixel. However the true position of the object may lie in a slightly shifted fraction of it. The offset to the central pixel is computed and provided for an accurate positioning of the source.

It is also possible that the PSF is polychromatic and varies in wavelength. The solution we follow is to divide the computation of the stamp into N sections of the wavelength range, each with its corresponding PSF and flux and combine the final stamp as the sum of the individual steps. This provides very accurate PSF convolution based on the spectra of the source, but it multiplies the computing time of this process.

4.2.4.6 Optimization

The computing time of our simulations is driven by the processing of the galaxy stamp and PSF convolution. Therefore optimizing these two steps allows us to perform larger and more

complex simulations. We found three ways to speed up this process: First we enable a mode where only the size of the stamp is provided without all complex calculations, allowing us to quickly discard sources that do not overlap with the detector pixel array. Secondly, we discretize to the first decimal digit of the variable Sérsic index in the galaxy bulges. With this imperceptible simplification many galaxies share the exact same profile and we can reuse the internal 2D profiles generated by Galsim and speed up the rendering process. And finally, we realized that very faint galaxies (below the target magnitude where shear aims to be estimated) could be replaced by simpler models that require fewer computation. These changes have almost no impact on the quality and use of the images and can reduce the computation time by orders of magnitude.

In situations where profiles deliver a very large galaxy stamp (sometimes larger than few thousand pixels in diameter), the memory necessary to build the convolution exceeds the limit established by the environment and raises an exception. In those situations we came out with a down-up sampling iterative process. When the memory limit is reached, we downsample the actual stamp by a factor of two, providing an equivalent WCS with a scale multiplied by the factor. As long as the memory limit is reached we keep downsampling the stamp with additional factors (most situations are already fixed in the first iteration). Once the stamp is successfully rendered, we upsample the image with a bi-quadratic interpolation which resulted to deliver the least residuals in this kind of profiles. Even though some degradation will occur in this down-up sampling process, these are extremely large nearby galaxies with high Sérsic index which do not belong to the targets for the main science in Euclid. A sample of the down-up sampling process can be seen in Figure 4.11.

4.2.4.7 Alternative models

The profiles detailed previously were used for flagship galaxies and high-z objects, with the equivalent parameters. The QSO sources are expected to be point-like and we render them as stars, just with the shape of the PSF. The Strong Lensed objects require more complex rendering that take place out of this library and are introduced from a pre-computed stamp library.

In order to get more realistic profiles we are integrating a machine learning framework (Euclid Collaboration et al., 2021) that can generate complex galaxy stamps learned from real Hubble Space Telescope observations by deep generative models. These galaxy thumbnails include spiral arms and other features that cannot be represented with analytical profiles as the ones we currently use. The advantage of this tool is that it can render those complex stamps from parametrized information such as the one available in the True Universe galaxy catalogue, without major changes in the flagship mock. These realistic stamps will enable the development of more complex algorithms to measure shear and to deblend larger objects.

Another key feature that will affect the lensing measurements is color gradients. This is the effect that galaxies do not have a constant color across its surface and due to PSF chromaticity, the shape of the galaxy would be affected. A simple model to represent that is to apply a different

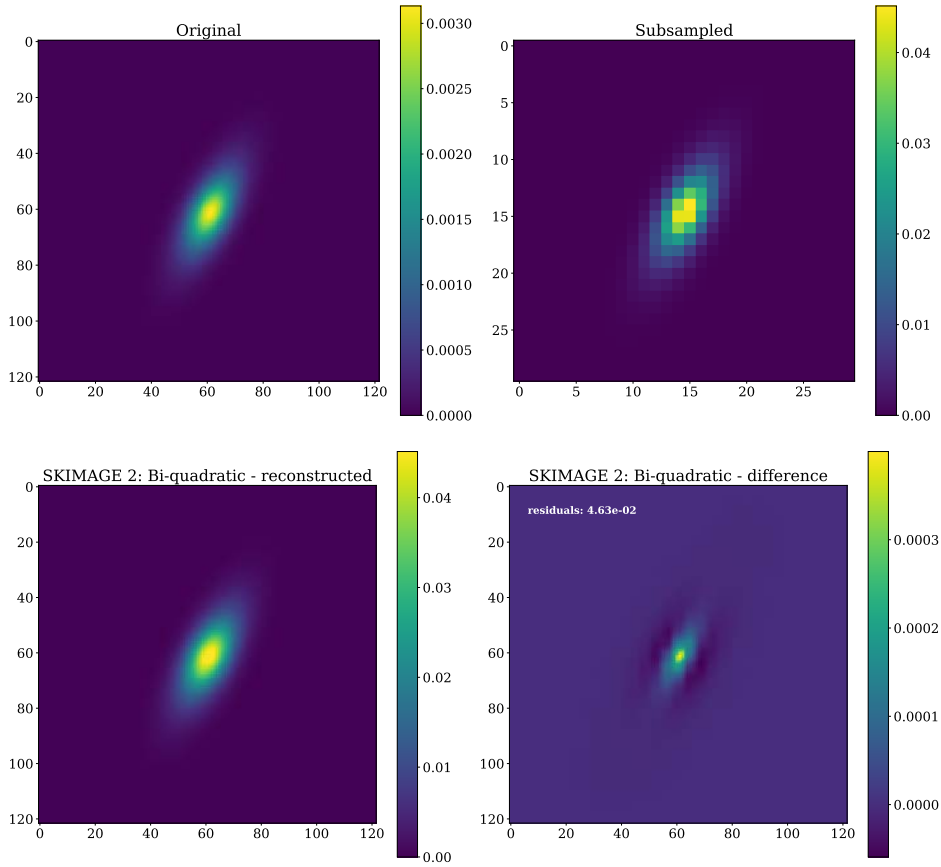


Figure 4.11: The down/up sampling iterative process we use to enable rendering very large galaxies. In this example we show a smaller sample of 120 pixels width for visualization purposes, but the galaxy stamps that require this are in the order of thousands of pixels wide. In this case we show a factor 4 downsampling and how the bi-quadratic 2D interpolation results in a very similar profile to the original and minimizes the residual compared to other interpolation methods.

spectra or extinction to the bulge and disk components and sum the two profiles with their corresponding PSF. The current limitation is to have a realistic model that distinguishes the spectra or color excess from both profiles. Due to the tight shape measurement requirements in Euclid, even this minor effect will also have to enter into our simulation pipeline.

4.3 The mission database

In such a complex mission, having a proper parameter configuration management is essential. From the instrumental models, to reference sky maps, telescope parameters, survey plan, etc., these will all evolve through the different phases of the mission and we need to track what was assumed and when. For this purpose we created a Mission Data Base (MDB) which holds all this

information and tracks the changes of each parameter. This is particularly relevant for OU-SIM as we build all our simulations with models or parameters represented in the MDB.

From the definition phase, through the development, qualification and operation, our knowledge of the mission evolved and this is represented in the MDB parameters. For each model we have set one of the following status

- **as required:** This specifies the limit in which a related requirement tolerates a particular value. This requirement derives from the top-level requirements that ensure Euclid can reach its scientific goals. Typically this is a pessimistic value as it stays in the edge of where the value can be.
- **as designed:** This provides the expected performance of a given model with the current design. It must perform better than the required limit.
- **as built:** At this level the model has been measured in built parts, ideally around the design value.
- **current best estimate - CBE:** This stage replaces the "as built" model with newer or more accurate measurements that contain our best knowledge of the model. Ideally the simulation would be done with all CBE models, which better represent the status of the system.
- **as simulated:** This is a particular configuration for a test, where the model is defined with a given value on the simulation and this value does not particularly relate to a measurement or a design. Even it could be interesting to explore hypothetical models, this is generally avoided to prevent optimizing a system on something that has not been proven.

All versions of the same parameter through its evolution are stored in the MDB. However it is important to identify which of the various representations of the parameter was used at a given time and so we define the Mission Configurations (MC), which gather a subset of the MDB with a unique status on each parameter. Therefore we require to define a Mission Configuration before running any simulation, and the MC used will be attached to the simulation run so anyone can perform its analysis knowing the true models that were used.

4.3.1 The reference survey

The reference survey details how Euclid will map the sky, specifying the set of fields observed that cover the entire wide footprint, the dither steps inside each field, when each pointing is observed and the various calibration fields including the deep fields. This information allows SIM to reproduce the required observation sequence that covers the area we want to simulate. Figure 4.12 shows the fields described in the reference survey that cover the Wide survey footprint.

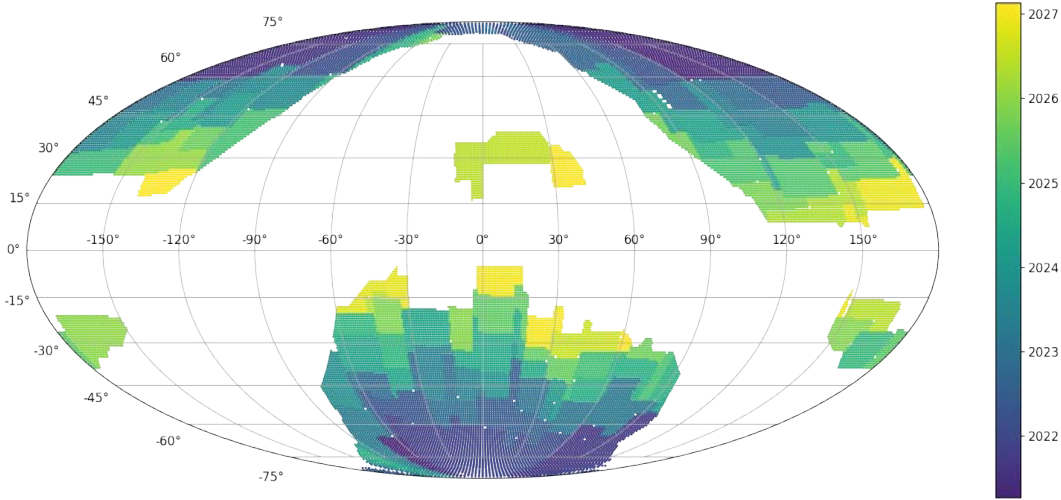


Figure 4.12: The Euclid reference wide survey from 2017 covering 15,000 deg² of the sky in Ecliptic coordinates, avoiding the Galactic and Ecliptic planes due to the high stellar density and high zodiacal light respectively. The color code indicate the observation time where Euclid is planning to point at each of the fields. More recent surveys have been released with further improvements and the updated launch date.

Euclid has defined four dithers inside each field, where each pointing results in a VIS, Y, J and H image, and one spectral exposure. The NISP-S dithers alternate different Grisms at different dispersion angles that enable spectral decontamination. The observing sequence we implemented is shown in Figure 4.13. Simulating a realistic observing sequence is very relevant for instrumental effects such as detector persistence, where charges slowly decay in the detector after reading it, leaving a signal for the next sequences observed whose intensity mostly depends on the interval between exposures.

The multiple exposures allow us to cover the detector gaps, reach the desired depth and correct for transitory effects such as the cosmic rays. We implemented the dither slews following the pointing accuracy model of the spacecraft, with minimal step jumps plus the associated pointing uncertainty of 11". The current model follows an S shape and implements the following minimal slews:

- Dither step 1: $\Delta X_{sc} = 50''$, $\Delta Y_{sc} = 100''$
- Dither step 2: $\Delta X_{sc} = 0''$, $\Delta Y_{sc} = 100''$
- Dither step 3: $\Delta X_{sc} = 50''$, $\Delta Y_{sc} = 100''$

plus a random error of 11" at 3σ after the minimal step. X_{sc} and Y_{sc} correspond to the X and Y axis in the spacecraft reference frame. With the survey ingredients previously explained we can perform realistic observations of Euclid in our simulation pipeline.



Figure 4.13: The reference observing sequence (ROS) in the four dithers for the VIS and NISP channels. At the beginning of each dither VIS and NISP-S observe simultaneously after the Filter Wheel Assembly (FWA) and few seconds to allow the telescope to stabilize. After the exposure, the Grism Wheel Assembly (GWA) lets the light pass for the filter observations of J, H and Y. The first dithers includes an additional short VIS exposure while the last dither ends with a brief flat field frame.

4.3.2 Models and Parameters

The second main block of the Mission Data Base is the models and parameters that allows us to reproduce how Euclid (and External surveys) see the sky. As explained previously, the models in the MDB constantly evolve as we construct and test the instruments and telescope. At this stage of the mission, less than a year before launch, we have a pretty accurate model of everything that could be tested on the ground. As explained previously, we prioritize the models labelled with "Current best estimate" (CBE), as these are closer to the spacecraft that will fly.

The MDB is organized in sections, each covering an important area of the mission. First we have the *Environment* section, containing physical constants such as Electron mass, Astronomical Units in meters or the velocity of light. Even these are publicly known values, it is important that all calculations are done with the exact same values provided by the MDB. Additionally, provides solar system models such as the zodiacal light model and the expected out of field straylight assumed in the definition of the survey.

Then we have the *Space Segment* section, with all known models for the PLM, SVM and VIS and NISP instruments. In the SVM subsection we find models related with the spacecraft such as the dither steps or slew duration. In the PLM subsection we find all telescope related information like wave front errors, collecting surface areas or platescale, as examples. And in the NISP and VIS instrument sections all parameters and models describing the instruments such as detector layout, QE and read noise, CTI models, wavelength dispersion and distortions, etc. We cannot enumerate them all here as there are hundreds of models to describe both instruments. More details are given in the Instrument Simulators section §4.4. However we find interesting to

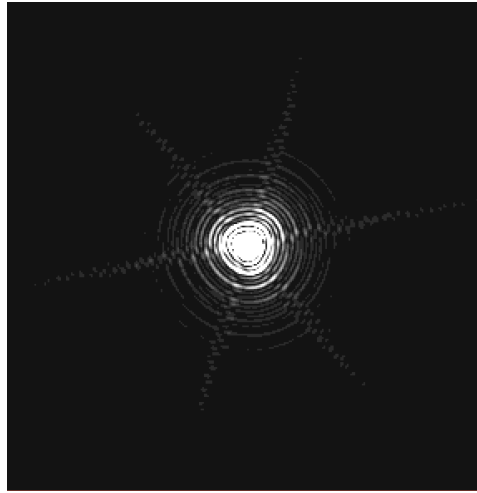


Figure 4.14: A single monochromatic VIS PSF sample out of the 460.000 models that cover the entire focal plane at 250 by 250 positions and in 8 wavelengths along the VIS passband. Each PSF has a resolution of 300x300 pixels sampled every $4\mu\text{m}$ (a third of a VIS pixel)

highlight the largest model in the MDB, the VIS PSF. It is composed of an array of 250 by 250 positions in the focal plane, each at 8 wavelength positions, comprising a total of 460.000 PSFs of 300x300 pixels. These were constructed from the ZEMAX optical model of the telescope (yet without polishing errors nor high-frequency components in the surface). Such large and detailed PSF had to be built to meet the accuracy requirements necessary to measure weak lensing at the target precision. A sample of a single optical VIS PSF can be seen in Figure 4.14.

Next we have the *Survey* section, defining the wide survey described previously, the deep and other calibration surveys and particularly relevant calibration source SEDs such as Vega, a white dwarf or a planetary nebula.

Lastly we included the information of the *External* surveys, as these are necessary for Euclid to reach the photometric redshift accuracy and we include pixel simulation for the planned Ground Based surveys. These contain instrumental and survey parameters for DES, Rubin, CFIS, JEDIS, PanSTARRS and WISHES (more details in §4.4.4).

With all the information provided in the MDB we can now build the instrument simulators which implement the models described here.

4.4 Instrument simulators

In this section we explain our implementation of all 4 instrument simulators: ELViS (Euclid VIS), Imagem (Euclid NISP-P), TIPS (Euclid NISP-S) and SIM-EXT (External Ground Based Surveys). These instrument simulators were originally built for various purposes, such as first evaluations of the instrument performance, but have significantly evolved to become the main Science Ground Segment pixel simulators. They all integrate all the parts previously described such as the True

Universe star and galaxy catalogues, the common SIMSPECTRA and SIMTHUMBNAILS libraries, the reference survey and the particular models from the MDB that describe the instrument it simulate.

4.4.1 VIS Simulator

ELVIS is the Euclid VIS simulator capable of accurately reproducing how the VIS instrument will see the sky through Euclid. The biggest challenge of this simulator is to reproduce the optical response of the system and instrument with the precision required. Euclid plans to measure with unprecedented precision the weak lensing signal of billions of galaxies and to achieve such a task, the VIS instrument needs to deliver an extraordinary good and stable image quality. To be able to prove that the instrument and processing can achieve such task, the simulator is required to deliver images with sufficient precision such that simulation or model limitations do not drive the quality of the image, and it is the instrument performance or the shear measurement algorithms who define how well we can estimate the shape of the galaxies.

Input sources The VIS simulator is capable of rendering all True Universe stars, comprising the real Tycho2 bright stars, the model Besançon stars and the MLT type brown dwarfs. Similarly, it renders the TU galaxies from the Flagship simulation, the high-z quasars and galaxies and the strongly lensed galaxies. Recently we included Solar System Objects (SSO), asteroids that due to their fast and nearby motion, appear as traces in the VIS focal plane. Finally the simulator includes cosmic rays (CR) during the integration and during readout, which leave substantially different trace. CRs are critical in this simulation due to their high density resulting from the higher levels of radiation in space and to the Charge Transfer Inefficiency (CTI - explained later). They significantly affect the image quality and therefore the ability to measure precise lensing parameters.

Background models There are five main contributions to the background in the VIS channel.

- Zodiacal light, a diffused light from scattered sun light in the interplanetary dust. This is based in an MDB model that provides the intensity as a function of ecliptic latitude and its average spectrum.
- Thermal irradiance, a background signal due to radiation from the heat of the various elements of the PLM and instrument.
- Diffuse scattered light, unwanted reflections at the telescope and optical path from the stellar sources in the field that spread into the focal plan resulting in an increase of the background. The intensity of this signal is related with the density of bright stars in the field of view and thus increases at lower galactic latitudes.

- In-field straylight, similarly to the diffuse scattered light, but end up in localized background increase due to very bright stars following a particular path in the optical system.
- Optical reflections, also known as ghosts, an out of focus image of an on-sky source inside or in the immediate vicinity of the instrument field of view. It mainly depends on the light path through the dichroic and telescope baffling and the position and brightness of the sources.

While the first three components are constant across the focal plane and rise its value depending on the telescope thermal state or where we are pointing in the sky, the last two create abrupt changes in the background which complicate the calibration and photometry.

PSF As mentioned, VIS simulations require great care in the reconstruction of the PSF, as it distorts the true shape of the galaxies, that is what we need to measure with great precision to estimate the weak lensing signal. In the VIS channel we have four main contributions that affect the final Point Spread Function.

First we have the optical PSF, which varies as a function of position in the focal plane and wavelength. The model has been described in §4.3.2 and it is the largest parameter in the MDB. The model has so many positions in the focal plane (250x250) because it cannot be interpolated spatially due to its irregular non-predictable pattern. Therefore when we render a particular star or galaxy, we take the closest PSF to its focal plane position. Each PSF is generated from the Zemax model at 8 wavelength positions over the VIS passband, and in this case we can increase the sampling by interpolating up to 1nm. With the reconstructed spectra from the source, we convolve the light profile with the PSF at each interpolated wavelength position, resulting a chromatic PSF that varies as a function of the color for each object.

Secondly we have the Attitude and Orbital Control System (AOCS), the system that the spacecraft uses to stabilize its orbit. This complex system performs tiny corrections to keep the attitude stable though the exposure, and slightly increases the width of the final PSF. The jitter of the AOCS can be replicated as a time series that mimic how the correction behaves. The AOCS jitter can be modeled by a Gaussian of 0.025 arcsec RMS.

The detector degrades the PSF by diffusing the charges in each pixel into its adjacent pixels. This effect can be described in the Pixel Response Function, which gives the distribution of fluxes around each pixel. Due to variations in the CCD thickness, charge diffusion is not constant across the field of view.

A final effect is the consequence of the distortions of the electric field by the charges accumulated in the CCD during the exposure. The pixel boundaries become dynamic as overfilled pixels become smaller than its neighbours, causing brighter sources to appear larger than fainter which end up more concentrated. This is known as the brighter-fatter effect.

The combination of all four PSF components result in a complex and variable PSF across the VIS field of view.

Optics and system The combined optical system of the PLM plus VIS instrument is accurately reproduced in ELViS. First we have the response of the optical system which comprises the reflection of the mirrors and the transmission of the dichroic element. The PLM not only affects the transmission but its geometry distorts the incoming light beam, modifying the plate scale as a function of focal plane position. The results into objects being displaced from its nominal position and a non-uniform illumination of the VIS array. The geometric distortions affect also the shape of the galaxies which require to be adjusted in the rendering process with SimThumbnails.

Furthermore, we reproduced how the detector array is mounted on the focal plane, with accurate gaps, misalignments and rotations as measured in the ground tests. This will add further realism and complexity in the astrometric solutions derived in the calibration process.

Additionally, we included the calibration unit with its highly non-flat illumination to replicate the flat field calibration images. Other subtle effects are also included such as the effect of variable illumination when the shutter opens and closes or the trails caused by the open shutter during the readout process, as sources keep integrating light along the detector columns while charges are being shifted.

Detectors The 36 e2v VIS back-illuminated, red-enhanced CCD detectors will provide a great dynamic range at a precise spatial resolution with their $12\mu\text{m}$ squared pixels. However, even though these are state of the art detectors, there are numerous effects that degrade the image and need to be accurately simulated.

The response of the pixels is not uniform due to variations in the quantum efficiency during the construction process and intrinsic to the material. This is known as Photon Response Non Uniformity or PRNU, and we can obtain realistic PRNU maps from the ground calibration campaigns. Additional cosmetic defects from the process of construction or dust are also included. Furthermore, each pixel does not respond uniformly across its square area, being the center of the pixel more responsive than the corners, effect that we call intra-pixel quantum efficiency. When the number of electrons exceed the full well (the limit of electrons that a pixel can hold), charges overflow in the readout direction into the neighbouring pixels, causing the effect of bleeding. The effect of Dark Current is also included with dependence of the VIS focal plane temperature.

Finally a last detector effect is the Charge Transfer Inefficiency (CTI) which creates traces in the readout direction by electrons that remain in the previous pixel. This effect increases as the radiation from space damages the detectors with time, being significantly worse at the end of the mission (and the most relevant degradation that limits the lifetime of the VIS instrument). We implemented a model described in Massey et al. (2014).

Electronics There are additional instrumental characteristics due to the electronics such as

As any optical CCD detector, a bias voltage is applied before the Analog to Digital Converter (ADC) to step up the signal moving away from the 0 value. This requires models for the electronic bias, the ADC gain conversion and proper simulation of the prescan and overscan sections. An



Figure 4.15: A summary of all the sources and instrumental effects in the VIS simulator.

additional complication is that the amplifier does not respond equally at all flux intensities, of about 1% in the lower regime and a 5% when getting closer to the ADC saturation. The ADC has a limited number of bits to sample the charges in each pixel, causing the discretization of the signal and setting the highest value at which the converter saturates. In the case of the electronics for Euclid with 16 bits per pixel, it is set to $2^{16} - 1$ or 65535 ADU. In the process of reading the sensor, the electronics add scatter in the measured values, which vary from detector to detector, typically around $5 e^-$ RMS. As all 36 detectors are being read at the same time, the induction of the current from the wires that transport the signal can produce reflected images on neighbouring amplifiers or detectors (known as electronic crosstalk).

A summary of all sources and instrumental effects simulated in the VIS simulator can be seen in Figure 4.15 and a sample image in Figure 4.16.

4.4.2 NISP-P Simulator

The NISP Photometric channel is simulated with IMAGEM which reproduces the near-infrared imaging in the three broad bands of the instrument (Y, J and H). Being the primary source of information for the photometric redshifts in the infrared bands, it is important to accurately simulate the flux of the sources and any possible instrumental effects that could bias the photo-

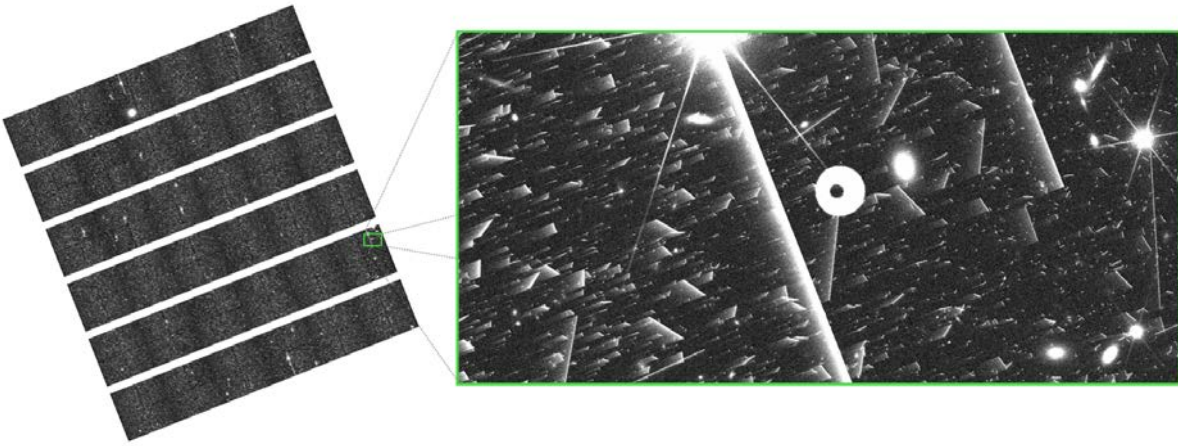


Figure 4.16: A simulated VIS exposure. The left image shows the entire 6x6 mosaic of detectors and the right image contains a zoomed in fraction of a quadrant, where the multiple effects can be seen such as bright stars, galaxies, cosmic rays, a complex PSF model, a ghost from a the nearby star, saturation bleeding and the trails of the charge transfer inefficiency increasingly visible in the readout direction.

metric intensity and color measurements. Below we highlight the main ingredients included in the NISP-P simulator.

Sources In addition to the sources already included in the VIS simulator in §4.4.1, we simulate two standard stars and a planetary nebulae. The standard stars are sources with a particularly smooth spectrum, ideal for photometric calibrations where synthetic photometry can be estimated accurately. We used the spectrum of Vega and a white dwarf, scaled to magnitude 17.5 in the Y band, which delivers a good signal to noise without getting close to the saturation or non-linear regime of the detectors. These standard stars are simulated in addition to the rest of the TU sources, at 5 locations in each detector, providing calibration points through the whole NISP focal plane. The planetary nebula aims to calibrate the NISP-S simulations and will be explained in the next section.

Background The main components of the background are Zodiacal light, thermal irradiance, out of field diffused straylight, in-field localized straylight and reflection ghosts. Even though these have been already explained in the VIS section, it is important to mention that due to the redder spectrum of the zodiacal light, it becomes a much more intense source of background for NISP.

There is an additional component affecting only the background of the NISP instrument. It is the effect of trapped charges that are released on later acquisitions of the image, which we call Persistence. The intensity of the persistence signal $I_{persistence}$ in the next image can be modelled as a double decay:

$$(4.20) \quad I_{persistence} = I_{-1}(p_{slow}e^{-t\sigma_{slow}} + p_{fast}e^{-t\sigma_{fast}})$$

where I_{-1} is the previous image and p and σ are the decay parameters for a faster and slower exponential functions for the persistence. The p_{fast} holds approximately 80% of the signal while p_{slow} holds the remaining 20%. The decay times are ~ 10 times longer for the σ_{slow} . The longer decay, even fainter, can be detected up to two hours after the signal was integrated. As a result, images contain in its background past signal from both photometric and spectroscopic channels.

PSF The Point Spread Function contains similar effects as the one explained for the VIS instrument, although with its less strict requirements, it can be simplified. Just a single model at the center of each detector is sufficient to represent the spatial variations. It is chromatic with 3 positions in wavelength, that are integrated with the incident spectra of each true universe source. It is also convolved with the AOCS jitter and the detector PSF.

Optics In addition to the common optical effects like filter and PLM transmission, FPA metrology or geometric distortions, it implements the effect of the filter wheel tilt. This is the inaccuracy of the filter to be positioned always at the exact position, resulting in a small angle error. This impacts the final distortions which are partly affected by the optical power of the filter element.

Similarly to the calibration unit of VIS, the NISP instrument can also illuminate its focal plane with a set of Luminescent Electronic Diodes (LED) that are set at different intensities to produce flat fields, which we also simulate.

Detector The 4x4 set of Teledyne H2RG infrared detectors are simulated with measured quantum efficiency and dark current maps from ground tests. We also include an oversampled map of intra pixel quantum efficiency to model the internal pixel response function. At the detector level we also include the Poisson noise from the incident sources and background.

Electronics The readout of the H2RG detectors follows a particular procedure (Kubik et al., 2016) called multiple accumulated sampling (MACC), where the incoming signal is sampled multiple times during the exposure. The sampling is divided in frames, which are accumulated in groups, and between groups frames are dropped. In this configuration we can estimate the ramp of the signal in each pixel, as well as the quality of the fit to the ramp, allowing to identify pixels which do not follow a stable rate (like cosmic rays that will produce a sudden bump in the accumulated flux). The whole readout process, including the intermediate frames are simulated in Imagem, providing a very accurate reproduction of the image and the associated on-board flagging of the fits to the ramps (also transmitted as an additional image layer).

A summary of all sources and instrumental effects simulated in the NISP-P simulator can be seen in Figure 4.17 and a sample image in Figure 4.18.

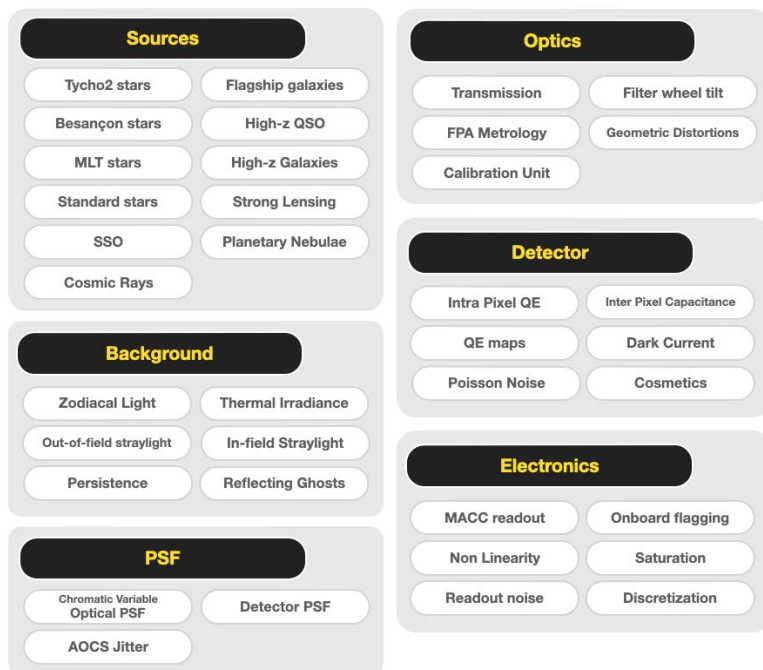


Figure 4.17: A summary of all the sources and instrumental effects in the NISP-P simulator.

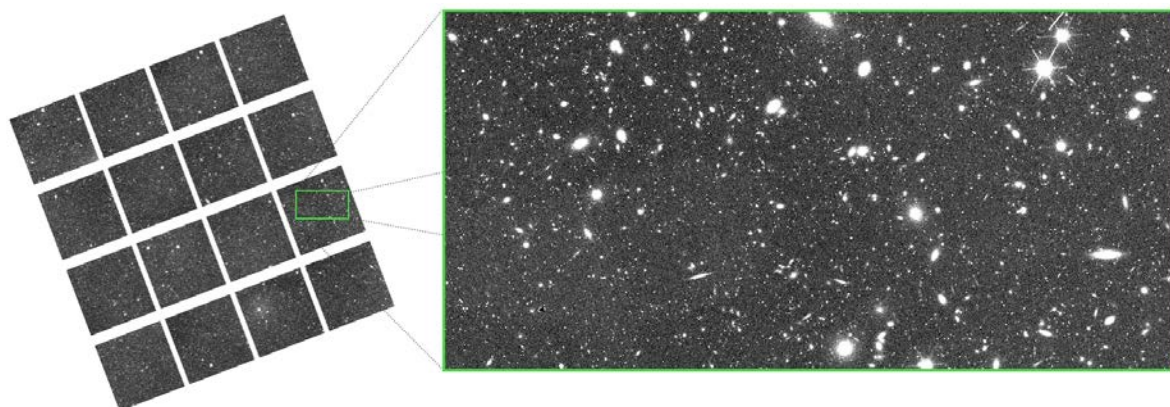


Figure 4.18: A simulated NISP-P exposure in the H-band. The left image shows the entire 4x4 mosaic of infrared detectors and the right image contains a zoomed in fraction where the most relevant elements of the simulation can be seen.

4.4.3 NISP-S Simulator

The NIPS Spectroscopic simulator is implemented in TIPS (Zoubian et al., 2014), a program built to simulate the slitless spectroscopy images from the NISP instrument. It is based in the more generic simulator AXESIM which performs the core spectra dispersion and convolution with the chromatic PSF.

Common sources and models The NISP-S simulator implements the same sources, background models, PSF, optics, detector and electronics as the NISP-P simulator described in §4.4.2. Consistency across the two NISP channels is particularly important as the spectroscopic analysis relies on measurements from the imaging channel.

There are few aspects that are specific in NISP-S and that TIPS has included in addition to the large set of common effects with the NISP-P simulator.

Spectral distortions The simulation of the slitless spectra requires a complete characterisation of the trace dispersion, with sensitivity, diffraction coefficients and PSFs at each dispersion order (from its Taylor expansion). The dispersion order of interest is the first, where the main spectrum can be recovered. However a constant 0th and a 2nd order dispersions need to be simulated as well, as they provide reference points and contamination in the image. While the average dispersion is 700, 13.4 and 6.7 Å/pixel for the 0th, 1st and 2nd order dispersions respectively, the wavelength steps $\Delta\lambda$ are not constant across the field of view. Thus we included in the simulation spectral distortions, a spatially varying $\Delta\lambda$. This significantly complicates the extraction of the spectra, which needs to be calibrated precisely to enable accurate spectroscopic redshifts.

An efficient way to calibrate the wavelength solution is with a Planetary Nebula, with very precise emission lines that serve as calibration points for the varying spectral dispersion. NISP-S simulations are now delivered with Planetary Nebulae sources as calibration frames.

Grism tilt Similarly to the filter wheel tilt effect described in the photometric channel, the grism wheel also has a limited positioning accuracy with an error of 0.1 degrees RMS. In this case the positioning error translates into a slightly tilted dispersed spectra with respect to the its nominal angle. Due to an optical issue in the red grism 270, only the red grisms 000 and 180 could be used. An alternative configuration was proposed that allows us to compute an effective decontamination procedure. The suggested solution implied setting a tilt of $+4^\circ$ and -4° in dithers 3 and 4 (as shown in Figure 4.13) in a sequence of 0° , 184° , -4° and 180° orientation of the dispersion direction. The adjustable dispersion angle was implemented in TIPS, which provided essential simulations in the optimization of the alternative configuration.

A summary of all sources and instrumental effects simulated in the NISP-S simulator can be seen in Figure 4.19 and a sample image in Figure 4.20.

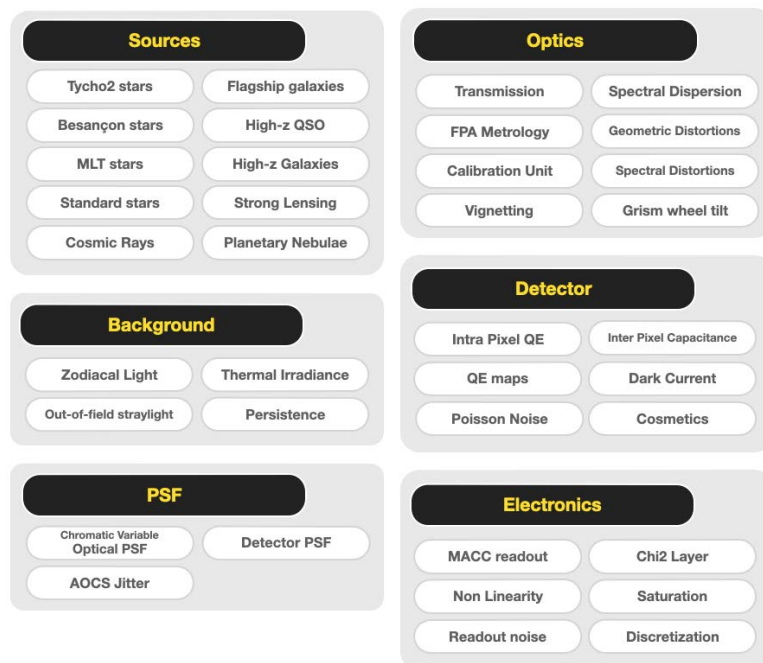


Figure 4.19: A summary of all the sources and instrumental effects in the NISP-S simulator.

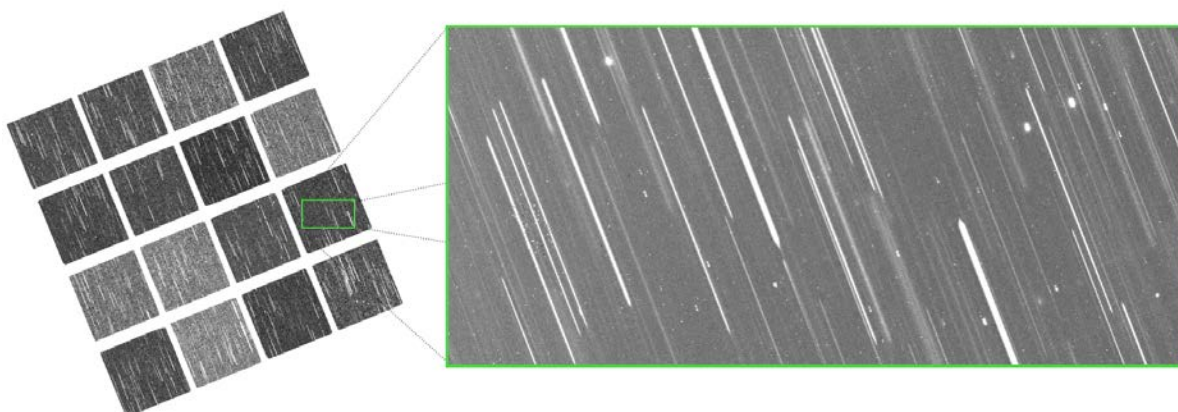


Figure 4.20: A simulated NISP-S exposure. The left image shows the entire 4x4 mosaic of infrared detectors and the right image contains a zoomed in fraction of a single detector, showing a typical slit-less spectroscopy of the NISP-S with 0th, 1st and 2nd order dispersed spectra.

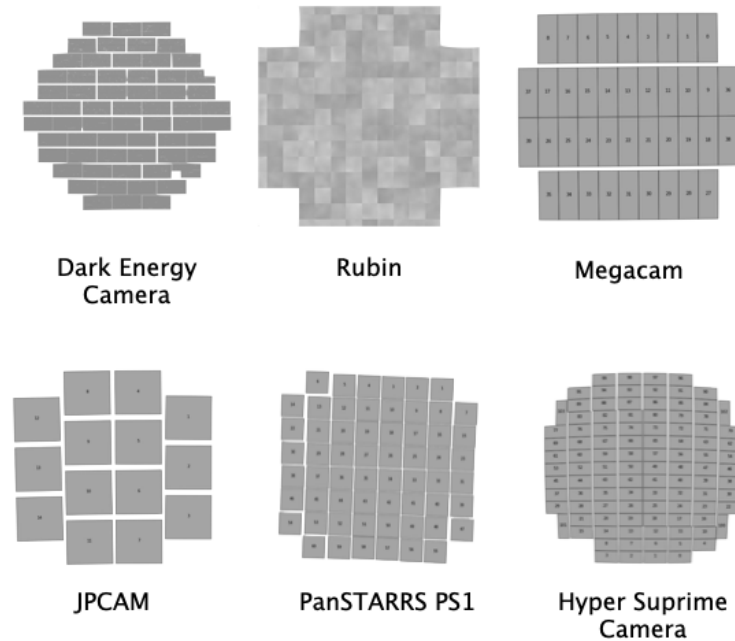


Figure 4.21: The six ground-based external surveys simulated by SIM-EXT.

4.4.4 EXT Ground Based Simulator

Lastly we have the Ground-Based simulator, called SIMEXT, capable of producing accurate pixel data for various Earth telescope + instrument systems. As explained before, Euclid relies on external optical photometry, complementing its own infrared bands in order to get accurate photometric redshifts. Earth observatories are not capable of seeing all the sky as Euclid does and so we require surveys in both North and South hemispheres that overlap with all the Euclid Wide survey. The Euclid Consortium reached an agreement with the Dark Energy Survey (DES) and the Rubin Observatory to provide all ugriz photometry in the southern sky. In the north UNIONS provide the necessary photometry, a joint effort between the Canada-France Imaging Survey (CFIS) for the g and r bands, the Javalambre-Euclid Deep Imaging Survey (JEDIS) in g band, the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) in the i band and the Wide Imaging with Subaru HSC of the Euclid Sky (WISHES) for the z-band. The focal plane of all six EXT surveys is shown in Figure 4.21.

The goal of SimEXT is not to produce extremely detailed simulations as the other Euclid instrument simulators. Instead we only need realistic imaging data that can flow inside the ground segment pipeline with the same True Universe inputs as the other Euclid channels. In this way we can perform joint analysis using the main Euclid pipeline with external data that should deliver equivalent performance as the real observations.

Sources The external simulator is capable of rendering all True Universe sources such as real and model stars, flagship and primordial universe galaxies and high- z and strong lensed sources. Using the same SimThumbnails as the Euclid simulators, is capable of rendering the exact same light profiles as the rest, including weak lensing distortions. As the PSF is monochromatic, it does not require the input spectra and uses the precomputed band fluxes, which speedup the computation of these simulations.

Background and Transmission In most cases, external surveys have already started their observations. Therefore we have realistic background magnitudes and zeropoints from the reduced image analysis. We used those real measurements as inputs to the simulation, which provide a very realistic distribution of sky brightness and atmospheric extinction. In the case of surveys that have not started their observations, we can model realistic distributions from the available models of sky, passbands and environmental conditions.

PSF The Point Spread Function is significantly simpler in most EXT simulations, with a simple Gaussian or Moffat profile. From a realistic distribution of seeing FWHM, we can replicate the average optical quality for each survey. In the case of DES we had a large sample of PSF models extracted from real data analysis using PSFEx (Bertin, 2011). We randomly select a model, enabling variable and complex PSF simulations for the ground based photometry.

Optics and Detectors In most surveys, we deliver a perfectly calibrated WCS, as the reference solution matches the one we deliver in the FITS header, requiring no further astrometric calibrations. However and similarly to the approach we followed with the real PSF modelling, we obtained the WCS solutions obtained with the astrometry tool SCAMP (Bertin, 2006) and used these to render the objects in the focal plane. The FITS headers were filled with the default WCS, which mocks a proper optical distortion simulation.

Additionally, we used master bias and master flats from the actual observations to include a realistic pixel-to-pixel variation and faithful readout patterns. With the same model we could create individual flat field and bias calibration images that had their own noise realization.

A summary of all sources and instrumental effects simulated in the SIMEXT simulator can be seen in Figure 4.22 and a sample image in Figure 4.23.

With all four instrument simulators presented in this section we can produce complex and detailed simulations in a large variety of channels and configurations under the same universe, environment and shared models.

4.5 Operation and production of simulations

Now that we described all ingredients in the simulation pipeline it is time to put them together to produce mock observation sets of data that can be processed and analyzed. In this section we

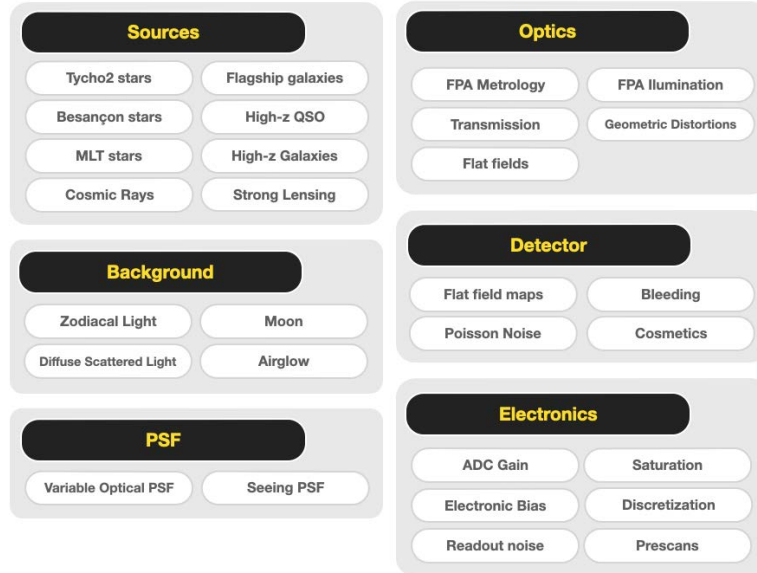


Figure 4.22: A summary of all the sources and instrumental effects in the EXT simulator.

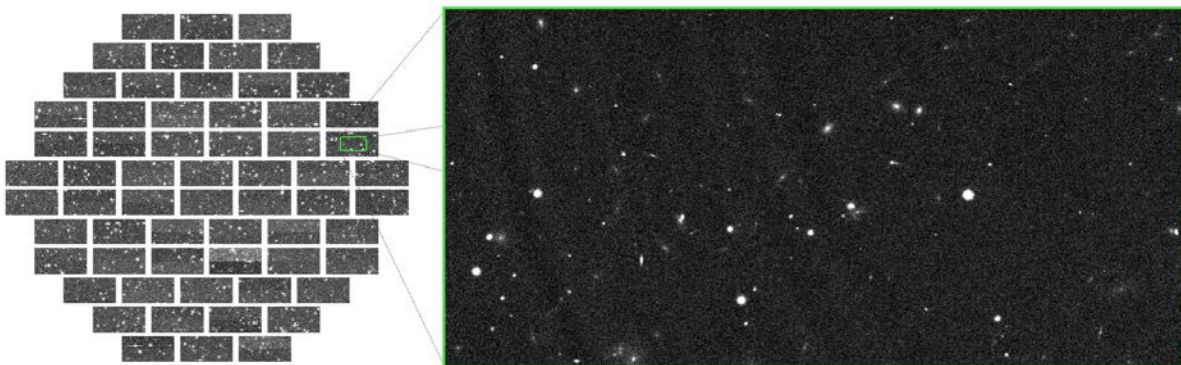


Figure 4.23: A simulated DES exposure in the r band. The left image shows the entire 62 detector mosaic and the right image contains a zoomed in fraction of a single detector, showing all TU sources and various detailed instrumental effects such as variable PSF model and astrometric distortions from real observations and real QE variations and cosmetic defects from its master flat field.

describe the Sim Planner, a tool to orchestrate production of simulations, the infrastructure of the Science Data Centers where simulations will run and the various sets of simulations we have produced up to date.

4.5.1 Infrastructure and Science Data Centers

The massive data processing required in the Euclid Science Ground Segment (including simulations) takes place in a distributed system across multiple Science Data Centers (SDCs). Even though most SDCs have a High Throughput Computing (HTC) design, the underlying infrastructure differs from one to each other. It is then essential to have common tools and rules such that development, storage and computing happen smoothly independent of where it happens.

Here we highlight the most relevant elements in the common infrastructure system in which we had to adapt our simulation pipeline to operate in the SGS:

- The **Euclid Archive System (EAS)**, containing all data and metadata processed in Euclid, it allows to index, search and access any information. It contains all images, catalogues, processing status, etc..
- **Distributed Storage System (DSS)**, allows to have the EAS across the distributed local storages in the SDCs. It manages the necessary transfers to have the data in the place when is needed.
- A **Common Data model** establishes the format of the data and metadata, such that interfaces between pipelines and with the archive are stable.
- An **Infrastructure Abstraction Layer (IAL)** orchestrates the data processing and adds a common layer such that jobs can run independently of the underlying IT infrastructure. It defines the Pipeline Processing Order (PPO) which describes the elements, configuration and inputs of a specific processing task.
- A common **Euclid Development Environment (EDEN)** establishes the set of libraries and its versions to be used by any Euclid software. This prevents inconsistencies or changes in behaviour of different libraries between development and production or at different pipelines.

With the previously described pipelines we can operate smoothly at multiple SDCs, with the advantages in computing power and storage that it enables. On the contrary, adapting a pipeline to run under the SGS infrastructure is a massive amount of work, with major changes in the design of the pipeline and a fix data model and environment that slowly changes, limiting rapid development of new features.

4.5.2 The Sim Planner

With all the ingredients to build the simulation we need a tool that can orchestrate large sets of simulations, covering particular areas on the sky and with a given configuration. For this purpose we build the SIMPLANNER, which from a given Simulation Request, it produces the corresponding PPOs that can be run under any SGS infrastructure. The simulation requests will specify aspects like channel to be simulated, area, input true universe catalogues, mission configuration and any other specific parameters relevant to the simulation. The SimPlanner then interprets the request and configures the corresponding simulator with the appropriate configuration.

Before a simulation can be produced, the True Universe catalogues that overlap the target area will need to be produced. The production of the catalogues is a more manual process that even takes place outside the SGS, like the generation of the Flagship mocks. However these need to be ingested and registered in the EAS to be able to use them in the simulation and in later processing for validation. A last step prior to start the simulation is to define the Mission Configuration with a set of parameters from the MDB.

With all necessary inputs and the simulation request, we can run the SimPlanner which produces the PPOs that can be launched to the IAL, and later processed at the various data centers. Once the processing is complete, the simulated frames are registered and ingested to the EAS, allowing the rest of the processing to access and continue its analysis, as if Euclid data would be. An overview of the SIM Pipeline orchestrated by the SimPlanner is shown in Figure 4.24.

The SimPlanner has implemented additional features such as implementing particular dithers from the reference survey (when they were not available for the deep fields for instance) or enabling complex parameter exploration, configuration of flat field sequences and other high-level tasks that can be done at this level.

4.5.3 Scientific Challenges

Even though some groups at the SGS can request simulations for a particular purpose (such as develop a new algorithm), the production of simulations is generally planned in advance in orchestrated tests that we call Scientific Challenges (SC). These exercises have been extremely valuable to incrementally develop the SGS pipeline and validate the requirements in each Organization Unit during the preparation of the mission. Here we give an overview of the first eight challenges with the goals and associated simulations required in each.

SC1 (2015) The first scientific challenge were only SIM took part in it. Its main goal was to test a very early SGS infrastructure and tools with the simulation pipeline only, mostly for integration purposes. The simulated VIS, NISP-P and NISP-S channels had few instrumental effects and only stars were included. A single Euclid FoV was simulated.

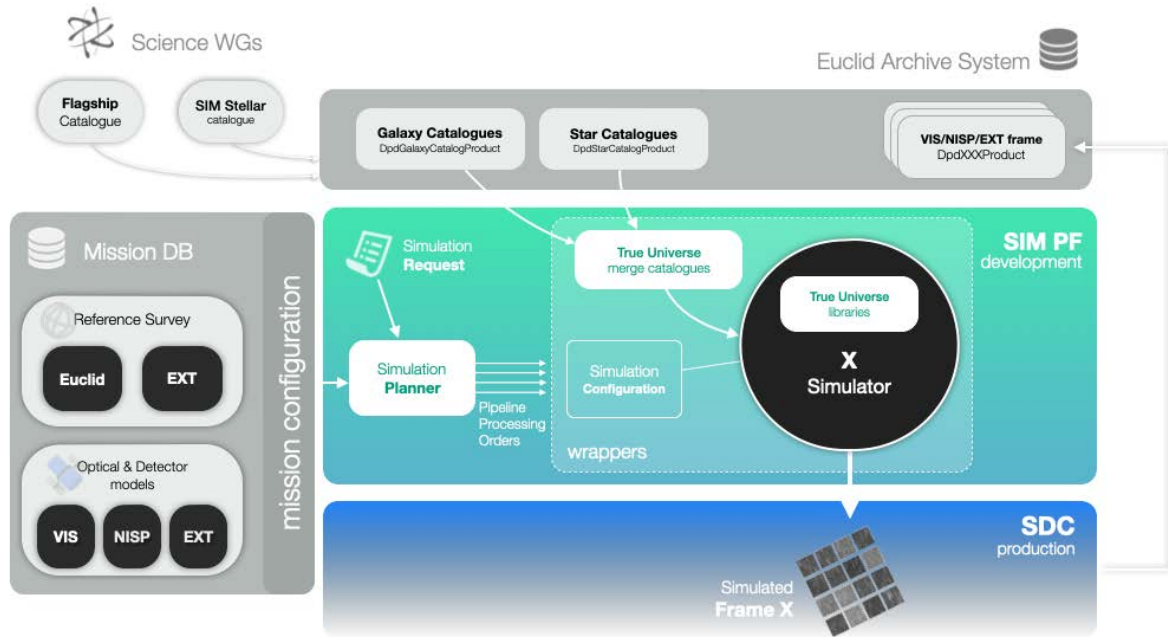


Figure 4.24: An overview of the SIM pipeline, orchestrated by the SimPlanner. The input catalogues which provide from the Science working groups are ingested in the Euclid Archive System, as well as the output products of SIM. With the models from the MBD (under a given mission configuration), the input TU catalogues and a simulation request, the SimPlanner produces the corresponding Pipeline Processing Orders (PPOs) that are executed under the distributed SDC infrastructure.

SC2 (2016) In the second exercise, three new OUs entered into the game, VIS (VIS calibration), NIR (NISP-P calibration) and SIR (NISP-S calibration and extraction of 1D spectra). Thus, the challenge consisted in SIM producing simulations in the three Euclid channels and the OUs processing the simulated data. Galaxies were now available in the 11 fields of Euclid we simulated. Additional complexity in all simulations was included such a CTI in the VIS images or geometric distortions in the NISP channels. The MDB was established and used as input.

SC3 (2017) With EXT (External calibration) and MER (Merging and photometry) participating for the first time in the Scientific Challenge, we delivered new DES and KIDS simulations as test data for their analysis. The Euclid channels were significantly more complex than in SC2 and thus calibration frames had to be delivered to enable the required processing. The area did not increase with just nine Euclid FoV. With MER combining for the first time data from different channels (VIS, NISP-P and EXT), valuable feedback was provided to SIM capturing issues in the simulation that could be rapidly solved.

SC4,5,6 (2018) This triple challenge aiming to feed data to SHE (shear measurements), PHZ (photo-z estimates) and SPE (spec-z estimates) significantly increased the complexity and area of the input simulation. We included the spatially and wavelength variable PSF in VIS and NISP as well as much denser simulation of the True Universe sources (down to VIS 26). A first deep field was simulated to enable calibration data for SHE as well as a new self-calibration field, where a specific pattern allowed accurate astrometric and photometric calibrations. As NISP-S now included spectral distortions, we had to deliver an additional Planetary Nebula calibration frame so SIR could deal with these wavelength variations. The external data set included an additional survey with Rubin simulations. The simulated wide fields covered a total of 52 Euclid FoV at three different locations, with a large field having average sky conditions and two small fields with alternated high/low stellar density and zodiacal light. With this variation we could evaluate the performance of each OU at different scenarios. Just to simulate 30 deg² of Euclid data, it took 620.000 hours of CPU, generated 22.4 TB of data and rendered 3.6 billion galaxies. This gives an idea of the complexity required to simulate at this precision.

SC7 (2019) With such a large and complex exercise in SC4,5,6, numerous issues were found in the simulation and later analysis. Therefore we decided to keep a similar configuration as in the previous challenge, focusing on fixing known issues and repeat the process with a cleaner set.

SC8 (2020) The last Scientific challenge was another major step up. The only remaining piece of the SGS pipeline LE3 finally entered, enabling a full end-to-end simulation, from the raw images until the scientific analysis. The whole exercise was operated by the IT group, without manual intervention from the OUs. The area was increased up to 270 Euclid fields covering an area of ~ 150 deg². Such increase in area was required by LE3 who could not perform certain analysis on smaller area. True Universe got significant new sources such as QSO, High-z galaxies, Strong Lensing, Brown Dwarfs and standard stars. Simulators also were improved with 23 new and updated models. External data was extended with 4 new surveys covering the northern sky. An overview of the simulated frames produced for this challenge is shown in Figure 4.25. The SC8 simulation is the current reference for the ground segment and provides key information in the validation of requirements and IT operation of the mission.

Simulations produced by SIM have also been essential in identifying solutions for critical problems found during the qualification of the mission, such as the Grism 270 optical issue where our simulations could evaluate and validate alternatives that were finally adopted. Furthermore, the outcome of the Scientific Challenges has proven to be an efficient procedure to evaluate the status and readiness of the ground segment through the various ESA reviews and technical keypoints.

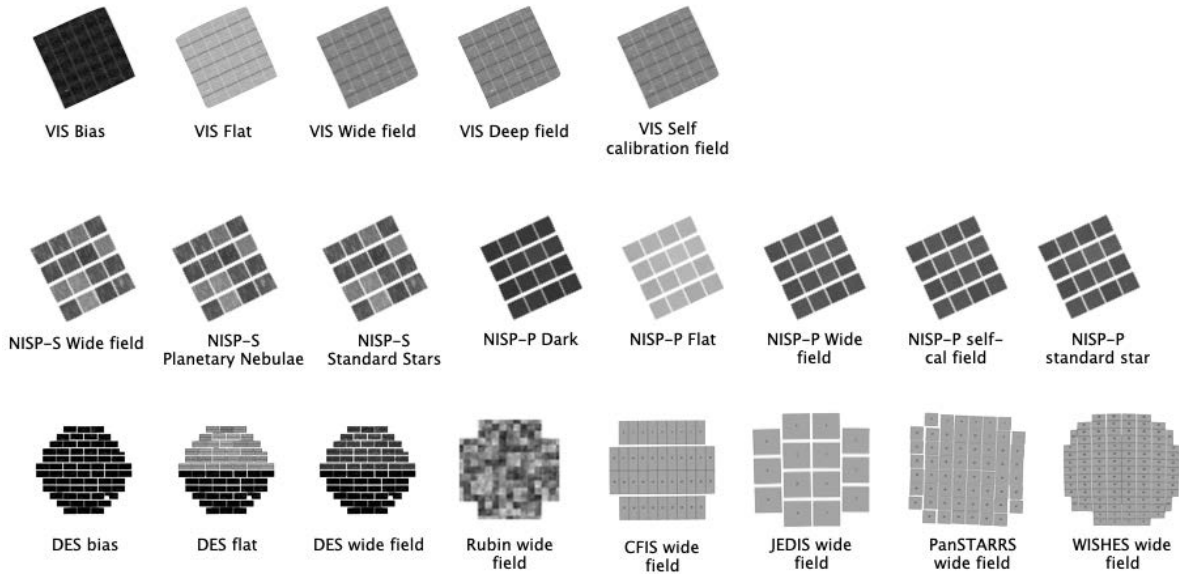


Figure 4.25: The simulation set in SC8, with all Euclid channels and its associated calibration fields. Additionally simulations for all external surveys were produced.

4.6 Summary

In this chapter we have explained our approach to develop a full pixel simulation pipeline for a Space Based mission.

First we had an overview of the complex mission, reviewing both instruments and how the ground segment is organized. We detailed the organization group in charge of implementing the simulations presented here.

As input sources for the simulation, we had an overview of the True Universe. The galaxy mocks are built from the largest cosmology n-body run, the Flagship simulation. Additional galactic sources were added to complement this massive simulation such as high redshift QSOs or strong lensed galaxies. The stars were built from a combination of a very large and deep Besançon model and real all-sky stellar catalogues. To provide consistency across all instrument simulators, we designed libraries to reconstruct the spectra and shape of each galaxy with SimSpectra and SimThumbnails respectively.

Other key ingredients for the simulation are the instrument models and the reference survey, hosted at the mission data base. We explained how we organized the different status of the parameters and how we interface with it.

The last step in the simulation process is the pixel simulators implemented for each instrument. We presented and detailed the four developed by SIM, the VIS, NISP-P, NISP-S and External simulators.

All these pipelines had to be implemented in the Science Ground Segment, an international

distributed set of data centers with different infrastructure components to enable its operation. We explained the Simulation Planner, a tool to orchestrate large sets of simulations under the SGS. We described the eight Scientific Challenges we participated with increasing size and complexity.

The presented approach allows the Euclid mission to perform large and accurate tests to train the ground segment pipeline, validate complex algorithms and estimate the mission performance throughout its construction. We also expect it to enable key studies once the spacecraft flies and sends real observations.

SYNOPSIS AND FINAL CONCLUSIONS

In this thesis we aimed to illustrate how new technologies and accurate simulations enable new possibilities in the analysis of cosmic survey data.

We have studied three different areas of a cosmic survey: the development of three auto-guiding systems at major 4-meter class telescopes, the overall data processing of a cosmology survey and the complex simulations in a space mission aiming to map the universe.

Auto-Guiders First we explored our implementation of the autoguiding systems for the Dark Energy Camera, the Physics of the Accelerating Universe camera and the Dark Energy Spectroscopy Instrument. The experience gained from the first two instruments allowed us to develop the most complex algorithms required for the spectrograph. Thanks to an evolving programming language, higher level functions allowed us to develop simpler codes, which resulted into clearer algorithms. For instance, the integrated centroiding algorithm in the DESI guider allowed us to have better control of the aligning process.

The simulations we performed in all three projects put us in an advanced state at the beginning of the commissioning, with an already working code that could be improved with real observations. Simulations also provided key information to estimate true guide errors, not possible with only sky data. As we were including more detailed and subtle instrumental effects, the guiding performance in the simulated data resembled closer to the one we measured at the telescope.

Mining all observed guide data allowed us to identify subtle patterns such as the field rotation residuals or guider accuracy dependence with declination and hour angle. These were the kind of systematics that could only be detected and corrected with large-scale analysis as the ones proposed in this work.

In all three experiments we verified that we met the tracking performance necessary to meet the higher level scientific requirements. As a result, major cosmological results have been published by the DES and PAU projects that started their survey observations. Many of these results (i.e the weak lensing measurements) could only be achieved with the exquisite guiding performance developed in this work.

Data mining the PAU Survey PAU is an innovative cosmological survey, with a complex data management system. We presented all the astronomy and technical work required to go from the telescope observations to its scientific analysis.

The simulations we designed prior to the beginning of the observations were essential to construct the entire data management system. These simulations allowed the collaboration to be ready to process the large volume of observations produced by PAUCam from the start of the survey. Furthermore, we could start developing and validating custom algorithms such as the particular photometric calibration of the narrow band system. The simulation effort is continuing for the entire survey and data analysis period, being key to understand very subtle effects and develop more accurate algorithms, such as the scattered light correction.

The infrastructure at the data center where all operations were performed, significantly evolved since the beginning of the project. We had to build a flexible system that could adapt to the underlying system changes. Also the astronomy codes had to evolve since the first light, as our understanding of the data improved with the continuous analysis runs.

Every day scientific programming libraries perform more complex and higher level analysis. With a large fraction of the scientific community using, debugging and improving these functions, they become a safe and efficient option to perform higher level astronomy analysis. Thus, we could concentrate on our more specific problem without the need of re-implementing over and over the same code.

The parallelized design of the PAU pipelines enabled quick reprocessing of large volumes of data. This allowed us to get rapid science feedback over the earlier processing steps, which helped us to accept or reject certain changes in our data processing pipeline. At the telescope, the (almost) automatic observing system with adaptive exposure time delivered a homogeneous survey coverage across the whole filter set and fields in the sky.

New emerging technologies like machine and deep learning allowed us to extract information from the data not possible before with more traditional algorithms and with less volume of data to train with. A clear example is the proposed background subtraction with neural networks, which is able to understand the stochastic behaviour of the scattered light enabling more accurate photometry. However not all new technology was appropriate in every situation and "old-style" relational databases as the PostgreSQL implemented for PAU provided an efficient solution, centralizing all survey metadata.

A deep understand of the final scientific goal and the new technologies available has been essential for the design of the data management system. Not all the technical solutions reach the

science goals in the same way, and depending on the chosen path, reaching the goal is simply not possible. We had to bring scientists with technicians together, discussing the problem in the same language (not always straightforward) and continue the development in close interaction to obtain an optimal solution. Thanks to a well established bridge between both teams, we could build CosmoHub that serves as an astronomy data distribution service for hundreds of scientists across the world. Other attractive web-based applications such as the operations monitor or the forced aperture inspector (that could have been done through a terminal window), facilitated the productivity of scientists allowing faster and more pleasant work.

As a result, PAU did and continues to do state-of-the-art science, only possible with the combination of an excellent camera, optimized observations and an efficient and evolving data processing.

Precise simulations for Space Euclid is a large and ambitious space mission. With so many systems, groups and organization units, it requires very complex interfaces (both technical and personal). OU-SIM plays a key role as the main contributor of simulations for the entire ground segment. In such key position, it has interfaces with almost every single group in Euclid. It interacts with the science working groups to feed the cosmological information and Galaxy models to the simulators. It interacts with the VIS and NISP teams as well as with the Payload Module (PLM) manufacturers to obtain the knowledge of the instruments. It interacts with every single Organization Unit in the ground segment to obtain the simulation requirements at their development stage. It interacts with the survey team to understand how Euclid will observe the sky. And it interacts again with the rest of the science working groups to report the possible scientific exploitation of the simulations. A careful and well defined interface had to be set to efficiently communicate with so many groups. This was greatly improved with a recent role of Instrument Model Owners (IMO), an interface person inside each instrument enabling a good communication with the rest of the ground segment.

Thanks to the experience gained with the pixel simulation pipeline built for PAU, we could develop a well defined design from the beginning, with clear and separate areas for True Universe inputs (from simulated stars and galaxy mocks), instrument models and pixel-level simulators. Contrary to the approach in PAU, OU-SIM required multiple channels to be simulated under the same environment. It was critical the design with common tools for spectra and shape, enabling a homogeneous and consistent set of simulations. Furthermore, the common libraries gave flexibility to introduce new sources or more complex light profiles. The long list of instrumental features had to be gradually incorporated into the simulation, in sync with the development of the corresponding counterpart in the reduction pipeline. Little by little we have been implementing the instrumental and environmental effects that cover the extensive list of requirements for all organization units in the ground segment. Even we implemented multiple validation tests in our simulation pipeline to ensure the correctness of the released data, the feedback from the organization units that processed the simulation was very valuable for us.

The scientific challenge exercises, organized almost every year by the collaboration put clear goals to everyone in the science ground segment, motivating scientific and technical developments. The early implementations of OU-SIM under an immature infrastructure resulted very inefficient from our scientific point of view, with development that had to go back and forth multiple times. Once the infrastructure was stabilized in Scientific Challenge 3, the development of the simulators grew at a faster pace. Predictions on future simulations have been hard to plan, as small instrumental changes drastically changed the performance of the simulation codes. For instance, the inclusion of a wavelength dependent PSF for the NISP-S simulator multiplied by 9 the processing time of each detector. Furthermore the instrument simulators grow in sync with the rest of the OUs and with the knowledge of the detectors as they are constructed and tested. This resulted into requests of new effects to be included in the simulation with little notice. Thanks to a flexible simulation design, we could keep the pace and deliver simulations shortly after an instrument model was made available.

The Euclid simulation that we have implemented has produced reference data for the implementation and qualification phase, where all requirements are being checked prior to the start of the mission. We expect more simulations after launch, that will allow us to verify what Euclid sees once real sky data arrives.

Final conclusions As shown in all areas of this study, accurate simulations enable more efficient and secure algorithms, allowing to validate each step in the development process and to better interpret the real data. This is even more true in large cosmic surveys where so much data needs to be processed almost blindly.

New technologies like big data, machine learning or high-level libraries allow to produce more efficient algorithms and pipelines, enabling new possibilities not available before.

Cosmic surveys are reaching new levels of accuracy and new methods are required in all areas, from a more autonomous observing systems to data processing that needs to extract information invisible to previous generation algorithms. A bridge between science and technology is needed to achieve the technological and scientific new challenges that cosmology presents nowadays.



SYNTHETIC SPECTRA SAMPLES

In this appendix we highlight some interesting synthetic spectra samples that illustrate the possibilities of the analysis and validation with this reference data set.

SDSS spectra 0267-51608-0459 - SN: 22.69 z: 0.14

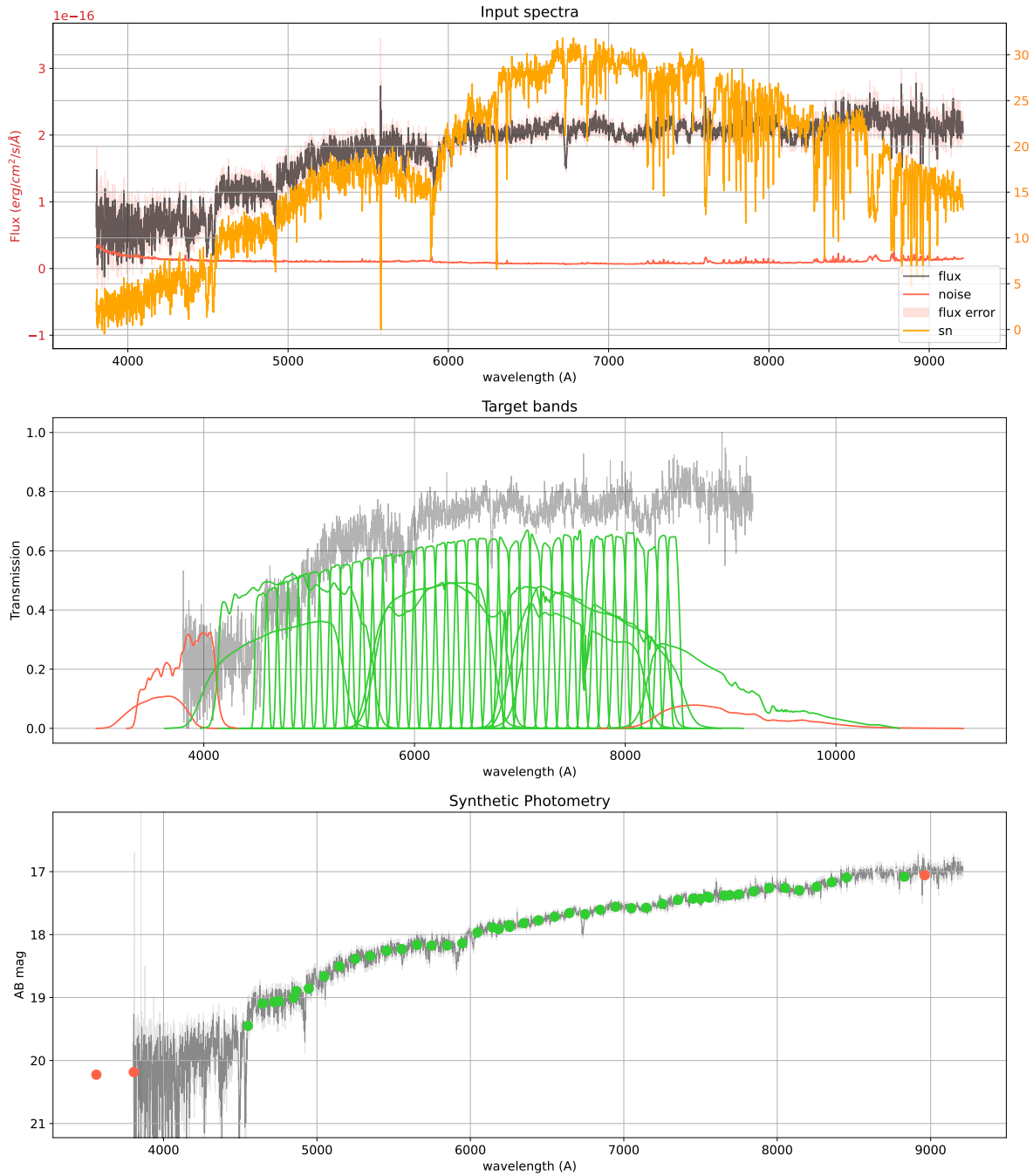


Figure A.1: A typical galaxy representative from the SDSS dataset.

SDSS spectra 3829-55300-0938 - SN: 17.25 z: 3.09

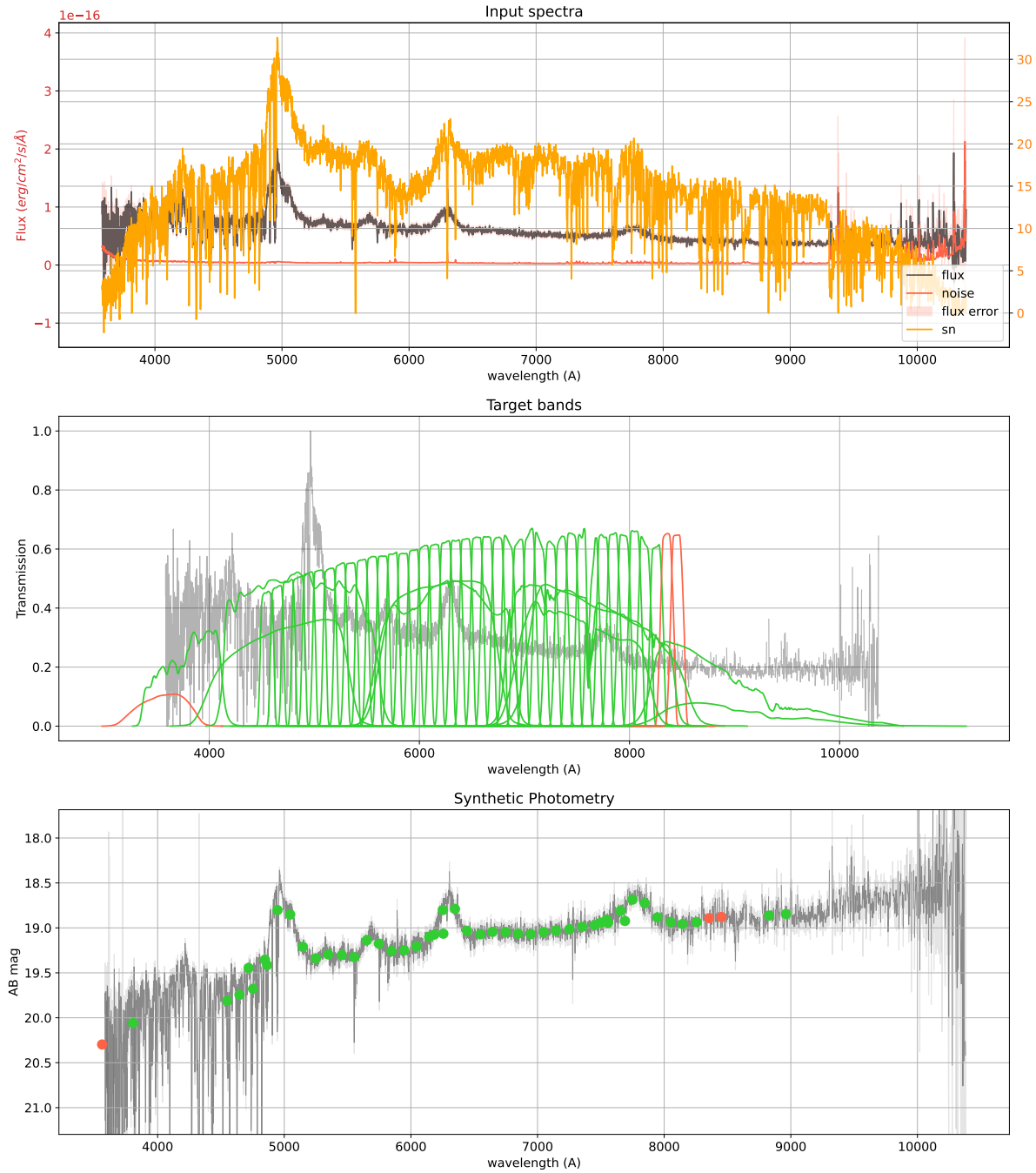


Figure A.2: A high redshift QSO with various broad lines, well sampled by the PAUS narrow band set.

SDSS spectra 3829-55300-0888 - SN: 3.02 z: 4.38

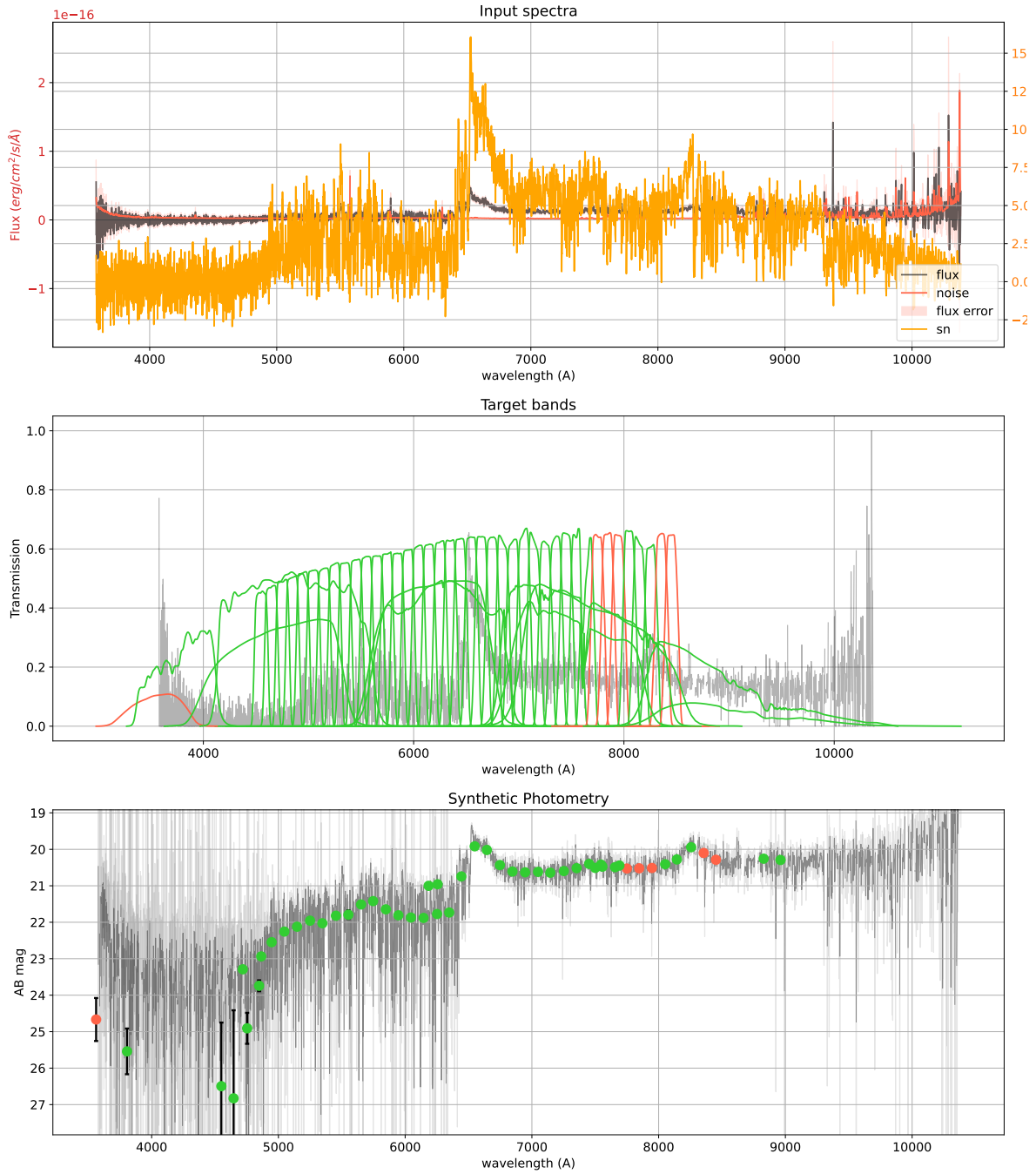


Figure A.3: A very high redshift QSO at redshift 4.38 a clear $L\gamma\alpha$ line in the middle of the PAUS wavelength range.

SDSS spectra 0501-52235-0366 - SN: 58.78 z: 0.02

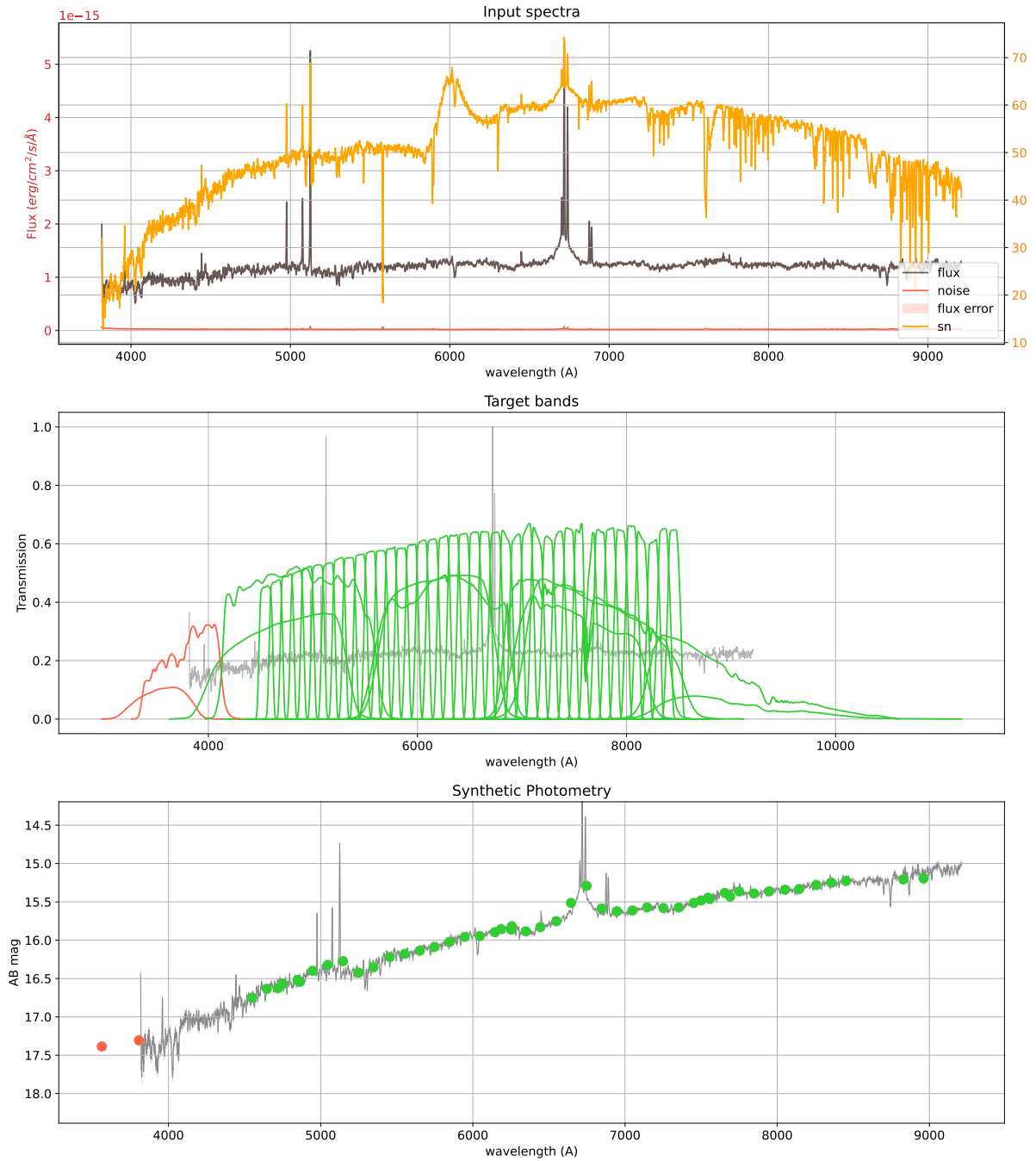


Figure A.4: A nearby bright AGN galaxy resulting into a high SN spectra.

SDSS spectra 2867-54479-0335 - SN: 90.52 z: -0.00

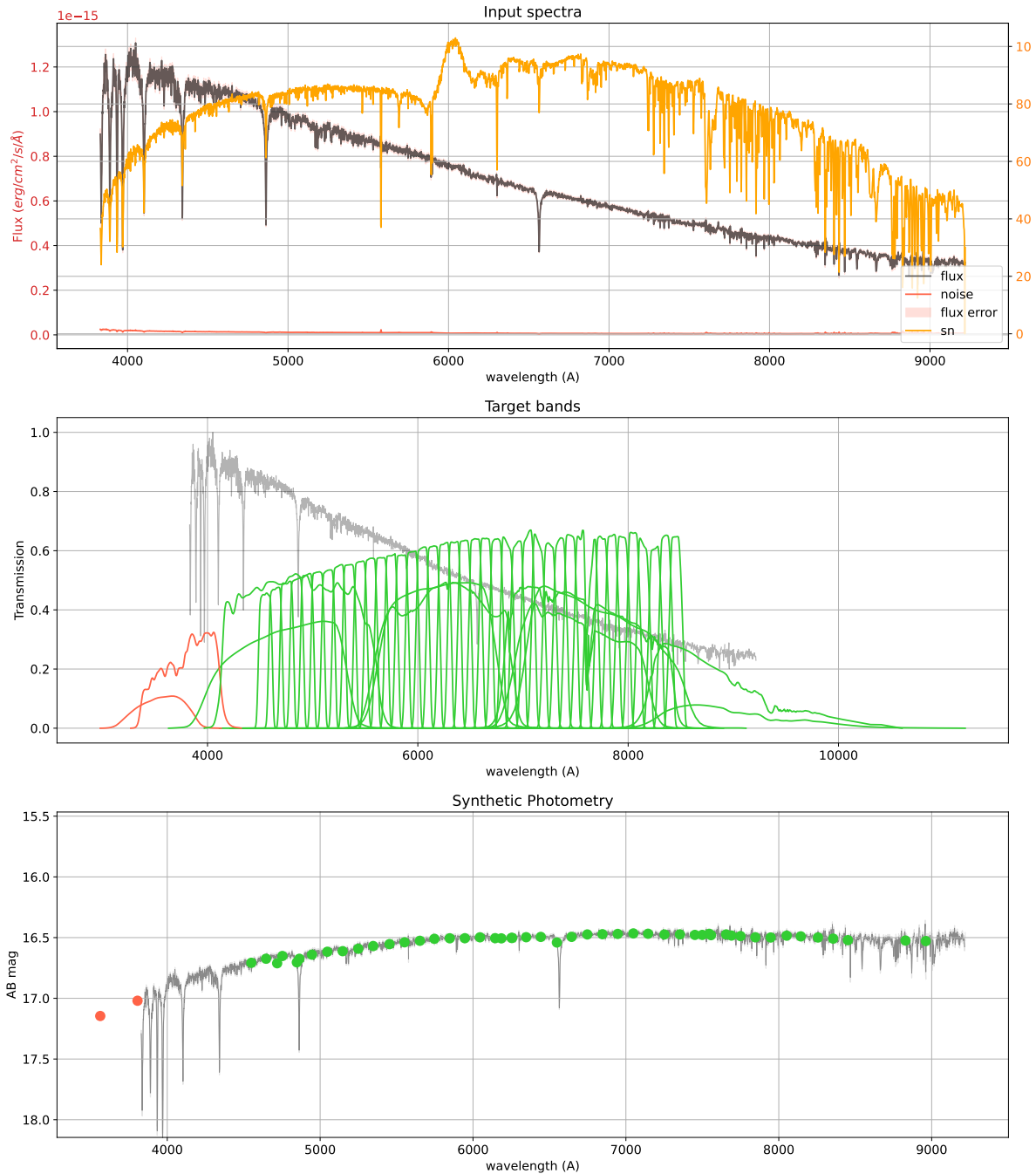


Figure A.5: A bright blue star of class A0 with a smooth spectrum.

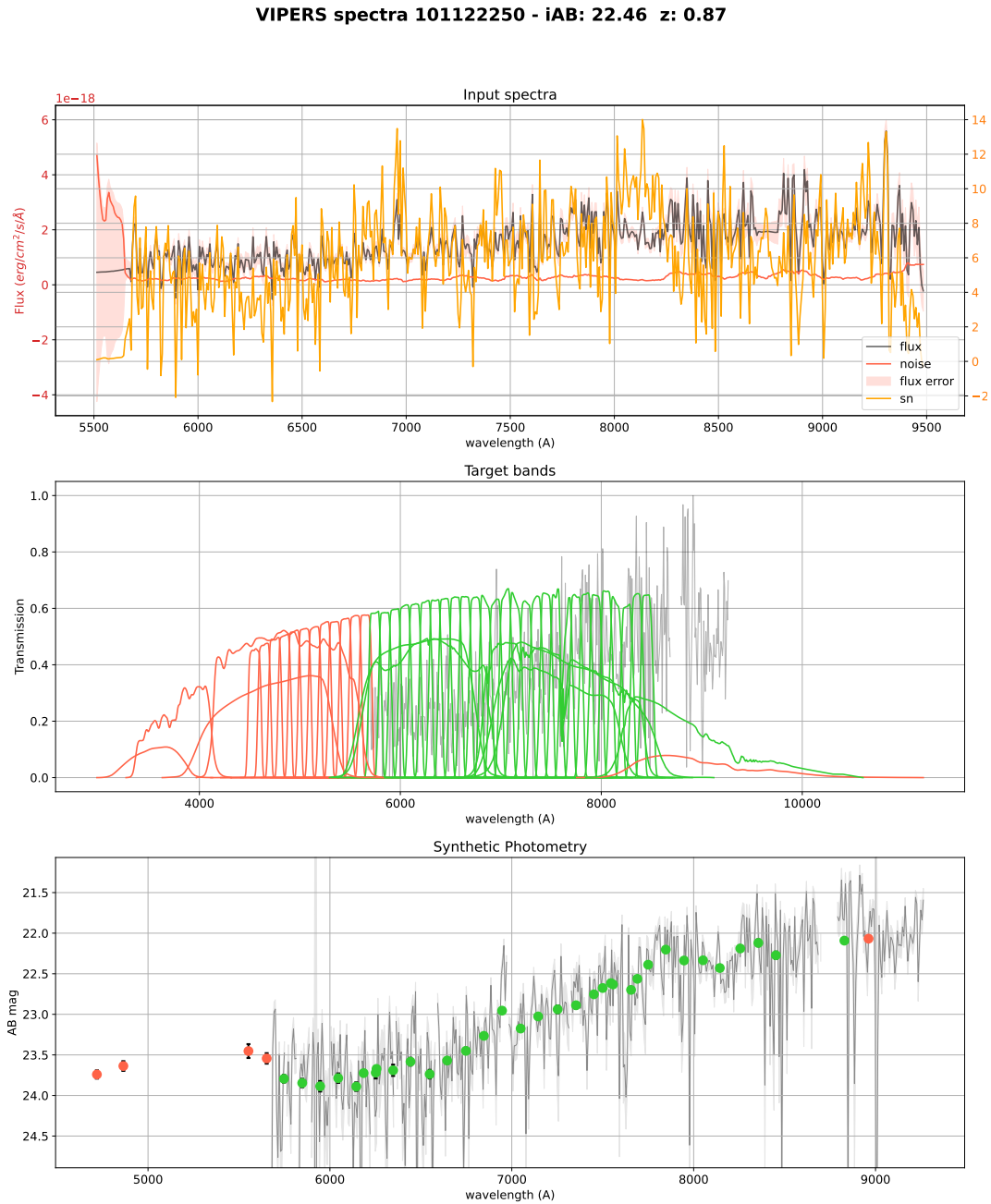


Figure A.6: An average VIPERS galaxy with an apparent magnitude close to the magnitude limit of the PAU Survey at $i_{AB}22.5$. Unfortunately the VIPERS spectra do not overlap with the bluest narrow bands in PAU, leaving approximately 30 NB of coverage for comparative analysis purposes.

DATA BASE SCHEMA

In this appendix we describe the PAU database schema. The current database contains the following tables in the following groups:

- Photometric Calibration
 - **image_zp**: Contains the image zeropoint measurements for each photometry-calibration method
 - **phot_method**: Method used during photometry for image calibration
 - **phot_zp**: Photometric Zeropoints
 - **star_photometry**: Contains the individual photometry measurements for each star matched with the reference catalogue during the nightly photometry
 - **star_zp**: Contains the individual zeropoint measurements for each star matched with the reference catalog during the nightly photometry
- MEMBA
 - **forced_aperture**: Single epoch forced aperture photometry
 - **forced_aperture_coadd**: Coadded forced aperture fluxes per band
 - **forced_aperture_report**: Reports for forced aperture inspector
 - **mask_image**: Contains the list of mask images (band and field)
 - **membra_ref_cat**: Reference catalogue used in each MEMBA production
- Crosstalk
 - **crosstalk_diff**: Crosstalk differences between raw images

- **crosstalk_ratio**: Crosstalk ratios between amplifiers
- Photo-z
 - **photoz_bcnz**: Photometric redshifts from BCNz code
- Nightly
 - **detection**: Contains the detections measured directly on the image after the nightly data reduction.
 - **image**: Contains the list of images associated to the mosaics (CCD and single amplifier images).
 - **mosaic**: Contains the list of mosaic exposure images (raw and reduced).
 - **obs_set**: Contains the list of Observation Sets registered in the database.
 - **obs_set_project**: Projects associated to observation sets.
 - **project**: List of projects associated to PAUCam observations
 - **quality_control**: Contains quality control entries measured during the data reduction process.
- Survey Strategy
 - **ss_target**: Survey Strategy targets from observations (updated to 20A)
- Extenal
 - **cfhtlens**: External CFHTLenS catalogue for Forced Photometry
 - **cosmos**: External table from zCOSMOS (DR3). Sources with accurate redshifts for forced photometry and validation.
 - **deep2**: The DEEP2 DR4 redshift catalog
 - **gaia_dr2**: Gaia DR2 stellar catalogue
 - **kids**: KIDS KV450-G9 reference catalogue
 - **sdss_spec**: SDSS Spectra catalogue over W3
 - **sdss_spec_photo**: External table from SDSS DR12 (Spec_Photo view). Sources with spectra for forced photometry and validation.
 - **sdss_star**: External table from SDSS DR12 (Star view). Stars for simulation and calibration.
- Synthetic Photometry

- **match_to_spec**: Match table between forced aperture catalogues and spectra catalogues.
- **spec_conv**: Contains the convolved fluxes derived from spectra observations from external surveys (i.e. SDSS, COSMOS, DEEP2...)
- **synth_sdss**: Synthetic photometry over SDSS Spectra (over Cosmos and W1)
- **synth_vipers**: Synthetic photometry over VIPERS Spectra (over W1)
- Brownthrower
 - **dependency**: Tracks the dependency between Brownthrower jobs (Operation table).
 - **job**: Tracks the list of Brownthrower computing jobs (Operation table).
 - **tag**: Contains tags for Brownthrower jobs (Operation table).
- Production
 - **production**: Tracks the different processing production runs for all pipelines.

ACRONYMS

ADU Analog-Digital Units.

AOCS Attitude and Orbital Control System.

BAO Baryon Acoustic Oscillations.

CANDELS Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey.

CBE Current Best Estimate.

CCD Charged Couple Devices.

CFHT Canada French Hawaii Telescope.

CFHTLenS Canada France Hawaii Lensing Survey.

CFHTLS Canada French Hawaii Telescope Legacy Survey.

Dec Declination.

DECam Dark Energy Survey Camera.

DES Dark Energy Survey.

DESI Dark Energy Spectroscopy Instrument.

ESA European Space Agency.

ESO European Southern Observatory.

ESOC European Space Operations Centre.

EXT EXT.

FGS (Euclid) Fine Guidance Sensor.

FITS Flexible Image Transport System.

FoV Field of View.

FWA Filter Wheel Assembly.

FWHM Full-Width Half-Max.

GAMA Galaxy And Mass Assembly survey.

GFA Guiding, Focusing and Alignment camera.

GUI Graphical User Interface.

GWA Grism Wheel Assembly.

HA Hour Angle.

HDU Header Data Unit.

HOD Halo Occupation Distribution.

HST Hubble Space Telescope.

IAL Infrastructure Abstraction Layer.

IT Information Technology.

MDB Mission Data Base.

MICE Marenstrum Institut de Ciències del Espai.

NISP Near Infrared Spectrometer and Photometer instrument.

NISP-P NISP Photometry.

NISP-S NISP Spectroscopy.

OGS Operations Ground Segment.

PA Position Angle.

PAU Physics of the Accelerating Universe.

PAUCam Physics of the Accelerating Universe Camera.

PAUCCS PAU Camera Control System.

PAUdm Physics of the Accelerating Universe Data Management.

PIC Port d'Informació Científica.

PID Proportional Integral Derivative.

PLM (Euclid)PayLoad Module.

PPO Pipeline Processing Order.

PSF Point Spread Function.

QE Quantum Efficiency.

QSO Quasi Stellar Object.

RA Right Ascension.

RMS Root Mean Square.

ROI Region of Interest.

ROS Reference Observation Sequence.

S/N Signal to Noise.

SC Scientific Challenge.

SDC Science Data Center.

SDSS Sloan Digital Sky Survey.

SED Spectrum Energy Distribution.

SGS Science Ground Segment.

SMARTS Small and Moderate Aperture Research Telescope System.

SSO Solar System Object.

STM Structural and Thermal Model.

SVM (Euclid) SerVice Module.

TCS Telescope Control System.

TCSi Telescope Control System interface.

TU True Universe.

VIS Visible Instrument.

WCS World Coordinate System.

WHT William Herschel Telescope.

BIBLIOGRAPHY

Abbott T. M. C., Adamów M., Agüena M., Allam S., Amon A., Annis J., Avila S., Bacon D., Banerji M., Bechtol K., Becker M. R., Bernstein G. M., Bertin E., Bhargava S., Bridle S. L., Brooks D., Burke D. L., Carnero Rosell A., Carrasco Kind M., Carretero J., Castander F. J., Cawthon R., Chang C., Choi A., Conselice C., Costanzi M., Croce M., da Costa L. N., Davis T. M., De Vicente J., DeRose J., Desai S., Diehl H. T., Dietrich J. P., Drlica-Wagner A., Eckert K., Elvin-Poole J., Everett S., Evrard A. E., Ferrero I., Ferté A., Flaughner B., Fosalba P., Friedel D., Frieman J., García-Bellido J., Gaztanaga E., Gelman L., Gerdes D. W., Giannantonio T., Gill M. S. S., Gruen D., Gruendl R. A., Gschwend J., Gutierrez G., Hartley W. G., Hinton S. R., Hollowood D. L., Honscheid K., Huterer D., James D. J., Jeltema T., Johnson M. D., Kent S., Kron R., Kuehn K., Kuropatkin N., Lahav O., Li T. S., Lidman C., Lin H., MacCrann N., Maia M. A. G., Manning T. A., Maloney J. D., March M., Marshall J. L., Martini P., Melchior P., Menanteau F., Miquel R., Morgan R., Myles J., Neilsen E., Ogando R. L. C., Palmese A., Paz-Chinchón F., Petravick D., Pieres A., Plazas A. A., Pond C., Rodriguez-Monroy M., Romer A. K., Roodman A., Rykoff E. S., Sako M., Sanchez E., Santiago B., Scarpine V., Serrano S., Sevilla-Noarbe I., Smith J. Allyn, Smith M., Soares-Santos M., Suchyta E., Swanson M. E. C., Tarle G., Thomas D., To C., Tremblay P. E., Troxel M. A., Tucker D. L., Turner D. J., Varga T. N., Walker A. R., Wechsler R. H., Weller J., Wester W., Wilkinson R. D., Yanny B., Zhang Y., Nikutta R., Fitzpatrick M., Jacques A., Scott A., Olsen K., Huang L., Herrera D., Juneau S., Nidever D., Weaver B. A., Adean C., Correia V., de Freitas M., Freitas F. N., Singulani C., Vila-Verde G., Linea Science Server .

The Dark Energy Survey Data Release 2

// *ApJS*. VIII 2021. 255, 2. 20.

Alarcon Alex, Gaztanaga Enrique, Eriksen Martin, Baugh Carlton M., Cabayol Laura, Casas Ricard, Carretero Jorge, Castander Francisco J., De Vicente Juan, Fernandez Enrique, Garcia-Bellido Juan, Hildebrandt Hendrik, Hoekstra Henk, Joachimi Benjamin, Manzoni Giorgio, Miquel Ramon, Norberg Peder, Padilla Cristobal, Renard Pablo, Sanchez Eusebio, Serrano Santiago, Sevilla-Noarbe Ignacio, Siudek Malgorzata, Tallada-Crespí Pau.

The PAU Survey: an improved photo-z sample in the COSMOS field

// *MNRAS*. III 2021. 501, 4. 6103–6122.

Astropy Collaboration , Astropy Contributors .

The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package
// AJ. IX 2018. 156, 3. 123.

Bertin E.

Automatic Astrometric and Photometric Calibration with SCAMP
// Astronomical Data Analysis Software and Systems XV. 351. VII 2006. 112.
(Astronomical Society of the Pacific Conference Series).

Bertin E.

SkyMaker: astronomical image simulations made easy.
// Memorie della SAIT. I 2009. 80. 422.

Bertin E.

Automated Morphometry with SExtractor and PSFEx
// Astronomical Data Analysis Software and Systems XX. 442. VII 2011. 435.
(Astronomical Society of the Pacific Conference Series).

Bertin E., Arnouts S.

SExtractor: Software for source extraction.
// A&AS. VI 1996. 117. 393–404.

Bertin Emmanuel, Mellier Yannick, Radovich Mario, Missonnier Gilles, Didelon Pierre, Morin Bertrand.

The TERAPIX Pipeline
// Astronomical Data Analysis Software and Systems XI. 281. I 2002. 228.
(Astronomical Society of the Pacific Conference Series).

Bradley Larry, Sipócz Brigitta, Robitaille Thomas, Tollerud Erik, Vinícius Zè, Deil Christoph, Barbary Kyle, Wilson Tom J, Busko Ivo, Günther Hans Moritz, Cara Mihai, Conseil Simon, Bostroem Azalee, Droettboom Michael, Bray E. M., Bratholm Lars Andersen, Lim P. L., Barentsen Geert, Craig Matt, Pascual Sergio, Perren Gabriel, Greco Johnny, Donath Axel, Val-Borro Miguel de, Kerzendorf Wolfgang, Bach Yoonsoo P., Weaver Benjamin Alan, D'Eugenio Francesco, Souchereau Harrison, Ferreira Leonardo.
astropy/photutils: 1.0.0. IX 2020.

Bruzual G., Charlot S.

Stellar population synthesis at the resolution of 2003
// MNRAS. X 2003. 344, 4. 1000–1028.

Cabayol-Garcia L., Eriksen M., Alarcón A., Amara A., Carretero J., Casas R., Castander F. J., Fernández E., García-Bellido J., Gaztanaga E., Hoekstra H., Miquel R., Neissner C., Padilla C., Sánchez E., Serrano S., Sevilla-Noarbe I., Siudek M., Tallada P., Tortorelli L.

- The PAU Survey: background light estimation with deep learning techniques
// MNRAS. II 2020. 491, 4. 5392–5405.
- Calabretta M. R., Greisen E. W.*
Representations of celestial coordinates in FITS
// AAP. XII 2002. 395. 1077–1122.
- Calzetti Daniela, Armus Lee, Bohlin Ralph C., Kinney Anne L., Koornneef Jan, Storchi-Bergmann Thaisa.*
The Dust Content and Opacity of Actively Star-forming Galaxies
// ApJ. IV 2000. 533, 2. 682–695.
- Carretero J., Castander F. J., Gaztañaga E., Crocce M., Fosalba P.*
An algorithm to build mock galaxy catalogues using MICE simulations
// MNRAS. II 2015. 447, 1. 646–670.
- Colless Matthew, Dalton Gavin, Maddox Steve, Sutherland Will, Norberg Peder, Cole Shaun, Bland-Hawthorn Joss, Bridges Terry, Cannon Russell, Collins Chris, Couch Warrick, Cross Nicholas, Deeley Kathryn, De Propris Roberto, Driver Simon P., Efstathiou George, Ellis Richard S., Frenk Carlos S., Glazebrook Karl, Jackson Carole, Lahav Ofer, Lewis Ian, Lumsden Stuart, Madgwick Darren, Peacock John A., Peterson Bruce A., Price Ian, Seaborne Mark, Taylor Keith.*
The 2dF Galaxy Redshift Survey: spectra and redshifts
// MNRAS. XII 2001. 328, 4. 1039–1063.
- Crocce M., Castander F. J., Gaztañaga E., Fosalba P., Carretero J.*
The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues
// MNRAS. X 2015. 453, 2. 1513–1530.
- DESI Collaboration .*
The DESI Experiment Part I: Science, Targeting, and Survey Design
// arXiv e-prints. X 2016. arXiv:1611.00036.
- Dolag K., Reinecke M., Gheller C., Imboden S.*
Spotch: visualizing cosmological simulations
// New Journal of Physics. XII 2008. 10, 12. 125006.
- Erben T., Hildebrandt H., Miller L., van Waerbeke L., Heymans C., Hoekstra H., Kitching T. D., Mellier Y., Benjamin J., Blake C., Bonnett C., Cordes O., Coupon J., Fu L., Gavazzi R., Gillis B., Grocutt E., Gwyn S. D. J., Holhjem K., Hudson M. J., Kilbinger M., Kuijken K., Milkeraitis M., Rowe B. T. P., Schrabback T., Semboloni E., Simon P., Smit M., Toader O., Vafaei S., van Uitert E., Velandier M.*

CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products

// MNRAS. VIII 2013. 433, 3. 2545–2563.

Eriksen M., Alarcon A., Gaztanaga E., Amara A., Cabayol L., Carretero J., Castander F. J., Crocce M., Delfino M., De Vicente J., Fernandez E., Fosalba P., Garcia-Bellido J., Hildebrandt H., Hoekstra H., Joachimi B., Norberg P., Miquel R., Padilla C., Refregier A., Sanchez E., Serrano S., Sevilla-Noarbe I., Tallada P., Tonello N., Tortorelli L.

The PAU Survey: early demonstration of photometric redshift performance in the COSMOS field

// MNRAS. IV 2019. 484, 3. 4200–4215.

Euclid Collaboration , Barnett R., Warren S. J., Mortlock D. J., Cuby J. G., Conselice C., Hewett P. C., Willott C. J., Auricchio N., Balaguera-Antolínez A., Baldi M., Bardelli S., Bellagamba F., Bender R., Biviano A., Bonino D., Bozzo E., Branchini E., Brescia M., Brinchmann J., Burigana C., Camera S., Capobianco V., Carbone C., Carretero J., Carvalho C. S., Castander F. J., Castellano M., Cavuoti S., Cimatti A., Clédassou R., Congedo G., Conversi L., Copin Y., Corcione L., Coupon J., Courtois H. M., Cropper M., Da Silva A., Duncan C. A. J., Dusini S., Ealet A., Farrens S., Fosalba P., Fotopoulou S., Fourmanoit N., Frailis M., Fumana M., Galeotta S., Garilli B., Gillard W., Gillis B. R., Graciá-Carpio J., Grupp F., Hoekstra H., Hormuth F., Israel H., Jahnke K., Kermiche S., Kilbinger M., Kirkpatrick C. C., Kitching T., Kohley R., Kubik B., Kunz M., Kurki-Suonio H., Laureijs R., Ligi S., Lilje P. B., Lloro I., Maiorano E., Mansutti O., Marggraf O., Martinet N., Marulli F., Massey R., Mauri N., Medinaceli E., Mei S., Mellier Y., Metcalf R. B., Metge J. J., Meylan G., Moresco M., Moscardini L., Munari E., Neissner C., Niemi S. M., Nutma T., Padilla C., Paltani S., Pasian F., Paykari P., Percival W. J., Pettorino V., Polenta G., Poncet M., Pozzetti L., Raison F., Renzi A., Rhodes J., Rix H. W., Romelli E., Roncarelli M., Rossetti E., Saglia R., Sapone D., Scaramella R., Schneider P., Scottez V., Secroun A., Serrano S., Sirri G., Stanco L., Sureau F., Tallada-Crespí P., Tavagnacco D., Taylor A. N., Tenti M., Tereno I., Toledo-Moreo R., Torradeflot F., Valenziano L., Vassallo T., Wang Y., Zacchei A., Zamorani G., Zoubian J., Zucca E.

Euclid preparation. V. Predicted yield of redshift $7 < z < 9$ quasars from the wide survey

// AAP. XI 2019. 631. A85.

Euclid Collaboration , Bretonnière H., Huertas-Company M., Boucaud A., Lanusse F., Jullo E., Merlin E., Castellano M., Brinchmann J., Conselice C. J., Dole H., Cabanac R., Courtois H. M., Castander F. J., Duc P. A., Fosalba P., Guinet D., Kruk S., Kuchner U., Serrano S., Soubrie E., Tramacere A., Wang L., others .

Euclid preparation: XIII. Forecasts for galaxy morphology with the Euclid Survey using Deep Generative Models

// arXiv e-prints. V 2021. 1. arXiv:2105.12149.

Fagrelus Parker, Abareshi Behzad, Allen Lori, Ballester Otger, Baltay Charles, Besuner Robert, Buckley-Geer Elizabeth, Butler Karen, Cardiel Laia, Dey Arjun, Duan Yutong, Elliott Ann, Emmet William, Gershkovich Irena, Honscheid Klaus, Illa Jose M., Jimenez Jorge, Joyce Richard, Karcher Armin, Kent Stephen, Lambert Andrew, Lampton Michael, Levi Michael, Manser Christopher, Marshall Robert, Martini Paul, Paat Anthony, Probst Ronald, Rabinowitz David, Reil Kevin, Robertson Amy, Rockosi Connie, Schlegel David, Schubnell Michael, Serrano Santiago, Silber Joseph, Soto Christian, Sprayberry David, Summers David, Tarlé Greg, Weaver Benjamin A.

ProtoDESI: First On-Sky Technology Demonstration for the Dark Energy Spectroscopic Instrument

// PASP. II 2018. 130, 984. 025005.

Flaugher B., Diehl H. T., Honscheid K., Abbott T. M. C., Alvarez O., Angstadt R., Annis J. T., Antonik M., Ballester O., Beaufore L., Bernstein G. M., Bernstein R. A., Bigelow B., Bonati M., Boprie D., Brooks D., Buckley-Geer E. J., Campa J., Cardiel-Sas L., Castander F. J., Castilla J., Cease H., Cela-Ruiz J. M., Chappa S., Chi E., Cooper C., da Costa L. N., Dede E., Derylo G., DePoy D. L., de Vicente J., Doel P., Drlica-Wagner A., Eiting J., Elliott A. E., Emes J., Estrada J., Fausti Neto A., Finley D. A., Flores R., Frieman J., Gerdes D., Gladders M. D., Gregory B., Gutierrez G. R., Hao J., Holland S. E., Holm S., Huffman D., Jackson C., James D. J., Jonas M., Karcher A., Karliner I., Kent S., Kessler R., Kozlovsky M., Kron R. G., Kubik D., Kuehn K., Kuhlmann S., Kuk K., Lahav O., Lathrop A., Lee J., Levi M. E., Lewis P., Li T. S., Mandrichenko I., Marshall J. L., Martinez G., Merritt K. W., Miquel R., Muñoz F., Neilsen E. H., Nichol R. C., Nord B., Ogando R., Olsen J., Palao N., Patton K., Peoples J., Plazas A. A., Rauch J., Reil K., Rheault J. P., Roe N. A., Rogers H., Roodman A., Sanchez E., Scarpine V., Schindler R. H., Schmidt R., Schmitt R., Schubnell M., Schultz K., Schurter P., Scott L., Serrano S., Shaw T. M., Smith R. C., Soares-Santos M., Stefanik A., Stuermer W., Suchyta E., Sypniewski A., Tarle G., Thaler J., Tighe R., Tran C., Tucker D., Walker A. R., Wang G., Watson M., Weaverdyck C., Wester W., Woods R., Yanny B., DES Collaboration .

The Dark Energy Camera

// AJ. XI 2015. 150, 5. 150.

Fosalba P., Crocce M., Gaztañaga E., Castander F. J.

The MICE grand challenge lightcone simulation - I. Dark matter clustering

// MNRAS. IV 2015a. 448, 4. 2987–3000.

Fosalba P., Gaztañaga E., Castander F. J., Crocce M.

The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps

// MNRAS. II 2015b. 447, 2. 1319–1332.

Fosalba Pablo, Gaztañaga Enrique, Castander Francisco J., Manera Marc.

The onion universe: all sky lightcone simulations in spherical shells

// MNRAS. XI 2008. 391, 1. 435–446.

Gaia Collaboration .

Gaia Data Release 2. Summary of the contents and survey properties

// AAP. VIII 2018. 616. A1.

Gatti M., Sheldon E., Amon A., Becker M., Troxel M., Choi A., Doux C., MacCrann N., Navarro-Alsina A., Harrison I., Gruen D., Bernstein G., Jarvis M., Secco L. F., Féré A., Shin T., McCullough J., Rollins R. P., Chen R., Chang C., Pandey S., Tutusaus I., Prat J., Elvin-Poole J., Sanchez C., Plazas A. A., Roodman A., Zuntz J., Abbott T. M. C., Agüena M., Allam S., Annis J., Avila S., Bacon D., Bertin E., Bhargava S., Brooks D., Burke D. L., Carnero Rosell A., Carrasco Kind M., Carretero J., Castander F. J., Conselice C., Costanzi M., Crocce M., da Costa L. N., Davis T. M., De Vicente J., Desai S., Diehl H. T., Dietrich J. P., Doel P., Drlica-Wagner A., Eckert K., Everett S., Ferrero I., Frieman J., García-Bellido J., Gerdes D. W., Giannantonio T., Gruendl R. A., Gschwend J., Gutierrez G., Hartley W. G., Hinton S. R., Hollowood D. L., Honscheid K., Hoyle B., Huff E. M., Huterer D., Jain B., James D. J., Jeltema T., Krause E., Kron R., Kuropatkin N., Lima M., Maia M. A. G., Marshall J. L., Miquel R., Morgan R., Myles J., Palmese A., Paz-Chinchón F., Rykoff E. S., Samuroff S., Sanchez E., Scarpine V., Schubnell M., Serrano S., Sevilla-Noarbe I., Smith M., Soares-Santos M., Suchyta E., Swanson M. E. C., Tarle G., Thomas D., To C., Tucker D. L., Varga T. N., Wechsler R. H., Weller J., Wester W., Wilkinson R. D.

Dark energy survey year 3 results: weak lensing shape catalogue

// MNRAS. VII 2021. 504, 3. 4312–4336.

Gonneau A., Lyubenova M., Lançon A., Trager S. C., Peletier R. F., Arentsen A., Chen Y. P., Coelho P. R. T., Dries M., Falcón-Barroso J., Prugniel P., Sánchez-Blázquez P., Vazdekis A., Verro K.

The X-shooter Spectral Library (XSL): Data release 2

// AAP. II 2020. 634. A133.

Grogin Norman A., Kocevski Dale D., Faber S. M., Ferguson Henry C., Koekemoer Anton M., Riess Adam G., Acquaviva Viviana, Alexander David M., Almaini Omar, Ashby Matthew L. N., Barden Marco, Bell Eric F., Bournaud Frédéric, Brown Thomas M., Caputi Karina I., Casertano Stefano, Cassata Paolo, Castellano Marco, Challis Peter, Chary Ranga-Ram, Cheung Edmond, Cirasuolo Michele, Conselice Christopher J., Roshan Cooray Asantha, Croton Darren J., Daddi Emanuele, Dahlen Tomas, Davé Romeel, de Mello Duília F., Dekel Avishai, Dickinson Mark, Dolch Timothy, Donley Jennifer L., Dunlop James S., Dutton Aaron A., Elbaz David, Fazio Giovanni G., Filippenko Alexei V., Finkelstein Steven L., Fontana Adriano, Gardner Jonathan P., Garnavich Peter M., Gawiser Eric, Giavalisco Mauro, Grazian Andrea, Guo Yicheng, Hathi Nimish P., Häussler Boris, Hopkins Philip F., Huang Jia-Sheng, Huang Kuang-Han, Jha

Saurabh W., Kartaltepe Jeyhan S., Kirshner Robert P., Koo David C., Lai Kamson, Lee Kyoung-Soo, Li Weidong, Lotz Jennifer M., Lucas Ray A., Madau Piero, McCarthy Patrick J., McGrath Elizabeth J., McIntosh Daniel H., McLure Ross J., Mobasher Bahram, Moustakas Leonidas A., Mozena Mark, Nandra Kirpal, Newman Jeffrey A., Niemi Sami-Matias, Noeske Kai G., Papovich Casey J., Pentericci Laura, Pope Alexandra, Primack Joel R., Rajan Abhijith, Ravindranath Swara, Reddy Naveen A., Renzini Alvio, Rix Hans-Walter, Robaina Aday R., Rodney Steven A., Rosario David J., Rosati Piero, Salimbeni Sara, Scarlata Claudia, Siana Brian, Simard Luc, Smidt Joseph, Somerville Rachel S., Spinrad Hyron, Straughn Amber N., Strolger Louis-Gregory, Telford Olivia, Teplitz Harry I., Trump Jonathan R., van der Wel Arjen, Villforth Carolin, Wechsler Risa H., Weiner Benjamin J., Wiklind Tommy, Wild Vivienne, Wilson Grant, Wuyts Stijn, Yan Hao-Jing, Yun Min S.

CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
// *ApJS*. XII 2011. 197, 2. 35.

Hanuschik R. W.

A flux-calibrated, high-resolution atlas of optical sky emission from UVES
// *AAP*. IX 2003. 407. 1157–1164.

Heavens Alan, Refregier Alexandre, Heymans Catherine.

Intrinsic correlation of galaxy shapes: implications for weak lensing measurements
// *MNRAS*. XII 2000. 319, 2. 649–656.

Heymans Catherine, Van Waerbeke Ludovic, Miller Lance, Erben Thomas, Hildebrandt Hendrik, Hoekstra Henk, Kitching Thomas D., Mellier Yannick, Simon Patrick, Bonnett Christopher, Coupon Jean, Fu Liping, Harnois Déraps Joachim, Hudson Michael J., Kilbinger Martin, Kuijken Koenraad, Rowe Barnaby, Schrabback Tim, Semboloni Elisabetta, van Uitert Edo, Vafaei Sanaz, Velander Malin.

CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey
// *MNRAS*. XI 2012. 427, 1. 146–166.

Hildebrandt H., Erben T., Kuijken K., van Waerbeke L., Heymans C., Coupon J., Benjamin J., Bonnett C., Fu L., Hoekstra H., Kitching T. D., Mellier Y., Miller L., Velander M., Hudson M. J., Rowe B. T. P., Schrabback T., Semboloni E., Benítez N.

CFHTLenS: improving the quality of photometric redshifts with precision photometry
// *MNRAS*. IV 2012. 421, 3. 2355–2367.

Høg E., Fabricius C., Makarov V. V., Urban S., Corbin T., Wycoff G., Bastian U., Schwekendiek P., Wicenec A.

The Tycho-2 catalogue of the 2.5 million brightest stars
// *AAP*. III 2000. 355. L27–L30.

Janesick James R.

Scientific charge-coupled devices. 2001. 680.

Kabsch W.

A solution for the best rotation to relate two sets of vectors

// Acta Crystallographica Section A. IX 1976. 32, 5. 922–923.

Kubik Bogna, Barbier Remi, Chabanat Eric, Chapon Arnaud, Clemens Jean-Claude, Ealet Anne, Ferriol Sylvain, Gillard William, Secroun Aurelia, Serra Benoit, Smadja Gerard, Tilquin André.

A New Signal Estimator from the NIR Detectors of the Euclid Mission

// PASP. X 2016. 128, 968. 104504.

Kuijken K., Heymans C., Dvornik A., Hildebrandt H., de Jong J. T. A., Wright A. H., Erben T., Bilicki M., Giblin B., Shan H. Y., Getman F., Grado A., Hoekstra H., Miller L., Napolitano N., Paolilo M., Radovich M., Schneider P., Sutherland W., Tewes M., Tortora C., Valentijn E. A., Verdoes Kleijn G. A.

The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR photometry over 1000 square degrees

// AAP. V 2019. 625. A2.

Kuijken Konrad, Heymans Catherine, Hildebrandt Hendrik, Nakajima Reiko, Erben Thomas, de Jong Jelte T. A., Viola Massimo, Choi Ami, Hoekstra Henk, Miller Lance, van Uitert Edo, Amon Alexandra, Blake Chris, Brouwer Margot, Buddendiek Axel, Conti Ian Fenech, Eriksen Martin, Grado Aniello, Harnois-Déraps Joachim, Helmich Ewout, Herbonnet Ricardo, Irisarri Nancy, Kitching Thomas, Klaes Dominik, La Barbera Francesco, Napolitano Nicola, Radovich Mario, Schneider Peter, Sifón Cristóbal, Sikkema Gert, Simon Patrick, Tudorica Alexandru, Valentijn Edwin, Verdoes Kleijn Gijs, van Waerbeke Ludovic.

Gravitational lensing analysis of the Kilo-Degree Survey

// MNRAS. XII 2015. 454, 4. 3500–3532.

LSST collaboration .

LSST: From Science Drivers to Reference Design and Anticipated Data Products

// ApJ. III 2019. 873, 2. 111.

Laigle C., McCracken H. J., Ilbert O., Hsieh B. C., Davidzon I., Capak P., Hasinger G., Silverman J. D., Pichon C., Coupon J., Aussel H., Le Borgne D., Caputi K., Cassata P., Chang Y. Y., Civano F., Dunlop J., Fynbo J., Kartaltepe J. S., Koekemoer A., Le Fèvre O., Le Floch E., Leauthaud A., Lilly S., Lin L., Marchesi S., Milvang-Jensen B., Salvato M., Sanders D. B., Scoville N., Smolcic V., Stockmann M., Taniguchi Y., Tasca L., Toft S., Vaccari Mattia, Zabl J.

The COSMOS2015 Catalog: Exploring the $1 < z < 6$ Universe with Half a Million Galaxies

// ApJS. VI 2016. 224, 2. 24.

Lastennet E., Lejeune T., Oblak E., Westera P., Buser R.

BaSeL: A Library of Synthetic Spectra and Colours for GAIA
// AP&SS. I 2002. 280. 83–87.

Laureijs R. et al.

Euclid Definition Study Report
// arXiv e-prints. X 2011. 1. arXiv:1110.3193.

Le Fèvre O., Cassata P., Cucciati O., Garilli B., Ilbert O., Le Brun V., Maccagni D., Moreau C., Scodreggio M., Tresse L., Zamorani G., Adami C., Arnouts S., Bardelli S., Bolzonella M., Bondi M., Bongiorno A., Bottini D., Cappi A., Charlot S., Ciliegi P., Contini T., de la Torre S., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., Lemaux B., López-Sanjuan C., McCracken H. J., Marano B., Marinoni C., Mazure A., Mellier Y., Merighi R., Merluzzi P., Paltani S., Pellò R., Pollo A., Pozzetti L., Scaramella R., Tasca L., Vergani D., Vettolani G., Zanicelli A., Zucca E.

The VIMOS VLT Deep Survey final data release: a spectroscopic sample of 35 016 galaxies and AGN out to $z \sim 6.7$ selected with $17.5 \leq i_{AB} \leq 24.75$
// AAP. XI 2013. 559. A14.

Lilly Simon J., Le Brun Vincent, Maier Christian, Mainieri Vincenzo, Mignoli Marco, Scodreggio Marco, Zamorani Gianni, Carollo Marcella, Contini Thierry, Kneib Jean-Paul, Le Fèvre Olivier, Renzini Alvio, Bardelli Sandro, Bolzonella Micol, Bongiorno Angela, Caputi Karina, Coppa Graziano, Cucciati Olga, de la Torre Sylvain, de Ravel Loic, Franzetti Paolo, Garilli Bianca, Iovino Angela, Kampeczyk Pawel, Kovac Katarina, Knobel Christian, Lamareille Fabrice, Le Borgne Jean-Francois, Pello Roser, Peng Yingjie, Pérez-Montero Enrique, Ricciardelli Elena, Silverman John D., Tanaka Masayuki, Tasca Lidia, Tresse Laurence, Vergani Daniela, Zucca Elena, Ilbert Olivier, Salvato Mara, Oesch Pascal, Abbas Umi, Bottini Dario, Capak Peter, Cappi Alberto, Cassata Paolo, Cimatti Andrea, Elvis Martin, Fumana Marco, Guzzo Luigi, Hasinger Gunther, Koekemoer Anton, Leauthaud Alexei, Maccagni Dario, Marinoni Christian, McCracken Henry, Memeo Pierdomenico, Meneux Baptiste, Porciani Cristiano, Pozzetti Lucia, Sanders David, Scaramella Roberto, Scarlata Claudia, Scoville Nick, Shopbell Patrick, Taniguchi Yoshiaki.

The zCOSMOS 10k-Bright Spectroscopic Sample
// ApJS. X 2009. 184, 2. 218–229.

Martínez-Lombilla Cristina, Trujillo Ignacio, Knapen Johan H.

Discovery of disc truncations above the galaxies' mid-plane in Milky Way-like galaxies
// MNRAS. II 2019. 483, 1. 664–691.

Massey Richard, Schrabback Tim, Cordes Oliver, Marggraf Ole, Israel Holger, Miller Lance, Hall David, Cropper Mark, Prod'homme Thibaut, Niemi Sami-Matias.

An improved model of charge transfer inefficiency and correction algorithm for the Hubble Space Telescope

// MNRAS. III 2014. 439, 1. 887–907.

Miller L., Heymans C., Kitching T. D., van Waerbeke L., Erben T., Hildebrandt H., Hoekstra H., Mellier Y., Rowe B. T. P., Coupon J., Dietrich J. P., Fu L., Harnois-Déraps J., Hudson M. J., Kilbinger M., Kuijken K., Schrabback T., Semboloni E., Vafaei S., Velander M.

Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey

// MNRAS. III 2013. 429, 4. 2858–2880.

Newman Jeffrey A., Cooper Michael C., Davis Marc, Faber S. M., Coil Alison L., Guhathakurta Puragra, Koo David C., Phillips Andrew C., Conroy Charlie, Dutton Aaron A., Finkbeiner Douglas P., Gerke Brian F., Rosario David J., Weiner Benjamin J., Willmer C. N. A., Yan Renbin, Harker Justin J., Kassin Susan A., Konidaris N. P., Lai Kamson, Madgwick Darren S., Noeske K. G., Wirth Gregory D., Connolly A. J., Kaiser N., Kirby Evan N., Lemaux Brian C., Lin Lihwai, Lotz Jennifer M., Luppino G. A., Marinoni C., Matthews Daniel J., Metevier Anne, Schiavon Ricardo P.

The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts

// ApJS. IX 2013. 208, 1. 5.

O'Donnell James E.

R v-dependent Optical and Near-Ultraviolet Extinction

// ApJ. II 1994. 422. 158.

Oke J. B.

Faint Spectrophotometric Standard Stars

// AJ. V 1990. 99. 1621.

Padilla Cristóbal, Castander Francisco J., Alarcón Alex, Aleksic Jelena, Ballester Otger, Cabayol Laura, Cardiel-Sas Laia, Carretero Jorge, Casas Ricard, Castilla Javier, Croce Martin, Delfino Manuel, Díaz Carlos, Eriksen Martin, Fernández Enrique, Fosalba Pablo, García-Bellido Juan, Gaztañaga Enrique, Gaweda Javier, Grañena Ferran, María Ílla José, Jiménez Jorge, López Luis, Martí Pol, Miquel Ramon, Neissner Christian, Pío Cristóbal, Sánchez Eusebio, Serrano Santiago, Sevilla-Noarbe Ignacio, Tallada Pau, Tonello Nadia, de Vicente Juan.

The Physics of the Accelerating Universe Camera

// AJ. VI 2019. 157, 6. 246.

Padmanabhan Nikhil, Schlegel David J., Finkbeiner Douglas P., Barentine J. C., Blanton Michael R., Brewington Howard J., Gunn James E., Harvanek Michael, Hogg David W., Ivezić Željko, Johnston David, Kent Stephen M., Kleinman S. J., Knapp Gillian R., Krzesinski Jurek,

Long Dan, Neilsen Jr. Eric H., Nitta Atsuko, Loomis Craig, Lupton Robert H., Roweis Sam, Snedden Stephanie A., Strauss Michael A., Tucker Douglas L.

An Improved Photometric Calibration of the Sloan Digital Sky Survey Imaging Data

// ApJ. II 2008. 674, 2. 1217–1233.

Pickles A., Depagne É.

All-Sky Spectrally Matched UBVRI - ZY and u g r i z Magnitudes for Stars in the Tycho2 Catalog

// PASP. XII 2010. 122, 898. 1437.

Pickles A. J.

A Stellar Spectral Flux Library: 1150-25000 Å

// PASP. VII 1998. 110, 749. 863–878.

Planck Collaboration .

Planck 2013 results. XI. All-sky model of thermal dust emission

// AAP. XI 2014. 571. A11.

Polletta M., Tajer M., Maraschi L., Trinchieri G., Lonsdale C. J., Chiappetti L., Andreon S., Pierre M., Le Fèvre O., Zamorani G., Maccagni D., Garcet O., Surdej J., Franceschini A., Alloin D., Shupe D. L., Surace J. A., Fang F., Rowan-Robinson M., Smith H. E., Tresse L.

Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey

// ApJ. VII 2007. 663, 1. 81–102.

Potter Douglas, Stadel Joachim, Teyssier Romain.

PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys

// Computational Astrophysics and Cosmology. V 2017. 4, 1. 2.

Pozzetti L., Hirata C. M., Geach J. E., Cimatti A., Baugh C., Cucciati O., Merson A., Norberg P., Shi D.

Modelling the number density of H α emitters for future spectroscopic near-IR space missions

// AAP. V 2016. 590. A3.

Prevot M. L., Lequeux J., Maurice E., Prevot L., Rocca-Volmerange B.

The typical interstellar extinction in the Small Magellanic Cloud.

// AAP. III 1984. 132. 389–392.

Richards Gordon T., Fan Xiaohui, Schneider Donald P., Vanden Berk Daniel E., Strauss Michael A., York Donald G., Anderson Jr. John E., Anderson Scott F., Annis James, Bahcall Neta A., Bernardi Mariangela, Briggs John W., Brinkmann J., Brunner Robert, Burles Scott, Carey Larry, Castander Francisco J., Connolly A. J., Crocker J. H., Csabai István, Doi

Mamoru, Finkbeiner Douglas, Friedman Scott D., Frieman Joshua A., Fukugita Masataka, Gunn James E., Hindsley Robert B., Ivezić Željko, Kent Stephen, Knapp G. R., Lamb D. Q., Leger R. French, Long Daniel C., Loveday Jon, Lupton Robert H., McKay Timothy A., Meiksin Avery, Merrelli Aronne, Munn Jeffrey A., Newberg Heidi Jo, Newcomb Matt, Nichol R. C., Owen Russell, Pier Jeffrey R., Pope Adrian, Richmond Michael W., Rockosi Constance M., Schlegel David J., Siegmund Walter A., Smee Stephen, Snir Yehuda, Stoughton Chris, Stubbs Christopher, SubbaRao Mark, Szalay Alexander S., Szokoly Gyula P., Tremonti Christy, Uomoto Alan, Waddell Patrick, Yanny Brian, Zheng Wei.

Colors of 2625 Quasars at $0 < Z < 5$ Measured in the Sloan Digital Sky Survey Photometric System

// AJ. V 2001. 121, 5. 2308–2330.

Robin A., Creze M.

Stellar populations in the milky way: a synthetic model.

// AAP. III 1986. 157. 71–90.

Rowe B. T. P., Jarvis M., Mandelbaum R., Bernstein G. M., Bosch J., Simet M., Meyers J. E., Kacprzak T., Nakajima R., Zuntz J., Miyatake H., Dietrich J. P., Armstrong R., Melchior P., Gill M. S. S.

GALSIM: The modular galaxy image simulation toolkit

// Astronomy and Computing. IV 2015. 10. 121–150.

SDSS Collaboration .

The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III

// ApJS. VII 2015. 219, 1. 12.

Schlegel David J., Finkbeiner Douglas P., Davis Marc.

Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds

// ApJ. VI 1998. 500, 2. 525–553.

Schultheis M., Chen B. Q., Jiang B. W., Gonzalez O. A., Enokiya R., Fukui Y., Torii K., Rejkuba M., Minniti D.

Mapping the Milky Way bulge at high resolution: the 3D dust extinction, CO, and X factor maps

// AAP. VI 2014. 566. A120.

Tallada P., Carretero J., Casals J., Acosta-Silva C., Serrano S., Caubet M., Castander F. J., César E., Croce M., Delfino M., Eriksen M., Fosalba P., Gaztañaga E., Merino G., Neissner C., Tonello N.

CosmoHub: Interactive exploration and distribution of astronomical data on Hadoop
 // *Astronomy and Computing*. VII 2020. 32. 100391.

The Dark Energy Survey Collaboration .

The Dark Energy Survey
 // arXiv e-prints. X 2005. 1. astro-ph/0510346.

Tody Doug.

The IRAF Data Reduction and Analysis System
 // *Instrumentation in astronomy* VI. 627. I 1986. 733.
 (Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series).

Tonello N., Tallada P., Serrano S., Carretero J., Eriksen M., Folger M., Neissner C., Sevilla-Noarbe I., Castander F. J., Delfino M., De Vicente J., Fernandez E., Garcia-Bellido J., Gaztanaga E., Padilla C., Sanchez E., Tortorelli L.

The PAU Survey: Operation and orchestration of multi-band survey data
 // *Astronomy and Computing*. IV 2019. 27. 171.

Tonry J., Davis M.

A survey of galaxy redshifts. I. Data reduction techniques.
 // *AJ*. X 1979. 84. 1511–1525.

Tully R. B., de Marseille O., Fisher J. R.

A New Method of Determining Distances to Galaxies
 // *Bulletin of the American Astronomical Society*. 7. VI 1975. 426.

York Donald G., Adelman J., Anderson Jr. John E., Anderson Scott F., Annis James, Bahcall Neta A., Bakken J. A., Barkhouser Robert, Bastian Steven, Berman Eileen, Boroski William N., Bracker Steve, Briegel Charlie, Briggs John W., Brinkmann J., Brunner Robert, Burles Scott, Carey Larry, Carr Michael A., Castander Francisco J., Chen Bing, Colestock Patrick L., Connolly A. J., Crocker J. H., Csabai István, Czarapata Paul C., Davis John Eric, Doi Mamoru, Dombeck Tom, Eisenstein Daniel, Ellman Nancy, Elms Brian R., Evans Michael L., Fan Xiaohui, Federwitz Glenn R., Fiscelli Larry, Friedman Scott, Frieman Joshua A., Fukugita Masataka, Gillespie Bruce, Gunn James E., Gurbani Vijay K., de Haas Ernst, Haldeman Merle, Harris Frederick H., Hayes J., Heckman Timothy M., Hennessy G. S., Hindsley Robert B., Holm Scott, Holmgren Donald J., Huang Chi-hao, Hull Charles, Husby Don, Ichikawa Shin-Ichi, Ichikawa Takashi, Ivezić Željko, Kent Stephen, Kim Rita S. J., Kinney E., Klaene Mark, Kleinman A. N., Kleinman S., Knapp G. R., Korienek John, Kron Richard G., Kunszt Peter Z., Lamb D. Q., Lee B., Leger R. French, Limmongkol Siriluk, Lindenmeyer Carl, Long Daniel C., Loomis Craig, Loveday Jon, Lucinio Rich, Lupton Robert H., MacKinnon Bryan, Mannery Edward J., Mantsch P. M., Margon Bruce, McGehee Peregrine, McKay Timothy A., Meiksin Avery, Merelli Aronne, Monet

David G., Munn Jeffrey A., Narayanan Vijay K., Nash Thomas, Neilsen Eric, Neswold Rich, Newberg Heidi Jo, Nichol R. C., Nicinski Tom, Nonino Mario, Okada Norio, Okamura Sadanori, Ostriker Jeremiah P., Owen Russell, Pauls A. George, Peoples John, Peterson R. L., Petravick Donald, Pier Jeffrey R., Pope Adrian, Pordes Ruth, Prosapio Angela, Rechenmacher Ron, Quinn Thomas R., Richards Gordon T., Richmond Michael W., Rivetta Claudio H., Rockosi Constance M., Ruthmansdorfer Kurt, Sandford Dale, Schlegel David J., Schneider Donald P., Sekiguchi Maki, Sergey Gary, Shimasaku Kazuhiro, Siegmund Walter A., Smee Stephen, Smith J. Allyn, Snedden S., Stone R., Stoughton Chris, Strauss Michael A., Stubbs Christopher, SubbaRao Mark, Szalay Alexander S., Szapudi Istvan, Szokoly Gyula P., Thakar Anirudda R., Tremonti Christy, Tucker Douglas L., Uomoto Alan, Vanden Berk Dan, Vogeley Michael S., Waddell Patrick, Wang Shu-i., Watanabe Masaru, Weinberg David H., Yanny Brian, Yasuda Naoki, SDSS Collaboration .

The Sloan Digital Sky Survey: Technical Summary

// *AJ*. IX 2000. 120, 3. 1579–1587.

Zoubian J., Kümmel M., Kermiche S., Apostolakos N., Chapon A., Ealet A., Franzetti P., Garilli B., Jullo E., Paioro L.

Instrument Simulations of the EUCLID/NISP Spectrometer

// *Astronomical Data Analysis Software and Systems XXIII*. 485. V 2014. 509.

(Astronomical Society of the Pacific Conference Series).

van Dokkum Pieter G.

Cosmic-Ray Rejection by Laplacian Edge Detection

// *PASP*. XI 2001. 113, 789. 1420–1427.