

UNIVERSIDAD DE LA LAGUNA  
Departamento de Astrofísica



*The low surface brightness Universe:  
the last frontier*

A dissertation submitted by  
**Raúl Infante Sainz**  
as a requirement for the degree of  
*Doctor of Philosophy in Astrophysics*  
by University of La Laguna



INSTITUTO DE ASTROFÍSICA DE CANARIAS  
San Cristóbal de La Laguna, Tenerife  
May 2021

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Thesis snapshot: 93e8af3  
Thesis examination date: (RAUL): May XX, 2021  
Thesis supervisor: Dr. Ignacio Trujillo Cabrera  
© Raúl Infante Sainz 2021

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

iii

*A Virgi, estrella guía que me ilumina en este Universo tan oscuro.  
A mis padres, hermanos, y abuelos, por su apoyo incondicional.  
A Yuri, por enseñarme a vivir el momento.  
Esta tesis es tan mía como vuestra.*

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### *Resumen*

El estudio del Universo de bajo brillo superficial supone una de las mayores oportunidades de descubrimientos en la Astronomía actual. Sin embargo, las técnicas tradicionales para reducir y tratar los datos astronómicos no son capaces de proporcionar imágenes en las que las estructuras de bajo brillo superficial puedan ser analizadas y estudiadas con suficiente nivel de detalle. Ir más profundo de 30 mag arcsec<sup>-2</sup> implica observar estructuras  $\sim 1500$  veces más débiles que el brillo del cielo más oscuro de la Tierra, y múltiples efectos sistemáticos lo hacen muy difícil (por ejemplo, el campo plano, la sobre-substracción del fondo del cielo, las reflexiones internas, la luz dispersada, los cirros Galácticos, etc.). En esta tesis se ha llevado a cabo una exploración del Universo de bajo brillo superficial con especial énfasis en los apartados más técnicos. Para corregir los efectos sistemáticos y poder estudiar las estructuras de bajo brillo superficial, se han desarrollado, mejorado y aplicado muchas técnicas a una gran cantidad de datos de telescopios muy diferentes. Se ha prestado especial atención al efecto de la función de dispersión de punto (PSF) para corregir la luz dispersada en las imágenes astronómicas. Un resultado notable es la detección de una estructura estelar de marea muy débil de la galaxia NGC 1052-DF4. Esto revela que la galaxia NGC 1052-DF4 está interactuando con otra galaxia cercana, explicando así su sorprendente falta de materia oscura. Las técnicas y herramientas desarrolladas están pensadas para ser utilizadas en la próxima generación de telescopios y cartografiados profundos con el objetivo de mejorar la calidad de los datos. Junto al desarrollo de esta tesis, también se ha desarrollado y madurado un programa para realizar investigaciones reproducibles: Maneage (del inglés, Managing data Lineage). El objetivo de Maneage es proporcionar un entorno totalmente controlado para llevar a cabo estudios científicos reproducibles. En este sentido, la práctica totalidad de las reducciones y el análisis de los datos astronómicos en esta tesis se han realizado utilizando Maneage.

### *Abstract*

The study of the low surface brightness Universe is one of our best opportunities for discovery in present-day Astronomy. However, the traditional techniques to reduce and treat astronomical data are not able to provide datasets in which low surface brightness structures can be analyzed and studied with a sufficient level of detail. Going beyond 30 mag arcsec<sup>-2</sup> requires the ability to observe structures  $\sim 1500$  times fainter than the darkest sky on Earth, and multiple systematic effects make it very hard (e.g., flat fielding, sky background oversubtraction, internal reflections, scattered light, Galactic cirrus, etc.). This thesis presents an extensive exploration of the low surface brightness Universe with especial emphasis on the most technical aspects. To correct the systematic effects and study the low surface brightness structures, many techniques have been developed, improved, and applied to numerous datasets from different telescopes. Special attention has been placed on the effect of the extended point spread function (PSF) to correct the scattered light in astronomical images. A remarkable result is the detection of the very faint tidal stellar structure of the galaxy NGC 1052-DF4. This reveals that NGC 1052-DF4 is interacting with another nearby galaxy, thus explaining the surprising lack of dark matter of the former. The techniques and tools developed are intended to be used in the next generation of wide and deep surveys with the goal of improving the quality of the data. Throughout the development of this thesis, a framework for reproducible research has also been developed and matured: Maneage (Managing data Lineage). The goal of Maneage is to provide a fully controlled environment to conduct reproducible scientific research. Almost all the reduction and analysis of astronomical data in this thesis has been performed using Maneage.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# Contents

|   |            |
|---|------------|
| <b>1 Introduction</b>                                       | <b>1</b>   |
| 1.1 The low surface brightness                              | 1          |
| 1.2 Low surface brightness: basic concepts                  | 13         |
| 1.3 Challenges in ultradeep imaging                         | 20         |
| 1.4 Deep imaging data processing: raw data reduction        | 23         |
| 1.5 Reproducibility in Science                              | 32         |
| 1.6 Thesis outline  | 32         |
| <b>2 Reproducibility in Science: Maneage</b>                | <b>35</b>  |
| 2.1 Introduction  | 35         |
| 2.2 Definition of important terms                           | 38         |
| 2.3 Principles of the proposed solution                     | 41         |
| 2.4 Implementation of Maneage                               | 45         |
| 2.5 Discussion and conclusions                              | 64         |
| <b>3 The point spread function</b>                          | <b>67</b>  |
| 3.1 Introduction  | 67         |
| 3.2 Constructing the extended SDSS PSFs                     | 68         |
| 3.3 Results   | 78         |
| 3.4 Discussion and conclusions                              | 86         |
| <b>4 The galaxy lacking dark matter NGC 1052-DF4</b>        | <b>89</b>  |
| 4.1 Introduction  | 89         |
| 4.2 Data  | 90         |
| 4.3 The spatial distribution of the globular cluster system | 94         |
| 4.4 Tidal stripping of the stellar distribution             | 98         |
| 4.5 Discussion  | 106        |
| 4.6 Conclusions   | 108        |
| <b>5 Replicability in ultradeep imaging</b>                 | <b>109</b> |
| 5.1 Introduction  | 109        |
| 5.2 The galaxy NGC 493                                      | 110        |
| 5.3 Datasets description                                    | 111        |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

|  |            |
|--|------------|
| 5.4 Results . . . . .                                    | 114        |
| 5.5 Conclusions . . . . .                                | 118        |
| <b>6 Ongoing and future work</b>                         | <b>121</b> |
| 6.1 Ultradeep imaging with amateur telescopes . . . . .  | 121        |
| 6.2 Star counting versus integrated photometry . . . . . | 126        |
| <b>7 Summary and conclusions</b>                         | <b>131</b> |
| <b>Bibliography</b>                                      | <b>135</b> |
| <b>A Appendix: Maneage tutorial</b>                      | <b>147</b> |
| A.1 Maneage tutorial for dummies . . . . .               | 147        |
| <b>B Appendix: Contributions during this thesis</b>      | <b>157</b> |
| B.1 Publications during this thesis . . . . .            | 157        |
| B.2 Conference proceedings . . . . .                     | 157        |
| B.3 Oral contributions . . . . .                         | 157        |
| <b>Acknowledgements</b>                                  | <b>159</b> |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

## Contents (detailed)

|   |          |
|---|----------|
| <b>1 Introduction</b>   | <b>1</b> |
| 1.1 The low surface brightness  | 1        |
| 1.1.1 The importance of low surface brightness astronomy                            | 2        |
| 1.1.2 Early studies   | 5        |
| 1.1.2.1 Galactic stellar halos and intracluster light                               | 5        |
| 1.1.2.2 Galactic cirrus and ultra-diffuse galaxies                                  | 7        |
| 1.1.3 State-of-the-art surveys  | 10       |
| 1.1.4 Ongoing and future deep surveys   | 12       |
| 1.2 Low surface brightness: basic concepts  | 13       |
| 1.2.1 Sources of noise in astronomical imaging                                      | 13       |
| 1.2.1.1 Photon noise  | 14       |
| 1.2.1.2 Dark noise  | 15       |
| 1.2.1.3 Readout noise   | 15       |
| 1.2.1.4 Total noise and signal-to-noise ratio                                       | 15       |
| 1.2.2 Magnitude system and surface brightness                                       | 16       |
| 1.2.3 Depth limits  | 18       |
| 1.2.3.1 Point source limiting magnitude   | 18       |
| 1.2.3.2 Surface brightness limiting magnitude                                       | 18       |
| 1.2.4 Expected surface brightness limit for different telescopes and exposure times | 19       |
| 1.3 Challenges in ultradeep imaging   | 20       |
| 1.3.1 Sky brightness (airglow)  | 20       |
| 1.3.2 Internal reflections  | 20       |
| 1.3.3 The point spread function: scattered light                                    | 21       |
| 1.3.4 Flat fielding   | 21       |
| 1.3.5 Galactic cirrus   | 22       |
| 1.4 Deep imaging data processing: raw data reduction                                | 23       |
| 1.4.1 Proper observational strategy   | 23       |
| 1.4.2 Bias and dark correction  | 24       |
| 1.4.2.1 Bias  | 24       |
| 1.4.2.2 Dark current  | 25       |
| 1.4.3 Flat field correction   | 26       |
| 1.4.3.1 Difficulties to create an auto flat field image                             | 28       |
| 1.4.4 Sky background correction   | 28       |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

|          |   |           |
|----------|---|-----------|
| 1.4.5    | Astrometric solution . . . . .  | 31        |
| 1.4.6    | Photometric calibration . . . . .   | 31        |
| 1.5      | Reproducibility in Science . . . . .                                      | 32        |
| 1.6      | Thesis outline . . . . .  | 32        |
| <b>2</b> | <b>Reproducibility in Science: Maneage</b>                                | <b>35</b> |
| 2.1      | Introduction . . . . .  | 35        |
| 2.2      | Definition of important terms . . . . .                                   | 38        |
| 2.2.1    | Definition: input . . . . .   | 38        |
| 2.2.2    | Definition: output . . . . .  | 39        |
| 2.2.3    | Definition: project . . . . .   | 39        |
| 2.2.4    | Definition: data provenance . . . . .                                     | 39        |
| 2.2.5    | Definition: data lineage . . . . .  | 40        |
| 2.2.6    | Definition: reproducibility and replicability . . . . .                   | 40        |
| 2.3      | Principles of the proposed solution . . . . .                             | 41        |
| 2.3.1    | Principle 1: completeness . . . . .                                       | 41        |
| 2.3.2    | Principle 2: modularity . . . . .   | 42        |
| 2.3.3    | Principle 3: plain text . . . . .   | 43        |
| 2.3.4    | Principle 4: minimal complexity . . . . .                                 | 43        |
| 2.3.5    | Principle 5: verifiable inputs and outputs . . . . .                      | 44        |
| 2.3.6    | Principle 6: history and temporal provenance . . . . .                    | 44        |
| 2.3.7    | Principle 7: free and open-source software . . . . .                      | 44        |
| 2.4      | Implementation of Maneage . . . . .                                       | 45        |
| 2.4.1    | Job orchestration with Make . . . . .                                     | 46        |
| 2.4.2    | General implementation structure . . . . .                                | 47        |
| 2.4.3    | Project configuration . . . . .   | 47        |
| 2.4.3.1  | Setting local directories . . . . .                                       | 49        |
| 2.4.3.2  | Checking for a C compiler . . . . .                                       | 49        |
| 2.4.3.3  | Building necessary software from source . . . . .                         | 49        |
| 2.4.3.4  | Software citation . . . . .   | 51        |
| 2.4.4    | High-level organization of the analysis . . . . .                         | 51        |
| 2.4.4.1  | Isolated analysis environment . . . . .                                   | 52        |
| 2.4.4.2  | Preparation phase . . . . .   | 52        |
| 2.4.5    | Low-level organization of analysis . . . . .                              | 53        |
| 2.4.5.1  | Non-recursive Make . . . . .  | 55        |
| 2.4.5.2  | Ultimate target: the report of the project ( <i>paper.pdf</i> ) . . . . . | 55        |
| 2.4.5.3  | Values within the report ( <i>project.tex</i> ) . . . . .                 | 55        |
| 2.4.5.4  | Verification of outputs ( <i>verify.mk</i> ) . . . . .                    | 56        |
| 2.4.5.5  | Project initialization ( <i>initialize.mk</i> ) . . . . .                 | 57        |
| 2.4.5.6  | Importing and validating inputs ( <i>download.mk</i> ) . . . . .          | 57        |
| 2.4.6    | The analysis . . . . .  | 58        |
| 2.4.7    | Projects as Git branches of Maneage . . . . .                             | 59        |
| 2.4.8    | Multi-user collaboration on single build directory . . . . .              | 61        |
| 2.4.9    | Publishing the project . . . . .  | 61        |
| 2.4.9.1  | Automatic creation of publication tarball . . . . .                       | 61        |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

|          |  |            |
|----------|--|------------|
| 2.4.9.2  | What to publish, and where? . . . . .                                  | 62         |
| 2.4.9.3  | Worries about getting scooped . . . . .                                | 63         |
| 2.4.10   | Past, present, and future of Maneage . . . . .                         | 64         |
| 2.5      | Discussion and conclusions . . . . .                                   | 64         |
| <b>3</b> | <b>The point spread function</b> . . . . .                             | <b>67</b>  |
| 3.1      | Introduction . . . . .   | 67         |
| 3.2      | Constructing the extended SDSS PSFs . . . . .                          | 68         |
| 3.2.1    | Constructing the outer part of the PSFs . . . . .                      | 69         |
| 3.2.2    | Constructing the intermediate part of the PSFs . . . . .               | 73         |
| 3.2.3    | Constructing the central part of the PSFs . . . . .                    | 74         |
| 3.2.4    | PSF junction (central-intermediate-outer parts) . . . . .              | 76         |
| 3.2.5    | Final refinements to the extended PSFs . . . . .                       | 78         |
| 3.3      | Results . . . . .  | 78         |
| 3.3.1    | PSFs characteristics . . . . .   | 80         |
| 3.3.2    | Scattered light subtraction . . . . .                                  | 83         |
| 3.4      | Discussion and conclusions . . . . .                                   | 86         |
| <b>4</b> | <b>The galaxy lacking dark matter NGC 1052-DF4</b> . . . . .           | <b>89</b>  |
| 4.1      | Introduction . . . . .   | 89         |
| 4.2      | Data . . . . .   | 90         |
| 4.2.1    | <i>HST</i> imaging . . . . .   | 90         |
| 4.2.2    | Deep ground-based imaging . . . . .                                    | 91         |
| 4.2.2.1  | GTC HiPERCAM images . . . . .  | 92         |
| 4.2.2.2  | IAC80 . . . . .  | 93         |
| 4.2.2.3  | PSF modeling and bright star subtraction of the IAC80 images . . . . . | 93         |
| 4.3      | The spatial distribution of the globular cluster system . . . . .      | 94         |
| 4.3.1    | GC selection . . . . .   | 95         |
| 4.3.2    | The distribution of GCs . . . . .                                      | 97         |
| 4.4      | Tidal stripping of the stellar distribution . . . . .                  | 98         |
| 4.4.1    | Removal of NGC 1035 . . . . .  | 99         |
| 4.4.2    | Surface brightness radial profiles . . . . .                           | 100        |
| 4.4.3    | Defining $r_{\text{break}}$ and $r_{\text{distort}}$ . . . . .         | 102        |
| 4.4.4    | The S shape of NGC 1052-DF4 . . . . .                                  | 103        |
| 4.4.5    | Tidal radius . . . . .   | 104        |
| 4.4.6    | Stellar mass in the tidal tails . . . . .                              | 105        |
| 4.5      | Discussion . . . . .   | 106        |
| 4.5.1    | The distribution of light and GCs . . . . .                            | 106        |
| 4.5.2    | Can tidal stripping explain the missing dark matter? . . . . .         | 107        |
| 4.6      | Conclusions . . . . .  | 108        |
| <b>5</b> | <b>Replicability in ultradeep imaging</b> . . . . .                    | <b>109</b> |
| 5.1      | Introduction . . . . .   | 109        |
| 5.2      | The galaxy NGC 493 . . . . .   | 110        |
| 5.3      | Datasets description . . . . .   | 111        |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

|  |            |
|--|------------|
| 5.3.1 Sloan Digital Sky Survey (SDSS) . . . . .                              | 111        |
| 5.3.2 The IAC Stripe 82 Legacy Project . . . . .                             | 112        |
| 5.3.3 Las Cumbres Observatory Telescope . . . . .                            | 112        |
| 5.3.4 Gran Telescopio Canarias . . . . .                                     | 113        |
| 5.4 Results . . . . .  | 114        |
| 5.4.1 Typical distance between sources . . . . .                             | 114        |
| 5.4.2 How far are current optical images from the confusion limit? . . . . . | 116        |
| 5.5 Conclusions . . . . .  | 118        |
| <b>6 Ongoing and future work</b>   | <b>121</b> |
| 6.1 Ultradeep imaging with amateur telescopes . . . . .                      | 121        |
| 6.1.1 Telescope and observing strategy . . . . .                             | 122        |
| 6.1.2 Data reduction . . . . .   | 123        |
| 6.1.3 Results and conclusions . . . . .                                      | 124        |
| 6.2 Star counting versus integrated photometry . . . . .                     | 126        |
| 6.2.1 Data analysis . . . . .  | 127        |
| 6.2.2 Results and conclusions . . . . .                                      | 128        |
| <b>7 Summary and conclusions</b>   | <b>131</b> |
| <b>Bibliography</b>  | <b>135</b> |
| <b>A Appendix: Maneage tutorial</b>  | <b>147</b> |
| A.1 Maneage tutorial for dummies . . . . .                                   | 147        |
| <b>B Appendix: Contributions during this thesis</b>                          | <b>157</b> |
| B.1 Publications during this thesis . . . . .                                | 157        |
| B.2 Conference proceedings . . . . .   | 157        |
| B.3 Oral contributions . . . . .   | 157        |
| <b>Acknowledgements</b>  | <b>159</b> |

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 1

## Introduction

Intending to be complete and modular, each chapter of this thesis is self-contained. This means that the reader is free to read any part independently of the others. But, of course, all chapters have some relations between them.

- In Chapter 1, basic concepts about astronomy and low surface brightness, as well as the reproducibility problem, are introduced.
- In Chapter 2, Maneage is presented as a framework to solve the problem of reproducibility in scientific research.
- Chapter 3 treats one of the most important challenges in low surface brightness studies: the point spread function (PSF).
- In Chapter 4, the “problem” of the galaxy missing dark matter (NGC 1052-DF4) is solved by obtaining and analyzing ultra-deep imaging data.
- In Chapter 5, deep imaging data from different telescopes and instruments are used to analyze the galaxy NGC 493, study the replicability of the detected low surface brightness features and extract useful results.
- In Chapter 6, future and ongoing work is presented.
- Finally, Chapter 7 lists the main conclusions of the thesis.

### 1.1 The low surface brightness

In this thesis, the term *low surface brightness* is going to be used almost in every phrase, and it is actually the fundamental concept around which this thesis orbits. However, there is not a formal definition of what “low surface brightness” is. While the words “surface” and “brightness” have a clear definition (see Section 1.2), the ill-defined term is “low”. What does “low” mean? Some entries in the dictionary<sup>1</sup> are: not measuring much from the base to the top; close to the ground

<sup>1</sup>Cambridge dictionary (25/02/2021): <https://dictionary.cambridge.org/dictionary/english/low>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

or the bottom of something; below the usual level. Intuitively, it is clear what it means. However, it is important to note that when using “low”, a reference level is required. Indeed, the key point for the discussion is that the reference level changes over time. Consider, for example, the Persian astronomer Abd al-Rahman al-Sufi (903–986 A.D.), who was the earliest person describing the Andromeda galaxy (M31) as a small cloud. For this astronomer, a low surface brightness structure would be anything below the limit of what he was able to detect with the naked eye. To put things into context, such a structure would be the core of the galaxy M31, which has a central surface brightness of  $\mu_V \sim 16 \text{ mag arcsec}^{-2}$ . The typical sky brightness on a dark night with no light pollution (for example, in professional observatories) is  $\mu_V \sim 22 \text{ mag arcsec}^{-2}$  (Patat 2003). This has been historically what astronomers have considered low surface brightness: the regime where the brightness of an object is a small fraction below the brightness of the night sky.

With the invention of the telescope, such a reference was pushed toward much fainter objects. Essentially, telescopes are instruments whose capacity for collecting photons is big just because the area of the mirror is large. Nowadays optical telescopes have apertures from a few centimeters up to the 10.4 meters (in diameter) of the Gran Telescopio Canarias (GTC). Compared with the pupil size of the human eye ( $\sim 4 \text{ mm}$  in diameter), it is a huge step forward in the power of collecting photons. Therefore, the possibility of detecting fainter objects increases as telescopes become larger and analysis techniques evolve.

### 1.1.1 The importance of low surface brightness astronomy

In general, being in the frontier of what can be studied is always a source of discovery and a way of increasing both fundamental and practical knowledge. In particular, the information encoded in the low surface brightness Universe is absolutely key in the understanding of the ultimate nature of galaxy formation and evolution. One of the current and well-established theories to explain how the Universe is formed and evolves is the Lambda Cold Dark Matter ( $\Lambda$ CDM) model (Blumenthal et al. 1984; Navarro et al. 1996; Peebles & Ratra 2003). Many observational facts support this particular model, but reviewing them is not the main topic of the current thesis. The relevant point is that independently of the model considered, it must give valuable predictions that should be empirically testable. Richard P. Feynman resumed this fact in one of his famous lectures at Cornell University in 1964:

*“I’m going to discuss how we would look for a new law. First, we guess it [...] Then we compute the consequences of the guess to see what it would imply. And then we compare those computation results [...] directly to observation to see if it works. If it disagrees with the experiment, it is wrong. In that simple statement, is the key to science. It doesn’t make any difference how beautiful your guess is, it doesn’t make any difference how smart you are, who made the guess, or what his name is. If it disagrees with the experiment, it’s wrong. That’s all there is to it.”*

—Richard P. Feynman (Cornell University, 1964)

Richard P. Feynman was absolutely right. It is important to bear in mind that in astronomy, it is very difficult to make laboratory experiments to test hypotheses and predictions. To do that, it is almost only possible to observe the Universe and make numerical simulations.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

In particular, numerical simulations considering the  $\Lambda$ CDM model predict that almost all galaxies have experienced mergers (see e.g., Bullock & Johnston 2005; Purcell et al. 2007; Cooper et al. 2010). The merging of galaxies occurs along with the evolution of the galaxies as a consequence of collisions with neighbor galaxies. Structures formed as a consequence of such events are shells and tidal streams in the outskirts of galaxies, as well as the smooth accretion of material in the halos. This can be seen in the simulated Milky Way mass stellar halos from Cooper et al. (2010), where the majority of sub-structures belonging to the stellar halo have surface brightness  $\mu > 28$  mag arcsec<sup>-2</sup> (see Figure 1.1). From the observational point of view, to be able to detect and characterize such galaxy stellar halos with the aim of test the simulated predictions is absolutely key (see e.g., Bullock & Johnston 2005; Johnston et al. 2008; Martínez-Delgado et al. 2010; Duc et al. 2015; Trujillo & Fliri 2016). The recent case in which the detection of a faint tidal stream of NGC 1052-DF4 explained the missing dark matter of this galaxy is a perfect example of the necessity of going extremely deep. This study is analyzed in detail in Chapter 4.

Also predicted by the  $\Lambda$ CDM cosmological simulations is the intracluster light (ICL) (see e.g., Rudick et al. 2011; Contini et al. 2014; Pillepich et al. 2018). The ICL comes from the stars that are bound to the gravitational potential of the cluster. As a consequence, the study of the ICL probes the theoretical models (see e.g., Uson et al. 1991; Mihos et al. 2005; Rudick et al. 2010; Montes & Trujillo 2014, 2018). In addition to this, by comparing the amount of light and how it is distributed within the cluster it is possible to trace the distribution of the dark matter of the cluster (Montes & Trujillo 2019; Alonso Asensio et al. 2020).

The  $\Lambda$ CDM cosmological model also predicts the number of low-mass satellites in galaxies. In the case of Milky Way-like galaxies, the number of dark matter subhalos predicted is about one order of magnitude higher than the number of satellites observationally detected. This problem is known as “The missing satellite problem”, and it is one of the fundamental questions to test in the  $\Lambda$ CDM model (e.g., Moore et al. 1999; Klypin et al. 1999; Springel et al. 2008; Madau et al. 2008). With the advent of large and deep astronomical surveys, the number of known satellite galaxies in the Local Group has increased. But it is still an ongoing debate whether the problem of the missing satellites is still there or not (Kim et al. 2018; Read & Erkal 2019).

There are two different approaches to study the stellar population of extended low surface brightness sources: individual star counting and integrated light photometry. The most reliable one is to obtain physical information from individual stars and studying their properties. This technique has important advantages such as avoiding the confusion of sources when the field is crowded, as well as not mixing flux from the foreground and background sources that cannot be accounted for. Some of the most relevant observations for star counting measurements are the Pan-Andromeda Archaeological Survey (PAndAS; Martin et al. 2006; Ibata et al. 2007; McConnachie et al. 2008) and the Galaxy Halos, Outer disks, Substructure, Thick disks, and Star clusters survey (GHOSTS; Radburn-Smith et al. 2011; Monachesi et al. 2016). However, the use of the star counting technique is only possible when the astronomical source is relatively nearby. For observations with the *Hubble Space Telescope (HST)*, this distance is  $\sim 16$  Mpc (Zackrisson et al. 2012). With any observations from ground-based telescopes, this distance becomes smaller as a consequence of the poorer spatial resolution of the telescope. Therefore, for objects further than 16 Mpc, it is only possible to use integrated light photometry to measure its properties. Taking into account that the vast majority of objects are further away than this

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

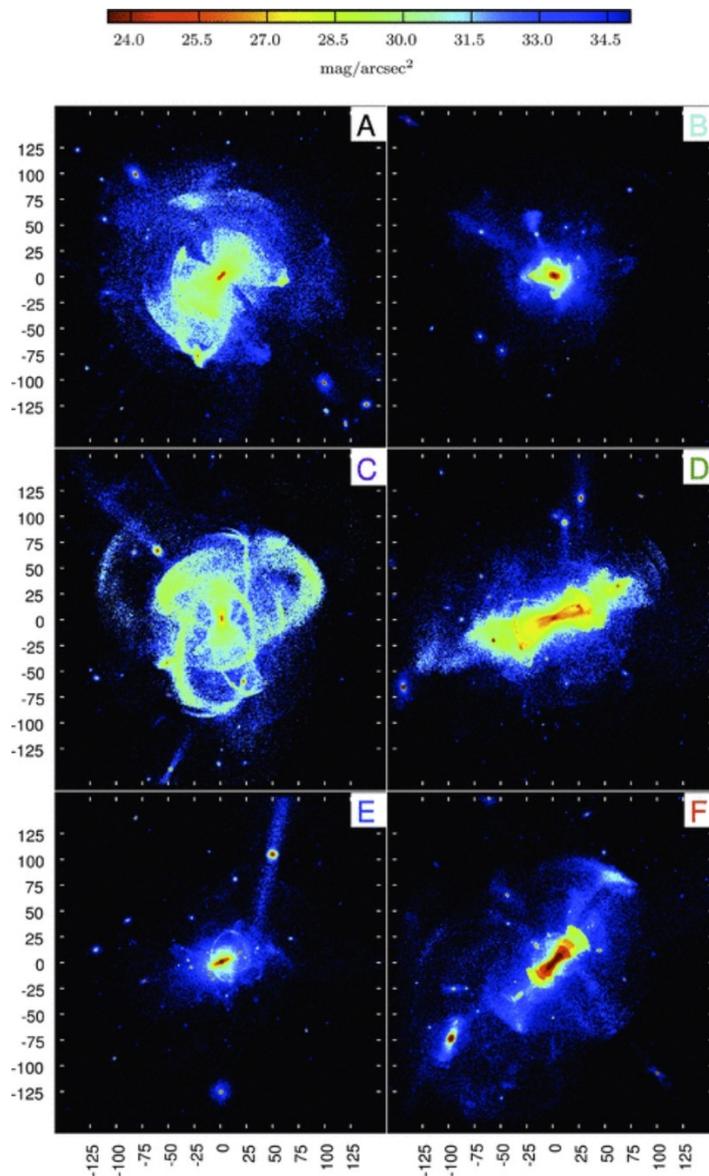


Figure 1.1: Simulated V-band surface brightness stellar maps of halos in Milky Way-like galaxies ( $z = 0$ ). Limiting depth is  $35 \text{ mag arcsec}^{-2}$ , axes are in units of kiloparsecs. Only stars formed in satellites are present in these plots. To avoid confusion, there is no contribution to these maps from a central galactic disk or bulge formed in-situ. Credit: Cooper et al. (2010).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

distance (16 Mpc and it will be only possible to observe them by doing very deep observations), it is absolutely key to ensure that both techniques give the same results when studying the same objects. This analysis is work in progress and it is explained in more detail in Chapter 6, where we reduce very deep data from the William Herschel Telescope (WHT) with the Physics of the Accelerating Universe Camera (PAUCam; Padilla et al. 2019) to obtain the characteristics of the stellar halo and the warp of the disk of the galaxy NGC 4565, and compare the results with the GHOSTS survey.

The reasons discussed above are only a few of the most important points for studying the low surface brightness Universe. In what follows, some of the first discoveries and state-of-the-art surveys are reviewed.

### 1.1.2 Early studies

#### 1.1.2.1 Galactic stellar halos and intracluster light

In 1969, two independent works found the same light envelope around the giant elliptical galaxy M87, with low surface brightness and extremely large in extension (diameter of 1 deg on the sky, which corresponds to 0.3 Mpc at the distance of the Virgo Cluster). de Vaucouleurs (1969) using a photoelectric photometer at the 36 inch telescope in the McDonald Observatory claimed:

*“A faint outer corona surrounding the main body of M87, [...], has been traced out to a minimum diameter in excess of one degree at a brightness level of 1 per cent of the night sky,  $\mu_B \simeq 27.3 \text{ mag arcsec}^{-2}$ .”*

—Gerard de Vaucouleurs (1969)

At the same time, Arp & Bertola (1969) used the 48 inch Palomar Schmidt telescope to obtain photographic plates and made the first image of such structure (see Figure 1.2). Note that in both cases, such deep observations were possible due to the use of the most advanced instrumentation at that time. After these first studies of M87, Arp & Bertola (1971) analyzed several nearby elliptical galaxies. Kormendy & Bahcall (1974) studied groups of galaxies and also isolated galaxies of various morphological types.

*“Photographic surface photometry is presented for 12 fields containing small and moderately rich groups of galaxies, as well as isolated galaxies of various morphological types. Surface brightnesses of 26–28 green mag per square arcsec are reached using isodensity tracings with small density steps. [...] Observations confirm the existence of faint envelopes  $\gtrsim 100 \text{ kpc}$  in diameter previously reported for elliptical galaxies by de Vaucouleurs (1969) and Arp and Bertola (1969, 1971). When measured at the same surface brightness, spiral galaxies are not substantially smaller than elliptical of the same luminosity. Some groups of galaxies dominated by ellipticals are imbedded in common luminous envelopes up to 1.3 Mpc in diameter.”*

—John Kormendy & John N. Bahcall (1974)

As cited above, these works showed the existence of envelopes or stellar halos of very low surface brightness around almost any type of galaxy (not only in ellipticals). The diffuse light was also found in the space between galaxies forming a group or cluster, and because of that, it is known as intracluster light. Interestingly, two decades earlier, the intracluster light had also been observed in the Coma Cluster by Zwicky (1951).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

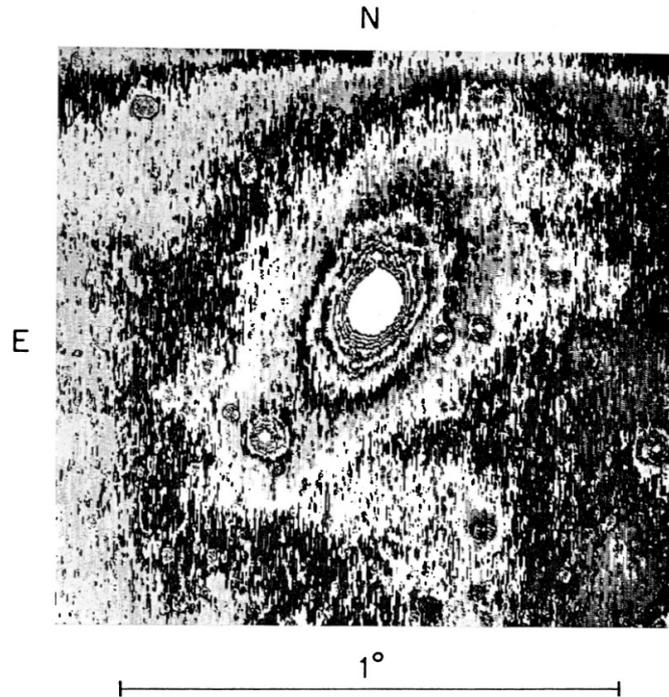


Figure 1.2: Image of the galaxy M87 obtained by [Arp & Bertola \(1969\)](#). This image corresponds to the “isodensity tracing” of the IIIa-J plate. It is the first image of a galactic stellar halo. The outermost isophote encloses a diameter of  $\sim 1$  deg. Credit: [Arp & Bertola \(1969\)](#).

*“One of the most interesting discoveries made in the course of this investigation is the observation of an extended mass of luminous intergalactic matter of very low surface brightness. The objects which constitute this matter must be considered as the faintest individual members of the cluster.”*

—Fritz Zwicky (1951)

Surprisingly, from the analysis of observations and theoretic predictions, [Zwicky \(1933\)](#) had previously concluded that the amount of non-luminous matter in the Universe should be greater than that of luminous matter (becoming an early precursor of the dark matter idea). As explained above (Section 1.1.1), the analysis of the intracluster light is one of the ways to study the nature of the dark matter and its distribution.

All of the previous studies show the systematic presence of diffuse structures such as envelopes around galaxies or light between galaxies of a cluster that are only visible in very deep

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

observations, indicating the presence of a considerable amount of previously hidden material.

### 1.1.2.2 Galactic cirrus and ultra-diffuse galaxies

The emission of light from the interstellar material of our own Galaxy (Galactic cirrus) is probably one of the earliest observations of what can be called low surface brightness astronomy. Observations from [Elvey & Roach \(1937\)](#) showed evidence of what they named “galactic light, which is probably produced by the scattering of starlight by interstellar matter”. Then, [Henryey & Greenstein \(1941\)](#) conducted observations (using a Fabry photometer attached to the 40 inch refractor at the Yerkes Observatory) aiming to “verify the existence of diffuse interstellar radiation”.

The observations of [Sandage \(1976\)](#) are considered the first report of Galactic cirrus using photographic plates. Allan Sandage used the Palomar 1.2m Schmidt telescope to observe “Extended regions of faint nebulosity at high galactic latitudes”. With more precision than previous studies (especially with larger spatial resolution), Galactic cirrus were described as “diffuse and filamentary structures on scales of 30 arcsec in connective patterns over scales of degrees” ([Sandage 1976](#)).

The fact that the night sky brightness is not zero is one of the main constraints to the detectability and observability of astronomical sources. Consider for example a galaxy whose apparent brightness is fainter than the sky brightness. In this scenario, it would be extremely difficult to observe and detect such an object. Depending on the relative difference in brightness, it would be necessary to spend more or less time observing the object to finally detect it over the sky signal (in Section 1.2 it is explained why it is possible to observe sources fainter than the sky brightness). It was [Disney \(1976\)](#) who noticed this effect and concluded that there could be a galactic population of low surface brightness that astronomers were not observing as they would be hidden. It is necessary to cite his own words to show the absolutely clever and simple way of arriving at such a conclusion:

*“It is well known that our counts of galaxies could be seriously biased by selection effects, largely determined by the brightness of the night sky. To illustrate this, suppose the Earth were situated near the centre of a giant elliptical galaxy, then the mean surface brightness of the sky would appear some 8–9 mag brighter than is observed from our position in the Galaxy ( $\mu_V \sim 23 \text{ mag arcsec}^{-2}$  looking toward the galactic pole, discounting atmospheric and zodiacal contributions). Optical astronomers would then find extragalactic space an empty void; spiral and irregular galaxies would be quite invisible and all they would easily detect of galaxies would be the core regions of ellipticals very similar to their own. They would be blinded to much of the Universe by the surface brightness of their parent galaxy. But this blinding is clearly a relative matter and we should ask to what extent we are blinded by the spiral galaxy in which we exist, faint as it may appear by comparison. I will argue that strong indirect evidence already exists that our knowledge of galaxies is heavily biased by the sky background, and that the true population of extra-galactic space may be very different from the one we can see.”*

—Michael J. Disney, (1976)

In the reasoning of Michael J. Disney, the main point is the relative brightness of the sky

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

due to being inside a galaxy. However, the fundamental limitation for observing such faint structures is the brightness of the sky due to the emission of the atmosphere during the night. Independently of the origin, the fact is that there is a bias. In this sense, the consequence is not only that there are objects that are not detected, but also structures forming part of already detected objects that remain hidden (as previous studies had shown): “Galaxies are like icebergs and what is seen above the sky background may be no reliable measure of what lies underneath.” (Disney 1976).

Several years later Disney’s prediction was confirmed. Sandage & Binggeli (1984) found very extended dwarf-like low surface brightness galaxies in the Virgo Cluster, with central surface brightness more than ten times fainter than the night sky brightness ( $\mu_V > 25 \text{ mag arcsec}^{-2}$ ) and very extended. These kinds of objects were not seen until deep observations using photographic plates were obtained. Moreover, Sandage & Binggeli (1984) pointed out the difficulty in observing such objects not only because of their low brightness but also because of their extension on the sky (this is, in fact, one of the challenges in low surface brightness studies nowadays, see Section 1.4.4):

*“They will be particularly hard to discover because of the large diameters of  $\sim 10\,000 \text{ pc}$  and central SB fainter than  $\sim 25 \text{ mag arcsec}^{-2}$ . Examples as close as M31 would, then, subtend  $\sim 3/4$  degrees on the plane of the sky and would be  $\sim 10$  times fainter in surface brightness than the night sky. These would surely have been missed to date.”*  
 —Allan Sandage & Bruno Binggeli (1984)

In the following years, a set of large and systematic surveys were designed with the goal of detecting and measuring extremely low surface brightness galaxies (see e.g., Bothun et al. 1987; Schombert & Bothun 1988; Ferguson & Sandage 1988; Impey et al. 1988; Bothun et al. 1991). In this context, Bothun et al. (1987) discovered the iconic Malin 1, the most extended low surface brightness disk ever detected to date. Figure 1.3 shows the image of this galaxy, both the raw data and the photographically amplified. It is important to remark that this galaxy becomes apparent only after the raw image has been amplified, indicating that the process of reducing or treating the data is absolutely key in detecting such faint objects.

*“The galaxy of interest, which we christen Malin 1, ( $\alpha = 12^h 34^m 27.3^s$ ,  $\delta = 14^\circ 36' 15''$ ), is shown in Fig. 1 [Plate 1], where the photographically amplified image (derived from three deep IIIa-J plates) is shown in comparison to the normal image from the U.K. Schmidt plate. As is apparent, the process of amplification has detected an extended LSB image.”*

—Bothun et al. (1987)

Several large-scale astronomical surveys have been conducted since the 1950s. One of the most famous was the Palomar Observatory Sky Survey (POSS-I; Minkowski & Abell 1963), which used the 1.2 m Schmidt telescope in the Palomar Observatory in California. POSS-I consists of  $\sim 1000$  photographic plates, each with a field of view of  $6.5^\circ \times 6.5^\circ$  in two wavelengths (optical blue and red bands). Similar surveys were conducted in the Southern Sky. The Anglo Australian Observatory (AAO) using the 1.2 m U.K. Schmidt telescope, and the European Southern Observatory (ESO) using the 1.0 m Schmidt telescope observed the skies in the 1970s–90s (see e.g., Reshetnikov 2005; Djorgovski et al. 2013).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

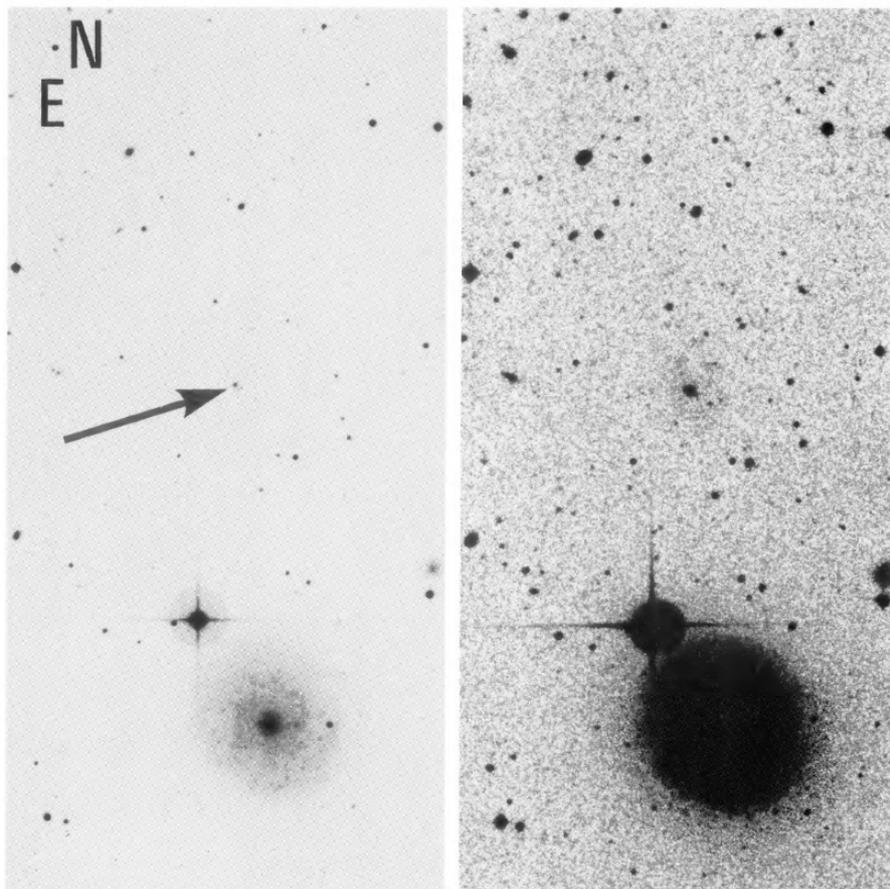


Figure 1.3: Image of Malin 1. On the left, the raw image from the photographic plate of the U.K. Schmidt telescope. On the right, the same image but photographically amplified. The arrow points to the object of interest: Malin 1. The prominent galaxy to the south is NGC 4571. Credit: [Bothun et al. \(1987\)](#).

New techniques with the aim of detecting fainter objects began to be developed. The most relevant techniques in this sense were photographic amplification (see [Figure 1.3](#)) and the stacking of many different images to increase the signal-to-noise ratio of the final image ([Malin 1978](#); [Malin & Carter 1980](#); [Malin 1981](#)). At the same time these new techniques were explored, pho-

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

tographic plates started to be digitized (see e.g., the review of Reshetnikov 2005). The first digital surveys come from the digitization of photographic plates: Digital Sky Survey (DSS; Lasker et al. 1989) and Palomar Digital Sky Survey (DPOSS; Djorgovski et al. 1998).

Charge-coupled devices (CCDs) appeared in 1969 (Boyle & Smith 1970; Amelio et al. 1970), later awarded the Nobel prize of physics in 2009. However, it was not until the 1970s when this technology developed rapidly and it started to be clear that it was going to replace photographic techniques. The higher quantum efficiency, as well as the reduced noise and the improved cosmetics, demonstrated that this technology was superior in all fronts for astronomical data acquisition (see e.g., Lesser 2015, and references therein for a review of the CCD history). The *Hubble Space Telescope* mission at the NASA Jet Propulsion Laboratory (JPL) between 1973–1979 developed this new technology (McLean 2002). For that reason, the Hubble wide field and planetary camera (WFPC) is considered the first CCD camera in an astronomy observatory. With the fast and wide adoption of CCD cameras, the era of digital astronomy had begun.

### 1.1.3 State-of-the-art surveys

The adoption of CCD cameras was a revolution in astronomy. With this technology, the astronomical surveys come directly digitized, which makes the process of data analysis much easier.

One of the most important astronomical surveys of all times is the Sloan Digital Sky Survey (SDSS; York et al. 2000). The SDSS used the 2.5 m telescope at the Apache Point Observatory in New Mexico. The SDSS imaging camera scanned the sky in strips along great circles in drift scan or time-delay integration mode (TDI; McGraw et al. 1980; Wright & Mackay 1981). This mode consists in opening the shutter for extended periods and imaging a continuous strip of the sky. Although this method is not free of technical complications (Gibson & Hickson 1992), it allows high accuracy in the flat field correction. The survey has gone through several phases, starting with the imaging and spectroscopic observations (SDSS-I 2000–2005, SDSS-II 2005–2008; York et al. 2000). During these phases, it observed more than 15 000 square degrees of the northern sky in five photometric bands ( $u$ ,  $g$ ,  $r$ ,  $i$ , and  $z$ ), and obtained spectra for hundred of thousands of astronomical objects. As a result of this huge amount of data taken in a systematic way, statistical analysis of astronomical objects became possible, thus giving us a better view of how the Universe is constituted (see e.g., Shen et al. 2003; Blanton et al. 2003; Tegmark et al. 2004; Bruzual & Charlot 2003). The imaging camera was retired in late 2009, and since then, the telescope only observed in spectroscopic mode (SDSS-III 2008–2014, SDSS-IV 2014–2020; Eisenstein et al. 2011; Blanton et al. 2017) with the aim of increasing the precision of cosmological measurements, as well as obtaining high-resolution spectra of stars and galaxies (APOGEE, SEGUE, BOSS, MaStar, eBOSS, MaNGA). In terms of surface brightness, the SDSS has a depth of  $\mu_g \sim 27 \text{ mag arcsec}^{-2}$  (see Section 1.2.3.2 for more details on the surface brightness limit and how it is calculated). Data from SDSS have been extensively analyzed throughout this thesis. For example, in Chapter 3, images of very bright stars are used to construct the extended point spread function of the SDSS telescope. The SDSS is not considered a deep survey by itself, but because of its importance, it is widely used and considered as the reference for other surveys.

A by-product of the SDSS imaging is the Stripe 82 survey. The Stripe 82 is a 2.5-degree wide stripe along the celestial equator in the southern Galactic cap ( $-50^\circ < \text{R.A.} < 60^\circ$ ,  $-1.25^\circ < \text{dec} < 1.25^\circ$ ) with a total area of 275 square degrees observed in all the five SDSS bands. The region was imaged multiple times by the SDSS between 1998 and 2004 as part of the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Legacy Survey in 82 runs under photometric conditions with generally good seeing and low sky background (Annis et al. 2014; Fliri & Trujillo 2016; Román & Trujillo 2018). Because different regions of the sky have been observed at different times, the depth of the survey varies slightly, but on average the surface brightness limit of Stripe 82 is  $\mu_r \sim 28.5 \text{ mag arcsec}^{-2}$  (see Chapter 5 for more details). Stripe 82 does not cover as much sky area as SDSS, but many works have taken advantage of its depth to study fundamental questions related to the low surface brightness Universe (see e.g., Frieman et al. 2008; Annis et al. 2014; Peters et al. 2017; Martínez-Lombilla & Knapen 2019; Román et al. 2020; Chamba et al. 2020; Trujillo et al. 2020). Other dedicated surveys can go deeper than SDSS, or roughly reach the same depth as Stripe 82 (e.g., Mihos et al. 2005; Ferrarese et al. 2012; Merritt et al. 2014; Capaccioli et al. 2015; van Dokkum et al. 2015; Duc et al. 2015; Koda et al. 2015; Martínez-Delgado et al. 2016; Mihos et al. 2017).

The Dragonfly Telephoto Array (Abraham & van Dokkum 2014) is a ground-based optical telescope. It consists of a set of 48 commercial telephoto lenses with simple optics to avoid systematics. In its current configuration, the array is equivalent to a 1 m aperture telescope with a field of view of 6 square degrees and a pixel scale of 2.8 arcsec/pixel (Abraham & van Dokkum 2014). In terms of depth, the galaxy profiles go down to  $\mu_g \sim 28 \text{ mag arcsec}^{-2}$  ( $3\sigma$  in  $12 \times 12 \text{ arcsec}^2$  boxes; Merritt et al. 2016).

Another example of state-of-the-art imaging for low surface brightness science is the *Hubble* Ultra Deep Field (HUDF). The HUDF is one of the deepest regions ever observed with the *Hubble Space Telescope* (Beckwith et al. 2006; Koekemoer et al. 2013; Illingworth et al. 2013). The main goal of this image is to study the nature of high-redshift galaxies ( $z \sim 7-8$ ) rather than analyze the outskirts of galaxies and extended objects. As a consequence, in the data reduction process, the sky background was oversubtracted, and low surface brightness structures were removed. Borlaff et al. (2019) took this issue into account and reduced the images completely from scratch with special care in the flat field and sky background correction (see Section 1.4). They reached the deepest image ever in terms of surface brightness ( $\mu_{F160W} > 32.5 \text{ mag arcsec}^{-2}$ ,  $3\sigma$ ;  $10 \times 10 \text{ arcsec}^2$ ), and enabled the analysis of the low surface brightness structures and objects within the field.

Finally, collaborations between professional and amateur astronomers have been also important for detecting low surface brightness structures (e.g., Martínez-Delgado et al. 2010). The advantage of amateur telescopes is that the field of view of the camera is large, and therefore, it is possible to observe extended objects that otherwise will not fit in the field of view of professional instruments. In this thesis, the galaxy M101 has been observed using an 8 cm amateur telescope with a total exposure time of 100 hours (see Chapter 6 for further details).

The possibility of characterizing low surface brightness objects by obtaining spectra of such sources is extremely hard due to its low signal-to-noise. As a consequence, only a few but very necessary studies have been done in this sense (see e.g., Kadowaki et al. 2017; Ferré-Mateu et al. 2018; Gu et al. 2018; Ruiz-Lara et al. 2018; Aguerri et al. 2020). Taking into account the difficulty in obtaining spectra of such faint objects, an alternative would be to use panchromatic imaging surveys such as the Javalambre-Physics of the Accelerated Universe Astrophysical Survey (J-PAS; Benitez et al. 2014) and the Javalambre Photometric Local Universe Survey (J-PLUS; Cenarro et al. 2019) to obtain low-resolution “spectra” of these objects. These surveys are not as deep as previously mentioned surveys. However, it may be possible to treat the data appropriately (i.e., binning pixels). In this case, the spatial resolution will be decreased, but on the other hand, the extraction of Spectral Energy Distribution (SED) of faint objects in

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

a systematic way for thousands of objects could provide very valuable information about the nature of these sources.

#### 1.1.4 Ongoing and future deep surveys

The previous studies discussed above have shown that the strong frontier imposed by the sky brightness has started to be pushed toward fainter limits. Nowadays, there are several wide and deep surveys dedicated to observing systematically the sky that enable the study of the low surface brightness Universe. Some of them are reviewed in what follows.

The Dark Energy Camera Legacy Survey (DECaLS; [The Dark Energy Survey Collaboration 2005](#); [Dark Energy Survey Collaboration et al. 2016](#); [Dey et al. 2019](#)) uses the Dark Energy Camera (DECam) on the Blanco 4m telescope (located at the Cerro Tololo Inter-American Observatory) to provide the optical imaging for the Dark Energy Spectroscopic Instrument survey (DESI; [Dey et al. 2019](#)) in the  $g$ ,  $r$ , and  $z$  photometric bands. The camera consists of  $62\,2048 \times 4096$  pixel CCDs arranged in a hexagonal  $3.2\text{ deg}^2$  field of view with a pixel scale of  $\sim 0.262\text{ arcsec/pixel}$ . It covers both the North Galactic cap region at  $\text{dec} < 32^\circ$  and the South Galactic cap region at  $\text{dec} < 34^\circ$ , with a total sky coverage of  $\sim 9000\text{ deg}^2$ . The Dark Energy Survey data release 2 (DES DR2; [Abbott et al. 2021](#)) is the second public data release of DES. This release includes data from the DES wide area survey covering  $\sim 5000\text{ deg}^2$  of the southern Galactic cap in five broad photometric bands:  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $Y$ . In terms of depth, the median coadded catalog depth for a  $1.95\text{ arcsec}$  diameter aperture at  $10\sigma$  are 24.7, 24.4, 23.8, 23.1, and 21.7 mag for the  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $Y$  bands, respectively.

A deeper survey but covering a smaller region of the sky is the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP [Aihara et al. 2018a,b](#)). It is a multi-band ( $g$ ,  $r$ ,  $i$ ,  $z$ ,  $Y$ , and 4 narrow-band filters) imaging survey with the Hyper Suprime-Cam (HSC) on the 8.2m Subaru Telescope. The camera has 104 science CCDs with a field of view of  $1.5\text{ deg}$  of diameter, and a pixel size of  $0.168\text{ arcsec/pixel}$ . It has targeted different fields with different exposure times depending on the filters, therefore obtaining different depths. In the second Public Data Release ([Aihara et al. 2019](#)) they fixed several problems such as the sky background subtraction (see Section 1.4.4). As an example, for the “Deep” field in the  $r$ -band with a total time of exposure of 84 min, it reaches a limiting magnitude of 27.4 mag ( $5\sigma$  within an aperture of  $2\text{ arcsec}$  of diameter). The available fields are Wide ( $300\text{ deg}^2$ ), Deep ( $26\text{ deg}^2$ ), and UltraDeep ( $4\text{ deg}^2$ ).

The Vera C. Rubin Legacy Survey of Space and Time (LSST; [LSST Science Collaboration et al. 2009](#); [Ivezić et al. 2019](#)) will be a next step forward regarding the depth, the sky coverage, and the number of sources available to be studied. The commissioning data of the LSST survey is expected by 2022. The telescope has an effective aperture of  $6.7\text{ m}$  and the camera has wide field of view of  $9.6\text{ deg}^2$ . The survey is planned to be a ten year imaging project to cover  $\sim 18\,000\text{ deg}^2$  of the southern sky. Each pointing will be imaged 2000 times with 15 s of exposure time in six different photometric bands ( $u$ ,  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $y$ ). In terms of depth for the coadded images, LSST will reach a limiting magnitude of 27.5 mag in the  $r$ -band for  $5\sigma$  point-like sources.

Regarding the space-based telescopes, one of the most important is the *Euclid* mission (to be launched in 2022). *Euclid* is a medium-class mission in ESA’s Cosmic Vision programme to investigate the expansion of the Universe ([Laureijs et al. 2011](#)). The *Euclid* payload consists of a 1.2m telescope with two instruments: the visual imager (VIS) and the near-infrared spectrometer and photometer (NISF) for the  $Y$ ,  $J$ , and  $H$  photometric bands. They have a common

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

field of view of  $0.53 \text{ deg}^2$ , with a pixel scale of  $0.1 \text{ arcsec/pixel}$  for VIS and  $0.3 \text{ arcsec/pixel}$  for NISP. The mission is expected to conduct two surveys: Wide ( $15\,000 \text{ deg}^2$ , excluding Galactic and ecliptic planes) and Deep ( $40 \text{ deg}^2$ , 3 different fields). The wide survey will reach limiting magnitudes  $m_{\text{VIS}} \sim 24.5 \text{ mag}$  ( $10\sigma$ , extended) and  $m_{Y,J,H} \sim 24.0 \text{ mag}$  ( $5\sigma$ , point-like). On the other hand, the deep survey is expected to reach limiting magnitudes of  $m_{\text{VIS}} \sim 26.5 \text{ mag}$  ( $10\sigma$ , extended) and  $m_{Y,J,H} \sim 26.0 \text{ mag}$  ( $5\sigma$ , point-like).

Finally, the MESSIER surveyor (Valls-Gabaud & MESSIER Collaboration 2017) is a proposed small mission designed to explore the very low surface brightness Universe. It aims to observe the full sky in drift scanning mode. The telescope consists of a  $50 \text{ cm}$  aperture of diameter and  $2 \text{ deg}^2$  of field of view camera in the optical and ultraviolet bands ( $150\text{--}1000 \text{ nm}$ ) with a pixel scale of  $1 \text{ arcsec/pixel}$ . It takes into account the design of the telescope from the basis to optimize the observation of very low surface brightness structures. To do that, the telescope is designed with very simple optics to avoid the internal reflections and a very small PSF (see Hugot et al. 2014; Muslimov et al. 2017, for further details).

All of the previously described surveys (present or future) have several difficulties that need to be solved for low surface brightness studies (see Section 1.3). Indeed, this thesis can be considered as preparatory work with the goal of developing tools and techniques to solve these difficulties for the upcoming wide and very deep surveys.

## 1.2 Low surface brightness: basic concepts

Real astronomical data is always a combination of signal with noise, and the origin of the noise is varied (sky background, instrumental, etc.). As explained in Section 1.1, the low surface brightness regime can be defined as the regime when the level of signal is similar to the amount of noise. In that situation, the noise plays a major role, and consequently, it has to be properly analyzed and understood.

The more time an object is observed, the more signal and also the more noise is obtained. However, signal and noise do not increase at the same speed. While the signal increases proportionally to the time of exposure ( $signal \propto time$ ), the noise increases as the square root of the exposure time ( $noise \propto \sqrt{time}$ ). In total, the signal-to-noise ratio ( $SN$ ) increases as the root square of the observed time:

$$SN = \frac{signal}{noise} \propto \frac{time}{\sqrt{time}} \propto \sqrt{time}. \quad (1.1)$$

The fact that the signal increases faster than the noise is the reason why it is possible to detect sources under the sky brightness. This is because while the noise is random (following a given distribution), the signal always comes from the astronomical source. However, it is necessary to spend a lot of time observing the objects when low surface brightness (low signal-to-noise regime) features are the targets.

### 1.2.1 Sources of noise in astronomical imaging

In general, noise affecting astronomical data comes from three primary sources: photon noise, dark noise, and readout noise. Therefore, a good understanding of the origin and nature of these

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

types of noise is fundamental. For further information about CCDs, see the Handbook of CCD Astronomy by [Howell & Tavackolimehr \(2019\)](#) and references therein. To simulate and study the properties in terms of signal and noise of astronomical images in practice, consider the GNU Astronomy Utilities<sup>2</sup> book (Gnuastro; [Akhlaghi & Ichikawa 2015](#)).

### 1.2.1.1 Photon noise

The vast majority of astronomical sources emit photons randomly due to the intrinsic quantum nature of the emission of light. This produces an inherent statistical variation in the arrival rate of photons incident on the CCD called photon noise. In modern CCD detectors, a photon from an astronomical source arrives at a particular pixel after a long travel through the Universe. In the pixel, that photon is captured and converted to a given amount of electrons (also called photo-electrons), which is finally measured (see Section 1.4.2.1 for further details on how the photons are converted into counts). If the number of photons received in a given interval of time (i.e., its rate) has an average value  $\lambda$ , the probability of detecting  $k$  photons in that fixed time interval is given by the Poisson distribution:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \dots\}. \quad (1.2)$$

The Poisson distribution accounts for discrete operations, such as the counting number of something (in this case, electrons within a pixel). This distribution is very skewed when  $\lambda$  is close to 0. On the other hand, as  $\lambda$  becomes larger, the distribution becomes symmetric. Moreover, when  $\lambda$  is large enough, a good approximation of the Poisson distribution is the normal distribution:

$$P(k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(k-\mu)^2/2\sigma^2}. \quad (1.3)$$

In this case, the normal distribution approximating the Poisson distribution has only one parameter ( $\lambda$ ):  $\mu = \lambda$  and  $\sigma = \sqrt{\lambda}$ . In real observations, the flux (photons) from astronomical objects ( $S_O$ , after convoluted with the PSF) is added to a given sky background flux ( $S_S$ ). That sky background flux is often referred as sky value (see Section 1.3.1). If the background ( $S_S$ ) is constant and high enough (a typical situation in ground-based optical astronomy), then the Poisson distribution can be approximated by a Gaussian distribution, and in this case, the mean and standard deviation of the distribution are  $\mu = S_O + S_S$  and  $\sigma = \sqrt{S_O + S_S}$ .

This is the typical situation when using an integrating instrument (e.g., CCDs). However, in photon counting devices, the number of events is normally small, and therefore Equation 1.2 must be used. It is important to note that photon noise does not just apply to photons from astronomical sources, but to any process involving the count of random events that occur at an average rate. For example, it also applies to the counting of photons from the sky (i.e., sky background), or the production of thermally-generated electrons in a semiconductor (i.e., dark current).

<sup>2</sup>GNU Astronomy Utilities (Gnuastro) is a set of astronomical software that can be run on the command line. It comes with a very extensive and detailed documentation: <https://www.gnu.org/software/gnuastro/>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 1.2.1.2 Dark noise

The dark current is the number of electrons that are spontaneously generated in the detector as a consequence of the thermal excitation within the silicon structure of the CCD pixel. It is independent of photon-induced signal but highly dependent on device temperature. Dark noise comes from the statistical variation of that generation of electrons. Similar to photon noise, dark noise is described by the Poisson distribution. Thus, it is equivalent to the square-root of the number of thermal electrons generated within the image exposure time. As the dark (current) noise depends on the temperature, the cooling of the CCD reduces its contribution dramatically (see Section 1.4.2.2).

### 1.2.1.3 Readout noise

Readout noise comes from two different components. The first is the conversion from the analog signal to a digital number, which is not perfectly repeatable but a statistical process. Second, the electronics themselves introduce spurious electrons into the entire process, yielding unwanted random fluctuations in the output. See Section 1.4.2.1 for further details on how the analog signal is converted to a digital number.

### 1.2.1.4 Total noise and signal-to-noise ratio

Taking into account the discussion above, for astronomical observations using CCDs, there are several sources of noise. Assuming that the number of photo-electrons detected are from the astronomical object ( $S_O$ ), the sky background ( $S_S$ ) and the dark current ( $S_D$ ), their correspondent noise amplitudes are:

- Photon noise in the detected photo-electrons from the source:  $\sqrt{S_O}$ .
- Photon noise in the detected photo-electrons from the sky background:  $\sqrt{S_S}$ .
- Photon noise in the thermally excited electrons (i.e., dark current):  $\sqrt{S_D}$ .
- Readout noise (time independent):  $R$ . This contribution does not go with the square root because it is the standard deviation in the number of electrons measured (it is not a Poissonian counting process).

Assuming that all sources of noise are independent, the total variance ( $\sigma_T^2$ ) is the sum of the variances,  $\sigma_T^2 = \sum_i^n \sigma_i^2$ . In this case, the total noise,  $N$ , is the standard deviation  $\sigma_T$ , given by:

$$N = \sqrt{S_O + S_S + S_D + R^2}. \quad (1.4)$$

Finally, the signal-to-noise ratio ( $SN$ ) is given by:

$$SN = \frac{S_O}{N} = \frac{S_O}{\sqrt{S_O + S_S + S_D + R^2}}. \quad (1.5)$$

From this equation, it can be seen that there are three limiting cases.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- **Object limited.** It occurs when  $S_O$  is much greater than  $S_S$ ,  $S_D$ , and  $R^2$ . In this case,  $SN = S_O/\sqrt{S_O} = \sqrt{S_O}$ . Since the flux from the objects goes linearly with the time of exposure ( $S_O \propto t$ ), the signal-to-noise ratio goes with the square root of the exposure time:  $SN \propto \sqrt{t}$ . Similarly,  $S_O \propto D^2$ , where  $D$  is the aperture of the telescope. Therefore,  $SN \propto D$ .
- **Background limited.** In this situation,  $S_S$  is much greater than  $S_O$ ,  $S_D$ , and  $R^2$ . Therefore,  $SN = S_O/\sqrt{S_S}$ . As both  $S_O$  and  $S_S$  scale in the same way with the exposure time and aperture of the telescope, the  $SN$  scales with respect to these values as in the object limited case ( $SN \propto \sqrt{t}$ ). This is the typical case in low surface brightness observations, in which the brightness of the astronomical sources is of the same order or fainter than the sky background. From this case, the importance of observing faint objects during dark nights can be clearly seen, to avoid the scattered light from the Moon (which is an important source of sky background).
- **Readout noise limited.** This situation happens when  $R^2$  is much greater than  $S_O$ ,  $S_S$ , and  $S_D$ . Since these quantities depend on the exposure time but  $R$  does not, this is the typical situation of short exposure time images. Note that  $R$  is also independent of the telescope aperture. In this case:  $SN = S_O/R$ . Consequently, the  $SN$  increases linearly with the exposure time ( $SN \propto t$ ), and as the square of the telescope aperture ( $SN \propto D^2$ ).

### 1.2.2 Magnitude system and surface brightness

The *luminosity* ( $L$ ) of an object is the amount of energy it is emitting into space per unit of time (energy/time). The luminosity is an intrinsic property of the object that does not depend on the distance. The *flux* ( $f$ ) of an object (or apparent brightness) depends on how far away it is from the observer, and it is the amount of energy per unit of time and area that arrives at the detector. Both quantities are related through the distance between source and observer ( $d$ ) by the following equation:

$$f = \frac{L}{4\pi d^2}, \tag{1.6}$$

where  $L$  is the intrinsic luminosity of the source [J/s],  $d$  is the distance of the source [m], and  $f$  is the flux of the source [J/s/m<sup>2</sup>]. This equation can be interpreted as the photons being spread out to space over a sphere whose area is  $4\pi d^2$ . The discussion above has been done considering the integrated flux in all wavelengths or frequency. However, in reality, it is only possible to use instruments that are able to measure the photons within a range of wavelength or frequencies<sup>3</sup>. This is due to the ultimate nature of the detectors, which are only able to measure photons within a specific wavelength range. In addition to this, specific filters are used to restrict the flux from astronomical sources to a particular range of wavelengths of interest.

Images of astronomical objects span a very large range of brightness. Treating directly with fluxes in physical units would involve a large range of values. To solve this inconvenience, from the very beginning astronomers have chosen to use a logarithmic scale when treating astronomical objects brightness. An important characteristic of the logarithm is that the value

<sup>3</sup>Some instruments measure the flux from a very wide range of wavelengths, they are called bolometers.

|   |                                  |
|---|----------------------------------|
| Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.<br><i>Su autenticidad puede ser contrastada en la siguiente dirección <a href="https://sede.ull.es/validacion/">https://sede.ull.es/validacion/</a></i> |                                  |
| Identificador del documento: 3258585  | Código de verificación: 8GKKDuAS |
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA   | Fecha: 04/03/2021 11:59:31       |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA  | 04/03/2021 12:00:29              |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA   | 14/04/2021 15:46:33              |

has to be greater than zero (positive) and the units have to be dimensional-less. The *magnitude* system is then defined as this logarithmic scale:

$$m - m_r = -2.5 \log_{10} \left( \frac{f}{f_r} \right), \quad (1.7)$$

where  $m$  and  $f$  are the magnitude and flux of the object, while  $m_r$  and  $f_r$  are the magnitude and flux (both known) of a reference object. The factor  $-2.5$  comes from the definition of ancient astronomers, where an increment of magnitude by one unit is an increase in brightness by the fifth root of 100 ( $\sqrt[5]{100} \sim 2.5$ ). Note that it also has a negative sign, which means that a brighter source will have a lower value for its magnitude: a 1 mag star is 100 times brighter than a 6 mag star.

One particular situation is when the reference flux is unity ( $f_r = 1$ ). In this situation, all the instrumental effects like the conversion between photons to electrons, as well as the photometric calibration of the image, are summarized in the *zeropoint* ( $zp$ ) magnitude. Using this zeropoint magnitude, the Equation 1.7 becomes:

$$m = -2.5 \log_{10} (f) + zp. \quad (1.8)$$

The setting of the zeropoint determines the connection between observed counts in the detector and a standard photometric system. For example, the AB system (Oke & Gunn 1983) is constructed in such a way that a source with  $f_\nu = 3631$  Jy (1 Jy = 1 Jansky =  $10^{-26}$  W Hz $^{-1}$  m $^{-2}$  =  $10^{-23}$  erg s $^{-1}$  Hz $^{-1}$  cm $^{-2}$ ) will have a magnitude equal to  $m = 0$  mag, by definition. The *color* of an astronomical object is defined as the ratio of its flux through two filters that cover different ranges of wavelength. Consider for example the flux using the Sloan filters  $g$  and  $r$  (two different wavelengths), then the color  $g-r$  is defined as the difference in magnitudes:

$$g-r = m_g - m_r = -2.5 \log_{10} \left( \frac{f_g}{f_r} \right) + (zp_g - zp_r). \quad (1.9)$$

As the magnitude is logarithmic, note that the ratio of fluxes becomes a difference in magnitudes. Note also that different bands could have different zeropoint values.

When considering extended astronomical objects like galaxies, a vital quantity is the *surface brightness* ( $\mu$ ), which is the brightness over its area. The area that covers the objects over the celestial sphere is commonly expressed in units of arcsec $^2$ . As a consequence, the surface brightness of a given object is expressed in units of mag arcsec $^{-2}$ . If a source has an extension (projected on the celestial sphere) of  $A$  arcsec $^2$ , then its surface brightness becomes:

$$\mu = -2.5 \log_{10} \left( \frac{f}{A} \right) + zp. \quad (1.10)$$

In other words, the surface brightness (in units of mag arcsec $^{-2}$ ) of an astronomical object is related with its magnitude ( $m$  in mag) and its area ( $A$  in arcsec $^2$ ) by the following expression:

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

$$\mu = m + 2.5 \log_{10}(A) + zp. \quad (1.11)$$

The concept of surface brightness is usually a source of confusion, but it is key for the next concepts and chapters.

### 1.2.3 Depth limits

To compare different datasets or astronomical images in terms of depth, it is important to have a reliable way of measuring the depth of a given image. Intuitively, the depth of an image or survey (quantified through a limiting magnitude) gives us an idea of what the brightness is of the faintest sources that can be detected. In general, there are two different brightness limits, depending on the source: point-like or extended sources.

#### 1.2.3.1 Point source limiting magnitude

Point source limiting magnitude is expressed for a certain significance level (number of sigmas) within an aperture size. It is the quantity that the telescopes and instrument exposure times calculator use to make the observing plan. It can be used as a measure of the brightness for the faintest source that can be reliably detected over an image, in which case it is expressed as  $n\sigma$  using aperture sizes of the full width at half maximum (FWHM; or a multiple of the FWHM to have enough pixels in the aperture). This is done, for example, in the Xtreme Deep Field (XDF) of the *Hubble Ultra Deep Field (HUDF)* using a sigma level of  $5\sigma$  in apertures of 0.35 arcsec of diameter (Illingworth et al. 2013).

In practice, the way of obtaining the point source limiting magnitude of an image is to consider the noise or sky background of that image. This can be done by making the photometry within randomly distributed apertures over the image. If the image is dominated by Poisson noise, then the distribution of values from the photometry will be a Gaussian-like distribution with a well-defined standard deviation ( $\sigma$ ), from which the limiting magnitude is computed. To put some surveys into context, SDSS has a  $5\sigma$  limiting magnitude of  $\mu_g \sim 23.3 \text{ mag arcsec}^{-2}$  (Ivezić et al. 2000), while for the Stripe 82 it is  $\mu_g \sim 25.1 \text{ mag arcsec}^{-2}$  (Jiang et al. 2014). In the case of the XDF cited above, the limiting magnitudes are between 29.1 and 30.3  $\text{mag arcsec}^{-2}$ , depending on the filter (Illingworth et al. 2013). Limiting magnitudes of different telescopes are obtained in this thesis for the data of the galaxy NGC 493, see Chapter 5 for further details.

#### 1.2.3.2 Surface brightness limiting magnitude

In the case of the surface brightness limit, the idea is essentially the same but for extended sources. If the image is Gaussian-noise dominated, then the distribution of pixel values of a given region follows a Gaussian-like distribution (with all pixels considered as signal previously detected and masked). By fitting a Gaussian function to the sky distribution, it is possible to obtain the standard deviation of the noise values ( $\sigma$ ). Considering Equation 1.10, the surface brightness limit at  $n\sigma$  level per pixel will be given by the following expression:

$$\mu_{\text{lim}}(n\sigma, p \times p) = -2.5 \log_{10} \left( \frac{n\sigma}{p^2} \right) + zp, \quad (1.12)$$

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

where  $\mu_{\text{lim}}(n\sigma, p \times p)$  is the  $n\sigma$  surface brightness limiting magnitude (in mag arcsec<sup>-2</sup>) of one single pixel with area  $p^2$  (in arcsec<sup>2</sup>).  $zp$  is the zeropoint given by the photometric calibration of the image. Note that it is assumed that pixels are squares with the same dimension of both sides ( $p = p_x = p_y$ ). The previous expression has been derived for one single pixel, and it is only useful when all datasets come from the same instrument (telescope and camera). However, the main reason to define such a quantity is to be able to compare the depths of different datasets, images, or surveys. In addition to this, the typical size of extended astronomical sources is larger than one pixel, and because of that, it is necessary to adapt the expression for the case of arbitrary aperture sizes and pixel scales. Considering an arbitrary area of size  $S^2$  (in arcsec<sup>2</sup>), and a pixel scale of  $p^2$  (in arcsec<sup>2</sup>/pixel), the previous equation will transform to:

$$\mu_{\text{lim}}(n\sigma, S \times S) = -2.5 \log_{10} \left( \frac{n\sigma}{\sqrt{p^2 \times S^2}} \right) + zp. \quad (1.13)$$

This equation is only valid under the assumption that the image is Poisson noise limited, so it is true that the noise is added in root mean square (RMS):

$$\sigma_{\text{total}} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_N^2} = \sqrt{N} \times \sigma, \quad (\text{if } \sigma_1 = \sigma_2 = \dots = \sigma_N). \quad (1.14)$$

For extended sources, a usual metric is  $3\sigma; 10 \times 10$  arcsec<sup>2</sup>. This means that when obtaining the surface brightness limiting magnitude, it will be a measure of the depth for fluctuations of  $3\sigma$  over angular extensions of  $10 \times 10$  arcsec<sup>2</sup> (compared to the sky noise). Areas like that are typical of the width of streams of nearby galaxies (see e.g., [Kniazev et al. 2004](#); [Martínez-Delgado et al. 2008](#); [Trujillo & Fliri 2016](#)).

#### 1.2.4 Expected surface brightness limit for different telescopes and exposure times

Under the assumption that telescope images are limited purely by Poissonian noise, it is possible to theoretically estimate the limiting surface brightness of an image, if the limiting surface brightness of another reference image is shown. The two limiting surface brightness would be related as follows:

$$\mu_{\text{lim}}(\Phi, A, N_\sigma, t) = 2.5 \log_{10} \left( \frac{\Phi}{\Phi_r} \frac{N_\sigma}{N_{\sigma,r}} \sqrt{\frac{A \times t}{A_r \times t_r}} \right) + \mu_{r,\text{lim}}(\Phi_r, A_r, N_{\sigma,r}, t_r), \quad (1.15)$$

where  $\Phi$ ,  $A$ ,  $N_\sigma$  and  $t$  represent the telescope aperture, the area that is used to estimate the limiting surface brightness, the number of  $\sigma$  that represents the confidence level of the surface brightness limiting magnitude, and the exposure time, respectively. The same applies to the rest of the symbols related to the reference image with the subindex  $r$ .

For example, the 2.5 m SDSS telescope has a limiting surface brightness (measured as  $3\sigma$  fluctuation in  $10 \times 10$  arcsec<sup>2</sup>) of 26.5 mag arcsec<sup>-2</sup> using an exposure time of 1 minute ( $r$ -band) ([Trujillo & Fliri 2016](#)). Based on this, the expected limiting surface brightness (as  $1\sigma$  in  $60 \times 60$  arcsec<sup>2</sup>) of an image taken with the 10.4 m GTC is 27.8 mag arcsec<sup>-2</sup> (for the same exposure time). Note that surveys with large sky coverage have different depths depending on the sky position, and the magnitudes usually expressed are averaged values.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 1.3 Challenges in ultra-deep imaging

Obtaining ultra-deep imaging is full of complexities. This is the main reason why there are no large astronomical surveys with depths going beyond  $30 \text{ mag arcsec}^{-2}$  (about 1500 times fainter than the darkest sky on Earth). In this section, the different processes and effects that make it really complex to obtain ultra-deep images are reviewed (see [Knapen & Trujillo 2017](#), and references therein for further details).

#### 1.3.1 Sky brightness (airglow)

The best astronomical observatories on the planet are located in the darkest places of the Earth. But even in those regions where the light pollution is minimal, the sky still has a brightness (in the visible range of the spectrum) of about  $\mu_V \sim 22 \text{ mag arcsec}^{-2}$ . That brightness is mainly due to the airglow, which is the recombination of atoms photoionized by the Sun during the day. The sky brightness during the night is variable, both temporally and spatially. As a consequence, if the exposure time is low, the variability from exposure to exposure will be high. In the same way, the spatial variations (gradients and structures) will be more important and harder to correct as the field of view of the camera is larger. One way to mitigate the variability is to obtain images of long exposure time. By doing this, the variations are averaged so they can be corrected during the sky background subtraction step of the data reduction. In addition to this, obtaining images with long exposure times makes it possible to do an auto flat field correction because the brightness of the sky background is high enough (see Section 1.4.3).

Observatories are often placed in mountains and in regions where there are very special conditions (low humidity and atmospheric stability to avoid as much as possible the turbulence). Building observatories at such special places is not by chance. The next step in that direction is to put the observatories in the space. This is extremely hard in terms of technical difficulties as well as budget and long-term program schedule, but on the other hand, several of the previous issues are minimized. Because in space there is no atmosphere, effects like the airglow and turbulence disappear. However, there is a diffuse component caused by the reflection of sunlight on the dust of the Solar system. It is known as Zodiacal light and it has a brightness of  $\mu_V \sim 23.5 \text{ mag arcsec}^{-2}$  ([Jouvel et al. 2011](#)). The intensity of the Zodiacal light is variable and depends on Solar activity.

#### 1.3.2 Internal reflections

Internal reflections come from the different parts of the dome, telescope, and instrument, and they can be seen as ghosts and structures around very bright objects. These structures are really difficult to model and correct because both the intensity and the place where they appear in the field of view of the camera depend on multiple parameters: the particular direction of the pointing, position angle, the filter used, etc. The reflections appear at different surface brightness levels (depending on the quality of the instrument), but they are common when reaching levels of  $\mu_V \sim 26 \text{ mag arcsec}^{-2}$ . One way to minimize the effect of internal reflections is to construct telescopes with simple optics (see e.g., [Abraham & van Dokkum 2014](#)). Essentially, the idea is to avoid complicated structures and make the light path as simple as possible. In this sense, amateur telescopes such as camera lenses are good options (see Section 6.1). Another way to reduce the effect of the internal reflections is to make observations with significant dithering

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

and rotation of the camera. The goal of this strategy is to avoid the repetition of a similar orientation of the camera on the sky (e.g., Trujillo & Fliri 2016).

### 1.3.3 The point spread function: scattered light

The point spread function (PSF) is defined as the image that the instrument (telescope and camera) obtain of a point-like source object. The effect is to blur sources by spreading the light from one place to another with a given probability (characterized by the PSF). The sum of this effect due to all sources of the image generates a screen of light: the scattered light field. The contribution of this scattered light becomes the dominant source of light at brightnesses fainter than  $\mu_V \sim 29 \text{ mag arcsec}^{-2}$  (Slater et al. 2009). The scattered light field strongly depends on how the sources are distributed on the image and on the shape of the PSF. As a consequence, the scattered light contribution will be more intense if the number of bright objects is large and if the PSF has significant wings. The effect of the PSF is not only related to the point sources (the stars) but also to the light of extended objects. The scattered light background is, consequently, the contribution of light due to extended objects in addition to the point-like sources.

Overall, the scattered light is the sum of the flux of all sources convolved with the PSF that results in an artificial excess of light, spread over the entire image. Therefore, an exquisite characterization of the PSF is mandatory for low surface brightness studies in order to explore the properties of the outskirts of galaxies (halos, stellar streams, etc.). In this thesis, a significant effort has been put into the characterization of the extended PSF (see Chapter 3). These techniques are applied in Chapter 4, where the correction of the PSF of bright stars is a key step in the analysis of the galaxy NGC 1052-DF4 and in the detection of its faint tidal streams. The next step is the correction of the scattered light field due to extended objects. This has been done for the galaxy NGC 4565 with the goal of extracting the age and metallicity of the outskirts structures of this galaxy (see Chapter 6 for more details).

Nowadays, the construction of the PSF of a given telescope is done by standard pipelines. They are usually modeled considering analytical functions such as Gaussian, Moffat, or a combination of them. However, the size of these PSFs tend not to be larger than a few arcseconds, and they are typically used for separating galaxies and stars and other high-level processes. In this sense, it is important to note that the key point for low surface brightness studies is to have a good characterization of the *extended* (over many arcminutes) PSF.

### 1.3.4 Flat fielding

A major advantage of CCD cameras is the linear response of the system: i.e., the output counts registered by the CCD are proportional to the incoming flux (number of photons) that the pixels have received. In addition, all pixels are expected to have the same efficiency. This means that if the CCD is illuminated with a flat source, then all pixels should count the same number of electrons. However, in practice the sensitivity of different pixels across the camera varies, and this is the reason why the flat field correction is needed: to account for pixel-to-pixel variations. The flat field corrects the pixel-to-pixel variations to homogenize the image.

The key point for making an accurate flat field correction is to have a uniform illumination of the CCD. In the vast majority of cases, a dome or twilight flat field is good enough. However, in low surface brightness research the goal is to observe very faint structures and they can

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

be accidentally removed if the flat field correction is not done very carefully. Exposures of the sky during the twilight sound as a good approximation of flat field illumination. But the real situation is that the side of the CCD closer to the location of the Sun receives a larger number of photons than the opposite side of the CCD. In addition to this, during twilight the brightness of the sky varies very fast and depending on the particular pointing, some bright stars could start to appear. As a consequence, the twilight flat field is not perfectly flat since it has gradients coming from the different illumination due to the position of the Sun, or because of other effects. The same happens for dome flat fields, where multiple reflections of light in different parts of the dome and structures can generate gradients in the flat field exposures. These artificial gradients will be printed on the individual exposures when correcting for the flat field, thus, adding parasitic gradients and light that will affect, and even outshine, the low surface brightness structures we want to observe.

To make a more accurate flat field correction, it is possible to use the science images to construct a master flat field image. To do that, it is necessary to design a good observing strategy consisting of observing the target with shifts over the different exposures (i.e., dithering). The main goal of the dithering pattern is to avoid having the same region of the sky observed by the same pixel more than once. Moreover, with the aim of having enough background pixels to obtain the flat field image (i.e., not the galaxy target), it is necessary to make the offsets of the dithering, at least, of the size of the galaxy (Trujillo & Fliri 2016). If the observations are taken over different nights or runs, then it is better to obtain a master flat field image for each observing block. The reason is that observing conditions could have varied, thus introducing some structures that an overall master flat field image could not account for. In any case, the number of images has to be high enough to ensure that the pixel value of the flat field is accurate and represents the sensitivity of the pixel.

It is important to note that usually the different illumination of the CCD or vignetting is corrected using the flat field image. However, the vignetting is not flat field (difference pixel-to-pixel sensitivity), but a different amount of light arriving at different parts of the CCD. The vignetting can be seen as poor illumination in the outer parts of the camera (borders and corners). Using the flat field to correct this effect is only an acceptable approximation if the vignetting is not very strong. In such cases, the lack of light in the outer regions will be corrected by the correspondent multiplicative factor. However, in many cases, it is much better to crop the images to avoid the outer regions with poor illumination.

### 1.3.5 Galactic cirrus

Galactic cirrus is very faint filamentary structure of dust that belong to our own Galaxy. The presence of these dust clouds is very important when observing objects in the Galactic plane, where the amount of dust is higher. However, it is important to note that since they are filamentary structures of the Milky Way at different distances from Earth, they are located everywhere in the sky, even at high Galactic latitudes. The presence of Galactic cirrus has to be taken into account because at low surface brightness it is a source of contamination. For example, if the goal is to study the stellar halo of a galaxy that is located in a region of the sky where there are Galactic cirrus, to distinguish between the galaxy halo and the Galactic cirrus can be really hard (see e.g., Sollima et al. 2010).

Contrary to what happens with the internal reflections or other systematics effects that can

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

be corrected in the reduction process, the Galactic cirrus is actually a source of flux that is always in the same position of the sky. In this sense, it is the same as the scattered light, but it is much more difficult to model because it does not correspond to point-like or instrumental sources. Taking into account the Galactic cirrus contamination and correcting them is one of the most difficult tasks that must be accounted for with the next generation of large and deep surveys. For more information and references about the Galactic cirrus and how to deal with this source in ultradeep imaging, we refer to the analysis done by Román et al. (2020), who make use of the extended PSFs of the Stripe 82 survey to correct for the scattered light and then analyze the properties of the Galactic cirrus.

Due to the intrinsic filamentary nature of the Galactic cirrus, they are not only a source of contamination in the final analysis of the data but also during the reduction of the raw data. Consider for example a particular region of the sky where there are a lot of Galactic cirrus. Then, even with a proper dithering pattern, the structures will be affecting the auto flat field correction if they are not properly masked. Also in the sky background correction, if they are not properly identified and masked, the sky value will be significantly overestimated, thus oversubtracting the sky background from individual images (see e.g., Figure 4 of Fliri & Trujillo (2016)).

#### 1.4 Deep imaging data processing: raw data reduction

In previous sections, the most relevant and difficult challenges of ultradeep imaging have been introduced. In this section, the technical aspects and possible solutions are reviewed as a general outline with valuable lessons learned during the development of the current thesis. To do that, a general reduction of data is followed, showing with examples the problems and solutions. More information and details will be provided for each particular case in the next chapters.

##### 1.4.1 Proper observational strategy

In ultradeep imaging, the design of a good observational strategy is key. The importance of dithering is well known, and the basic idea is to avoid observing the same region of the sky with the same pixel many times. For example, if there is a hot pixel or some reflection that comes from a contaminant source, then it is necessary to move the object to different parts of the CCD to not have these systematics always on the same pixel (that cannot be corrected). Typical dithering offsets are just a tiny (few arcsec) displacement in successive exposures. These small offsets are enough for the vast majority of observations if a small source is the target. Indeed, a small dithering pattern makes the reduction of the data easier in terms of, e.g., astrometry. However, these small displacements of the object around the CCD are not enough in low surface brightness observations because they do not allow an auto flat field correction. As explained above, auto flat field correction is a must because it does not introduce gradients or artificial structures that come from dome or twilight flat field exposures. But, to obtain an accurate auto flat field image, the offsets of the dithering pattern have to be of the same order as the size of the target (Trujillo & Fliri 2016). Another important point is to have a balance between the number of exposures to have an accurate flat field, and the time of exposure to reach enough level of counts.

The rotation of the camera is also another possibility. It reduces the internal reflections generated by very bright objects on the images. In the same way, it mitigates the presence of

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

diffraction spikes around stars because, in different orientations, they are averaged out in the final coadded image. However, the rotation of the camera can introduce some differences in the illumination of the CCD (for example, if the optical axis of the illumination is not aligned with the center of the CCD).

As for non uniform illumination (vignetting), one has to compute the offsets of the dithering pattern so that target does not fall into the poorly illuminated area (borders and outer part of the CCD). To summarize, for deep imaging observations, the most relevant points to take into account are:

- Dome or twilight flats can introduce structures and gradients. A better option is to make an auto flat field correction.
- Auto flat field correction is only possible if enough number of exposures (with relatively high signal-to-noise of the sky background) are taken.
- Design a proper dithering pattern with the goal of correcting the systematics. To do that, do not repeat the same pointing and avoid the poorly illuminated areas of the CCD.
- Rotation of the camera is good for removing systematics, but it can also add other effects (changes in the illumination).
- Check the stability of the flat field: if necessary, obtain a flat field image per observing block (same night and same CCD orientation).

#### 1.4.2 Bias and dark correction

Bias and dark correction are the first steps necessary to apply to each raw image when processing the data. To understand them, it is necessary to know how a CCD works and the basic underlying physics.

##### 1.4.2.1 Bias

A CCD camera measures the number of photo-electrons  $N_e$  produced in each pixel (as a measure of the incoming photons). In one pixel, the total charge due to these electrons is  $Q = eN_e$ , where  $e$  represents the electric charge of a single electron. Then, the charge is dumped into a capacitor of capacitance  $C$ , and the voltage is measured  $V = Q/C$ . This means that the number of electrons is proportional to the voltage measured:  $N_e = CV/e$ . To store the data, it is necessary to convert the voltage (analog quantity) to a number of counts  $N_c$  (discrete or digital quantity). This is done by an analog-to-digital converter (ADC), and it implies a discretization of the original voltage measurement. In an ADC, the number of counts is proportional to the voltage ( $N_c \propto V$ ). Therefore, the number of counts is proportional to the number of electrons through a proportional constant  $G$ :  $N_e = GN_c$ . The proportional constant is the gain, in units of electrons per Analog-Digital Units (eADU). The number of bits of the ADC used to store the count values sets the limits of the values that can be stored. For example, with a 16-bit ADC, it is possible to represent counts from 0 to  $2^{16} - 1$ .

We can now discuss the necessity of the bias level. Consider the ideal situation in which an observation is done in a very dark observatory, and the detector gain is  $G = 1 \text{ e/ADU}$  (so the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

number of photo-electrons is equal to the number of counts in ADUs). The time of exposure is low enough to only have on average 1 photo-electrons generated in each pixel. Moreover, consider that the readout noise of the CCD is 3 ADUs. It means that each pixel is going to count not 1 ADU, but a distribution of values given by a Gaussian<sup>4</sup> distribution. The mean of the distribution will be 1 ADU (generated because of the observation) and the standard deviation will be 3 ADUs (generated as a consequence of the readout noise). In such a situation, it is clear that many pixels should contain negative count values. However, this is not possible at all. Due to the intrinsic nature of counting, it is not possible to count values less than 0 (see Section 1.2.1.1). Physically, what happens is that the data would be corrupted by the digitization process in the ADC producing an unrealistic distribution. To solve this problem, the bias level is generated with a large constant value. This is done by applying an extra voltage to the capacitor before the analog to digital conversion.

The bias correction consists of removing this bias level applied to each raw exposure. The way of doing this correction is to take several images with zero exposure time. Because the time of exposure is zero, the images do not contain photo-electrons or thermally excited electrons: only the constant value and the readout noise are registered in such exposures. To increase the signal-to-noise ratio of the master bias image, it is necessary to obtain a set of images and average them. To avoid outlier values, it is recommended to use a robust operator (such as the median or sigma clipping mean) for stacking the bias exposures. Once the master bias image has been obtained, it can be subtracted from all science images to remove the bias level from each pixel.

#### 1.4.2.2 Dark current

Photo-electrons are generated in CCDs as a consequence of photons exciting electrons from the valence band to the conduction band of the semiconductor. The promotion of such electrons can be caused by an astronomical photon arriving at the pixel, or by a thermal excitation. The thermal excitation of electrons leads to the so-called dark current, and it produces photo-electrons that are not distinguishable from the others.

The dark current is strongly dependent on the temperature, and it can be very important. For example, at room temperature, the dark current of a standard CCD is typically 100 000 e/pixel/s (this is high enough to saturate most CCDs in a very short time). To reduce the contribution of the dark current, the CCDs are cooled down to low temperature, using liquid nitrogen to get temperatures around or below 200 K ( $-70^{\circ}\text{C}$ ). The dark current has to be measured and subtracted from the science images, in a similar way to the bias level. It is done by obtaining dark frames. They are exposures with the same integration time as the science images, but with the shutter of the camera closed (to avoid any illumination into the CCD). As a consequence, dark exposures will take into account the bias level and the dark current generated from the thermal excitation. Therefore, the bias level has to be subtracted from the dark frames before stacking them to obtain the master dark image. The correction of the dark current is done in the same way as the bias level: the master dark frame is subtracted from each raw exposure

<sup>4</sup>In this example, since the number of counts is very small, the distribution would be Poisson-like. However, for simplicity, here it is assumed that the distribution is Gaussian-like to have a well-defined standard deviation (see Section 1.2.1.1). The fundamental idea of the reasoning does not depend on the particular distribution considered.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

(previously bias-corrected).

The dark contribution is usually not taken into account in professional CCDs. The reason is that with the cooling system of professional cameras nowadays, the dark current is negligible. However, it is important to note that in ultradeep imaging, this contribution could be important. Not only for the absolute values but for the complex 2-dimensional structure of the dark current across the CCD that can introduce gradients and structures if they are not properly corrected. For example, in Figure 1.4 it is showed a master bias and master dark images of an amateur telescope (see Section 6.1). It can be clearly seen that they are not the same, and consequently, the corrected images could have residuals of such structures.

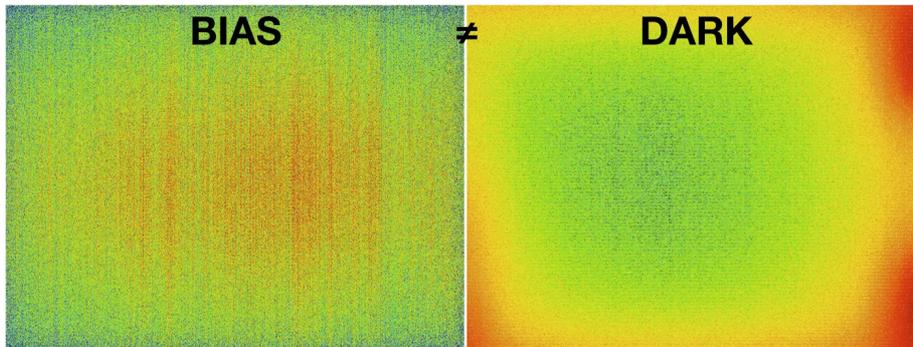


Figure 1.4: Differences between a master bias and a master dark images for the amateur astronomer camera ASI1600MM (see Section 6.1 for further details). The dark image has higher values since the integrating time is always higher than bias exposures. Note that the 2-dimensional structures of both images are completely different. While the bias image shows a column pattern (due to the readout), the dark frame shows a different structure across the image.

### 1.4.3 Flat field correction

Once the raw exposures have been corrected for bias level and dark current, it is necessary to make the flat field correction. As explained in Section 1.3.4, flat field correction is needed to account for pixel-to-pixel variations. To avoid structures from twilight or dome flat field images, the best option is to make an auto flat field correction. To do that, the science images are used to construct the master flat field image. The flat field correction has to be done independently for each filter. The process is explained in what follows. First of all, it is necessary to detect and mask all the pixels that are not considered sky background. The detection of the signal can be done with several astronomical software packages. In this thesis, the Gnuastro programs are extensively used. In particular, `NoiseChisel` (Akhlaghi & Ichikawa 2015) is used to characterize the different pixels. Then, once the image has been separated between pixels with astronomical sources and pixels without them, the astronomical source mask is used to replace the former

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

pixels with Not a Number (NaN) values.

The process of detecting and masking the signal is done over the whole set of images that are going to be used to construct the master flat field image. As an example, consider Figure 1.5. In this figure, we show several dithered and already masked images of the galaxy M101 obtained with an amateur telescope. Note how the galaxy M101 as well as all other sources have been detected and masked (using NaN values, shown in white).

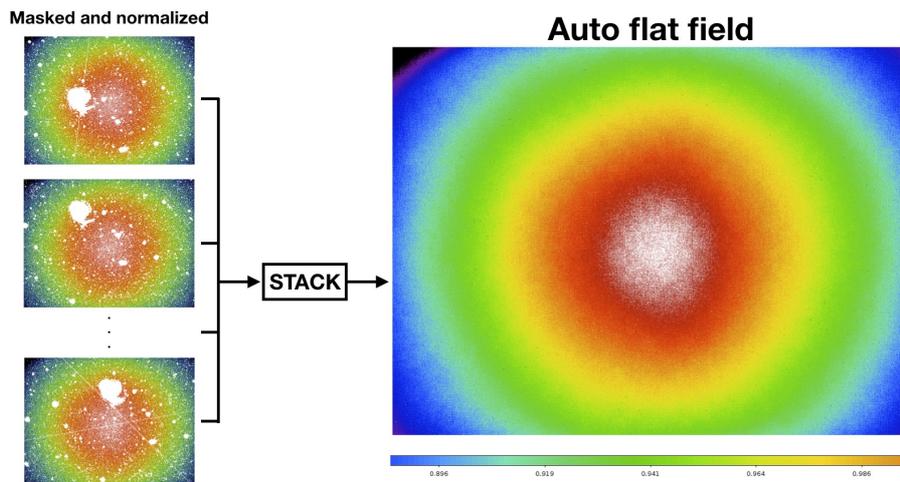


Figure 1.5: The process of constructing an auto flat field image. On the left, there are several images of the field surrounding the galaxy M101. The pixels with signal have been detected and masked (white regions are NaNs). After the signal has been masked, the images are normalized to have all pixel values on a common scale. Finally, all masked and normalized images are stacked to obtain the auto flat field image.

Since the sky background values can vary from exposure to exposure, each masked science image has to be normalized. Therefore, the normalization of each image is necessary to put all images into the same pixel scale values. Once each image has been masked and normalized, they are stacked together to obtain the auto flat field image. The stacking has to be computed using a robust operator in order to reject possible non-masked outlier pixel values, for example,  $3\sigma$  clipping mean or similar. It is important to be aware of the data type (unsigned integer, integer, float, double precision) and the stacking operator since they can discretize the final pixel value to some particular number. Finally, the flat field image obtained previously is used to correct the individual exposures by dividing by the flat field image.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 1.4.3.1 Difficulties to create an auto flat field image

In the previous sections, the basic steps to construct an auto flat field image have been reviewed. In some cases, however, there are extra difficulties. Consider for example the case of having a very strong vignetting or fringing<sup>5</sup> patterns on the images. These effects would make reliable detection of the signal (astronomical sources) by common software almost impossible. The reason is that the strong variations across different parts of the camera would be mixed with the objects and background, making the separation between the signal from astronomical sources and artifacts of the camera very complex.

Obtaining masks in such situations is hard but not impossible, and it is necessary to consider a couple of extra steps in the flat field construction process. As the main problem is the strong background gradients that impede a proper mask determination, the idea is to correct (at least as a first approximation) such background gradients. This can be done with a very crude flat field image constructed from images that are not masked at all. Then, once the strong background structures have been removed, it is possible to obtain accurate masks of astronomical sources and construct the flat field image as explained above.

### 1.4.4 Sky background correction

Once the bias, dark, and auto flat field corrections have been applied, the remaining pixel values correspond to the flux that has come from astronomical sources and the atmosphere (i.e., flux not generated by the telescope/instrument). The sky background subtraction then consists of modeling and subtracting the sky background flux from the images. In principle, it is a well-defined problem, but it is also a very difficult task in practice. The reason is that the sky background is a combination of many different physical phenomena: actual sky background brightness of the atmosphere during the night (i.e., airglow), Zodiacal light, scattered light of all objects on the image, scattered flux from sources outside the field of view, flux from background sources that are not resolved or deblended, etc. (see Section 1.3.1 for an overview of the sky background challenge).

To understand the difficulty of the sky background correction, consider Figure 1.6, which shows the same region of the sky, the outskirts of the galaxy NGC 493, observed by different telescopes using different depths. It can be clearly seen that new objects start to appear as the images are deeper. To put the problem into context, focus on the two regions marked with a green circle and a red rectangle. It is clear how in the SDSS image (top left panel) there are almost no detectable sources in those regions, while in the others there are certain objects. Not only the small or point-like sources, but note also how the scattered light of diffuse structures such as the galaxy halo become larger as the image is deeper. This is just an illustration of the sky background challenge (for more information about these images see Chapter 5). In this case, the presence of Galactic cirrus has not been taken into account, if this were the case, the sky background estimation would be even more difficult.

The modeling of the sky background has traditionally been done by considering those pixels in which there is no signal and obtaining from them a smooth surface. In practice, boxes of a given size are considered for obtaining the averaged pixel value of the sky background over

<sup>5</sup>Fringing are interference patterns due to multiple reflections of light in the different layers of the CCD. As it happens with the vignetting, it is not pixel-to-pixel variation (see Section 1.3.4).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

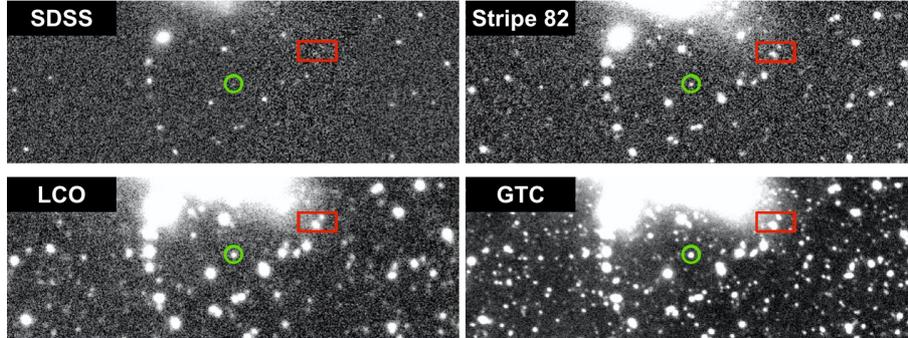


Figure 1.6: Outskirts of the galaxy NGC 493 as seen by different telescopes. The center of the circular region corresponds to R.A. (J2000) = 1:22:04.45 and dec (J2000) = +0:55:13.61 (the radius of the aperture is  $R = 3$  arcsec). The images correspond to the  $r$ -band of the different surveys labeled in the top-left corner (Sloan Digital Sky Survey, Stripe 82, Las Cumbres, Gran Telescopio Canarias).

different meshes across the image. Then, such a boxed image is convoluted with a kernel to make the sky model smoother. In this process, there are two main points: the detection of objects, and the size of the boxes. The detection of objects is extremely important, as seen before. If the objects are not properly detected and masked, their flux will be taken into account in the sky background modeling. This is called oversubtraction: in other words, flux from astronomical sources is considered as sky background brightness. This is typically what happens in the regions close to objects (e.g., outskirts of extended objects such as galaxies that are very difficult to detect and mask). The oversubtraction prevents the accurate analysis of the data because it has actually removed the physical brightness of the objects. The other important point is the selection of the box size to estimate the sky background. If the the box is very large, then, possible gradients on the image will not be corrected. On the other hand, if the box is small, then it will account for the sky background on small scales, but also for low surface brightness structures that were not masked (for example galaxy stellar halos, stellar streams, Galactic cirrus, etc.). These structures are extremely difficult to detect and mask, and consequently, they will be removed if the box is smaller than the typical scale of these structures.

An additional consideration is the relative size of the observed object and the field of view of the camera (or the size of the individual CCD if the camera is composed of multiple CCDs). If the observed object is smaller than the CCD, then it is possible to use the regions of the sky that are far away from the object to model the sky background. On the contrary, if the observed object has approximately the same or larger size than the CCD, then it is more complicate (if not impossible) to estimate the sky background. The reason is that a significant fraction of all the flux collected by the CCD has surely come from the source. This is, indeed, the situation described in Section 4.2.2.1 for the HiPERCAM data of NGC 1052-DF4. In such situations, the sky will be oversubtracted and the only thing to be done is to try to mitigate the oversubtraction

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

somehow (for example using bigger boxes). In this sense, a good lesson learned after dealing with observations from different telescopes is: try to not use multiple CCDs cameras for observing very extended objects. Note that this advice will be also of help in the auto flat field construction and correction.

The oversubtraction can be clearly seen as “holes” of negative values around extended sources in large astronomical surveys. In Figure 1.7 we show the sky oversubtraction for the Public Data Release 1 (PDR1) of the Hyper Suprime-Cam (HSC). This issue was fixed later in the PDR2 of the same survey by considering a global sky background correction (not per CCD), and considering larger boxes for estimating the sky background. See Aihara et al. (2019) for further details.

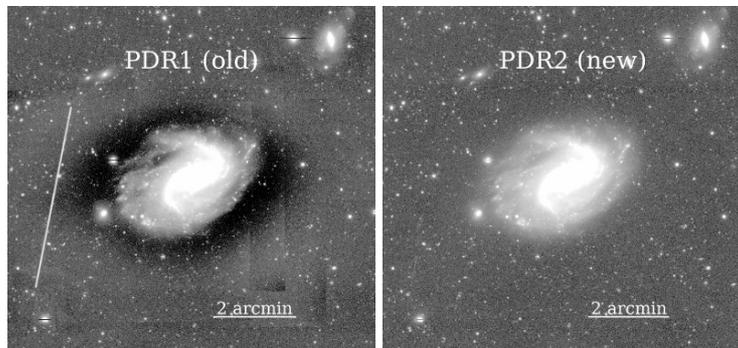


Figure 1.7: Images of a nearby galaxy in the *i*-band of the Hyper Suprime-Cam Survey. Left: coadded image from PDR1 in which the sky background has been over subtracted as indicated by the dark halo around the galaxy. Right: same image but from PRD2, in which the sky background correction has been improved by considering a “global sky” and bigger boxes in the sky background estimation. Credit: Aihara et al. (2019).

As explained above, the sky background correction is a very difficult task. However, it is possible to address it by being very careful and trying to mitigate as much as possible the over-subtraction. By doing this, the photometry of all sources is improved and low surface brightness structures are less likely to be removed. One way of making the sky background correction is to use programs like *NoiseChisel* (Akhlaghi & Ichikawa 2015) or *MTOjects* (Teeninga et al. 2016) for detecting the signal on the images. These programs are optimized for detecting the signal in the low signal-to-noise regime. Once the signal has been detected and masked, consider a 2-dimensional polynomial function with a low degree to model the remaining pixels. The degree of the polynomial surface should be as small as possible ( $\lesssim 3$ ) to not model and remove local and faint structures (hidden as sky background).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 1.4.5 Astrometric solution

In the reduction of the data, the next natural step is to obtain the astrometric solution of each image. The astrometric solution consists of knowing which part of the sky corresponds to each image. Formally, the astrometric solution is obtained once the celestial coordinates of a reference pixel and the pixel scale are known. In practice, the astrometric solution of one image is obtained by comparing a catalog of sources ( $X$  and  $Y$  position, image coordinates) with a reference catalog, from which the position of the objects are known and expressed in celestial coordinates (R.A. and dec). This task is relatively easy when some astrometric information is already present in the header of the raw images. This is the case when the telescope adds information of the pointing as metadata. In such cases, that crude astrometric information can be used by astronomical software like SCAMP (Bertin 2006) to obtain the final and refined astrometric solution.

However, obtaining the astrometric solution from scratch is really hard in some situations, and the main problem tends to be always the same: not having a sufficient number of objects (in the image or the reference catalog). The lack of detectable objects can be due to multiple reasons: small field of view, low signal-to-noise ratio, use of a filter in which the objects are not bright enough, among others. In addition to this, if the observation has been done using a big dithering pattern, then the situation would be even worse because relative astrometry is more difficult.

To obtain the absolute astrometry from scratch, one of the best programs so far is **Astrometry** (Lang et al. 2010). It has the possibility of uploading an image to its website (almost in any format), and it will obtain the astrometric solution. For more professional and systematic use, the command line version is needed. Basically, it is necessary to use two programs:

- **build-astrometry-index**: used for splitting the reference catalog into several indexes. These indexes will be used by the other program to obtain the astrometric solution.
- **solve-field**: computes the astrometric solution.

Obtaining an accurate astrometric solution for the different images is necessary in all cases, but in low surface brightness studies, it is absolutely crucial. Otherwise, small differences in the alignment of the images will cause a loss of signal-to-noise ratio in the final stacked image.

### 1.4.6 Photometric calibration

The photometry of a source using a set of filters gives crude spectral information about the object (flux as a function of wavelength). A well-defined set of filters is known as a photometric system. A photometric system is used to compare different observations from different telescopes and instruments. Consider the following situation. Two astronomers observe the same source with the same filter, the same instrument at the same time, but with telescopes of different sizes. The astronomer with the large telescope will measure more counts than the astronomer with the small telescope. Therefore, directly comparing the results in counts is meaningless. The way they have to compare their observation is to calibrate the data into a common photometric system. The necessity is even more clear if exposure time, telescope size, camera, filter, and other considerations are different.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

The filter bandpass limits the number of photons that are recorded by the CCD. A narrow bandpass will make it difficult to detect sufficient photons from a source (especially if it is faint). On the contrary, a photometric system with wide bandpass filters will allow more photons to be collected by the CCD. As a consequence, in ultradeep imaging or low surface brightness studies, it is more common to use wide bandpass filters. One of the most widely used photometric system is the Sloan  $(u, g, r, i, z)$  set of filters covering the optical spectrum.

In practice, photometric calibration consists of obtaining the conversion between instrumental magnitudes and calibrated magnitudes. To do that, several considerations have to be taken into account, such as extinction because of the airmass, aperture or PSF photometry of objects, color terms if the filter response is not the same. Once the photometric calibration has been done, the zeropoint value is used for later and high-level analysis of the data.

## 1.5 Reproducibility in Science

Reproducibility is a major principle of the scientific method. Throughout the development of this thesis, a lot of data from different telescopes and instruments have been treated and analyzed using different software and techniques. As the volume and variety of datasets increases, the complexity of the data analysis also increases (the number of new features, considerations, improvements, steps, etc.). Consequently, reproducing the same output even using the same computer and techniques becomes very hard. This is something that everybody in charge of doing research has experienced. Dealing with this situation in which everything was working fine, but suddenly it does not and apparently nothing has changed, is extremely frustrating and requires a lot of time. Probably, the origin of the problem is just an update of specific software that you did not even notice.

History shows that many scientific claims or discoveries were in fact systematic effects due to human error or software bugs. Having a good framework in which all software, as well as the history of the research, is recorded properly is fundamental. Many workers have tried to account for this fact by constructing frameworks and methodologies to do reproducible research as reviewed in the Appendix of Akhlaghi, Infante-Sainz et al. (2020). However, none of these projects go deep enough, to the most basic level, to solve the problem of reproducibility. All of them rely on high-level software that has many dependencies or are not sufficiently transparent. With the aim of accounting for all of these difficulties, in Chapter 2 we present Maneage: a framework to make research reproducible. This project started some time ago and was developed during this thesis. In fact, almost all pipelines, reduction of the data, high-level analysis of the data, and all computations presented in the next chapters have been obtained using Maneage. The reproducibility of results is a fundamental part of this thesis, and because of that, Chapter 2 is entirely dedicated to this issue. It corresponds to the extended version of Akhlaghi, Infante-Sainz et al. (2020), where the importance of the reproducibility and how it is accounted practically is presented. In this thesis, Maneage is mainly used to make astrophysical research. However, this framework has been designed to be used in any field of science.

## 1.6 Thesis outline

The main goal of this thesis is to study the low surface brightness Universe, and in particular, to develop methods and techniques to solve the most important difficulties in this field. All of these

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

tasks are done in a reproducible way. To do that, data from many telescopes and instruments have been reduced in a very careful way, to avoid destroying the low surface brightness features. The general structure of the thesis is as follows.

Chapter 1 is a short introduction of the low surface brightness astronomy, as well as its challenges and the problem of the reproducibility. In this chapter, we also present general and useful lessons acquired during the development of this thesis. In Chapter 2, Maneage is presented as a framework to work in a closed and fully controlled environment. This framework is not only designed for astronomical research but in general to do any kind of research or data analysis. This framework is the general structure in which the next works are developed. Chapter 3 is dedicated to the effect of the PSF. Data from the SDSS is used to build the extended PSFs of that survey. In addition to this, it is also explained how they can be used to correct for the scattered light. This work has been done under Maneage. In Chapter 4, the strange galaxy “missing dark matter” NGC 1052-DF4 is analyzed. This chapter shows in detail how ultradeep imaging and low surface brightness research can be used to study the ultimate nature of astronomical objects. In this case, deep imaging data from several telescopes reveals that this galaxy is interacting with another nearby galaxy. Therefore, the apparent lacking of dark matter can be naturally explained without relying on exotic physics. In Chapter 5, data from different surveys and telescopes of the galaxy NGC 493 are reduced and analyzed. The main goal of this work is to study how the properties and physical parameters of astronomical sources depend on the depth of the images. The preliminary results from the analysis of the data reveal that the low surface brightness features are consistent and replicable throughout the different telescopes considered. In Chapter 6 we describe the ongoing and future work that could not be finished during the duration of this thesis because of time limits. Two main projects are presented: ultradeep imaging with an amateur telescope, and the comparison between the star count and integrated photometry techniques. Finally, the main conclusions of this thesis are presented in Chapter 7.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 2

## Reproducibility in Science: Maneage

This chapter explains the problem of reproducibility in science, and our proposal on how to treat data properly. Here, “Maneage” (**Man**aging data **Lineage**) is presented.<sup>1</sup> It is a framework designed for managing data, having full control of the software, and reproducing scientific results. This chapter corresponds to the original and long version of the paper “*Towards Long-term and Archivable Reproducibility*” (submitted to CiSE, [arXiv:2006.03018](https://arxiv.org/abs/2006.03018), Akhlaghi, Infante-Sainz et al. 2020), in which I am the second author. The importance of reproducibility in science is vital. Consequently, this chapter treats this issue from a general point of view and not only centered in astronomy or the low surface brightness Universe. The reason for dedicating an entire chapter to Maneage is that it has been highly matured along with the development of this thesis. In addition to this, it is also heavily used for the analysis and treatment of the data of all the next chapters.

### 2.1 Introduction

The increasing volume and complexity of data analysis have been highly productive, giving rise to a new branch of big data in many fields of the sciences and industry. However, given its inherent complexity, the mere results are barely useful alone. Questions such as these commonly follow any such result: what inputs were used? what operations were done on those inputs? How were the configurations or training data chosen? how did the quantitative results get visualized into the final demonstration plots, figures, or narrative/qualitative interpretation? may there be a bias in the visualization? See Figure 2.1 for a more detailed visual representation of such questions for various stages of the workflow.

In data science and database management, this type of metadata is commonly known as *data provenance* or *data lineage*. Their definitions are elaborated with other basic concepts in Section 2.2. Data lineage is being increasingly demanded for integrity checking from both the scientific and industrial/legal domains. Notable examples in each domain are respectively the “Reproducibility crisis” in the sciences that was claimed by the Nature journal (Baker 2016), and the General Data Protection Regulation (GDPR) by the European Parliament and the California Consumer Privacy Act (CCPA), implemented in 2018 and 2020 respectively. The former argues

<sup>1</sup><https://maneage.org/>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

that reproducibility (as a test on sufficiently conveying the data lineage) is necessary for other scientists to study, check and build-upon each other's work. The latter requires the data industry to give individual users control over their data, effectively requiring thorough management and knowing the lineage of the data. Besides regulation and integrity checks, having robust data governance (management of data lineage) in a project can be very productive: it enables easy debugging, experimentation on alternative methods, and optimization of the workflow.

In the sciences, the results from the analysis are published as scientific papers which have also been the primary conveyor of the result of the lineage, usually in narrative form within the "Methodology" section of the paper. From our own experiences, this section is usually most discussed during peer review and conference presentations, showing its importance. After all, a result is defined as "scientific" based on its *method* (the "scientific method"), or lineage in data-science terminology. In the industry, however, data governance is usually kept as a trade secret and it is not publicly published or scrutinized. Therefore, while the proposed approach discussed in this chapter (Maneage) is also useful in industrial contexts, the main practical focus would be on the scientific front, which has traditionally been more open to publishing the methods and anonymous peer scrutiny.

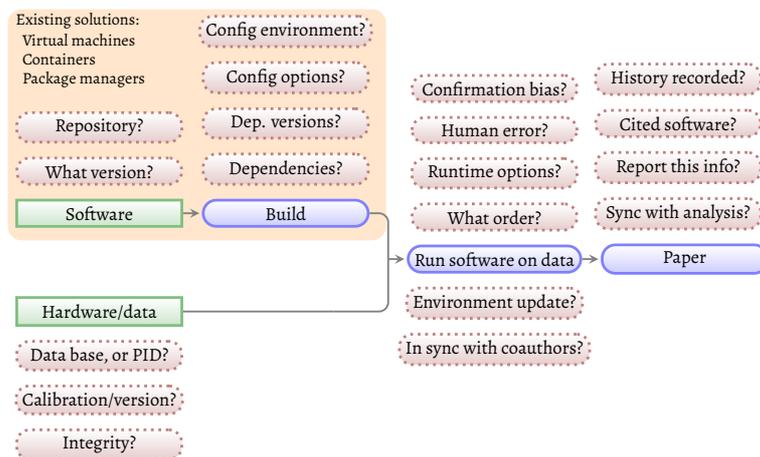


Figure 2.1: Graph of a generic workflow (connected through arrows), highlighting the issues on each step. The green boxes with sharp edges are inputs, and the blue boxes with rounded corners are intermediate or final outputs. The red boxes with dashed edges highlight the main questions at each respective stage. The box covering software download and build phases shows common tools that software developers use for this phase, but a scientific project is much more involved.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

The traditional format of a scientific paper has been very successful in conveying the method with the result in the last centuries. However, the complexity mentioned above has made it impossible to describe all the technical steps of a project to a sufficient level of detail. Citing this difficulty, many authors try to describe the very high-level aspects of their analysis, while even the most basic calculations (like the mean of a distribution) can depend on the software implementation. In this sense, due to the complexity of modern scientific analysis, a small deviation in the final result can be due to many different steps.

Nature is already a black box that we are trying to unlock and understand. Not being able to experiment on the methods of other researchers is an artificial and self-imposed black box wrapped over the original. Publishing the precise codes of the analysis is the only guarantee. For example, Miller (2006) found a mistaken column flipping, leading to the retraction of 5 papers in major journals, including Science. Baggerly & Coombes (2009) highlighted the inadequate narrative description of the analysis and showed the prevalence of simple errors in published results, ultimately calling their work “forensic bioinformatics”. Herndon et al. (2014) and Horvath (2015, a self-correction) also reported similar situations, and Ziemann et al. (2016) concluded that one-fifth of papers with supplementary Microsoft Excel gene lists contain erroneous gene name conversions. Smart (2018) describes how a 7-year old conflict in theoretical condensed matter physics was only identified after the relative codes were shared. Such integrity checks tests are a critical component of the scientific method but are only possible with access to the data and codes.

The completeness of a paper’s published metadata (or “Methods” section) can be measured by the question: given the same input datasets, can another researcher reproduce the same result automatically, without needing to contact the authors? Several studies have attempted to answer this question with different levels of detail. For example, Allen et al. (2018) found that roughly half of the papers in astrophysics do not even mention the names of any analysis software they have used, while Menke et al. (2020) found that the fraction of papers explicitly mentioning their tools and software has greatly improved in medical journals over the last two decades.

Ioannidis et al. (2009) attempted to reproduce 18 published results by two independent groups but only fully succeeded in 2 of them and partially in 6. Chang & Li (2015) attempted to reproduce 67 papers in well-regarded economic journals with data and code. Only 22 could be reproduced without contacting authors, and more than half could not be replicated at all. Stodden et al. (2018) attempted to replicate the results of 204 scientific papers published in the journal Science *after* that journal adopted a policy of publishing the data and code associated with the papers. Even though the authors were contacted, the success rate was 26%. Generally, this problem is unambiguously felt in the community: Baker (2016) surveyed 1574 researchers and found that only 3% did not see a “reproducibility crisis”.

This is not a new problem in the sciences. In 2011, Elsevier conducted an “Executable Paper Grand Challenge” (Gabriel & Capone 2011). The proposed solutions were published in a special edition, but most of them have not been continued since then. Before that, Ioannidis (2005) proved that “most claimed research findings are false”. In the 1990s, Schwab et al. (2000); Buckheit & Donoho (1995); Claerbout & Karrenbach (1992) describe this same problem very eloquently and also provided some solutions that they used. While the situation has improved since the early 1990s, these papers still resonate strongly with the frustrations of scientists nowadays. Even earlier, through his famous quartet, Anscombe (1973) qualitatively

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

showed how distancing researchers from the intricacies of algorithms and methods can lead to a misinterpretation of the results. One of the earliest such efforts was [Roberts \(1969\)](#), which discussed conventions in FORTRAN programming and documentation to help in publishing research codes.

From a practical point of view, for those who publish the data lineage, a major problem is the fast-evolving and diverse software technologies and methodologies that are used by different teams in different epochs. [Zhao et al. \(2012\)](#) describe it as “workflow decay” and recommend preserving these auxiliary resources. But in the case of software, it is not as straightforward as data: if preserved in binary form, the software can only be run on certain hardware and, if kept as source-code, their build dependencies and build configuration must also be preserved. [Gronenschild et al. \(2012\)](#) specifically study the effect of software version and environment and encourage researchers to not update their software environment. However, this is not a practical solution because software updates are necessary, at least to fix bugs in the same research software. Generally, the software is not a secular component of projects, where one software can easily be swapped with another. Projects are built around specific software technologies, and research in software methods and implementations is itself a research topic in many domains ([Di Cosmo & Pellegrini 2019](#)).

This chapter introduces Maneage as a solution to address these important issues. Section 2.2 defines the necessary concepts and terminology, leading to a discussion of the necessary guiding principles in Section 2.3. Section 2.4 introduces the implementation of Maneage, going into lower-level details in some cases. Finally, in Section 2.5, the prospects of using systems like this template are discussed.

## 2.2 Definition of important terms

The concepts and terminologies of reproducibility and project workflow management and design are commonly used in different ways by different research communities. As a consequence, before starting with the technical details, it is important to clarify the specific terms used throughout this chapter.

### 2.2.1 Definition: input

Any computer file that may be usable in more than one project. The inputs of a project include data, software source code, etc. (see [Hinsen \(2016\)](#) on the fundamental similarity of data and source code). Inputs may be encoded in plain text (for example tables of comma-separated values, CSV, or processing scripts), custom binary formats (for example JPEG images), or domain-specific data formats (e.g., FITS in astronomy, see [Pence et al. 2010](#)).

Inputs may have initially been created (e.g., software source code) or collected (e.g., data) for one specific project. However, they can also be used in other or later projects. Following the principle of modularity, it is optimal to treat the inputs of any project as independent entities, not mixing them with how they are managed (how software is run on the data) within the project (see Section 2.2.3).

Inputs are nevertheless necessary for building and running any project. Some inputs may already be archived or published independently prior to the publication of the project. In this case, they can easily be downloaded and used by independent projects. Otherwise, they can be

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

published with the project, but as independent files, for example see [zenodo.3408481](#) (Akhlaghi 2019) or [zenodo.3524937](#) (Infante-Sainz et al. 2020).

### 2.2.2 Definition: output

Any computer file that is obtained as a result of the project. The outputs can be datasets (terabyte-sized, small tables or images, a single number, a true/false (boolean) outcome, etc.), automatically generated software source code or any other kind of file. The raw output files are commonly supplemented with a paper or report that summarizes them in a human-friendly readable format. The report commonly includes highlights of the input and output datasets (or intermediate products) as graphs, figures, tables, or simple numbers blended into the text.

The outputs can either be published independently on data servers that assign specific persistent identifiers (PIDs) to be cited in the final published paper. Alternatively, the datasets can be published with the project source, for example, [zenodo.1164774](#) (Bacon et al. 2017, Sections 7.3 and 3.4).

### 2.2.3 Definition: project

The most high-level series of operations done on the inputs to produce the outputs. Because the report of the project is also defined as an output (see above), besides the high-level analysis, the source of the project also includes the recipes to produce plots, figures, or tables.

With this definition, this concept of a “project” is similar to “workflow”. However, it is important to emphasize that the source code of a project and the inputs are different entities. For example, the project may be written in the same programming language as one analysis step. Generally, the project source is defined as the most high-level source file that is unique to that individual project (its language is irrelevant). The project is thus only in charge of managing the inputs and outputs of each analysis step (take the outputs of one step, and feed them as inputs to the next), not to analyze by itself. A well-designed project will follow the modularity principle: analysis scripts should be well-defined as an independently managed software source (see Section 2.3.2). For example modules in Python, packages in R, or libraries and programs in C/C++ that can be imported in higher-level project sources.

### 2.2.4 Definition: data provenance

The set of metadata (in any ontology, standard, or structure) that connects it to the components (other datasets or scripts) that produced it. Data provenance is a very generic term that points to slightly different technical concepts in different fields like databases, storage systems, and scientific workflows. For example, within a database, an SQL query from a relational database connects a subset of the database entries to the output (*why*-provenance), their more detailed dependency (*how*-provenance), and the precise location of the input sources (*where*-provenance), for more see [Cheney et al. \(2009\)](#). In scientific workflows, provenance goes beyond a single database and its datasets to link the analysis with the text of the paper (see e.g., [Bavoil et al. 2005](#); [Moreau et al. 2008](#); [Malik et al. 2013](#)). Data provenance thus provides a high-level view of the genealogy of the data.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 2.2.5 Definition: data lineage

The low-level and fine-grained record of the source of the data and operations that occur to it, down to the exact command that produced each intermediate step. Data lineage is commonly used interchangeably with data provenance (for example Cheney et al. 2009, among others). This *recording* does not necessarily have to be in a formal metadata model. But data lineage must be complete (see completeness principle in Section 2.3.1), and allow extraction of data provenance metadata, and thus higher-level operations like the visualization of the workflow.

### 2.2.6 Definition: reproducibility and replicability

These terms have been used in the literature with various meanings, sometimes in a contradictory way. It is, therefore, necessary to clarify the precise usage of both terms. But before that, it is important to bear in mind that this chapter only considers a computational analysis. In other words, analysis after the data has been collected and stored as a file on a filesystem. Therefore, many of the definitions reviewed in Plesser (2018), which are about data collection, are out of context here. As a consequence, here it is considered the same definition as Leek & Jager (2017); Fineberg et al. (2019), among others:

- **Reproducibility:** (same inputs → consistent result). Formally: “obtaining consistent [not necessarily identical] results using the same input data; computational steps, methods, and code; and conditions of analysis” (Fineberg et al. 2019). This is thus synonymous with “computational reproducibility”.

Fineberg et al. (2019) allow non-bitwise or non-identical numeric outputs within their definition of reproducibility, but they also acknowledge that this flexibility can lead to complexities: what is an acceptable non-identical reproduction? Exactly reproducible outputs can be precisely and automatically verified without statistical interpretations, even in a very complex analysis involving many CPU architectures, and random operations (see Section 2.3.5).

- **Replicability:** (different inputs → consistent result). Formally: “obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data” (Fineberg et al. 2019).

Generally, since replicability involves new data collection, it can be expensive. Note that replicability does not have to be limited to different input data. Using the same data, but with different implementations of methods, is also a replication attempt (also known as “in silico” experiments, see Stevens et al. 2003).

Some authors have oppositely defined these terms. Examples include Hinsen (2015) and the policy guidelines of the Association of Computing Machinery (ACM, dated April 2018).<sup>2</sup> ACM has itself adopted the 2008 definitions of Vocabulaire International de Métrologie (VIM).

Besides the two terms above, “repeatability” is also sometimes used in regards to the concept discussed here and must be clarified. For example, Ioannidis et al. (2009) use “repeatability” to encompass both the terms above. However, the ACM/VIM definition for repeatability is

<sup>2</sup><https://www.acm.org/publications/policies/artifact-review-badging>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

“a researcher can reliably repeat her own computation”. Hence, in the ACM terminology, the only difference between replicability and repeatability is the “team” that is conducting the computation. In the context of this chapter, inputs are precisely defined (Section 2.2.1): files with specific/registered checksums (see Section 2.3.5). Therefore the inputs are team-agnostic, allowing to safely ignore “repeatability” as defined by ACM/VIM.

### 2.3 Principles of the proposed solution

The core principle behind this solution is simple: science is defined by its method, not by its result. Statements that convey a “result” abound in all aspects of human life (e.g., in fiction, religion, and science). What distinguishes one from the other is the “method” that the result was derived. Science is the only case that attempts to be as objective as possible through the “scientific method”.

The most important point is that starting a new project with a robust data management strategy is much more effective than imposing it in the end. The researchers play a crucial role in making their research more Findable, Accessible, Interoperable, and Reusable (the FAIR principles). However, scientists and researchers, in general, do not have the necessary knowledge to be able to make their research in such a way, and this is simply because it is a complete field of research by itself. As a consequence, in this chapter, it is proposed a framework that is optimally designed for both designing and executing a project, as well as publication of the methods along with the published paper report with the results.

This is not the first attempted solution to this fundamental problem. Various solutions have been proposed since the early 1990s, see the Appendix of Akhlaghi, Infante-Sainz et al. (2020) for a review. To better highlight the differences between those methods, and the foundations of this method (which help in understanding certain implementation choices), in the subsections below, the core principle above is expanded by breaking it into logically independent sub-components.

The proposed solution is a low-level skeleton that is designed to be easily adapted to any high-level, project-specific, choice. For example, in terms of hardware choice, a large simulation project simply cannot be run on smaller machines. However, when such a project is managed in the proposed system, the complete project (see Section 2.3.1) is published and readable by peers, who can be sure that what they are reading, contains the full and exact environment with the commands that produced the result.

#### 2.3.1 Principle 1: completeness

A project should be complete (or self-contained), without needing any particular features from the host operating system (OS), and not affecting the host OS. At build-time (when the project is building its necessary tools), the project should not need anything beyond a minimal Portable Operating System Interface Unix (POSIX) environment on the host, which is available in Unix-like operating system like GNU/Linux<sup>3</sup>, Berkeley Software Distribution (BSD) based or macOS. At run-time (when the environment is built), the project should not use or affect any host operating system programs or libraries.

In general, the source of a project should include the whole project: access to the inputs (see Section 2.2), instructions to configure and build the necessary software, instruction to make

<sup>3</sup>GNU is a recursive acronym for “GNU” is Not Unix!

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

the analysis (run the software on the data), and construct the final paper report in its final format. As recommended by Zhao et al. (2012), a complete project automatically builds all its necessary third-party tools, it does not just assume that they exist. This principle has important consequences:

- A complete project does not need any privileged or root permissions for system-wide installation, or environment preparations. Even when the user does have root privileges, interfering with the host operating system for a project may lead to many conflicts with the host or other projects. This principle ensures the safe execution of the project, and that it will not cause any security problem. It is particularly important for high-performance computing (HPC) facilities, where because of security reasons, HPC users commonly do not have privileged permissions.
- A complete project does not need any user interaction and can complete itself automatically. This is because manual interaction is incompleteness. Interactivity is also an inherently irreproducible operation, exposing the analysis to human error, and requiring expert knowledge.

### 2.3.2 Principle 2: modularity

A project should be compartmentalized or partitioned into independent modules or components with well-defined inputs and outputs having no side-effects. In a modular project, communication between the independent modules is explicit, providing optimizations on multiple levels:

- Execution: independent modules can be run in parallel, or modules that do not need to be run (because their dependencies do not have changed) will not be re-done.
- Data lineage and data provenance extraction (recording the origin of any dataset).
- Citation: allowing others to credit specific parts of a project.

This principle does not just apply to the analysis phase, it applies to the whole project. Within the analysis phase, this principle can be summarized with the Unix philosophy, best described by McIlroy et al. (1978) in the “Style” section: “Make each program do one thing well. To do a new job, build afresh rather than complicate old programs by adding new features”. Independent parts of the analysis can be maintained as independent software (for example shell, Python, or R scripts, or programs written in C, C++, or FORTRAN, among others). This core aspect of the Unix philosophy has been the cause of its continued success (particularly through GNU and BSD) and development in the last half-century.

For the most high-level analysis and operations, the boundary between the “analysis” and “project” can become blurry. It is thus inevitable that some highly project-specific, and small, analysis steps are also kept within the project and not maintained as a separate software package (that is built before the project is run). This is not a problem, because inputs are defined as files that are usable by other projects (see Section 2.2.1). If necessary, such highly project-specific software can later spin-off into a separate software package.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 2.3.3 Principle 3: plain text

The primarily stored or archived format for the project should be plain text with human-readable encoding, for example, ASCII or Unicode.<sup>4</sup> The reason behind this principle is that opening, reading, or editing non-plain text (executable or binary) file formats need specialized software. Binary formats will complicate various aspects of the project: its usage, archival, automatic parsing, or human readability. This is a critical principle for long-term preservation and portability: when the software to read binary format has been deprecated or become obsolete and is not installable on the running system, the project will not be readable or usable anymore.

A project that is solely in plain text format can be put under version control as it evolves, with easy tracking of changed parts, using already available and mature tools in software development. Note that software source code is also in plain text. After publication, independent modules of a plain text project can be used and cited through services like Software Heritage (Di Cosmo et al. 2018, 2020), enabling future projects to easily build on top of old ones.

Archiving a binary version of the project is like archiving a well-cooked dish itself, which will be inedible with hardware changes (temperature, humidity, and the natural world in general). But archiving the recipe of the dish, also kept in plain text, will allow you to re-cook it any time.

A fundamental aspect that keeping the project in plain text makes possible is that any interested person would be able to read the deepest lines of code to be able to understand and check that everything is fine. Even when the researcher is not able to run the project or there is any problem in executing any of the steps.

### 2.3.4 Principle 4: minimal complexity

An important measure of the quality of a project is how much it avoids complexity. In principle, this is similar to Occam's razor extrapolated to project management: "Never posit pluralities without necessity" (Schaffer 2015). In this context, Occam's razor can be interpreted as the following cases: minimize the number of dependency software (there are often multiple ways of doing something), avoid complex relations between analysis steps (which is related to the principle of modularity discussed in Section 2.3.2), or avoid the programming language that is currently in vogue because it will fall out of fashion and it will take the project down with it. This principle has several important consequences:

- Easier learning curve. Scientists can not spend so much time adopting new tools and methods as fast as software developers. They have to invest the majority of their time in their own research domain. Because of this, researchers usually continue their careers with the languages and tools they learned when they started.
- Future usage. Scientific projects require longevity. Those that depend too much on an ever-evolving or high-level software developing toolchain, will be harder to archive, run, or even study for their immediate and future peers.

<sup>4</sup>Plain text format does not include document container formats like .odf or .doc, for software like OpenOffice or Microsoft Office.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Compatible and extensible. A project that has minimal complexity can easily adapt to any kind of data, programming language, host hardware, or software. It can also be easily extended for new inputs and environments.

### 2.3.5 Principle 5: verifiable inputs and outputs

The project should contain automatic verification checks for its inputs (software source code and data) and outputs. When applied, expert knowledge will not be necessary to confirm the correct reproduction. It is important to emphasize that in practice, exact or bit-wise reproduction is very hard to implement at the level of a file. This is because many specialized scientific software commonly prints the running date on their output files (which is very useful in its own context).

For example, in plain text tables, such metadata are commonly printed as commented lines (usually starting with #). Therefore, when verifying such plain text tables, the checksums that are used to validate the data can be recorded after removing all commented lines. Fortunately, the tools to operate on specialized data formats also usually have ways to remove requested metadata (like creation date) or ignore metadata altogether. For example, the FITS standard in astronomy (Pence et al. 2010) defines a special DATASUM keyword, which is a checksum calculated only from the raw data and ignoring all metadata.

### 2.3.6 Principle 6: history and temporal provenance

No exploratory research project is done in the first attempt. Projects evolve as they are being completed. Naturally, earlier phases of a project are redesigned only after later phases have been completed. The “history” of a project is as scientifically relevant as the final snapshot of the project. All the outputs (datasets or narrative papers) need to contain the exact point in the history of the project that produced them.

For a complete project (see Section 2.3.1) that is under version control (for example using Git), this would be the unique commit checksum or hash. This principle thus benefits from the plain text principle (Section 2.3.3). Note that with the definition of project (Section 2.2.3), “changes” in the project include changes in the software building or versions, changes in the running environment, changes in the analysis, or changes in the narrative. After publication, the history of a project can also be published on services like Software Heritage (Di Cosmo et al. 2018), enabling precise citation and archival of all stages of the evolution of the project.

Taking this principle to a higher level, newer projects are built upon the shoulders of previous projects. A project management system should be able to provide this temporal connection between projects. Quantifying how newer projects relate to older projects (for example through Git branches) will enable: 1) scientists to simply use the relevant parts of an older project, 2) quantify the connections of various projects, which is primarily of interest for meta-research or historical studies. In data science, “provenance” is used to track the analysis and original datasets that were used in producing a higher-level dataset.

### 2.3.7 Principle 7: free and open-source software

Technically, reproducibility is also possible with non-free and non-open-source software (a black box). This principle is thus necessary to complement the definition of reproducibility (Sec-

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

tion 2.2.6). This is because software freedom is an important pillar for the sciences as shown below:

- Based on the completeness principle (Section 2.3.1), it is possible to trace the provenance of the outputs back to the exact source code lines within the analysis software. If the source code of the software is not available, such important and useful provenance information is lost.
- A non-free software may not be runnable on a given hardware. Since free software is modifiable, others can modify (or hire someone to modify) it and make it runnable on their particular platform.
- A non-free software cannot be distributed by the authors, making the whole community reliant only on the proprietary owner’s server (even if the proprietary software does not ask for payments). A project that uses free software can also release the necessary tarballs of the software it uses. For example see [zenodo.3408481](#) (Akhlaghi 2019) or [zenodo.3524937](#) (Infante-Sainz et al. 2020).
- A core component of reproducibility is that anonymous peers should be able to confirm the result from the same datasets with minimal effort, and this includes financial cost beyond hardware.

## 2.4 Implementation of Maneage

The proposed solution is an implementation of the principles discussed in Section 2.3: it is complete and automatic (Section 2.3.1), modular (Section 2.3.2), fully in plain text (Section 2.3.3), having minimal complexity (Section 2.3.4), with automatically verifiable inputs and outputs (Section 2.3.5), preserving temporal provenance (Section 2.3.6), and it is free software (Section 2.3.7).

In practice, Maneage is a collection of plain text files that are distributed in pre-defined sub-directories by context, and are all under version-control (currently with Git). In its raw form (before customizing for different projects), it is a fully working skeleton of a project without much flesh: containing all the low-level infrastructure, with just a small demonstrative “delete-me” analysis. To start a new project, the users will clone<sup>5</sup> the core skeleton, create their own Git branch, and start customizing the core files (adding their high-level analysis steps, scripts to generate figures, and narrative).

In this section, the current implementation of Maneage is reviewed. Generally, the job orchestration is implemented with Make (a POSIX software), this choice is elaborated in Section 2.4.1. A general outline of the file structure of the project is explained in Section 2.4.2. As described there, it is important to remark the cosmetic distinction between “configuration” (or building the necessary software) and “execution” (or running the software on the data). These two phases are discussed in Sections 2.4.3 and 2.4.4.

<sup>5</sup>In Git, “cloning” is the process of copying all files of a project (and their history) into the host system.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 2.4.1 Job orchestration with Make

When non-interactive processing is needed (see Section 2.3.1), shell scripts are usually the first solution that comes to mind. However, the inherent complexity and non-linearity of the progress in a scientific project (where experimentation is fundamental) make it hard and inefficient to manage the scripts as the project evolves. For example, a script will start from the top every time it is run. Therefore, even if 90% of a research project has been already done, and only the remaining 10% must be executed, a script will always start from the start to execute all the steps.

It is possible to manually ignore (by using conditionals) or comment parts of a script to only do specific parts. However, such conditionals and comments will only add complexity and will discourage experimentation on an already completed part of the project. This is also prone to very serious bugs in the end (e.g., due to human error, some parts may be left-out or not up to date) when re-running from scratch. Such bugs are very hard to notice during the work and frustrating to find in the end. These problems motivated the creation of Make in the early Unix operating system (Feldman 1979).

**In the Make paradigm, process execution starts from the end: the final target.** Through the Make syntax, the user specifies the prerequisites of each target, and a recipe (a small shell script) to create the target from the prerequisites. With this lineage, Make can build a dependency tree internally and find the rules that need to be executed on each run. This has many advantages:

- **Only executing the necessary steps.** In the scenario above, a researcher that has just added the final 10% of the research, will only have to run those extra steps, without any modification to the previous parts. With Make, it is also trivial to change the processing of any intermediate (already written) rule (or step) in the middle of an already written analysis. In this case, next time Make is run, only rules that are affected by the modifications or additions will be re-run, not the whole analysis/project. Most importantly, this enables full reproducibility from scratch with no changes in the code of the project that was working fine during the research. This allows robust results and lets scientists do experiments to be critical with the analysis itself, without having to waste energy on the added complexity of experimentation in scripts.
- **Parallel processing.** Since the dependencies are demarcated in Make, it can identify independent steps and run them in parallel. This greatly speeds up the processing, with no cost in terms of complexity.
- **Codifying data lineage and provenance.** In many systems, data provenance has to be manually added. However, in Make, it is part of the design and no extra manual step is necessary to fully track the series of steps that generated the data.

Make has been a fixed component of POSIX (or Unix-like operating systems including Unix, GNU/Linux, BSD, and macOS, among others) from the very early days of Unix, almost 40 years ago. It is, by far, the most reliable, commonly used, well-known, and well-maintained workflow manager nowadays. Because the core operating system components are built with it, Make is expected to keep this unique position in the foreseeable future. Make is also well known by many outside of the software developing communities.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

In summary, Make satisfies all the principles described above (see Section 2.3), while avoiding the well-known problems of using high-level languages for project management.

### 2.4.2 General implementation structure

As described above, a project using this template is a combination of plain text files that are organized in various directories by context. Figure 2.2 shows this directory structure and some representative files in each directory. The top-level source only has two main directories: `tex/` (containing  $\LaTeX$  files) and `reproduce/` (containing all other parts of the project) as well as several high-level files. Most of the top project directory files are only intended for human readers (as narrative text, not scripts or programming sources): `COPYING` is the high-level copy-right license, `README.md` is a basic introduction to the specific project, and `README-hacking.md` describes how to customize the template for creators of new projects.

In the top project directory, there are two non-narrative files: `project` (which should have been under `reproduce/`) and `paper.tex` (which should have been under `tex/`). The former is necessary for the top project directory because it is the high-level user interface, with the `./project` command. The latter is necessary for many web-based automatic paper generating systems like arXiv, journals, or systems like Overleaf.

`project` is a simple executable POSIX-compliant shell script. It is the high-level wrapper script to call the Makefiles of the project. Recall that the main job orchestrator in this system is Make (see Section 2.4.1). In the current implementation, the execution of the project consists of the following two calls to the `project` script:

```
./project configure # Build the software
./project make      # Do the analysis
```

The operations of both are managed by files under the top-level `reproduce/` directory. When the first command is called, the contents of `reproduce/software` are used. Similarly, the second command executes the files under `reproduce/analysis`. This highlights the cosmetic distinction adopted in Maneage between the two main steps of a project: 1) building the full software environment of the project, and 2) doing the analysis (running the software). Technically there is no difference between the two and they could easily be merged under one directory. However, during a research project, researchers commonly just need to focus on their analysis steps and will rarely need to edit the software environment settings (maybe only once at the very beginning of the project). Therefore, having the files mixed under the same directory can be confusing.

In summary, the same structure governs both aspects of a project: software building and analysis. This is an important and unique feature in this template. Researchers that have become familiar with Makefiles for orchestrating their analysis, will also easily be able to modify the Makefiles for the software that is built in the project and feel free to customize their own software.

### 2.4.3 Project configuration

A critical component of any project is the set of the necessary software to do the analysis. Verifying an already built software environment (which is critical to reproduce the results) is very hard. This fact has forced most projects to move around the whole built software environment (a black box) as virtual machines or containers. Because these black boxes are almost impossible

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

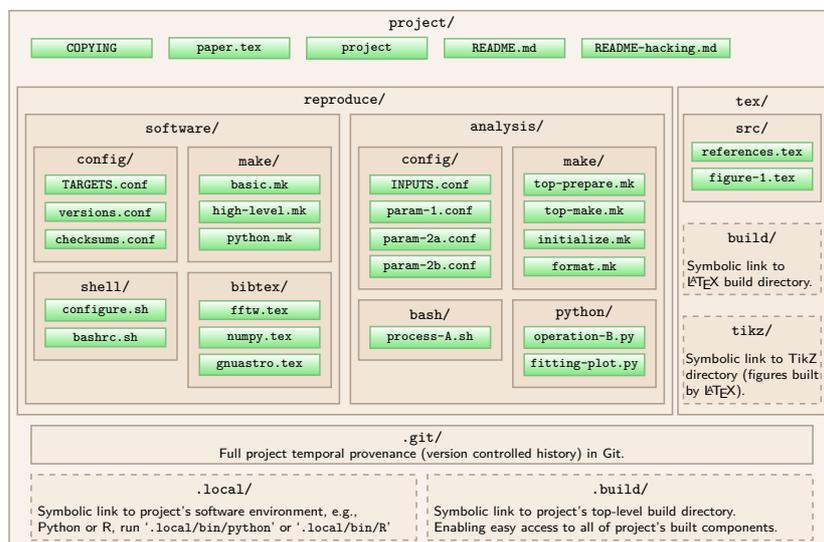


Figure 2.2: Directory and file structure in a hypothetical project using Maneage. Files are shown with small, green boxes that have a suffix in their names (for example `format.mk` or `download.tex`). Directories (containing multiple files) are shown as large, brown boxes, where the name ends in a slash (`/`). Directories with dashed lines and no files (just a description) are symbolic links that are created after building the project, pointing to commonly needed built directories. Symbolic links and their contents are not considered part of the source and are not under version control. Files and directories are shown within their parent directory. For example the full address of `format.mk` from the top project directory is `reproduce/analysis/make/format.mk`.

to reproduce themselves, they need to be archived, even though they can take gigabytes of space. Package managers like Nix or GNU Guix do provide a verifiable software building environment, but because they aim to be generic package managers, they have their own limitations on a project-specific level.

Based on the principles of completeness and minimal complexity (Sections 2.3.1 and 2.3.4), a project using Maneage also contains the full instructions to build its necessary software in the same language that the analysis is orchestrated: Make. The project configuration (building the software environment) is managed by the files under `reproduce/software`. The project configuration involves three high-level steps which are discussed in the next subsections: setting the local directories, checking a working C compiler, and the software source code download, build and install.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 2.4.3.1 Setting local directories

All files built by the project (software or analysis) will be under a “build directory” (BDIR) on the host filesystem. No other location on the running operating system will be affected by the project. Following the modularity principle (Section 2.3.2), this directory should be separate from the source directory. Therefore, at configuration time, the first thing to specify is the BDIR directory on the host system. The build directory can be specified in two ways: 1) on the command-line with the `--build-dir` option, or 2) manually giving the directory after running the configuration: it will stop with a prompt and some explanation.

Two other local directories can optionally be specified by the project when inputs are present locally and do not need to be downloaded: 1) software tarball directory and 2) input data directory. The project just needs reading permissions on these directories: when given, nothing will be written inside of them. The project will only look into them for the necessary software tarballs and input data. If they are not found, the project will attempt to download any necessary file from the recoded URLs/PIDs within the project source. These directories are therefore primarily tailored to scenarios where the project must run offline.

After the project has been configured, a symbolic link is generated on the top project source directory. It points to the build directory, BDIR. The symbolic link is a hidden file named `.build`, see Figure 2.2. This symbolic link allows access to the built files very easily, no matter where the build directory is located on the filesystem.

### 2.4.3.2 Checking for a C compiler

Maneage builds all its necessary software internally to avoid dependency issues with various software versions on different hosts. A working C compiler is thus mandatory and the configure script will abort if a working C compiler is not found. In particular, on GNU/Linux systems, the project builds its own version of the GNU Compiler Collection (GCC), therefore a static C library is necessary with the compiler. If not found, an informative error message will be printed and the project will abort.

The custom version of GCC is configured to also build FORTRAN, C++, objective-C, and objective-C++ compilers. Python and R running environments are themselves written in C, therefore they are also automatically built afterward if the project uses these languages. On macOS systems, Maneage currently does not build a C compiler, but it is planned to do so in the future.

### 2.4.3.3 Building necessary software from source

All necessary software for the project and its dependencies are installed from the source. Researchers using the template only have to specify the most high-level analysis software they need in `reproduce/software/config/TARGETS.conf` (see Figure 2.2). Based on the completeness principle (Section 2.3.1), on GNU/Linux systems the dependency tree is automatically traced down to the GNU C Library and GCC. Thus creating identical high-level analysis software on any system. When the C library and compiler can not be installed (for example on macOS systems), the users are forced to rely on the C compiler and library of the host, and this may hamper the exact reproducibility of the final result. Because the main output of the project is currently a L<sup>A</sup>T<sub>E</sub>X-built PDF, the project also contains an internal installation of T<sub>E</sub>XLive,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

providing all the necessary tools to build the PDF, independently of the L<sup>A</sup>T<sub>E</sub>X version and packages already present in the host operating system.

To build the software from source, the project needs access to its source tarball or zip file. If the tarballs are already present on the system, the user can specify the respective directory at the start of project configuration (Section 2.4.3.1). If not, the software tarballs will be downloaded from pre-defined servers. Ultimately, the origin of the tarballs is irrelevant for this project, what matters is the content of the tarball: checked through the SHA-512 checksum (part of the SHA-2 algorithms, see Romine 2015). If the SHA-512 checksum of the tarball is different from the checksum stored for it, the project will complain and abort.

Software tarball access, unpacking, building, and installation is managed through Makefiles, see Sections 2.4.1 and 2.4.2. The software of the project is classified into two classes: 1) basic and 2) high-level. The former contains meta-software: software needed to build other software, for example, GNU Gzip, GNU Tar, GNU Make, GNU Bash, GNU Coreutils, GNU SED, GNU Binutils, GNU Compiler Collection (GCC), etc.<sup>6</sup> The basic software is built for any project with the tools of the host operating system. The high-level software is used directly in the scientific analysis and can differ from project to project. However, because the basic software has already been built by the project, the higher-level software is built with them and independently of the host operating system.

The building of the software is managed by two top-level Makefiles that follow the same classification. Both are under the `reproduce/software/make/` directory (Figure 2.2): `basic.mk` and `high-level.mk`. Because `basic.mk` can not assume anything about the host, it is written to comply with POSIX Make and POSIX shell, which are very limited compared to GNU Make and GNU Bash, respectively. However, after it is finished, a specific version of GNU Make (among other basic software) is installed, thus making it possible to use more advanced features of GNU tools in `high-level.mk`.

The software of the project is installed under `BDir/software/installed`. The `.local` symbolic link in the top project source directory points to this directory for easy access (see Figure 2.2). It contains the top-level POSIX filesystem hierarchy subdirectories for the project, including `bin/`, `lib/`, `include/`, among others. For example, the custom-built GNU Make executable is placed under the directory `BDir/software/installed/bin/make` or alternatively in `.local/bin/make`.

To orchestrate software building with Make, the building of each software is represented as a file. In the Makefiles, that file should be used as a target in the rules that build the software, or as a prerequisite in the rules of the software that depends on it. Therefore, once a program is built, a simple plain text file is created under a sub-directory of `.local/version-info`. The representative files for C/C++ programs or libraries are placed under the `proglib` sub-directory. The Python or T<sub>E</sub>XLive representative files are placed under the `python` and `tex` subdirectories respectively. Make uses this file to refer to the software and arrange the order of software execution. The contents of this plain text file are the name and possible citation of the software that is directly imported into the final paper in the end.

<sup>6</sup>Note that almost all these GNU software are also installable on non-GNU/Linux operating systems like BSD or macOS, exceptions include GNU Binutils.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

#### 2.4.3.4 Software citation

Based on the completeness principle (Section 2.3.1), Maneage contains the full list of installed software, their versions, and their configuration options. However, this information is buried deep into the source of the project. A distilled fraction of this information must also be printed in the final report blended into the narrative. Furthermore, when a published paper is associated with the used software, it is important to cite that paper. The citations help software authors gain more recognition and grants, encouraging them to further develop it. This is particularly important in the case of research software, where the researcher invests significant time in building the software and requires official citation to justify continued work on it.

One notable example that highlights this issue is GNU Parallel (Tange 2018). Every time it is run, it prints the citation information before it starts. This does not cause any problem in automatic scripts, but can be annoying when reading or debugging the outputs. Users can disable the notice with the `--citation` option and accept to cite its paper, or support its development directly by paying 10 000 euros. This is justified by an uncomfortably true statement: “history has shown that researchers forget to [cite software] if they are not reminded explicitly. [...] If you feel the benefit from using GNU Parallel is too small to warrant a citation, then prove that by simply using another tool”.<sup>7</sup> In bug 905674, the Debian developers argued that because of this extra condition, GNU Parallel should not be considered as free software, and they are using a patch to remove that part of the code for its build under Debian-based operating systems.<sup>8</sup> Most other research software does not resort to such drastic measures, however, citations are important for them. For a review of the necessity and basic elements in software citation, see Katz (2014) and Smith et al. (2016). Given the increasing number of the software used in scientific research, the only reliable solution is to automatically cite the used software in the final paper.

As mentioned above in Section 2.4.3.3, a plain text file is built automatically at the end of the successful installation of the software. This file contains the name, version, and possible citation of that software. At the end of the configuration phase, all these plain text files are merged into one  $\LaTeX$  macro that can be imported directly into the final paper or report. The paragraph produced by this macro will not be too large, but it will greatly help in the recording of the used software environment and will automatically cite the software where necessary.

In the current version of Maneage, it is assumed that the published report of the project is built by  $\LaTeX$ . Therefore, every software that has an associated paper, has a Bib $\TeX$  file under the `reproduce/software/bibtex` directory. When the software is built for the project (possibly as a dependency of another software specified by the user), the Bib $\TeX$  entries are copied to the build directory, and the command to cite that Bib $\TeX$  record is included in the  $\LaTeX$  macro with the name and version of the software.

#### 2.4.4 High-level organization of the analysis

Once a project is configured (Section 2.4.3), the software is ready to be used. The analysis phase of the project (running the software on the data) is also orchestrated through Makefiles.

<sup>7</sup>GNU Parallel’s FAQ on the need to cite software: <http://git.savannah.gnu.org/cgi/parallel.git/plain/doc/citation-notice-faq.txt>

<sup>8</sup>Debian bug on the citation notice of GNU Parallel: <https://bugs.debian.org/cgi-bin/bugreport.cgi?bug=905674>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

In order to best follow the principle of modularity (Section 2.3.2), the analysis is not done in one phase or with a single Makefile. Here, the two high-level phases of the analysis are reviewed. The organization of the lower-level analysis, in many modular Makefiles, is discussed in Section 2.4.5.

After running `./project make`, the analysis is done in two sequential phases: 1) preparation and 2) main analysis. The purpose of the preparation phase is further elaborated in Section 2.4.4.2. Technically, these two phases are managed by the two high-level Makefiles: `top-prepare.mk` and `top-make.mk`. Both are under `reproduce/analysis/make` (see Figure 2.2) and both have an identical lower-level analysis structure. But before that, in Section 2.4.4.1 the isolation of the analysis environment from the host is discussed.

#### 2.4.4.1 Isolated analysis environment

By default, the analysis part of the project is not exposed to any variable environment of the host. This is accomplished through the `env -i` command, which will remove the host environment.<sup>9</sup> The project will define its own values for standard environment variables to avoid using the system or user defaults. Combined with the fact that all the software were configured and compiled from source for each project at configuration time (Section 2.4.3.3), this completely isolates the analysis from the host operating system, creating an exactly reproducible result on any machine that the project can be configured.

For example, the project builds its own fixed version of GNU Bash (a command-line shell environment). It also has its own `bashrc` startup script<sup>10</sup>, and the `BASH_ENV` environment variable is set to load this startup script. Furthermore, the `HOME` environment variable is set to `BDIR` to avoid the penetration of any existing Bash startup file of the user's home directory into the analysis.

#### 2.4.4.2 Preparation phase

When `./project make` is called, the first Makefile that is run is `top-prepare.mk`. It is designed for any selection steps that may be necessary to optimize `top-make.mk`, or to “prepare” for it. It is mainly useful when the research targets are a sub-sample of the inputs, and may not be necessary for many scenarios. Its role is described here with an example.

Consider that the raw input data (that the project received from a database) has 5000 rows (potential targets to analyze). However, this particular project only needs to work on 100 of them, not the full 5000. If the full 5000 targets are given to `top-make.mk`, Make will need to create a data lineage for all 5000 targets and project authors have to add checks in many places to ignore those that are not necessary. This will add more complexity to the project and it is prone to many bugs. Furthermore, if the filesystem is not fast (for example a filesystem that exists over a network), checking all the intermediate and final files over the full lineage can be slow.

In this scenario, the preparation phase finds the IDs of the 100 targets of interest and saves them as a Make variable in a file under `BDIR`. Later, this file is loaded into the analysis phase, precisely identifying the targets of interest. This selection phase can not be done within

<sup>9</sup>Note that the project's self-built `env` program is used, not the one provided by the host operating system. Within the project, `env` is installed as part of GNU Coreutils. The option `-i` stands for `--ignore-environment`.

<sup>10</sup>The Bash startup script of the project is under `reproduce/software/shell/bashrc.sh`, see Figure 2.2.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

`top-make.mk` because the full data lineage (all input and output files) must be known to Make before it starts to execute the necessary operations. Make can call itself as another Makefile, but this practice is strongly discouraged here because it makes the flow very hard to read. However, if the project authors insist on calling Make within Make, it is certainly possible (Section 2.4.5.1).

The preparation phase thus allows `top-make.mk` to optimally organize the complex set of operations that must be run on each input and the dependencies (possibly in parallel). It also greatly simplifies the coding for the project authors. Ideally `top-prepare.mk` is only for the “preparation phase”. However, projects can be complex, and ultimately, the choice of which parts of an analysis being a “preparation” can be highly subjective. Generally, the internal design and concepts of `top-prepare.mk` are identical to `top-make.mk`. Therefore in Section 2.4.5, where the lower-level management is discussed, we will only focus on the latter to avoid confusion.

#### 2.4.5 Low-level organization of analysis

A project consists of many steps, including data access (possibly by downloading), running various steps of the analysis on the obtained data, and creating the necessary plots, figures, or tables for a published report, or output datasets for a database. If all of these steps are organized in a single Makefile, it will become very large, or long, and will be hard to maintain, extend/grow, read, reuse, and cite. Generally, large files are a bad practice because it is against the modularity principle (Section 2.3.2).

The proposed template is thus designed to encourage and facilitate modularity by distributing the analysis in many Makefiles that contain contextually-similar (or modular) analysis steps. In the rest of this paper these modular, or lower-level, Makefiles will be called subMakefiles. The subMakefiles are loaded into `top-make.mk` in a certain order and executed in one instance of Make without recursion. In other words, this modularity is just cosmetic for Make: Make sees all the subMakefiles as if they were parts of one file. However, this modularity plays a critical role for the human reader/author of the project and is necessary for re-using or citing parts of the analysis in other projects.

Within the source of the project, the subMakefiles are placed in `reproduce/analysis/make` (in the same directory than `top-make.mk`), see Figure 2.2. Therefore, by design, `top-make.mk` is very simple: it just defines the ultimate target (`paper.pdf`), and the name and order of the subMakefiles that should be loaded into Make.

The precise organization of the analysis steps highly depends on each particular project. However, many aspects of the project management are the same, irrespective of the particular project, here we will focus on those. Figure 2.3 is a general overview of the analysis phase in a hypothetical project using Maneage. As described above and shown in Figure 2.3, `top-make.mk` imports the various Makefiles under the `reproduce/` directory that are in charge of the different phases of the analysis. Each of the subMakefiles builds intermediate targets, or outputs (files), which are shown there as blue boxes. In the subsections below, the analysis is described using this graph. Following the Make’s paradigm (see Section 2.4.1), the description of the process will be done by starting from the ultimate target (`paper.pdf`) and tracing back its lineage all the way up to the inputs and configuration files.

To avoid getting too abstract in the subsections below, where necessary, a basic analysis on the data of Menke et al. (2020) is done.<sup>11</sup> The reason for using this analysis example is

<sup>11</sup>The data were published as supplementary material on bioRxiv.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

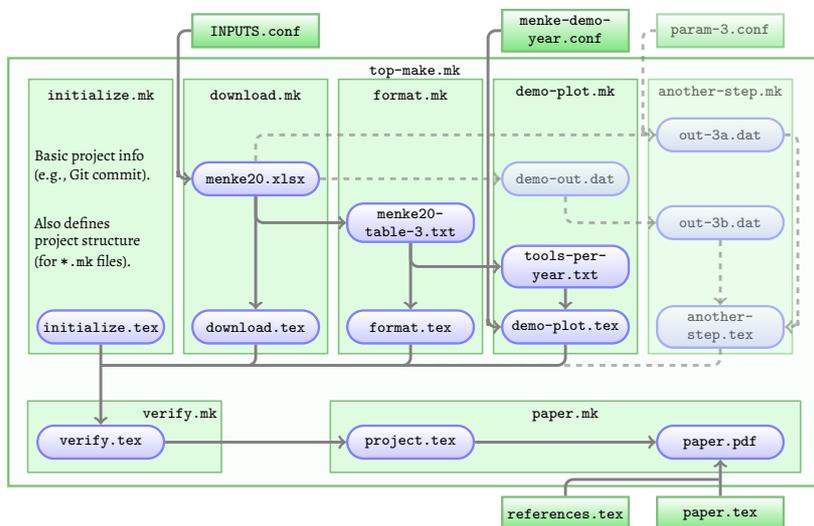


Figure 2.3: Schematic representation of data lineage in a hypothetical project/pipeline using Maneage. Each colored box is a file in the project and the arrows show the dependencies between them. Green files/boxes are plain text files that are under version control and into the source-directory. Blue files/boxes are output files of various steps in the build-directory, located within the Makefile (\*.mk) that generates them. For example `paper.pdf` depends on `project.tex` (in the build directory and generated automatically) and `paper.tex` (in the source directory and written by hand). In turn, `project.tex` depends on all the \*.tex files at the bottom of the Makefiles above it. The solid arrows and built boxes with full opacity are described in the context of a demonstration project in this paper. The dashed arrows and lower opacity built boxes, just show how adding more elements to the lineage is also easily possible, making this a scalable tool.

just to illustrate that Maneage has been designed to be used in any research field, not only in astrophysics. However, note that because not the same software is being used, this is not a reproduction (see Section 2.2.6). It is not possible to use the same software because they use Microsoft Excel for the analysis, which violates several principles: 1) Completeness (as a graphic user interface program, it needs human interaction, Section 2.3.1), 2) Minimal complexity (even free software alternatives like LibreOffice involve many dependencies and are extremely hard to build, Section 2.3.4), and 3) Free software (Section 2.3.7).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

### 2.4.5.1 Non-recursive Make

It is possible to call a new instance of Make within an existing Make instance. This is known as recursive Make.<sup>12</sup> Recursive Make is used by many Make users, especially in the software development communities. However, recursive Make is discouraged and not used in Maneage.

All the subMakefiles mentioned above are included<sup>13</sup> into `top-make.mk` (i.e., their contents are read into Make as it is parsing `top-make.mk`). From the point of view of Make, this is identical to having one long file with all the subMakefiles concatenated one to each other. Make is only called once and there is no recursion.

Furthermore, in Maneage there is the convention that only `top-make.mk` (or `top-prepare.mk`) can include subMakefiles. SubMakefiles should not include other subMakefiles. The main reason behind this convention is the Minimal Complexity principle (Section 2.3.4): a simple glance at `top-make.mk` will immediately show all the subMakefiles of the project and their loading order. When the names of the subMakefiles are descriptive enough, this enables both the project authors and project readers to get a complete view of the various stages of the project.

### 2.4.5.2 Ultimate target: the report of the project (`paper.pdf`)

The ultimate purpose of a project is to report the data analysis results. In scientific projects, this report is the published paper. In the industry, it is a quality-check and analysis of the final data products. In both cases, the report contains many visualizations of the final data product of the project, for example as plots, figures, tables, or numbers blended into the narrative description.

In Figure 2.3 it is shown as `paper.pdf`, note that it is the only built file (blue box) with no arrows leaving it. In other words, nothing depends on it: highlighting its unique ultimate target position in the lineage. The instructions to build `paper.pdf` are in `paper.mk`. The source code of the report is `paper.tex` and it contains the main narrative as well as that of the figures and tables. To build the final report in PDF format, `references.tex` and `project.tex` are also loaded into  $\LaTeX$ .

### 2.4.5.3 Values within the report (`project.tex`)

Figures, plots, tables, and the narrative are not the only analysis outputs that go into the paper. In many cases, quantitative values from the analysis are also blended into the sentences of the report. For example, this sentence in the abstract of Akhlaghi (2019): “[...] detect the outer wings of M51 down to S/N of 0.25 [...]”. The reported signal-to-noise ratio value “0.25” depends on the analysis and is an output of the analysis just like the figures and tables. Manually typing the number in the  $\LaTeX$  source is prone to very important bugs. The author may forget to check it after a change in the analysis (e.g., using a newer version of the software, or changing an analysis parameter for another part of the paper). Given the evolution of scientific projects, this type of human error is very hard to avoid when such values are manually written. As a consequence, such values must also be automatically generated.  $\LaTeX$  macros are used to automatically generate and blend them in the text.

<sup>12</sup>[https://www.gnu.org/software/make/manual/html\\_node/Recursion.html](https://www.gnu.org/software/make/manual/html_node/Recursion.html)

<sup>13</sup>[https://www.gnu.org/software/make/manual/html\\_node/Include.html](https://www.gnu.org/software/make/manual/html_node/Include.html)

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

In the quote above, the  $\LaTeX$  source<sup>14</sup> looks like this: “detect the outer wings of M51 down to S/N of  $\$demofoptimizedsn\$$ ”. The  $\LaTeX$  macro “ $\$demofoptimizedsn\$$ ” is automatically calculated and recorded during in the project and expands to the value “0.25”. The automatically generated file `project.tex` stores all such inline output macros. Furthermore, Figure 2.3 shows that it is a prerequisite of `paper.pdf` (as well as the manually written  $\LaTeX$  sources that are shown in green). Therefore `paper.pdf` will not be built until this file is ready and up-to-date.

However, managing all the necessary  $\LaTeX$  macros for a full project in one file is against the modularity principle and can be frustrating and buggy. To address this problem, all sub-Makefiles must contain a fixed target with the same base-name, but with a `.tex` suffix. For example, in Figure 2.3, assume `menke20-table-3.txt` is a table and the mean of its third column must be reported in the paper. Therefore, in `format.mk`, a prerequisite of `format.tex` is `menke20-table-3.txt` (as shown by the arrow in Figure 2.3). The recipe of this rule will calculate the mean of the column and put it in the  $\LaTeX$  macro which is written in `format.tex`. In a similar way, any other reported calculation from `format.mk` is stored as a  $\LaTeX$  macro in `format.tex`.

These  $\LaTeX$  macro files form the core skeleton of the project. As shown in Figure 2.3, the outward arrows of all built files of any subMakefile ultimately lead to one of these  $\LaTeX$  macro files. Note that built files in a subMakefile do not have to be a prerequisite of its `.tex` file. They may point to another Makefile’s  $\LaTeX$  macro file. For example even though `demo-out.dat` is a target in `demo-plot.mk`, it is not a prerequisite of `demo-plot.tex`, it is a prerequisite of `out-3b.dat` (a target in `another-step.mk`). The lineage ultimate ends in a  $\LaTeX$  macro file in `another-step.tex`.

#### 2.4.5.4 Verification of outputs (verify.mk)

An important principle is that outputs should be automatically verified, see Section 2.3.5. However, simply confirming the checksum of the final PDF, or figures and datasets is not generally possible. As mentioned in Section 2.3.5, many tools that produce datasets or PDF files write the creation date into the produced files. Therefore it is necessary to verify the outputs before the PDF is created. To facilitate output verification, the project has a `verify.mk` Makefile, see Figure 2.3. It is the only prerequisite of `project.tex` that was described in Section 2.4.5.2. Verification is therefore the connection-point, or bottleneck, between the analysis steps of the project and its final report.

Prior to publication, the project authors should add the MD5 checksums of all the  $\LaTeX$  macro files in the recipe of `verify.tex`. The necessary structure is already there, so adding or changing the values is trivial. If any  $\LaTeX$  macro is different in future builds of the project, the project will abort with a warning of the problematic file. When projects involve other outputs (for example images, tables, or datasets that will also be published), their contents should also be validated. To do this, prerequisites should be added to the `verify.tex` rule that automatically check the contents of other project outputs. Recall that many tools print the creation date automatically when creating a file and this kind of metadata must be ignored. `verify.tex`

<sup>14</sup>Akhlaghi (2019) uses Manage to be reproducible, so its  $\LaTeX$  source is available in multiple ways: 1) direct download from arXiv:1909.11230, by clicking on “other formats”, or 2) the Git or zenodo.3408481 links are also available on arXiv.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

contains some Make functions to facilitate the checking with some file formats, others can be added easily.

#### 2.4.5.5 Project initialization (initialize.mk)

The `initialize.mk` subMakefile is present in all projects and it is the first subMakefile that is loaded into `top-make.mk` (see Figure 2.3). Authors rarely need to modify or edit this file, it is part of the low-level infrastructure of Maneage. Nevertheless, project authors are strongly encouraged to study it and use all the useful variables and targets that it defines. `initialize.mk` does not contain any analysis or major processing steps, it just initializes the system. For example, it sets the necessary environment variables, internal Make variables and defines generic rules like `./project make clean` (to clean/delete all built products, not software) or `./project make dist` (to package the project into a tarball for distribution) among others.

It also adds one special L<sup>A</sup>T<sub>E</sub>X macro in `initialize.tex`: the current Git commit that is generated every time the analysis is run. It is stored in the `\projectversion` macro and can be used anywhere within the final report. One good place to put it is at the end of the abstract for any reader to be able to identify the exact point in history that the report was created. It also uses the `--dirty` feature of Git's `--describe` output. If any version-controlled file is not already committed, the value to this macro will have a `-dirty` suffix. If it is in a prominent place (like the abstract), it will always remind the author to commit their work.

#### 2.4.5.6 Importing and validating inputs (download.mk)

The `download.mk` subMakefile is present in all Maneage projects and contains the common steps for importing the input datasets into the project. All necessary input datasets to the project are imported through this subMakefile. This helps in modularity and minimal complexity (Sections 2.3.2 and 2.3.4). To see which external datasets were used in a project, this is the only necessary file to manage/read. Also, a simple call to a downloader (for example `wget`) is not usually enough. Irrespective of where the dataset is used, some operations are always necessary when importing datasets:

- The file may already be present on the host, or the user may not have an internet connection. Hence it is necessary to check the given input directory on the host before attempting to download it.
- The network might temporarily fail but connect with an automatic re-trial. Crashing the whole analysis because of a temporary network issue requires human intervention and is against the completeness principle (Section 2.3.1).
- Before it can be used, the integrity of the imported file must be confirmed with its stored checksum.

In some scenarios, the generic download script may not be useful. For example when the database takes queries and generates the dataset for downloading on the fly. In such cases, the users can add their own Make rules in this `download.mk` to import the file. They can use its predefined structure to do the extra steps like validating it. Note that in such cases the servers often encode the creation date and version of their database system in the resulting

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

file as metadata. Even when the actual data is identical, this metadata (which is in the same file) will differ based on the moment the query was done. Therefore a simple checksum of the whole downloaded file could not be used for validation in such scenarios, see Section 2.3.5. Each external dataset has some basic information, including its expected name on the local system (for offline access), the necessary checksum to validate it (either the whole file or just its main data), and its URL/PID. In Maneage, such information is in the `INPUTS.conf` file.

#### 2.4.6 The analysis

The analysis subMakefiles are loaded into `top-make.mk` after the initialization and download steps (see Section 2.4.5.5). However, the analysis phase involves much more complexity. If done without modularity in mind from the start, research project sources can become very long, thus becoming hard to modify, debug, improve, or read. Maneage is therefore designed to encourage and facilitate splitting the analysis into multiple and modular subMakefiles. Theoretical discussion of this phase can be hard to follow, and because of that, here only the more important and basic considerations will be exposed. In Appendix A.1 there is a hands-on tutorial for starting with Maneage. There, authors will be able to do an easy research example to learn the basic steps from the very beginning, without assuming any previous knowledge.

Consider the data lineage graph of Figure 2.3, it is clear that the analysis is broken into three subMakefiles: `format.mk`, `demo-plot.mk`, and `another-step.mk`. The first issue is that `menke20.xlsx` should be converted to a simple plain text table, which is generically usable by simple tools (see the principle of minimal complexity in Section 2.3.4). This is done in `format.mk`, where the original downloaded table is converted into a simple white-space separated, plain text table (`menke20-table-3.txt`), and some basic calculations are obtained. In order to keep the source and output-directories separate, if necessary, it makes a new directory (for example, `subdir`) under the build-directory (BDIR, see Section 2.4.3.1). As a consequence, the output file of `format.mk` (`menke20-table-3.txt`) is saved into the directory `BDIR/subdir`. Generally, the rule to create the output directory will be the first rule of each Makefile.

Since each Makefile is supposed to make a particular and well-defined task, each output will be in a different output directory. Following the same example, `demo-plot.mk` and `another-step.mk` will generate their own sub-directories, where the outputs will be saved. Note how this way of organizing the outputs follows the principle of modularity. In addition to this, the analysis subMakefiles should only contain the organization of the analysis, they should not contain any fixed numbers, settings, or parameters. Such elements should only be used as variables that are defined elsewhere. In the data lineage plot of Figure 2.3, configuration files are shown as the sharp-edged, green `*.conf` files in the top row. All the configuration files of a project are placed under the `reproduce/analysis/config` (see Figure 2.2) subdirectory, and are loaded into `top-make.mk` before any of the subMakefiles. The configuration files greatly simplify the project management from multiple perspectives as listed below:

- If an analysis parameter is used in multiple places within the project, simply changing the value in the configuration file will change it everywhere in the project. This is cortical in more complex projects and, if not done like this, it can lead to significant human error.
- Configuration files enable the logical separation between the low-level implementation and high-level running of a project. For example, after writing the project, the authors do

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

not need to remember where the number/parameter was used, they can just modify the configuration file. Other co-authors or readers of the project also benefit. They just need to know that there is a unified place for high-level project settings, parameters, or numbers without necessarily having to know the low-level implementation.

- A configuration file will be a prerequisite to any rule that uses its value. If the configuration file is updated (the value/parameter is changed), Make will automatically detect the data lineage branch that is affected by it and re-execute only that branch, without any human interference.

This is a great leap compared with the current, mostly manual, project management that many scientists employ. Manual management is prone to serious human error factors. At the later phases of a project, scientists are least likely to experiment on their project's configurations. However, the later phases of a project are precisely the times where the lower-level parts of the project are complete and the authors can look at the bigger picture. This style of managing project parameters, therefore, produces a much more healthy scientific result where experimentation is very cheap during all phases of a project: before its publication (by the authors) and after it (by the authors and readers).

#### 2.4.7 Projects as Git branches of Maneage

Maneage is fully composed of plain text files distributed in a directory structure (see Sections 2.3.3, 2.4.2, and Figure 2.2). Therefore, it can be maintained under version control systems like Git. Every commit in the version-controlled history contains a complete snapshot of the data lineage. Maneage is maintained by its developers in a central branch, which will be called **maneage** hereafter. The **maneage** branch contains all the low-level infrastructure, or skeleton, that is necessary for any project. This skeleton is primarily the configuration features discussed in Section 2.4.3. The **maneage** branch only contains a minimal demonstration analysis in order to be complete.<sup>15</sup>

To start a new project, users simply clone it from its reference repository and build their own Git branch over the most recent commit. This is demonstrated in the first phase of Figure 2.4. The project started by branching of from commit 0c120cb in **maneage**. They can start customizing Maneage for the project by adding their high-level analysis in their own branch (**project**) and pushing it to their own Git repository. Maneage contains a file called **README-hacking.md** that has a complete checklist of steps to start a new project and remove the demonstration parts. In the same way, it also contains a file called **README-tutorial.md**, which is a complete hands-on tutorial to start using Maneage with a very simple research example. This tutorial is also present in this thesis as in Appendix A.1. Note that both files are updated on the **maneage** branch and will always be up-to-date with the low-level infrastructure. Consequently, as Maneage evolves, the tutorial in Appendix A.1 and the up-dated one may differ.

After a project starts, Maneage will evolve. For example, new features will be added, low-level bugs will be fixed that are useful for any project. Because all the changes in Maneage are committed on the **maneage** branch (that projects also branch-off from), updating the project's low-level infrastructure is as easy as merging the **maneage** branch into the project's branch. For

<sup>15</sup>The names of all the files related to the demonstration of the **maneage** branch have a **delete-me** prefix to highlight that they must be deleted when starting a new project.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

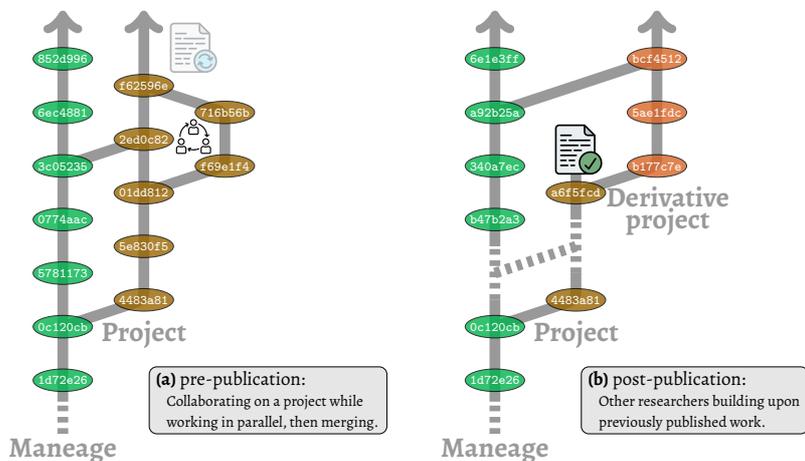


Figure 2.4: Projects start by branching off the main Maneage branch and developing their high-level analysis over the common low-level infrastructure. The low-level infrastructure can always be updated (keeping the added high-level analysis intact), with a merge between branches. Two phases of a project are shown here. In phase 1, a co-author has made two commits in parallel to the main project branch, which has later been merged. In phase 2, the project has finished. Note the identical first project commit and the Maneage commits it branches from. The dashed parts of scenario 2 can be any arbitrary history after those shown in phase 1. Another team now wants to build upon that published work in a derivate branch or project. The second team applies two commits and merges its branch with Maneage to improve the skeleton and continue their research. The Git commits are shown on their branches as colored ellipses, with their hash printed in them. The commits are colored based on the team that is working on that branch. The collaboration and paper icons are respectively made by ‘mynamepong’ and ‘iconixar’ and downloaded from [www.flaticon.com](http://www.flaticon.com).

example, see how Maneage’s 3c05235 commit has been merged into the branch project trough commit 2ed0c82 in Figure 2.4 (phase 1).

This does not just apply to the pre-publication phase. When done in Maneage, a project can be revived at any later date by other researchers as shown in phase 2 of Figure 2.4. In that figure, a new team of researchers have decided to experiment on the results of the published paper and have merged it with the Maneage branch (commit a92b25a) to fix some possible portability problem for their operating system that was fixed as a bug in Maneage after the publication of the paper. Propagating bug fixes or improvements in the low-level infrastructure to all projects using Maneage has been one of the reasons it evolved so well over the last 5 years, see Section 2.4.10. As it has become more used in different projects by many users, a lot of bugs were found and improvements made. These bug fixes and improvements would propagate to all other projects automatically in their next merge with the maneage branch.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Other scenarios include a third project that can easily merge various high-level components from different projects into its own branch, thus adding a temporal dimension to its data lineage. Modern version control systems provide many more capabilities that can be exploited through Maneage in project management. They are possible thanks to the shared branch it has with all projects that use Maneage.

#### 2.4.8 Multi-user collaboration on single build directory

Because the source of the project and build directories are separate, different users can share a build directory while working on their own separate project branches during a collaboration. Similar to the parallel branch that is later merged in phase 1 of Figure 2.4.

To give all users privilege, Maneage assumes that they are in the same (POSIX) user group of the system. All files built in the build directory are then automatically assigned to this user group, with read-write permissions for all group members (`-rwxrwx---`), through the `sg` and `umask` commands that are prepended to the call to Make. The `./project` script has a special `--group` option which activates this mode in both configuration and analysis phases. It takes the user group name as its argument and the built files will only be accessible by the group members, even when the shared location is accessible by people outside the project.

When multiple project members are contributing to a shared build directory, they usually work on independent parts of the project which will not cause any conflict. When there is a conflict, a member can temporarily change the name of the part of the top directory within their branch. The project already applies this strategy for part of the project that runs  $\text{\LaTeX}$  to build the final report. This is because project members will usually also be editing their parts of the report/paper as they progress. To fix this, when the project is configured and built with `--group`, the username of each member of the project will be appended to the  $\text{\LaTeX}$  build directory (which is under  $\$(\text{BDIR})/\text{tex}$ ).

#### 2.4.9 Publishing the project

Once the project is complete, publishing the project is the final step. In a scientific scenario, it is public. As discussed in the various steps before, the source of the project (the software configuration, data lineage, and narrative text) is fully in plain text, greatly facilitating the publication of the project.

##### 2.4.9.1 Automatic creation of publication tarball

To facilitate the publication of the project source, Maneage has a special `dist` target during the build process which is activated with the command `./project make dist`. In this mode, Maneage will not do any analysis, it will simply copy the full source of the project (in the current commit) into a temporary directory and compress it into a `.tar.gz` file. If a Zip compression is necessary, the `dist-zip` target can be called instead `dist`.

The `dist` tarball contains the full data lineage and is enough to reproduce the full project. It can build the software, download the data, run the analysis, and build the final PDF. However, it does not contain the Git history, it is just a checkout of one commit. Instead of the history, it contains all the necessary built products that go into building the final paper without the analysis. For example, the used plots, figures, tables, and `project.tex`, see Section 2.4.5.3.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

As a result, the tarball by itself can build the final report with a simple `pdflatex paper` command without running `./project`. When the project is distributed as a tarball (not as a Git repository), building the report may be the main purpose, like the arXiv distribution scenario discussed below, the data lineage (under the `reproduce/` directory) is likely just a supplement.

#### 2.4.9.2 What to publish, and where?

The source of the project, which is fully in hand-written plain text, has a very small volume, usually much less than one megabyte. However, the necessary input files (see Section 2.2.1) and built datasets may be arbitrarily large, from megabytes to petabytes or more. Therefore, there are various scenarios for the publication of the project as described below:

- **Only source.** Publishing the project source is very easy because it only contains plain text files with a very small volume. A commit will usually be on the scale of  $\sim 100$  kB. With the Git history, it will usually only be on the scale of  $\sim 5$  MB.
  - **Public Git repository.** This is the simplest publication method. The project will already be on a (private) Git repository before publication. In such cases, the private configuration can be removed so it becomes public.
  - **arXiv.** On arXiv it is possible to upload the  $\LaTeX$  source of the paper. arXiv will run its own internal  $\LaTeX$  engine on the uploaded files and produce the PDF that is published. When the project is published, arXiv also allows users to anonymously download the  $\LaTeX$  source tarball that the authors uploaded.<sup>16</sup> Therefore, simply uploading the tarball from the `./project make dist` command is sufficient for arXiv, and will allow the full project data lineage to also be published there with the  $\LaTeX$  source. Some of the very first papers in which this procedure were done are Akhlaghi (2019, arXiv:1909.11230) and Infante-Sainz et al. (2020, arXiv:1911.01430). Since arXiv is mirrored in many institutes over the planet, this is a robust way to preserve the reproducible lineage.
  - **In output datasets.** Many data storage formats support an internal structure with the data file. One commonly used example today is the Hierarchical Data Format (HDF), and in particular, its HDF5 which can host a complex filesystem in POSIX syntax. Since the volume of the project source is so insignificant compared to the output datasets of most projects, the whole project source can be stored with each published data file if the format supports it.
- **Source and data.** The project inputs (including the software tarballs, or possible datasets) may have a large volume. Publishing them with the source is thus not always possible. However, based on the definition of inputs in Section 2.2.1, they are usable in other projects: another project may use the same data or software source differently. Therefore, even when published with the source, it is encouraged to publish them as separate files.

<sup>16</sup>In the current arXiv user interface, the tarball is available by clicking the “Other formats” link on the paper’s main page, and then clicking “Download source”, it can be checked with <https://arxiv.org/abs/1909.11230> of Akhlaghi (2019) or <https://arxiv.org/abs/1911.01430> of Infante-Sainz et al. (2020)

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

For example, this strategy was also followed in [Akhlaghi \(2019\)](#) and [Infante-Sainz et al. \(2020\)](#).

- **Final PDF.** For an easy understanding of the project.
- **Git history.** As the Git “bundle” of the project. This single file contains the full Git history of the project until its publication date (only 4 MB), see Section 2.4.7.
- **Project source tarball.** Output of `./project make dist`, as explained above.
- **Tarballs of all necessary software.** This is necessary in case the software webpages are not accessible for any reason at a later date, or if the project must be run with no internet access. This is only possible because of the free software principle discussed in Section 2.3.7.

Note that [Akhlaghi \(2019\)](#) used previously published datasets that are automatically accessed when necessary. Also, that paper did not produce any output datasets beyond the figures shown in the report, therefore the Zenodo upload does not contain any dataset. When a project involves data collection or added-value data products, they can also be uploaded with the files above.

#### 2.4.9.3 Worries about getting scooped

Publishing the source of the project with the paper can have many benefits for the researcher and the community. For example, if the source is published with a pre-print, others may help the authors to find bugs, or improvements to the source that can affect the validity or precision of the result, or simply optimize it so it does the same work in half the time, for example.

However, one particular feedback raised by a minority of researchers is that publishing the reproducible data lineage immediately after publication may hamper their ability to continue harvesting from all their hard work. Because others can easily reproduce the work, others may take the next follow-up project they originally intended to do. This is informally known as getting scooped.

The level that this may happen is an interesting subject to be studied once many papers become reproducible. But it is a valid concern that must be addressed. Given the strong integrity checks in Maneage, we believe it has features to address this problem in the following ways:

1. This worry is essentially the phase 2 of Figure 2.4. The commits of the other team are built upon the commits of the original authors. It is therefore clear (with the precision of a character) how much of their result is purely their own work (qualitatively or quantitatively). In this way, Maneage can contribute to a new concept of authorship in scientific projects and help to quantify Newton’s famous “standing on the shoulders of giants” quote. However, this is a long-term goal and requires major changes to academic value systems.
2. The authors can be given a grace period where the journal, or some third authority, keeps the source and publishes it a certain time after publication. Journals can create specific policies for such scenarios. For example, setting that all project sources will be available publicly,  $N$  months or years after publication while allowing authors to opt-out

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

of it if they like, so the source is published immediately with the paper. However, journals cannot expect exclusive copyright to distribute the project source in the same way they do with the final paper. As discussed in the free software principle of Section 2.3.7, the project source must be free for the community to use, modify, and distribute.

This can also be done by the authors using servers like Zenodo, where authors can obtain the final DOI of the dataset, and publish it at a later date. Reproducibility is indeed very important for the sciences, but the hard work that went into it should also be acknowledged for the authors that would like to publish the source at a later date.

#### 2.4.10 Past, present, and future of Maneage

As with any software, the core architecture of Maneage will inevitably evolve. The current version introduced here has already experienced 5 years of evolution and several reincarnations. Its primordial implementation was written for Akhlaghi & Ichikawa (2015). This paper described a new detection algorithm in astronomical image processing. An automated sequence of commands to build the figures, and update the paper/report was a practical necessity as the algorithm was evolving. In particular, it did not just reproduce figures, it also used L<sup>A</sup>T<sub>E</sub>X macros to update numbers printed within the text.

The system later evolved in Bacon et al. (2017), in particular the two sections of that paper that were done by M. Akhlaghi: Akhlaghi & Bacon (2018, zenodo.1163746) and Akhlaghi et al. (2018, zenodo.1164774). With these projects, the skeleton of the system was written as a more abstract “template” that could be customized for separate projects. The template later matured by including installation of all necessary software from source and used in Akhlaghi (2019, zenodo.3408481), Infante-Sainz et al. (2020, zenodo.3524937) (see Chapter 3). It has been later used in several other papers as in Montes, Infante-Sainz et al. (2020) (see Chapter 4), Peper & Roukema (2020, zenodo.4062461), and chapters of this thesis (see Chapter 5 and 6). The future of Maneage is wide and not only focused on astronomy-related fields but to be adopted in all kinds of scientific studies. Roukema (2020) used Maneage in a Covid-19 related work, thus, demonstrating that Maneage can also be useful in non-astronomy fields. As Maneage will evolve, a list of the notable changes will be kept in the README-hacking.md file of the project. In the same way, the tutorial README-tutorial.md (Appendix A.1), will be also kept updated to include all changes and improvements.

## 2.5 Discussion and conclusions

To optimally extract the potentials of big data in science, it is necessary to have a complete view of its lineage. Scientists are, however, rarely trained sufficiently in data management or software development, and the plethora of high-level tools that change every few years does not help. Such high-level tools are primarily targeted at software developers, who are paid to learn and use them effectively for short-term projects. Scientists, on the other hand, need to focus on their own research fields and need to think about longevity. In this chapter, a new approach to solve these problems has been introduced: Maneage (Managing data Lineage) (Akhlaghi, Infante-Sainz et al. 2020). Maneage is designed as a complete framework, providing scientists with a pre-built low-level skeleton, using simple and robust tools that have withstood the test of time and are actively maintained. Scientists can customize Maneage’s existing data management

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

for their own projects, enabling them to learn and master the lower-level tools. This improves their efficiency and the robustness of their scientific results, while also enabling future scientists to reproduce and build upon their work.

Maneage satisfies all the proposed criteria and it has been implemented in several projects and works, thus demonstrating that it is possible to use it in order to account for the problem of reproducibility. In fact, the data reduction of all telescopes and instruments, as well as the high-level analysis of this thesis, have been done using Maneage (see all next chapters). With a larger user-base and wider application in scientific (and hopefully industrial) applications, Maneage will grow and become even more robust, stable, and user-friendly.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 3

## The point spread function

In this chapter, we develop and make a practical application of one of the fundamental tools of the low surface brightness astronomer: the point spread function (PSF). This chapter corresponds to the already published paper “*The Sloan Digital Sky Survey extended point spread functions*” (Infante-Sainz et al. 2020, MNRAS, 491, 5317), in which it is showed how to obtain and use the PSF in low surface brightness studies. The construction of the extended PSFs has been done using Maneage (see Chapter 2). The techniques and tools developed in this chapter are also applied in the works presented in Chapters 4 and 6.

### 3.1 Introduction

The point spread function (PSF) describes the response of an imaging system to the light produced by a point source. Real PSFs have complex structures as their shapes depend on the optical path that light takes as it travels through the atmosphere and multiple optical elements, mirrors, lenses, detectors, etc. For the vast majority of astronomical works, only a tiny portion of the PSF (i.e. normally a few inner arcseconds; see e.g. Trujillo et al. 2001a,b) is characterized. In practice, however, the light of both point and extended sources are spread over the entire detector due to the effect of the PSF at large radii. Therefore, it is necessary to have a good understanding of its structure along the entire detector (typically extending over arcminutes or more).

Extended PSFs have become a vital tool to obtain precise photometric information in modern astronomical surveys. For instance, Slater et al. (2009) modelled the extended PSF and the internal reflections produced by the stars of the Burrell Schmidt telescope and showed that virtually all the pixels of the image are dominated by the scattered light by both stars and galaxies at 29.5 mag arcsec<sup>-2</sup> (V-band). Trujillo & Fliri (2016) also characterized and used the extended PSF of the 10.4m Gran Telescopio Canarias (GTC) telescope to model and remove the scattered light in ultradeep observations of the UGC 00180 galaxy. Even more troublesome for low-surface brightness studies is the finding (see e.g. Trujillo & Bakos 2013; Sandin 2014, 2015) that the outer regions of astronomical objects are severely affected by their own scattered light produced by the convolution with the PSF. In order to correct this effect, Karabal et al. (2017) generated the PSF and models of the internal reflections from images of the Canada-

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

France-Hawaii Telescope (CFHT) to de-convolve a sample of three galaxies and correct them from instrumental scattered light. More recently, Román et al. (2020) characterized the PSFs of the Stripe 82 survey and used them to model and correct the scattered light field produced by stars to study the optical properties of the Galactic cirri. All the above works have shown that having an extended PSF is crucial when accurate photometric and structure properties of astronomical objects at low-surface brightness levels are required.

One of the most commonly used surveys for measuring photometric properties of astronomical objects is the Sloan Sky Digital Survey (SDSS; York et al. 2000), covering  $14\,555\text{ deg}^2$  on the sky (just over 35 per cent of the full sky) in five photometric bands ( $u$ ,  $g$ ,  $r$ ,  $i$ , and  $z$ ). Although SDSS is a relatively shallow survey compared to current (see e.g. IAC Stripe 82 Legacy Survey; Fliri & Trujillo 2016; Román & Trujillo 2018) and future wide-area surveys (see e.g. Large Synoptic Survey Telescope (LSST), Euclid, Messier; LSST Science Collaboration et al. 2009; Laureijs et al. 2011; Valls-Gabaud & MESSIER Collaboration 2017), its massive use has motivated the exploration of their extended PSFs as a useful exercise that can be repeated in future deeper surveys.

Previous attempts to characterize the extended SDSS PSFs were conducted by Zibetti et al. (2004); de Jong (2008); Bergvall et al. (2010), and Tal & van Dokkum (2011). All of them are extensively analysed and discussed in Sandin (2014). In general, the goal of these studies was to analyse the outer part of SDSS galaxies statistically to check the role of the scattered light in their low-surface brightness regions. de Jong (2008) obtained extended  $g$ ,  $r$ , and  $i$  SDSS PSFs. These PSFs have an extension of  $\sim 1$  arcmin in radius. Those PSFs were created by the stacking of averaged radial profiles of stars and then assuming (following the shape of the intermediate region) an  $r^{-2.6}$  power-law function to extend their outer parts to a larger distance. Similarly, Tal & van Dokkum (2011) stacked all the bright stars ( $8.0 < m_r < 8.2$ ) from images of SDSS DR7 (Abazajian et al. 2009) in each of the five optical bands to characterize the outer parts of the PSFs, while the inner parts were obtained from non-saturated SDSS PSF models (Lupton et al. 2001). These authors also reached an extension of  $\sim 1$  arcmin in radius. Even though 1 arcmin is a relatively large extension, a more extended PSF is needed when dealing with nearby galaxies. In fact, as shown by Sandin (2014), PSFs of at least 1.5 times the size of objects are needed to make a proper analysis of the effect of the PSF on the light distribution of astronomical sources. In other words, the study of a large number of interesting objects whose size exceeds the 1 arcmin scale requires significantly larger PSFs. Motivated by this requirement and its utmost importance for studies involving deep imaging analysis, we have characterized the SDSS PSFs down to  $\sim 8$  arcmin (i.e. approximately an order of magnitude larger in extension than previous works) have been characterized. These PSFs are made publicly available to the astronomical community (see Section 3.3).

This chapter is structured as follows. In Section 3.2, we describe all the steps followed in order to build the SDSS extended PSFs. Section 3.3 is dedicated to the characteristics of the PSFs and how they can be used to model and correct for the scattered light of stars in the Coma Cluster SDSS image. In Section 3.4, we provide our discussion and conclusions.

### 3.2 Constructing the extended SDSS PSFs

This work aims to build, for the SDSS filter set, PSFs as extended as possible. To achieve this goal, it is necessary to use stars within a wide brightness range. Very bright ( $< 7$  mag) stars

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

(although saturated and with bleeding patterns in their central part) are useful for creating the outer part of the PSFs. The stacking of bright stars ( $\sim 9$  mag) is done to construct the intermediate region of the PSF, whereas faint ( $> 14$  mag; non-saturated) stars are used for assembling the central core (inner part).

The stars considered to construct the PSFs are selected from the US Naval Observatory (USNO-B1) Catalog (Monet et al. 2003). We only take stars having low proper motions (pmRA and pmDEC  $< 600$  mas/yr) and errors in their central spatial coordinates equal to zero (e\_RAJ2000 and e\_DEJ000 = 0). Each star is selected to construct a different region of the PSF according to its magnitude. Once the stars from the USNO-B1 Catalog are chosen, SDSS images of these stars in each filter are downloaded from the SDSS archive. We explored the possibility of using Gaia (Gaia Collaboration et al. 2018) for our selection of the brightest stars, but we found that this catalogue is incomplete for objects with magnitude brighter than 7 mag, and no stars with magnitudes below 1.7 mag are present.

The SDSS camera has an array of 30 CCDs (located in 5 rows and 6 columns; each row of the camera corresponds to a filter). Each CCD has a size of 10 arcmin along the drift scanning direction and 13 arcmin perpendicular to that direction (see Gunn et al. 1998, for a detailed description of the SDSS camera). We use single SDSS images to create the PSFs. We avoided using mosaic images generated from the combination of multiple images, because each CCD can have a different sky background value. None the less, since each star is observed at a different CCD position, it is possible to obtain a PSF larger than the size of single CCDs. This is because each star will contribute to generate a different region of the extended PSF depending on its location over the CCD image. Naturally, to have enough coverage of the outermost region of the PSF, it is necessary to have a large number of stars.

For  $g$ ,  $r$ , and  $i$  filters, in order to reach a signal-to-noise ratio ( $S/N$ ) above 3 along the entire radial profiles of the extended SDSS PSFs, we need to combine about 1000 individual star images. The images in  $u$  and  $z$  bands are noisier and shallower. Therefore, the same number of stars only allow us to obtain extended PSF profiles with  $S/N \sim 1$  at a radius of 8 arcmin for these two filters. As the brightest stars from the SDSS footprint are considered, the addition of fainter stars does not help to get better  $S/N$  in the final stacked PSFs. For the intermediate and inner parts of the PSFs we also combine  $\sim 1000$  stars. In practical terms, we find that in order to have a final number of 1000 useful stars we typically need to explore 1050 stars, as some of them are rejected due to quality cuts, which will be discussed next sections.

In order to make our results reproducible, in the next subsections we describe in detail the steps followed to build the PSFs. The construction of the outer part of the PSFs is described in Section 3.2.1. The intermediate and inner parts will be briefly described in Sections 3.2.2 and 3.2.3. The methodology used for the junction between the inner, intermediate, and outer parts of the PSFs is discussed in Section 3.2.4.

### 3.2.1 Constructing the outer part of the PSFs

We use the brightest stars from the SDSS footprint in each band to build the outer part of the PSFs. These stars are not necessarily the same for all the SDSS filters. In fact, we have selected stars from the USNO-B1 Catalog with brightness  $< 7.0$  mag as followed:  $B1\text{mag} < 7.0$  mag for the  $u$  SDSS PSF,  $R1\text{mag} < 7.0$  mag for the  $g$  and  $r$  SDSS PSFs, and  $I\text{mag} < 7.0$  mag for the  $i$  and  $z$  SDSS PSFs. The magnitudes of the stars in  $B1\text{mag}$  (optical  $B$  band between 400 and

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|  |                            |
|--|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA        | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA               | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguilár<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

500 nm),  $R1mag$  (optical  $R$  band between 600 and 750 nm), and  $Imag$  (optical  $I$  band between 750 and 1000 nm) are given in the USNO-B1 Catalog. This catalogue presents magnitudes in the photographic system (O, E, J, F, N)<sup>1</sup>, and not in standard systems (see [Monet et al. 2003](#), for a detailed description on how USNO-B1 Catalog was obtained).

The construction of the PSF's outer part is as follows. First, we recentre each individual star into a newer and larger image ( $im1$ ) with an odd number of pixels in order to have the centre of the star in the central pixel of  $im1$ . To do that, we run `SWarp` ([Bertin 2010](#)) using as the central position of the stars the coordinates provided by the USNO-B1 Catalog. Aiming to create a model for the outer part of the PSF as a result of the stacking of many individual stars, it is critical to detect and mask all the external sources. On doing this, we avoid the contribution of this undesired flux to the PSF model. To implement such a step, we run `NoiseChisel` and `Segment` ([Akhlaghi & Ichikawa 2015](#); [Akhlaghi 2019](#)) to create a segmentation map of each re-centred image ( $im1$ ). We used the following `NoiseChisel` and `Segment` parameters<sup>2</sup> to optimize the source detection and segmentation: `tilesize=100,100`; `interpnumgb=1`; `minnumfalse=1`; `qthresh=0.5`. The segmentation map is used to mask all flux contributions that does not belong to the central bright star.

To stack all the individual stars, the flux of each image is normalized to the flux within a ring of 1-pixel width (SDSS pixel scale is  $0.396 \text{ arcsec pixel}^{-1}$ ) at a radial distance of 60 arcsec from the centre of the star. This radial distance is selected because it is close enough to the centre, therefore a good  $S/N$  is obtained, but is also distant enough to avoid the region where most of the internal reflections produced by the optics of the telescope are found. Each image is divided by the mean flux ( $m1$ ) of this ring.

Bright stars are severely affected by bleeding and saturated pixels in SDSS images. The SDSS team has allocated a significant effort to correct this issue. None the less, the correction of the bleeding and saturated pixels done by SDSS pipeline ([Stoughton et al. 2002](#)) still shows (for some of the brightest stars) a depletion along the saturated pixel columns which are not part of the PSF. This effect becomes visible approximately at a radial distance of 300 pixels from the centre along the drift scanning direction. To avoid being affected by this issue at creating the PSF models, we reject those stars which have this depletion in the SDSS filters  $g$ ,  $r$ , and  $i$ . In  $u$  and  $z$  SDSS bands, the correction done by the SDSS pipeline of the saturated pixels is good enough and does not affect the creation of the PSFs. The depletion is especially relevant in the  $i$  band, and because of that, we illustrate this issue in [Figure 3.1](#).

The rejection of stars affected by the depletion is done by analysing the pixels of the saturated region and comparing them with the surrounding pixels close to this region. To do that, first of all, we mask a central band of 600 pixels width perpendicular to the drift scanning direction (cross lined region in the upper panel of [Figure 3.1](#)). Then we collapse vertically the image into a 1-dimensional vector using the median values of the non-masked pixels. This 1-dimensional vector is equivalent to an average profile of the outer part of the star along the perpendicular direction to the drift scanning direction (bottom panel of [Figure 3.1](#)). The central part of the profile is divided into three regions. Depending on the SDSS filter, these regions are placed at different positions motivated by the location of the depletion. For  $g$  and  $r$  SDSS bands the

<sup>1</sup>O, E, J, F, N photographic system magnitudes correspond to  $B1mag$ ,  $B2mag$ ,  $R1mag$ ,  $R2mag$ , and  $Imag$  in the [USNO-B1 VizieR catalogue](#).

<sup>2</sup>The meaning of the parameters used can be found in the documentation book of this software: <https://www.gnu.org/software/gnuastro/manual/>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

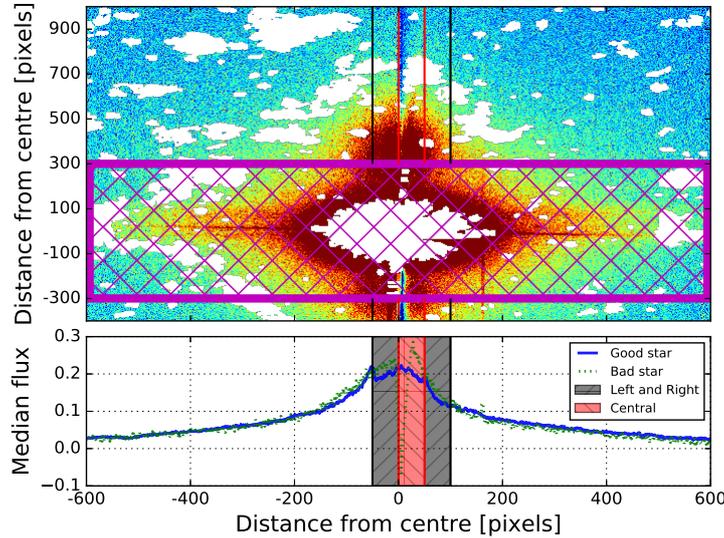


Figure 3.1: An illustration of how the stars affected by the depletion in the saturated pixels are rejected. Upper panel: SDSS  $i$  band image of a bright star (R.A. (J2000) = 22:55:10.96; Dec (J2000) = -04:59:16.37). The drift scanning direction of the survey goes from top to bottom. White areas correspond to masked pixels using `NoiseChisel` and `Segment` object masks. Inaccurately corrected saturated pixels by the SDSS pipeline can be seen above and below the star in the central region as a decrease in the flux of the image. The cross lined magenta rectangle encloses pixels that are masked in order to obtain the 1-dimensional profile of the star. Bottom panel: collapsed 1-dimensional profiles of two stars perpendicular to the drift scanning direction using the median value. With solid blue line, we plot the profile of one star (R.A. (J2000) = 14:19:45.23; Dec (J2000) = 16:18:25.02) unaffected by the problem of the depletion of saturated pixels ( $p = 0.84$ ). With dotted green line is plotted the profile of the star of the upper panel with the problem of the depletion ( $p = -14.36$ ). This can be clearly seen as a sharp drop of the profile in the central region. Parameter  $p$  is defined in Equation 3.1 (see the text for more details).

depletion is located in the middle of the image. Consequently, we put the central region in the middle of the image with a width of 10 pixels. For the  $i$  SDSS band, the depletion is not exactly at the central position, therefore the central region is moved 5 pixels from the centre to the right and has a width of 50 pixels. The other two regions are located at the left and right sides of the central region and have 50 pixels width for all bands. Using these regions, we define the parameter  $p$  as the ratio between the minimum value of the star profile in the left and right regions, and the minimum value of the central region. That is,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

$$p = \frac{\min(\text{left}, \text{right})}{\min(\text{centre})}. \quad (3.1)$$

With this definition, if the correction of the saturated pixels is reasonable,  $p$  is expected to be between 0 and 1. Note that  $p < 0$  implies that the profile of the star in the central region has a minimum value with a different sign than the profile values on the regions in each side. Alternatively,  $p > 1$  means that the profile has a strong depletion in the central region compared to the profile in the other two explored regions. Consequently, we reject all stars with  $p < 0$  or  $p > 1$  until the accumulated number of acceptable downloaded stars reaches 1050 in each band. For  $r$  and  $g$  filters, 1 out of 5 stars are rejected because of a strong depletion problem, while in the  $i$  band, this issue affects 1 out of 3 stars. Once we have done all the previous steps, we stack all masked and normalized star images using  $3\sigma$  clipping median.<sup>3</sup>

After the stacking of the brightest SDSS stars, we obtain an image which is a first version of the outer part of the PSF ( $\text{psf1}$ ). However, due to the brightness of the considered stars we have used, many contaminant sources (particularly those radially close to the central region of the stars) were not properly deblended during the first run using `NoiseChisel` and `Segment`. For this reason, to improve our masks, we do the following: we subtract the  $\text{psf1}$  from each original image ( $\text{im1}$ ) using the normalization value estimated before ( $n1$ ), leaving a residual image ( $\text{im2}$ ) in which we run `NoiseChisel` and `Segment` again to obtain better object masks. That is,

$$\text{im2} = \text{im1} - (n1 \times \text{psf1}). \quad (3.2)$$

`NoiseChisel` and `Segment` do the detection of pixels using boxes with a size as the one provided by the `tilesize` parameter, used to estimate the sky background and the  $S/N$ . For this reason, sometimes the detected pixels have the shape of these `tilesize`s. In order to avoid this effect, we choose random `tilesize`s. The parameters of `NoiseChisel` and `Segment` in the second iteration are `qthresh=0.5`, `minnumfalse=1`, `interpnumgb=1`, and `tilesize=irandom,irandom`, where `irandom` is a random integer between 70 and 100. Thanks to this second iteration, it is possible to mask the internal reflections and the diffraction spikes of stars in a much better way. With the new masks, the normalization of each image is improved and the stacking of all  $\text{im2}$  images (with the improved masks and normalization) results in a much better PSF ( $\text{psf2}$ ) than the initial one ( $\text{psf1}$ ).

By following the previous steps, the outer part of the PSFs have been built stacking about 1050 stars in each filter. However, not all the stars are equally bright. In fact, this difference in brightness plays a major role in the outermost part of the PSF if we want to create reliable PSFs. To guarantee that the PSFs are accurate enough along the entire radial range of exploration, we have modified our stacking procedure such at all radial distances the brightness of the profiles of the stars that are contributing to the stacking have brightness above the surface brightness limit of the images. This is done as follows. We extract the circularly averaged surface brightness radial profile of each star to estimate at which radial distance ( $R_{\text{mlim}}$ ) the surface brightness profile reaches  $\mu = 26.5 \text{ mag arcsec}^{-2}$ . This value is chosen as is the typical surface brightness limit of the SDSS survey (i.e.  $3\sigma$  in  $10 \times 10 \text{ arcsec}^2$  boxes; Pohlen & Trujillo 2006; Trujillo &

<sup>3</sup>In general, the median is a much more robust operator than the mean because it is less affected by outlier values. We compared stacked images using mean and median operator with  $3\sigma$  clipping, and the results are quite similar with sigma clipping median operator stacked images having slightly less noise.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Fliri 2016). Beyond  $R_{\text{mlim}}$  the signal of the star is so poor that if we were using these regions, the PSF model would have been affected by the inclusion of noise. Consequently, we mask all pixels beyond  $R_{\text{mlim}}$  for each star that is used for the final stacking.

An additional advantage of estimating  $R_{\text{mlim}}$  is that, combined with the normalization value we have used before to combine the stars, it can be used to remove the presence of stars heavily affected by contaminating sources. Example of such contamination could be the presence of another close-by bright star, the presence of Galactic cirri, a very crowded region with many sources, etc. The way we identify these potentially problematic cases is by plotting  $R_{\text{mlim}}$  versus the normalization factor (see Figure 3.2). In an ideal case (as all the stars used for creating the final stacked PSF have similar shapes), there should be a strong relation between the brightness at a radial distance of 60 arcsec (i.e. the normalization factor we have used) and the radial distance where the brightness is  $26.5 \text{ mag arcsec}^{-2}$  (i.e.  $R_{\text{mlim}}$ ). Figure 3.2 shows that this is in fact the case. The vast majority of stars (blue points) follow a nice trend between these two quantities. We can characterize this trend by fitting a linear relation (black solid line). The  $3\sigma$  outliers to this relation are plotted using red crosses. As expected, most of the outliers correspond to stars that have extra (parasitic) light from other nearby sources (that have not been properly masked), which artificially enlarge the location of  $R_{\text{mlim}}$ . By removing those stars, we create an accurate representation of the outer part of the PSFs. The final stacking of all “good” (normalized and masked) stars is done through a  $3\sigma$  clipped median (pixel-by-pixel). The removal of the outliers decreases our number of used stars to approximately  $1000^4$ .

### 3.2.2 Constructing the intermediate part of the PSFs

To build the intermediate part of the PSFs, we have followed essentially the same steps as the ones described in the previous Section 3.2.1. This time we have used 1050 stars of magnitude  $B1\text{mag} \sim 8.3 \text{ mag}$  for  $u$  SDSS band,  $R1\text{mag} \sim 9.0 \text{ mag}$  for  $g$  and  $r$  SDSS bands, and  $I\text{mag} \sim 9.3 \text{ mag}$  for  $i$  and  $z$  SDSS bands. Stars with this brightness have saturated pixels in the centre region. However, the bleeding pixels have been well corrected by the SDSS pipeline, and they do not have the problem of the depletion explained in the previous section, and thus, no rejection of stars was done because of this issue.

The normalization of the stars for creating the intermediate part of the PSF is conducted in a ring of 1-pixel width at a radial distance from the centre of 7 arcsec. As it was done for the outer part of the PSFs, two iterations are carried out in order to obtain better masks for undesired objects and also to improve the normalization of each star. Finally, we stacked all the masked and normalized stars using  $3\sigma$  clipping median to obtain the intermediate part of the PSF.

In this case, contrary to what happened for the outer part of the PSFs, all stars have the same brightness and therefore, their contribution in signal and noise to the final stacked image is very similar. Because of that, it was not necessary to mask the outer regions of these stars according to their  $S/N$ .

<sup>4</sup>The exact number of stars considered for the outer, intermediate, and inner parts can be found in the header of the PSF images with the keyword NOUTER, NINTER, and NINNER, respectively.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

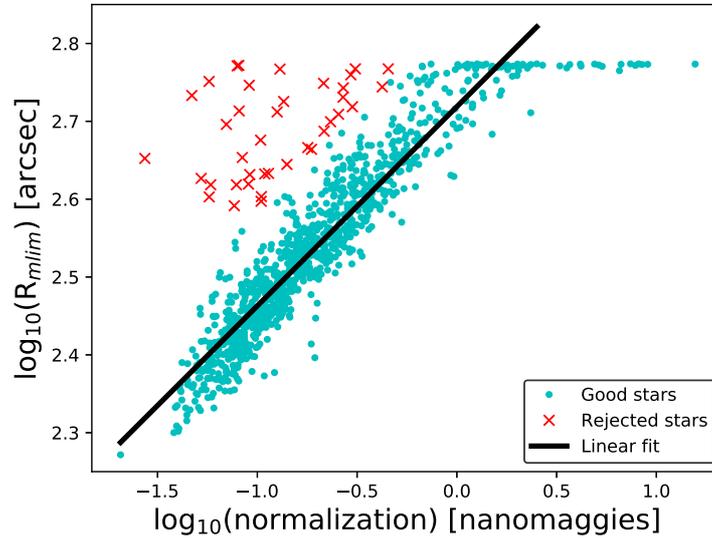


Figure 3.2: Removal of stars that are potentially affected by external (undesired) light in the outermost parts for the  $g$  SDSS band. We plot the radial distance ( $R_{\text{mlim}}$ ) at which the surface brightness profile of each star reaches  $26.5 \text{ mag arcsec}^{-2}$  (the surface brightness limit of the SDSS survey) versus the normalization value estimated at the radial distance of  $60 \text{ arcsec}$ . As expected, the vast majority of the stars (blue points) follow a linear correlation trend between the two quantities, except for some cases, where the stars have an artificially large  $R_{\text{mlim}}$  caused by parasitic light (red crosses). The outliers are identified through a  $3\sigma$  clipping mean taking the linear fit as reference (black solid line). The saturation of stars in the upper part of the plot with high-normalization values correspond to those stars which reach  $R_{\text{mlim}}$  at the border of the image (which has the same size for all the individual frames).

### 3.2.3 Constructing the central part of the PSFs

To build the central part of the PSFs, 1050 stars with brightness  $> 14 \text{ mag}$  in all bands from USNO-B1 Catalog ( $B1\text{mag}$  for  $u$  SDSS band,  $R1\text{mag}$  for  $g$  and  $r$  SDSS bands, and  $I\text{mag}$  for  $i$  and  $z$  SDSS bands) were considered. With this selection criteria, all the retrieved stars were unsaturated, and therefore we can obtain the core of the PSFs by simply stacking them after a proper treatment.

The creation of the outer and intermediate parts of the PSFs using the centre given by the USNO-B1 Catalog (using low proper motion stars) was accurate enough for our purpose. Indeed, due to the strong saturated and bleeding pixels, it was impossible to obtain the centre with more precision than the one given by the catalogue. However, when constructing the central part of the PSF, we can obtain the centre of each star more accurately, as we are considering

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

non-saturated stars. Note that this is important because small errors on the location of the central part create (after stacking many stars) wider PSFs than the true ones for the core. We use `Imfit` (Erwin 2015) to fit a 2-dimensional Moffat function to each star (see e.g. Trujillo et al. 2001b). `Imfit` provides the spatial coordinates of the centre of the stars. Then we use `SWarp` to resample the star into a new grid with the centre obtained from `Imfit`.

In the case of the stars used to create the central part of the PSFs, `NoiseChisel` and `Segment` was run only once as the stars are faint enough, so a first detection and segmentation is able to properly mask all external sources. The normalization of each star contributing to the stacked PSF is done in a ring of 1-pixel width at a radial distance of 2 arcsec. Once all images are ready, we stack them using  $3\sigma$  clipping median to obtain a first version of the inner part of the PSF (`psf1`).

An additional consideration that must be taken into account is the contamination of the USNO-B1 Catalog by sources that are not stars when selecting faint objects. That problem was not an issue in the construction of the outer and intermediate parts of the PSFs. This is because when considering very bright sources, the contaminants of the catalogue by sources that are not stars is negligible. The main contaminant source of the catalogue at fainter magnitudes is produced by galaxies. After obtaining the first version of the inner part of the PSF (`psf1`), we compare the stacked radial profile with the profiles of all the individual sources used to create `psf1`. In Figure 3.3 we show the radial profile of the `psf1` in solid black line, the radial profile of one bonafide star in dashed blue line, and the radial profile of one genuine galaxy in dot-dashed red line. We use the fact that the radial profile of galaxies and stars are different in a region centred at the normalization radius (2 arcsec) to reject galaxies from the sample. To remove the extended sources we measure the following parameter:

$$s = \text{std} \left( \frac{\text{object}}{\text{psf1}} \right)_{\text{Control area}} \quad (3.3)$$

Defined in this way,  $s$  is the standard deviation of the ratio between the intensity radial profile of the object (star or galaxy) and the intensity radial profile of the `psf1` in the “Control area” (shaded region in Figure 3.3). This region has 6 pixels width and is centred at the normalization radius (2 arcsec). The “Control area” is chosen such that it maximizes the difference at measuring  $s$  between the profile of a star and the profile of an extended source. The radial profile of a genuine star will have a very similar shape compared to the radial profile of `psf1`, and therefore a low value of  $s$  is expected. On the other hand, the radial profile of a galaxy would differ significantly from the radial profile of `psf1`, and because of that the  $s$  parameter will be larger. The distribution of  $s$  values is shown in Figure 3.4 and is used to distinguish between stars and galaxies. There are two clear peaks on the  $s$  distribution. The vast majority of objects are point-like sources and consequently, they are grouped around low values of  $s$  ( $\sim 0.07$ ). In addition to this, we find another group of extended objects with  $s$  values 10 times higher. These objects are removed from the final sample by using  $2.5\sigma$  clipped mean. Interestingly, in the USNO-B1 Catalog there is a star-galaxy estimator that measures the probability of a particular source to be a star (point source like object) or galaxy (extended source). However, we did not find an obvious correlation between the measured value of  $s$  and the star-galaxy estimator in the USNO-B1 Catalogue. Moreover, we found some sources classified as stars by USNO that, after visual inspection, turned out to be galaxies. Because of that, in this work, we only rely in the parameter  $s$  for rejecting extended sources from our sample. The number of objects removed

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

from the stacking due to this galaxy contamination depends on the filter, but on average the amount of objects rejected is about 20% of the original sample (1050 stars). Once they were removed, we repeat the entire process of stacking the images using  $3\sigma$  clipped median to obtain the central part of the PSFs.

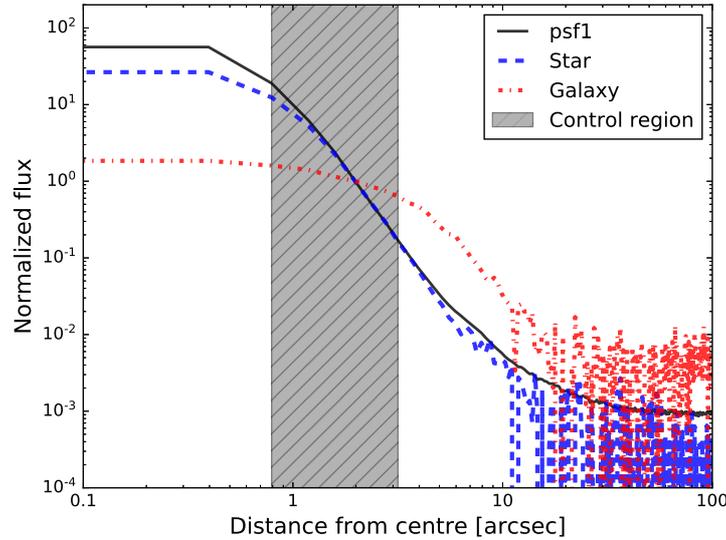


Figure 3.3: Radial profile of the stacked PSF (psf1; solid black line), an individual star (dashed blue line), and a contaminant galaxy (dot-dashed red line). The profiles are shown for  $r$  SDSS filter. The chosen star is at R.A. (J2000) = 16:13:49.55; Dec (J2000) = +32:21:27.43 while the galaxy is at R.A. (J2000) = 13:17:51.44; Dec (J2000) = +33:49:34.45. The grey region represents the radial location where the parameter  $s$  (Equation 3.3) is computed. All profiles intersect at a radial distance of 2 arcsec because it is the normalization radius. The parameter  $s$  is used to distinguish between stars and extended objects. For this particular case, the values are  $s = 0.13$  (star) and  $s = 0.88$  (galaxy).

### 3.2.4 PSF junction (central-intermediate-outer parts)

In this section, we describe how the different parts of the PSF are joined together in order to produce a representative final PSF. To conduct such a process, we use the outer part of the PSF as the reference. Then, the intermediate PSF is modified in such a way that its flux and reference sky level matches the shape of the outer PSF at a given radius, called the junction radius ( $r_{\text{jun}}$ ). Once this junction is done, the same process is done for joining the outer-intermediate PSF with

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

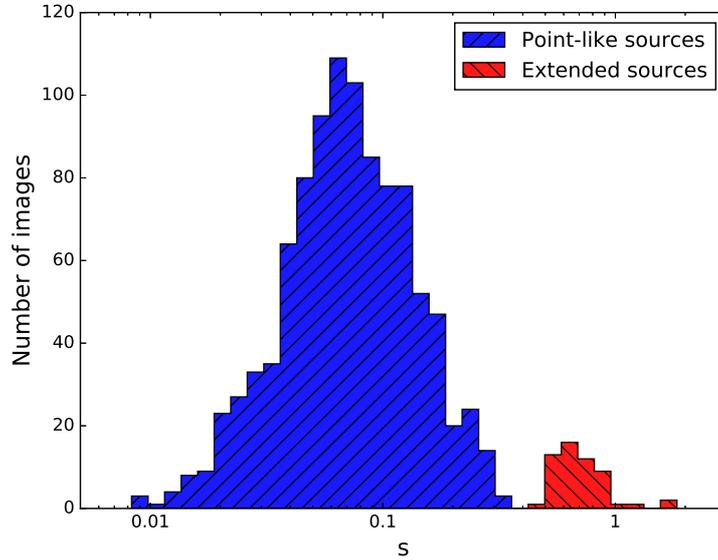


Figure 3.4: The removal of extended sources to create the central part of the PSFs. This figure shows the distribution of the parameter  $s$  in the SDSS  $r$  filter. The vast majority of the sources has a small value of  $s$  with a median  $s$  value equal to 0.07. Note that x-axis is in logarithmic scale. We have selected the dip between the blue and red distributions to separate our sample into bonafide stars and galaxy contaminants (see the text for details).

the central PSF<sup>5</sup>.

The region in which the junction is done depends on each SDSS band, but the criteria is the same for all bands. We require equal  $S/N$  between the different PSF regions at the junction radius in order to have a smooth transition between the different parts. Therefore, the choice of  $r_{\text{jun}}$  is calculated from the  $S/N$  radial profiles of the different sections of the PSF. Figure 3.5 shows the joining methodology for the PSF in the SDSS  $r$  band. As commented above, the PSF radial profile of the outer part is used as the reference (solid line with circles). For the  $r$  band, the saturation of the brightest stars appears at a radial distance of  $\sim 10$  arcsec. This can be seen as a flat radial profile for radii smaller than 10 arcsec. The radial profile of the intermediate part is plotted using a dashed line with down oriented triangles. This intermediate part has a different flux compared to the outer PSF part. The junction of these two PSF pieces consists in modifying the intermediate PSF to have the same flux (a multiplicative factor  $f$ ) and reference sky background level (additive factor  $c$ ) than the profile of the outer PSF. Having two

<sup>5</sup>The exact junction radius for joining the outer and the intermediate parts can be found in the header of the images with the keyword `ROUTER`. For the junction radius of the outer-intermediate with the central part, the keyword is `RINNER`.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

parameters to determine ( $f$  and  $c$ ) we need two independent equations to solve the problem. We do the following, we use the junction radius to separate two different regions in the profiles A region located at a radius  $r_{-1}$  (which is  $r_{\text{jun}} - 1$  pixel) and another region located at  $r_{+1}$  (which is  $r_{\text{jun}} + 1$  pixel). The equations to solve are

$$\begin{aligned} O_{-1} &= I_{-1} \times f + c \\ O_{+1} &= I_{+1} \times f + c \end{aligned} \quad (3.4)$$

where  $O_{-1}$  and  $O_{+1}$  are the values of the radial profile of the outer PSF part at  $r_{-1}$  and  $r_{+1}$ , respectively.  $I_{-1}$  and  $I_{+1}$  are the values of the radial profile of the intermediate PSF part at  $r_{-1}$  and  $r_{+1}$ , respectively. Consequently,  $f$  and  $c$  are

$$\begin{aligned} f &= \frac{O_{+1} - O_{-1}}{I_{+1} - I_{-1}} \\ c &= O_{-1} - I_{-1} \times \frac{O_{+1} - O_{-1}}{I_{+1} - I_{-1}} \end{aligned} \quad (3.5)$$

The junction between the central PSF part (dashed line with right oriented triangles in Figure 3.5) and the matched outer-intermediate PSF is done in the same way as in the case of the outer and intermediate PSFs junction. Note that the plateau region observed in the most inner radial profile (solid line with left oriented triangles) is not saturation but the value of the central pixel of the PSF. Considering these three PSF parts, we are able to obtain the entire extended PSFs without saturation in the inner part and high  $S/N$  in the outer part.

### 3.2.5 Final refinements to the extended PSFs

The outer parts of the PSFs have been built using very bright stars. As a consequence, the final PSF still contains a rectangular region along the central part of the PSF and perpendicular to the drift scanning direction that presents a depletion. This is an artificial effect due to the saturation of the brightest stars and remains after all the corrections made by the SDSS pipeline. To correct for this effect in the final PSFs, we interpolate the values of the pixels in these depleted regions using the flux values of adjacent (non-saturated) pixels.

Finally, the PSFs are rotated 90 deg counterclockwise to orientate the drift scanning direction with the left to right direction. That means that in our final PSFs the sky would move from left to right. To use our PSFs, they must be rotated to match the drift scanning direction of the given SDSS dataset. In the particular case of the Stripe 82 survey, since it observes an equatorial band, the sky is moving from left to right and consequently, this is the same direction as the orientation of the final PSFs. In any other case, the PSFs will have to be rotated to align the drift scanning direction of the image with the drift scanning direction of the PSF. To finish the process, we normalize each PSF to have the sum of all pixels equal to 1.

## 3.3 Results

By following the previous steps it has been possible to obtain PSFs in the five SDSS bands with an unprecedented radial extension ( $\sim 8$  arcmin). All PSFs obtained in this work are

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

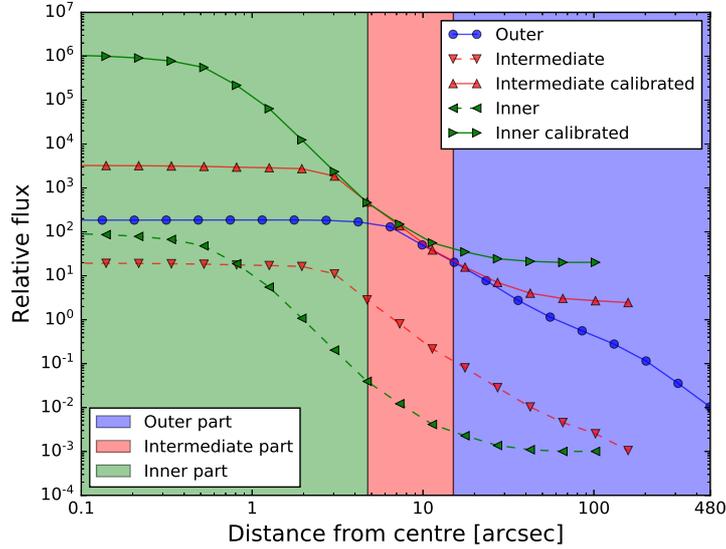


Figure 3.5: Radial profiles of the different PSF parts in the case of the SDSS  $r$  band. Dashed lines are radial profiles before the junction and solid lines are radial profiles after the matching. The radial profile of the outer PSF part (solid line with circles) is taken as the reference. The different shaded regions are colour coded in the bottom legend and correspond to the three pieces of the PSF. The vertical lines that divide the different coloured regions indicate the radius at which these parts have been joined.

publicly available and they can be found at the IAC Stripe 82 web page in FITS format<sup>6</sup>: <http://research.iac.es/proyecto/stripe82>. They are also available at CDS via anonymous ftp to <http://cdsarc.u-strasbg.fr/> (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/MNRAS> (Ochsenbein et al. 2000). It is important to bear in mind that PSFs provided above have to be rotated in order to align the drift scanning direction with the drift scanning direction of the image considered. Section 3.3 is divided in two parts. Section 3.3.1 is dedicated to describe the characteristics and features of the extended PSFs. In Section 3.3.2, the PSFs are used to model and remove the scattered light field of the stars in the SDSS image of the Coma Cluster central region as a practical example.

<sup>6</sup>The version number of each PSF is stored into the header of the images and can be checked with the keyword VERSION.

|   |                                  |
|---|----------------------------------|
| Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.       |                                  |
| Su autenticidad puede ser contrastada en la siguiente dirección <a href="https://sede.ull.es/validacion/">https://sede.ull.es/validacion/</a> |                                  |
| Identificador del documento: 3258585  | Código de verificación: 8GKKDuAS |
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA   | Fecha: 04/03/2021 11:59:31       |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA  | 04/03/2021 12:00:29              |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA   | 14/04/2021 15:46:33              |

### 3.3.1 PSFs characteristics

The extended SDSS PSFs can be seen in Figure 3.6. Along the entire structure, the  $S/N$  radial profiles of the  $g$ ,  $r$ , and  $i$  PSFs attain values above 3 while the  $u$  and  $z$  PSFs have  $S/N \sim 1$  at a radial distance of 8 arcmin. Beyond that radius, the quality of the PSFs drops significantly. This is because at larger distances, the low number of images used to create the PSFs are manifested through the appearance of residuals left by the object masks. Because of that, we have cropped the PSFs to have an extension in radius of 8 arcmin for all SDSS bands. All pixels beyond that distance from the centre (8 arcmin) have been masked with NaN (Not a Number) values. In any case, depending on the user's need, it is possible to crop the PSFs to a smaller size but taking into account that the normalization of the extended PSFs are done such that the sum of all (not NaN) pixels is equal to 1.

It is important to note that the structural properties of the PSF vary depending on multiple parameters, such as the camera column number, seeing, spatial location on the CCD, and others (see Gunn et al. 2006; Xin et al. 2018, for further details about the variation of SDSS PSFs). In this sense, the extended PSFs we present in this work represent the properties of averaged PSFs for the entire SDSS survey.

Although we have done our best during the masking process, including a double iteration in order to avoid internal reflections (i.e. the reflections of light produced after reflecting of the CCD and reflecting again in optical elements such as the filter, etc.), some residuals remain in the final PSFs. In each band, they appear at different positions. For example, in the case of  $g$  and  $r$  PSFs, the internal reflection appears at a radial distance of  $\sim 25$  arcsec from the centre. In the case of the  $i$  band PSF, the internal reflections are located in the centre. The fact that the internal reflections of  $g$ ,  $r$ , and  $i$  filters are near to their centres make it possible to subtract the scattered light produced by the stars without leaving behind strong residuals (we will show a practical example in the next section). In the case of the  $u$  and  $z$  bands, the PSFs present two big lobes of light on the left and right sides. These structures are aligned parallel to the drift scanning direction and they vary from exposure to exposure. It is therefore challenging to subtract the light of stars in  $u$  and  $z$  filters without producing strong residuals, although these residuals are stronger in the  $u$  band.

The diffraction spikes, produced by the secondary mirror structure of the telescope that are visible in individual images of stars in the SDSS survey, do not appear in our combined PSF models. This is because of the random orientation of the spikes in individual exposures. For this reason, they are averaged out in the final stacked PSFs. Additionally, in order to increase the  $S/N$  of the final models, the spikes were masked in the second masking iteration. We originally tried to include the spikes in the final PSFs, but the variability in flux and shape made impossible to model them realistically in the end.

#### The effect of the drift scanning in the PSFs

The drift scanning observing mode of the SDSS survey makes the PSFs somewhat asymmetric as is seen in Figure 3.6. In order to quantify this asymmetry, we study the flux along the drift scanning direction and along the perpendicular direction. We have done that by measuring the flux along these two directions using slits 50 pixels wide. As is seen in Figure 3.7, the  $g$  and  $r$ -band PSFs have an excess of flux in the drift scanning direction beyond 2 arcmin of radial distance from the centre (dotted lines compared with solid lines). This excess of flux is

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

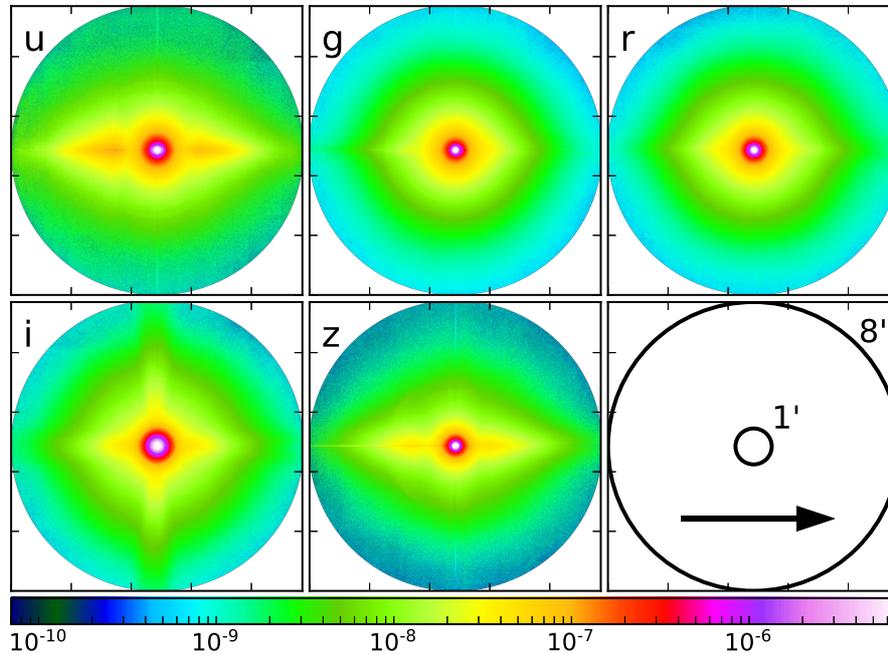


Figure 3.6: The extended SDSS PSFs. The bottom right panel illustrates the size of the PSFs showing, with two circles, the location of the radial distances of 1 and 8 arcmin. The arrow orientation indicates the drift scanning direction. All PSFs have been normalized to have a total flux of 1. The colour bar indicates the pixel values of each PSF. Note that the intensity range of the pixel extends near 5 orders of magnitude.

quantified by computing the ratio between the two profiles and the results are that along the drift scanning direction the flux is 1.6 and 1.7 times higher than in the perpendicular direction, respectively. The *i* band PSF presents a cross-like shape structure whose main arms correspond to the drift scanning and perpendicular directions. In the case of *u* and *z*, PSFs there are two bright lobes along the drift scanning direction. Along such direction, the fluxes are 5.6 and 4.1 times higher than in the perpendicular direction, respectively.

All these features demonstrate that an extrapolation of the PSFs to larger radii assuming a power law or any other particular function is too crude an approximation to represent the structural richness of the SDSS PSFs. Because of the non-symmetric shape of the PSFs, it is also necessary to rotate the PSF according to the orientation of each particular field when PSF models are used to subtract stars (see Section 3.3.2). The same is also necessary when the PSFs are used to model the effect of the PSF in astronomical objects with the aim of obtaining their

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

intrinsic physical parameters (e.g. when using *Imfit*, *GALFIT*, or similar astronomical software).

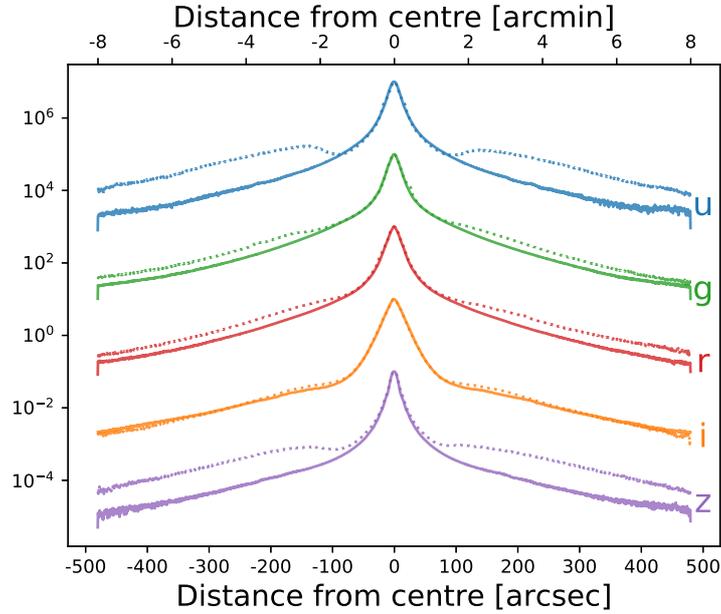


Figure 3.7: PSF profiles along the drift scanning direction (dotted lines) and perpendicular to it (solid lines). Each SDSS filter is annotated on the right side of the figure. To ease the reading, PSF profiles have been shifted vertically. The vertical axis is shown in arbitrary units. The horizontal axis corresponds to the radial distance from the centre in arcsec (bottom) and in arcmin (top).

### Radial profiles of SDSS PSFs

Figure 3.8 shows the circularly averaged radial profiles of all SDSS PSFs. For comparison, we also plot the radial profiles of the SDSS PSFs obtained by *de Jong (2008)*. To facilitate the comparison of the two works, we have re-normalized the flux of our PSFs to 1 only, accounting for the flux inside the maximum radius of the profiles produced by *de Jong (2008)* (i.e.  $R_{\max} \sim 46$  arcsec). The dynamical range of our PSFs covers 20 mag, and the PSF radial profiles decline as power laws ( $r^{-\alpha}$ ) with an exponent close to  $\alpha = -2.49$ .

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

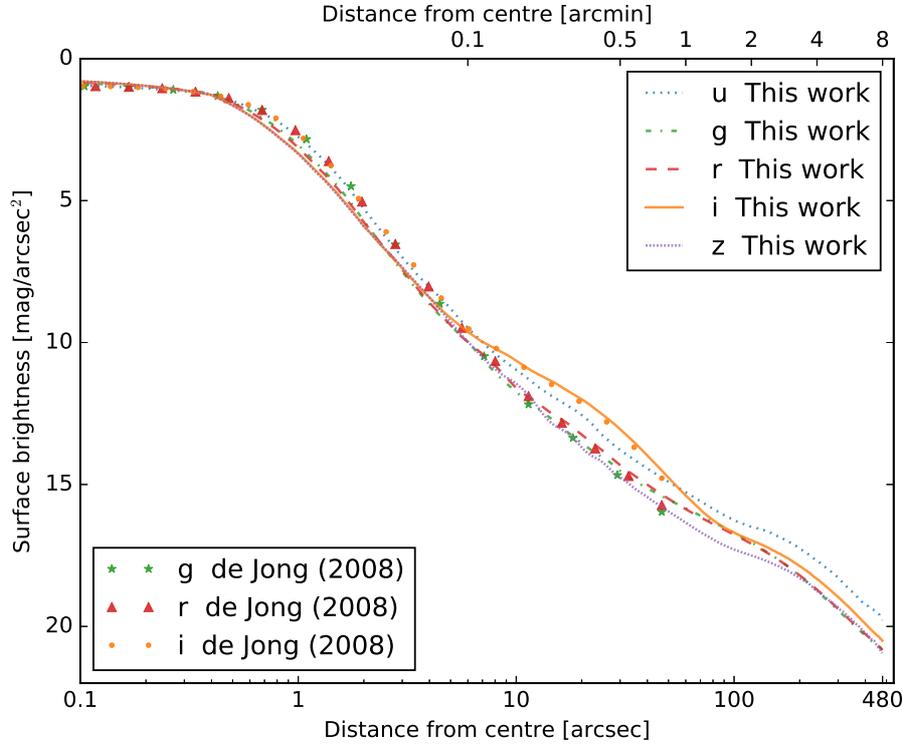


Figure 3.8: Circularly averaged surface brightness radial profiles of SDSS PSFs of this work compared with those obtained by de Jong (2008). The SDSS filters are line coded as explained in the legends. A remarkable feature is the bump in the *i* band that can be seen as an extended halo of light between 7 and 70 arcsec.

### 3.3.2 Scattered light subtraction

In this section, we show an example of how the PSF models can be used to remove the scattered light produced by the stars in the central part of the Coma Cluster field. We select this region because it is a classic example where studying intracluster light or low-surface brightness structures in general is impossible without an extended and well-characterized PSF.

We have removed the scattered light field produced by all the stars in the Gaia DR2 catalogue (Gaia Collaboration et al. 2018) with  $G_{\text{mag}} < 13$  mag. We only show this exercise using *g*, *r*, and *i* filters. These images are deeper compared with the other two filters (*u* and *z*) in the SDSS survey, and the effect of scattered light towards the faintest regions is therefore easier to

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

show. Note that in Section 3.2.1 USNO-B1 Catalog was considered because we found Gaia was incomplete among the brightest stars. In this particular field, the brightest star has  $G_{mag} = 7.0$  mag and it is part of the Gaia DR2 Catalogue. Therefore, we used the more accurate star positions provided by Gaia.

Once a list of stars is ready, the scattered light field is modelled hierarchically, that is, we start by fitting and subtracting the brightest star, then we go to the second brightest star, and so on. The process is repeated until all stars of a given magnitude are fitted and subtracted. The reason for doing this iterative process is because the light of very bright stars affect significantly the fitting and subtraction of other near and fainter stars. The fitting of each star requires two steps: first, to obtain an accurate centre of the star, and secondly, to perform the flux calibration of the PSF. The details of how the process works are as follow.

First, we run `NoiseChisel` and `Segment` on the image to get the object masks as well as the background value and the saturation level. As a first approximation, the centre of the stars to be modelled are obtained from the Gaia DR2 catalogue. We refined the the centre estimation by using `Imfit` (as explained in Section 3.2.3). The calibration in flux of the PSF model to each individual star is done through the radial profile of the star. To calibrate the PSF, we use the radial range between 0.1 times the saturation level and 3 times the value of the background level of the image (see Figure 3.9). In this region, for each pixel we compute the ratio between the radial profile of the star and the radial profile of the PSF. The exact factor  $F$  is computed by taking the  $3\sigma$  clipped median of all the pixels ratio.  $F$  is the factor by which the PSF model has to be scaled up to get the same flux level as the star. Once we have the centre position of the star and  $F$ , the star is completely modelled, we subtract the model from the original image. As we have said above, this process is done star by star, until the last star considered is modelled and subtracted.

For modelling the scattered light field, a good masking of the sources in the image is mandatory. Otherwise, the excess of flux from objects close to the stars that are being modelled could alter the measurement of their centre and flux. To minimize these problems, we perform the scattered light modelling in two iterations. The first one is the one that has been described above. Once we have removed the brightest stars from the image, we run again `NoiseChisel` and `Segment` over the image to obtain a better objects mask. By doing this, the masking process takes into account the diffraction spikes, internal reflections, and nearby objects that were not possible to obtain in the first masking. Finally, we repeat again the scattered light field fitting process taking into account the new object masks. Both the centre position and the scaling factor  $F$  are better determined with the improved masks in the second iteration. Note that this two-step methodology is the same as described in Section 3.2.1 for constructing the outer and intermediate part of the PSFs, but in this case for subtracting the stars.

In Figure 3.9 we show an example of how the PSF modelling process works for a real star. The radial profile of a real star in the  $r$  band SDSS Coma Cluster central field is plotted in solid blue line. The innermost region of the star is clearly saturated as can be seen from the flatness of the radial profile within  $R < 5$  arcsec. A small bump at  $R \sim 25$  arcsec that corresponds to the internal reflection of the star is also seen. The model PSF does not have this feature, because they were masked out during the second iteration of the PSF construction process. As our PSF model cannot model all the particular features of a given individual star, after the subtraction process, some residuals are left. They correspond mainly to the central region due to the saturated pixels and internal reflections. In the same way, the diffraction spikes of each

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

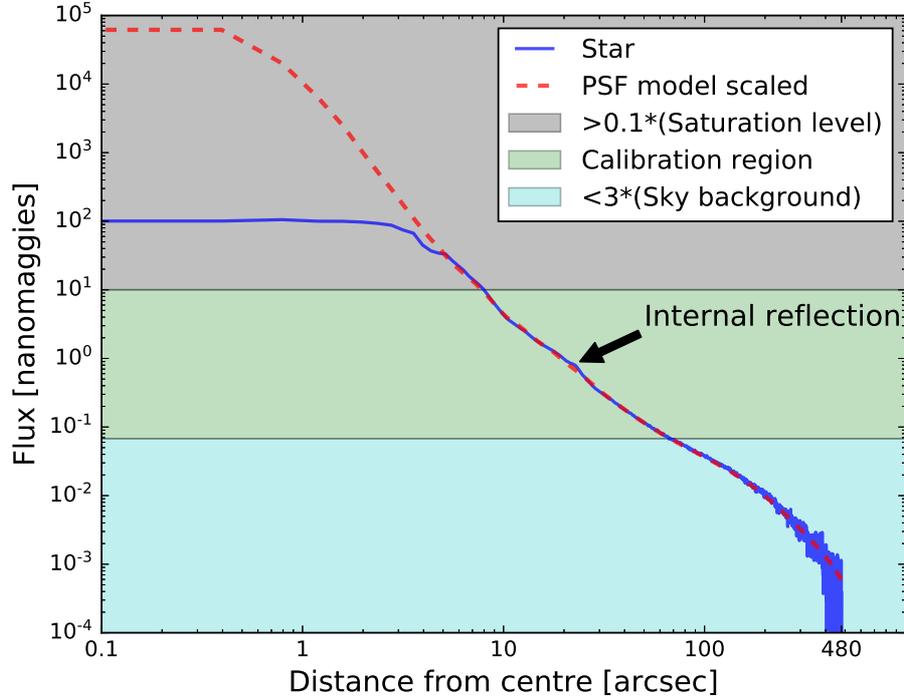


Figure 3.9: An example of PSF model fitting to a real star (R.A. (J2000) = 12:59:32.975, Dec (J2000) = +28:03:56.295, Gmag = 7.0 mag) in SDSS  $r$  band. Solid blue line is the radial profile of the star to be modelled. The grey region corresponds to the radial profile values which are 0.1 times above the saturation level of the image (saturation=100 nanomaggies). The light blue area represents the flux level which is 3 times below the intensity value of the background value of the image. The light green area corresponds to the region used to obtain the scaling factor between the PSF model and the real star. The dashed red line represents the radial profile of the PSF once it has been scaled up to the same flux level as the real star. See the text for more details.

star will remain as residuals since the PSF models do not have diffraction spikes. All of these effects can be seen in Figure 3.10, where we show a colour composite ( $g$ ,  $r$ , and  $i$ ) image of the central part of the Coma Cluster field in three different panels. Top panel shows the original image with the scattered light of near and bright stars contaminating the central parts of the galaxy cluster. In the middle panel, we show the scattered light field model of the brightest stars. In this particular case, we have modelled all stars brighter than 13 mag in the  $G$  band of Gaia

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

DR2 Catalogue. The bottom panel shows the stars subtracted image. After the removal of the scattered light field, the low-surface brightness features of the Coma Cluster and its intracluster light are ready to be explored.

### 3.4 Discussion and conclusions

We present extended SDSS PSFs (8 arcmin in radius) for all five SDSS filters. We describe the methodology followed in order to obtain these large PSFs and we show a particular example of how the PSFs can be used to remove the scattered light field produced by the brightest stars in the images. Our approach of deriving PSFs can be applied to other telescopes and instruments to obtain extended PSFs. This work shows that the PSFs have complex bidimensional structures as it can be seen in Figure 3.6. In that figure, we plot the radial profiles produced by de Jong (2008) with a maximum radius of  $\sim 46$  arcsec for the  $g$ ,  $r$ , and  $i$  SDSS filters together with the radial profiles of the PSFs of this work. The reader can see that the agreement between these two works is rather good. Compared to de Jong (2008), we enlarge the model of SDSS PSFs an order of magnitude in radius. In addition to this, we also provide their 2-dimensional structures.

The PSFs models presented in this work can be used to remove the scattered light of point-like sources, as we have described in Section 3.3.2. In addition to this, the PSFs can be also used to model galaxies and other astronomical objects with the goal of correcting the scattered light of the object itself (see e.g. Martínez-Lombilla et al. 2019). In fact, this is what it is done in next chapters (see Chapters 4 and 6, where the PSF is used to model and correct the point-like sources as well as the galaxies themselves). How the scattered light is affecting the outskirts of extended objects is crucial in low surface brightness science, and the only way of accounting for this effect is through having an extended and well characterization of the PSF of the telescope. The next generation of large and deep sky surveys such as LSST, Euclid mission, James Webb Space Telescope (JWST), and many others would require an exquisite characterization of their PSFs for studies involving low surface-brightness structures.

#### Note on reproducibility

Being able to reproduce results is in the core of science, and because of that, the construction of the PSFs have been developed in a reproducible way. This means that the construction of the PSFs can be done by anyone from scratch, just downloading the project publicly available at <https://gitlab.com/infantesainz/sdss-extended-psfs-paper/> or <https://doi.org/10.5281/zenodo.3524937>, and following the instructions described in the README.md file of that repository. Note that the subtraction of the scattered light described in Section 3.3.2 is not included, because it is not part of the PSF construction but a particular example on how to remove the scattered light produced by the stars in a given region of the sky. See Chapter 2 to know more details about the reproducibility problems and how Maneage is able to account for them.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

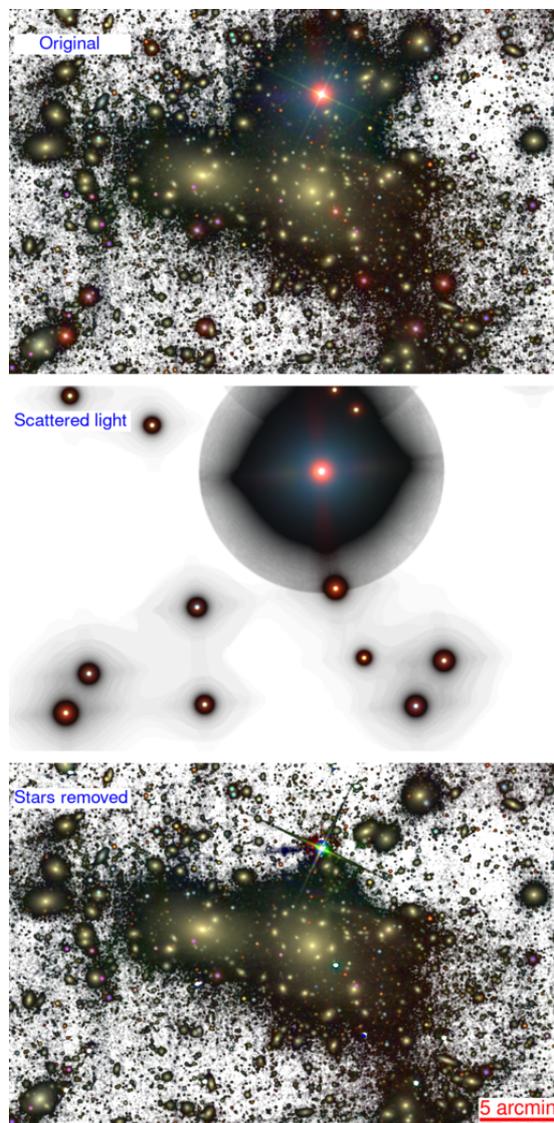


Figure 3.10: Colour composite image of the central region of the Coma Cluster as seen by SDSS (R.A. (J2000) = 12:59:48.658, Dec (J2000) = +27:58:49.443), using  $g$ ,  $r$ , and  $i$  filters. The background of the image (in grey) corresponds to the sum of  $g$ ,  $r$ , and  $i$  SDSS band images. The images have been rebinned to 2 arcsec per pixel and convoluted with a Gaussian kernel of 3 pixel width to enhance the low-surface brightness features of the data. Upper panel: SDSS original image. Note the enormous contamination produced by the brightest stars in the central part of the Coma Cluster. Middle panel: scattered light field of all stars brighter than 13 mag in the  $G$  band of Gaia DR2 Catalogue. Bottom panel: scattered light field subtracted image.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 4

## The galaxy lacking dark matter NGC 1052-DF4

In this chapter, we explore the galaxy NGC 1052-DF4 in detail. This galaxy has recently received a huge amount of attention as it was claimed to be devoid of dark matter. If confirmed, this would be a very strange object given the current model of galaxy formation and evolution (see Section 1.1.1). By obtaining very deep imaging data from several facilities and through a careful analysis, we have been able to detect very faint tidal stellar structure. This reveals that the galaxy NGC 1052-DF4 is interacting with another nearby galaxy, thus explaining its lack of dark matter. Here, the extreme importance of ultra-deep imaging is clearly seen. This chapter corresponds to the already published paper “*The Galaxy ‘Missing Dark Matter’ NGC 1052-DF4 is Undergoing Tidal Disruption*” (Montes, Infante-Sainz et al. 2020, ApJ, 904, 114), of which I am the second author.

### 4.1 Introduction

Dark matter (DM) is a key constituent in current models of galaxy formation and evolution. In fact, without the presence of DM, the primordial gas would lack enough gravitational pull to start collapsing and forming new galaxies. Because this DM is thought to interact mostly gravitationally, its presence can only be inferred from its effect on the visible matter (e.g., Freeman 1970; Rubin & Ford 1970).

Although transient, DM-free, stellar aggregations (also called tidal dwarf galaxies) have been known for some time (see e.g., Duc et al. 2000; Lelli et al. 2015; Ploekinger et al. 2015, 2018), the existence of long-lived non-isolated galaxies lacking DM would be a challenge to the current galaxy formation paradigm. For this reason, the recent report of two old potentially stable galaxies with low (or null) DM content, [KKS2000]04 (also known as NGC 1052-DF2) and NGC 1052-DF4 (van Dokkum et al. 2018, 2019, hereafter vD19), has triggered an intense debate about the nature of these objects. While in the case of [KKS2000]04 (NGC 1052-DF2) its unusual properties can be simply resolved if the object is located at 13 Mpc (instead of the originally assumed 20 Mpc, see e.g., Trujillo et al. 2019), the low velocity dispersion of the globular clusters (GCs) of NGC 1052-DF4 is compatible with the absence of DM, even at a

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

closer distance (13.5 Mpc, Monelli & Trujillo 2019).

An appealing alternative to explain the low DM content of NGC 1052-DF4 is that this galaxy is undergoing tidal stripping (e.g., Ogiya 2018; Lokas 2020; Nusser 2020; Yang et al. 2020; Jackson et al. 2021; Macciò et al. 2021). In fact, NGC 1052-DF4 is very close both in velocity ( $\Delta v \sim 180 \text{ km s}^{-1}$ ) and in spatial projection ( $\Delta R \sim 4'$ ) to the galaxy NGC 1035. If NGC 1052-DF4 is undergoing tidal disruption, this could explain the apparent lack of DM as this is preferentially stripped over the stars of the galaxy or its GC system, due to its more extended spatial distribution (Smith et al. 2013, 2016).

In this work, we explore the tidal stripping scenario and whether it can explain the “lack” of DM observed in NGC 1052-DF4. We explore this scenario in two ways. First, we study the spatial distribution of the GCs in NGC 1052-DF4. If NGC 1052-DF4 is being tidally stripped by NGC 1035, their stripped GCs would be placed along the orbit traced by the satellite. Therefore, exploring how the spatial distribution of GCs in NGC 1052-DF4 is can give us clues about the present state of the galaxy.

Second, we obtained very deep imaging of NGC 1052-DF4 in order to search for any signs of tidal interactions in the stellar light around this galaxy (see also Müller et al. 2019). Tidal effects in dwarf galaxies have been studied in simulations (e.g., Johnston et al. 1996; Read et al. 2006; Peñarrubia et al. 2008, 2009), and they have found that if, for example, tidal tails are present in the outer parts of this galaxy, their shape and direction can provide information of the current orbit of the satellite (e.g., Johnston et al. 2002; Klimentowski et al. 2009).

In this work, we will demonstrate that both the GCs and the stellar distribution of NGC 1052-DF4 are in agreement with the tidal stripping scenario. We will also discuss how this stripping scenario explains the apparent “lack” of DM, making this galaxy perfectly compatible with the current galaxy formation paradigm. All magnitudes in this work are given in the AB magnitude system.

## 4.2 Data

The data used in this work come from three different facilities: the *Hubble Space Telescope* (*HST*) and two ground-based telescopes: the Gran Telescopio Canarias (GTC) and the IAC80 telescope. We describe the details of each observation in what follows.

### 4.2.1 HST imaging

NGC 1052-DF4 was observed with the *HST* ACS Wide-Field Channel (WFC) as part of the programmes GO-14644 and GO-15695 (PI: van Dokkum). We retrieved the data from the MAST archive.<sup>1</sup> The observations consist of a total of four orbits in the F606W band and eight in the F814W band. The total exposure time is 8240 s for the F606W and 16 760 s in F814W.

To obtain the final coadded images, we follow the approach detailed below. The individual frames were precalibrated in a standard way, i.e., bias- and dark-current-subtracted, flat-fielded, and charge transfer efficiency (CTE)-corrected producing `flt` files. A preliminary luminance coadd using the images of both filters was made using `Astrodrizzle` (Gonzaga et al. 2012) with a median constant sky subtraction. We used this first coadd to create a deep master mask using `NoiseChisel` (Akhlaghi & Ichikawa 2015; Akhlaghi 2019), later applied to each individual

<sup>1</sup><https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

frame in order to derive a more robust sky background. This deep master mask was manually expanded to include any remaining light from NGC 1052-DF4 and other large sources in the field of view, and then drizzled to the field of view of each individual exposure. For each one of them, we subtract a robust median sky-background level (using bootstrapping to avoid the effect of outliers and cosmic rays) from masked frames (see [Borlaff et al. 2019](#), for further details).

At this point, the rough astrometry provided by the headers of the `f1c` files was refined using the `TweakReg` utility. After this, we derived the final coadd for the images with `Astrodrizzle` into astrometrically corrected drizzled images with the F814W and F606W filters. The relevant drizzling parameters used are `pixfrac=1`, `kernel=lanzcos3`, and `combine type=iminmed`. Note that the `combine type` option, `iminmed`, is generally the same as the median, except when the median is significantly higher than the minimum good pixel value. In those cases, `iminmed` will choose the minimum value. The final coadd of the images still shows a significant gradient across the detector, consistent in both filters, lower close to the gap between the CCDs. To mitigate this problem, we used `NoiseChisel` to generate a final sky-background model for the coadded images, following the optimized methodology for the detection of the low surface brightness wings of extended sources.<sup>2</sup> The optimal configuration was set at `tilesize=300,300`, `interpnunmb=3`. As a final step, after this background model has been subtracted, we readjust the median sky-background level by a constant once again using the deep master mask.

We take advantage of these high-resolution *HST*/ACS observations, together with the photometry from the GTC HiPERCAM data (see Section 4.2.2.1) to characterize the GC population of NGC 1052-DF4.

#### 4.2.2 Deep ground-based imaging

The study of low surface brightness light in deep images is extremely challenging. It is not enough to have very deep data, but a careful data reduction is crucial. In fact, it is common to find “dips” around the brightest extended objects in very deep surveys (see, e.g., CFHTLS and the HSC-SSP DR1; [Goranova et al. 2009](#); [Aihara et al. 2018a](#)) due to aggressive sky subtraction. In addition, the study of diffuse and extended objects is susceptible to biases due to flat-field inaccuracies and the scattered light from bright stars, to name a few.

NGC 1052-DF4 was observed with the GTC and the IAC80 telescope. Most of the reduction steps are common for both datasets and, therefore, described in the following. Specifics for each of the telescopes are discussed later. The aim of this reduction process is to correct all systematics effects introduced by the instrument/telescope as well as to calibrate the data. The whole reduction process was done within a controlled and enclosed environment as described in [Akhlaghi, Infante-Sainz et al. \(2020\)](#).

Deep observations require an observing strategy that produces a background as flat as possible around the galaxy target. Another key step is the dithering pattern. For this work, we followed the dithering strategy in [Trujillo & Fliri \(2016\)](#). In short, this consists of a dithering pattern with large steps (typically the size of the source under investigation) and whenever possible rotation of the camera. This ensures we achieve the flat background required for our goals.

<sup>2</sup>Gnuastro 2.3 Tutorial - Detecting large extended targets: [https://www.gnu.org/s/gnuastro/manual/html\\_node/Detecting-large-extended-targets.html](https://www.gnu.org/s/gnuastro/manual/html_node/Detecting-large-extended-targets.html)

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Bias frames were taken the same night of the observations as part of the different observing programs. The master bias frames were created as the sigma-clipped ( $3\sigma$ ) mean and subtracted from the original raw images.

Dome flats are not suitable for our goals due to inhomogeneities in the dome illumination that can result in gradients across the image. Consequently, the flat-field frames derived here are based on the same science exposures obtained for this work. To solve this, the flat-field correction was performed in two steps. As the bias-corrected CCD frames present steep gradients,<sup>3</sup> we first stack each bias-corrected image after normalizing them by its  $3\sigma$ -clipped mean. This produces a rough master flat-field frame that is used to correct the CCD frame. These corrected CCD images were used to build an object mask using `NoiseChisel`'s detection maps. Again, for each CCD, the masked and normalized images are combined to create the final, and more accurate, master flat-field frames for each band.

We, then, performed the astrometric calibration of the different frames. We used the `Astrometry` software (Lang et al. 2010) to produce an approximate astrometric solution, later refining it with `SCAMP` (Bertin 2006). Later, we performed the sky-background correction by masking all signal pixels detected with `NoiseChisel` for each frame. The remaining pixels are used to compute a first-degree 2D polynomial fit that is then subtracted from the entire frame. This ensures the correction of any remaining gradient while preserving local low surface brightness structures.

Once all frames were corrected of all systematic effects and in a common astrometric solution, they were resampled into a common grid using `SWarp` (Bertin 2010) and stacked using a  $3\sigma$ -clipped mean into a final image, one per filter. The photometric calibration of both sets of images, HiPERCAM and IAC80, was done using SDSS DR12 (Alam et al. 2015).

#### 4.2.2.1 GTC HiPERCAM images

Deep multiband imaging with GTC was requested through Director Discretionary Time to observe NGC 1052-DF4 with HiPERCAM. HiPERCAM (Dhillon et al. 2018) is a quintuple-beam, high-speed astronomical imager able to obtain images of celestial objects in five different filters ( $u, g, r, i, z$ ) simultaneously. The image area of each of the five CCDs is  $2048 \times 1024$  pixels ( $2.7' \times 1.4'$ ; 1 pixel =  $0.08''$ ) divided into four channels of  $1024 \times 512$  pixels. NGC 1052-DF4 was observed between 2019 September 4 and 8. During these observations, there was a problem with the electronics of the  $i$ -band CCD, making it impossible to use this band. The average seeing of the images was  $1.1''$ .

For HiPERCAM, the standard calibration of each individual frame, i.e., bias and flat-field correction, was performed in each channel of the CCD. After the flat-field correction, we assembled each set of four channels into a single image per filter. To do this, a fine gain correction was done by computing the average value along the adjacent pixels for the different channels.

We obtained  $\sim 55$  exposures per filter using a constant exposure time of 106 s per frame. The different exposures that went into the final images were visually inspected and rejected if rejecting those with low-quality or strong gradients were still present. The final exposure times in the innermost  $R < 1'$  of the images are 1.6, 1.7, 1.5, and 1.7 hr for the  $u, g, r,$  and  $z$  bands, respectively.

<sup>3</sup>This is due to illumination inhomogeneities as well as fringing, particularly for the GTC HiPERCAM  $z$  band.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

Unfortunately, the modest field of view of HiPERCAM is similar to the size of the galaxy (with a diameter of  $\sim 1.5'$ ). This precludes a reliable sky subtraction and, therefore, the analysis of the extended low surface brightness emission around NGC 1052-DF4. Consequently, even though the GTC data are very deep, we only used them to study the GC system around NGC 1052-DF4 (which, being point-like sources, are not significantly affected by the sky subtraction). The point-like source limiting magnitudes ( $5\sigma$  in  $2''$  diameter apertures) are  $m_u = 27.11 \pm 0.06$ ,  $m_g = 26.98 \pm 0.07$ ,  $m_r = 26.44 \pm 0.07$ , and  $m_z = 24.81 \pm 0.10$  mag.

#### 4.2.2.2 IAC80

In order to explore the distribution of light around NGC 1052-DF4 and NGC 1035, we use the IAC80 telescope. The IAC80 telescope is an 82 cm telescope located at the Teide Observatory in Tenerife. The camera currently installed at its Cassegrain primary focus, and used here, is CAMELOT2 (the Spanish acronym of the “Teide Observatory Light Improved Camera”). It contains an E2V 2048  $\times$  2048 back-illuminated chip with a pixel scale of  $0''.336$  per pixel corresponding to a  $10'.4 \times 10'.4$  field of view. This field of view is enough to get a reliable sky estimation around NGC 1052-DF4. NGC 1052-DF4 was observed with CAMELOT2 between 2019 December 19 and 2020 February 20 in the  $g$ ,  $r$ , and  $i$  bands. The final exposure times are 22.33, 23.33, and 15.66 hr for the  $g$ ,  $r$ , and  $i$  bands respectively. The surface brightness limits of these images are  $\mu_g = 29.8$ ,  $\mu_r = 29.3$ , and  $\mu_i = 28.8$  mag arcsec $^{-2}$  ( $3\sigma$ ;  $10'' \times 10''$  boxes). They were estimated by measuring the rms of the final masked images by randomly placing 3000 boxes of  $10'' \times 10''$  across the images.

Figure 4.1 shows the  $13.3' \times 13.3'$  region around NGC 1052-DF4 taken with the IAC80 telescope. In the figure, we have also plotted the field of view of the IAC80/CAMELOT2, GTC/HiPERCAM, and HSC/ACS cameras for ease of comparison.

#### 4.2.2.3 PSF modeling and bright star subtraction of the IAC80 images

A key aspect in low surface brightness science is to remove the scattered light field produced by the brightest objects of the images. To do this, an accurate characterization of the point-spread function (PSF) of the images is required. As we are only studying the distribution of light around NGC 1052-DF4 in the IAC80 images, this was done only for these images.

To build and correct for the extended ( $R = 7.5'$ ) PSF models of the IAC80 telescope, we followed the methodology outlined in Román et al. (2020) and Infante-Sainz et al. (2020). The outermost regions of the PSFs were obtained through dedicated observations of bright stars with magnitudes between 3.2 and 4.8 mag, in the  $V$  band. The exposure times vary with the star magnitude and the observed band between 7 and 22 s. The observations of the bright stars were conducted with a large dithering pattern in order to cover the entire field of view of the camera. Individual frames were masked to avoid the contamination of background sources and finally combined to create the outer parts of the PSF models.

Because we only use bright stars to build this model, the final PSF models are saturated in the center. However, as the goal is to eliminate the contribution of the outer parts of bright stars surrounding NGC 1052-DF4, we did not derive the inner part of the PSF model. We fit and subtracted eight bright stars in the field of view of the camera that could affect our photometry. To subtract the stars in our images, we follow the steps outlined in Román et al. (2020).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

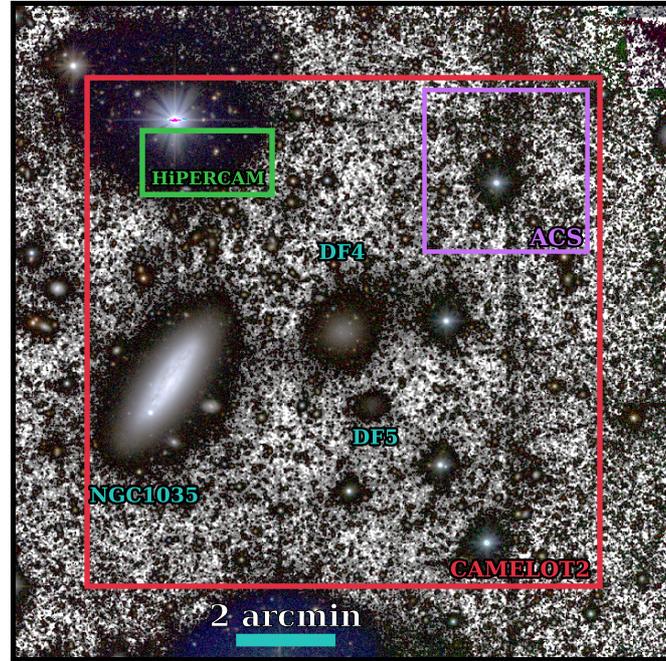


Figure 4.1: We show a  $13.3' \times 13.3''$  region around NGC 1052-DF4 imaged with the IAC80 telescope. The figure is a composite of an RGB color image using the  $g$ ,  $r$ , and  $i$  IAC80 bands and a black and white  $g+r+i$  image for the background. The figure highlights the main galaxies in the field of view: NGC 1052-DF4, and their neighbors NGC 1035 and NGC 1052-DF5. The fields of view of the different cameras used in this work are overplotted. The surface brightness limit of this composite image is equivalent to  $\mu_r \sim 29.9 \text{ mag arcsec}^{-2}$  ( $3\sigma$ ;  $10'' \times 10''$  boxes). North is up, east is left.

### 4.3 The spatial distribution of the globular cluster system

Globular clusters (GCs) are thought to form in the episodes of intense star formation that shaped galaxies (see Brodie & Strader 2006, and references therein). Their typical masses ( $10^4\text{--}6M_\odot$ ) and compact sizes (half-light radii of a few parsecs; Harris 1996) make them easily observable in external galaxies and, therefore, good tracers of the properties of their host system even at large radial distances, unlike stellar dynamics studies that are limited to the inner regions of galaxies.

The spatial distribution of the GC system of a galaxy might reveal clues about its present

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

state. For instance, in an accretion event, the stripped material is placed along the orbit of the parent satellite galaxy, forming tidal streams (e.g., Toomre & Toomre 1972; Johnston et al. 1996). Therefore, during this process, the stripped GCs will also align along the orbit (e.g., Mackey et al. 2010; Hughes et al. 2019) and, consequently, their observed spatial distribution can indicate the direction of this orbit.

vD19 identified seven GCs associated with NGC 1052-DF4 by confirming them spectroscopically after a first selection based on their *HST* F606W–F814W color. In this section, we use our deep HiPERCAM imaging to identify additional GC candidates and determine the spatial distribution of the GCs of NGC 1052-DF4. The goal is to explore whether the GCs align through a particular direction, which could indicate that NGC 1052-DF4 is undergoing tidal stripping or, on the contrary, the GCs are more spherically distributed around NGC 1052-DF4 as it would be expected if the galaxy is in isolation.

#### 4.3.1 GC selection

The task of identifying the GCs in galaxies is nontrivial. The uncertain level of foreground and background contamination limits its utility in characterizing the overall GC system of the host galaxy with a large degree of confidence. A common technique for identifying GCs in galaxy systems is to use their colors. However, recent works have shown the increased effectiveness of identifying GCs using more information from their spectral energy distributions (e.g., Montes et al. 2014; Muñoz et al. 2014).

For this reason, in order to identify more GC candidates in NGC 1052-DF4, we use the  $(u-r)-(r-z)$  color-color diagram, as described in Taylor et al. (2017). In this case, the effectiveness of this selection criteria is based on the inclusion of the  $u$ -band, which allows a cleaner separation of metal-poor GCs from the foreground stars of our Galaxy. The presence of hot horizontal branch stars, contributing to the flux in the  $u$ -band, in these metal-poor GCs will make them bluer in  $u-r$  than foreground stars for a given  $r-z$ .

For a preliminary selection, we took advantage of the high spatial resolution of *HST* to preselect potential GCs based on their morphology (FWHM and ellipticity) and F606W magnitude,  $m_{F606W}$ . This selection was modeled upon the properties of the confirmed GCs in vD19. First, we run SExtractor (Bertin & Arnouts 1996) on the *HST* images in dual mode using F814W as the detection image. We also ran in the HiPERCAM images of NGC 1052-DF4 using in this case the  $r$  image as our detection image. Once the catalogs for the different instruments, *HST* and HiPERCAM, are obtained, we matched them based on their position in the sky.

The initial GC selection consists of objects that simultaneously fulfill the following properties:  $0.1'' < \text{FWHM} < 0.2''$ ,  $0 < \text{ellipticity} < 0.2$  and  $20 < m_{F606W} < 25$  mag. This observed magnitude range corresponds to the absolute magnitude range  $-10.65 < M_{F606W} < -5.65$  if the galaxy is at 13.5 Mpc or  $-11.5 < M_{F606W} < -6.5$  if the galaxy is at 20 Mpc. The aim of this selection is to find GCs that are similar to the confirmed GCs while allowing for sources that might have been missing from the spectroscopic-based sample. Figure 4.2 shows the histograms of the FWHM, ellipticity, and  $m_{F606W}$  of all the sources detected with SExtractor (gray), the confirmed GCs in vD19 (magenta), and the GC candidates based on our preliminary selection (mint green). Twenty objects fall within this preselection, including the seven objects in vD19.

The next step is to narrow down the selection based on the colors of the GCs. Figure 4.3 shows the  $(u-r)-(r-z)$  color-color diagram for the preselected GC candidates (mint green)

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

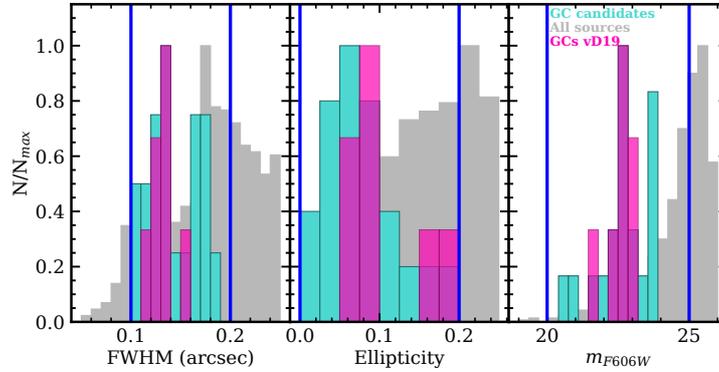


Figure 4.2: Histograms of FWHM (left), ellipticity (middle), and F606W magnitude  $m_{F606W}$  (right) for all the sources simultaneously detected sources in the *HST* and HiPERCAM images (gray). The magenta histograms show the distribution of the spectroscopically confirmed GCs in vD19. The blue vertical lines mark the range in each parameter of the initial selection. The mint green histograms show the sources that simultaneously satisfy the conditions of FWHM, ellipticity, and magnitude enclosed by the vertical blue lines.

from our preliminary selection. We highlighted the seven GCs in vD19 in magenta. As seen in Figure 4.3, the confirmed GCs fill a very narrow space in the color-color diagram, and therefore, we further narrowed down the initial selection to objects with  $1.5 < u-r < 2.2$  and  $0.55 < r-z < 0.75$ , shown as the red box in Figure 4.3. These ranges in color are based on the  $\pm 3\sigma$  around the median colors of the confirmed GCs (magenta), with  $\sigma$  being their dispersion in color. Using this color selection, our final sample consists of a total of 11 GCs, 4 new candidates in addition to the confirmed GCs in vD19. The radius containing half of the GCs is  $\sim 155''$ , 9 times larger than the  $R_e$  of the stellar light ( $17''$ ; vD19).

The new GC candidates reported here are located to the West of NGC 1052-DF4, the side facing away from NGC 1035, which is less likely to be contaminated by the GC system of the disk galaxy. In addition, we expect very low GC contamination from this low-mass disk galaxy. Using the Milky Way as a reference, and the correlation between the size of the GC system and the stellar mass of the host (Forbes 2017), NGC 1035 is expected to have half of their GCs contained within a radius of  $< 4$  kpc, or  $61''/42''$  (Hudson & Robison 2018) at 13.5/20 Mpc, respectively. For reference, the projected distance between the centers of the galaxies is  $222''$ .

The coordinates and magnitudes of all the objects are listed in Table 4.1. The magnitudes are corrected by the extinction of our Galaxy:  $A_v = 0.11$ ,  $A_g = 0.08$ ,  $A_r = 0.06$ ,  $A_z = 0.03$ ,  $A_{F606W} = 0.06$ , and  $A_{F814W} = 0.04$ .

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

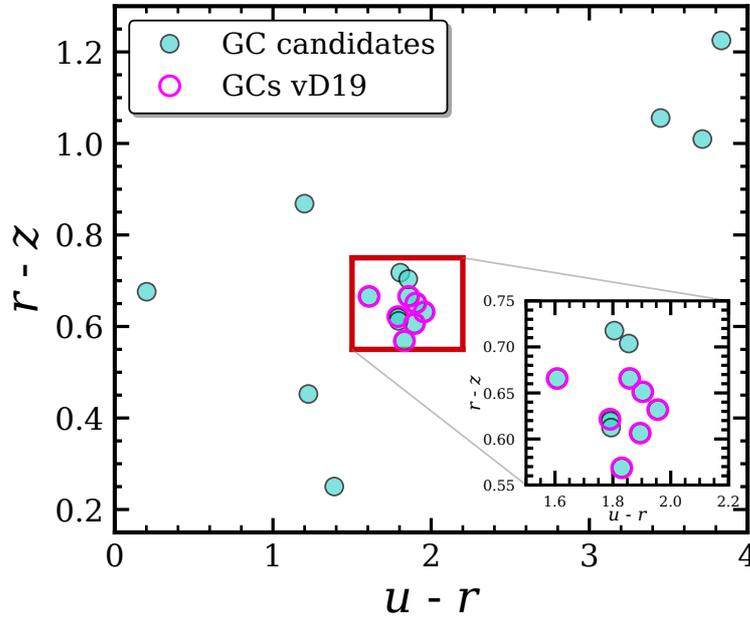


Figure 4.3: The  $(u-r)-(r-z)$  color-color diagram of the initial sample of candidate GCs (mint green). The spectroscopically confirmed GCs of vD19 are highlighted in magenta. The red box indicates the color-color region we have selected to create our final sample of GC candidates. The inset shows a zoom-in into the red box for ease of viewing.

#### 4.3.2 The distribution of GCs

In this section, we address the spatial distribution of the GCs in NGC 1052-DF4 in order to obtain clues about the interaction state of the galaxy. As mentioned before, in an accretion event, the stripped material (stars and GCs) will be deposited along the orbit of the parent satellite. In Figure 4.4, we show an RGB image using the HiPERCAM  $g$ ,  $r$ , and  $z$  bands. The highlighted sources are the confirmed GCs from vD19 (magenta) and the new candidates obtained in the previous section (mint green). The GCs tend to align in a particular axis along with the galaxy.

In order to derive this axis, we use `scipy.odr` to perform an orthogonal distance regression to all the GCs, confirmed and candidates. We use the orthogonal distance regression as we are interested in calculating the orthogonal distance to the axis instead of predicting the value of  $Y$  for any given  $X$ . The range of possible axes, i.e., directions of the orbit, will be given by the slope of the regression and its error. This is indicated by the blue-shaded region in Figure 4.4. This region encompasses 68% of the axes compatible with the GC distribution. One of the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

| Name              | R.A.  | Decl.        | $m_u$<br>(mag) | $m_g$<br>(mag) | $m_r$<br>(mag) | $m_z$<br>(mag) | $m_{F606W}$<br>(mag) | $m_{F814W}$<br>(mag) |
|-------------------|---|--------------|----------------|----------------|----------------|----------------|----------------------|----------------------|
| C-1 <sup>a</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 14.52 <sup>s</sup> | -8°08'27.56" | 25.69 ± 0.19   | 24.35 ± 0.03   | 23.88 ± 0.04   | 23.17 ± 0.08   | 23.78 ± 0.02         | 23.46 ± 0.02         |
| C-2 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 16.96 <sup>s</sup> | -8°08'04.35" | 25.00 ± 0.10   | 23.49 ± 0.01   | 23.04 ± 0.02   | 22.41 ± 0.04   | 22.93 ± 0.01         | 22.55 ± 0.01         |
| C-3 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 18.22 <sup>s</sup> | -8°07'23.17" | 24.40 ± 0.06   | 22.97 ± 0.01   | 22.50 ± 0.01   | 21.84 ± 0.02   | 22.45 ± 0.01         | 22.06 ± 0.01         |
| C-4 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 17.23 <sup>s</sup> | -8°07'05.75" | 24.47 ± 0.06   | 23.12 ± 0.01   | 22.68 ± 0.01   | 22.06 ± 0.03   | 22.63 ± 0.01         | 22.24 ± 0.01         |
| C-5 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 15.24 <sup>s</sup> | -8°06'57.94" | 23.74 ± 0.03   | 22.34 ± 0.01   | 21.91 ± 0.01   | 21.34 ± 0.02   | 21.70 ± 0.01         | 21.25 ± 0.01         |
| C-6 <sup>c</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 14.39 <sup>s</sup> | -8°06'59.81" | 23.74 ± 0.03   | 22.38 ± 0.01   | 21.95 ± 0.01   | 21.33 ± 0.01   | 21.87 ± 0.01         | 21.48 ± 0.01         |
| C-7 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 15.21 <sup>s</sup> | -8°06'52.24" | 24.88 ± 0.09   | 23.45 ± 0.01   | 22.99 ± 0.02   | 22.38 ± 0.04   | 22.88 ± 0.01         | 22.40 ± 0.01         |
| C-8               | 2 <sup>h</sup> 39 <sup>m</sup> 16.34 <sup>s</sup> | -8°06'44.37" | 25.53 ± 0.16   | 24.24 ± 0.03   | 23.73 ± 0.03   | 23.12 ± 0.08   | 23.66 ± 0.02         | 23.25 ± 0.01         |
| C-9 <sup>b</sup>  | 2 <sup>h</sup> 39 <sup>m</sup> 12.52 <sup>s</sup> | -8°06'40.50" | 24.38 ± 0.06   | 22.99 ± 0.01   | 22.53 ± 0.01   | 21.85 ± 0.02   | 22.46 ± 0.01         | 22.04 ± 0.01         |
| C-10              | 2 <sup>h</sup> 39 <sup>m</sup> 10.55 <sup>s</sup> | -8°05'45.92" | 24.53 ± 0.06   | 23.23 ± 0.01   | 22.67 ± 0.01   | 21.97 ± 0.03   | 22.67 ± 0.01         | 22.27 ± 0.01         |
| C-11 <sup>b</sup> | 2 <sup>h</sup> 39 <sup>m</sup> 16.79 <sup>s</sup> | -8°06'15.91" | 22.00 ± 0.01   | 21.12 ± 0.01   | 20.39 ± 0.01   | 19.73 ± 0.01   | 22.35 ± 0.01         | 21.92 ± 0.01         |

Table 4.1: Confirmed and Candidate GCs of NGC 1052-DF4.

<sup>a</sup> This candidate might be associated to NGC 1052-DF5.

<sup>b</sup> GC spectroscopically confirmed in [vD19](#).

<sup>c</sup> This candidate has been identified as a star in [Shen et al. \(2020\)](#).

|                                       | Angle ± Error |
|---------------------------------------|---------------|
| All the GCs in this work              | 58 ± 12       |
| Removing DF5's GC                     | 46 ± 9        |
| Confirmed GCs in <a href="#">vD19</a> | 50 ± 18       |

Table 4.2: Angles and the Errors of Their Axes Defined by the Spatial Distribution of the GCs of NGC 1052-DF4. Note. These angles are defined from the X-axis counterclockwise.

GCs, marked by a dashed circle, might be associated with NGC 1052-DF5 given its proximity to this galaxy and, consequently, might be biasing the computed axis. Taking into account this consideration, we computed the range of possible axes, but this time removing this GC, shown by the region enclosed by the dashed blue line. These two regions are compatible given the small number of sources. They are also compatible with the range of axes calculated if only using the confirmed GCs in [vD19](#). The angles defined by these axes and their errors, derived from the slopes of the orthogonal distance regressions, can be found in [Table 4.2](#).<sup>4</sup>

Consequently, the nonisotropic distribution of GCs around NGC 1052-DF4 is suggesting that it might be experiencing tidal interaction with a neighboring galaxy.

#### 4.4 Tidal stripping of the stellar distribution

While the study of the spatial distribution of the GCs in NGC 1052-DF4 can provide insights into whether the galaxy is undergoing tidal stripping, the strongest evidence is contained in the outermost part of the object. In fact, in the case of a tidal stripping event, the stellar distribution of NGC 1052-DF4 is expected to be distributed following a characteristic S shape (e.g., [Johnston et al. 2002](#); [Klimontowski et al. 2009](#)). We explore this using the IAC80 ultra-deep imaging.

<sup>4</sup>Removing C-6, identified as a star in [Shen et al. \(2020\)](#), does not change the results.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|  |                            |
|--|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA        | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA               | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguilar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

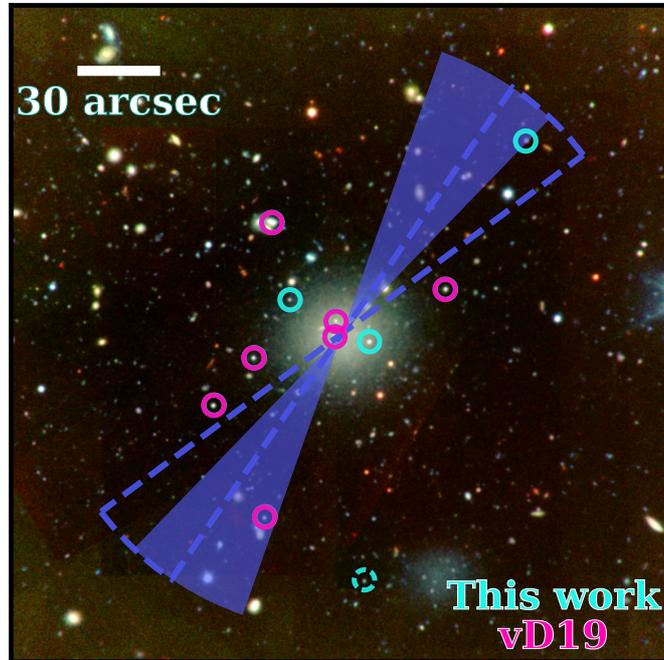


Figure 4.4: HiPERCAM postage stamp of  $240'' \times 240''$  around NGC 1052-DF4. The mint green circles highlight the candidate GCs found in this work while the magenta circles are the GCs found in vD19. The dashed mint circle at the bottom points to a GC probably associated with NGC 1052-DF5. The blue area indicates the range of directions of the orbit of NGC 1052-DF4 defined by the spatial distribution of the GCs. The area enclosed by the dashed blue line is the same but removing the GC probably associated with NGC 1052-DF5 (dashed mint circle). In this image,  $30''$  corresponds to a distance of  $\sim 2$  kpc at a distance of 13.5 Mpc and 2.9 kpc at 20 Mpc.

#### 4.4.1 Removal of NGC 1035

Figure 4.1 already shows that the presence of an excess of light around NGC 1052-DF4 is evident. This excess is centered around the galaxy and has the characteristic S shape of an object undergoing tidal stripping (e.g., Johnston et al. 2002). In order to study how this light is distributed, it is crucial to remove possible contaminants that can bias our results. In Section 4.2.2.3, we have already modeled and subtracted the brightest stars in the field of view, including the nearest star to the west of NGC 1052-DF4. However, even though NGC 1035 is  $\sim 4'$  away, the faint

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

halo of this galaxy can contaminate the light of NGC 1052-DF4 in very deep observations.

In order to eliminate this source of contamination, we modeled the galaxy using `ellipse` and `bmodel` in IRAF. While `ellipse` fits elliptical isophotes to the image, `bmodel` uses its output to create a 2D model of the galaxy. The advantage is that the ellipticity and position angle can vary at different radius, achieving a more authentic model of the galaxy and minimizing the residuals.

Prior to using `ellipse`, we need to carefully mask all foreground and background sources to reduce contamination that can affect the fitting procedure. As a single `SExtractor` setup for the detection and masking in deep observations is appropriate, we used a two-step approach: a “hot+cold” mode (e.g., Rix et al. 2004; Montes & Trujillo 2014). The “cold” mode will mask extended sources, while the “hot” mode is optimized to detect the more compact and faint sources that are embedded within the galaxy light.

`SExtractor` was run in a deep combined  $g+r+i$  image, and both masks were built from the generated segmentation maps. In the case of the “hot” mode, we unsharp-masked the original image in order to facilitate the masking of the dusty lanes of the galaxy. To create this unsharp-masked image, we convolved the image with a box filter with a side of 6 pixels and then subtracted it from the original. The “cold” mask was further expanded by 6 pixels while the “hot” was expanded by 3 pixels. Both masks were combined to create the final mask for our IAC80 images, unmasking NGC 1035 on the “cold” mask. This guarantees that the dust lanes and other sources within the galaxy are masked but not the diffuse stellar light of NGC 1035. The final mask was visually inspected to manually mask any remaining light that was missed by the process described above.

We run `ellipse` in the deep  $g+r+i$  image in two steps. In the first run, all of the parameters were allowed to vary freely, while in the second run, we fixed the center of the galaxy to the median centers returned in the first run. The geometry of the ellipses obtained in the fitting of the deep image was then used to measure the individual  $g$ ,  $r$ , and  $i$  images and create a model of the galaxy using `bmodel`.<sup>5</sup> The fit performed by `ellipse` reaches down to  $\sim 3'$  from the center of NGC 1035.

Figure 4.5 shows the result of subtracting both the light of NGC 1035 and the nearest bright star toward the west of the image. The S shape of the light around NGC 1052-DF4 becomes even more conspicuous after subtracting possible contaminants.

#### 4.4.2 Surface brightness radial profiles

The goal of this work is to investigate whether NGC 1052-DF4 is interacting with NGC 1035 and if that could explain the lower content of DM in this ultradiffuse galaxy. The presence of tidal tails and other asymmetries that indicate interactions will appear as an excess of light at a large radius and deviations from the morphology of the inner parts of the galaxy. To that end, we derived the radial profiles for the IAC80 images using the software `ellipse` in IRAF. `ellipse` fits elliptical isophotes to the 2D images of galaxies using the method described in Jedrzejewski (1987). It provides the median intensity, ellipticity, and position angle for each of the fitted isophotes.

We used the same methodology as in the previous section to build the mask for fitting NGC 1052-DF4. We run `SExtractor` in the deep  $g+r+i$  image and obtain the segmentation

<sup>5</sup>The results of the fit are consistent if we model each band separately.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

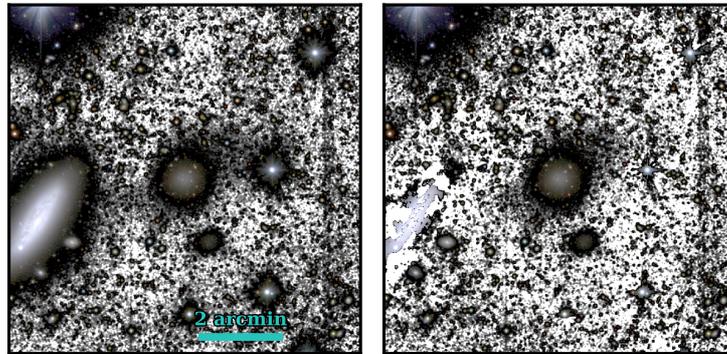


Figure 4.5: Left panel shows the original IAC80 postage stamp of the  $500'' \times 500''$  region around NGC 1052-DF4. The right panel shows the results of subtracting the nearest bright star (west) and the fitting and removal of NGC 1035. After subtracting these sources of contamination, the S shape of the galaxy becomes more evident.

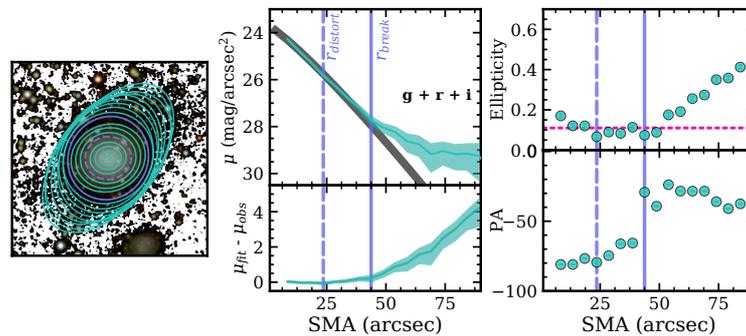


Figure 4.6: Output from `ellipse` for the  $g + r + i$  image of NGC 1052-DF4. The left panel shows the  $200'' \times 200''$  region around NGC 1052-DF4 with the fitted ellipses overplotted. The middle panel shows the radial surface brightness profile as a function of the semimajor axis (top) and the residuals from fitting this profile with a Sérsic model (bottom). The rightmost panel presents the profiles of ellipticity (top) and PA (bottom).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

maps to build the masks. In this case, the “hot” model was run in the original  $g+r+i$  image. The “cold” mask was further expanded by 5 pixels while the “hot” was expanded by 3 pixels, and NGC 1052-DF4 was left unmasked on the “cold” mask. That leaves the light of the galaxy unmasked while masking all compact sources like background galaxies and GCs that are superposed.

Once we have all contaminant sources masked, we derived the radial profiles for the IAC80  $g+r+i$  image with `ellipse`. This image was smoothed with a Gaussian of  $\sigma = 2$  pix to improve the signal-to-noise ratio of the outer parts of the galaxy. The isophote fitting process was done in two steps: (1) allowing all parameters to vary freely and (2) fixing the centers to the median centers of the isophotes returned by the first iteration. The `ellipse` geometry obtained was then used to extract the photometry from the original  $g+r+i$  image (no smoothing applied) and for each of the individual bands.

Figure 4.6 shows the output of `ellipse` for the original  $g+r+i$  image. In the leftmost panel, a postage stamp of a  $200'' \times 200''$  region around NGC 1052-DF4 is shown with the fitted ellipses overlotted. The 1D radial surface brightness profile as a function of the semimajor axis (SMA) is shown in the top middle panel, up to  $\sim 90''$ . The shaded regions represent the median error in each elliptical isophote. The right panel shows the ellipticity (top) and PA (bottom) as a function of SMA. The magenta dashed line in the ellipticity profile marks the value of the ellipticity of NGC 1052-DF4 as reported in [vD19](#).

#### 4.4.3 Defining $r_{\text{break}}$ and $r_{\text{distort}}$

One of the most striking features in the 1D surface brightness profile of NGC 1052-DF4 is its sudden change in the slope at  $\text{SMA} \sim 45''$ . In order to investigate this, Figure 4.6 also shows a [Sérsic \(1968\)](#) fit to the 1D surface brightness profile of NGC 1052-DF4 (gray line in the top middle panel). This Sérsic model ( $R_e = 18''$  and<sup>6</sup>  $n = 0.86$ ) nicely fits the inner ( $R < 40''$ ) parts of the galaxy, as mentioned in [vD19](#), while at large radius, it is not a good description of the surface brightness profile. The bottom middle panel shows the residuals of subtracting the best Sérsic fit from the  $g+r+i$  surface brightness. At a radius of  $\sim 45''$ , there is a significant excess of light with respect to the model fit. Although this excess starts earlier, the slope of these residuals at this radius becomes steeper. This is accompanied by a steep increase in the ellipticity of the isophotes reaching a value of  $\sim 0.4$  at  $\sim 90''$ . In addition, there is a sudden change in the PA trend.

Following [Johnston et al. \(2002\)](#), we tentatively define the radius where the surface brightness profile of NGC 1052-DF4 departs ( $\Delta\mu \gtrsim 0.2 \text{ mag arcsec}^{-2}$ ) from the Sérsic model as the break radius ( $r_{\text{break}}$ ) marked as the solid vertical blue line in the profiles at  $43.8''$ . We also marked this particular ellipse in the postage stamp (left panel of Figure 4.6).

Although this break is the most prominent feature in the surface brightness profile, the satellite starts losing its shape well within this point. [Johnston et al. \(2002\)](#) refers to this point where the morphology of the galaxy is affected by tidal forces as  $r_{\text{distort}}$ . In the case of NGC 1052-DF4, at  $\sim 23.5''$ , the PA starts increasing, i.e., twisting of the isophotes, up to  $r_{\text{break}}$ . This radius is where we define  $r_{\text{distort}}$ , and it is marked in Figure 4.6 as the dashed blue vertical line in the profiles and the dashed blue ellipse in the postage stamp.

<sup>6</sup>This is 1.2 kpc at 13.5 Mpc or 1.7 kpc at 20 Mpc.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

#### 4.4.4 The S shape of NGC 1052-DF4

In simulations, tidal debris is seen to spread along the orbit of the satellite in thin streams (e.g., Johnston et al. 1996). This suggests that the morphology of these loose populations can be used to constrain the direction in which the satellite is moving, which in turn can tell us something about the satellite's orbit. In the previous section, we have seen that the change of shape of NGC 1052-DF4 with radius in the  $g + r + i$  image suggests that the outer parts of this galaxy are fully compatible with being tidally disrupted (Section 4.4.3). In order to assess if signs of this interaction can be seen in the individual images, we binned  $4 \times 4$  pix the IAC80 images to improve the signal-to-noise ratio of the faint outer parts of the galaxy. Consequently, the new pixel is the sum of 16 original pixels. In addition, we convolved the images with a Gaussian kernel of  $\sigma = 1$  (rebinned) pix to further enhance structures. Figure 4.7 shows the postage stamps of the  $250'' \times 250''$  region centered on NGC 1052-DF4 of the three bands:  $g$  (left),  $r$  (middle), and  $i$  (right).

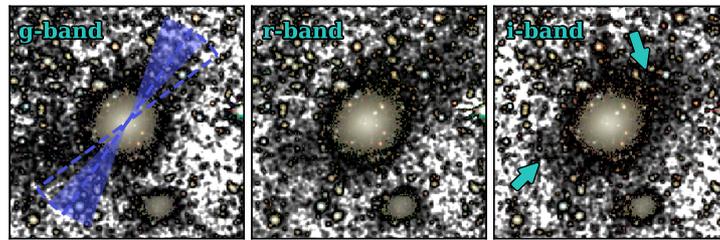


Figure 4.7: Postage stamps of the  $250'' \times 250''$  region centered on NGC 1052-DF4 for the  $g$  (left),  $r$  (middle) and  $i$  (right) IAC80 images. The blue-shaded region indicates the range of the different directions of the orbit of this galaxy as derived by the spatial distribution of GCs in Section 4.3.2, overplotted on the  $g$ -band stamp. The region outlined by the dashed line corresponds to the same range of orbits but without including the GC possibly associated with NGC 1052-DF5. The excess of light to the northwest and southeast of NGC 1052-DF4 is marked by two teal arrows in the rightmost panel ( $i$  band).

In each of the images, there is an excess of light around NGC 1052-DF4 aligned with the range of orbit direction derived from the distribution of GCs in Section 4.3.2, overplotted on the  $g$ -band stamp (left), that is, to the northwest and southeast of the galaxy, marked by the teal arrows in the rightmost panel of Figure 4.7. This excess is seen in all three bands independently. By confirming the presence of this excess in the three different bands, we ensure that it is a real feature.

To explore in more detail this observed S shape and therefore constrain the direction in which NGC 1052-DF4 is moving, we derived the isocontours of the light of this galaxy. To do that, we use the rebinned IAC80  $g + r + i$  and convolve it with a Gaussian of  $\sigma = 2.5$  pix to further enhance structures that might be buried in the noise. We constructed a new mask for this image

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

by visually masking all of the sources. In addition, we generously masked the residuals of the star to the west and of NGC 1035.

We derived the isocontours of light using `matplotlib`'s `contour` function in a similar way to Montes & Trujillo (2019). `contour` provides and draws isocontour lines at different given intensities in the image. In this case, we derived the radial light profiles of the binned  $g + r + i$  image in elliptical apertures, assuming a fixed ellipticity of 0.11 ( $b/a = 0.89$ ) and PA of  $7^\circ$  (counterclockwise from the X-axis), the properties of the inner parts of the galaxy.<sup>7</sup> Using this radial profile, we interpolated the intensity values at five physical radial distances:  $10''$ ,  $20''$ ,  $30''$ ,  $40''$ , and  $50''$  from the center of NGC 1052-DF4. The different distances correspond to surface brightness values of 24.6, 25.8, 26.9, 27.8, and 29.1 mag arcsec<sup>-2</sup>. These contours are plotted in Figure 4.8. In the left panel, the underlying image is a composite of an RGB color image created using the binned ( $4 \times 4$ )  $g$ ,  $r$ , and  $i$  filters and a black and white  $g + r + i$  binned image for the background.

On the right panel, we plotted the same contours on a black background to facilitate visualization. We can clearly see that the last contour (lightest teal shade, at 29.1 mag arcsec<sup>-2</sup>) shows the presence of tidal tails from material moving away from NGC 1052-DF4. Tentatively, we have drawn a schematic S shape (in purple) tracing the shape of this last contour. We also plotted the region containing the axes from the distribution of GCs (Section 4.3.2). The S shape observed here is common in simulations where a dwarf galaxy is interacting with a more massive galaxy (e.g., Figure 5 in Johnston et al. 2002) and also in observations (e.g., Koch et al. 2012).

#### 4.4.5 Tidal radius

The tidal radius identifies the radius where the tidal effects are expected to be important in the satellite. To compute the tidal radius ( $r_{\text{tidal}}$ ) for NGC 1052-DF4, we follow Equation (5) in Johnston et al. (2002). This equation is for the tidal radius at the pericenter of the satellite orbit. Instead, as we do not know where NGC 1052-DF4 is in relation to its orbit around NGC 1035, we compute the instantaneous  $r_{\text{tidal}}$ , which is the tidal radius at the current position of the satellite from the parent galaxy and assuming zero ellipticity, i.e., a circular orbit (Johnston et al. 2002). Consequently, the equation for the instantaneous  $r_{\text{tidal}}$  is given by

$$r_{\text{tidal,inst}} = D \times \left( \frac{m_{\text{dyn}}}{M_{1035,\text{dyn}}} \right)^{1/3}, \quad (4.1)$$

where  $D$  is the radial distance between the galaxies and  $m_{\text{dyn}}$  and  $M_{1035,\text{dyn}}$  are the dynamical masses for NGC 1052-DF4 and NGC 1035, respectively. The dynamical mass of NGC 1035 is  $M_{1035,\text{dyn}} = 1.9 \pm 0.1 \times 10^{10} M_\odot$  (Truong et al. 2017),<sup>8</sup> while for NGC 1052-DF4 we took the estimation of the total mass from vD19;  $m_{\text{dyn}} = 4_{-3}^{+12} \times 10^7 M_\odot$ , assuming a distance of 20 Mpc (Danieli et al. 2020). At a distance of 13.5 Mpc, this dynamical mass would be around half of that value. Assuming that the projected distance between the galaxies is the actual distance

<sup>7</sup>As shown in Figure 1 of Montes & Trujillo (2019), the purpose of this is to obtain approximate values at a certain distance in order to derive a contour, and therefore, the real shape of the galaxy light at that distance.

<sup>8</sup>The dynamical mass of NGC 1035 in Truong et al. (2017) was estimated at a distance of 17 Mpc. As the way this dynamical mass was estimated is proportional to the size of NGC 1035, we converted converted it according to the distance in our calculations.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

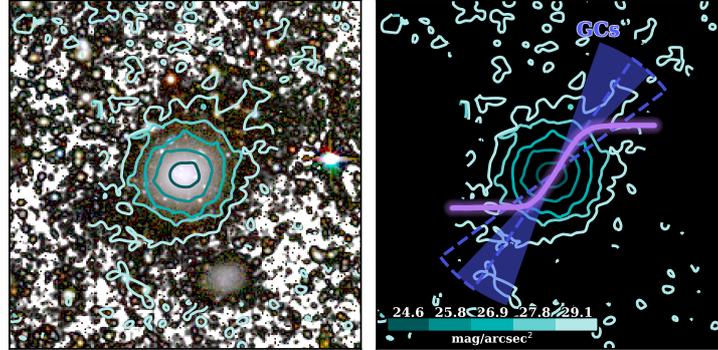


Figure 4.8: Contours at five different surface brightness levels of NGC 1052-DF4. The left panel shows the contours overlotted on a composite from an RGB color image and a black and white  $g + r + i$  image for the background. The images are  $4 \times 4$  binned. The right panel shows the contours over a black background with a tentative S shape (purple solid line) tracing the shape of the last contours (at  $29.1 \text{ mag arcsec}^{-2}$ ). We also plotted the range of directions of the orbit of this galaxy obtained from the GC distribution (blue solid area) and without the GC probably associated with NGC 1052-DF5 (dashed line) in Section 4.3.2. The S shape is compatible with NGC 1052-DF4 moving along the direction defined by the position of the GCs.

between their centers,  $D = 222''$ , we obtained  $r_{\text{tidal}} = 33 \pm 8''$  and  $28 \pm 12''$ , depending on the distance assumed, 13.5 and 20 Mpc, respectively.

This is similar to the  $r_{\text{distort}}$  identified in Figure 4.6 ( $23.5''$ ).  $r_{\text{distort}}$  is defined as the radius where the morphology of the galaxy departs from its original shape. In this case, at  $r_{\text{distort}}$ , the PA of the isophotes starts increasing, i.e., twisting, marking the point where the tidal forces are starting to affect the morphology of the galaxy although the stars still remain bound to the satellite (e.g., Peñarrubia et al. 2009).

Note that if NGC 1052-DF4 was located at 20 Mpc (Danieli et al. 2020), the most likely disturber would be NGC 1052. Assuming the following dynamical mass for NGC 1052  $M_{1052, \text{dyn}} = (1.7 \pm 0.9) \times 10^{12} M_{\odot}$  (at 18 Mpc; Pierce et al. 2005) and a distance of  $28.7'$  between both galaxies, we obtain a tidal radius of  $51 \pm 15''$ . This is around twice larger than if we assume NGC 1035 is the main galaxy responsible for the tidal distortion of NGC 1052-DF4. This larger value for the tidal radius would be in tension with the  $r_{\text{distort}}$  measured in this work. This suggests that NGC 1035 is likely responsible for the disruption of NGC 1052-DF4.

#### 4.4.6 Stellar mass in the tidal tails

$r_{\text{break}}$  identifies the radius where observations become dominated by the unbound populations of stars and therefore where we can find breaks in the light profiles (e.g., Johnston et al. 2002; Peñarrubia et al. 2009). The lack of strong distortions in the stellar light at the center of

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

NGC 1052-DF4 indicates that the stars of this galaxy might only be starting to be stripped now. Consequently, we can infer a lower limit to the amount of stellar mass stripped so far by NGC 1052-DF4 by measuring the stellar mass outside  $r_{\text{break}}$ . The fraction of the stellar light of NGC 1052-DF4 beyond  $r_{\text{break}}$ , including the tidal tails down to  $\mu_r = 29.1 \text{ mag arcsec}^{-2}$ , is  $7 \pm 1\%$ , averaging the individual estimates from the  $g$ ,  $r$  and  $i$  bands. This relatively low value for the stripped stellar mass, together with the fact that the central part of the galaxy appears undisturbed, is compatible with the idea that the stellar body of the galaxy is only starting to be disrupted now. Note that this quantity is a lower limit of the true fraction of mass lost by the galaxy.

## 4.5 Discussion

In this work, we show that NGC 1052-DF4 presents clear signs of interaction. In the following, we discuss how this interaction can affect the amount of DM present in this galaxy. For all the calculations, we assume that the distance to the galaxy is 13.5 Mpc (Monelli & Trujillo 2019), although qualitatively the main results presented here do not depend on this assumption.

Note that the previous deep imaging of NGC 1052-DF4 reached around  $\mu_r = 28.5 \text{ mag arcsec}^{-2}$  ( $3\sigma$ ,  $10'' \times 10''$ , Müller et al. 2019), still inconclusive for addressing the problem of the interaction between these two systems. Similarly, the data from The Dragonfly Telephoto Array, where this galaxy was identified (Cohen et al. 2018), reach between  $\mu_{g/r} = 27.4$  and  $28.0 \text{ mag arcsec}^{-2}$  ( $3\sigma$ ,  $12'' \times 12''$ ) as reported in Merritt et al. (2016).

### 4.5.1 The distribution of light and GCs

Matter stripped from satellite galaxies forms tidal tails which will eventually become completely unbound and form larger tidal extensions tracing approximately the original orbit of the satellite in its host halo. In the same manner, the GCs of these satellites will be stripped from the galaxy and deposited along the orbit.

Using  $N$ -body simulations, Klimentowski et al. (2009) studied the properties of tidal tails of a dwarf galaxy orbiting in a Milky Way-like potential. They showed that there are always two tidal tails emanating from the two opposite sides of the dwarf galaxy and that, for most of the time, these tails are oriented radially toward the host galaxy and not along the orbit. This shape, reminiscent of an S, is produced by the particles that, once unbound, are seen to move in the direction of the tidal forces, which are perpendicular to the orbit of the satellite, before being dispersed along with it.

The outskirts of NGC 1052-DF4 present an excess of light to the northwest and southeast of the galaxy (marked by arrows in Figure 4.7). In Figure 4.8, we explored this in more depth using a binned  $g + r + i$  and plotting isocontours of light at different surface brightness values. The faintest surface brightness ( $29.1 \text{ mag arcsec}^{-2}$ ) clearly shows an S shape (tentatively drawn in purple). The tails of the S are oriented directly to NGC 1035 (see Figure 4.1), indicating that this disk galaxy (Truong et al. 2017) is causing the tidal perturbation. As shown in Figure 5 in Monelli & Trujillo (2019), NGC 1035 is the closest galaxy to NGC 1052-DF4 at a projected distance of  $\sim 222'' = 14.5 \text{ kpc}$ , at a distance of 13.5 Mpc (Sorice et al. 2014; Monelli & Trujillo 2019).

In addition, in Section 4.3.2, we identify four new GC candidates of NGC 1052-DF4. These,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

along with the confirmed GCs in [vD19](#), align in a particular direction. This preferential alignment of the GCs suggests that they are also being stripped in the interaction, and therefore, they are distributing in the direction of the orbit.

Putting all the evidence together, what [Figure 4.8](#) is telling us is that NGC 1052-DF4 is moving roughly along the direction defined by the axis of the GCs, roughly perpendicular to the tails of the S.

#### 4.5.2 Can tidal stripping explain the missing dark matter?

Deep imaging of NGC 1052-DF4 has revealed that this galaxy is undergoing tidal disruption. Now, the question that arises is whether this tidal interaction could explain the low DM content measured in this galaxy.

[Smith et al. \(2013, 2016\)](#) used numerical simulations to explore the effect of tidal stripping in the DM content of dwarf galaxies in clusters of galaxies. They found that only when the remaining DM fraction falls under 10% – 15% is when the stars and GCs of the galaxy are significantly stripped. We warn the reader that these simulations cannot be directly compared to the NGC 1052-DF4 scenario, but they provide a qualitative understanding of the problem.

To assess whether the tidal stripping scenario explains the properties of NGC 1052-DF4, we need to calculate the remaining fraction of DM. For this, we first determined the total stellar mass of the galaxy (including tidal tails). The colors were obtained by integrating the light of the galaxy and tidal tails for each of the bands,  $g$ ,  $r$ , and  $i$ . We used the relations between color and mass-to-light ratios from [Roediger & Courteau \(2015\)](#) for three combinations of colors  $g-r$ ,  $g-i$ , and  $r-i$ . The total stellar mass of the galaxy is  $M_* = 3.6 \pm 1.3 \times 10^7 M_\odot$ , the average and dispersion from the three individual estimates, at 13.5 Mpc.

Using the stellar mass to halo mass relationship in [Brook et al. \(2014\)](#), we derived an expected initial total mass for NGC 1052-DF4 of  $M_{h,ini} = 2.2 \pm 0.3 \times 10^{10} M_\odot$ .<sup>9</sup> Taking the current estimate of the total dynamical mass measured in [vD19](#), the average remaining fraction of DM of  $\lesssim 1\%$ . However, the dynamical mass-to-light ratios of isolated UDGs vary widely (e.g., [Toloba et al. 2018](#); [Müller et al. 2020](#)) affecting our estimate of the initial total mass of the galaxy. So, the remaining DM fraction could be as high as 10%.

This is consistent with the simulations in [Smith et al. \(2013, 2016\)](#). The bulk of the stars in a satellite are located in the very central regions of its DM halo. Hence, the stars are more shielded from the tidal forces of the host galaxy with respect to the outer parts of the DM halo. This means that, during the interaction, the DM from the outer parts will be preferentially stripped over the stars of the galaxy ([Smith et al. 2016](#)). In other words, the mass stripping of DM halos proceeds gradually from the outside in (e.g., [Stoehr et al. 2002](#); [Peñarrubia et al. 2008](#); [Smith et al. 2016](#)).

In NGC 1052-DF4, the fraction of stellar mass in the tidal tails is  $\sim 7\%$  ([Section 4.4.6](#)), meaning that the bound stellar mass is 93%, in contrast with the  $\lesssim 1\%$  of DM still bound to the galaxy. This indicates that the galaxy has lost significant amounts of DM before it started losing its stars, in agreement with simulations.<sup>10</sup> Therefore, tidal stripping is a very likely

<sup>9</sup>Alternatively, if we use the number of GCs to infer the initial halo mass of this galaxy as in [Saifollahi et al. \(2020\)](#), the result is  $M_{h,ini} = 3.6 \times 10^{10} M_\odot$  for the seven confirmed GCs in [vD19](#) or  $M_{h,ini} = 5.6 \times 10^{10} M_\odot$  for all the GCs identified in this work, a factor of 1.8 – 2.5 higher.

<sup>10</sup>Even if the fraction of stellar mass unbound was of 50% of the current stellar mass of the galaxy, this scenario will still be true.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

explanation for the extremely low DM content of NGC 1052-DF4 (see also Ogiya 2018; Nusser 2020; Yang et al. 2020; Jackson et al. 2021; Macciò et al. 2021). Note that we are assuming that the measured DM content of this galaxy in vD19 is not affected by the tidal interaction. According to Smith et al. (2013), the observed GC velocity dispersion can be used to measure the true enclosed total mass to within a factor of 2, even when the remaining DM fraction falls as low as  $\sim 3\%$ . However, when it drops below 3%, the total mass is overestimated due to the presence of large numbers of unbound GCs that boost the inferred velocity dispersion of the system (Smith et al. 2013).

To sum up, the preferential tidal stripping of DM over the stars in NGC 1052-DF4 is a very plausible explanation for the measured low content of DM in this galaxy.

#### 4.6 Conclusions

The recent discovery of a second galaxy “missing” dark matter had the potential to revolutionize how we understand galaxy formation. In this work, we have shown evidence that this galaxy is undergoing tidal disruption. Therefore, the “absence of dark matter” in NGC 1052-DF4 is almost certainly caused by its interaction with its neighbor, NGC 1035. Dark matter is less concentrated than stars and, therefore, during interactions, is preferentially stripped from satellite galaxies. In the case of NGC 1052-DF4, as the central parts of the galaxy remain untouched and only  $\sim 7\%$  of the stellar mass of the galaxy is in the tidal tails, we can assume that the stellar component is now starting to be stripped.

This work shows the necessity of very deep imaging in order to detect and characterize these faint substructures around dwarf satellites in nearby galaxies in order to explain peculiarities that, otherwise, remain inexplicable in shallower data.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 5

## Replicability in ultradeep imaging

The present chapter describes ongoing work where we explore the outskirts of the galaxy NGC 493 and the replicability of the low surface brightness features in deep imaging. We use data from the Sloan Digital Sky Survey, Stripe 82, Las Cumbres Observatory Telescope, and Gran Telescopio Canarias. We show that the study of the stellar halo requires observations deeper than  $30 \text{ mag arcsec}^{-2}$ . Also, we will show that current technology allows us to explore reliably low surface brightness features down to  $31 \text{ mag arcsec}^{-2}$ . The work in this chapter has been done using Maneage (see Chapter 2).

### 5.1 Introduction

Getting deeper and deeper astronomical images is a difficult task. The first challenge with ground-based images is caused by the intrinsic brightness of the nighttime atmosphere (which in the darkest spots on Earth is around  $22 \text{ mag arcsec}^{-2}$  in the optical bands, see e.g., Patat 2003). This brightness is mainly produced by the airglow, followed by Zodiacal light and scattered light from astronomical sources. The airglow is the brightness resulting from the recombination of atoms excited by daylight that emits photons back at night. The airglow changes the brightness of the atmosphere on time-scales of minutes, so exposures with a long enough integration time or the sum of many short exposures should produce a relatively flat image background. Removing this flat background should be relatively easy, and for this reason, the vast majority of present-day optical astronomical images reach, without major problems, surface brightness limits that are about 100 times dimmer than the night sky. A good example of this is the astronomical images produced by the current standard on survey imaging, i.e., the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2003). These images have a typical depth of  $26.5 \text{ mag arcsec}^{-2}$  (Pohlen & Trujillo 2006). This is also the traditional depth reached using photographic plates (de Vaucouleurs et al. 1976).

Breaking the barrier of  $26.5\text{--}27 \text{ mag arcsec}^{-2}$  requires a significant amount of telescope time. Currently, there is a good number of examples showing images with depth (measured as  $3\sigma$  and  $10 \times 10 \text{ arcsec}^2$  boxes) between  $28 \text{ mag arcsec}^{-2}$  (e.g., Merritt et al. 2016) to  $29 \text{ mag arcsec}^{-2}$  (Duc et al. 2015; Fliri & Trujillo 2016), and even some reaching  $29.5 \text{ mag arcsec}^{-2}$  (see e.g., Mihos et al. 2017). In dealing with those datasets, whose surface brightness limits are around

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

1000 times fainter than the night sky, one has to deal with many observational problems. One of these difficulties is the light reflected by the structure of the telescope including its dome. Potential solutions to this problem are to simplify the optics of the telescope to avoid, as much as possible, the problem connected to the internal reflections (see e.g., Abraham & van Dokkum 2014; Muslimov et al. 2017) or to use observing strategies that minimize these reflections by applying a combination of dithering and rotation patterns (see e.g., Trujillo & Fliri 2016). In addition and related to this, the point spread function (PSF) of the images produces dispersion of the astronomical sources, whose brightness is distributed in a complicated structure (with surface brightness around  $29 \text{ mag arcsec}^{-2}$  even in relatively empty sky regions; Slater et al. 2009). This intricate background can be subtracted from the images when the PSF of the data is known with high accuracy (see Chapter 3).

Ultra-deep images reaching depths beyond  $30 \text{ mag arcsec}^{-2}$  is very scarce (see Chapter 1 for further details about the low surface brightness challenges and how to deal with them). In fact, only the largest telescopes (VLT, GTC) have been able to produce images at such extreme surface brightness (nearly a million times fainter than the night sky brightness). For this reason, due to the very limited number of very deep observations, it is very timely to probe how reliable are the detections of structures on the sky reaching surface brightness fainter than  $30 \text{ mag arcsec}^{-2}$ . Moreover, it is worth comparing very deep imaging with shallower data to see what (if any) are the effects of insufficient depth on the inferred properties of the objects under study. The goal of this work is to explore these issues by comparing images of the same galaxy taken by different telescopes and filters.

This chapter is structured as follows. In Section 5.2, there is a general description of the galaxy NGC 493 indicating its main characteristics. Section 5.3 is dedicated to describing the datasets from different facilities and the methodology used for obtaining the results. In Section 5.4, the results are presented. In Section 5.5, we provide our conclusions.

## 5.2 The galaxy NGC 493

NGC 493 is an Sc galaxy (Nilson 1973) located at R.A. (J2000) = 01:22:08.985 and dec (J2000) = +00:56:43.21. The galaxy is at a distance of  $D = 23.1 \text{ Mpc}$  (Sorce et al. 2014). At such a distance 1 arcsec corresponds to 0.112 kpc. The galaxy was selected to fulfill three observational conditions:

- It is located in a region of the sky with relatively low dust contamination:  $A_g = 0.108$  and  $A_r = 0.074$  (according to NASA/IPAC Extragalactic Database (NED); Schlafly & Finkbeiner 2011).
- There are no bright foregrounds objects in the vicinity whose scattered light can affect the galaxy.
- Its spatial dimensions ( $1.73 \times 0.46 \text{ arcmin}$ ) are small enough that it can be observed by cameras with a modest field of view.

Additionally, the galaxy was selected to be within the Stripe 82 region, yielding an extra observation to compare with. Using the SDSS we derive the following apparent magnitudes of: 12.7 mag ( $g$ -band) and 12.3 mag ( $r$ -band), which after correction by Galactic extinction implies

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

of a color of  $g-r = 0.37$  mag. The absolute magnitude in the  $r$ -band is  $-19.6$  mag. Using [Roediger & Courteau \(2015\)](#), we estimate a mass-to-light ratio  $(M/L)_r = 0.65\Upsilon_{\odot}$ . Consequently, the estimated stellar mass is  $3 \times 10^9 M_{\odot}$ . This is about 1/20th of the stellar mass of the Milky Way. The amount of  $H_I$  mass can be obtained from its integrated  $H_I$  flux density ( $S_{H_I} = 31.14 \text{ Jy km s}^{-1}$ ; [Springob et al. 2005](#)). The following equation is applied to get the  $H_I$  mass:  $M_{H_I} = 2.36 \times 10^5 \times D^2 \times S_{H_I} = 3.9 \times 10^9 M_{\odot}$  (see e.g., [Filho et al. 2013](#)).

The dynamical mass within the region dominated by the baryonic disk can be crudely measured by using the size of the galaxy and its rotational velocity ( $M_{\text{dyn}} = 2.326 \times 10^5 R_{25}^2 v_{\text{rot}}^2$ ; see e.g., [Pohlen & Trujillo 2006](#)). The radial distance corresponding to the  $25 \text{ mag arcsec}^{-2}$  isophote is  $R_{25} = 1.73$  arcmin, or equivalently 11.6 kpc. Its maximum rotational velocity corrected by inclination is 125 km/s (HyperLeda; [Makarov et al. 2014](#)). Consequently, the dynamical mass of NGC 493 in the innermost  $\sim 10$  kpc is  $M_{\text{dyn}} = 4.2 \times 10^{10} M_{\odot}$ . In this sense, the galaxy is very similar to M33 (see e.g., [Corbelli & Salucci 2000](#)).

### 5.3 Datasets description

In this section, the different datasets are reviewed. They were obtained using the following facilities: SDSS (including also Stripe 82), Las Cumbres Observatory (LCO), and the Gran Telescopio Canarias (GTC). A general description of the SDSS and Stripe 82 surveys can be found in Section 1.1.3 (for further details of SDSS see Chapter 3, where the SDSS survey was used for constructing extended PSFs). The reduction of the raw data from the LCO and GTC facilities are also fully detailed in this section. It is important to recall that the reduction of the data, as well as the later analysis, have been done using Maneage (see Chapter 2). In Table 5.1, we summarize all the limiting depths and exposure times of the data.

| Telescope | Diameter<br>(m) | FWHM <sub>g</sub><br>(arcsec) | FWHM <sub>r</sub><br>(arcsec) | $t_{g,\text{exp}}$<br>(h) | $t_{r,\text{exp}}$<br>(h) | $\mu_{g,\text{lim}}$<br>(mag arcsec <sup>-2</sup> ) | $\mu_{r,\text{lim}}$<br>(mag arcsec <sup>-2</sup> ) | $m_{g,\text{lim}}$<br>(mag) | $m_{r,\text{lim}}$<br>(mag) |
|-----------|-----------------|-------------------------------|-------------------------------|---------------------------|---------------------------|---|---|-----------------------------|-----------------------------|
| SDSS      | 2.5             | 1.2                           | 1.1                           | 0.03                      | 0.03                      | 27.66±0.02  | 27.23±0.02  | 23.66±0.02                  | 23.37±0.03                  |
| Stripe 82 | 2.5             | 1.3                           | 1.3                           | 0.75                      | 0.75                      | 29.25±0.01  | 28.67±0.01  | 25.15±0.03                  | 24.64±0.05                  |
| LCO       | 1.0             | -                             | 1.9                           | -                         | 45.3                      | -   | 30.04±0.01  | -                           | 25.46±0.04                  |
| GTC       | 10.4            | 1.2                           | 1.1                           | 5.1                       | 6.1                       | 31.46±0.01  | 31.05±0.02  | 27.25±0.03                  | 26.93±0.04                  |

Table 5.1: Main characteristics of the different datasets used in this work.

#### 5.3.1 Sloan Digital Sky Survey (SDSS)

SDSS  $g$ ,  $r$  and  $i$ -band imaging data were retrieved from the SDSS DR14 Sky Server ([Abolfathi et al. 2018](#)). The magnitude zeropoint for the whole dataset is the same: 22.5 mag. The exposure time of the images is 53.9s, and the pixel size is 0.396 arcsec. The observations were taken in drift scanning mode, providing accurate photometry down to  $\mu_r \sim 26.5\text{--}27 \text{ mag arcsec}^{-2}$  ( $3\sigma$ ;  $10 \times 10 \text{ arcsec}^2$  in the  $r$ -band; [Pohlen & Trujillo 2006](#)).

In the particular case of the region where NGC 493 is located, the SDSS survey repeated the observation twice, producing images that are effectively equivalent to 1.8 min of integration time. As a result of this deeper integration, this field is deeper than regular SDSS images. Two estimations of the depth of the different datasets were computed: the limiting magnitudes for point-like sources and the surface brightness limits (see Section 1.2.3). This is done in a circular

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

region centered on the galaxy with a diameter of 10 arcmin. This size is chosen to guarantee that the same area is explored by the different telescopes. The images were profusely masked by using `NoiseChisel` (Akhlaghi & Ichikawa 2015) to detect the background and foreground sources. In addition to the galaxy itself, very bright objects close to NGC 493 are manually masked.

The limiting magnitudes were estimated by randomly placing circular apertures on the background of the images. The diameter of the apertures ( $D$ ) is twice the full width half maximum (FWHM) of point-like sources. This aperture is taken to measure around 93% of the flux of a point-like source under the assumption of Gaussian seeing (see e.g., Trujillo et al. 2001a). In the region around NGC 493, the SDSS images have a seeing of FWHM  $\sim 1.2$  arcsec in  $g$  and  $r$  filters. The limiting magnitudes are estimated as the  $5\sigma$  fluctuations with respect to the background. The values computed are  $m_{g,\text{lim}} = 23.66 \pm 0.02$  mag and  $m_{r,\text{lim}} = 23.37 \pm 0.03$  mag.

The limiting surface brightness was estimated in the same region of the image as the limiting magnitude (i.e., in a circle of 10 arcmin surrounding NGC 493 with the galaxy and the background and foreground sources properly masked). The following surface brightness limits were obtained:  $\mu_{g,\text{lim}} = 27.66 \pm 0.02$  and  $\mu_{r,\text{lim}} = 27.23 \pm 0.02$  mag arcsec $^{-2}$  ( $3\sigma; 10 \times 10$  arcsec $^2$ ).

### 5.3.2 The IAC Stripe 82 Legacy Project

The Stripe 82 is a 2.5-degree wide stripe along the celestial equator in the southern Galactic cap with a total area of 275 square degrees observed in all the five SDSS bands.

The Stripe 82 survey is a 2.5-degree wide equatorial stripe ( $-50^\circ < \text{R.A.} < 60^\circ$ ,  $-1.25^\circ < \text{dec} < 1.25^\circ$ ) covering a total area of 275 square degrees in all five SDSS bands. It has been imaged by the SDSS telescope multiple times under generally photometric conditions with good seeing and low sky background (Abazajian et al. 2009).

In the IAC reduction of this dataset (Fliri & Trujillo 2016; Román & Trujillo 2018), the main goal was to preserve the characteristics of the background (sky and diffuse light) on the individual images by using a non-aggressive sky subtraction strategy. Incorporating such a strategy in the co-addition of these images has allowed us to reach two magnitudes deeper than the regular SDSS. The co-added images reach a mean surface brightness limit of  $\mu(3\sigma; 10 \times 10 \text{ arcsec}^2) = 27.9, 29.1, 28.6, 28.1,$  and  $26.7$  mag arcsec $^{-2}$  for the  $u, g, r, i,$  and  $z$  bands, respectively. These depths can change slightly depending on the spatial location as some regions of the Stripe 82 were observed more frequently than others.

In the particular case of the region where NGC 493 is, the total exposure time on the source is 0.75 h. The FWHM seeing of these images is very similar in both bands:  $\sim 1.3$  arcsec. The limiting magnitudes and surface brightness limits around NGC 493 were computed as explained in the previous section. In this case, the results are:  $m_{g,\text{lim}} = 25.15 \pm 0.03$  mag and  $m_{r,\text{lim}} = 24.64 \pm 0.05$  mag. The limiting surface brightness are:  $\mu_{g,\text{lim}} = 29.25 \pm 0.01$  and  $\mu_{r,\text{lim}} = 28.67 \pm 0.01$  mag arcsec $^{-2}$  ( $3\sigma; 10 \times 10$  arcsec $^2$ ).

### 5.3.3 Las Cumbres Observatory Telescope

Las Cumbres Observatory (LCO) Telescope is a network of observatories with identical telescopes of 0.4, 1, and 2 m apertures (Brown et al. 2013). For the present work, 1 m telescopes were used. The 1 m telescopes are f/7.95 Ritchey-Chrétien optical systems on equatorial mounts, equipped

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

with “Sinistro” optical cameras with a field of view of  $26.5 \text{ arcmin} \times 26.5 \text{ arcmin}$ . The pixel size is  $0.389 \text{ arcsec}$ . We used the Sloan  $r$  filter.

The observations presented in this work were conducted by two different 1 m telescopes during fall 2015, both located at the Cerro Tololo site. We took 176 different frames of 1200 s each, producing a total integration time of 58.67 h. Images were obtained during different non-consecutive nights and different seeing conditions, although all of them were taken during photometric and dark nights.

The observing dithering pattern ( $3 \times 3$  grid) was designed so the galaxy was placed at different positions of the CCDs in each run. The dithering pattern step was 1 arcmin. This allowed a precise determination of the sky background and the removal of artifacts (see Section 1.3.4). 21% of the individual frames were discarded due to different issues, mostly problems with the guiding system. The final co-addition is an image composed of 136 frames with a total exposure time of 45.33 h. The FWHM seeing of the final LCO image is  $\sim 1.9 \text{ arcsec}$ .

The final image was calibrated photometrically using as a reference the SDSS image. To do this, after subtracting the sky from the image, the radial profile of the galaxy NGC 493 was computed and matched with the radial profile obtained for the SDSS image. Both profiles were extracted using the same position angle,  $PA = 148.0 \text{ deg}$ , and axis ratio,  $q = 0.32$ . The region used to match both profiles is  $5 < R < 50 \text{ arcsec}$ . These limits are selected to avoid the innermost region of the galaxy which is affected differently by the diverse PSF FWHMs of each telescope and also to guarantee a high signal-to-noise ratio.

Using the same region as before, the limiting magnitude for the LCO image is  $m_{r,\text{lim}} = 25.46 \pm 0.04 \text{ mag}$ . The limiting surface brightness is:  $\mu_{r,\text{lim}} = 30.04 \pm 0.01 \text{ mag arcsec}^{-2}$  ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ).

### 5.3.4 Gran Telescopio Canarias

The Gran Telescopio Canarias (GTC) is a 10.4 m telescope located at the Roque de los Muchachos Observatory (ORM), in La Palma, Spain. The Optical System for Imaging and low-intermediate-Resolution Integrated Spectroscopy (OSIRIS; Cepa et al. 2003) camera was used to get the deepest optical data presented in this chapter. OSIRIS has a total unvignetted field of view of  $7.8 \times 7.8 \text{ arcmin}$ . This camera consists of two CCDs with a gap of  $9.4 \text{ arcsec}$  between them. The pixel scale of the camera is  $0.254 \text{ arcsec}$ .

The requested time was 8 h in both Sloan  $g$  and  $r$  filters. Data were taken during several (non-consecutive) nights. Different technical problems reduced the amount of useful time on source to 5 h in  $g$ -band and 6 h in  $r$ -band. The final co-added images have a FWHM seeing of  $\sim 1.1 \text{ arcsec}$ .

The observing strategy is explained in full detail in Trujillo & Fliri (2016). In short, a dithering and rotation pattern is used to minimize the effect of residual light on the images. Thanks to this strategy, the collective power of a 10.4 m telescope and a careful data reduction, the expected surface brightness limits in both bands should be fainter than  $31 \text{ mag arcsec}^{-2}$  ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ).

As was the case for LCO, the GTC final coadded images in  $g$  and  $r$  were calibrated using the NGC 493 profiles of SDSS. For the GTC, the limiting magnitudes are  $m_{g,\text{lim}} = 27.25 \pm 0.03 \text{ mag}$  and  $m_{r,\text{lim}} = 26.93 \pm 0.04 \text{ mag}$ . The limiting surface brightness are  $\mu_{g,\text{lim}} = 31.46 \pm 0.01 \text{ mag arcsec}^{-2}$  and  $\mu_{r,\text{lim}} = 31.05 \pm 0.02 \text{ mag arcsec}^{-2}$  ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

## 5.4 Results

The different telescope imaging of NGC 493 are shown in Figure 5.1. This figure illustrates how the observed features of the galaxy change as a function of depth.

The color part of the images corresponds to a colour composite of the SDSS filters  $g$ ,  $r$  and  $i$  taken from the SDSS data. The black and white part of each dataset corresponds to the image on the  $r$ -band filter obtained in each telescope. Some remarkable results are shown in Figure 5.1. The first one is that extremely deep observation are required to observe the rich stellar halo surrounding this modestly massive galaxy (NGC 493 is very similar to M33). In fact, it is necessary to go beyond  $29 \text{ mag arcsec}^{-2}$  to find the characteristic filamentary structure of the stellar halos. Another important result is that prominent streams visible in the GTC image are starting to be noticeable in the LCO image. This indicates that these very diffuse ( $29\text{--}30 \text{ mag arcsec}^{-2}$ ) structures are reproducible using different telescopes. The comparison shows that surface brightness features around  $30 \text{ mag arcsec}^{-2}$  are well within the capabilities of a 1 m telescope using 50 h on-source once a good observing strategy and a careful reduction job is conducted.

To explore the reproducibility of even deeper data, it is necessary to compare the observations taken with the GTC telescope in the different filters. This comparison is done in Figure 5.2. The comparison using different filters is not as straightforward as it was above using the same filter, as the lowest surface brightness features could have different colors, and therefore, their brightness across the image can change depending on the color of such feature.

At first sight the  $g$  and the  $r$ -band GTC images of NGC 493 look very much alike. However, it can be seen that in general the low surface brightness features in the  $r$ -band are more prominent. A striking example of this is the appearance of the intracluster light in a filament of galaxies at  $z = 0.65$  in the images. At such a redshift, the  $g$ -band is equivalent to the near-ultraviolet rest-frame, while the  $r$ -band corresponds to  $u$ -band rest-frame. Only when approaching the optical rest-frame it is possible to appreciate the intracluster light along the filament. However, the rest of the structures around NGC 493 are visible in both filters with similar characteristics, likely indicating that they correspond to stellar streams associated with the galaxy itself. The faintest visible features around NGC 493 have a brightness of  $\sim 31.5 \text{ mag arcsec}^{-2}$ , showing the replicability of structures of this brightness in our current deepest optical imaging datasets.

Taking into account that the different datasets correspond to the same galaxy and its surroundings, it is possible to study how the properties of the background sources, as well as those of the galaxy NGC 493 itself, change as a function of the imaging depth. In the next subsections, the outskirts of the galaxy will be analyzed with the goal of studying the properties of the background sources, and how they affect the low surface brightness estimations.

### 5.4.1 Typical distance between sources

As images get deeper, the average distance between the outer regions of the sources becomes smaller. In this work, the average distance between all the sources on the images has been computed for the different datasets. Here, a source is defined as  $1.0\sigma$  detection on areas of 10.0 pixels above the sky background noise level (DETECT\_THRESH and DETECT\_MINAREA parameters of SExtractor). The average distance between sources is estimated as follows. For each source, the distance to the rest of the sources is computed. Then, the distance to its closest neighbor

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

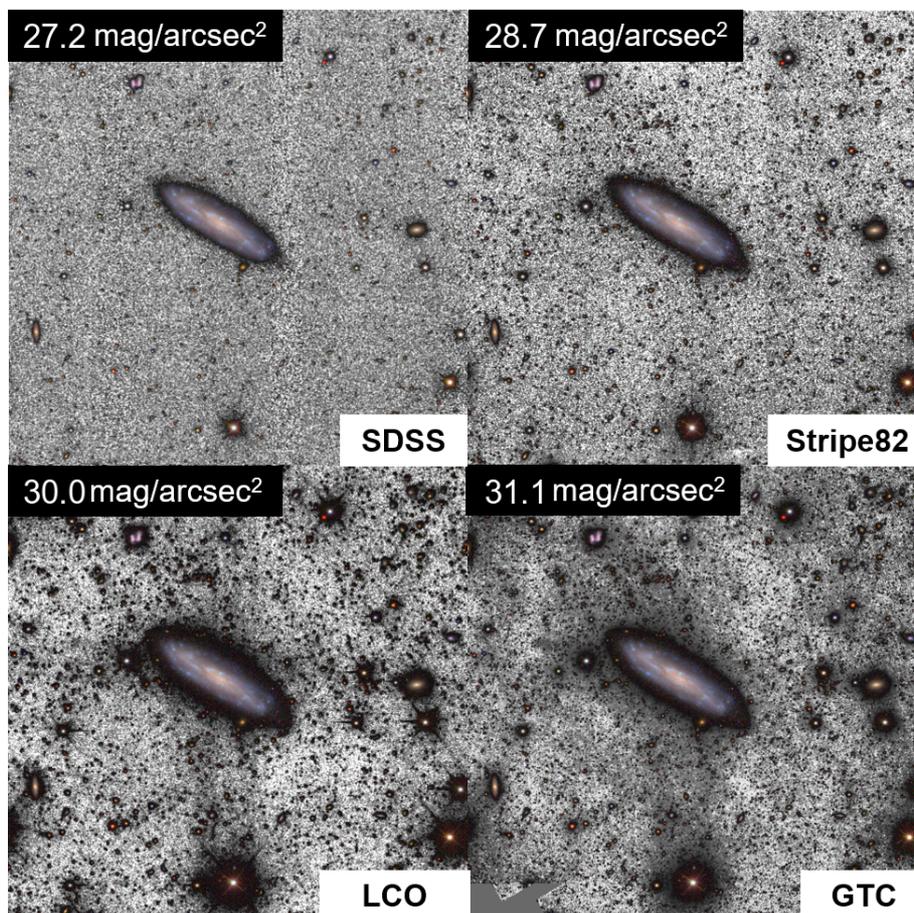


Figure 5.1: The galaxy NGC 493 as seen by different telescopes and integration times. SDSS (2.5 m;  $2 \times 53.9$  s), Stripe 82 (2.5 m; 0.75 h), LCO (1 m, 45.3 h), and GTC (10.4 m; 6.1 h). The images are a color composite of the SDSS filters  $g$ ,  $r$ , and  $i$  from the SDSS survey, with the black and white background the image corresponding to the  $r$ -band filter at the different telescopes. The size of the images corresponds to  $12.6 \times 12.6$  arcmin. North is up, east is left.

is taken. The distance between sources follows a log-normal distribution. To determine the mean distance between sources, a Gaussian function is fitted to the base 10 logarithm of the distribution of the distances (see Figure 5.3). The average distance is measured as the location

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

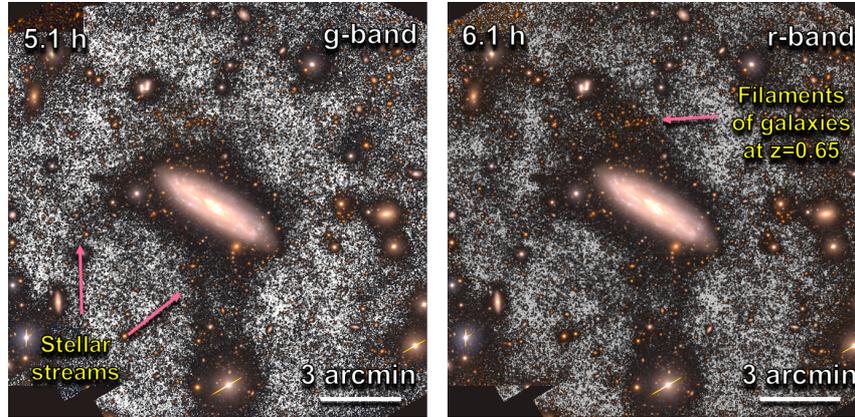


Figure 5.2: The galaxy NGC 493 as seen by the 10.4 m GTC telescope using two different filters:  $g$ -band and  $r$ -band. The faintest features observable in these images correspond to  $\mu_g = 31.3 \text{ mag arcsec}^{-2}$  and  $\mu_r = 31.0 \text{ mag arcsec}^{-2}$  ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ).

of the peak of the Gaussian distribution.

From Figure 5.3, it can be seen that the peak of the distribution of distances is different for each dataset. As the images become deeper, the typical distance between sources is smaller, thus indicating that deeper images tend to have a higher density of astronomical objects. In particular, for the  $r$ -band, the typical distances between sources are  $10.01 \pm 0.22$ ,  $7.36 \pm 0.10$ ,  $6.47 \pm 0.07$ , and  $3.62 \pm 0.02 \text{ arcsec}$ , for the SDSS, the Stripe 82, the LCO, and the GTC images, respectively. It is important to note that this analysis is the study of how the typical distance between sources varies as a function of the depth of the images, and not absolute or averaged values for a given depth or dataset. In other words, the particular values computed represent only the typical distance between sources for the outskirts of the galaxy NGC 493.

#### 5.4.2 How far are current optical images from the confusion limit?

In Figure 5.3, the distribution of typical distances between sources for the different datasets in  $r$ -band are shown. Interestingly, in the deepest images (GTC), the average distance between sources is  $3.62 \pm 0.02 \text{ arcsec}$ . This result raises the following question: how far are our current deepest optical images from reaching the confusion limit?

Figure 5.4 shows the typical distances between sources as a function of the limiting magnitude for the different datasets and filters ( $g$  and  $r$ -band). This figure shows an expected result: as the images get deeper, the objects are found to be closer to each other.

Assuming that the spatial distance between sources continues to decrease as observed in the data, it is possible to extrapolate at which brightness limit there will be a source confusion in the optical. We define that an image has reached its confusion limit when there are no pixels on the image free of being affected by the light of detected sources. A very conservative way of

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

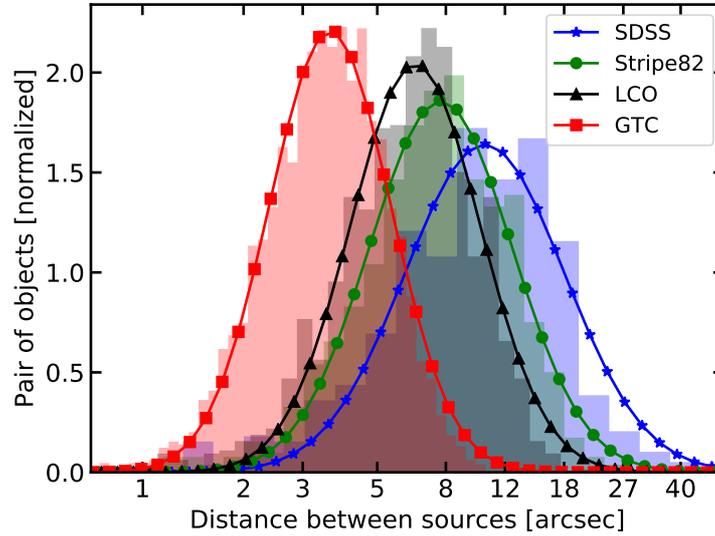


Figure 5.3: Distribution of the typical distances between sources for the different datasets using  $r$ -band images. Measured distances between sources are shown as transparent histograms. The typical distances between sources are given by the location of the peak of the Gaussian fitting, plotted with solid lines. Note that the x-axis is not in linear scale.

estimating this value is by assuming that the light distribution of point sources can be described with a Gaussian shape. If an object follows a Gaussian profile, its light affects basically those pixels located within a radial distance of  $R_S = 3\sigma$ , where  $\sigma$  can be estimated from the FWHM of the point spread function of the image, i.e.,  $\sigma = \text{FWHM}/\sqrt{8 \ln 2}$  (see e.g., Trujillo et al. 2001a). Consequently, the confusion limit will be reached when the centers of two sources are separated  $2R_S$ .

For an image obtained with telescopes on Earth, a reasonable value for the FWHM is 1 arcsec. In that case, the confusion limit would be reached when the distance between sources is  $\sim 2.55$  arcsec. For a telescope like the GTC, that will be reachable after an observation of  $\sim 20.3$  and  $\sim 43.8$  h in the  $g$  and  $r$  bands, respectively. For the next generation of telescopes like the Rubin Observatory Legacy Survey of Space and Time (LSST,  $D = 8.4$  m), the Thirty Meter Telescope (TMT,  $D = 30$  m), or the Extremely Large Telescope (ELT,  $D = 30$  m), that will be obtained in  $\sim 31$ ,  $\sim 2.5$ , and  $\sim 1.5$  h, respectively ( $g$  band). These numbers have been obtained by using the Equation 1.15, assuming  $5\sigma$  and taking the GTC  $g$ -band as the reference. The limiting magnitude considered is  $m_{\text{lim}} = 28$  mag, according to Figure 5.4. This result shows that current technology is capable of reaching the confusion limit in the optical band if sufficient exposure time is allocated in the observation.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

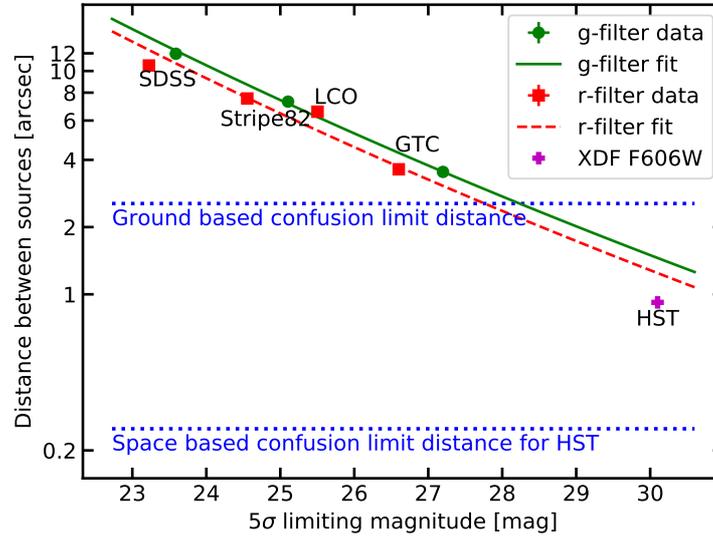


Figure 5.4: Typical distance between sources as a function of the limiting magnitude for all datasets and filters explored in this work. Measured data are shown as circles and squares, while fitted curves are shown with solid and dashed lines, for  $g$  and  $r$ -filter, respectively. Ground-based and space-based ( $HST$ ) confusion limit distances are plotted as horizontal dashed lines. Data for the space based measurement comes from the Xtreme Deep Field (XDF) of the  $HST$  using the filter F606W (Illingworth et al. 2013).

The situation improves, of course, with space-based instrumentation. In space-based deep images, the FWHM seeing is much smaller than ground-based observations because they are not affected by the atmosphere blurring. As a consequence, the confusion limit decreases toward much smaller distances between sources (see Figure 5.4).

## 5.5 Conclusions

In an era where ultradeep optical imaging is crossing the 30 mag arcsec<sup>-2</sup> surface brightness frontier, a natural question is how reliable the faintest features one that we see in those images. Most of the images at such depth are unique (as the time investment for getting them is enormous) and consequently hardly replicable by different teams. Here, we have shown the results of ongoing work in which we have studied the same galaxy (NGC 493) using different telescopes (SDSS, Stripe 82, LCO, and GTC). We have explored the replicability of the faintest surface brightness features on those images, obtaining consistent results from datasets of different telescopes and photometric bands (see Figure 5.1 and 5.2). Thus, demonstrating that the current technology

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

allows us to safely recover features on the  $31 \text{ mag arcsec}^{-2}$  regime ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ). Still, it is necessary to conduct a quantitative study by obtaining the photometry of the very faint stellar halo of the galaxy NGC 493, as well as other objects on the images to extract their physical parameters.

In this chapter, we have compared the results obtained from different facilities for the same region of the sky (NGC 493 and its surroundings). In this sense, it is important to remark that the values provided here are not absolutes but particular for this region of the sky. For example, the typical distance between sources (Figure 5.3) will be different for other regions of the sky (in a more crowded region with a galaxy cluster or towards the central region of the Galactic disk, the averaged distance between sources will be smaller than the ones obtained in this work). However, the relevant point raised here is to analyze how this parameter varies as a function of the depth of the images, and study what is the limit of the current ground-based telescopes.

In pushing the analysis of the images to their limits, it is extremely important to quantify the systematics limits because they provide valuable information regarding how to design the observational strategy, reduce the raw data, and analyze the datasets. For example, obtaining an auto flat field correction or making the sky background subtraction on individual images of 120s of exposure time with the GTC (as in this work) is not trivial because there are detectable objects in almost all pixels of the images. The current large area surveys such as the SDSS or the next generation of wide surveys such as the LSST will not be in trouble. However, the situation would become more difficult with the advent of larger telescopes in the near future. With the arrival of larger telescopes such as the TMT or the ELT, it will be necessary to take into account the limits discussed above when performing very deep observations to be able to reduce the data properly without removing the low surface brightness structures.

The reproducibility (obtaining the same outputs using the same inputs) and replicability (obtaining consistent results using different inputs) of results is fundamental in science. Setting a good example in this context, the reduction of all data, as well as the high-level analysis presented in this chapter, have been done under Maneage (see Chapter 2 for further details).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 6

## Ongoing and future work

Throughout this thesis we have presented a large amount of technical work, which will be key for exploiting the next generation of deep surveys. In this chapter, we show some practical examples of all our tools being applied to ongoing and future works. They correspond mainly to projects that have been started during the thesis period, but that have not yet been published. The first one corresponds to the use of an amateur telescope to obtain ultra-deep imaging of the galaxy M101, exploring the possibility of using that technology. The second is a comparison of star count and integrated light photometry techniques for exploring the faint outskirts of nearby galaxies. In this case, ultra-deep imaging using PAUCam at the William Herschel Telescope of the galaxy NGC 4565 is used to compare both techniques and check if they agree.

### 6.1 Ultra-deep imaging with amateur telescopes

All over the world, amateur astronomers take striking pictures of celestial objects. However, obtaining data that can be used for scientific research is hard and requires a lot of dedication (see e.g., [Martínez-Delgado et al. 2010](#)). The reason is that it needs a carefully designed observing strategy (including calibration data such as bias, darks, and flat field exposures) as well as a good reduction and analysis of the data. In particular, when trying to study the low surface brightness Universe on large scales (e.g., sources with an apparent size of tens of arcminutes or more, like nearby galaxy halos, tidal streams, Galactic cirrus) the main limitation of the current professional facilities is the small field of view of the instruments. Telescopes with small apertures could be better suited for this purpose because they normally have a wider field of view and, in general, they have a single CCD, which makes the data reduction significantly easier. This is the context in which the collaboration between professional (Raúl Infante-Sainz and Ignacio Trujillo) and amateur (Aleix Roig) astronomers was born. With the goal of testing the possibilities of using amateur telescopes for low surface brightness studies, we focus on the galaxy M101. This galaxy is a good target since it has been observed in depth by professional astronomers (see e.g., [Mihos et al. 2018](#)) so we can compare our results with those deep observations.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 6.1.1 Telescope and observing strategy

The amateur telescope we have used is shown in Figure 6.1. It has the following characteristics; mount: MESU 200, telescope: FSQ85ED, camera: ASI1600MM-Cooled, guiding: OAG, software: Sequence Generator Pro, aperture diameter: 8.5 cm, focal ratio:  $f/5.3$ , field of view:  $2.25 \times 1.70$  deg, pixel scale: 1.74 arcsec/pixel, filter: Baader Luminance 31 mm.



Figure 6.1: Picture of the amateur telescope used in this work. It is located in Prades, Catalunya, Spain.

This telescope is located in a small town called Prades (in Catalunya, Spain), where typical observing conditions during dark nights are sky brightness of  $21 \text{ mag arcsec}^{-2}$  and a seeing of between 1.5 and 2.5 arcsec.

As the goal of this work is to reach the limits of amateur telescopes for low surface brightness science, a good understanding of the systematics of the instrument is mandatory. In this sense, we took bias and dark exposures as calibration data. Note that taking dark images nowadays is not very common because the cameras are expected to be cold enough (e.g., typical temperature of CCD cameras for professional telescopes is about  $-90^\circ\text{C}$ ) to make the dark current contribution almost negligible. However, the cooling system of this amateur telescope is only able to cool the camera down to  $-20^\circ\text{C}$ . Consequently, the dark current at this temperature is not negligible and must be accounted for. Figure 1.4 (Chapter 1) shows the master bias and

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585      Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

dark images of this telescope.

The observing strategy we follow consists of a dithering and rotation pattern with the aim of not having the same pixel of the camera observing the same part of the galaxy more than once (see e.g., [Trujillo & Fliri 2016](#)). This observing strategy makes it possible to do flat field correction using the data itself. Dome or twilight flat field exposures have structures and gradients (i.e., they are not perfectly flat, see Section [1.3.4](#)). This issue becomes more important as the field of view of the camera gets larger. Consider, for example, a very large field of view camera of some square degrees. If a twilight exposure for flat field is taken, that image will have a gradient with those pixels near the location of the Sun having more counts. Taking into account all the above, the best option is to use a flat field correction based on the brightness of the sky using the data itself. See Section [1.3.4](#) for further details.

### 6.1.2 Data reduction

We have developed a pipeline for processing the data of this telescope that makes use of `Maneage` (see Chapter [2](#); [Akhlaghi, Infante-Sainz et al. 2020](#)). To process the data, GNU Astronomy Utilities (`Gnuastro`; [Akhlaghi & Ichikawa 2015](#)) are also widely used. Data reduction consists of correcting all the systematics of the instrument (bias level, dark current, flat field) as well as the sky background subtraction, astrometric solution, and photometric calibration (see Section [1.4](#) for more on the data reduction process).

As the main goal is to detect and characterize the low surface brightness structures, it is really important to prevent the destruction of these features during the sky background correction. To do this, the pipeline takes special care in the detection of low surface brightness structures using `NoiseChisel` ([Akhlaghi & Ichikawa 2015](#); [Akhlaghi 2019](#)) on each individual image. After masking those pixels, a 2-dimensional polynomial function of degree 1 is fitted to the non-masked pixels. Modeling the sky background with such a low degree of freedom makes it possible to correct potential structures over the entire image due to sky gradients, but at the same time, the low polynomial degree prevents locally fitting low surface brightness structures and removing them.

The astrometric solution of each individual exposure is obtained by using the software `Astrometry` ([Lang et al. 2010](#)). We use Gaia DR2 ([Gaia Collaboration et al. 2018](#)) as the reference catalog (objects around the galaxy M101 (R.A. (J2000) = 14:03:12.50; dec (J2000) = +54:20:53.7) within a box of  $2 \times 2$  deg<sup>2</sup>). Then, the catalog is split into different catalog indexes using the program `build-astrometry-index`. By doing this, the main catalog is divided into different catalogs that the program which computes the astrometry solution can read as inputs. Finally, the program `solve-field` is run over each individual image in order to obtain the astrometric solution.

The photometric calibration is done using SDSS Data Release 13 ([Albaret et al. 2017](#)) in the *r*-band because this is the filter that is more similar to the transmission curve of the luminance filter that has been used for the observations. By doing the photometry of the stars in the individual images and matching them with the SDSS catalog, we obtain the factor by which we need to multiply each individual image to have the same photometric calibration as the SDSS survey. Once the photometric calibration has been done, each image is resampled into a new pixel grid using `SWarp` ([Bertin 2010](#)). Finally, all images are stacked using a  $3\sigma$  clipping median pixel per pixel to obtain the coadded image of the M101 galaxy.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### 6.1.3 Results and conclusions

The final stacked image of M101 has a limiting surface brightness of  $27.8 \text{ mag arcsec}^{-2}$  ( $r$ -band;  $3\sigma$  in  $10 \times 10 \text{ arcsec}^2$  boxes). In Figure 6.2, we compare our result with other two deep observations carried out by F. Neyer<sup>1</sup> (amateur astronomer) and Mihos et al. (2018).

As it can be seen in Figure 6.2, F. Neyer obtained similar results with a larger amateur telescope but less exposure time. The work done by Mihos et al. (2018) using a professional and larger telescope got a deeper image of M101, although the total exposure time is lower. In that work, they have also corrected the scattered light produced by the stars by modeling them using the PSF (see Section 3.3.2). Consequently, the low surface brightness structures around the galaxy can be studied in more detail. Our results demonstrate that current amateur telescopes can be used for frontier low surface brightness science if a good observing strategy and data processing is done.

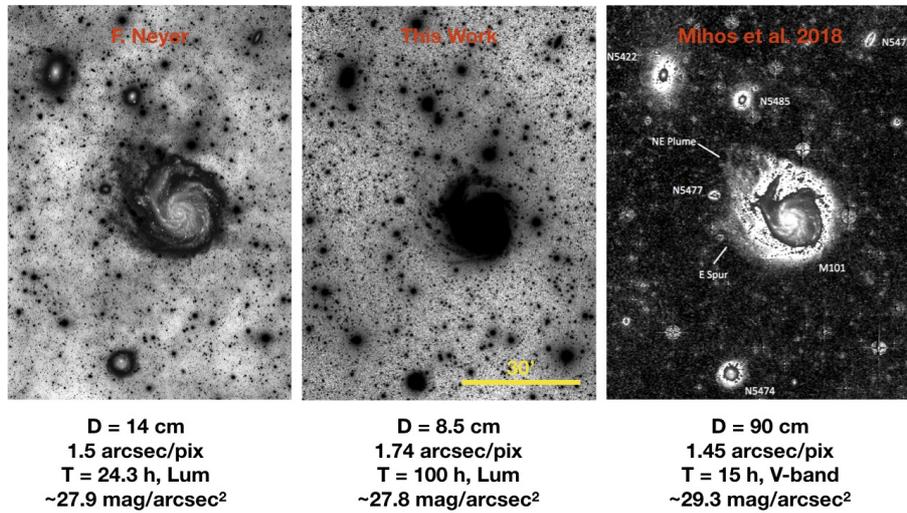


Figure 6.2: Left: image obtained by F. Neyer with an amateur telescope. Center: image obtained in this work with an amateur telescope. Right: image obtained by Mihos et al. (2018) with a professional telescope. The characteristics and details of the telescopes and observations are shown below each image.

We have analyzed the surface brightness limiting magnitude ( $\mu_{\text{limit}}$ ) as a function of the total exposure time (i.e., we have computed the limiting magnitude as a function of the number of images used for the final stacked image) and we have found that it follows the theoretical prediction ( $\mu_{\text{limit}} \propto \sqrt{\text{exposure time}}$ ). As a consequence, we can say that as far as we have checked, this telescope is behaving as expected. This is a work in progress, and more extensive

<sup>1</sup>Fabian Neyer personal webpage: <http://www.starpointing.com/ccd/m101.html>

|  |                                  |
|--|----------------------------------|
| Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.<br>Su autenticidad puede ser contrastada en la siguiente dirección <a href="https://sede.ull.es/validacion/">https://sede.ull.es/validacion/</a> |                                  |
| Identificador del documento: 3258585   | Código de verificación: 8GKKDuAS |
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA  | Fecha: 04/03/2021 11:59:31       |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA   | 04/03/2021 12:00:29              |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA  | 14/04/2021 15:46:33              |

6.1 Ultradeep imaging with amateur telescopes

125

and detailed analysis will be done. The conclusion of this preliminary work is to encourage the community to carry out collaborative projects between amateur and professional astronomers in order to take advantage of the capabilities that amateur telescopes offer, in particular, their large field of view.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

## 6.2 Star counting versus integrated photometry

The star counting technique consists of resolving individual stars with the goal of studying the stellar populations. Thus, it can be used for extracting relevant information such as the age and metallicity. The star counting technique is able to detect extended surface brightness distributions with a brightness as low as  $\sim 32 \text{ mag arcsec}^{-2}$  (see e.g., [Ibata et al. 2009, 2014](#)). However, this technique requires the detection of individual stars in external galaxies, and because of that, it is limited to only a few nearby galaxies, in particular, to those galaxies that are nearer than 16 Mpc when using the *Hubble Space Telescope (HST)* ([Zackrisson et al. 2012](#)). As an alternative, integrated photometry allows us to measure the ages and metallicities of more distant objects. But it is necessary to note that integrated light takes into account the flux of all sources along the line of sight. In this context, a crucial test is to ensure that both techniques give the same results, so it will be safe to use integrated light photometry in those objects where the star counting technique is not feasible. This ongoing work aims to conduct such a test.

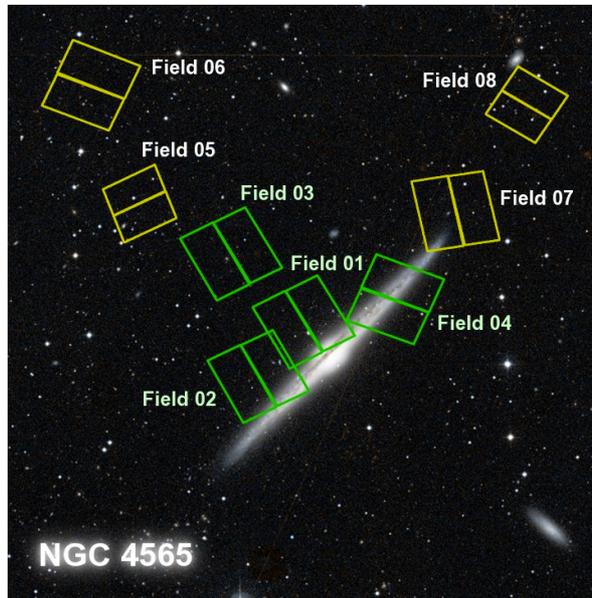


Figure 6.3: Digital Sky Survey (DSS) colored image of NGC 4565, showing the location of the *HST* ACS/WFC and WFC3/UVIS fields. The different fields correspond to the locations of the GHOST survey pointings with the *HST*. The different color (yellow and green) of the regions indicates that the observations come from different runs. North is up and east is to the left. Credit: [Monachesi et al. \(2016\)](#).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

The galaxy NGC 4565, at a distance of 12 Mpc (Tully et al. 1992), has been observed as part of the GHOST survey (Radburn-Smith et al. 2011; Monachesi et al. 2016). They observed this galaxy with the *HST* at different locations to make a radial profile along the halo direction (perpendicular to the plane of the disk), and also along the disk (see Figure 6.3). By resolving individual stars they compute the age and metallicity of the different regions.

### 6.2.1 Data analysis

We took ultradeep imaging data in three filters (*g*, *r*, and *i* bands) with the Physics of the Accelerating Universe Camera (PAUCam; Padilla et al. 2019) at the 4 m William Herschel Telescope (WHT). PAUCam instrument is a camera composed of 18 CCDs. Each CCD has four channels. PAUCam has a large field of view ( $\sim 1^\circ \times 1^\circ$ ) that makes it possible to observe the galaxy NGC 4565 and its surrounding. The total amount of time observed in each filter was 4.8, 3.1, and 4.4 hours for the *g*, *r*, and *i*-bands, respectively. Very bright stars were also observed with the aim of constructing the extended PSF and correct the scattered light field.



Figure 6.4: Left: color composite image taken with the WHT of NGC 4565 using the *g*, *r*, and *i* filters. Right: scattered light field model obtained using the extended PSFs. Only stars brighter than 13 mag in the Gaia DR2 catalog have been considered. The size of each image is  $\sim 1^\circ \times 1^\circ$ . North is up and east is to the left.

The processing of the raw data was done as explained in Section 1.4. The observational strategy consisted of dithering and rotation of the camera with the aim of correct systematic effects and auto flat field correction (see e.g., Trujillo & Fliri 2016). After the correction of all systematic effects as well as the photometric and astrometric calibration, the final co-addition of all images results in very deep images of the galaxy NGC 4565.

In addition to the correction of the scattered light field due to the bright stars, the scattered

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

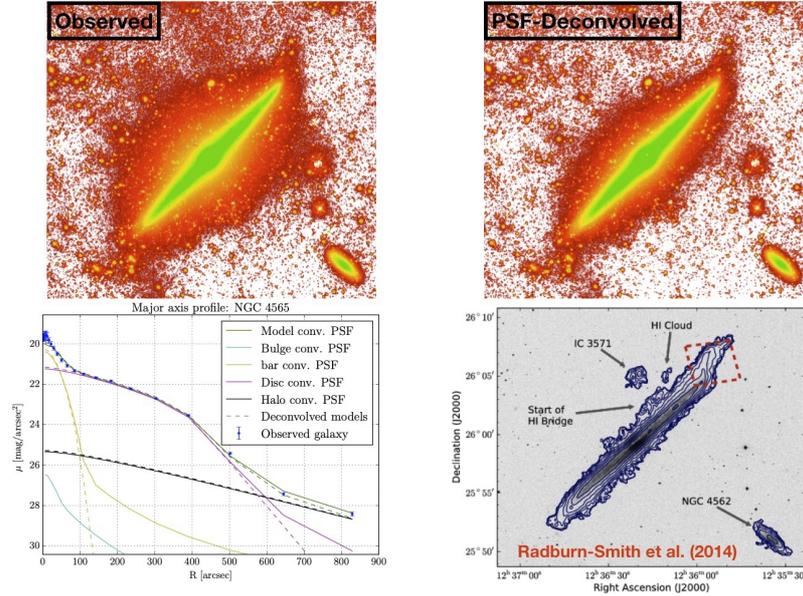


Figure 6.5: Modeling the effect of the PSF on NGC 4565. Top left: observed image of the galaxy NGC 4565 once the scattered light field model produced by the bright stars has been subtracted. Top right: image of the galaxy NGC 4565 once it has been deconvolved by the effect of the PSF. Note how the amount of light in the outskirts of the “deconvolved” image is much smaller. Thus, indicating that the effect of the PSF is extremely important. Bottom left: profiles of the different components considered for modeling the galaxy (credit: Cristina Martínez-Lombilla). Bottom right:  $H_I$  map from Radburn-Smith et al. (2014). The region enclosed with the red dashed rectangle correspond to the warp structure at the edge of the disk.

light from the galaxy NGC 4565 was also modeled and corrected. This was done by modeling the galaxy using `Imfit`, see Figure 6.5. This kind of deconvolution is necessary because otherwise the physical parameters derived for the galaxy would be affected by the extra flux from the scattered light (see how the width of the galaxy changes after the deconvolution process).

## 6.2.2 Results and conclusions

The final images of NGC 4565 have a surface brightness limiting magnitude of 30.5, 29.9, and 29.3 mag arcsec<sup>-2</sup> for the  $g$ ,  $r$ , and  $i$  filters, respectively ( $3\sigma; 10 \times 10$  arcsec<sup>2</sup>). The depth of the images as well as the extended PSFs characterization allow us to study the outskirts of the galaxy in detail. In this preliminary work, only two small regions in the halo and the warp of the galaxy with the *HST* data have been analyzed. After correcting the scattered light of very

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

bright stars and the galaxy itself, it is possible to make the photometry of these regions. To obtain the age and metallicity, the colors of these regions ( $g-r$  and  $r-i$ ) were compared with the E-MILES stellar population models (Vazdekis et al. 2016). This is shown in Figure 6.6 for the warp region.

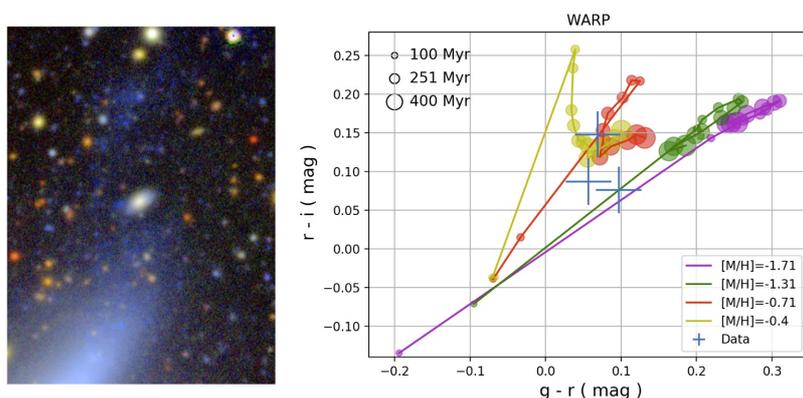


Figure 6.6: Left: zoom to the warp region at the edge of the NGC 4565 disk (color image). The warp can be identified as an extended and diffuse structure in blue. Right: color-color diagram of data measured from the warp and stellar population models (Vazdekis et al. 2016) corresponding to some ages and metallicities. By comparing the data with the models, it is possible to obtain a rough estimation of the age and metallicity of the warp.

According to the color values obtained for the warp, the stellar population of this region is young (ages between 100 and 400 Myr) and metal-poor with  $[M/H]$  between  $-1.71$  and  $0.4$ . Using the star count techniques, Radburn-Smith et al. (2014) found similar values: age  $< 600$  Myr and metallicity  $[M/H] \sim -1$ . For the stellar halo region, the agreement between the two different techniques is also rather good. From integrated photometry and comparing the measured colors with stellar population models, the age is between 2 and 10 Gyr, while the metallicity  $[M/H]$  is between  $-1.31$  and  $-0.71$ . The values from the GHOST survey for the same location of the stellar halo (Monachesi et al. 2016) correspond to a structure of age  $\sim 10$  Gyr and metallicity  $[M/H] \sim -1.2$ .

It is important to recall that this is a work in progress. The results described here are preliminary, and the analysis of the data will be revised to check if everything has been done properly. Nonetheless, they indicate that the star counting technique and deep integrated photometry are in agreement. Thus, it is possible to use the integrated photometry technique in those cases where the star counting technique is not feasible. In practice, this will happen in the vast majority of astronomical objects because resolving individual stars is only possible for a few nearby objects.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
 UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
 UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
 UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# 7

## Summary and conclusions

The study of the low surface brightness Universe is a source of discoveries, a way of increasing fundamental knowledge, and also pushing the current technology beyond its limits. In astronomy, it is almost impossible to do laboratory experiments to test the predictions from theoretical models. Therefore, it is crucial to observe and obtain accurate data in order to analyze and extract physical information to compare with the predictions from theoretical models.

In this thesis, a very technical and detailed study of the low surface brightness Universe has been done. The importance of studying the low surface brightness Universe is explained in the introduction in Chapter 1. In that chapter, the main challenges (sky background correction, flat fielding, point spread function, among others) are discussed. The next chapters are dedicated to accounting for these challenges.

Starting with the general challenge of reproducing scientific results, Chapter 2 introduces Maneage as a framework to carry out reproducible research. The problem of the point spread function (PSF) is fully explained in Chapter 3, where the Sloan Digital Sky Survey (SDSS) is used to construct the extended PSF and correct the scattered light field of very bright stars in the Coma cluster region. In that chapter, detailed information is given with the aim of being useful for future wide surveys. The construction of the PSFs is entirely done using Maneage. Chapter 4 is dedicated to the study of the galaxy NGC 1052-DF4. This galaxy was claimed to be devoid of dark matter, thus creating a strong tension with the current  $\Lambda$ CDM model. By using very deep imaging data from several facilities and applying techniques to preserve and study the low surface brightness features, we discovered a very faint tidal stream. The interaction of this galaxy with another neighbor galaxy naturally explains its lack of dark matter. In Chapter 5, data from different telescopes and instruments were used to make a study of the outskirts of the galaxy NGC 493. This study shows the analysis of how the depth of the astronomical data can affect the physical parameters measured for astronomical objects. Finally, in Chapter 6, ongoing and future low surface brightness studies are shown. These studies correspond to work in progress so the results are preliminary. They correspond to the use of amateur telescopes for low surface brightness, and a comparison of the star counting techniques with the integrated light photometry. Chapter 7 corresponds to the summary and conclusions of the thesis.

In what follows, the main technical conclusions obtained during the development of this thesis are summarized.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- **The reproducibility problem.** Science is characterized by its method, not by particular results or discoveries. Having a framework that allows full control over the data lineage as well as the software and programs that are used to treat the data is key to rely upon the results. Maneage is a framework that has been demonstrated to account for such matters. It goes to the fundamental concepts and has been constructed following the principles explained in Section 2.3. In practice, Maneage has been used to reduce and analyze the vast majority of the data presented along with this thesis.
- **Observational strategy.** Low surface brightness studies require very careful observations and data processing. The observational strategy has to be carefully designed to be able to correct the systematic effects. In this sense, one of the most important points is to consider cameras with CCD field of view that is larger than the size of the object to be studied. Otherwise, even with big dithering, the auto flat field and sky background correction will not be enough.
- **Data reduction.** Data reduction consists of removing or minimizing the systematic effects to finally have the data ready to be analyzed. Nowadays, bias correction is something standard that does not pose a major challenge. In the same way, the dark current correction is not usually done because cameras are cooled down to low temperatures, thus the dark current noise is negligible. However, it can be an important contribution to be considered, for example, in amateur telescopes, as shown in Section 6.1. Flat field correction is challenging since it can introduce gradients and artificial structures. As a consequence, the best solution so far is to do an auto flat field correction. To do that, the data has to be acquired with large dithering. It is important to be aware that the auto flat field could vary from night to night, with position angle, or with other factors. Thus, it is necessary to check that the auto flat field correction is stable and can be done. The sky background correction is also one of the major challenges in low surface brightness science. To avoid overestimating the sky background, it is absolutely crucial to detect and mask very accurately all objects and the signal from the images. As the low surface brightness structures can be hidden under the image noise, even after the detection and masking of the objects, there will be undetected faint structures on the images. It is absolutely crucial not to take these structures into account during the estimation of the sky background. This can be done by using a constant value as the sky level if there are no gradients on the images. If some gradients or structures need to be corrected, then the function or model to be used should be as smooth as possible to avoid taking into account variations and small local structures.
- **Point spread function.** The PSF is key to have a deep understanding of the telescope and the instrument. Moreover, in low surface brightness studies it is important to characterize the PSF as extended as possible. In Chapter 3, SDSS data were used to develop the necessary techniques to construct the extended (8 arcmin in radius) PSFs of this survey. The modeling of the brightest stars on the Coma cluster region has been done as a practical example of how the PSFs can be used for correcting the scattered light field. These techniques were used in the following chapters to construct the extended PSFs and correct the scattered light field of very bright stars (Chapter 4), and also the galaxy itself (Chapter 6).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- **High-level analysis and measurements.** All technical considerations explained above were considered to reduce the data from different telescopes and instruments with the goal of preserving the low surface brightness structures. In Chapter 4, these techniques were applied to finally discover a very faint tidal stream in the galaxy NGC 1052-DF4. The discovery explained its lack of dark matter and reconciled the nature of this object with the current model of galaxy formation and evolution. In addition to this, it was also necessary to investigate new techniques for quantifying and measuring low surface brightness structures. For example, traditional photometry measurements using circular or elliptical apertures are not suitable for low surface brightness structures that do not have an analytical shape (e.g. Galactic halos, dust cirrus, ultra-diffuse galaxies, etc.). In Chapter 5, the study of the outskirts of the galaxy NGC 493 reveals that the analysis becomes extremely hard as the depth of the dataset increases. To be able to detect, deblend, and obtain accurate measurements from astronomical sources is not a trivial task when the depth of the images reaches  $\mu \sim 31 \text{ mag arcsec}^{-2}$  ( $3\sigma; 10 \times 10 \text{ arcsec}^2$ ). As a consequence, new software and algorithms such as Gnuastro are necessary. The development of high-level measurements is still a work in progress, but its importance is crucial and they will be put into practice in the ongoing and future works (Chapter 6).

Starting from the very beginning by planning the observational strategy, then correcting all the systematic effects, and finally designing new ways of measuring high-level quantities for very faint structures, the present work has helped to understand the low surface brightness regime. In the process, important developments to construct a framework for managing data and its lineage (Maneage) has been developed. Overall, all work presented in this thesis is a contribution to have a better understanding to unveil the low surface brightness Universe.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

## Bibliography

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, AJ, 126, 2081
- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
- Abbott, T. M. C., Adamow, M., Aguena, M., et al. 2021, arXiv e-prints, arXiv:2101.05765
- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
- Abraham, R. G., & van Dokkum, P. G. 2014, PASP, 126, 55
- Aguerri, J. A. L., Girardi, M., Agulli, I., et al. 2020, MNRAS, 494, 1681
- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018a, PASJ, 70, S8
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018b, PASJ, 70, S4
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
- Akhlaghi, M. 2019, IAU Symposium 355, arXiv:1909.11230
- Akhlaghi, M., & Bacon, R. 2018, Zenodo, DOI:10.5281/zenodo.1163746
- Akhlaghi, M., Bacon, R., & Inami, H. 2018, Zenodo, DOI:10.5281/zenodo.1164774
- Akhlaghi, M., & Ichikawa, T. 2015, ApJS, 220, 1
- Akhlaghi, Infante-Sainz, Roukema, B. F., Valls-Gabaud, D., & Baena-Gallé, R. 2020, arXiv e-prints, arXiv:2006.03018
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
- Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25
- Allen, A., Teuben, P. J., & Ryan, P. W. 2018, ApJS, 236, 10
- Alonso Asensio, I., Dalla Vecchia, C., Bahé, Y. M., Barnes, D. J., & Kay, S. T. 2020, MNRAS, 494, 1859
- Amelio, G. F., Tompsett, M. F., & Smith, G. E. 1970, The Bell System Technical Journal, 49, 593

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120
- Anscombe, F. 1973, The American Statistician, 27, 17
- Arp, H., & Bertola, F. 1969, Astrophys. Lett., 4, 23
- . 1971, ApJ, 163, 195
- Bacon, R., Conseil, S., Mary, D., et al. 2017, A&A, 608, A1
- Baggerly, K. A., & Coombes, K. R. 2009, The Annals of Applied Statistics, 3, 1309
- Baker, M. 2016, Nature, 533, 452
- Bavoil, L., Callahan, S. P., Crossno, P. J., et al. 2005, VIS 05. IEEE Visualization, 135
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
- Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints, arXiv:1403.5237
- Bergvall, N., Zackrisson, E., & Caldwell, B. 2010, MNRAS, 405, 2697
- Bertin, E. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 112
- Bertin, E. 2010, SWarp: Resampling and Co-adding FITS Images Together, ascl:1010.068
- Bertin, E., & Arnouts, S. 1996, Astronomy and Astrophysics Supplement, 117, 393
- Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 592, 819
- Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28
- Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517
- Borlaff, A., Trujillo, I., Román, J., et al. 2019, A&A, 621, A133
- Bothun, G. D., Impey, C. D., & Malin, D. F. 1991, ApJ, 376, 404
- Bothun, G. D., Impey, C. D., Malin, D. F., & Mould, J. R. 1987, AJ, 94, 23
- Boyle, W. S., & Smith, G. E. 1970, The Bell System Technical Journal, 49, 587
- Brodie, J. P., & Strader, J. 2006, ARA&A, 44, 193
- Brook, C. B., Di Cintio, A., Knebe, A., et al. 2014, ApJ, 784, L14
- Brown, T. M., Baliber, N., Bianco, F. B., et al. 2013, PASP, 125, 1031
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Buckheit, J. B., & Donoho, D. L. 1995, Wavelets and Statistics, 1, 55

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931
- Capaccioli, M., Spavone, M., Grado, A., et al. 2015, A&A, 581, A10
- Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, A&A, 622, A176
- Cepa, J., Aguiar-Gonzalez, M., Bland-Hawthorn, J., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1739–1749
- Chamba, N., Trujillo, I., & Knapen, J. H. 2020, A&A, 633, L3
- Chang, A. C., & Li, P. 2015, Finance and Economics Discussion Series 2015-083, 1
- Cheney, J., Chiticariu, L., & Tan, W.-C. 2009, Foundations and Trends in Databases, 1, 379
- Claerbout, J. F., & Karrenbach, M. 1992, SEG Technical Program Expanded Abstracts, 1, 601
- Cohen, Y., van Dokkum, P., Danieli, S., et al. 2018, ApJ, 868, 96
- Contini, E., De Lucia, G., Villalobos, Á., & Borgani, S. 2014, MNRAS, 437, 3787
- Cooper, A. P., Cole, S., Frenk, C. S., et al. 2010, MNRAS, 406, 744
- Corbelli, E., & Salucci, P. 2000, MNRAS, 311, 441
- Danieli, S., van Dokkum, P., Abraham, R., et al. 2020, ApJ, 895, L4
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
- de Jong, R. S. 2008, MNRAS, 388, 1521
- de Vaucouleurs, G. 1969, Astrophys. Lett., 4, 17
- de Vaucouleurs, G., de Vaucouleurs, A., & Corwin, J. R. 1976, Second reference catalogue of bright galaxies, 1976, 0
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
- Dhillon, V., Dixon, S., Gamble, T., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Proc. SPIE, 107020L
- Di Cosmo, R., Gruenpeter, M., & Zacchiroli, S. 2018, Proceedings of iPRES 2018, 204.4
- . 2020, Computing in Science & Engineering, 22, 33
- Di Cosmo, R., & Pellegrini, F. 2019, Ouvrir la science
- Disney, M. J. 1976, Nature, 263, 573
- Djorgovski, S. G., Gal, R. R., Odewahn, S. C., et al. 1998, in Wide Field Surveys in Cosmology, ed. S. Colombi, Y. Mellier, & B. Raban, Vol. 14, 89

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Djorgovski, S. G., Mahabal, A., Drake, A., Graham, M., & Donalek, C. 2013, Sky Surveys, ed. T. D. Oswalt & H. E. Bond, 223
- Duc, P. A., Brinks, E., Springel, V., et al. 2000, AJ, 120, 1238
- Duc, P.-A., Cuillandre, J.-C., Karabal, E., et al. 2015, MNRAS, 446, 120
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
- Elvey, C. T., & Roach, F. E. 1937, ApJ, 85, 213
- Erwin, P. 2015, ApJ, 799, 226
- Feldman, S. I. 1979, Journal of Software: Practice and Experience, 9, 255
- Ferguson, H. C., & Sandage, A. 1988, AJ, 96, 1520
- Ferrarese, L., Côté, P., Cuillandre, J.-C., et al. 2012, ApJS, 200, 4
- Ferré-Mateu, A., Alabi, A., Forbes, D. A., et al. 2018, MNRAS, 479, 4891
- Filho, M. E., Winkel, B., Sánchez Almeida, J., et al. 2013, A&A, 558, A18
- Fineberg, H. V., Allison, D. B., Barba, L. A., et al. 2019, The National Academies Press, 1
- Fliri, J., & Trujillo, I. 2016, MNRAS, 456, 1359
- Forbes, D. A. 2017, MNRAS, 472, L104
- Freeman, K. C. 1970, ApJ, 160, 811
- Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338
- Gabriel, A., & Capone, R. 2011, Procedia Computer Science, 4, 577
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
- Gibson, B. K., & Hickson, P. 1992, MNRAS, 258, 543
- Gonzaga, S., Hack, W., Fruchter, A., & Mack, J. 2012, The DrizzlePac Handbook
- Goranova, Y., Hudelot, P., Magnard, F., & et al. 2009, The CFHTLS T0006 Release
- Gronenschild, E. H. B. M., Habets, P., Jacobs, H. I. L., et al. 2012, PLoS ONE, 7, e38234
- Gu, M., Conroy, C., Law, D., et al. 2018, ApJ, 859, 37
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
- Harris, W. E. 1996, AJ, 112, 1487
- Heney, L. G., & Greenstein, J. L. 1941, ApJ, 93, 70

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Herndon, T., Ash, M., & Pollin, R. 2014, Cambridge Journal of Economics, 38, 257
- Hinsen, K. 2015, F1000Research, 3, 289
- Hinsen, K. 2016, The Self Journal of Science, 1: arXiv:1605.02960
- Horvath, S. 2015, Genome Biology, 16, 96
- Howell, s., & Tavackolimehr, A. 2019, Handbook of CCD Astronomy
- Hudson, M. J., & Robison, B. 2018, MNRAS, 477, 3869
- Hughes, M. E., Pfeffer, J., Martig, M., et al. 2019, MNRAS, 482, 2795
- Hugot, E., Wang, X., Valls-Gabaud, D., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, ed. J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, & H. A. MacEwen, 91434X
- Ibata, R., Martin, N. F., Irwin, M., et al. 2007, ApJ, 671, 1591
- Ibata, R., Mouhcine, M., & Rejkuba, M. 2009, MNRAS, 395, 126
- Ibata, R. A., Lewis, G. F., McConnachie, A. W., et al. 2014, ApJ, 780, 128
- Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
- Impey, C., Bothun, G., & Malin, D. 1988, ApJ, 330, 634
- Infante-Sainz, R., Trujillo, I., & Román, J. 2020, MNRAS, 491, 5317
- Ioannidis, J. P. A. 2005, PLoS Medicine, 2, e124
- Ioannidis, J. P. A., Allison, D. B., Ball, C. A., et al. 2009, Nature Genetics, 41, 149
- Ivezić, Ž., Goldston, J., Finlator, K., et al. 2000, AJ, 120, 963
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
- Jackson, R. A., Kaviraj, S., Martin, G., et al. 2021, MNRAS, 502, 1785
- Jedrzejewski, R. I. 1987, MNRAS, 226, 747
- Jiang, L., Fan, X., Bian, F., et al. 2014, ApJS, 213, 12
- Johnston, K. V., Bullock, J. S., Sharma, S., et al. 2008, ApJ, 689, 936
- Johnston, K. V., Choi, P. I., & Guhathakurta, P. 2002, AJ, 124, 127
- Johnston, K. V., Hernquist, L., & Bolte, M. 1996, ApJ, 465, 278
- Jouvel, S., Kneib, J. P., Bernstein, G., et al. 2011, A&A, 532, A25
- Kadowaki, J., Zaritsky, D., & Donnerstein, R. L. 2017, ApJ, 838, L21

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Karabal, E., Duc, P.-A., Kuntschner, H., et al. 2017, A&A, 601, A86
- Katz, D. S. 2014, Journal of Open Research Software, 2, e20
- Kim, S. Y., Peter, A. H. G., & Hargis, J. R. 2018, Phys. Rev. Lett., 121, 211302
- Klimentowski, J., Lokas, E. L., Kazantzidis, S., et al. 2009, MNRAS, 400, 2162
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
- Knapen, J. H., & Trujillo, I. 2017, Ultra-Deep Imaging: Structure of Disks and Haloes, ed. J. H. Knapen, J. C. Lee, & A. Gil de Paz, Vol. 434, 255
- Kniazev, A. Y., Grebel, E. K., Pustilnik, S. A., et al. 2004, AJ, 127, 704
- Koch, A., Burkert, A., Rich, R. M., et al. 2012, ApJ, 755, L13
- Koda, J., Yagi, M., Yamanoi, H., & Komiyama, Y. 2015, ApJ, 807, L2
- Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, ApJS, 209, 3
- Kormendy, J., & Bahcall, J. N. 1974, AJ, 79, 671
- Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, AJ, 139, 1782
- Lasker, B. M., McLean, B. J., Shara, M. M., et al. 1989, Bulletin d'Information du Centre de Données Stellaires, 37, 15
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
- Leek, J. T., & Jager, L. R. 2017, Annual Review of Statistics and Its Application, 4, 109
- Lelli, F., Duc, P.-A., Brinks, E., et al. 2015, A&A, 584, A113
- Lesser, M. 2015, PASP, 127, 1097
- Lokas, E. L. 2020, A&A, 638, A133
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
- Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. J. Harnden, F. R., F. A. Primini, & H. E. Payne, 269
- Macciò, A. V., Prats, D. H., Dixon, K. L., et al. 2021, MNRAS, 501, 693
- Mackey, A. D., Huxor, A. P., Ferguson, A. M. N., et al. 2010, ApJ, 717, L11
- Madau, P., Diemand, J., & Kuhlen, M. 2008, ApJ, 679, 1260
- Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, A&A, 570, A13
- Malik, T., Pham, Q., & Foster, I. 2013, Implementing Reproducible Research, Chapter 2, 1.  
URL: <https://osf.io/ns2m3>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Malin, D. F. 1978, *Nature*, 276, 591
- . 1981, *AAS Photo Bulletin*, 27, 4
- Malin, D. F., & Carter, D. 1980, *Nature*, 285, 643
- Martin, N. F., Ibata, R. A., Irwin, M. J., et al. 2006, *MNRAS*, 371, 1983
- Martínez-Delgado, D., Peñarrubia, J., Gabany, R. J., et al. 2008, *ApJ*, 689, 184
- Martínez-Delgado, D., Gabany, R. J., Crawford, K., et al. 2010, *AJ*, 140, 962
- Martínez-Delgado, D., Läsker, R., Sharina, M., et al. 2016, *AJ*, 151, 96
- Martínez-Lombilla, C., & Knapen, J. H. 2019, *A&A*, 629, A12
- Martínez-Lombilla, C., Trujillo, I., & Knapen, J. H. 2019, *MNRAS*, 483, 664
- McConnachie, A. W., Huxor, A., Martin, N. F., et al. 2008, *ApJ*, 688, 1009
- McGraw, J. T., Angel, J. R. P., & Sargent, T. A. 1980, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 264, *Conference on Applications of Digital Image Processing to Astronomy*, ed. D. A. Elliott, 20–28
- McIlroy, M. D., Pinson, E. N., & Tague, B. A. 1978, *Bell System Technical Journal*, 57, 6, [ark:/13960/t0gt6xf72](https://doi.org/10.1002/t0gt6xf72)
- McLean, I. S. 2002, *Experimental Astronomy*, 14, 25
- Menke, J., Roelandse, M., Ozyurt, B., Martone, M., & Bandrowski, A. 2020, *bioRxiv*, 2020.01.15.908111
- Merritt, A., van Dokkum, P., & Abraham, R. 2014, *ApJ*, 787, L37
- Merritt, A., van Dokkum, P., Abraham, R., & Zhang, J. 2016, *ApJ*, 830, 62
- Mihos, J. C., Durrell, P. R., Feldmeier, J. J., Harding, P., & Watkins, A. E. 2018, *ApJ*, 862, 99
- Mihos, J. C., Harding, P., Feldmeier, J., & Morrison, H. 2005, *ApJ*, 631, L41
- Mihos, J. C., Harding, P., Feldmeier, J. J., et al. 2017, *ApJ*, 834, 16
- Miller, G. 2006, *Science*, 314, 1856
- Minkowski, R. L., & Abell, G. O. 1963, *The National Geographic Society-Palomar Observatory Sky Survey*, ed. K. A. Strand, 481
- Monachesi, A., Bell, E. F., Radburn-Smith, D. J., et al. 2016, *MNRAS*, 457, 1419
- Monelli, M., & Trujillo, I. 2019, *ApJ*, 880, L11
- Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, *AJ*, 125, 984

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Montes, M., Acosta-Pulido, J. A., Prieto, M. A., & Fernández-Ontiveros, J. A. 2014, MNRAS, 442, 1350
- Montes, M., & Trujillo, I. 2014, ApJ, 794, 137
- . 2018, MNRAS, 474, 917
- . 2019, MNRAS, 482, 2838
- Montes, Infante-Sainz, Madrigal-Aguado, A., Román, J., et al. 2020, ApJ, 904, 114
- Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJ, 524, L19
- Moreau, L., Ludscher, B., Altintas, I., et al. 2008, Concurrency Computation: Practice and Experiment, 20, 473
- Muñoz, R. P., Puzia, T. H., Lançon, A., et al. 2014, ApJS, 210, 4
- Müller, O., Rich, R. M., Román, J., et al. 2019, A&A, 624, L6
- Müller, O., Marleau, F. R., Duc, P.-A., et al. 2020, A&A, 640, A106
- Muslimov, E., Valls-Gabaud, D., Lemaître, G., et al. 2017, Appl. Opt., 56, 8639
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
- Nilson, P. 1973, Nova Acta Regiae Soc. Sci. Upsaliensis Ser. V, 0
- Nusser, A. 2020, ApJ, 893, 66
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
- Ogiya, G. 2018, MNRAS, 480, L106
- Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
- Padilla, C., Castander, F. J., Alarcón, A., et al. 2019, AJ, 157, 246
- Patat, F. 2003, A&A, 400, 1183
- Peñarrubia, J., Navarro, J. F., & McConnachie, A. W. 2008, ApJ, 673, 226
- Peñarrubia, J., Navarro, J. F., McConnachie, A. W., & Martin, N. F. 2009, ApJ, 698, 222
- Peebles, P. J., & Ratra, B. 2003, Reviews of Modern Physics, 75, 559
- Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, A&A, 524, A42
- Peper, M., & Roukema, B. F. 2020, arXiv e-prints, arXiv:2010.03742
- Peters, S. P. C., van der Kruit, P. C., Knapen, J. H., et al. 2017, MNRAS, 470, 427
- Pierce, M., Brodie, J. P., Forbes, D. A., et al. 2005, MNRAS, 358, 419

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, MNRAS, 475, 648
- Plesser, H. E. 2018, Frontiers in Neuroinformatics, 11, 76
- Ploekinger, S., Recchi, S., Hensler, G., & Kroupa, P. 2015, MNRAS, 447, 2512
- Ploekinger, S., Sharma, K., Schaye, J., et al. 2018, MNRAS, 474, 580
- Pohlen, M., & Trujillo, I. 2006, A&A, 454, 759
- Purcell, C. W., Bullock, J. S., & Zentner, A. R. 2007, ApJ, 666, 20
- Radburn-Smith, D. J., de Jong, R. S., Seth, A. C., et al. 2011, ApJS, 195, 18
- Radburn-Smith, D. J., de Jong, R. S., Streich, D., et al. 2014, ApJ, 780, 105
- Read, J. I., & Erkal, D. 2019, MNRAS, 487, 5799
- Read, J. I., Wilkinson, M. I., Evans, N. W., Gilmore, G., & Kleyna, J. T. 2006, MNRAS, 367, 387
- Reshetnikov, V. P. 2005, Physics Uspekhi, 48, 1109
- Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, ApJS, 152, 163
- Roberts, K. 1969, Computer Physics Communications, 1, 1
- Roediger, J. C., & Courteau, S. 2015, MNRAS, 452, 3209
- Román, J., & Trujillo, I. 2018, Research Notes of the American Astronomical Society, 2, 144
- Román, J., Trujillo, I., & Montes, M. 2020, A&A, 644, A42
- Romine, C. H. 2015, Federal Information processing standards publication, 180, 4
- Roukema, B. F. 2020, arXiv e-prints, arXiv:2007.11779
- Rubin, V. C., & Ford, W. Kent, J. 1970, ApJ, 159, 379
- Rudick, C. S., Mihos, J. C., Harding, P., et al. 2010, ApJ, 720, 569
- Rudick, C. S., Mihos, J. C., & McBride, C. K. 2011, ApJ, 732, 48
- Ruiz-Lara, T., Beasley, M. A., Falcón-Barroso, J., et al. 2018, MNRAS, 478, 2034
- Saifollahi, T., Trujillo, I., Beasley, M. A., Peletier, R. F., & Knapen, J. H. 2020, MNRAS, arXiv:2006.14630
- Sandage, A. 1976, AJ, 81, 954
- Sandage, A., & Binggeli, B. 1984, AJ, 89, 919
- Sandin, C. 2014, A&A, 567, A97

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
*Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>*

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- . 2015, *A&A*, 577, A106
- Schaffer, J. 2015, *Australasian Journal of Philosophy*, 93, 644
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schombert, J. M., & Bothun, G. D. 1988, *AJ*, 95, 1389
- Schwab, M., Karrenbach, N., & Claerbout, J. 2000, *Computing in Science and Engineering*, 2, 61
- Sérsic, J. L. 1968, *Atlas de galaxias australes*
- Shen, S., Mo, H. J., White, S. D. M., et al. 2003, *MNRAS*, 343, 978
- Shen, Z., van Dokkum, P., & Danieli, S. 2020, *arXiv e-prints*, arXiv:2010.07324
- Slater, C. T., Harding, P., & Mihos, J. C. 2009, *PASP*, 121, 1267
- Smart, A. 2018, *Physics Today*, Aug, DOI:10.1063/PT.6.1.20180822a
- Smith, A. M., Katz, D. S., & Niemeyer, K. E. 2016, *PeerJ Computer Science*, 2, e86
- Smith, R., Choi, H., Lee, J., et al. 2016, *ApJ*, 833, 109
- Smith, R., Sánchez-Janssen, R., Fellhauer, M., et al. 2013, *MNRAS*, 429, 1066
- Sollima, A., Gil de Paz, A., Martínez-Delgado, D., et al. 2010, *A&A*, 516, A83
- Sorce, J. G., Tully, R. B., Courtois, H. M., et al. 2014, *MNRAS*, 444, 527
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, *MNRAS*, 391, 1685
- Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, *ApJS*, 160, 149
- Stevens, R., Glover, K., Greenhalgh, C., et al. 2003, *Proceedings of UK e-Science All Hands Meeting*, 43
- Stodden, V., Seiler, J., & Ma, Z. 2018, *Proceedings of the National Academy of Sciences*, 115, 2584
- Stoehr, F., White, S. D. M., Tormen, G., & Springel, V. 2002, *MNRAS*, 335, L84
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, *AJ*, 123, 485
- Tal, T., & van Dokkum, P. G. 2011, *ApJ*, 731, 89
- Tange, O. 2018, *Zenodo*, 1146014, DOI:10.5281/zenodo.1146014
- Taylor, M. A., Puzia, T. H., Muñoz, R. P., et al. 2017, *MNRAS*, 469, 3444
- Teeninga, P., Moschini, U., Trager, S. C., & Wilkinson, M. H. 2016, *Mathematical Morphology - Theory and Applications*, 1, doi:doi:10.1515/mathm-2016-0006

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, Phys. Rev. D, 69, 103501
- The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro
- Toloba, E., Lim, S., Peng, E., et al. 2018, ApJ, 856, L31
- Toomre, A., & Toomre, J. 1972, ApJ, 178, 623
- Trujillo, I., Aguerri, J. A. L., Cepa, J., & Gutiérrez, C. M. 2001a, MNRAS, 321, 269
- . 2001b, MNRAS, 328, 977
- Trujillo, I., & Bakos, J. 2013, MNRAS, 431, 1121
- Trujillo, I., Chamba, N., & Knapen, J. H. 2020, MNRAS, 493, 87
- Trujillo, I., & Fliri, J. 2016, ApJ, 823, 123
- Trujillo, I., Beasley, M. A., Borlaff, A., et al. 2019, MNRAS, 486, 1192
- Truong, P. N., Newman, A. B., Simon, J. D., et al. 2017, ApJ, 843, 37
- Tully, R. B., Shaya, E. J., & Pierce, M. J. 1992, ApJS, 80, 479
- Uson, J. M., Boughn, S. P., & Kuhn, J. R. 1991, ApJ, 369, 46
- Valls-Gabaud, D., & MESSIER Collaboration. 2017, in IAU Symposium, Vol. 321, Formation and Evolution of Galaxy Outskirts, ed. A. Gil de Paz, J. H. Knapen, & J. C. Lee, 199–201
- van Dokkum, P., Danieli, S., Abraham, R., Conroy, C., & Romanowsky, A. J. 2019, ApJ, 874, L5
- van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Nature, 555, 629
- van Dokkum, P. G., Abraham, R., Merritt, A., et al. 2015, ApJ, 798, L45
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, MNRAS, 463, 3409
- Wright, J. F., & Mackay, C. D. 1981, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 290, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 160
- Xin, B., Ivezić, Ž., Lupton, R. H., et al. 2018, AJ, 156, 222
- Yang, D., Yu, H.-B., & An, H. 2020, Phys. Rev. Lett., 125, 111105
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579
- Zackrisson, E., de Jong, R. S., & Micheva, G. 2012, MNRAS, 421, 190
- Zhao, J., Gomez-Perez, J. M., Belhajjame, K., et al. 2012, IEEE 8th International Conference on E-Science, 1

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

- Zibetti, S., White, S. D. M., & Brinkmann, J. 2004, MNRAS, 347, 556
- Ziemann, M., Eren, Y., & El-Osta, A. 2016, Genome Biology, 17, 177
- Zwicky, F. 1933, Helvetica Physica Acta, 6, 110
- . 1951, PASP, 63, 61

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

# A

## Appendix: Maneage tutorial

### A.1 Maneage tutorial for dummies

After reading Chapter 2, a particular researcher may have the following thought: “Alright, there is a problem about reproducibility, and I understand that there is a good solution that I could use. But, how can I start using Maneage in my own research?” The answer is in this tutorial.

Here, through this hands-on tutorial, the reader will be able to follow a very easy practical example, step by step, in order to start learning how to use Maneage.

From my own experience (and those of many people that have already started to use Maneage), reading the documentation books of programs and software is hard at the very beginning. However, following practical tutorials with basic examples is much easier and the concepts are much better diggested. Later, with the acquired knowledge and experience, anyone will be able to include new tools, more efficient solutions to the problems, methodologies, and so on. It is important to recall that Maneage will evolve, and consecuntly, this tutorial and the general documentation could be modified in the future. The last versions of the tutorial and the Maneage documentation is always up to date in its web site: <https://maneage.org/>.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

## Maneage tutorial

Copyright (C) 2020 Raul Infante-Sainz [infantesainz@gmail.com](mailto:infantesainz@gmail.com)  
Copyright (C) 2020 Mohammad Akhlaghi [mohammad@akhlaghi.org](mailto:mohammad@akhlaghi.org)  
See the end of the file for license conditions.

This document is a tutorial in which it is described how Maneage (management + lineage) works in practice. It is highly recommended to read the `README-hacking.md` in order to have a clear idea of what is this project about. Actually, in this tutorial it is assumed you have the project already set up and working properly. In order to do it, please, read and follow all the steps described in the sections `Customization checklist` up to the section `Title, short description and author` (including the last one).

With the current tutorial, the reader will be able to have a fully reproducible paper describing a small research example carried out step by step. The research example is very simple: it will consist in analyse a dataset with two columns (time and population). The analysis will be just to make a linear fitting of the data, and then, write the results in a small paragraph into the final paper.

In the following, the tutorial assume you have three different directories. You had to set up them in the configure step:

- `input-directory` : Necessary input data for the project is in this directory.
- `project-directory` : This directory contains the project itself (source codes), it is under `Git` control.
- `build-directory` : Output directory of the project, it is where all the necessary software and the results of the project are saved.

**IMPORTANT NOTE:** the tutorial assume you are always in `project-directory` when considering command lines.

**In short:** this hands on tutorial will guide you through a simple research example in order to show the workflow in Maneage . The tutorial describes by step how to download a small file containg data, analyse the data (by making a linear fitting), and finally write a small paragraph with the fitting parameters into the final paper. All of this will be done in the same Makefile.

### Installing available software: Matplotlib

If all steps above have been done successfully, you are ready to start including your own analysis scripts. But, before that, let's install `Matplotlib` Python package, which will be used later in the analysis of the data when obtaining the linear fit figure. This Python package will be used as an example on how to install programs that are already available in Maneage . Just open the Makefile `reproduce/software/config/installation/TARGETS.mk` and add to the `top-level-python` line, the word `matplotlib` .

```
# Python libraries/modules.  
top-level-python = astropy matplotlib
```

After that, run the configure step again with the option `-e` to continue using the same configuration options given before (input and build directories). Also, run the prepare and make steps:

```
$/project configure -e  
$/project prepare  
$/project make
```

# Open 'paper.pdf' and see if everything is fine. Note that now, 'Matplotlib' is appearing in the software appendix at the end of the document.

Once you have verified that `Matplotlib` has been properly installed and it appears into the final `paper.pdf` , you are ready to make the first commit of the project. With the next commands, you will see which files have been modified, what are the modifications, prepare them to be committed, and make the commit. In the commit process, `Git` will open the text editor for writing the commit message. Take into account that all changes committed will be preserved in the history

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

of your project. So, it is a good practice to take some time to describe properly what have been done/changed/added. Finally, as this is the very first commit of the project, tag this as the zero-th version.

```
$ git status      # See which files have been changed.
$ git diff        # See the lines you have modified.
$ git add -u      # Put all tracked changes in staging area.
$ git status      # Make sure everything is fine.
$ git commit      # Your first commit, add a nice description.
$ git tag -a v0.0 # Tag this as the zero-th version of your project.
```

Now, have a look at the `Git` history of the project. Note that the local `master` branch is one commit above than the remote `origin/master` branch. After that, push your first commit and its tag to your remote repository with the next commands. Since you had setup your `master` branch to follow `origin/master`, you can just use `git push`.

```
$ git log --oneline --decorate --all --graph # Have a look at the Git history.
$ git push                                  # Push the commit to the remote/origin.
$ git push --tags                           # Push all tags to the remote/origin.
```

Now it is time to start including your own scripts to download and make the analysis of the data. It is important to bear in mind that the goal of this tutorial is to give a general view of the workflow in `Maneage`. In this sense, only a few basic concepts about `Make` and how it is used into this project will be given. `Maneage` is much more powerfull and much more things than the ones showed in this tutorial can be done. So, read carefully all the documentation and comments already available into each file, be creative and experiment making your own research.

In the following, the tutorial will be focused in download the data, analyse the data, and finally write the results into the final paper. As a consequence, there are a lot of things already done that are not necessary. For example, all the text of the final paper already written into the `paper.tex` file, some `Makefiles` to download images from the Hubble Space Telescope and analyse them, etc. In your own research, all of this work would be removed. However, in this tutorial they are not removed because we will only show how to do a simple analysis and include a small paragraph with the result of the linear fitting.

**In short:** in this section you have learnt how to install available software in `Maneage`. In this particular case, you installed `Matplotlib`

### Including Python script to make the analysis

You are going to use a small Python script to make the analysis of the data. This Python script will be invoked from a `Makefile` that will be set up later. For now, we are going to just create the Python script and put it in an appropriate location. All analysis scripts are kept into a subfolder with the name of the same file type in `reproduce/analysis`. For example, the `Makefiles` are saved into the `make` directory, and bash scripts are saved into the `bash` directory. Since there is any `python` directory, create it with the following command.

```
$ mkdir reproduce/analysis/python
```

After that, you need the Python script itself. The code is very simple: it will take an input file containing two columns (year and population), the name of the output file in which the parameters of the linear fit will be saved, and the name of the figure showing the original data and the fitted curve. Paste the next Python script into a new file named `linear-fit.py` into the directory generated in the above step (`reproduce/analysis/python`).

```
# Make a linear fit of an input data set
#
# This Python script makes a linear fitting of a data consisting in time and
# population. It generates a figure in which the original data and the
# fitted curve is plotted. Finally, it saves the fitting parameters.
#
# Original author:
# Copyright (C) 2020, Raul Infante-Sainz <infantesainz@gmail.com>
# Contributing author(s):
# Copyright (C) YEAR, YourName YourSurname.
#
# This Python script is free software: you can redistribute it and/or modify it
# under the terms of the GNU General Public License as published by the
```

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

```
# Free Software Foundation, either version 3 of the License, or (at your
# option) any later version.
#
# This Python script is distributed in the hope that it will be useful, but
# WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General
# Public License for more details. See <http://www.gnu.org/licenses/>.

# Necessary packages
import sys
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit

# Fitting function (linear fit)
def func(x, a, b):
    return a * x + b

# Define input and output arguments
ifile = sys.argv[1] # Input file
ofile = sys.argv[2] # Output file
ofig = sys.argv[3] # Output figure

# Read the data from the input file.
data = np.loadtxt(ifile)

# Time and population:
# time ----- x
# population ---- y
x = data[:, 0]
y = data[:, 1]

# Make the linear fit
params, pcov = curve_fit(func, x, y)

# Make and save the figure
plt.clf()
plt.figure()

plt.plot(x, y, 'bo', label="Original data")
plt.plot(x, func(x, *params), 'r-', label="Fitted curve")

plt.title('Population along time')
plt.xlabel('Time (year)')
plt.ylabel('Population (million people)')
plt.legend()
plt.grid()

plt.savefig(ofig, format='PDF', bbox_inches='tight')

# Save the fitting parameters
np.savetxt(ofile, params, fmt='%.3f')
```

Have a look at this Python script. At the very beginning, it has a block of commented lines with a descriptive title, a small paragraph describing the the script, and the copyright with the contact information. For each file, it is very important to have such kind of meta-data. Below these lines, there is the source code itself.

As it can be seen, this Python script ( `linear-fit.py` ) is designed to be invoked from the command line in the following way.

```
$ python /path/to/linear-fit.py /path/to/input.dat /path/to/output.dat /path/to/figure.pdf
```

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

`/path/to/input.dat` is the input data file, `/path/to/output.dat` is the output data file (with the fitted parameters), and `/path/to/figure.pdf` is the plotted figure.

You will do this invocation inside of a Make rule (that will be set up later). Now that you have included this Python script, make a commit in order to save this work. With the first command you will see the files with modifications. With the second command, you can check what are the changes. Correct, add and modify whatever you want in order to include more information, comments or clarify any step. After that, add the files and commit the work. Finally, push the commit to the remote/origin.

```
$ git status                # See which files you have changed.
$ git diff                  # See the lines you have added/changed.
$ git add reproduce/analysis/python/linear-fit.py # Put all tracked changes in staging area.
$ git commit                # Commit, add a nice descriptions.
$ git push                  # Push the commit to the remote/origin.
```

Check that everything is fine having a look at the `Git` history of the project. Note that the `master` branch has been increased in one commit, while the `template` branch is behind.

```
$ git log --oneline --decorate --all --graph # See the 'Git' history.
```

In short: in this section you have included a `Python` script that will be used for making the linear fitting.

### Downloading data

As it was said before, there are multiple things that are already included into the project. One of them is to use a dedicated Makefile to manage all necessary download of the input data ( `reproduce/analysis/make/download.mk` ). By appropriate modifications of this file, you would be able to download the necessary data. However, in order to keep this tutorial as simple as possible, we will describe how to download the data you need more explicitly.

The data needed by this tutorial consist in a simple plain text file containing two rows: time (year) and population (in million of people). This data correspond to Spain, and it can be downloaded from this URL:  
`http://akhlaghi.org/data/template-tutorial/ESP.dat` . But don't do that using your browser, you have to do it into `Maneage` !

Let's create a Makefile for downloading the data. Later, you will also include (in the same Makefile) the necessary work in order to make the analysis. Save this Makefile in the dedicated directory ( `reproduce/analysis/make` ) with the name `getdata-analysis.mk` . In that Makefile, paste the following code.

```
# Download data for the tutorial
#
# In this Makefile, data for the tutorial is downloaded.
#
# Copyright (C) 2020 Raul Infante-Sainz <infantesainz@gmail.com>
# Copyright (C) YYYY Your Name <your-email@example.xxx>
#
# This Makefile is free software: you can redistribute it and/or modify it
# under the terms of the GNU General Public License as published by the
# Free Software Foundation, either version 3 of the License, or (at your
# option) any later version.
#
# This Makefile is distributed in the hope that it will be useful, but
# WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General
# Public License for more details. See <http://www.gnu.org/licenses/>.
```

```
# Download data for the tutorial
# -----
#
pop-data = $(indir)/ESP.dat
$(pop-data): | $(indir)
    wget http://akhlaghi.org/data/template-tutorial/ESP.dat -O $@
```

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

```
# Final TeX macro
# -----
#
# It is very important to mention the address where the data were
# downloaded in the final report.
$(mtdir)/getdata-analysis.tex: $(pop-data) | $(mtdir)
    echo "\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}" > $@
```

Have a look at this Makefile and see the different parts. The first line is a descriptive title. Below, include your name, contact email, and finally, the copyright. Please, take your time in order to add all relevant information in each Makefile you modify. As you can see, these lines start with # because they are comments.

After that information, there are five white lines in order to separate the different parts. Then, you have the Make rule to download the data. Remember the general structure of a Make rule:

```
TARGETS: PREREQUISITES
        RECIPE
```

In a rule, it is said how to construct the TARGETS from the PREREQUISITES, following the RECIPE. Note that the white space at the beginning of the RECIPE are not spaces but a single TAB. Take into account this if you copy/paste the code.

Now you can see this structure in our particular case:

```
$(pop-data): | $(indir)
    wget http://akhlaghi.org/data/template-tutorial/ESP.dat -O $@
```

Here we have:

- \$(pop-data) is the TARGET. It is previously defined just one line above: pop-data = \$(indir)/ESP.dat. As it can be seen, the target is just one file named ESP.dat into the indir directory.
- \$(indir) is the PREREQUISITE. In this case, nothing is needed for obtaining the TARGET, just the output directory in which it is going to be saved. This is the reason of having the pipe | at the beginning of the prerequisite (it indicates an order-only-prerequisite).
- wget http://akhlaghi.org/data/template-tutorial/ESP.dat -O \$@ is the RECIPE. It states how to construct the TARGET from the PREREQUISITE. In this case, it is just the use of wget to download the file specified in the URL (http://akhlaghi.org/data/template-tutorial/ESP.dat) and save it as the target: -O \$@. Inside of a Make rule, \$@ is the target. So, in this case: \$@ is \$(pop-data).

With this, you have included the rule that will download the data. Now, to finish, you have to specify what is the final purpose of the Makefile: download that data! This is done by setting \$(pop-data) as a prerequisite of the final rule. Remember that each Makefile will build a final target with the same name as the Makefile, but with the extension .tex. As a consequence, they will be TeX macros in which relevant information to be included into the final paper are saved. Here, you are saving the URL.

```
$(mtdir)/getdata-analysis.tex: $(pop-data) | $(mtdir)
    echo "\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}" > $@
```

In this final rule we have:

- \$(mtdir)/getdata-analysis.tex is the TARGET. It is the TeX macro. Note that it has the same name as the Makefile itself, but it will be saved into the \$(mtdir) directory. What do I need for constructing this target? The prerequisites.
- \$(pop-data) | \$(mtdir) are the PREREQUISITES. In this case you have two prerequisites. First, \$(pop-data), which indicates that the final TeX macro has to be generated after this file has been obtained. The second prerequisite is order-only-prerequisite, and it is the directory in which the target is saved: \$(mtdir).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

- `echo "\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}" > $@` is the RECIPE. Basically, it writes the text `\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}` into the TARGET ( `$@` ). As you can see, this is the definition of a new command in TeX . The definition of this new command `\popurl` will be used for writing the final paper.

Only one step is remaining to finally make the download of the data. You have to add the name (without the extension .mk) of this Makefile into the `reproduce/analysis/make/top-make.mk` Makefile. There it is defined which Makefiles have to be executed. You have to end up having:

```
makesrc = initialize \  
         download \  
         getdata-analyse \  
         delete-me \  
         paper
```

As always, read carefully all comments and information in order to know what is going on. Also, add your own comments and information in order to be clear and explain each step with enough level of detail. If everything is fine, now the project is ready to download the data in the make step. Try it!

```
$ ./project make
```

Hopefully, it will download and save the file into the folder called `inputs` under the `build-directory` . Check that it is there, and also have a look at the `TeX` macro in order to see that the new command has been included, it is into the top-build directory: `build-directory/teX/macros/getdata-analysis.tex` .

Now that all of this changes have been included and it works fine, it is time to check little by little everything and make a commit order to save this work. Remember to put a good commit title and a nice commit message describing what you have done and why. Then, push the commit to the remote/origin.

Congratulations! You have included you first Makefile and the data is now ready to be analysed!

In short, to download the data you did the following:

- Create a Makefile: `reproduce/analysis/make/getdata-analysis.mk`
- Write meta-data at the beginning: title, your name, email, copyright, etc.
- Define the file you want to download, and the rule to do it.
- Write the rule to generate the `TeX` macro, putting as prerequisite, the file you are downloading.
- Add the name of the Makefile (without the `.tex` ) into `reproduce/analysis/make/top-make.mk`
- `$ ./project make` in order to execute the project and download the data.
- Check that everything worked fine by looking at the downloaded file and the `TeX` macro.
- Commit and push all the work included.

### Adding the analysis rule

Until this point, you have included the Python script that will do the linear fitting, and the rule for downloading the data. Now, it is necessary to construct the Make rule in which this Python script is invoked to do the analysis. This rule will be put in the same Makefile you have already generated for downloading the data. But, before this, define the directory in which the target is going to be saved.

```
odir = $(BDIR)/fit-parameters
```

This is a folder under the `build-directory` called `fit-parameters` . After that, define the target: a plain text file in which the linear fit parameters are saved (by the Python script). Put it into the previously defined directory. As the data is from Spain, name it `ESP.txt` .

```
param-file = $(odir)/ESP.txt
```

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

Now, include a rule to construct the output directory `odir`. This is necessary because this directory is needed for saving the file `ESP.txt`.

```
$(odir):  
    mkdir $@
```

With all the previous definitions, now it is possible to set the rule for making the analysis:

```
$(param-file): $(indir)/ESP.dat | $(odir)  
    python reproduce/analysis/python/linear-fit.py $< $@ $(odir)/ESP.pdf
```

In this rule you have:

- `$(param-file)` is the TARGET. It is the file previously defined in which the fitting parameters will be saved.
- `$(indir)/ESP.dat | $(odir)` are the PREREQUISITES. In this case you have two prerequisites. First, `$(indir)/ESP.dat`, which is the input file previously downloaded by the rule above. In this file there is the input data that the Python script will use for making the linear fit. `$(odir)` is the second prerequisite. It is order-only-prerequisite (indicated by the pipe `|`), and it is the directory where the target is saved.
- `python reproduce/analysis/python/linear-fit.py $< $@ $(odir)/ESP.pdf` is the RECIPE. Basically, it call python to run the script `reproduce/analysis/python/linear-fit.py` with the necessary arguments: the input file `$<`, the target `$@`, and the name of the figure `$(odir)/ESP.pdf` (a PDF figure saved into the same directory than the target).

Finally, in order to indicate you want to obtain the target you have just included (`$(param-file)`), it is necessary to add it as a prerequisite of the final TARGET `$(mtexdir)/linear-fit.tex`. So, in the last rule (which creates the `TeX` macro), remove `$(pop-data)` and put `$(param-file)` instead. By doing this, you are telling to the Makefile that you want to obtain the file in which it is saved the fitted parameters. Inside of the rule, define a couple of bash variables (`a` and `b`) that are the fitted parameters extracted from the prerequisite. For `a`:

```
a=$(cat $< | awk 'NR==1{print $1}')
```

Similarly, for obtaining the parameter `b` (which is in the second row):

```
b=$(cat $< | awk 'NR==2{print $1}')
```

Then you have to specify the new `TeX` commands for these two parameters, just write them as it was done before for the URL:

```
echo "\newcommand{\afitparam}{$a}" >> $@  
echo "\newcommand{\bfitparam}{$b}" >> $@
```

So, at the end you will have the final rule like this:

```
$(mtexdir)/getdata-analysis.tex: $(param-file) | $(mtexdir)  
  
    echo "\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}" > $@  
  
    a=$(cat $< | awk 'NR==1{print $1}')  
    b=$(cat $< | awk 'NR==2{print $1}')  
  
    echo "\newcommand{\afitparam}{$a}" >> $@  
    echo "\newcommand{\bfitparam}{$b}" >> $@
```

Important notes: you have to use two `$` in order to use the bash `$` character inside of a Make rule. Also, note that you have to put `>>` in order to not create a new target each time you write something into the target. With the double `>` it will only add the line at the end of the file without generating a new file.

With all the above modifications, you are ready to obtain the fitting parameters. If you add the necessary comments and information, the final Makefile would look similar to:

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

```
# Download data and linear fitting for the tutorial
#
# In this Makefile, data for the tutorial is downloaded. Then, a Python
# script is used to make a linear fitting. Finally, fitted parameters as
# well as the URL is saved into a TeX macro.
#
# Copyright (C) 2020 Raul Infante-Sainz <infantesainz@gmail.com>
# Copyright (C) YYYY Your Name <your-email@example.xxx>
#
# This Makefile is free software: you can redistribute it and/or modify it
# under the terms of the GNU General Public License as published by the
# Free Software Foundation, either version 3 of the License, or (at your
# option) any later version.
#
# This Makefile is distributed in the hope that it will be useful, but
# WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General
# Public License for more details. See <http://www.gnu.org/licenses/>.

# Download data for the tutorial
# -----
#
# The input file is defined and downloaded using the following rule
pop-data = $(indir)/ESP.dat
$(pop-data): | $(indir)
    # Use wget to download the data
    wget http://akhlaghi.org/data/template-tutorial/ESP.dat -O $@

# Output directory
# -----
#
# Small rule for constructing the output directory, previously defined
odir = $(BDIR)/fit-parameters
$(odir):
    # Build the output directory
    mkdir $@

# Linear fitting of the data
# -----
#
# The output file is defined into the output directory. The fitted
# parameters will be saved into this directory by the Python script.
param-file = $(odir)/ESP.txt
$(param-file): $(indir)/ESP.dat | $(odir)
    # Invoke Python to run the script with the input data
    python reproduce/analysis/python/linear-fit.py $< $@ $(odir)/ESP.pdf

# TeX macros final target
# -----
#
# This is how we write the necessary parameters in the final PDF. In this
# rule, new TeX parameters are defined from the URL, and the fitted
# parameters.
$(mtekdir)/getdata-analysis.tex: $(param-file) | $(mtekdir)

    # Write the URL into the target
    echo "\newcommand{\popurl}{http://akhlaghi.org/data/template-tutorial}" > $@

    # Read the fitted parameters and save them into the target
```

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

```
a=$(cat $< | awk 'NR==1{print $1}')  
b=$(cat $< | awk 'NR==2{print $1}')  
  
echo "\newcommand{\afitparam}{$sa}" >> $@  
echo "\newcommand{\bfitparam}{$sb}" >> $@
```

Have look at this Makefile and note that it is what it has been described above. Take your time for making useful comments and modifying whatever you think it is necessary. If everything is fine, now the project is ready to download the data and make the linear fitting. Try it!

```
$ ./project make
```

Hopefully, now you will have the fitted parameters into the `build-directory/fit-parameters/ESP.txt` file, and the figure in the same directory. Do not pay to much attention at the quality of the fitting. It is just an example. Also, check that the TeX macro has been created successfully by having a look at `build-directory/tex/macros/getdata-analyse.tex`. Finally, now that you have ensured that everything is fine, make a commit in order to keep the work safe. In the next step, you will see how to include this data into the final paper.

**In short:** with the work included in this section, the project is able to download and make the linear fitting of the data. The result is the fitted parameters that are also saved in a TeX macro, and the figure showing the data with the fitted curve.

### Editing the final paper

With all the previous work, the project is able to download the file containing the data (two columns, year and population of Spain), and analyse them by making a linear fitting ( $y=ax+b$ ). The result is a TeX macro in which there are the information about the URL of the data and the linear fitting parameters ( `a` and `b` ). Now, it is time to add a small paragraph into the paper, just to illustrate how to write the relevant parameters from the analysis.

Before all, make a copy of the current `paper.pdf` document you have into the `project-directory`. This paper is an example that Maneage constructs by default. Now, you will modify it by adding a small paragraph including the fitting parameters and the URL. So, open `project-directory/paper.tex` and add the following paragraph just at the beginning of the abstract section.

```
By following the steps described in the tutorial, I have been able to obtain this reproducible paper!  
The project is very simple and it consists in download a file (from \popurl), and make an easy linear fit  
The linear fitting is  $y=a*x+b$ , with the following parameters:  $s_a=\backslashafitparam$  and  $s_b=\backslashbfitparam$ 
```

As you can see, the TeX definitions done before in the Makefiles, are now included into the paper: `\popurl`, `\afitparam`, and `\bfitparam`. If you do again the make step `$ ./project make`, you will re-compile the paper including this paragraph. Check that it is true and compare with the previous version, of the paper. Contratulations! You have complete this tutorial and now you are able to use Maneage for making your exciting research in a reproducible way!

### Copyright information

This file is part of the reproducible paper template <http://savannah.nongnu.org/projects/reproduce>

This template is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This template is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Template. If not, see <https://www.gnu.org/licenses/>.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

Firmado por: RAUL INFANTE SAINZ  
UNIVERSIDAD DE LA LAGUNA

Fecha: 04/03/2021 11:59:31

IGNACIO TRUJILLO CABRERA  
UNIVERSIDAD DE LA LAGUNA

04/03/2021 12:00:29

María de las Maravillas Aguiar Aguiar  
UNIVERSIDAD DE LA LAGUNA

14/04/2021 15:46:33

# B

## Appendix: Contributions during this thesis

### B.1 Publications during this thesis

1. *Towards Long-term and Archivable Reproducibility*; Akhlaghi, M., Infante-Sainz, R., Roukema, B.F., Valls-Gabaud, D., Baena-Gallé, R.; submitted to CiSE, arXiv:2006.03018.
2. *The Galaxy “Missing Dark Matter” NGC 1052-DF4 is Undergoing tidal Disruption*; Montes, M., Infante-Sainz, R., Madrigal-Aguado, A., Román, J., Monelli, M., Borlaff, A.S., Trujillo, I.; ApJ, 904, 114.
3. *The Sloan Digital Sky Survey extended point spread functions*; Infante-Sainz, R., Trujillo, I., Román, J.; MNRAS, 491, 5317.
4. *The missing light of the Hubble Ultra Deep Field*; Borlaff, A., Trujillo, I., Román, J., Beckman, J.E., Eliche-Moral, M.C., Infante-Sainz, R., Lumbreras-Calle, A., Takuro, S., Gómez-Guijarro, C., Cebrián, M., Dorta, A., Cardiel, N., Akhlaghi, M., Martínez-Lombilla, C.; A&A, 621, A133.

### B.2 Conference proceedings

1. *Reaching The Limits Of Amateur Telescopes For Low Surface Brightness Science*; Infante-Sainz, R., Roig, A., Trujillo, I.; IAU Symposium 355: The Realm of the Low Surface Brightness Universe. In press.
2. *The Importance of Having an Extended Point-spread Function in Low Surface-brightness Science*; Infante-Sainz, R., Trujillo, I., Román, J.; RNAAS, 4, 8, id.130.
3. *Extended Point-spread Functions for Deep Astronomical Imaging Surveys*, Baena-Gallé, R.; Infante-Sainz, R., Akhlaghi, M., Trujillo, I., Knapen, J.H.; RNAAS, 4, 7, id.124.

### B.3 Oral contributions

1. **Contributed talk.** Infante-Sainz, R., Trujillo, I., Román, J.; *The importance of the extended point spread function in low surface brightness science*. American Astronomical

Society Virtual Meeting (AAS 236th). June 2020, USA (virtual).

2. **Contributed talk.** Infante-Sainz, R.; *The importance of having an extended PSF low surface brightness science.* Group meeting of LSST pipeline team at Princeton University. March 2020, Princeton, USA.
3. **Contributed talk.** Infante-Sainz, R.; *The SDSS extended PSFs and star counting VS integrated photometry techniques.* The Low Surface Brightness Universe as seen by LSST. February 2020, Sesto, Italy.
4. **Contributed talk.** Infante-Sainz, R., Roig, A., Trujillo, I.; *Llevando los telescopios aficionados al límite con ciencia de bajo brillo superficial: Imagen profunda de 100 horas de M101 con un telescopio pequeño.* III Congreso Pro-Am. December 2019, Huesca, Spain.
5. **Invited talk.** Infante-Sainz, R., Roig, A., Trujillo, I.; *Hunting in the darkness: the adventure of unveiling the low surface brightness Universe.* 25 Convención de astronomía. November 2019, Sabadell, Spain.
6. **Contributed talk.** Infante-Sainz, R., Roig, A., Trujillo, I.; *A 100 h Imaging of M101 with a small telescope: reaching the limits of amateur telescopes for low surface brightness science.* Symposium IAU 355: The Realm of the low-surface brightness Universe. July 2019, Tenerife, Spain.
7. **Contributed talk.** Infante-Sainz, R.; *Building the extended PSF and removing the scatter light from SDSS imaging.* SUNDIAL Work Package 2 meeting. October 2018, Tenerife, Spain.
8. **Contributed talk.** Infante-Sainz, R., Trujillo, I., Román, J., Martínez-Lobilla, C., Borlaff, A., Chamba, N.; *A crucial test in stellar halo physics: star counting versus integrated photometry techniques.* Stellar halos across the cosmos at MPIA. July 2018, Heidelberg, Germany.
9. **Contributed talk.** Infante-Sainz, R.; *Low Surface Brightness Science with SDSS.* Exploring the Ultra-Low Surface Brightness Universe at ISSI. November 2017, Bern, Switzerland.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |

### *Acknowledgements*

En primer lugar quiero agradecer a mi director de tesis, Nacho Trujillo, su incansable apoyo y ayuda para sacar esta tesis adelante. Por siempre ir un paso más allá en el descubrimiento de la verdad y aportar a lo largo del camino puntos de vista que me han hecho crecer como persona y como científico. Gracias.

En el Instituto de Astrofísica de Canarias (IAC) agradezco a todo el personal que lo hace funcionar tan eficientemente. En particular, a todas las personas del grupo TRACES y sus numerosas reuniones en las que tanto he aprendido. Mención especial merecen Johan Knapen, Jesús Falcón y Mike Beasley, por brindar su apoyo de manera tan profesional y humana. A Javier Román, por ser el mejor hermano mayor que he tenido al empezar en este mundo de la investigación y en particular el bajo brillo superficial. To Mohammad Akhlaghi, for spending all the necessary time in showing the importance of reproducibility, as well as better ways of developing good codes/scripts and managing data. A Lourdes, Eva, Carmen, Judith, personal de secretaría y administración que tan eficiente y humanamente trabajan para que todo funcione como la seda. Gracias.

El doctorado no hubiese sido lo mismo sin mis colegas, compañeros de aventura en congresos y proyectos: Aleix Roig, Alejandro Borlaff, Cristina Martínez, Mireia Montes, Nushkia Chamba, Giulia Golini, Samaeh Sharbaf, Samane Raji, Roberto Baena, Alberto Madrigal, Carlos Morales, Raúl Castellanos, Fernando Buitrago, Josep Drudis. A David Valls por acogerme en la que fue mi primera charla técnica del doctorado en Berna. A Aleix Roig, por su infinito esfuerzo y paciencia en el proyecto de observar M101 más de 100 horas con su modesto telescopio. A la agrupación astronómica de Sabadell, Huesca y la SEA, Emili Capella y Cesca. To Robert Lupton, Yusra AlSayyad, Morgan Schmitz, Lee Kelvin, and all the people in Princeton, where I was so comfortable in such a short time because of the Covid-19! De los que tanto he aprendido y a los que también espero haber aportado algo. Gracias.

A todos mis colegas de carrera en Granada, especialmente Alize, con la que tantos momentos de reflexión y debates he tenido y tendremos. A mis compañeros del Máster de Astrofísica en la ULL: Núria Casasayas, Jaume, Paula Izquierdo, Elena, Núria Salvador, Juan Carlos, Lucía, Paula Sola. A mis colegas de corralón y verdadera familia en Tenerife: Felipe, Cris, Zaira, Simón, Camilo, Ana, David, Diego, Isa, Adur, Ana Belén, Salvo, Paz. En la distancia pero no menos importante, a mis shurs de toda la vida en Estepona: Sergio, Adrián, Manuel, Gaspi, Jorge, Luisfer, y chupiclimbers, Elena, Kike, Mafe, Fran, Marta, Alvarito, Christian, Dani. ¡Sois unos cracks!

A todos mis profesores, por enseñarme que conocer el mundo que nos rodea es una experiencia fascinante. A Isabel Senda, por explicar la asignatura de física en el bachillerato como nadie. Eva, tutora durante bachiller que siempre confió y me ayudó en todo lo que necesité. A Estrella y Ute, por ser tan buenas profesoras y mejores personas durante la carrera y trabajo fin de grado, guiando mis primeros pasos en la astrofísica. Gracias.

Para acabar, me gustaría agradecer a todos los que me han visto crecer y siempre han estado ahí, en la cercanía y en la distancia. A toda mi familia, tanto del norte como del sur de España, a los que están y los que no, Papá, Mamá, Sergio, Carlos, abuelos, tíos, y sobrinos. Por supuesto, a la que desde hace tiempo también es mi familia en Chiclana, Ángel, Cati, Adelaida, Ángel, y sobrinas. Finalmente, a la persona más importante en mi vida. Virgi. Gracias por apoyarme, estar siempre a mi lado, y hacer que esta tesis salga adelante. Better together.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.  
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 3258585 Código de verificación: 8GKKDuAS

|   |                            |
|---|----------------------------|
| Firmado por: RAUL INFANTE SAINZ<br>UNIVERSIDAD DE LA LAGUNA       | Fecha: 04/03/2021 11:59:31 |
| IGNACIO TRUJILLO CABRERA<br>UNIVERSIDAD DE LA LAGUNA              | 04/03/2021 12:00:29        |
| María de las Maravillas Aguiar Aguiar<br>UNIVERSIDAD DE LA LAGUNA | 14/04/2021 15:46:33        |