



universität  
wien

# DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„A Vision of Orion“

verfasst von / submitted by

Stefan Meingast, Bakk. MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of  
Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on the  
student record sheet:

A 796 605 413

Dissertationsgebiet lt. Studienblatt /  
field of study as it appears on the student  
record sheet:

Astronomie

Betreut von / Supervisor:

Univ. Prof. Dr. João Alves





*Things are only impossible until they are not.*

Jean-Luc Picard

### **The path to Orion**

The Orion constellation has been a center of wonder, imagination, and scientific exploration throughout the entire history of mankind. Here, the constellation decorates the night sky above a dry riverbed in the rocky Atacama desert in Chile. A bright moon is visible to the right along with the famous Hyades and Pleiades open star clusters closer to the ground. Image credit: Yuri Beletsky.



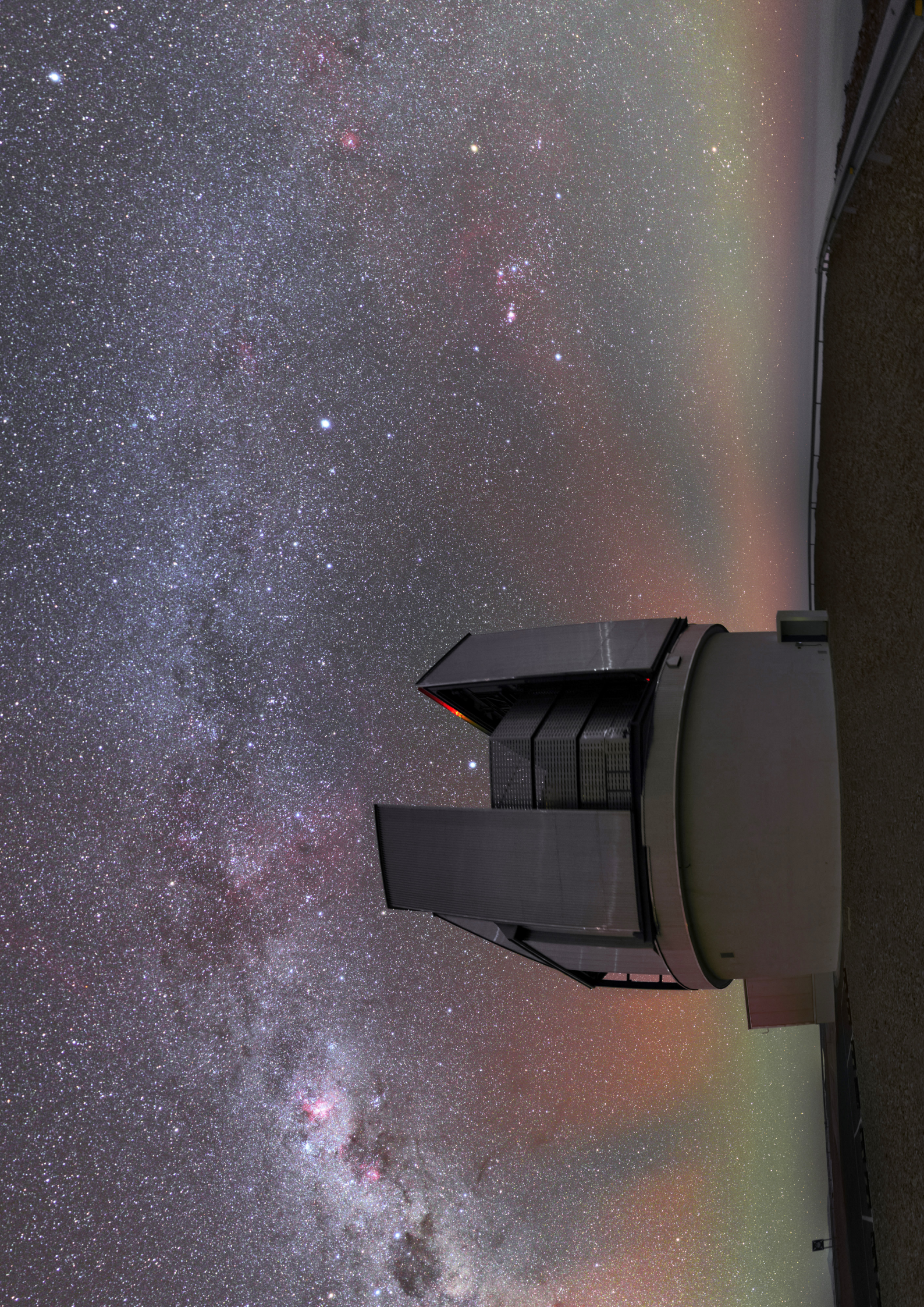




## **VISTA**

The Visual and Infrared Survey Telescope for Astronomy (VISTA) is part of the European Southern Observatory's Paranal facilities and explores the southern night sky in the Chilean desert. Above the telescope dome and the Orion constellation to the right, the Milky Way majestically arches across the sky in this picturesque panorama. Image credit: Yuri Beletsky/ESO.



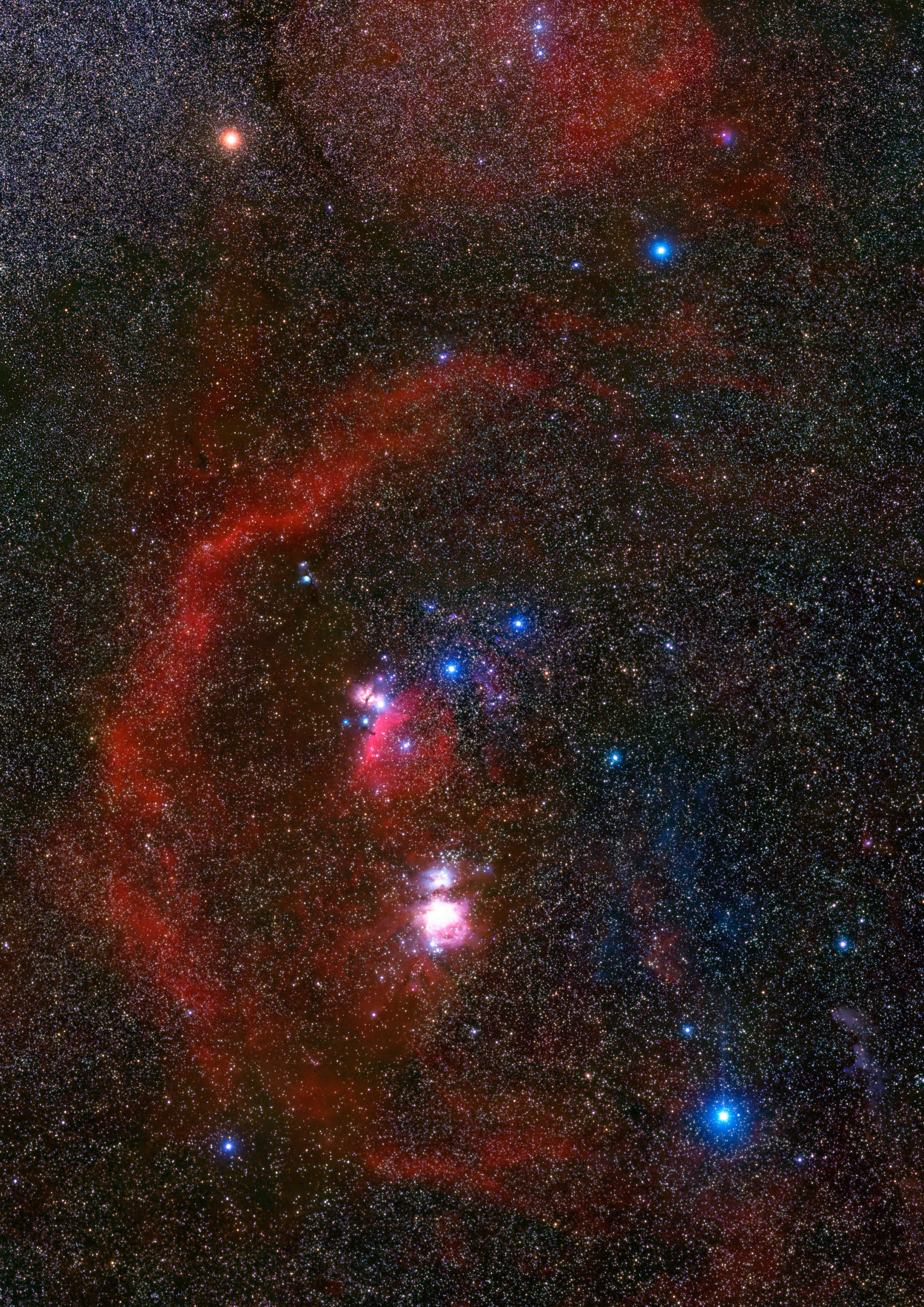




### **The Orion constellation**

The bright stars in Orion clearly outline one of the most prominent constellations in the night sky. This picture brilliantly shows the giant stars Betelgeuse, Bellatrix, Saiph, and Rigel in the corners of the image, as well as the Orion Belt stars in the center. The red glowing hydrogen is witness to the remarkable star formation history of this region, where today young stars are formed in the Flame Nebula to the left of Orion's Belt and in the Orion Nebula to the south. Image Credit: Wei-Hao Wang.







## **Orion A**

Located at a distance of about 1400 light years, the Orion A molecular cloud is the nearest birthplace of massive stars to the sun. The northern end of the cloud hosts the famous Orion Nebula where thousands of new stars are currently forming. The cloud, however, extends for several degrees to the South and is home to many other less well-known star-forming sites. Image credit: ESO/VISION survey.







## **The Orion Nebula**

An infrared look into the heart of the Orion Nebula reveals a massive cluster, composed of thousands of stars. This spectacular view is caused by stars ten-thousand times more luminous than the Sun which illuminate the surrounding interstellar medium. Image credit: Stefan Meingast/VISION survey.









# Contents

<b>Abstract</b>	<b>i</b>
<b>List of Publications</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context and Motivation . . . . .	1
1.1.1 Measuring column-densities in molecular clouds . . . . .	3
1.1.2 New opportunities . . . . .	5
1.1.3 Interstellar dust extinction . . . . .	7
1.2 Specific goals of the thesis . . . . .	12
1.2.1 The Visible and Infrared Survey Telescope for Astronomy . .	16
1.2.2 Thesis Overview . . . . .	20
<b>2 VISTA Orion A survey</b>	<b>23</b>
2.1 Overview . . . . .	23
2.2 Publication details . . . . .	24
2.3 Introduction . . . . .	25
2.4 Observations . . . . .	30
2.4.1 Instrumentation . . . . .	30
2.4.2 Survey design and strategy . . . . .	30
2.5 Data processing . . . . .	31
2.5.1 Motivation . . . . .	32
2.5.2 Master calibration frames . . . . .	34
2.5.3 Science data calibration . . . . .	35
2.5.4 Astrometric calibration . . . . .	35
2.5.5 Tile and Orion A mosaic assembly frames . . . . .	36
2.5.6 Photometric calibration . . . . .	37
2.5.7 Final catalog assembly . . . . .	38

2.6	VISTA Orion A survey data products . . . . .	39
2.6.1	VISTA Orion A source catalog . . . . .	40
2.6.2	VISTA Orion A mosaic L-RGB . . . . .	40
2.7	Young stellar populations toward Orion A . . . . .	41
2.7.1	Orion A population . . . . .	41
2.7.2	Foreground population . . . . .	46
2.8	Summary . . . . .	47
2.9	References . . . . .	48
2.10	Appendix A: Data characteristics . . . . .	49
2.10.1	Photometric properties . . . . .	49
2.10.2	Astrometric properties . . . . .	50
2.11	Appendix B: Tabela . . . . .	54
<b>3</b>	<b>Estimating extinction</b>	<b>57</b>
3.1	Overview . . . . .	57
3.2	Publication details . . . . .	58
3.3	Introduction . . . . .	59
3.4	Motivation . . . . .	60
3.5	Method description . . . . .	62
3.5.1	The multidimensional feature space . . . . .	62
3.5.2	Constructing probability density distributions . . . . .	62
3.5.3	Estimating extinction . . . . .	63
3.5.4	Creating smooth extinction maps . . . . .	64
3.6	Method validation . . . . .	65
3.6.1	Applying NICER and PNICER to real data . . . . .	65
3.6.2	The intrinsic color distribution and population bias . . . . .	65
3.6.3	Extinction maps . . . . .	67
3.6.4	Performance . . . . .	67
3.6.5	Software availability . . . . .	68
3.7	Summary . . . . .	68
3.8	References . . . . .	69
3.9	Appendix A: Software dependencies . . . . .	70
<b>4</b>	<b>Infrared Extinction in Orion A</b>	<b>71</b>
4.1	Overview . . . . .	71
4.2	Publication details . . . . .	72
4.3	Introduction . . . . .	73
4.4	Data . . . . .	75

4.4.1	Overview . . . . .	75
4.4.2	Cross-matching and cleaning . . . . .	76
4.4.3	Control field data . . . . .	76
4.4.4	Master catalog . . . . .	77
4.5	Methods . . . . .	77
4.5.1	Linear fitting . . . . .	77
4.5.2	Measuring extinction . . . . .	78
4.6	The infrared extinction law in Orion A . . . . .	80
4.6.1	Fitting color-color diagrams . . . . .	80
4.6.2	The average extinction law . . . . .	81
4.6.3	Spatial variations across the cloud . . . . .	82
4.6.4	The shape of the Orion A infrared extinction curve . . . . .	85
4.7	Orion A extinction map . . . . .	85
4.7.1	Preparations . . . . .	85
4.7.2	The extinction map . . . . .	87
4.7.3	Herschel-Planck re-calibration . . . . .	88
4.8	Conclusions and summary . . . . .	90
4.9	References . . . . .	91
<b>5</b>	<b>Conclusions &amp; Outlook</b>	<b>93</b>
5.1	Summary of results . . . . .	93
5.2	VISIONS - VISTA Star Formation Atlas . . . . .	96
5.2.1	Own contribution . . . . .	97
5.2.2	Motivation . . . . .	98
5.2.3	Objectives . . . . .	99
5.2.4	Survey Design & Observing strategy . . . . .	101
5.2.5	Project schedule . . . . .	105
5.2.6	Data processing . . . . .	106
5.2.7	Follow-up observations . . . . .	107
5.3	Concluding statement . . . . .	110
	<b>References</b>	<b>111</b>
<b>A</b>	<b>Appendix</b>	<b>127</b>
A.1	Zusammenfassung . . . . .	129
A.2	Acknowledgements . . . . .	131
A.3	Reprint permission . . . . .	133
A.4	VISION ESO proposal . . . . .	135
A.5	Curriculum Vitae . . . . .	145





# Abstract

Star formation is arguably one of the most fundamental processes in the observable universe. It drives and regulates the energy budget of entire galaxies on both small and large scales and so critically determines their evolution. New stars are born in giant accumulations of gas and dust, so-called molecular clouds, but little is still known about these intriguing objects. To expand our knowledge and understanding of these complexes, it is therefore critical to investigate the properties of molecular clouds in a homogeneous and self-consistent manner.

Much of our understanding of the formation of stars originates from the nearby ( $<500$  pc) star-forming complexes, most notably the Orion A molecular cloud. There, stars and planets are currently forming in a variety of physical condition, ranging from a dense massive cluster to dispersed groups and isolated young stars. In this thesis, I study the Orion A molecular cloud complex as a whole from an observational point of view and present the most comprehensive near-infrared imaging survey of the region. This survey is unmatched in completeness and sensitivity and called for the development of new methods and techniques to investigate the gas distribution of the molecular cloud.

My approach to studying Orion A is based on the fact that light is extinguished when traveling through the interstellar medium. Dust particles absorb and scatter photons, and measuring this effect allows deriving physical properties of molecular clouds. This includes, for example, the gas mass distribution and even properties of the dust grain population. In contrast to dust emission or molecular line observations, dust extinction provides more reliable mass estimates because fewer assumptions have to be made. I introduce a new method, PNICER, based on machine-learning, to calculate the effects of dust extinction. This new method avoids critical systematic errors found in other techniques, leading to more reliable results when used in combination with deep photometric data. I also address the decade-old question of the universality of the near-infrared extinction law and find statistically significant evidence of its variability. Moreover, I compare dust extinction with dust emission measurements which reveal previously unknown issues affecting often-used tracers of the gas column-density.

I conclude the thesis with an outlook to future projects which will study other nearby star-forming regions in similar detail. Specifically, the methods presented here will be the basis for this future endeavor to finally build a complete picture of the physical processes regulating the collapse of molecular clouds to form new stars.



## List of Publications

This thesis is mainly based on three first author publications. During the time as a doctoral student at the University of Vienna I have contributed to several additional published manuscripts. Here, I list all refereed publications in chronological order which I have either produced myself, or co-authored in this timeframe. An exception in this list is the last entry, which represents a manuscript that is in the peer-review stage at the time of thesis submission. Most of the below listed studies address similar topics as presented in this thesis. Nevertheless, I also pursued many other research themes out of my own passion for astronomical research. Apart from my main scientific interest of Galactic star formation, infrared extinction, and large-scale photometry (the topics of this thesis), the below listed publications also include studies on variable and pulsating stars, AGB stars, and even methods to search for extraterrestrial intelligence.

**Title:** Single-Photon Technique for the Detection of Periodic Extraterrestrial Laser Pulses

**Authors:** W. Leeb; A. Poppe; E. Hammel; J. Alves; M. Brunner; S. Meingast

**Status:** Accepted for publication in *Astrobiology*, March 3rd 2013

**Bibliographic reference:** *Astronomy & Astrophysics*, Volume 601, id.A137, 12 pp.

**DOI:** [10.1089/ast.2012.0951](https://doi.org/10.1089/ast.2012.0951)

**Title:** Pulsating stars in NGC 6231. Frequency analysis and photometric mode identification near the main-sequence

**Authors:** S. Meingast; G. Handler; R.R. Shobbrook

**Status:** Accepted for publication in *Astronomy & Astrophysics*, July 3th 2013

**Bibliographic reference:** *Astronomy & Astrophysics*, Volume 559, id.A108, 26 pp.

**DOI:** [10.1051/0004-6361/201220823](https://doi.org/10.1051/0004-6361/201220823)

**Title:** Simulated low-intensity optical pulsar observation with single-photon detector

**Authors:** W. Leeb; J. Alves; S. Meingast; M. Brunner

**Status:** Accepted for publication in *Astronomy & Astrophysics*, November 27th 2014

**Bibliographic reference:** *Astronomy & Astrophysics*, Volume 574, id.A9, 7 pp.

**DOI:** [10.1051/0004-6361/201424480](https://doi.org/10.1051/0004-6361/201424480)

**Title:** LX Cygni: A carbon star is born

**Authors:** S. Uttenthaler; S. Meingast; T. Lebzelter; B. Aringer; R. Joyce; K. Hinkle; L. Guzman-Ramirez; R. Greimel

**Status:** Accepted for publication in Astronomy & Astrophysics, November 3th 2015

**Bibliographic reference:** Astronomy & Astrophysics, Volume 585, id.A145, 9 pp.

**DOI:** [10.1051/0004-6361/201526619](https://doi.org/10.1051/0004-6361/201526619)

**Title:** VISION - Vienna survey in Orion. I. VISTA Orion A Survey

**Authors:** S. Meingast, J. Alves, D. Mardones, P. S. Teixeira, M. Lombardi, J. Großschedl, J. Ascenso, H. Bouy, J. Forbrich, A. Goodman, A. Hacar, B. Hasenberger, J. Kainulainen, K. Kubiak, C. Lada, E. Lada, A. Moitinho, M. Petr-Gotzens, L. Rodrigues, C. G. Román-Zúñiga

**Status:** Accepted for publication in Astronomy & Astrophysics, December 1st 2015

**Bibliographic reference:** Astronomy & Astrophysics, Volume 587, id.A153, 31 pp.

**DOI:** [10.1051/0004-6361/201527160](https://doi.org/10.1051/0004-6361/201527160)

**Title:** APOGEE strings: A fossil record of the gas kinematic structure

**Authors:** A. Hacar; J. Alves; J. Forbrich; S. Meingast; K. Kubiak; J. Großschedl

**Status:** Accepted for publication in Astronomy & Astrophysics, February 4th 2016

**Bibliographic reference:** Astronomy & Astrophysics, Volume 589, id.A80, 10 pp.

**DOI:** [10.1051/0004-6361/201527805](https://doi.org/10.1051/0004-6361/201527805)

**Title:** The Population of Compact Radio Sources in the Orion Nebula Cluster

**Authors:** J. Forbrich; V. Rivilla; K. Menten; M. Reid; C. Chandler; U. Rau; S. Bhatnagar; S. Wolk; S. Meingast

**Status:** Accepted for publication in the Astrophysical Journal, March 16th 2016

**Bibliographic reference:** The Astrophysical Journal, Volume 822, Issue 2, article id. 93, 13 pp. (2016).

**DOI:** [10.3847/0004-637X/822/2/93](https://doi.org/10.3847/0004-637X/822/2/93)

**Title:** Gas absorption and dust extinction towards the Orion Nebula Cluster

**Authors:** B. Hasenberger; J. Forbrich; J. Alves; S. Wolk; S. Meingast; K. Getman; I. Pillitteri

**Status:** Accepted for publication in Astronomy & Astrophysics, May 31st 2016

**Bibliographic reference:** Astronomy & Astrophysics, Volume 593, id.A7, 10 pp.

**DOI:** [10.1051/0004-6361/201628517](https://doi.org/10.1051/0004-6361/201628517)

**Title:** Orion revisited. III. The Orion Belt population

**Authors:** K. Kubiak.; J. Alves; H. Bouy; L. Sarro; J. Ascenso; A. Burkert; J. Forbrich; J. Großschedl; A. Hacar; B. Hasenberger; M. Lombardi; S. Meingast; R. Köhler; P. Teixeira

**Status:** Accepted for publication in Astronomy & Astrophysics, July 30th 2016

**Bibliographic reference:** Astronomy & Astrophysics, Volume 598, id.A124, 13 pp.

**DOI:** [10.1051/0004-6361/201628920](https://doi.org/10.1051/0004-6361/201628920)

**Title:** Search for variables in six Galactic open clusters

**Authors:** E. Paunzen; G. Handler; M. Lendl; B. Baumann; C. Rab; S. Meingast; M. Rode-Paunzen; M. Netopil; V. Antoci; L. Zhu; M. Zejda; H. Božić

**Status:** Accepted for publication in New Astronomy, October 31st 2016

**Bibliographic reference:** New Astronomy, Volume 52, p. 133-139.

**DOI:** [10.1016/j.newast.2016.10.012](https://doi.org/10.1016/j.newast.2016.10.012)

**Title:** Estimating extinction using unsupervised machine learning

**Authors:** S Meingast, M. Lombardi, J. Alves

**Status:** Accepted for publication in Astronomy & Astrophysics, February 24th 2017

**Bibliographic reference:** Astronomy & Astrophysics, Volume 601, id.A137, 12 pp.

**DOI:** [10.1051/0004-6361/201630032](https://doi.org/10.1051/0004-6361/201630032)

**Title:** VISION - Vienna survey in Orion. II. Infrared extinction in Orion A

**Authors:** S. Meingast, J. Alves, M. Lombardi

**Status:** Submitted to in Astronomy & Astrophysics, June 19th, 2017



## List of Figures

0.1	Orion above the Atacama Desert . . . . .	5
0.2	The Visible and Infrared Survey Telescope for Astronomy . . . . .	7
0.3	The Orion constellation . . . . .	9
0.4	The Orion A molecular cloud . . . . .	11
0.5	The Orion Nebula . . . . .	13
1.1	From Diffuse Gas to Stars and Planets . . . . .	2
1.2	Orion wide-field extinction map . . . . .	6
1.3	The interstellar extinction curve . . . . .	11
1.4	Extinction in Barnard 68 . . . . .	12
1.5	The Orion star-forming complex . . . . .	14
1.6	The Visible and Infrared Survey Telescope for Astronomy . . . . .	17
1.7	Schematic layout of the VIRCAM focal plane array . . . . .	18
2.1	Orion A overview . . . . .	26
2.2	Prominent objects in Orion A 1 . . . . .	27
2.3	Prominent objects in Orion A 2 . . . . .	28
2.4	VISTA survey coverage . . . . .	30
2.5	Image quality comparison . . . . .	32
2.6	CASU photometric calibration offset . . . . .	33
2.7	CASU noise map . . . . .	33
2.8	VISTA astrometric calibration . . . . .	36
2.9	Verification of source classification . . . . .	39
2.10	Raw calibrated VISTA photometry . . . . .	39
2.11	Color-magnitude and color-color diagrams . . . . .	41
2.12	Selected YSOs as viewed by VISTA . . . . .	42
2.13	New YSOs and galaxy clusters as viewed by VISTA . . . . .	43
2.14	K-band luminosity functions and YSOs in Orion A . . . . .	44
2.15	Foreground density map . . . . .	45
2.16	Foreground object analysis . . . . .	46
2.17	Luminosity functions . . . . .	49
2.18	Magnitude errors . . . . .	49
2.19	Comparison of color-magnitude diagrams . . . . .	50
2.20	Estimated completeness limits . . . . .	51
2.21	Astrometric calibration offsets . . . . .	51
2.22	Exposure time and image quality maps . . . . .	52

2.23	Seeing statistics of the VISTA survey . . . . .	53
3.1	Covariance in color-color diagrams . . . . .	60
3.2	PNICER method concept . . . . .	63
3.3	PNICER compared to NICER . . . . .	65
3.4	Population bias . . . . .	66
3.5	PNICER extinction map compared to NICER . . . . .	67
3.6	PNICER software performance . . . . .	68
4.1	Survey coverage of Orion A . . . . .	74
4.2	Near-infrared bandpasses . . . . .	75
4.3	VISTA photometry compared to HAWK-I . . . . .	75
4.4	Synthetic IRAC photometry with <i>WISE</i> . . . . .	76
4.5	Color-color diagrams for variations passband combinations . . . . .	77
4.6	MCMC fitting results for average extinction law . . . . .	81
4.7	MCMC fitting results for L1641 and the Orion Nebula . . . . .	81
4.8	Sub-region layout for Orion A . . . . .	83
4.9	Infrared extinction curve for Orion A . . . . .	84
4.10	Color excess ratios for subregions of Orion A . . . . .	84
4.11	Color excess ratios as a function of extinction thresholds . . . . .	85
4.12	Extreme deconvolution . . . . .	86
4.13	Orion A extinction map . . . . .	87
4.14	Source density gain for consecutive passbands . . . . .	88
4.15	Herschel calibration . . . . .	88
4.16	Pixel substructure . . . . .	89
5.1	VISIONS on-sky coverage . . . . .	103
5.2	Astropy reduction cascade . . . . .	108
5.3	VISTA non-linearity correction . . . . .	109
A.1	VISTA science verification extinction map . . . . .	137



## List of Tables

2.1	Near-infrared surveys toward Orion A . . . . .	29
2.2	VISTA tile properties and image quality . . . . .	31
2.3	Catalog data columns . . . . .	40
2.4	Foreground groups . . . . .	46
2.5	Estimated completeness limits . . . . .	49
2.6	Catalog sample data . . . . .	54
3.1	Adopted extinction law . . . . .	62
4.1	Properties of used passbands . . . . .	76
4.2	MCMC fitting results . . . . .	82
5.1	VISIONS observation itemized for all sub-surveys . . . . .	104



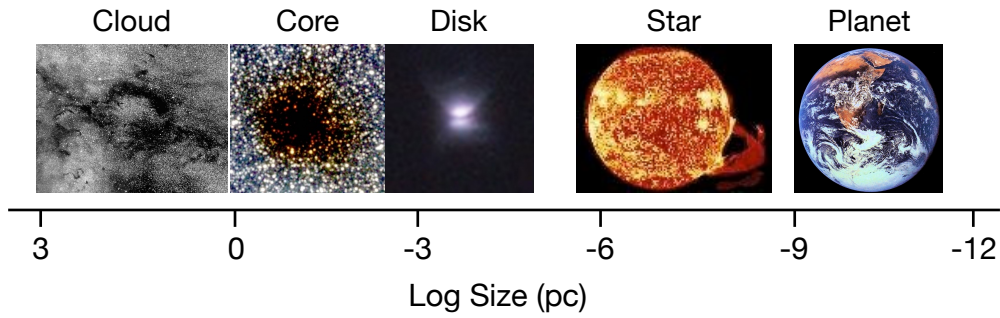
# Introduction

## 1.1 Context and Motivation

The process of star formation remains one of the greatest astrophysical challenges in the 21st century. Star formation in itself comprises a multitude of linked physical processes. In the earliest stage, giant accumulations of gas and dust, so-called molecular clouds, condense in the interstellar medium which can form filamentary structures and dense pockets of gas and dust, so-called cores. These cores are the precursors of new-born stars and can undergo gravitational collapse to form protostellar objects. Within this protostellar environment an accretion disk forms around the central pre-main-sequence object. While ongoing accretion in this Young Stellar Object (YSO) builds up mass in its center, the disk can further undergo fragmentation and thereby form a planetary system. During the evolution of such YSOs the environment is gradually cleared, eventually leaving only a remnant disk, which possibly harbor planets on which life may be able to develop.

The process of forming new stars spans orders of magnitudes in size scales, where molecular clouds can range from a few parsecs to dozens or even hundreds of parsecs. Dense cores, the birthplaces of single star systems (including binaries and higher order multiples), are typically 0.1 pc in diameter, accretion disks extend over several hundred au ( $1 \text{ pc} = 206\,265 \text{ au}$ ), and the resulting main sequence stars, like our sun, may have diameters on the order of about 1% of an au. Figure 1.1 illustrates this development over various stages as a function of size and shows that the formation of a new star can easily span ten orders of magnitude, or more. This fact alone presents a difficult obstacle in understanding and creating a theory overarching the entire evolution from a parental molecular cloud to single stars with planetary systems. For this reason, research under the general heading of star formation has fragmented into studying individual stages in the formation process. Nevertheless, many questions remain unsolved and connecting individual stages seems to be a logical step forward in the development of a comprehensive empirical foundation.

Following pioneering work in the 20th century, which identified molecular clouds as the birth sites of new stars, more and more discoveries about young stars have



**Figure 1.1.** Stars and planets form from diffuse gas, accumulated into giant clouds of mostly hydrogen, helium, and dust, in the interstellar medium. The entire process spans typically more than ten orders of magnitude in size.

been made. Largely thanks to technological advancements and the introduction of radio and infrared astronomy in the second half of the 20th century, certain empirical rules have been developed to describe the above outlined individual stages. For example, we know that the majority of stars is not formed in isolation, but rather in dense groups, so-called clusters (e.g. [Lada and Lada 2003](#); [Bressert et al. 2010](#)). These clusters dynamically evolve over time and, depending on their stellar mass content, can remain bound, or slowly disperse into the galactic field population.

Another, more controversial, example of such empirical rules is the initial mass function (IMF), i.e. the number of stars per mass interval forming within one star-forming complex. It is commonly assumed in modern astrophysics that this initial mass function is invariant at least in our Galaxy ([Bastian et al. 2010](#)). Recent results contradict this empirical rule (e.g. [Dib 2014](#)) and, in general, observational evidence is often biased by completeness problems or low number statistics (or both).

The issue of the IMF provokes the question of the origin of such a potentially universal distribution. Naturally, one explanation leads to the origin of the stars themselves: the molecular clouds. Some studies have shown an intimate connection between the IMF and the gas mass distribution in the form of the core mass function (CMF, the number of cores per mass interval). For example, [Alves et al. \(2007\)](#) find a direct connection between the CMF and the IMF (for a similar example see also [Motte et al. 1998](#)). In their analysis, the CMF takes the same shape as a typical IMF, albeit an ubiquitous efficiency conversion factor of  $\sim 30\%$ . This result suggests that the masses of the eventually forming stars are determined at an early stage in the evolution of a molecular cloud.

It therefore seems that the key to unlock some remaining mysteries, regarding the development of young stellar systems, lies in understanding the parental molecular clouds. These objects, however, are difficult to characterize due to the fact that most of their mass content can not be observed directly. Similar to essentially all barionic mass in the universe, the bulk of a molecular cloud is composed of hydrogen. In these clouds, however, most of the hydrogen is (as their designation already suggests) in molecular form. At the typical temperatures in such an environment (10 to 20 K) hydrogen does not emit electromagnetic radiation because even the lowest excited state of the molecule requires transition energies equivalent to several thousand Kelvin (Stahler and Palla 2005). For this reason, the bulk of mass in a molecular cloud is virtually invisible. Moreover, since the three-dimensional structure of molecular clouds is practically not measurable without theoretical assumptions, their mass and structure are typically explored via column densities (volume densities integrated along the line of sight).

### 1.1.1 Measuring column-densities in molecular clouds

Since the  $\text{H}_2$  molecule can not be directly measured in molecular clouds, indirect tracers need to be used to study their structure and mass distribution. To do so, typically three different methods are employed which are conceptually very different and rely each on a set of assumptions.

Firstly, dust extinction can be used to determine a molecular cloud's structure and mass. This method relies on the fact that dust in a molecular cloud absorbs and scatters light of background sources, leading to so-called interstellar reddening (see Sect. 1.1.3 for further details). Measuring this effect on sources in the background of a molecular cloud allows to construct two-dimensional maps of the gas column-density (Lada et al. 1994; Alves et al. 2001; Lombardi and Alves 2001). Here, the measured dust extinction has been shown to be directly proportional to the  $\text{H}_2$  column density (e.g. Reina and Tarengi 1973; Güver and Özel 2009, and references therein), connected via

$$\frac{N_{\text{H}_2}}{A_V} = 2.21 \times 10^{21} \text{ mag}^{-1} \text{ cm}^{-2}, \quad (1.1)$$

where  $N_{\text{H}_2}$  refers to the column density of molecular hydrogen and  $A_V$  to the visual extinction in magnitudes. In contrast to other popular methods (i.e. via dust emission and molecular lines; see below), dust extinction relies on far fewer

assumptions and has been shown to be the most reliable tracer of column-densities (Goodman et al. 2009). One obvious caveat, however, is the dependence on available background sources to measure the reddening effect. Specifically, the achievable spatial resolution depends on the location of the studied region in the sky and the sensitivity and wavelength of the observations. As a consequence, extinction mapping is far more efficient for regions projected against the galactic bulge or the plane. In addition, the method works best for nearby molecular clouds because of minimal contamination by foreground stars. Also, knowledge about the extinction law is required, which, however, can be determined much more reliably than the assumptions for the other column-density tracers. Since the original work of Lada et al. (1994), several other techniques, exploiting dust extinction as column-density tracer, have been developed. A more thorough overview of these methods, including a brief history in the developments of this field, will be given in Sect. 3.3.

Another popular method to estimate gas masses relies on molecular line emission. While molecular hydrogen can not be directly observed in molecular clouds, other molecular species appear in moderate abundances. Most prominently, carbon monoxide (CO) and its isotopologues get excited mostly by collisions with  $\text{H}_2$ , thereby producing emission lines in the radio regime. Here, the so-called X-factor relates the emission line brightness of CO to the molecular hydrogen mass (e.g. Pineda et al. 2008). While this approach, in principle, allows to construct high-resolution gas column density maps, it is also accompanied by a number of issues. Molecular line excitation may occur only under certain environmental conditions, where for example the CO emission profile depends on the density of the environment. Also, maximum emissivity occurs at  $\text{H}_2$  volume densities typical for the most compact parts of molecular clouds ( $n_{\text{H}_2} \sim 10^3 \text{ cm}^{-3}$ ), but emission rapidly drops for lower densities. Moreover, the abundance of the observed molecule relative to  $\text{H}_2$  may not be constant and, most importantly, molecular line emission can become optical thick<sup>1</sup>. The latter issue can be bypassed by using CO isotopologues (or other molecules) which are thought to remain optically thin across all environments of a molecular cloud (e.g.  $^{13}\text{CO}$ ). Such isotopologues, however, appear in even smaller abundances than  $^{12}\text{CO}$ .

The third popular method to determine a molecular cloud's structure and mass is based on dust emission. Here, thermal emission from dust grains can be modelled

---

<sup>1</sup>Optically thick media in reference to a certain emission line refers to the fact that the radiation is steadily absorbed and reradiated. If this happens too often, emission and column density cease to be proportional to each other.

with a modified blackbody, with the temperature and the optical depth as free parameters (e.g. [Stutz et al. 2010](#); [Galametz et al. 2012](#); [Lombardi et al. 2014](#)). This method relies on the assumption that the frequency dependence of the optical depth follows a power law, where the exponent  $\beta$  is referred to as dust emissivity index and is determined to be  $\beta \approx 2$  in the Galaxy ([Planck Collaboration et al. 2011a,b](#)). Using a combination of passbands, measuring the dust continuum emission, the two free parameters (temperature and optical depth) can be constrained. Prominent examples of such measurements are available from the PACS ([Poglitsch et al. 2010](#)) and SPIRE ([Griffin et al. 2010](#)) photometers aboard the *Herschel Space Observatory* ([Pilbratt et al. 2010](#)), providing several passbands from 70 to 500  $\mu\text{m}$ <sup>2</sup>. To obtain an estimate of the gas mass in molecular clouds from dust emission measurements, the obtained results on dust optical depth still need to be calibrated. [Stutz et al. \(2010\)](#), for example, used a given dust model and fixed values for the grain size distribution and the hydrogen-to-dust mass. With this setup, they calculated expected flux densities and extinction and compared these to the emission measurements to constrain the gas mass distribution. [Lombardi et al. \(2014\)](#), on the other hand, directly used an extinction map (see below) and performed a linear fit between dust optical depth and extinction to set the zero-point for the gas mass estimate. Both variants are heavily relying on generous assumptions, giving rise to calibration problems. This issue will be discussed in more detail in Sect. 4.7.3.

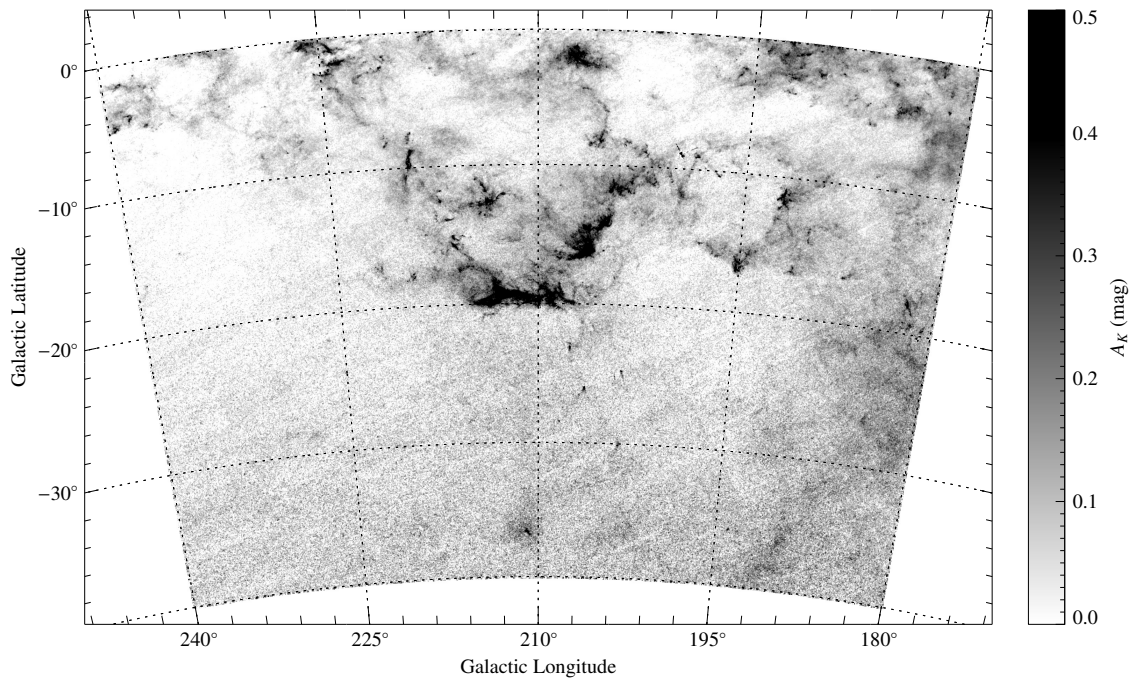
### 1.1.2 New opportunities

Most of our knowledge about molecular clouds, including their structure, and evolution has its foundation in the technological advancements over the last decades. In reference to dust extinction measurements, observations in the infrared have proven to be invaluable because they allow to observe highly reddened background sources which would be undetectable for visible light observations. Infrared cameras on large telescopes have been used for about two decades, but only recently large detector arrays have become available to the astronomical community. Such large arrays allow to systematically observe considerable portions of the night sky in unprecedented detail, opening a variety of new opportunities for observational astronomy. With respect to the dust extinction technique, such large-scale observations offer unique opportunities to improve our understanding of the gas and dust within molecular clouds and their spatial structure and evolution.

---

<sup>2</sup>The peak emission wavelength of a blackbody at a typical dust temperature of 10 K is  $\sim 290 \mu\text{m}$ .





**Figure 1.2.** A wide-field extinction map constructed from near-infrared photometry by [Lombardi et al. \(2011\)](#). Molecular clouds become visible at all scales, revealing the complex structure of the interstellar medium. Located at the center of this map, the Orion star-forming region is one of the most intriguing objects for studying dust extinction.

The most notable among large infrared surveys in recent years is the Two Micron All Sky Survey (2MASS, [Skrutskie et al. 2006](#)). This survey covers essentially the entire night sky in the near-infrared (NIR) bands  $J$  ( $1.25 \mu\text{m}$ ),  $H$  ( $1.65 \mu\text{m}$ ), and  $K_S$  ( $2.16 \mu\text{m}$ ), with  $10\sigma$  completeness limits of 15.8, 15.1, and 14.3 mag in these passbands, respectively. Based on this enormous data set, already several studies on dust extinction, including a series of manuscripts on nearby molecular clouds (e.g. the series started by [Lombardi et al. 2006](#)), have been published. An example of an extinction map is given in Fig. 1.2, which shows a  $60 \times 40 \text{ deg}^2$  wide field centered on the Orion star-forming region. Largely based on these dust extinction maps, for example, theories on probability density functions have been developed which are thought to be able to measure a molecular cloud's physical characteristics, like turbulence, and even predict star formation rates and efficiencies ([Kainulainen et al. 2009](#), cf. [Lombardi et al. 2015](#)). More recently, even larger observatories and more sophisticated detector technologies have again widened the playing field. These allow far more sensitive observations and provide yet again new possibilities in the field of dust extinction.



This thesis is devoted to investigating new opportunities in studying dust extinction with large-scale surveys, to review the practicality of well-established dust extinction methods, and to provide solutions for unavoidable difficulties when confronting existing ideas with the latest developments in observational astronomy. Before continuing with the specific goals of this thesis, I will give a concise overview of the most important aspects when studying interstellar extinction from an observational point of view. Moreover, I will put particular emphasis on the mathematical and physical principles behind measuring extinction with astronomical photometry. These concepts are an integral part of the work presented here and are furthermore not included in the published papers because they are usually assumed to be a prerequisite when reading manuscripts with astrophysical background.

### 1.1.3 Interstellar dust extinction

Dust accounts for about 1% of the mass of the interstellar medium in the Galaxy (e.g. [Draine and Li 2007](#)). Nevertheless, dust plays a key role in shaping the structural and chemical characteristics of galaxies, it is fundamental to the formation of stars and planets, and greatly influences the interstellar radiation field. Dust shields interiors of molecular clouds, it absorbs UV and optical light which is re-radiated at infrared wavelengths (e.g. [Bernstein et al. 2002](#)), it facilitates chemical reactions to form molecules (e.g. molecular hydrogen, [Hollenbach and Salpeter 1971](#)), and provides a cooling mechanism for dense clouds (e.g. [Krügel 2002](#)).

More than 200 years ago, the first observational account of the existence of absorbing interstellar dust was given by William Herschel, who observed stars and nebulae in the night sky with the best telescopes available at this time. He discovered shapes and structures in the arrangement of celestial objects and published a series of observational reports. In these reports, he recorded, among other things, that fewer stars were visible in some directions, compared to other sightlines. In 1784, he published an account of his observations on nebulae in the *Philosophical Transactions of the Royal Society of London* and noted:

*In my late observations on nebulae I soon found, that I generally detected them in certain directions rather than in others; that the spaces preceding them were generally quite deprived of stars, so as often to afford many fields without a single star in it.*  
([Herschel 1784](#))

The final proof of the existence of dark regions in the sky only came more than a century later, when Edward Emerson Barnard cataloged and published photographs of dark nebulae (Barnard 1919). Soon after, Wolf (1923) demonstrated an empirical connection between the total amount of extinction and the number density of observed stars. By comparing luminosities and angular diameters of open clusters, Trumpler (1930) finally demonstrated the physical nature of interstellar absorption, where he also found that distant clusters are redder. He gave a first quantitative estimate of interstellar extinction and attributed it to scattering processes due to the presence of gas.

Today, we know that interstellar extinction is the sum of two physical processes: absorption and scattering, where I will use the term extinction in this thesis to specifically refer to the attenuation of electromagnetic radiation in the interstellar medium<sup>3</sup>. Extinction is a fundamental characteristic which affects all flux measurements of astronomical sources and is furthermore the main subject of this thesis. The manuscripts, presented in Chapters 3 and 4, will give more insight into practical applications of extinction measurements. In contrast to the content of the manuscripts, in this section I will introduce the mathematical and physical concepts with respect to interstellar extinction which are used in this thesis. In particular, I will derive a quantity that can be measured by photometric observations to characterize extinction along the line of sight toward point source measurements, i.e. stars or galaxies.

I start by defining two fundamental parameters for astronomical photometry:

$$m_\lambda \equiv -2.5 \log F_\lambda(d) + c \quad (1.2)$$

$$M_\lambda \equiv -2.5 \log F_\lambda(d = 10 \text{ pc}) + c. \quad (1.3)$$

Here,  $m_\lambda$  refers to the apparent magnitude of an object,  $M_\lambda$  denotes the so-called absolute magnitude, while  $F_\lambda$  is the measured flux of an object, i.e. the energy at wavelength  $\lambda$  per unit area per time. While the apparent magnitude depends on the distance  $d$  to an object, the absolute magnitude measures its luminosity as if the object was viewed from a distance of 10 pc. In this notation, the constant  $c$  includes attenuation effects and generally also depends on the optical system that is used

---

<sup>3</sup>Attenuation of electromagnetic radiation not only occurs in the interstellar medium. In general, absorption and scattering effects need to be taken into account whenever radiative transfer is at work. This applies to the interior of stars and their atmospheres and also, for example, to the Earth's atmosphere which has an attenuation effect on measured fluxes, strongly depending on the object's altitude above the horizon. A full analytical description of the problem can be given with the equation of radiative transfer, which, in contrast to the discussion here, also takes emission effects into account.

(and so requires a calibration of the measurements). With these two equations, together with the fact that the measured luminosity of an object decreases with the squared distance ( $F \propto d^{-2}$ ), one can define the distance modulus  $\mu$

$$\mu = m_\lambda - M_\lambda = 5 \log \left( \frac{d}{10 \text{ pc}} \right). \quad (1.4)$$

To obtain a description of interstellar extinction, the monochromatic energy per area per unit time which passes through a surface, is defined as

$$F_\lambda = \int I_\lambda d\Omega, \quad (1.5)$$

where  $\Omega$  refers to the solid angle of the observed object (with  $\Omega = \pi R_\star^2/d^2$  for a star's disk with stellar radius  $R_\star$ ) and  $I_\lambda$  refers to the specific intensity. The specific intensity is the energy per frequency interval passing through a surface of unit area from a cone with unit solid angle. Following radiative transfer, the intrinsic intensity  $I_{\lambda,0}$  and the intensity measured by an observer at distance  $d$  are connected via

$$I_\lambda = I_{\lambda,0} \cdot e^{-\tau_\lambda}, \quad (1.6)$$

where  $\tau_\lambda$  is referred to as optical depth at wavelength  $\lambda$  and reflects attenuation effects between the emitter and the observer.

The solid angle of a star, located at distance  $d$ , is sufficiently small to write Equ. 1.5 as

$$F_\lambda = I_{\lambda,0} \cdot \frac{\pi R_\star^2}{d^2} \cdot e^{-\tau_\lambda}. \quad (1.7)$$

Furthermore, the flux density for a similar, but nearby star ( $\tau_\lambda \approx 0$ , distance  $d_0$ ) is given by

$$F_{\lambda,0} = I_{\lambda,0} \cdot \frac{\pi R_\star^2}{d_0^2}. \quad (1.8)$$

Dividing Equ. 1.7 by Equ. 1.8 eliminates the stellar radius and the specific intensity and yields

$$\frac{F_\lambda}{F_{\lambda,0}} = \frac{d_0^2}{d^2} \cdot e^{-\tau_\lambda}. \quad (1.9)$$

By rewriting this equation in logarithmic form

$$\log F_\lambda - \log F_{\lambda,0} = \log \left( \frac{d_0^2}{d^2} \right) + \log e^{-\tau_\lambda} \quad (1.10)$$

and by applying minor modifications

$$-2.5 \log F_\lambda + 2.5 \log F_{\lambda,0} = 5 \log \left( \frac{d}{d_0} \right) + 1.086 \tau_\lambda, \quad (1.11)$$

one obtains a similar form as given in Equations 1.2 and 1.3. Specifically, with  $d_0 = 10 \text{ pc}$

$$m_\lambda - M_\lambda = 5 \log \left( \frac{d}{10 \text{ pc}} \right) + 1.086 \tau_\lambda. \quad (1.12)$$

In contrast to the above definition of the distance modulus in Equ. 1.4, this equation includes another term which incorporates the attenuation of radiation between a source and the observer. Specifically, the last term in the above equation connects the optical depth  $\tau_\lambda$  to the magnitude system in astronomical photometry and defines the extinction at wavelength  $\lambda$  as

$$A_\lambda \equiv 1.086 \tau_\lambda. \quad (1.13)$$

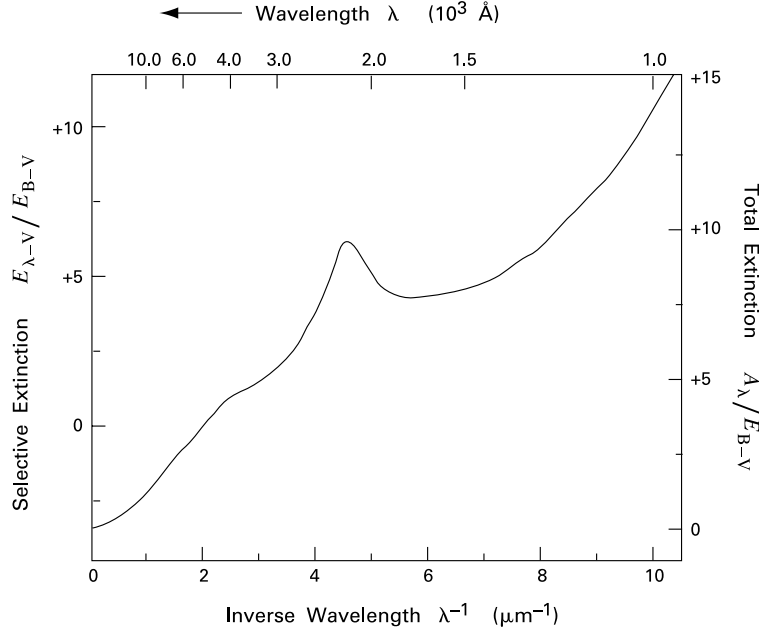
With Equ. 1.12 it becomes apparent, that measuring interstellar extinction requires, in principle, knowledge about intrinsic source characteristics (i.e. its distance, or equivalently its absolute magnitude). These parameters, however, are usually very difficult to obtain. For this reason, the characterization of interstellar extinction is often based on photometric colors, i.e. the difference of magnitudes at two different wavelengths (different filters). Writing Equ. 1.12 for two different passbands  $m_1$  and  $m_2$ , the distance to the star can be eliminated and one obtains

$$(m_1 - m_2) - (M_1 - M_2) = A_1 - A_2. \quad (1.14)$$

Here, the term  $m_1 - m_2$  can be measured and  $M_1 - M_2$  refers to the intrinsic color of the star, which can be easier estimated than a star's luminosity. The reason for this lies in the fact that stellar colors typically follow sequences and depend only on intrinsic characteristics and extinction, while luminosities (magnitudes) are also distance-dependent. This specific problem of estimating intrinsic colors of stars (and galaxies) will be the main subject of Chapter 3. Furthermore, Equation 1.14 represents the so-called color excess  $E_{m_1-m_2}$ . This parameter is often found in the literature to define the total to selective extinction ratio  $R$ . Specifically, for the  $V$  band this ratio is defined as

$$R_V \equiv \frac{A_V}{E_{B-V}}. \quad (1.15)$$

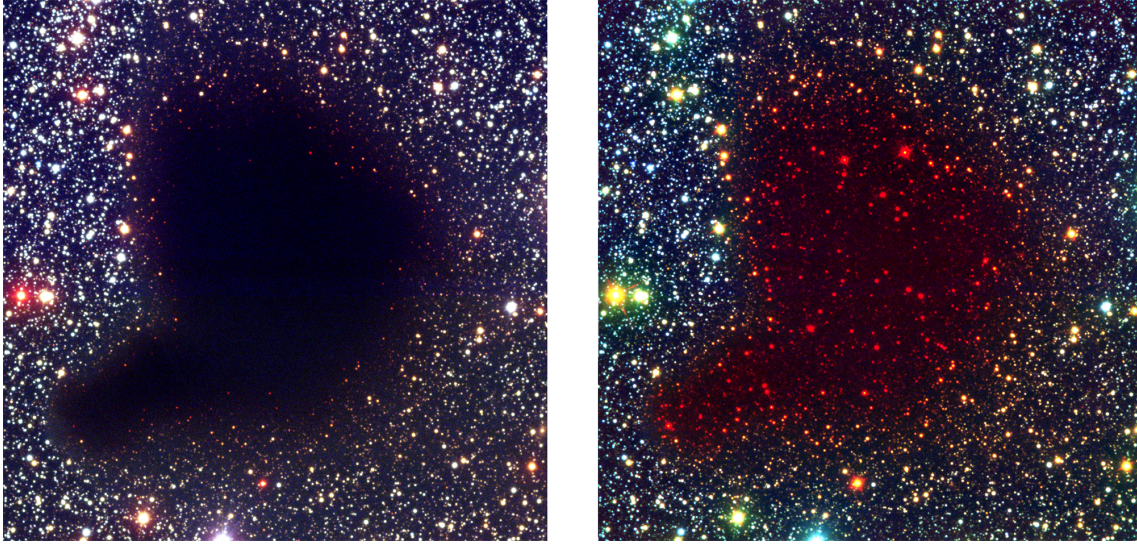
Other definitions of  $R$  at different wavelengths are also used. The wavelength



**Figure 1.3.** The interstellar extinction curve, parametrized with the color excess ratio (left axis) and the total extinction normalized by the  $E_{B-V}$  color excess (right axis). In general, the UV and optical spectral range are much more attenuated than longer wavelengths, giving rise to reddening. This figure was adopted from [Stahler and Palla \(2005\)](#).

dependence of extinction is shown in Fig. 1.3 where it becomes apparent that shorter wavelengths are much more affected by the attenuation. This general trend leads to the effect of reddening: Because shorter wavelengths preferably are subject to more extinction, sources appear red (compare Equ. 1.14). An example of the reddening of sources is given in Fig. 1.4 where the dark cloud Barnard 68 is shown in composite images at different wavelengths. While background sources do not appear at all in the visual bands, infrared light is able to penetrate through the dense layers due to less attenuation.

Observations of diffuse and dense environments showed that the extinction curve is characterized by large variations in the UV to optical wavelength range. [Cardelli et al. \(1989\)](#) introduced a parametrization of the extinction curve from the UV to NIR spectral range, which uses a total of seven variables. They furthermore continued to show that all seven parameters are strongly-correlated with  $R_V$  (Equ. 1.15), allowing to predict the shape of the extinction curve with just one single parameter. In general, a value of  $R_V=3.1$  is often taken to represent the average extinction law in the Galaxy, but studies have reported value ranging from 2.1 to 5.8. Sightlines associated with molecular clouds typically show larger-than-



**Figure 1.4.** Interstellar reddening visualized in the dark cloud Barnard 68. The left panel is a three-color composite made from observation in the  $B$  ( $0.44\ \mu\text{m}$ ),  $V$  ( $0.55\ \mu\text{m}$ ), and  $I$  ( $0.9\ \mu\text{m}$ ) bands. The image on the right-hand side includes data from the  $B$ ,  $I$ , and  $K$  ( $2.2\ \mu\text{m}$ ) bands. Image credit: João Alves/ESO.

average  $R_V$  values, which is oftentimes interpreted as evidence for the presence of large dust grains and consequently grain growth.

In contrast to UV and optical wavelengths, the situation in the near- and mid-infrared is more controversial. Here, some studies report a rather uniform extinction law between diffuse and dense environments, while others claim the opposite (Sect. 4.3 lists detailed examples from the literature)<sup>4</sup>. An absolute key element to estimating gas masses in molecular clouds is therefore to derive correct line-of-sight extinction toward sources in the background of molecular clouds. For this purpose it is necessary to (a) measure and understand the extinction law itself and (b) to derive correct estimates of interstellar reddening via the above given equations.

## 1.2 Specific goals of the thesis

Interstellar extinction is a powerful tool to study the earliest stages of star formation by enabling the investigation of molecular clouds in an unbiased and robust

<sup>4</sup>More details on individual characteristics of the extinction curve (such as the  $9.7\ \mu\text{m}$  absorption feature) will be given in Chapter 4 where I present an investigation of the infrared extinction law in Orion A molecular cloud complex.



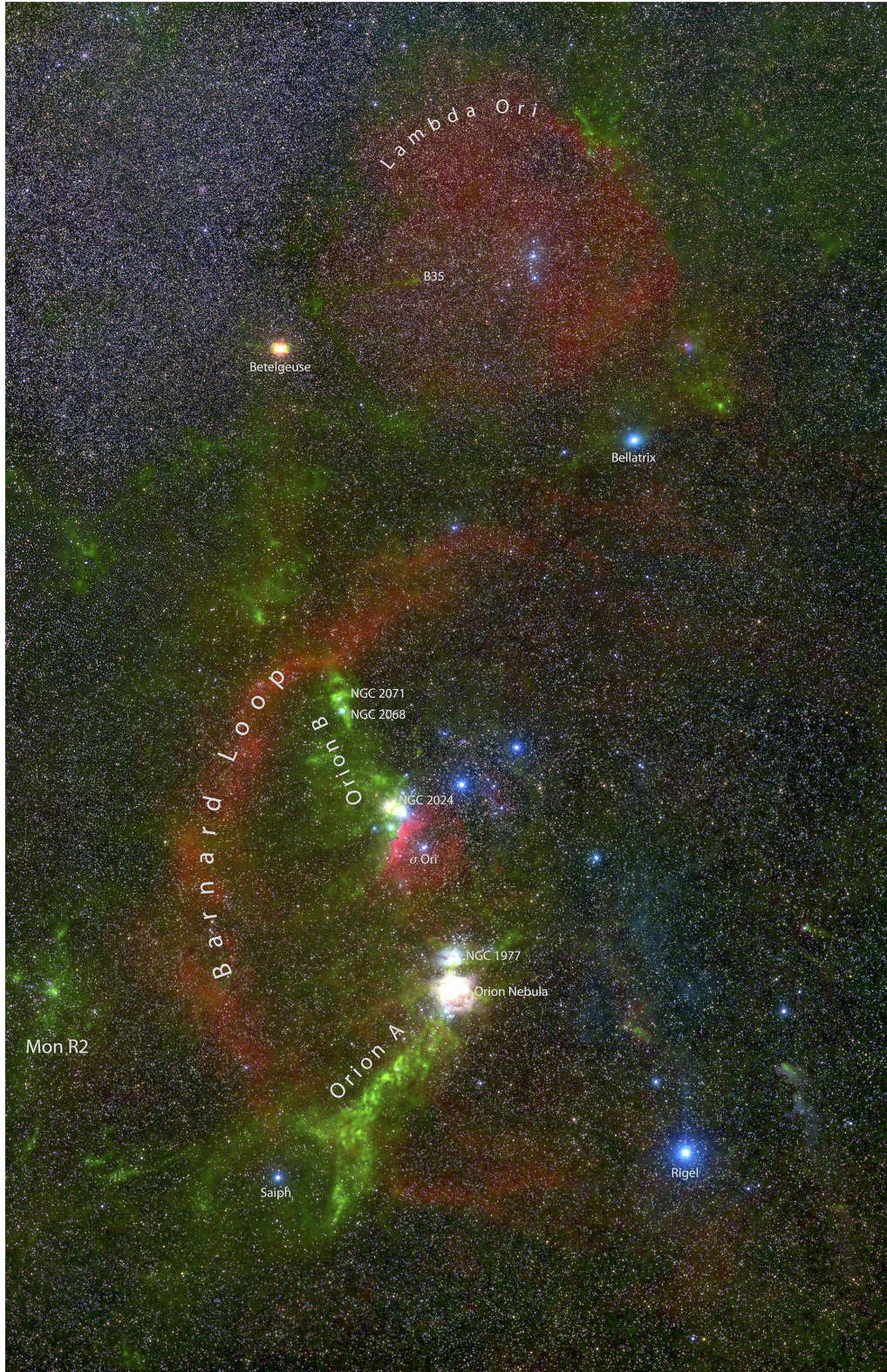
manner. Specifically, nearby molecular cloud complexes ( $d \lesssim 500$  pc) are unique laboratories to expand our knowledge in this research field. By studying the nearby complexes, structures can be resolved in detail and their young stellar population is observable down to the hydrogen burning limit and even below. Among the nearby molecular clouds, one particular star-forming complex is unique in itself: Orion. Situated near the galactic anti-center and well below the galactic plane, Orion not only is home to the nearest massive star-forming region, but also harbors a series of other star-forming environments comprising a range of physical conditions. While other nearby regions form at the most only a few dozen new stars (Evans et al. 2009), Orion hosts thousands of YSOs (Megeath et al. 2012), possibly up to above 5000 (Megeath et al. 2016). Thus, the vast majority of new stars in the solar neighborhood are forming in Orion.

The Orion star-forming complex spans several tens of degrees in the sky and comprises two major molecular cloud complexes: Orion A and Orion B (see Fig. 1.5), where about 80% of the population of young stars in this region are associated with Orion A. Moreover, Orion A is home to the famous Orion Nebula Cluster (ONC) which reaches stellar surface densities well above  $1000 \text{ pc}^{-2}$  (Hillenbrand and Hartmann 1998). Much of our current knowledge about star formation has its origin in observational efforts in this region. A detailed overview of the most important work and an outline of the importance of Orion as a star formation laboratory in general are part of the introduction of the first manuscript included in this thesis (Sect. 2.3).

Like many other pioneering studies in astrophysical research, also this thesis is devoted to investigating the Orion star-forming region. Specifically, the work presented here is dedicated to expand our understanding of interstellar dust extinction by using state-of-the-art instrumentation in combination with innovative approaches and methods. The goals of the thesis can be summarized in four individual points.

- The foundation of the thesis project are newly acquired wide-field NIR imaging data of the largest star-forming molecular cloud complex within 500 pc. The data were acquired with the ESO VISTA telescope (see Sect. 1.2.1 for details) and constitute the most sensitive NIR observations of the entire region so far. Most observational studies of this region in the past concentrated on the clustered environment, associated with the Orion Nebula, and often ignore the fact that the cloud complex extends for several degrees toward the south. Thus, in contrast to essentially all other preceding work, these new





**Figure 1.5.** The Orion star-forming complex comprises two major molecular clouds: Orion A and Orion B. The clouds are visualized in green on top of an optical image. This figure was adopted from [Lombardi et al. \(2011\)](#). Image Credit: Wei-Hao Wang.



observations cover the entire Orion A molecular cloud complex and are not restricted to specific sub-regions of the cloud. The first goal was to analyze the telescope data, to implement a robust data reduction framework, and to demonstrate the power of large photometric databases with respect to young stellar populations.

- Traditional techniques to study the gas and dust distribution in molecular clouds with extinction are based on methods that are designed mostly for shallow photometry where the bulk of the detected sources are of stellar nature. One particular problem arises due to the fact that new state-of-the-art photometric observations can detect a significant amount of extra-galactic sources. It is therefore far from obvious, whether these methods remain applicable for next-generation surveys which offer a tremendous improvement in sensitivity. Consequently, the second goal of this thesis was to investigate and quantify such issues and develop and implement a new technique which avoids problems of comparable methods and is robust and easily applicable at the same time.
- The third goal of the thesis was to use the new Orion A photometric database, together with new extinction tools, to investigate dust extinction in Orion A. Specifically, I first characterize the NIR extinction law with respect to its often-claimed universality. Moreover, also environmental differences in the dust grain composition, possibly influenced by radiative feedback from the massive cluster stars, were subject to a thorough investigation. Finally, also a comparison between dust emission and dust extinction measurements was to be done.
- The three items listed above constitute the major results of this thesis. In addition to these goals, which solely address the Orion A molecular cloud, all work was planned to prepare for even more extensive follow-up observations. Here, the goal was to prepare for other large-scale observing projects to obtain the best possible NIR photometric measurements for all other major nearby star-forming regions. Thus, this thesis can also be viewed as a pioneering study, before attempting to build a coherent picture of the star formation processes by investigating not only one single cloud, but by including all major nearby molecular cloud complexes.

To achieve these specific goals, highly sensitive, large-scale NIR observations were planned and requested in the P90 ESO observing semester (October 2012 – March

2013)<sup>5</sup>. Up to the point of offering large modern NIR survey telescopes to the astronomical research community, such an undertaking would have been an immensely large investment of observing time. This is because the entire cloud complex spans several degrees on the sky and smaller telescopes, equipped with cameras with limited fields-of-view, would have required several weeks (or even months) of observing time to homogeneously map the region as a whole with similar sensitivity.

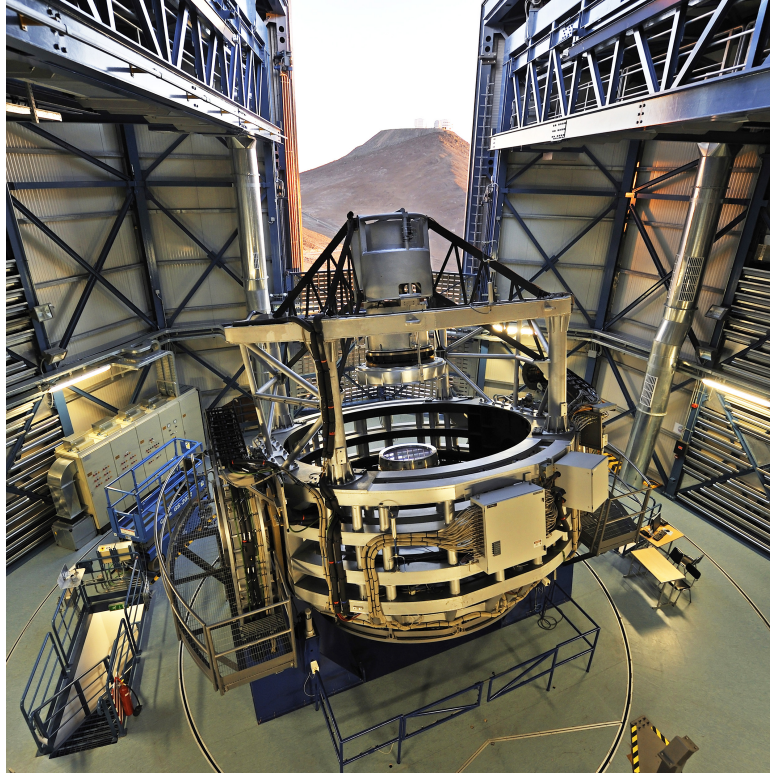
Apart from the above outlined objectives, the observations were motivated by several other science cases. These include, for example, the characterization of YSOs in the cloud which can be done in unprecedented detail via a combination of NIR and mid-infrared (MIR) photometry and visual indicators from the high resolution imaging data. All efforts with respect to the Orion A observing project were later aggregated under a unique project name: Vienna Survey in Orion, in short VISION. VISION constitutes the running title for two manuscripts in this thesis and further studies are planned to continue the series. Before outlining the individual chapters in this thesis, I will introduce the employed telescope in more detail. Similar to defining the fundamentals of interstellar dust extinction in Sect. 1.1.3, also the detailed description of the telescope technology is not part of the published manuscripts, but is included here for completeness.

### 1.2.1 The Visible and Infrared Survey Telescope for Astronomy

Large parts of this thesis are based on data collected with the Visible and Infrared Survey Telescope for Astronomy, in short VISTA. The telescope is located in the Atacama desert in Chile (longitude  $70^{\circ} 23' 51''$  west, latitude  $24^{\circ} 36' 57''$  south, 2518 m altitude). There, it is integrated into the ESO Paranal facilities, along with the Very Large Telescope array (VLT) and the VLT Survey Telescope (VST). In contrast to the other facilities, VISTA is located on a separate mountain peak, approximately 1.5 km northward of Cerro Paranal. Figure 1.6 shows the telescope inside its dome structure with the VLT telescopes in the background. The telescope itself is equipped with a 4.1 m f/1.0 main mirror. In combination with the secondary mirror, the telescope features a very fast optical design with an effective f/3.25 focal ratio in the Cassegrain focus. An active optics systems controls both the position of the secondary mirror, as well as the figure of the primary mirror. Here, low-order

---

<sup>5</sup>The original submitted ESO proposal is included in the Appendix in Sect. A.4.



**Figure 1.6.** The Visible and Infrared Survey Telescope for Astronomy, VISTA. The telescope is located in Chile and integrated into ESO’s VLT facilities. Also visible are the VLT Unit Telescopes in the background on top of Cerro Paranal. Credit: G. Hüdepohl ([atacamaphoto.com](http://atacamaphoto.com))/ESO

wavefront sensors, running in parallel to science observations, assure the correct positioning of the secondary mirror. With the active optics, the optical system allows for a theoretical image quality of  $0.51''$ , which will be degraded further by atmospheric turbulence. Typical seeing conditions for this location are well below  $1''$  with a median of  $0.66''$  (<https://www.eso.org/sci/facilities/paranal/astroclimate.html>).

Currently, VISTA is operated with a single instrument located in the Cassegrain focal station: the VISTA Infrared Camera, VIRCAM (Dalton et al. 2006). As the name suggests, VIRCAM is an imaging system which allows to take observations in the infrared spectral range. It is equipped with 16 sparsely placed Raytheon VIRGO  $2048 \times 2048$  pix detectors (for a total of 67 Mpix), which provide a nightly data-rate of several hundred GB. In addition, the camera offers a series of filters. These passbands<sup>6</sup> allow to take imaging data from  $\sim 0.9 \mu\text{m}$  to  $\sim 2.2 \mu\text{m}$ , and specifically include the broadband filters  $z$ ,  $y$ ,  $J$ ,  $H$ ,  $K_S$ , and the narrowband filters NB980, NB990, and NB118. The camera’s focal plane array is schematically displayed in

<sup>6</sup>Throughout this thesis, the terms *passband* and *filter* are used interchangeably.





years and typically use hundreds of hours of telescope time. The first cycle of VISTA public surveys started in P85 (April 2010) and has already been completed. In total six programs were carried out: (1) UltraVISTA ([McCracken et al. 2012](#)); (2) VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, [Edge et al. 2013](#)); (3) VISTA Magellanic Clouds Survey (VMC, [Cioni et al. 2011](#)); VISTA Variables in the Via Lactea (VVV, [Minniti et al. 2010](#)); VISTA Hemisphere Survey (VHS, [McMahon et al. 2013](#)); VISTA Deep Extragalactic Observations Survey (VIDEO, [Jarvis et al. 2013](#)). The call for the second generation public surveys with VISTA/VIRCAM has been published in 2015, for which the selection process has already been completed and the observations commenced in April 2017. Seven new programs have been selected: (1) VISTA Near-infrared Observations Unveiling Gravitational Wave Events (VINROUGE, <http://www.star.le.ac.uk/nrt3/VINROUGE/>); Completing the legacy of UltraVISTA (<http://home.strw.leidenuniv.nl/~ultravista/>); The VVV extended ESO Public Survey (VVVX, <http://vvvsurvey.org/>); The VISTA Extragalactic Infrared Legacy Survey (VEILS, [http://www.ast.cam.ac.uk/~mbanerji/VEILS/veils\\_index.html](http://www.ast.cam.ac.uk/~mbanerji/VEILS/veils_index.html)); Southern H-ATLAS Regions Ks-band Survey (SHARKS); VISTA Star Formation Atlas (VISIONS, <https://visions.univie.ac.at/>). The last project mentioned here, VISIONS, is led by a team at the University of Vienna. As part of my dissertation, I have significantly contributed to its eventual success and give a detailed description of the project in Sect. 5.2 of this thesis.

After the completion of these second generation public surveys, the VISTA telescope is scheduled for refitting in the timeframe between 2019 and 2021. Specifically, VIRCAM will be dismantled and replaced by the 4-meter Multi-Object Spectrograph, 4MOST ([de Jong et al. 2012](#)). In contrast to the now-installed infrared camera, VIRCAM, this new instrument will provide the ability for large-scale spectroscopic surveys for more than 2000 objects simultaneously. Thus, after VIRCAM will be decommissioned, there will no longer be large-scale NIR imaging capabilities available from the ESO infrastructure. Nevertheless, the new instrument will provide excellent follow-up observations for the previous generation of public surveys (Sect. 5.2.7).

The telescope's location on the southern hemisphere enables many synergies with other state-of-the-art observatories in the area. Specifically, the large VLT Unit Telescopes and the Atacama Large Millimeter/submillimeter Array (ALMA) provide observations in many other wavelength ranges with a large variety in available instrumentation configurations. Furthermore, VISTA observations will also play an important role in the future, when ESO Extremely Large Telescope (ELT) becomes operational in 2024. In particular, and with respect to the ELT, VISTA photometry

will play a fundamental role in enabling the photometric calibration of imaging observations with the ELT first-light imager MICADO (Davies et al. 2016; Leschinski et al. 2016). Typical external calibrators, such as 2MASS do not have the dynamic range to support a reliable calibration for such large telescope optics. For example, the  $K_S$  band saturation limit of typical MICADO observations will be at  $\sim 15$  mag, which is approximately the completeness limit of 2MASS observations. Moreover, other modern large-scale NIR survey facilities, such as those using the UKIRT Infrared Deep Sky Survey (UKIDSS Lawrence et al. 2007), are located in the northern Hemisphere.

## 1.2.2 Thesis Overview

The main results presented in this thesis are structured into four major chapters. While the Chapters 2 and 3 represent already published manuscripts for which I am the first author, Chapter 4 is a submitted manuscript which currently is subject to the journal’s reviewing process. Chapter 5 summarizes the results and presents an extensive outlook for follow-up work in the next few years. All manuscripts in the Chapters 2 through 4 are included in their original form and are preceded by an overview and publication details, including the full bibliographic reference. Finally, Appendix A contains supplementary material. In the following lines, I will give a short description of each chapter.

### Chapter 2 – Paper I

The first publication in this thesis represents the VISTA Orion A survey paper which also includes most of the original collaborators of the initial ESO observing proposal (available in Sect. A.4). I introduce the Orion A complex as a major opportunity to study star formation processes and continue to describe all major technical aspects which were developed for the presented VISTA observations. The main product of this paper is the, so far, largest photometric NIR catalog of the cloud complex, as well as publicly available image data. To this end, I developed a new approach to reduce the raw telescope data to avoid many caveats in the standard ESO pipeline. Furthermore, I perform a basic statistical analysis of the young stellar population of the cloud and demonstrate the data quality by identifying previously unknown YSOs. I also show, that this dataset can be used far beyond star formation studies by applying a simple algorithm to find new galaxy clusters.



### **Chapter 3 – Paper II**

The second chapter in this thesis builds on the photometric data as published in the manuscript in Chapter 2. During my work on the Orion A NIR catalog release, I performed preliminary tests which showed that well-established techniques to derive line-of-sight extinctions to single source are not suited for an application with deep photometry. I outline the main disadvantages of other methods and continue to introduce a new method, PNICER, which is capable of overcoming many of the mentioned difficulties. In the future, this method will be the baseline of extinction studies, including follow-up observations with VISTA (see Sect. 5.2).

### **Chapter 4 – Paper III**

With the developments from the first two manuscripts in Chapters 2 and 3, I use the VISTA photometry, together with PNICER to study dust extinction in the Orion A molecular cloud complex. The main findings of this chapter refer to a variable extinction law, suggesting a changing dust grain composition across the cloud. Furthermore, this trend seems to follow the picture that radiative feedback from the massive cluster stars impacts general grain properties. Moreover, I find a flat MIR extinction law, and uncover some previously unknown biases in the calibration of gas surface density maps based on dust emission measurements.

### **Chapter 5 – Conclusions & Outlook**

I conclude this thesis with a summary of the main findings in the presented manuscripts from Chapters 2 through 4. I also briefly outline possible follow-up work that could, in general, improve our understanding of star formation in Orion A and, in particular, also our understanding of the dust grain population in the cloud. Finally, I present a currently ongoing project which will be the main focus of my time during my first post-doctoral appointment (Sect. 5.2). The project's name is VISIONS, which already hints at the intimate connection to my thesis work. VISIONS will provide Orion A-like observations for other nearby cloud complexes. Furthermore this project will measure, for the first time, dynamical properties of embedded sources on large scales in the nearby star-forming clouds, opening up a new parameter space to study the evolution of young stellar systems.

### **Appendix**

Appendix A includes a German version of the Abstract (Sect. A.1), personal acknowledgements (Sect. A.2), the reprint permission for published articles (Sect. A.3), the

original ESO observing proposal for the Orion A VISTA project (Sect. [A.4](#)), as well as my Curriculum Vitae (Sect. [A.5](#)).

# VISTA Orion A survey

## 2.1 Overview

In the first publication presented in this thesis, I introduce the Vienna Survey in Orion, in short VISION. In this manuscript I present the Orion A molecular cloud as one of the most well-studied regions for research into fundamental questions regarding star formation and highlight important work over the past few decades. The main subject of the paper is a new ESO VISTA large-scale NIR imaging survey of the entire Orion A molecular cloud complex. The introduction puts this survey in context with other observational work in this region and clearly establishes the observations as the most extensive NIR imaging observing program of the cloud so far. The main body of the manuscript contains detailed descriptions on the applied methods, where it was necessary to implement a new software package to avoid several caveats in the standard ESO pipeline. Most notably, the new pipeline improves the spatial resolution of the final data products by about 20% compared to the reduction framework provided by ESO. This remarkable enhancement enables better photometry and was also a key motivation for the VISIONS public survey (see Chapter 5). Moreover, several other caveats in the original pipeline with respect to the photometric calibration of the image data are discussed and solutions are provided. The developed data reduction software environment is at the time of writing this thesis not available for third parties, but is currently being ported to Python for later publication.

After establishing all technical foundations for the data reduction recipes, I continue to describe the source detection and extraction methods and introduce an effective technique to separate resolved from point-like objects (for star-galaxy discrimination). This method is based on aperture corrections and uses machine learning classification algorithms. Next, I continue with a statistical analysis of the cloud's YSO population by comparing the luminosity functions of the extincted population to an extinction-free control field. Furthermore, by simple visual inspection, I identify several previously unknown YSO candidates. To further demonstrate the quality of the photometry, I also investigate the stellar foreground population to



the cloud, as discovered by [Alves and Bouy \(2012\)](#), and even identify new galaxy cluster candidates with a simple source density algorithm.

## 2.2 Publication details

**Title:** VISION - Vienna survey in Orion. I. VISTA Orion A Survey

**Authors:** Stefan Meingast, João Alves, Diego Mardones, Paula Stella Teixeira, Marco Lombardi, Josefa Großschedl, Joana Ascenso, Herve Bouy, Jan Forbrich, Alyssa Goodman, Alvaro Hacar, Birgit Hasenberger, Jouni Kainulainen, Karolina Kubiak, Charles Lada, Elizabeth Lada, André Moitinho, Monika Petr-Gotzens, Lara Rodrigues, Carlos G. Román-Zúñiga

**Status:** Accepted for publication in Astronomy & Astrophysics, December 1st 2015

**Bibliographic reference:** Astronomy & Astrophysics, Volume 587, id.A153, 31 pp.

**DOI:** [10.1051/0004-6361/201527160](https://doi.org/10.1051/0004-6361/201527160)

**Online data:** <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/587/A153>

**Own contributions:** Literature research, pipeline software development, data reduction, data analysis, source catalog generation, data distribution via VizeR, image combination, Orion A population analysis, preparation of figures and plots, paper writing.

## VISION – Vienna survey in Orion

### I. VISTA Orion A Survey<sup>\*,\*\*</sup>

Stefan Meingast<sup>1</sup>, João Alves<sup>1</sup>, Diego Mardones<sup>2</sup>, Paula Stella Teixeira<sup>1</sup>, Marco Lombardi<sup>3</sup>, Josefa Großschedl<sup>1</sup>, Joana Ascenso<sup>4,5</sup>, Herve Bouy<sup>6</sup>, Jan Forbrich<sup>1,7</sup>, Alyssa Goodman<sup>7</sup>, Alvaro Hacar<sup>1</sup>, Birgit Hasenberger<sup>1</sup>, Jouni Kainulainen<sup>8</sup>, Karolina Kubiak<sup>1</sup>, Charles Lada<sup>7</sup>, Elizabeth Lada<sup>9</sup>, André Moitinho<sup>10</sup>, Monika Petr-Gotzens<sup>11</sup>, Lara Rodrigues<sup>2</sup>, and Carlos G. Román-Zúñiga<sup>12</sup>

<sup>1</sup> Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, 1180 Wien, Austria  
 e-mail: [stefan.meingast@univie.ac.at](mailto:stefan.meingast@univie.ac.at)

<sup>2</sup> Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile

<sup>3</sup> University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

<sup>4</sup> CENTRA, Instituto Superior Tecnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal

<sup>5</sup> Universidade do Porto, Departamento de Engenharia Física da Faculdade de Engenharia, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

<sup>6</sup> Centro de Astrobiología, INTA-CSIC, Depto Astrofísica, PO Box 78, 28691 Villanueva de la Cañada, Madrid, Spain

<sup>7</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>8</sup> Max-Planck-Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

<sup>9</sup> Astronomy Department, University of Florida, Gainesville, FL 32611, USA

<sup>10</sup> SIM/CENTRA, Faculdade de Ciências de Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal

<sup>11</sup> European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany

<sup>12</sup> Instituto de Astronomía, UNAM, Ensenada, CP 22860, Baja California, Mexico

Received 10 August 2015 / Accepted 1 December 2015

#### ABSTRACT

**Context.** Orion A hosts the nearest massive star factory, thus offering a unique opportunity to resolve the processes connected with the formation of both low- and high-mass stars. Here we present the most detailed and sensitive near-infrared (NIR) observations of the entire molecular cloud to date.

**Aims.** With the unique combination of high image quality, survey coverage, and sensitivity, our NIR survey of Orion A aims at establishing a solid empirical foundation for further studies of this important cloud. In this first paper we present the observations, data reduction, and source catalog generation. To demonstrate the data quality, we present a first application of our catalog to estimate the number of stars currently forming inside Orion A and to verify the existence of a more evolved young foreground population.

**Methods.** We used the European Southern Observatory's (ESO) Visible and Infrared Survey Telescope for Astronomy (VISTA) to survey the entire Orion A molecular cloud in the NIR *J*, *H*, and *K<sub>S</sub>* bands, covering a total of  $\sim 18.3$  deg<sup>2</sup>. We implemented all data reduction recipes independently of the ESO pipeline. Estimates of the young populations toward Orion A are derived via the *K<sub>S</sub>*-band luminosity function.

**Results.** Our catalog (799 995 sources) increases the source counts compared to the Two Micron All Sky Survey by about an order of magnitude. The 90% completeness limits are 20.4, 19.9, and 19.0 mag in *J*, *H*, and *K<sub>S</sub>*, respectively. The reduced images have 20% better resolution on average compared to pipeline products. We find between 2300 and 3000 embedded objects in Orion A and confirm that there is an extended foreground population above the Galactic field, in agreement with previous work.

**Conclusions.** The Orion A VISTA catalog represents the most detailed NIR view of the nearest massive star-forming region and provides a fundamental basis for future studies of star formation processes toward Orion.

**Key words.** techniques: image processing – methods: data analysis – stars: formation – stars: pre-main sequence

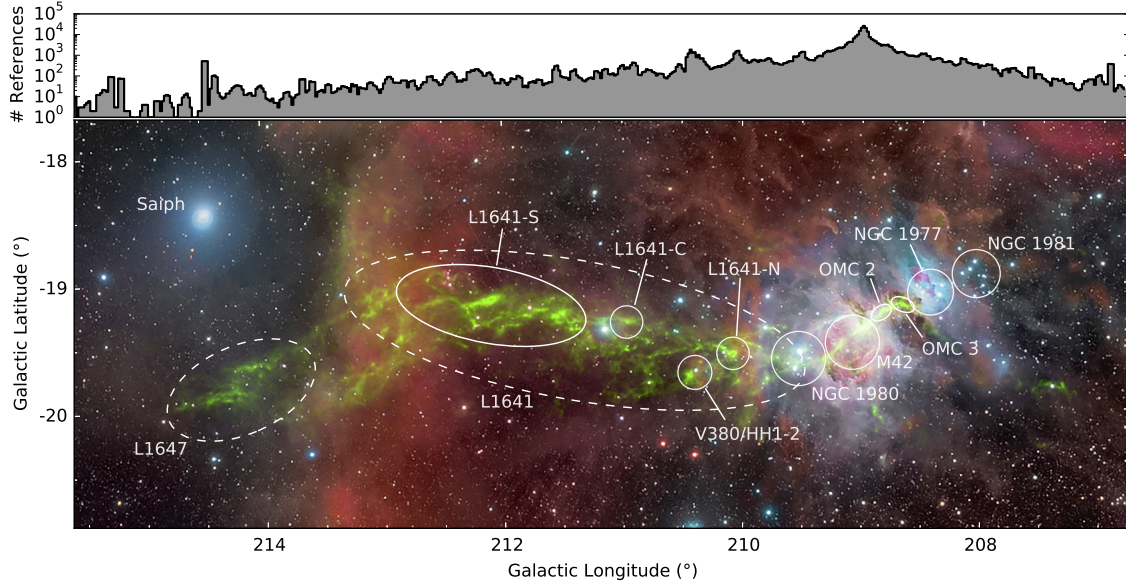
### 1. Introduction

One of the major obstacles since the beginning of star formation studies in the late 1940s is that stars are embedded in molecular gas and dust during their formation and early evolution, inaccessible to optical imaging devices. The deployment of infrared imaging cameras on optical- and infrared-optimized

telescopes during the past three decades has revolutionized the field, providing astronomers with the ability to detect, survey, and systematically study the earliest evolutionary phases of young stars within nearby molecular clouds. Since then, technological advancements have allowed us to constantly improve the sensitivity, resolution, and efficiency of infrared surveys culminating in the all-sky near-infrared (NIR) 2MASS survey (Skrutskie et al. 2006) and in space-borne observatories, such as the Infrared Astronomical Satellite (IRAS, Neugebauer et al. 1984), the *Spitzer* Space Telescope (Werner et al. 2004), the Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010), and the *Herschel* Space Observatory (Pilbratt et al. 2010). The

\* Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under program ID 090.C-0797(A).

\*\* Image data and full Table B.1 are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](http://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/587/A153>



**Fig. 1.** Composite of optical data (image courtesy of Roberto Bernal Andreo; deepskycolors.com) overlaid with *Planck-Herschel* column density measurements of Orion A in green. Approximate positions of noteworthy objects and regions are marked and labeled. On top, a histogram (note the logarithmic scaling) shows the number of references for all objects in the SIMBAD database at a given galactic longitude with a 3 arcmin bin size. We see an extreme gradient in attention paid to the various portions of the cloud, with the peak coinciding with M42 and the ONC. Prominent objects (e.g., the V380/HH 1–2 region) produce a local spike in the reference histogram whereas the bulk of the molecular cloud has been studied in comparatively few articles. The coordinates of L1647 in the SIMBAD database ( $l = 212.13$ ,  $b = -19.2$ ) do not match the original publication ( $l = 214.09$ ,  $b = -20.04$ ).

deployment of active- and adaptive-optics systems on 8–10 m class telescopes, combined with advanced instrumentation, has allowed ground-based NIR observations to match the supreme sensitivity of space-borne observatories but reaching higher spatial resolutions thanks to the larger apertures, which is critical to star formation research.

Over the past 25 years, systematic NIR imaging surveys of molecular clouds, in particular of the Orion giant molecular clouds, have revealed much of what we currently know about the numbers and distributions of young stars in star-forming regions. For example, the early foundational NIR surveys (e.g., Lada et al. 1991; Strom et al. 1993; Chen & Tokunaga 1994; Hodapp 1994; Ali & Depoy 1995; Phelps & Lada 1997; Carpenter 2000; Carpenter et al. 2000; Davis et al. 2009) revealed the importance of embedded clusters in the star-forming process. By combining information on the distribution of young stars with surveys of the distribution and properties of molecular gas, important insight into how nature transforms gas into stars have been gained (e.g., Lada 1992; Carpenter et al. 1995; Lada et al. 1997, 2008; Megeath & Wilson 1997; Carpenter et al. 2000; Teixeira et al. 2006; Román-Zúñiga et al. 2008; Evans et al. 2009; Gutermuth et al. 2009, 2011).

In the Orion star-forming complex, one finds a few of the best-studied testbeds for star formation theories, such as the embedded clusters NGC 2024, NGC 2068, and NGC 2071, as well as the optically visible young clusters  $\lambda$ ,  $\sigma$  Ori,  $\iota$  Ori, and NGC 1981. However, none of these regions have drawn nearly as much attention as the famous Orion nebula cluster (ONC), embedded in the Orion A molecular cloud. The ONC itself is the closest massive star factory and therefore a prime laboratory

for addressing many open questions of current star formation research. Many fundamental quantities regarding the formation of stars have been tested against this benchmark cluster, but orders of magnitude fewer studies have been published about objects in other parts of Orion A, and even fewer have addressed the molecular cloud as a whole, creating a biased view of the region. To help visualize this bias, we show in Fig. 1 a composite of an optical image overlaid on the *Herschel-Planck* column density map from Lombardi et al. (2014). Here we marked several objects and star-forming regions throughout Orion A, which is mentioned later in this paper. On top of the image we plot a histogram of the number of articles referenced in the SIMBAD (Wenger et al. 2000) database<sup>1</sup> for all objects at a given longitude slice in bins of 3 arcmin.

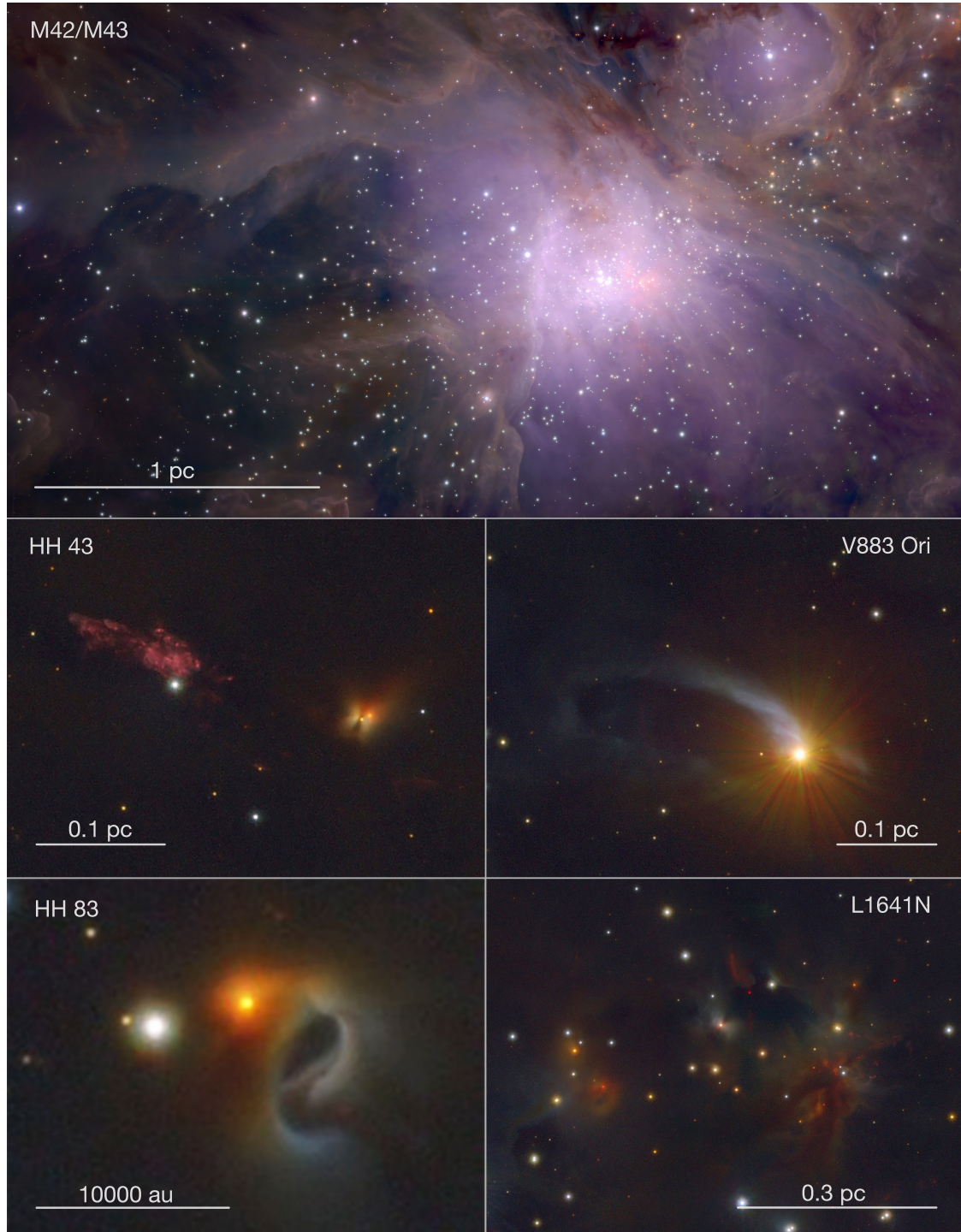
While the ONC and its surroundings are subject to various studies in thousands of articles, objects in the eastern region of the cloud (in galactic frame<sup>2</sup>) receive considerably less attention with a few tens of published studies. We note that the SIMBAD database is not complete, nonetheless these numbers are a good indicator of the bias in the astronomical community for some regions of Orion A. Figures 2 and 3 show examples of prominent objects observed in our survey of Orion A. Together, these objects alone have more than 5000 bibliographic references listed in the SIMBAD database.

While previous NIR surveys of Orion A have given us important insights, they are limited in their depth and sensitivity

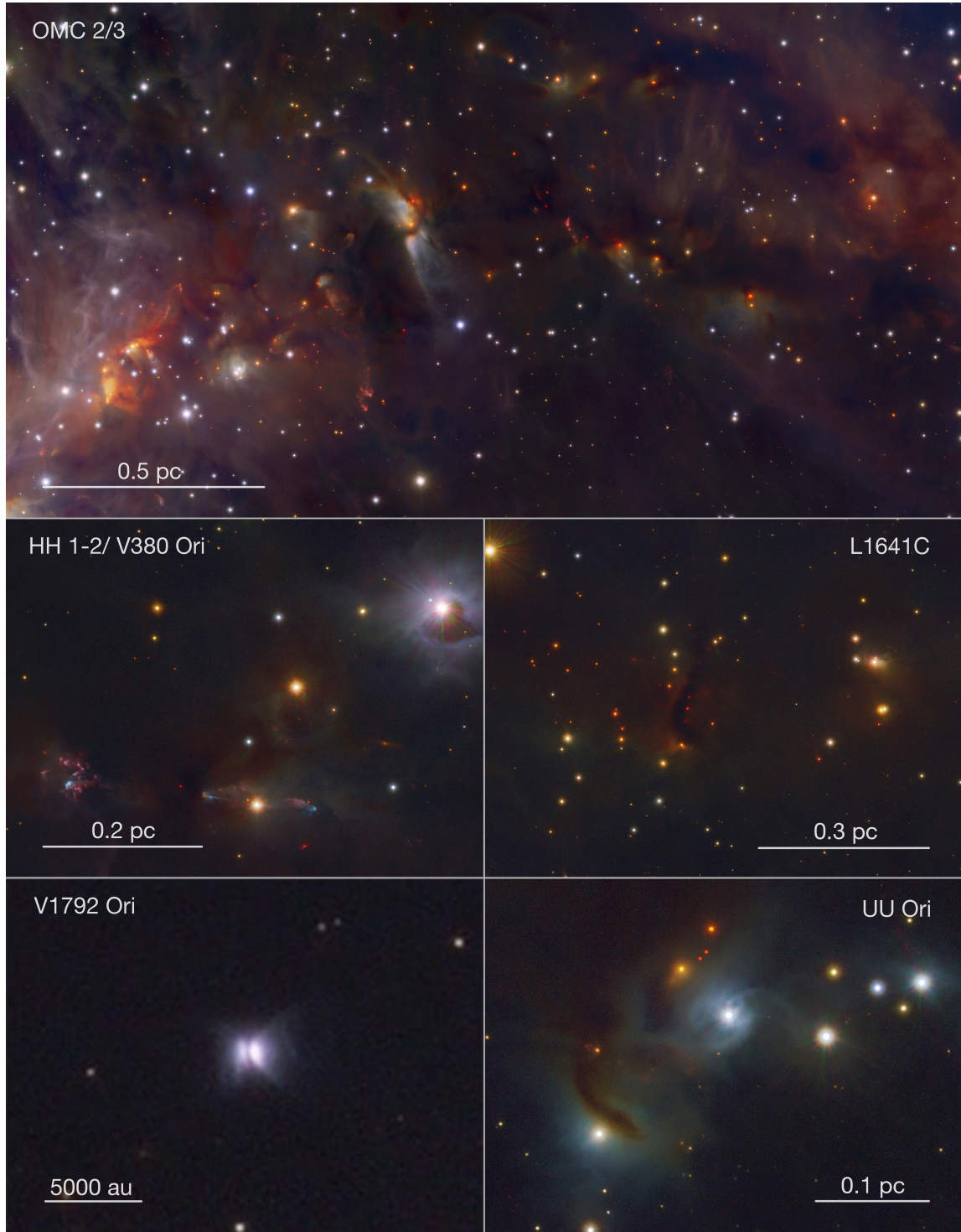
<sup>1</sup> The references were extracted from the SIMBAD database on 2015 May 10.

<sup>2</sup> During the rest of this paper, we always refer to the galactic coordinate frame when using cardinal directions.





**Fig. 2.** Detailed view of some prominent objects in Orion A as seen with VISTA. Here, the  $J$ ,  $H$ , and  $K_s$  bands were mapped to the blue, green, and red channels, respectively. All images are in a galactic projection (north is up, east is left). The physical length given in the scale bars was calculated with the adopted distance of 414 pc.



**Fig. 3.** Same as Fig. 2.



**Table 1.** On-sky coverage and relative gain in source counts for selected NIR surveys toward Orion A.

Reference	Coverage <sup>(a)</sup> (arcmin <sup>2</sup> )	Gain <sup>(b)</sup>	Bands
Strom et al. (1993)	2772	~4–6 <sup>(c)</sup>	<i>JHK</i>
Ali & Depoy (1995)	1472	4	<i>K</i>
Carpenter (2000) <sup>(d)</sup>	65 857	9.3	<i>JHK<sub>s</sub></i>
Lawrence et al. (2007) <sup>(e)</sup>	~26 500	1.4	<i>ZYJHK<sub>s</sub></i>
Robberto et al. (2010)	1200	1.4	<i>JHK<sub>s</sub></i>
This work	65 857		<i>JHK<sub>s</sub></i>

**Notes.** Our survey improves both coverage and sensitivity when compared to the literature. <sup>(a)</sup> Refers to the common on-sky area of the given survey with our VISTA coverage. <sup>(b)</sup> Approximate gain in source counts when restricted to the same on-sky coverage. <sup>(c)</sup> Estimate based on completeness limits since no source catalog is available. <sup>(d)</sup> Study based on the second incremental 2MASS data release. Source counts in this table were taken from the final 2MASS all-sky data release. <sup>(e)</sup> Data from UKIDSS DR10. Because of the many spurious detections of nebulousity in the UKIDSS survey, we estimated the gain in source counts by selecting a “clean” subregion.

and/or only cover a fraction of the entire molecular cloud. As a fundamental step toward a complete picture of the star formation processes in Orion A, we present the most sensitive NIR survey of an entire massive star-forming molecular cloud yet. Table 1 lists NIR surveys throughout the past two decades. Compared to the ONC surveys from the 1990s, our survey is about four times more sensitive (in terms of source counts) and covers a ~50 times larger area at the same time. Moreover, we also increase source counts by about 40% compared to the more recent dedicated NIR survey of the ONC by Robberto et al. (2010). Compared to the Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006), which obviously has a greater coverage, we gain almost a factor of 10 in sensitivity. For completeness we mention here that a similar survey has been conducted of the Orion B molecular cloud. These results are presented in Spezzi et al. (2015).

The target of our survey, the Orion A giant molecular cloud, extends for about 8 deg (~60 pc) and contains several well-studied objects and an extensive literature: we refer the reader to the review papers of Bally (2008), Briceno (2008), O’Dell et al. (2008), Allen & Davis (2008), Alcalá et al. (2008), Muench et al. (2008), and Peterson & Megeath (2008). Here we only list a selection of the many results for this important region, including studies of the ONC (Hillenbrand & Hartmann 1998; Lada et al. 2000; Muench et al. 2002; Da Rio et al. 2012), Herbig-Haro objects (HH; for a historic overview, see, e.g., Reipurth & Heathcote 1997), such as HH 1-2 (see, e.g., Herbig & Jones 1983; Lada 1985; Fischer et al. 2010) and HH 34 (e.g., Reipurth et al. 2002), and variable FU Ori type pre-main-sequence stars such as V883 (e.g., Strom & Strom 1993; Pillitteri et al. 2013). Along the “spine” of Orion A, there are also multiple noteworthy minor star-forming regions, such as L1641-N (e.g., Gálfaik & Olofsson 2008; Nakamura et al. 2012), which are themselves, however, much less prominent than the Orion nebula and its surroundings.

Studies referring to the entire cloud are rare. Megeath et al. (2012) present a *Spitzer*-based catalog of young stellar objects (YSO) for both Orion A and Orion B. They identify 2446 pre-main-sequence stars with disks and 329 protostars in Orion A. Pillitteri et al. (2013) present an *XMM-Newton* survey of L1641 where they investigate clustering properties of

Class II and Class III YSOs. They find an unequal spatial distribution in L1641, which suggests multiple star formation events along the line of sight, in agreement with the interpretation of Alves & Bouy (2012) and Bouy et al. (2014), and migration of older stars. More recently, Lombardi et al. (2014) have used a 2MASS dust extinction map (Lombardi et al. 2011), along with Planck dust emission measurements, to calibrate *Herschel* data and construct higher angular-resolution and high dynamic range column-density and effective dust-temperature maps. Stutz & Kainulainen (2015) investigate variations in the probability distribution functions of individual star-forming clouds in Orion A and suggest a connection between the shape of the distribution functions and the evolutionary state of the gas.

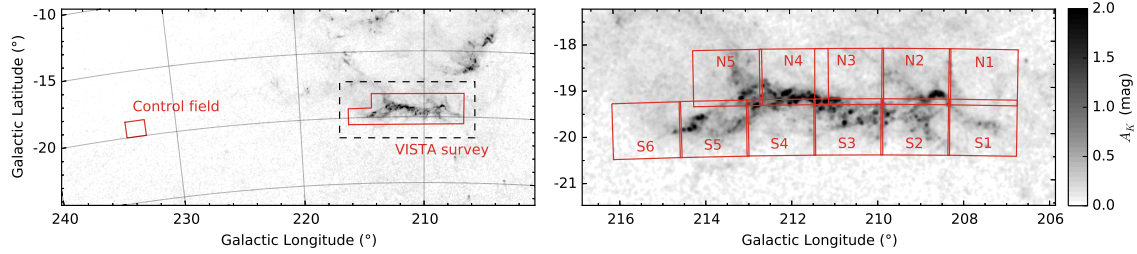
Regarding the overall evolution of the Orion star-forming region and following Blaauw (1964), Gomez & Lada (1998) speculated on the presence of multiple overlapping populations in the direction of the ONC with a possible triggered star formation scenario. As also mentioned by Bally (2008), recent studies by Alves & Bouy (2012) and Bouy et al. (2014) reveal a slightly older foreground population associated with NGC 1980 with distance and age estimates of ~380 pc and 5–10 Myr, respectively. They find 2123 potential members for this foreground population, which, however, is an incomplete estimate, because they did not cover the entire Orion A molecular cloud owing to lack of data in the eastern regions. Based on a shift in X-ray luminosity functions across Orion A, Pillitteri et al. (2013) also find evidence of a more evolved foreground population near NGC 1980 at a distance of 300–320 pc. Proposing an alternative view, Da Rio et al. (2015) find that sources near NGC 1980 do not have significantly different kinematic properties from the embedded population, concluding that NGC 1980 is part of Orion A’s star formation history and is currently emerging from the cloud.

Early distance estimates from Trumpler (1931) placed the ONC at 540 pc. Subsequent studies find distances of  $480 \pm 80$  pc (Genzel et al. 1981),  $437 \pm 19$  pc (Hirota et al. 2007),  $389^{+24}_{-21}$  pc (Sandstrom et al. 2007),  $440 \pm 34$  pc ( $392 \pm 32$  pc with a different subset of target stars, Jeffries 2007), and  $371 \pm 10$  pc (Lombardi et al. 2011). Based on optical photometry and a *Planck*-based dust screen model, Schlafly et al. (2014) find a distance of  $420 \pm 42$  pc toward the ONC, while the eastern edge of Orion A appears to be 70 pc more distant. For the remainder of this paper, we adopt the distance of  $414 \pm 7$  pc from Menten et al. (2007).

The VISION (VISTA Survey In Orion) data presented in this paper (Orion A source catalog and three-color image mosaic) are made available to the community via CDS. In future publications the survey data will allow us and the community to refine, extend, and characterize several critical properties of Orion A as a whole. This includes characterizing individual YSOs with the improved resolution and sensitivity, searching for HH objects and jets in a uniform manner, characterizing YSO clustering properties, determining IMFs down to the brown dwarf regime, and describing the gas mass distribution with respect to YSO positions.

This article is structured as follows. In Sect. 2 we present a survey overview, including its design and observing strategy. Section 3 describes all our data-processing procedures from basic image reduction, co-addition, astrometric, and photometric calibration to catalog generation and cleaning. In Sect. 4 we review the main data products of our survey and provide a first look at the resulting photometry, including the possibilities for accessing both the generated source catalog and image data. In addition, we present a catalog of interesting objects that includes some new YSO candidates based on their morphological





**Fig. 4.** VISTA survey coverage. The *lefthand side plot* shows the wide-field extinction map from Lombardi et al. (2011) with both the control field and Orion A coverage marked as red boxes. The *right hand side figure* shows a close-up of Orion A with the individual tiles labeled. The cutout region of this figure is marked with a black dashed box in the *left plot*.

appearance and new candidate galaxy clusters. In Sect. 5 we present first results obtained from this new database, where we derive an estimate for the YSO population in Orion A and investigate the foreground populations. Section 6 contains a brief summary, and Appendices A and B contain additional information on the quality of the data products and supplementary data tables, respectively.

## 2. Observations

### 2.1. Instrumentation

The observations of the Orion A molecular cloud have been carried out with the Visible and Infrared Survey Telescope for Astronomy (VISTA, Emerson et al. 2006), a 4 m class telescope that is operated by the European Southern Observatory (ESO) as part of its Cerro Paranal facilities. A single instrument, the VISTA Infrared Camera (VIRCAM, Dalton et al. 2006), is attached to the telescope’s Cassegrain mount, which offers a range of broadband and narrowband filters in the NIR covering a wavelength range from about  $0.85 \mu\text{m}$  to  $2.4 \mu\text{m}$ . VIRCAM features a set of sixteen  $2\text{k} \times 2\text{k}$  Raytheon VIRGO detectors arranged in a sparse  $4 \times 4$  pattern. Each detector covers about  $11.6 \times 11.6$  arcmin on sky with gaps of 10.4 arcmin and 4.9 arcmin between them in the instrument’s X/Y setup, respectively. Working at a mean pixel scale of  $0.339 \text{ arcsec/pix}$  in both axes, the instrument field of view in the telescope’s beam is  $1.292 \times 1.017 \text{ deg}$ .

The detectors offer a quantum efficiency above 90% across the  $J$ ,  $H$ , and  $K_S$  bands but suffer from significant cosmetic deficiencies (e.g., bad pixel rows and columns, as well as bad read-out channels) and nonlinearity effects, which need to be taken care of during data calibration. The gaps between the individual detectors make it necessary to observe multiple overlapping fields for a contiguous coverage. This is achieved by a six-step offset pattern that can be executed in several ways. As a consequence of this observing strategy, the effective coverage (hence exposure time) over a single field varies with position. The standard offset pattern offers a coverage of as little as just one frame on the edge of the field, two frames for most of the area and up to six overlapping exposures for only a tiny portion of the final frame. As is usual for NIR observations, a dither or jitter<sup>3</sup> pattern is usually executed at each offset position to mitigate saturation effects and to increase the total frame coverage to facilitate bad pixel rejection during co-addition.

<sup>3</sup> Here we use the term dither for user-defined offset positions, whereas jitter refers to random telescope positioning.

For the rest of this paper, we use VISTA terminology to describe the telescope’s data products and its parameters: a simultaneous integration from all sixteen detectors is called a “pawprint”, and a fully sampled image resulting from the co-added frames of the six-step offset pattern is called a “tile”. The integration time for a single readout from all detectors is referred to as DIT (detector integration time), whereas multiples of these single integrations can be stacked internally before readout. The number of integrations in such a stack is referred to as NDIT.

### 2.2. Survey design and strategy

The Orion A molecular cloud is centered at approximately  $l = 210^\circ$ ,  $b = -19^\circ$  and extends for about eight degrees, which is well aligned with the Galactic plane. The spine of the cloud (i.e., regions with high extinction) is very narrow with only about  $0.3 \text{ deg}$  at its widest point. However, shallower extinction levels are observed much more widely, which did not allow us to cover the entire cloud with only one series of pointings along the molecular ridge. Therefore we designed the survey to feature 11 individual pointings with two parallel sequences of tiles aligned with its spine. For each tile we also included overlaps with its neighboring field to ensure a contiguous coverage. Figure 4 shows the final tile coverage on top of the extinction map from Lombardi et al. (2011). Also shown are our designations for each tile labeled from N1, ..., N5, S1, ..., S6 indicating row (north/south) and column position (west to east). An observation of a tile in one of the three filters defined an observation block (OB). All 11 tiles were observed in  $J$ ,  $H$ , and  $K_S$ , where all except the tiles N2 and S2 were executed with a standard jitter pattern with a maximum random throw within a  $25 \text{ arcsec}$  wide box centered on the initially acquired position. The N2 and S2 tiles include the ONC and therefore a large amount of extended emission. These positions were observed with a separate sky offset field centered at  $l = 209.272^\circ$ ,  $b = -21.913^\circ$ . Because the sky offset field had to be observed in addition to the science fields, the total duration of these sequences was greater than the maximum allowed OB length of one hour in the  $H$  and  $K_S$  bands. Therefore each of these four OBs was executed twice. Starting in October 2012 and spreading out over about six months until early March 2013, a total of 37 individual OBs were executed to complete the observations of Orion A.

For statistical comparisons in the following analyses, we also observed a control field (CF) in the same filters in addition to the science field, centered at  $l = 233.252^\circ$ ,  $b = -19.399^\circ$  (see Fig. 4). Table 2 gives a comprehensive overview of all collected data of Orion A including basic parameters and statistics of each observing sequence. For the  $H$  and  $K_S$  bands, the DIT

**Table 2.** Observing dates and basic parameters for the Orion A VISTA survey.

Tile ID	filter	Start time UT	DIT (s)	NDIT (#)	NJitter (#)	Airmass range	Image quality <sup>a</sup> (arcsec)	90% Completeness <sup>b</sup> (mag)
S1	<i>J</i>	2012/10/02 08:38:24	5	8	3	1.074–1.088	0.78–0.92	20.47
S1	<i>H</i>	2013/02/27 02:09:44	2	27	5	1.308–1.657	0.68–0.83	20.06
S1	<i>K<sub>S</sub></i>	2013/01/20 00:23:40	2	20	5	1.110–1.197	0.63–1.02	18.91
S2	<i>J</i>	2012/12/25 05:20:26	5	9	6	1.122–1.328	0.75–0.86	20.08
S2 (a)	<i>H</i>	2013/02/08 01:17:15	2	17	5	1.058–1.136	0.72–0.89	19.58
S2 (b)	<i>H</i>	2013/02/17 01:24:44	2	17	5	1.089–1.245	0.75–0.89	19.58
S2 (a) <sup>c</sup>	<i>K<sub>S</sub></i>	2013/01/25 00:35:11	2	15	5	1.060–1.125	0.87–1.06	18.85
S2 (b)	<i>K<sub>S</sub></i>	2012/10/04 08:07:02	2	15	5	1.058–1.112	0.62–0.75	18.85
S3	<i>J</i>	2012/10/05 08:39:35	5	8	3	1.054–1.064	0.61–0.69	20.87
S3	<i>H</i>	2013/03/02 01:12:20	2	27	5	1.137–1.319	0.70–0.84	20.01
S3	<i>K<sub>S</sub></i>	2013/01/30 00:33:06	2	20	5	1.054–1.093	0.67–0.81	19.08
S4	<i>J</i>	2012/11/13 08:09:40	5	8	3	1.103–1.143	0.70–0.92	20.48
S4	<i>H</i>	2013/03/01 02:18:03	2	27	5	1.304–1.665	0.65–0.91	20.27
S4	<i>K<sub>S</sub></i>	2013/02/09 01:30:51	2	20	5	1.048–1.093	0.63–0.88	19.13
S5	<i>J</i>	2013/02/24 02:53:52	5	8	3	1.352–1.461	0.66–0.92	20.41
S5	<i>H</i>	2013/03/06 00:30:46	2	27	5	1.072–1.189	0.62–0.82	20.25
S5	<i>K<sub>S</sub></i>	2013/02/16 01:45:14	2	20	5	1.076–1.162	0.81–1.04	18.97
S6	<i>J</i>	2013/02/24 03:35:28	5	8	3	1.573–1.750	0.76–0.96	20.23
S6	<i>H</i>	2013/03/09 00:22:19	2	27	5	1.066–1.182	0.80–1.09	19.69
S6	<i>K<sub>S</sub></i>	2013/01/31 00:30:20	2	20	5	1.037–1.080	0.72–1.00	18.95
N1	<i>J</i>	2012/11/13 07:48:48	5	8	3	1.108–1.141	0.72–0.86	20.55
N1	<i>H</i>	2013/02/27 00:09:19	2	27	5	1.072–1.135	0.65–0.86	20.26
N1	<i>K<sub>S</sub></i>	2013/01/27 00:38:09	2	20	5	1.077–1.122	0.69–0.84	19.17
N2	<i>J</i>	2013/02/25 02:23:36	5	9	6	1.298–1.753	0.78–0.97	20.14
N2 (a)	<i>H</i>	2013/02/28 01:07:48	2	17	5	1.125–1.337	0.66–0.84	19.69
N2 (b)	<i>H</i>	2013/03/01 00:56:10	2	17	5	1.112–1.306	0.64–0.78	19.69
N2 (a)	<i>K<sub>S</sub></i>	2013/01/28 00:40:27	2	15	5	1.060–1.107	0.81–0.93	18.86
N2 (b)	<i>K<sub>S</sub></i>	2013/01/29 00:45:10	2	15	5	1.059–1.096	0.84–0.98	18.86
N3	<i>J</i>	2012/11/04 08:10:11	5	8	3	1.070–1.092	0.58–0.78	20.84
N3	<i>H</i>	2013/03/03 01:58:40	2	27	5	1.270–1.588	0.70–0.89	20.16
N3	<i>K<sub>S</sub></i>	2013/02/01 00:31:35	2	20	5	1.056–1.093	0.66–0.88	19.20
N4	<i>J</i>	2012/11/15 07:41:11	5	8	3	1.075–1.103	0.64–1.03	20.70
N4	<i>H</i>	2013/03/03 23:58:43	2	27	5	1.050–1.111	0.60–0.79	20.16
N4	<i>K<sub>S</sub></i>	2013/02/15 00:19:06	2	20	5	1.045–1.050	0.80–1.18	18.83
N5	<i>J</i>	2013/02/24 03:14:42	5	8	3	1.452–1.590	0.78–0.93	20.40
N5	<i>H</i>	2013/03/07 01:16:29	2	27	5	1.158–1.369	0.62–0.84	20.18
N5	<i>K<sub>S</sub></i>	2013/02/16 00:33:10	2	20	5	1.038–1.073	0.88–1.02	18.92
CF	<i>J</i>	2013/01/02 07:40:14	5	8	3	1.484–1.624	0.63–0.76	20.67
CF	<i>H</i>	2013/02/15 03:48:50	2	27	5	1.216–1.489	0.65–0.95	19.78
CF	<i>K<sub>S</sub></i>	2013/02/18 04:08:35	2	20	5	1.339–1.609	0.67–0.87	18.99

**Notes.** <sup>(a)</sup> The image quality refers to measured FWHM estimates of point-like sources, which varies across each tile because of camera distortion and variable observing conditions. <sup>(b)</sup> Completeness estimates are derived from the full combined Orion A mosaics and are calculated on the basis of artificial star tests. Details on the method are described in A.1. <sup>(c)</sup> Rejected in co-addition due to large differences in image quality with respect to Tile S2 (b.)

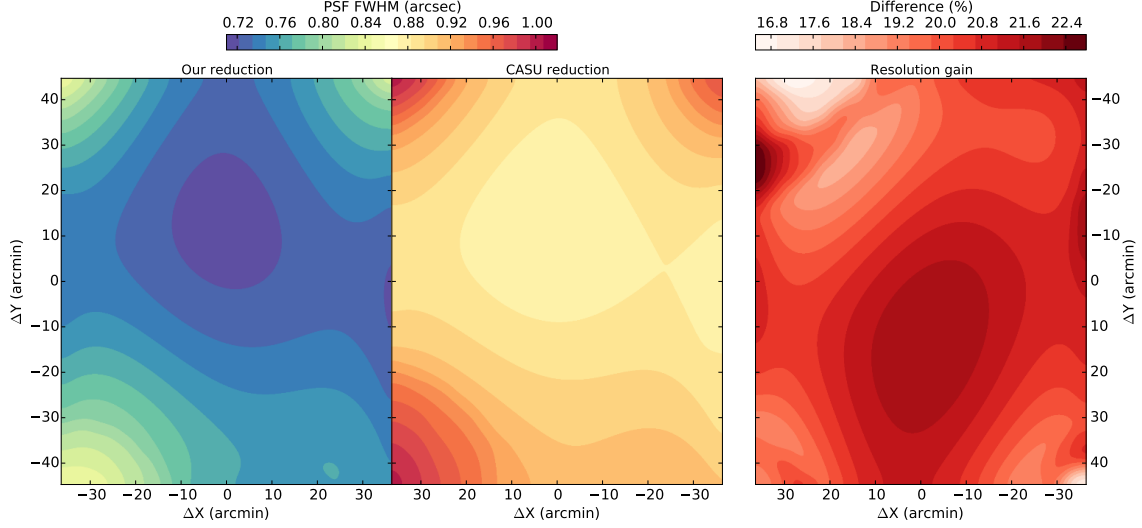
was chosen to be only 2 s owing to saturation issues and was compensated by increasing numbers of NDIT ranging from 15 to 27 for these bands. In the *J* band we reached a more efficient duty cycle with larger DITs since saturation is less critical at this wavelength. The number of jitter positions at each of the six telescope offsets to form a tile was chosen to be a minimum of 3 for the *J* band to allow for reliable bad pixel rejection. For the *H* and *K<sub>S</sub>* bands, we observed five jittered positions for each paw-print. The total on-source exposure time is given by the product of the minimum exposure time (DIT × NDIT) and the number of observations taken at this position determined by NJitter and the six-step offset pattern. For the large majority of sources in a tile, this is given by DIT × NDIT × NJitter × 2.

In addition to the science and control fields, calibration frames were also needed to process the raw files into usable data products. Dark frames, sky flat fields for each band, and dome

flats to measure detector nonlinearity were provided as part of ESO’s standard calibration plan for VIRCAM.

### 3. Data processing

A total amount of ~280 GB of science data, along with ~680 GB of calibration frames, was obtained for the Orion A VISTA survey. Together with the complex observing routine and camera setup, only a dedicated pipeline is able to handle the data reduction procedure. Calibrated science data products are available through the VISTA data flow system (Irwin et al. 2004) provided by the Cambridge Astronomical Survey Unit (CASU). However, since the pipeline is designed for stability and is optimized for reducing a much larger amount of data under many different observing conditions, we identified several drawbacks in this system. To achieve the best possible data quality of the



**Fig. 5.** FWHM maps for both our reduction and the standard CASU pipeline of tile S1 in *H* band. Clearly a significant gain in image quality is achieved in our reduction.

VISTA Orion A survey, we decided to implement all key data reduction procedures ourselves. Details on the methods, including a mathematical description of the CASU pipeline modules, can be found in the VISTA data reduction library design and associated documents<sup>4</sup>.

### 3.1. Motivation

Below we list the specific points that motivated us to develop our own customized reduction pipeline for the Orion A VISTA data.

- Owing to the observing strategy with VIRCAM, sources are sampled several times not only at different detector positions but also by different detectors. To optimally co-add all reduced paw prints, each input image needs to be resampled and aligned with a chosen final tile projection. The CASU pipeline uses a radial distortion model, together with fast bilinear interpolation, to remap the images for the final tiling step. Bilinear interpolation, however, has several drawbacks. Primarily it can introduce zero-point offsets and a significant dispersion in the measured fluxes. Typically a Moiré pattern is also seen on the background noise, and additionally it “smudges” the images, leading to lower output resolution (see Bertin 2010 for details and examples). To test for the absolute gain in resolution over the bilinear resampling kernel, full width half maximum (FWHM) maps for each observed tile were calculated with PSFex (Bertin 2011). Figure 5 shows a FWHM map for the tile S1 in *H* band for both the CASU and our reduction, as well as the gain in resolution. We typically achieve 20% higher resolution, i.e. a 20% smaller FWHM, by simply using more suitable resampling kernels<sup>5</sup> (for an interesting in-depth discussion

of the importance of resampling methods, see, e.g., Lang 2014). We find that our resampling method recovers the image quality of the pawprint level for the combined tiles, which is not the case for the CASU reduction.

- The CASU pipeline only produces source catalogs for individual tiles. As can be seen in Fig. 4, it would be beneficial to co-add all input tiles to increase the effective coverage on the tile’s edges and run the source extraction on the entire survey region. The spatially correlated noise (which is not traced by weight maps) and the necessity of yet another resampling pass make this step highly undesirable for the CASU tiles.
- Even with such short integration times as in our survey, stars brighter than ~12th magnitude in *J* (11.5 and 11 mag in *H* and *K<sub>s</sub>*, respectively) show saturation and residual nonlinearity effects when compared to the 2MASS catalog. Replacing these measurements with reliable photometry from 2MASS requires that both catalogs are calibrated toward the same photometric system. As already demonstrated by Gonzalez et al. (2011), among others, this is not the case, and a comparison with 2MASS requires the recalibration of the photometric zero point. Reliable color transformations can be found in Soto et al. (2013). We also tested this by producing magnitudes from the CASU tile catalogs via

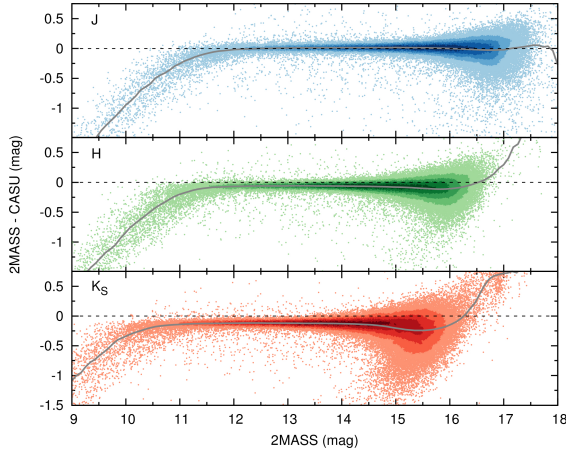
$$m_{i,j} = -2.5 \log \left( \frac{F_{i,j}}{t_j} \right) - \text{apcor}_j + \text{ZP}_j \quad (1)$$

where  $m_{i,j}$  are the calculated magnitudes for the  $i$ th source measured on the  $j$ th tile,  $F_{i,j}$  the flux measurements given in the CASU tile catalogs,  $t_j$  the exposure times,  $\text{apcor}_j$  the aperture corrections, and  $\text{ZP}_j$  the zero points as given in the tile headers. We also applied the appropriate zero-point transformation for the given atmospheric extinction and used the recommended aperture radius. We then concatenated all original tile catalogs and cross-matched the data with 2MASS. The comparison of the photometry is shown in Fig. 6 where magnitude differences between both data sets are displayed as a function of the 2MASS flux measurement.

<sup>4</sup> Accessible through <http://casu.ast.cam.ac.uk>, Lewis et al. 2010.

<sup>5</sup> We observed a dependency of the resolution gain on observing conditions. We get only 10–15% for bad seeing conditions (>1 arcsec) and up to almost 30% for excellent conditions (~0.6 arcsec)

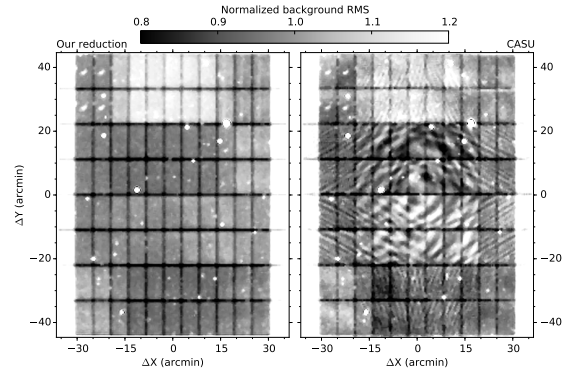




**Fig. 6.** Difference in the photometry between the CASU default reduction and the 2MASS catalog. The blue, green, and red data points show the  $J$ ,  $H$ , and  $K_S$  bands, respectively. The color shading represents source density in a  $0.2 \times 0.05$  box in the given parameter space and the gray solid line a running median along the abscissa with a box width of 0.5 mag. Several processing steps contribute to the apparent offsets, which are all avoided in our data reduction.

A systematic offset is visible with values around 0.1 mag in  $H$  and 0.12 mag in  $K_S$ .

- For the zero-point calculation of each observed field, the CASU pipeline applies a galactic extinction correction to all photometric measurements. To this end the pipeline uses the [Schlegel et al. \(1998\)](#) all-sky extinction maps (with a resolution of a few arc-minutes), together with the correction from [Bonifacio et al. \(2000\)](#). For each source, a bilinear interpolation yields the extinction correction factor for the zero point. This will also add systematic offsets with respect to photometric data for which no such correction was applied. More critically, for surveys covering multiple fields with variable extinction, systematic offsets are expected between the tiles. For studies concerned with the intrinsic color of stars (e.g., extinction mapping), however, it is critical not to be biased in any way by such systematic offsets.
- The CASU pipeline by default stacks all frames of an entire set to build a single background model for one tile. This only works well if spatial sky variations across the detector array are constant for the entire duration of the observations. In the NIR this typically applies for small sets of data with relatively short total exposure times, such as for the VISTA survey products (e.g., VVV, [Minniti et al. 2010](#)). However, since our OBs were at the limit of the maximum allowed execution time of 1 h, significant changes in atmospheric conditions are expected for nonphotometric nights. This can lead to residual gradients across single detector frames, which can result in cosmetically imperfect reductions and difficult sky level estimates. This is especially the case for fields with separate offset sky positions with large gaps in the sky sampling.
- In total there are two interpolation steps employed by the CASU pipeline. The first generates stacks for each jitter sequence, and the second is used during the tiling procedure to correct for astrometric and photometric distortions (the latter to account for variable on-sky pixel size due to field distortion). Both of these steps use bilinear interpolation, which can introduce spatially correlated noise. The



**Fig. 7.** Noise rms maps of the CF in  $K_S$  for our reduction and the CASU pipeline. The bilinear interpolation and the radial distortion model clearly leave spatially correlated noise in the tiled images. On the other hand, the variance in our reduction is only dominated by detector coverage and intrinsic detector characteristics. No sources were masked prior to noise calculations.

difference between the original bilinear interpolation and our data product, for which we use higher order resampling kernels, is illustrated in Fig. 7, where the background rms maps of the CF in the  $K_S$  band are shown. The radial distortion model, together with the fast interpolation, clearly leaves its mark. These rms maps were generated with SExtractor ([Bertin & Arnouts 1996](#)) and represent smoothed-noise rms models with a background mesh size of 64 pixels. For reliable source detection, however, one has to keep track of the variable noise throughout an image. As a consequence this makes it difficult to reliably run external source detection packages on the output CASU tiles in cases the pipeline does not work satisfactorily, as in regions with extended emission such as the ONC.

From all these points, only the photometric offset relative to the 2MASS system can be corrected for via color transformations. Bias-free photometry and high resolution are both critical for all further studies with the Orion A VISTA data. Therefore we have written a semi-automatic data-reduction package that is completely independent of the CASU pipeline. All functionalities of this package will be offered in open-source Python code in a future paper. The implemented reduction steps are discussed in detail during the following sections. In summary, the following capabilities have been implemented specifically for the VIRCAM reduction package:

- calculation of all required master calibration frames and parameters: bad pixel masks (BPM), dark frames, flat fields, and nonlinearity coefficients;
- basic image calibration: nonlinearity correction, removal of the dark current, and first-order gain harmonization with the master flat;
- accurate weight map generation for co-addition and source detection;
- static and dynamic background modeling;
- removal of cosmetic deficiencies (bad pixel masking, global background harmonization, etc.);
- illumination correction (second-order gain harmonization) using external standards;

- source detection, astrometric calibration, and co-addition via external packages;
- robust aperture photometry using variable aperture corrections;
- photometric calibration based on the 2MASS reference catalog (Vega magnitude system).

Many of the techniques are similar to the methods used in the CASU pipeline. However, the problems listed above are carefully avoided. All sequential data reduction procedures are described in the following sections.

### 3.2. Master calibration frames

The basic image reduction steps include the generation of all required calibration frames and parameters and their application to the raw science data to remove the instrumental signature from VIRCAM.

#### 3.2.1. Bad pixel masking

Before any other calibration step can be performed, a BPM is required to avoid introducing systematic offsets in, for instance, dark current calculations or linearity estimations. This step, however, has to be independent of any further calibration steps. Therefore we used a set of dome flats with constant exposure times that are first stacked at the detector level. The median of each detector served as a preliminary master flat and was then used to normalize each input image. Good pixels in each recorded flat field would then theoretically contain only values around unity due to the constant exposure time. Then, all pixels that deviated by more than 4% with respect to the expected unity value were marked. Finally, if a single pixel was marked in this way in more than 20% of all images in the sequence, it was propagated as a bad pixel to the final master BPM. Typical bad pixel-count fractions were found between 0.1% and 0.2% for the best detectors and around 2% for the worst.

#### 3.2.2. Nonlinearity correction

To correct for detector nonlinearities, we used the same method as for the CASU pipeline. For details on this method, the reader is referred to the VISTA data reduction library design document<sup>6</sup>. In principle a set of dome flat fields with increasing exposure time was first masked, i.e. with the BPM and pixels above the saturation level, and corrected for dark current with the accompanying dark frames. Then the flux was determined for the detector as the mode of each masked frame. The increasing exposure time should then provide a constant slope in the flux vs. exposure time relation for a completely linear detector with a given constant zero point (in double-correlated read mode used for our observations this offset should be close to 0). A least-squares fit to these data using a function of the form

$$\Delta I = \sum_{m=0}^3 b_m t_i^m [(1 + k_i)^m - k_i^m] \quad (2)$$

was performed, where  $i$  indicates each detector,  $m$  indicates the order of the function,  $\Delta I$  are the measured nonlinear fluxes for

the reset-corrected double-correlated read output,  $b_m$  are the coefficients to be solved for,  $t_i^m$  are the integration times, and  $k_i$  are the ratios between the reset-read overhead and the integration times. All least-squares fits in our reduction package made use of the MPFIT IDL library described in Markwardt (2009). These nonlinearity coefficients were stored in look-up tables and were later applied to each input frame by a simple nonlinear inversion. We also tested nonlinearity corrections on the channel level (each of the 16 detectors of VIRCAM hosts 16 separate readout channels) and found no significant differences in the output data quality.

#### 3.2.3. Dark current estimation

To estimate the dark current, a set of dark frames with the same exposure time parameters (i.e., DIT and NDIT) as the science frames was stacked at the detector level. The output of this procedure was a master dark pawprint that was calculated as the average of the pixel stack with a simple rejection of the minimum and maximum pixel value. We favored this method over a median because of the small number of available dark frames (typically five per unique DIT/NDIT combination for the VIRCAM calibration plan).

#### 3.2.4. First-order gain harmonization

For photometric consistency across all detectors, one has to calibrate all pixels to the same gain level. We used a series of twilight flats to correct for pixel-to-pixel gain variations. For camera arrays, however, it is usually not enough to create master flats for each detector separately since the detectors themselves also need to be brought to the same gain value with respect to each other. In a first step, all input flats were linearized, the dark current was removed, and bad pixels were masked. We then normalized all input pawprints by the median flux over all channels to account for the variable illumination, preserving detector-to-detector differences. The master flat field was then simply calculated as the median of the stacked, calibrated, and scaled input pawprints.

#### 3.2.5. Weightmaps

To accurately trace variable noise and bad pixels across the frames, we used weight maps initially generated from the master flat field. The normalized master flat field already accounted for variable sensitivity across the focal plane introduced by vignetting from filter holders and other detector/camera characteristics. We simply added bad pixels and rejected pixels with an unusually low/high response. These weights were later used for source detection to trace the spatial variations in background noise and during co-addition for an optimized weighting scheme.

#### 3.2.6. Saturation levels, read-noise, and gain

The saturation levels, read-noise, and gain of each detector are important parameters for any source detection method and error calculations. We determined the read-noise and gain following Janesick's method (e.g., Janesick 2001) on a set of dedicated calibration frames. Initial values for the saturation levels of each detector were taken from the VIRCAM user manual<sup>7</sup>. These were then checked during the calculation of the coefficients for the

<sup>6</sup> <http://casu.ast.cam.ac.uk/surveys-projects/vista/technical/data-processing/design.pdf/view>

<sup>7</sup> <https://www.eso.org/sci/facilities/paranal/instruments/vircam/doc.html>

nonlinear inversion as described in Sect. 3.2.2. The saturation level of Detector 6 had to be refined since we still saw significant nonlinearity below the given threshold. We lowered the original value of 36 000 ADU to 24 000 ADU. All calculated parameters were stored in look-up tables for later processing steps.

### 3.3. Science data calibration

After producing all the necessary calibration master files and parameters as described above, we consecutively applied the nonlinearity correction, dark current subtraction, gain harmonization, and bad pixel masking. In addition to these standard data reduction procedures, NIR data typically benefit from the removal of the (highly variable) background signature and detector-dependent cosmetic corrections.

#### 3.3.1. Background model

Additional additive background signatures (e.g., atmospheric emission, residual scattered light) can be removed by creating a background model. For our observing sequences with only a few individual exposures, we masked any contaminating sources prior to calculating the residual background. As a first step, we therefore created a static background model, calculated from a simple median of all stacked data, to allow for a rough first-pass source detection.

These temporary background models were applied to the science data, which in turn were used to create source masks with SExtractor. Very bright sources produced large halo structures on the VIRCAM detectors, which were simply masked by placing a circular mask with a radius proportional to a preliminary calculated magnitude:

$$m_{\text{preliminary}} = -2.5 \log\left(\frac{F}{t}\right) + ZP_{\text{VISTA}}, \quad (3)$$

where  $m_{\text{preliminary}}$  is the preliminary adopted magnitude,  $F$  is the measured flux from SExtractor,  $ZP_{\text{VISTA}}$  is the zero point for each band from the VISTA user manual, and  $t$  is the integration time (DIT  $\times$  NDIT). By manually comparing some sources to 2MASS, we typically found errors of only a few 10% for these estimates, which was sufficient for source masking. A star of magnitude eight received a mask with a 50 arcsec radius  $r$ . All other masks were calculated with  $\Delta r / \Delta \text{mag} = -10$  relative to this value. In addition, we also manually produced masks to cover regions of extended emission throughout Orion A.

Subsequently, dynamic background models with variable window sizes,  $w$ , were calculated, where  $w$  corresponds to the number of frames to include for each model. For our data, a compromise between accurate sky sampling and acceptable noise in the background model was found for values of  $w$  around 15 to 20 in the  $H$  and  $K_s$  bands, and  $w \approx 10$  for  $J$ . To this end we first normalized the input data by subtracting the mode of each frame and then calculated the median from the  $w$  closest input pawprints in time. In all cases with separate offset sky observations, the background models were calculated from the offset observations alone.

#### 3.3.2. Cosmetics

As a last step in the basic reduction, some cosmetic flaws, which were still visible after the preceding calibration stages, had to be taken care of. Mainly, a residual horizontal pattern could be seen on the background across all detectors. This, however, can

easily be removed by just subtracting the median of each detector row from all input frames and is referred to as “de-stripping” in the CASU reduction. Since de-stripping only works for images where regions of extended emission (if present) are smaller than a detector, we skipped this step for all frames that included the Orion nebula.

As a final step in the data reduction, bad pixels, as given in the BPM, were interpolated. This was necessary for successfully deploying our high-order resampling kernels owing to the large number of bad pixels on the VIRCAM detectors. Running such kernels on regions with bad pixels produces “holes” of the size of the kernel in the resampled frames, which were more difficult to reject in the final pixel stack and, in general, increased the overall noise level. Bilinear interpolation kernels, such as those used by the CASU pipeline, suffer considerably less from this problem, however at the cost of introducing more systematic errors. Bad pixel interpolation in general is not desirable since these should naturally be rejected during co-addition. To keep the impact at a minimum, we used a nonlinear bi-cubic spline interpolation code where only pixels that have fewer than 20% “bad neighbors” within a radius of four pixels were interpolated. For an overview of the impact of different interpolation methods see Popowicz et al. (2013), among others.

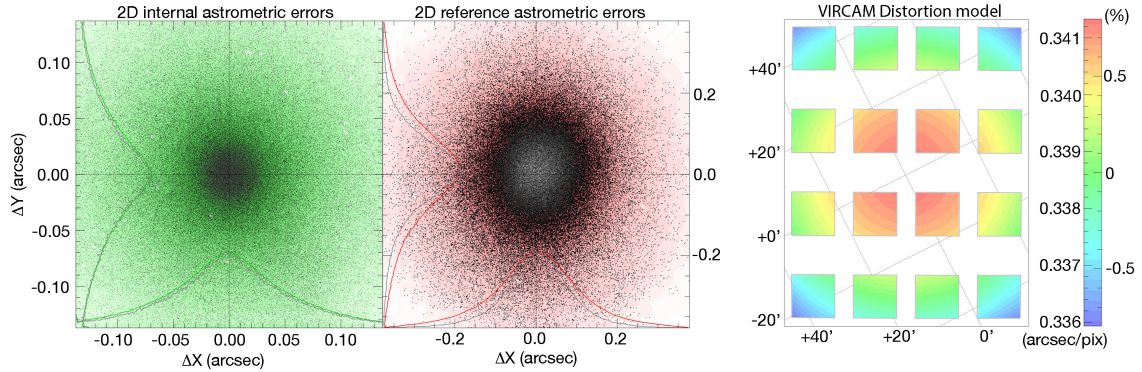
#### 3.3.3. Remarks

A first inspection of the data did not reveal any strong contamination by cosmic ray events. Also subsequent visual inspection of the reduced combined images showed only very few artifacts that might have originated in cosmic rays. Therefore, no attempt to identify and mask those was made. Also, no fringe correction was applied during any stage of the data processing. Fringes can occur for various reasons, such as interference effects in the detector or scattered light. If these patterns are not highly variable on spatial scales, they can be mistaken for sky background emission. They differ from them by variable amplitudes and different time scales and therefore, if present, must be removed in a separate reduction step. After inspecting many of the science frames in our survey, only very localized and low amplitude fringe patterns could be found, which were mostly taken care of during the background modeling and/or co-addition. Any attempts to correct for those small effects would have undoubtedly introduced more systematic errors, so they were neglected.

### 3.4. Astrometric calibration

Before any co-addition could be performed, the frames needed to be registered to a common reference frame. To calculate astrometric solutions, we used Scamp (v2.0.1, Bertin 2006). This software package performs pattern matching of arbitrary source catalogs with any available reference catalog, and subsequently determines accurate astrometric (and to some extent also photometric) solutions. From the reduced pawprints, relatively shallow source catalogs were generated with SExtractor to match the 2MASS dynamic range, which served as the basis for the astrometric calibration. Scamp offers several options for treating multi-extension Flexible Image Transport System (FITS, Pence et al. 2010) files, in our case focal plane arrays, depending on the reliability and completeness of the initial input parameters in the FITS headers. We tried several combinations of the available modes in Scamp, but only running the software in





**Fig. 8.** Verification plots created by Scamp for all data on Orion A in the  $K_S$  band. The two figures on the left show the 2-dimensional internal and external dispersions in the astrometric matching procedure. The symmetry of the Gaussian-shaped distributions indicates that the remaining systematic errors are negligible compared to other noise terms. The typical global errors (rms) were found around 40 mas internally and 70 mas with respect to the reference catalog. The plot on the righthand side shows the derived distortion model of VIRCAM, where the color indicates the variation in pixel scale across the focal plane. Only very small distortion levels are seen with an amplitude of  $\sim 2\%$  from the center to the edge.

LOOSE<sup>8</sup> mode offered an unbiased global astrometric solution without any systematics across the focal plane. In principle, the FIX\_FOCALPLANE<sup>9</sup> mode was also employed successfully, but in that case we observed systematic source clipping toward the outer detectors, resulting in mismatches between individual tiles.

All three bands were calibrated separately with a third-order distortion model over the focal plane. Figure 8 shows the internal (VISTA source-to-source scatter) and external (VISTA-to-2MASS scatter) astrometric errors along with the derived VISTA/VIRCAM distortion model as generated by Scamp. The dispersion (rms) for the global astrometric solution in all bands was between 40 and 45 mas with respect to internal source matches and about 70 mas with respect to external (2MASS) matches. This compares very well to the CASU mean rms value of 70 mas as given in the headers of the assembled tiles, which also uses 2MASS as an astrometric reference.

In addition to the astrometric solutions, Scamp can be used to derive photometric scaling factors to calibrate all input data to the same zero point. This method, however, has two major drawbacks: (a) Scamp only calculates zero-point offsets between entire pawprints and does not take residual detector-to-detector differences into account; and (b) Scamp requires single sources to be visible in all input data. The observing strategy, together with the sparse focal plane coverage, provides only a tiny overlapping field for all telescope pointings. For our jitter box width, we found overlaps smaller than 1 arcmin across. Therefore in most cases there would be no sources available in these overlaps. For these reasons Scamp cannot be used for a global fine-tuned gain harmonization based on relative internal source measurements alone. We therefore adjusted the relative zero points by comparing the source catalogs for each detector with 2MASS reference stars (see Sect. 3.5.3 for details). The typical internal photometric scatter at this stage was around 0.01 mag. For more details on the astrometric properties of our VISTA survey, see Appendix A.2.

<sup>8</sup> In this mode each detector is treated individually without a global focal plane model.

<sup>9</sup> Here Scamp attempts to derive a common WCS projection followed by computing the median of the detector positions with respect to the focal plane.

### 3.5. Tile and Orion A mosaic assembly

Prior to assembling the final mosaics, additional processing steps were required to produce science-ready data. These included resampling onto a common reference frame, global background modeling, and the fine-tuned gain harmonization. For quality control and computational reasons, we chose to co-add each single tile before assembling the final Orion A mosaic.

For all co-addition tasks during the data processing, we used the method of Gruen et al. (2014), who implemented an algorithm for optimized artifact removal while retaining superior noise characteristics in the co-added frame. In principle this method works in a similar way to a  $\kappa - \sigma$  clipping technique, but allows for an additional degree of freedom to account for variable point spread function (PSF) shapes. Not only did we observe excellent artifact removal, but also the standard deviation in the background was found to typically be 10–20% lower than a median-combined mosaic. The photometric calibrations referred to in the following sections are described in Sect. 3.6.

#### 3.5.1. Resampling

With the focal plane model and astrometrically calibrated science frames in place, the images were resampled onto a common reference frame using SWarp (v2.38.0, Bertin et al. 2002) using a third-order Lanczos kernel (Duchon 1979). To avoid complex flux-scaling applications across the tiles due to variable on-sky pixel sizes, we chose a conic equal area projection (COE, Calabretta & Greisen 2002) in equatorial coordinates with one standard parallel at  $\delta = -3^\circ$  as the projection type, the field center to be aligned with the center of each tile, and a pixel scale of 1/3 arcsec/pix. Choosing an equal area projection over the standard gnomonic tangential projection assured that every pixel covered the same area on-sky, which avoids further flux adjustments for the subsequent photometry.

#### 3.5.2. Global background modeling

During resampling, SWarp can fit a user-defined background model to each frame, but it neglects overlaps between the individual images. In cases of very crowded fields, particularly in

the presence of extended emission, this method introduces discontinuities across the full tiles. To correct for these last remaining offsets between overlapping images, we calculated a global background model with the Montage software package<sup>10</sup> while using very large mesh sizes (a constant offset for the tiles N2 and S2 and 1/3 of a detector for all other fields) with SWarp. It is important to note here that despite our efforts to apply as little spatial filtering as possible in the background correction, minor residuals are still visible throughout the assembled tiles. For this reason we do not encourage measurements of nebulous emission on our mosaics. We estimate that structures of few arcminutes in size should mostly be preserved in our reduction.

### 3.5.3. Illumination correction

Up to this point any zero-point offsets between the detectors were only corrected for during the calibration with the master flat field. This, however, proved to be mostly insufficient. Unaccounted-for scattered light in the optical train or imperfect flat fields are two examples of effects that can create variable photometric zero points over the field of view. As already mentioned above, it is not possible to attempt a gain harmonization based on internal photometric measurements from the science fields with VIRCAM owing to the nonexistent overlaps in the offset pattern. The calibration plan offers standard field observations specifically for this correction, but since these measurements could not be performed simultaneously with the Orion A field and are carried out only once per night or upon user request, it was safer to rely on external standard catalogs.

To this end, we defined subsets of the data for which we assumed stable photometric conditions with respect to the zero point and the PSF shape. Each of these subsets comprised one detector for each jitter sequence (5 frames for the  $H$  and  $K_S$  bands, 3 or 6 frames for the  $J$  band; compare with Table 2). Thus each tile was split into 96 subsets (16 detectors, 6 offset positions). The jitter box width and the execution time for each of these sequences were in a range where this assumption should hold. This assumption only breaks down for the tiles with offset sky fields (S2 and N2), where one of the six-step offset patterns was completed before any jitter was executed.

The images in each of these subsets were co-added, and for the resulting data we performed source extraction with SExtractor to calculate zero-point offsets relative to 2MASS. We then used the 96 determined zero points to calculate relative flux scaling factors.

### 3.5.4. Observing parameters

For quality control purposes, we also calculated several observing parameters for each of the given subsets as defined in Sect. 3.5.3. These include the local seeing conditions (FWHM; estimated with PSFEx), effective exposure time, frame coverage, and the local effective observing time (MJD). Most important, aperture correction maps were also generated for apertures with discrete radii of 2/3, 1, 2, 3, and 4 arcsec. The fluxes were corrected to an aperture of 5 arcsec for which no variation due to changing seeing conditions was expected. Only point-like sources (as classified by SExtractor) with a high signal-to-noise ratio (S/N) were included in the calculation of the aperture corrections. Examples of the quality control parameters are shown in Appendix A.3.

<sup>10</sup> <http://montage.ipac.caltech.edu/>

### 3.5.5. Co-addition

Once the photometric flux scaling was adjusted with the correct zero-point offsets, the original resampled frames were co-added to the final tiles with SWarp. We then again created shallow source catalogs and calculated relative zero-point offsets for each tile, and we finally merged all tiles into the Orion A mosaic. The final mosaic constructed from all data for each filter also features a COE projection with the same standard parallel and pixel scale as the individual tiles. For easier data access and three-color image assembly, we used the same projection for all filters.

Unfortunately, the two separate observations in  $K_S$  of tile S2 featured one of the best and one of the worst observing conditions in terms of image quality (FWHM), respectively. As a consequence, when co-adding these tiles, we saw a significant drop in S/N after source extraction. For this reason we decided to only include the data set taken during the better ambient conditions.

## 3.6. Photometric calibration

The recipes described in this section apply to all stages throughout the data processing where photometric calibration was performed.

### 3.6.1. Source detection and extraction

Source detection and extraction was performed with SExtractor where we tested several different detection thresholds with respect to the background noise level,  $\sigma$ . For the final source catalog we chose a threshold of  $1.5\sigma$ , requiring at least three connected pixels above this level, while lowering the default deblending threshold by two orders of magnitude to also detect sources in high-contrast regions. This combination proved to be optimal because a visual inspection of multiple regions in the mosaic showed only a few misdetections ( $<1\%$ ) of nebosity and residual artifacts. We interpreted the low threshold as a validation of the methods for creating the weight maps and co-added the data. For zero-point determinations and astrometric matching, the threshold was typically set to  $7\sigma$ .

The resulting source catalogs were cleaned by removing all bad measurements, i.e. sources with negative fluxes or a SExtractor flag larger than or equal to four (essentially saturated or truncated objects). For tiles and the final Orion A mosaic catalog, we applied the previously determined aperture corrections to all the extracted sources. In addition, each source was also assigned an effective observing MJD, exposure time, frame coverage, and local seeing value using the quality control data as described in Sect. 3.5.4.

### 3.6.2. Photometric zero point

For reliably determining the photometric zero point, only a limited dynamic range was used since bright stars still showed signs of nonlinearity, and for fainter stars we found a large dispersion relative to 2MASS owing to low S/N in the reference catalog. For the  $J$ ,  $H$ , and  $K_S$  bands, we used ranges of [12, 15], [11.5, 14.5], and [11, 14] mag, respectively. Also, we required an  $A$  quality flag in 2MASS for a reference source to be used in the zero-point determination, a SExtractor flag of 0 (i.e., no blending, truncation, or incomplete or corrupted data), and an error below 0.1 mag in our catalog. These requirements offer both good S/N values in our survey and 2MASS, and typically also several thousand available reference stars for a single tile. For the co-added

subset described in Sect. 3.5.3, we typically find several tens of sources matching these criteria. For cross-correlation between the catalogs, we searched for matches within a radius of 1 arcsec where, in cases of multiple possibilities, the nearest match was always selected.

The zero point was determined by applying a one-pass  $2\sigma$  clipping in the 2MASS – VISTA parameter space and by fitting a simple linear function with a forced slope of 1 to the data, weighted by the sum of the inverse measurement errors. Typical errors for the zero point were found to be around 0.01 mag. We also decided not to include color terms in the photometric calibration since (a) we did not see any significant dependency on those within the measurement errors; and (b) we aimed for separate calibrations for each individual filter without the need for detections in multiple bands.

### 3.6.3. Catalog magnitudes

In summary, magnitudes and errors were calibrated onto the 2MASS photometric system (in contrast to the CASU pipeline) and calculated using equations of the form

$$m_{i,r} = -2.5 \log \left( \frac{F_{i,r}}{t_i} \right) + \text{apcor}_{i,r} + \text{ZP}_{i,r} \quad (4)$$

$$\Delta m_{i,r} = 1.0857 \times \frac{\sqrt{A_{i,r} \sigma_{i,r}^2 + F_{i,r}/g}}{F_{i,r}} \quad (5)$$

where  $r$  refers to each aperture size,  $F_i$  are the measured fluxes,  $t_i$  the exposure times,  $\text{apcor}_i$  the aperture corrections,  $\text{ZP}_i$  the determined zero point,  $\Delta m_i$  the calculated errors,  $A_i$  the area of the aperture,  $\sigma_i$  the standard deviation of the noise, and  $g$  the gain. We note here that magnitude errors calculated in this way should only be taken as lower limits because (a) SExtractor does not include a term describing the absolute background flux in the aperture typically found in CCD S/N equations and (b) we do not include systematic errors. Furthermore, we do not include uncertainties from the determination of the zero point (typically  $\Delta m_{\text{ZP}} \approx 0.01$  mag.) for individual source magnitude errors. Since the errors are calculated independently for each source based on photon statistics alone, they can be considered random and may only show spatial correlations due to variable observing conditions and changes in the image quality across the focal plane array.

Finally, the adopted catalog magnitude for each source was chosen so that the selected aperture maximized the S/N among all measurements. After extensive tests, we found that the best overall measurement to represent fluxes for all sources can be achieved by selecting the catalog magnitude from only the two smallest apertures (2/3 and 1 arcsec).

### 3.7. Final catalog assembly

For the final source catalog, we applied the aforementioned source extraction procedures to the entire Orion A mosaic. In this way, we avoided the issue of multiple detections of the same sources and at the same time increased the S/N in the overlapping regions. In contrast to all intermediate catalogs, we included additional processing steps for assembling the final catalog of the full Orion A mosaic. We focused on four remaining issues:

1. morphological classification to distinguish between extended and point-like objects;

2. cleaning of spurious detections;
3. sources in the residual nonlinearity and saturation range;
4. source detection near the Orion Nebula due to significant and highly variable extended emission.

During the following sections we address these supplementary processing steps individually.

#### 3.7.1. Morphological classification

SExtractor itself is able to distinguish between extended and point-like objects on the basis of neural networks. The successful application of this method, however, critically depends on the input parameters; in particular, a correct guess of local seeing conditions (FWHM) is crucial. Since the FWHM varies by more than a factor of 2 across the entire Orion A mosaic, the classification with SExtractor shows residual systematics correlating with the variable PSF sizes. To mitigate this situation, we ran SExtractor several times on subsets of the mosaic with similar seeing (see Sect. A.3).

The morphological classification for each source was chosen among all three bands to minimize the effect of the local seeing. Despite all these efforts, systematic trends are still visible with the SExtractor classification that correlates with observing conditions. Nevertheless, for sources well above the detection threshold, this method produced reliable results.

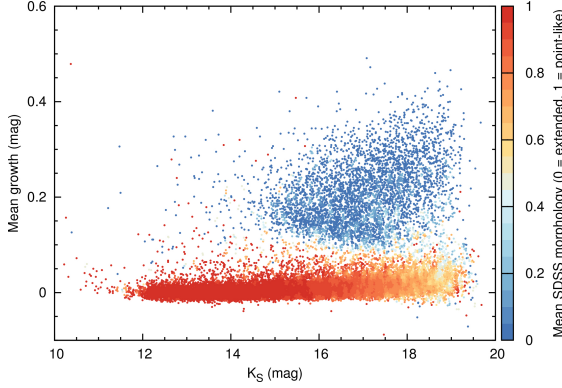
These issues led us to decide to implement an independent method for distinguishing sources with point-like or extended morphology. We found that a very robust parameter for describing the shape of a source was provided by the curve-of-growth analysis calculated earlier. Since our aperture corrections are only valid for point-like sources (only point-like sources were allowed in its calculation), any elliptically or irregularly shaped source should show a growth value different from 0 in our aperture-corrected magnitudes. Among all available apertures, the best parameter for the classification was the difference between aperture corrected magnitudes for the 1 and 2/3 arcsec apertures.

We then cross-matched (1 arcsec radius) all detected sources with the Sloan Digital Sky Survey (SDSS) catalog (DR7, Abazajian et al. 2009), which includes one of the most reliable galaxy classifications in this field down to very faint magnitudes. From the cross-matched sample, we constructed a relatively clean subset by selecting only those sources with magnitudes brighter than 23 mag in all available bands ( $u, g, r, i, z$ ), which seemed to be a good compromise between acceptable S/N and source counts. This subset contained about 47 000 objects, among which about 80% were classified as stars.

Figure 9 shows the curve-of-growth parameter as a function of  $K_S$  magnitude, with color indicating the corresponding SDSS morphology. Clearly, galaxies separate very well from stars. We then used the cross-matched SDSS subset as a training sample for a  $k$ -nearest-neighbor analysis ( $k\text{NN}$ ,  $k = 30$ ) applied to the entire survey catalog. For details on the classification method and the Python implementation we used, see Pedregosa et al. (2011). In reference to Fig. 9, this method tends to favor point-like sources for faint objects simply because the training set included about four times more stars than galaxies.

For completeness, we mention here that we also attempted a multivariate classification based on the available colors,  $(J - H)$  and  $(H - K_S)$ . Unfortunately, the limited available color parameters were not enough to separate galaxies from point sources reliably with these methods. Supplementary optical data of equivalent completeness would aid tremendously in identifying





**Fig. 9.** Magnitude growth when increasing photometric apertures from 2/3 to 1 arcsec as a function of  $K_S$  band magnitude. The color indicates the mean SDSS morphology in a box of  $\Delta K_S = 0.1$  mag,  $\Delta \text{growth} = 0.02$  mag. In this parameter space, galaxies are well separated from point-like objects, which was used for a refined morphological classification.

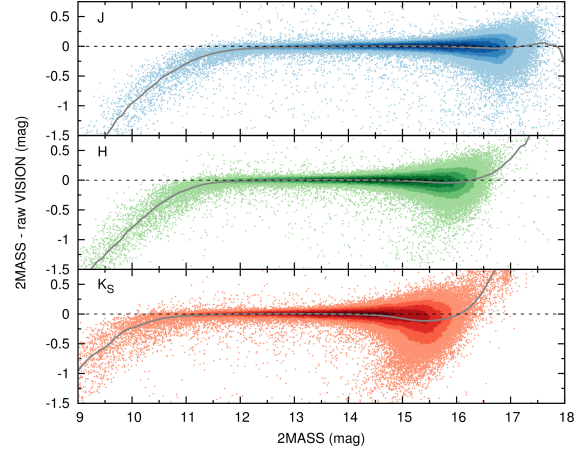
background galaxies. Unfortunately, SDSS covers only a portion of our Orion A field.

### 3.7.2. Spurious detections

SExtractor has its own, quite robust, implementation for cleaning spurious detections by assessing local detection thresholds for each individual source. These were mostly picked up in the vicinity of bright stars ( $\leq 10$  mag) and could be identified well by their morphological classification. To remove these, we used a cleaning radius proportional to the 2MASS magnitude to approximately fit the halo structures and simply removed all extended objects. In addition to this cleaning iteration, we also found a few hundred detections associated with extended emission. This subset, however, was easily identified by large curve-of-growth values, extended SExtractor morphology, and proximity to the detection limit. A visual inspection of the remaining sources only revealed very few spurious detections. We did not attempt to remove those manually, since such a task could not have been applied in a consistent manner for the entire mosaic.

### 3.7.3. Residual nonlinearity and saturation

Comparing the resulting catalog with 2MASS, as displayed in Fig. 10, one can see that stars brighter than a wavelength-dependent magnitude limit were either affected by residual nonlinearity problems or were saturated. We therefore replaced all the sources in our catalog that were located in the immediate vicinity of stars brighter than 13, 12, 11.5 mag in  $J$ ,  $H$ ,  $K_S$  in 2MASS, respectively, with the corresponding single clean reference catalog measurement from 2MASS. All clean, high S/N detections (quality flag A) were propagated from the 2MASS catalog. A few remaining very bright sources (quality flag B or worse) can therefore only be found in the 2MASS catalog. Most of them are detections of nebosity near the ONC and only very few ( $\sim 15$ ) are associated with saturated sources. In addition we found about 20 sources in 2MASS with A quality flags across all three bands that are fainter than the above-mentioned limits and were not detected in the VISTA images. These sources were not added to the final catalog.



**Fig. 10.** 2MASS vs. VISTA photometry. Clearly residual nonlinearity and saturation effects have an impact on the bright end of the measured magnitudes. For this reason we decided to replace VISION photometry with 2MASS photometry for sources brighter than (13, 12, 11.5) in  $(J, H, K_S)$ , respectively. The shading indicates source density in a  $0.2 \times 0.05$  box in this parameter space. The gray line indicates a running median with a box width of 0.5 mag, and the dotted horizontal line marks the reference value of zero mag difference between the 2MASS and VISTA catalogs.

### 3.7.4. Sources near the Orion nebula

Common source detection techniques unfortunately do not provide satisfactory results in regions where the background varies significantly on very small scales (i.e., on scales smaller than a few times the size of the core of the PSF). In this case the modeling of the background fails even for advanced methods. The SExtractor method (and also the CASU pipeline) fails for the regions around the ONC where we find background variation on sub-arcsec levels, even when using specialized filtering kernels. As a result some localized emission peaks get easily picked up as sources, producing a relatively large number of false detections. We therefore decided to make a  $2000 \times 2000$  pix ( $\sim 11 \times 11$  arcmin) cutout around the ONC for which we manually cleaned the SExtractor catalogs, while also adding missed sources. All sources in this subset were recentered by calculating a Gaussian least-squares fit at the input coordinates with IRAF (Tody 1986). From this new and cleaned coordinate list, we created an artificial image with Skymaker (v3.10.5, Bertin 2009) and used this image as input for SExtractor in double imaging mode, while extracting the sources from the original cutout. We note here, that this, of course, does not avoid the problems in the photometry associated with such highly variable background (e.g., flux over- or underestimations depending on aperture radius owing to imperfect removal of the extended emission and systematic offsets in measured source positions). Compared to automated 2MASS photometry that shows many misdetections, we are confident that our source catalog in this region is among the most reliable ones.

## 4. VISTA Orion A survey data products

In this section we describe the main data products of the VISTA Orion A survey: Sect. 4.1 contains a concise overview on the source catalog, together with the presentation of the resulting

**Table 3.** Accessible data columns for our VISION survey.

Column name	Description	Unit
VISION_ID	internal numbering	#
RAJ2000	Right ascension (J2000)	hh:mm:ss
DEJ2000	Declination (J2000)	dd:mm:ss
(J/H/K <sub>S</sub> )	corrected magnitude	mag
(J/H/K <sub>S</sub> )_err	magnitude error	mag
Class_cog <sup>a</sup>	kNN-morphology	{0, 1}
Class_sex <sup>b</sup>	SExtractor-morphology	{0, 1}
(J/H/K <sub>S</sub> )_mjd	effective MJD	d
(J/H/K <sub>S</sub> )_exptime	eff. exposure time	s
(J/H/K <sub>S</sub> )_fwhm	source FWHM	arcsec
(J/H/K <sub>S</sub> )_seeing	local seeing	arcsec
(J/H/K <sub>S</sub> )_coverage	frame coverage	# exposures
(J/H/K <sub>S</sub> )_aper	aperture radius	arcsec
(J/H/K <sub>S</sub> )_2mass_id	2MASS identifier	
(J/H/K <sub>S</sub> )_origin	original catalog	

**Notes.** Apart from astrometric and photometric measurements, we also include several quality control parameters and two morphological classification schemes. <sup>(a)</sup> Based on curve-of-growth (cog) characteristics and a subsequent kNN analysis with SDSS classifications in the training sample. <sup>(b)</sup> Based on neural networks and delivered with SExtractor; the published value is the one across all three bands for which the best seeing conditions were measured.

color–magnitude and color–color diagrams for both the Orion A and CF data. In Sect. 4.2 we present an L-RGB version of the entire mosaic and a catalog of interesting objects extracted from the provided image. This includes a selection of prominent, already known YSOs and, based on morphology criteria, an identification of five new YSO candidates. We also used the catalog to identify probable new galaxy clusters. Details on photometric, astrometric, and quality control properties can be found in Appendix A. Appendix B contains supplementary data tables.

#### 4.1. VISTA Orion A source catalog

After all calibration, merging, and cleaning steps, the final source catalog contained 161 parameters across all three bands. Since most of the calculated parameters were only used for calibration purposes and can only be fully understood with access to all details of the data reduction recipes, we decided to reduce the load by including the 35 most important columns in the published catalog. An overview of the available data is given in Table 3, and a sample of the VISION data is shown in Table B.1.

For a total on-sky coverage (including the jitter sequences and pixel rejection during co-addition) of 18.2935 deg<sup>2</sup>, we detected a total of 799 995 individual sources across all three observed bands; 505 339 of these were detected in all three filters, 653 888 in at least in two bands. For the individual *J*, *H*, and *K<sub>S</sub>* bands, we detect 571 458, 747 290, and 640 474 sources, respectively. In contrast, the 2MASS point source catalog contains 86 460 sources in the area covered by our survey, and the 2MASS Extended Catalog (Skrutskie et al. 2006) only a few hundred. For the CF we found a total of 93 909 sources with 65 665 detected sources in all three bands, and 80 526 sources were detected in at least two filters. For both the Orion A observations and the CF, we classified about 30% as extended objects with our curve-of-growth analysis. Above the residual nonlinearity and saturation limits, we added or replaced 7788, 6298, and 5355 sources from the 2MASS catalog in *J*, *H*, and *K<sub>S</sub>*,

respectively. The mean image quality (FWHM) for the Orion A data is 0.78, 0.75, and 0.8 arcsec with a standard deviation of 0.07, 0.08, and 0.1 arcsec in *J*, *H*, and *K<sub>S</sub>*. The survey catalog for all sources will be made available through the CDS.

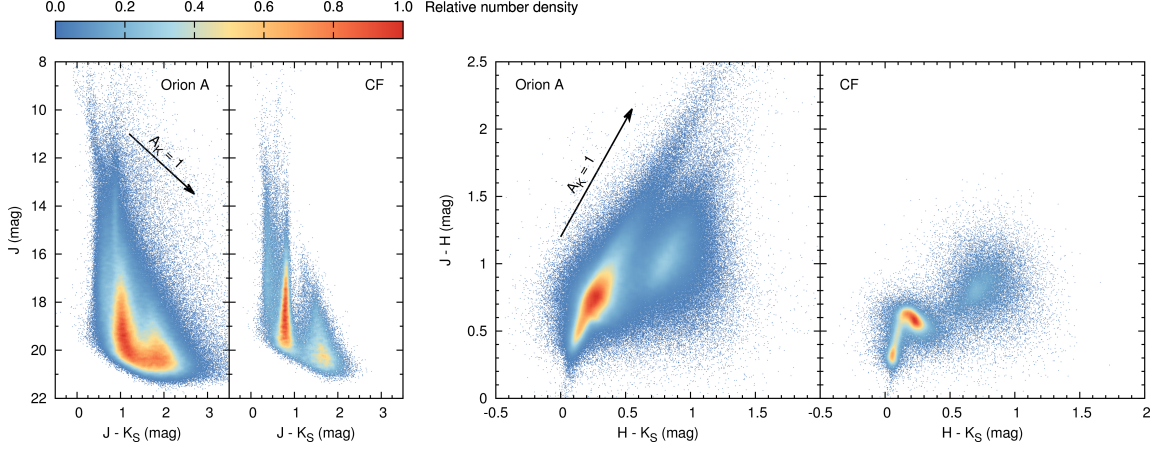
##### 4.1.1. VISTA photometry

We estimate the completeness to be 20.3 mag for *J*, 19.7 mag for *H*, and 18.7 mag for *K<sub>S</sub>* as determined by the histogram peaks in the three luminosity distributions. This, however, is highly variable throughout the mosaic because of unequal coverage, variable extinction, and extended emission in the region (for details see Appendix A.1). The median absolute deviations of the photometric errors are 0.056, 0.048, and 0.048 mag in *J*, *H*, and *K<sub>S</sub>*, respectively.

As an example of the calibration and photometric properties of the survey, we show color–color and color–magnitude diagrams for both Orion A and the CF in Fig. 11. The CF in the color–magnitude diagram exhibits the typical early-type dwarf sequence ( $J - K_S \sim 0.4$  mag) and the fainter but very visible branch of M dwarfs ( $J - K_S \sim 0.8$  mag), together with the fainter locus of galaxies at  $J - K_S \sim 1.5$  mag. The color–color diagram for the CF is very clean with a well-defined main sequence that is not obviously affected by extinction and a good separation between the main sequence and the somewhat extended locus of galaxies. The clump in the CF near  $H - K_S \sim 0.1$  mag,  $J - H \sim 0.3$  mag shows bright early-type dwarf stars and appear in overabundance here since these stars are visible at much further distances than the fainter late-type stars (Alves 1998). Since Orion A is not projected against the galactic plane or the bulge, the CF lacks the typical giant sequence, which separates from the dwarf sequence at  $H - K_S \sim 0.15$  mag,  $J - H \sim 0.7$  mag (Bessell & Brett 1988). In contrast to the CF, stars and galaxies toward the Orion A cloud can be substantially affected by dust extinction, introducing a color-excess due to reddening.

#### 4.2. VISTA Orion A mosaic L-RGB

Supplementarily to the Orion A source catalog, we also created a full resolution L-RGB image optimized for displaying the whole dynamic range of the data. To this end we created an artificial luminosity channel by co-adding all tiles including all filters at once. Since the total mosaic size exceeded the data limits for ordinary FITS conversion software, we converted each individual band from FITS to TIFF via STIFF (v2.4.0, Bertin 2012). We then mapped the *J*, *H*, *K<sub>S</sub>* bands to the blue, green, and red channels, respectively, and created a luminance channel from the combined image in Photoshop<sup>TM</sup>. Residual image defects (saturated stars, imperfect background subtraction, etc.) were modestly rectified for better artistic impressions. We then extracted the three RGB channels and subsequently built a Hierarchical Progressive Sky (HiPS, Fernique et al. 2015) with the Aladin Sky Atlas (Bonnarel et al. 2000; Boch & Fernique 2014). In this way we were able to preserve the optimized nonlinear curve stretch, together with the cosmetic corrections from Photoshop. Aladin resamples the input data onto a HEALPix grid (Górski et al. 2005), which only allows for discrete tile orders (and therefore pixel scales). Even though the original pixel scale of the reduced image was set to 1/3 arcsec, we chose a grid with a slightly coarser scale at 402.6 mas to reduce loading times and disk use. Given that the typical seeing of our survey is mostly around 0.8 arcsec, we only undersample our data in a few cases (see Fig. A.6).



**Fig. 11.** Color–magnitude and color–color diagrams for both the Orion A and the CF data. The colors indicate normalized source density within a  $0.1 \times 0.1$  mag box in the  $J$  vs.  $J - K_S$  parameter space and  $0.02 \times 0.02$  mag in  $J - H$  vs.  $H - K_S$ . The black arrows indicate the effect of an extinction of 1 mag in  $K_S$ . The presence of heavy dust extinction in Orion A pushes many sources toward redder colors when compared to the CF, which itself has a clearly defined main-sequence ( $0.2 \leq J - K_S \leq 1.2$ ,  $0 \leq H - K_S \leq 0.4$ ) and galaxy locus ( $1.2 \leq J - K_S \leq 2.2$ ,  $0.5 \leq H - K_S \leq 1.2$ ) in both diagrams.

The full L-RGB progressive sky map will be made available through the CDS at <http://alasky.u-strasbg.fr/VISTA/VISTA-Orion-A-Colored>. Until the integration of the data, the HiPS will also be available through [http://homepage.univie.ac.at/stefan.meingast/Orion\\_A\\_VISTA\\_RGB/](http://homepage.univie.ac.at/stefan.meingast/Orion_A_VISTA_RGB/). It will not be possible to perform photometry on these data.

Significant overdensities ( $>5\sigma$ ) were inspected visually. With this method we selected ten outstanding overdensities and classify them as potential new galaxy clusters. Postage stamps of these regions are displayed in the bottom matrix of Fig. 13. Cross identifications, coordinates and magnitudes for all objects shown in Figs. 12 and 13 are listed in Table B.2.

#### 4.2.1. Catalog of interesting objects

With the superior resolution and sensitivity of the Orion A VISTA survey, it becomes possible to investigate the morphology of some already known YSOs and even extend this view to identify new candidates. We have compiled a representative list of young stars associated with prominent features of scattered light and outflows in different evolutionary stages as given in the literature (e.g., Lada 1987; Andre et al. 2000; Evans et al. 2009). Figure 12 shows examples of YSOs at different ages extracted from the provided Orion A L-RGB. Class I and Class II identifications were taken from Megeath et al. (2012), and Class III identifications from Pillitteri et al. (2013). While many Class I and II sources show spectacular nebulous structure in their vicinity, Class III sources are virtually indistinguishable from stars on the main sequence in NIR colors. Based on a visual inspection of the morphology of these known YSOs, we were able to identify five new YSO candidates not mentioned in previous studies (e.g., due to the restricted coverage of the *Spitzer* survey). These candidate YSOs are shown in the top row of Fig. 13, and all together they show the typical morphological characteristics of our test sample.

In addition to the investigation of YSO morphology, we also used the VISTA Orion A catalog to identify overdensities in background galaxies, which are potential previously undiscovered galaxy clusters. Here we evaluated the spatial density distribution of extended sources as given in our catalog with a 2 arcmin wide Epanechnikov kernel on a  $1 \times 1$  arcmin grid. The sample was restricted to  $\text{class\_sex} \leq 0.3$ ,  $\text{class\_cog} = 0$ ,  $J/H/K_S > 15$  mag, and  $K_S < 18$  mag. The last requirement filters most misclassifications for faint unresolved sources.

### 5. Young stellar populations toward Orion A

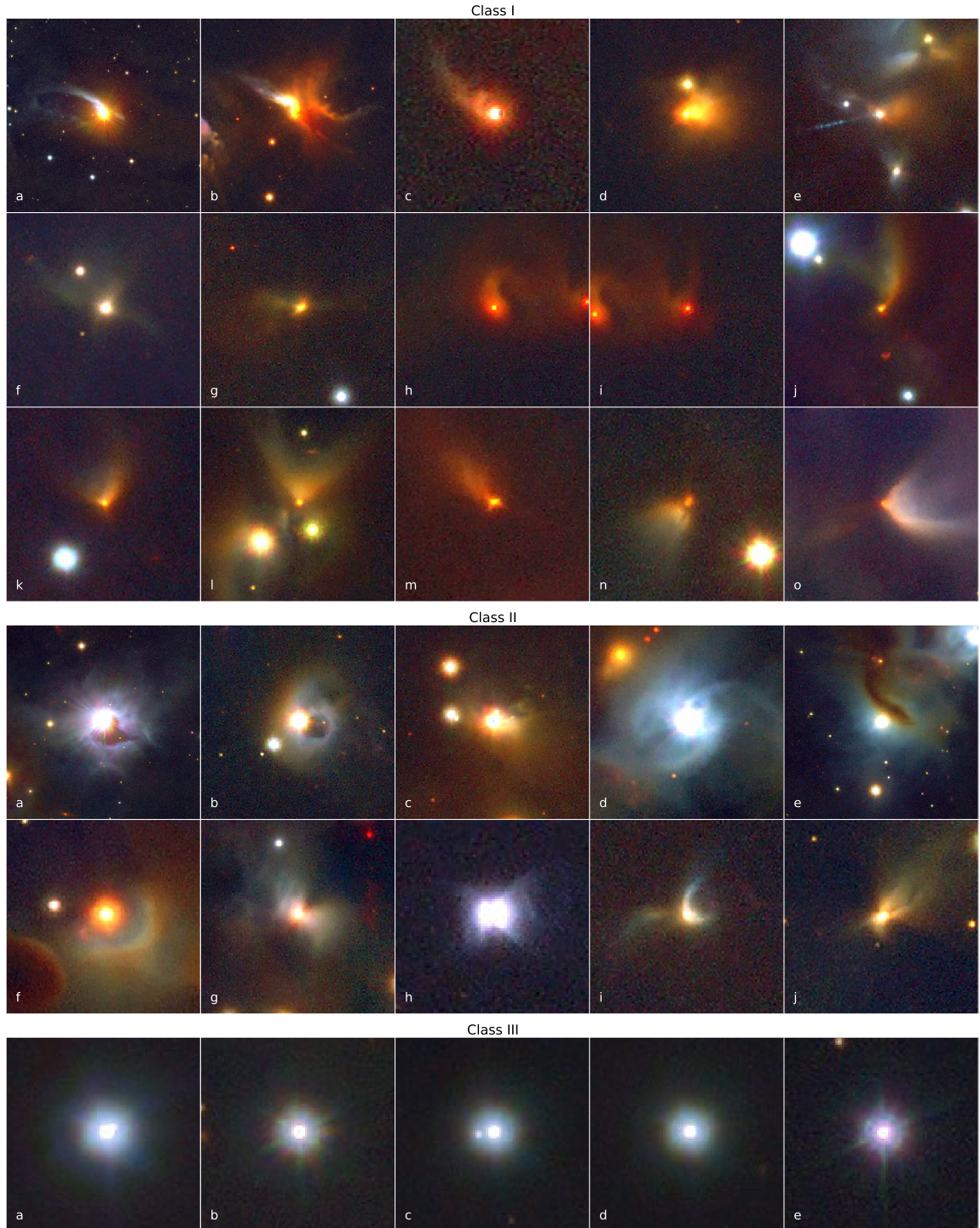
To discuss the stellar populations seen toward Orion A, we begin by deriving an estimate of the young stellar population associated with the molecular cloud via the  $K_S$  band luminosity function (KLF). Here we use the term KLF for the general  $K_S$  band magnitude distribution of sources in our survey and do not explicitly refer to the monochromatic luminosity distribution of a given sample (such as the YSOs associated with the molecular cloud). In the second part, we discuss the foreground populations.

#### 5.1. Orion A population

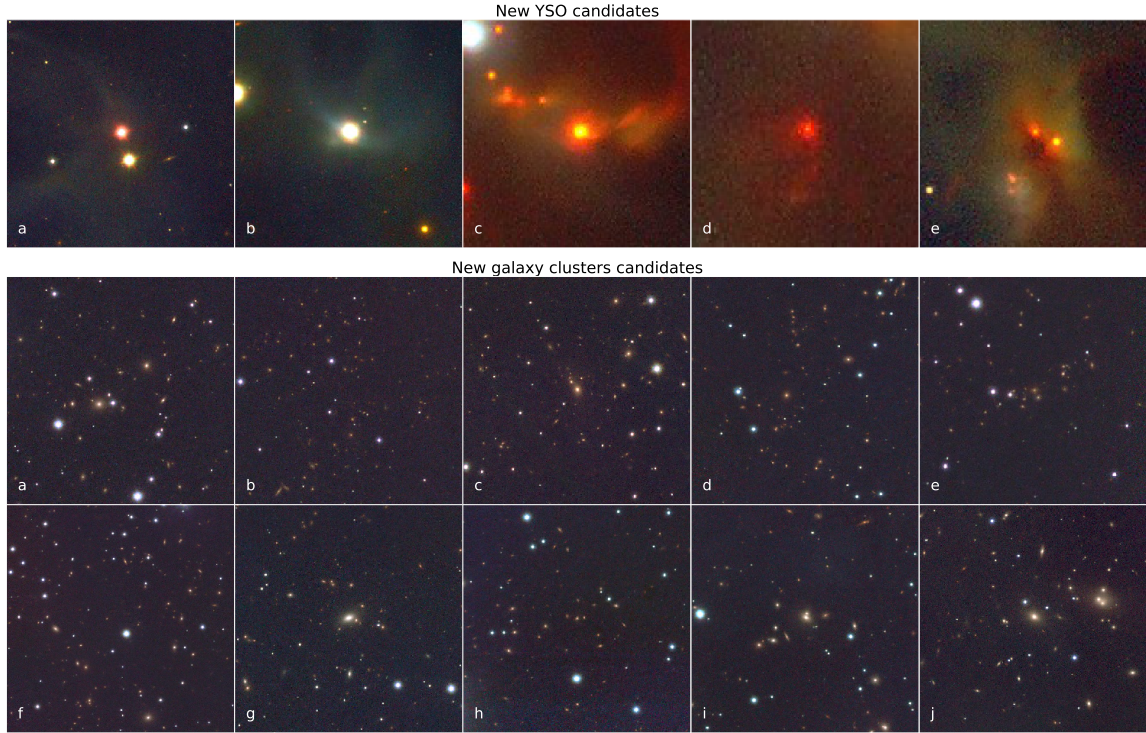
Orion A shows a significant gradient in star formation activity along its spine coinciding with the number of published articles (cf. Fig. 1) with about twice as many YSOs near the integral-shaped filament compared to the rest of the cloud (Megeath et al. 2012). To the west we find – among others – the ONC, the BN/KL region, and the OMC 2-3 complex, which are well-studied regions. Toward the east of the ONC, however, the cloud has drawn much less attention to itself. Here we attempt to statistically derive a complete census of the entire Orion A population by means of the KLF. To rule out critical systematic errors, we compared our survey data not only to the CF, but also to source counts given in the Besançon model of the Galaxy (Robin et al. 2003). Our analysis was split into the following consecutive steps that will be discussed in more detail individually:

1. comparing the source counts given in the Besançon model with our CF;
2. scaling the CF to fit the coverage of the Orion A survey;





**Fig. 12.** Selected YSOs in our VISTA Orion A survey. Class I and Class II sources were taken from [Megeath et al. \(2012\)](#), Class III objects from [Pillitteri et al. \(2013\)](#). Class III sources are virtually indistinguishable from more evolved stars in NIR colors, while Class I and Class II sources very often show characteristic structures of scattered light. The labels in the bottom left corners refer to Table B.2.



**Fig. 13.** Newly identified objects in the VISTA Orion A survey. The *top row* shows potential new YSOs based on similar morphology compared to sources listed in the literature (see Fig. 12). The *bottom matrix* shows new candidate galaxy clusters near Orion A, identified as over-densities in the galaxy distribution in the presented catalog. The labels in the bottom left corners refer to Table B.2.

3. dereddening the Orion A data;
4. calculating the excess population toward Orion A with respect to the CF, as well as the Besançon model. This excess includes both the young populations associated with Orion A, as well as the young foreground population unrelated to the galactic field;
5. estimating the total foreground population toward Orion A (galactic field + young foreground);
6. estimating the young stellar population in the foreground to Orion A unrelated to the galactic field;
7. estimating the total young stellar population of the Orion A molecular cloud.

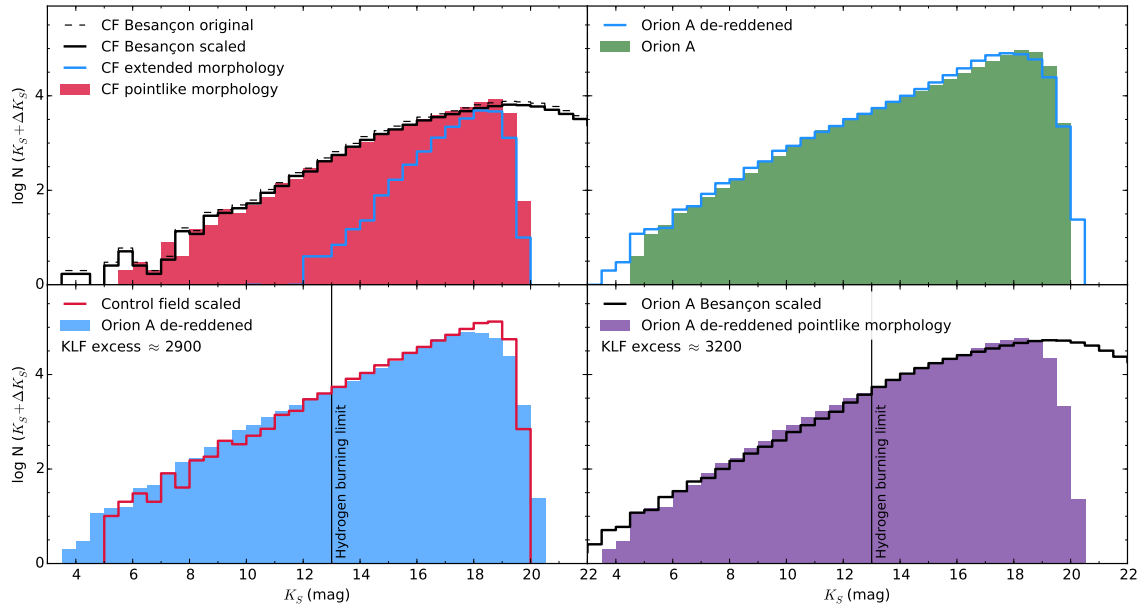
*Ad 1.* The Besançon model of the Milky Way plays a crucial role in estimating the young populations Orion A and the foreground because we can approximate the number of expected galactic field stars up to the adopted distance of Orion A of 414 pc (Menten et al. 2007). For this reason we wanted to verify that the source counts measured in our CF and the ones taken from the model match. The top lefthand plot in Fig. 14 shows the results of our comparison. Looking at the Besançon data we see an offset in source counts relative to the measurements in the CF. The CF histogram shows only about 85% of the source counts in the model when considering the bins with  $7 \leq K_S \leq 15$  mag for which we have a reasonable number of source counts and, at the same time, do not expect any problems from a morphological misclassification. For this reason we decided to scale all model data with the scaling factor derived from the difference

between the observed CF KLF and the Besançon model in our subsequent analysis. We split the data from the CF into extended and point-like morphology since the Besançon model does not include background galaxies.

Comparing the histograms we note a discrepancy in source counts between the CF KLF and the model KLF at the faint end ( $K_S = 19$  mag) where mostly galaxies are situated. This is readily explained by our morphological classification, which tends to classify sources as point-like for very noisy measurements. The (scaled) background model also overpredicts sources at the bright end of the spectrum ( $K_S \lesssim 10$  mag) relative to our measurements. This can be explained by the absence of very bright sources in our catalog since only sources with clean measurements (quality flag A) were adopted from the 2MASS catalog. For the further analysis we note that these potentially missing bright sources do not critically influence our population estimates because there are so few of them. Besides this seemingly constant offset and the aforementioned excesses, we find good agreement with respect to the shape of the two histograms.

*Ad 2.* To make statistical comparisons between the CF and the Orion A surveys, we needed to scale the source counts of the CF to match the survey coverage. We decided to scale based on field coverage where we find that the Orion A survey covers an area 10.149 times larger than the CF. Other scaling methods (e.g., total source counts) were ruled out because of dissimilar stellar populations and the presence of Orion A, which blocks many background sources.





**Fig. 14.**  $K_S$  band luminosity functions (histograms; 0.5 mag bin width) for the various analysis steps to estimate both young foreground and the total Orion A population. *Top left:* comparison of the CF KLF with the Besançon model of the Galaxy. We find that a slight adjustment to all source counts in the model was necessary to fit the observations (see text for details). *Top right:* KLFs for the final survey catalog and the de-reddened data. Sources are mostly pushed to brighter magnitudes, however, we also allow (for statistical reasons) negative values of  $A_K$  and therefore find a minor population at fainter magnitudes than in the original histogram (the last bin starting at  $K_S = 20$  in the de-reddened histogram). *Bottom left:* comparison of the de-reddened data with the scaled CF KLF. We clearly see an excess of bright sources, as expected from the young populations we find towards Orion A. Note here the good agreement of the CF and Orion A KLF for magnitudes fainter than  $K_S = 13$  mag up to the completeness limit. *Bottom right:* comparison of the de-reddened Orion A KLF for point sources to the Besançon model. Also here we find a clear excess of sources caused by the young populations seen towards the molecular cloud. Both comparisons deliver similar numbers which is interpreted as a validation for our statistical approach.

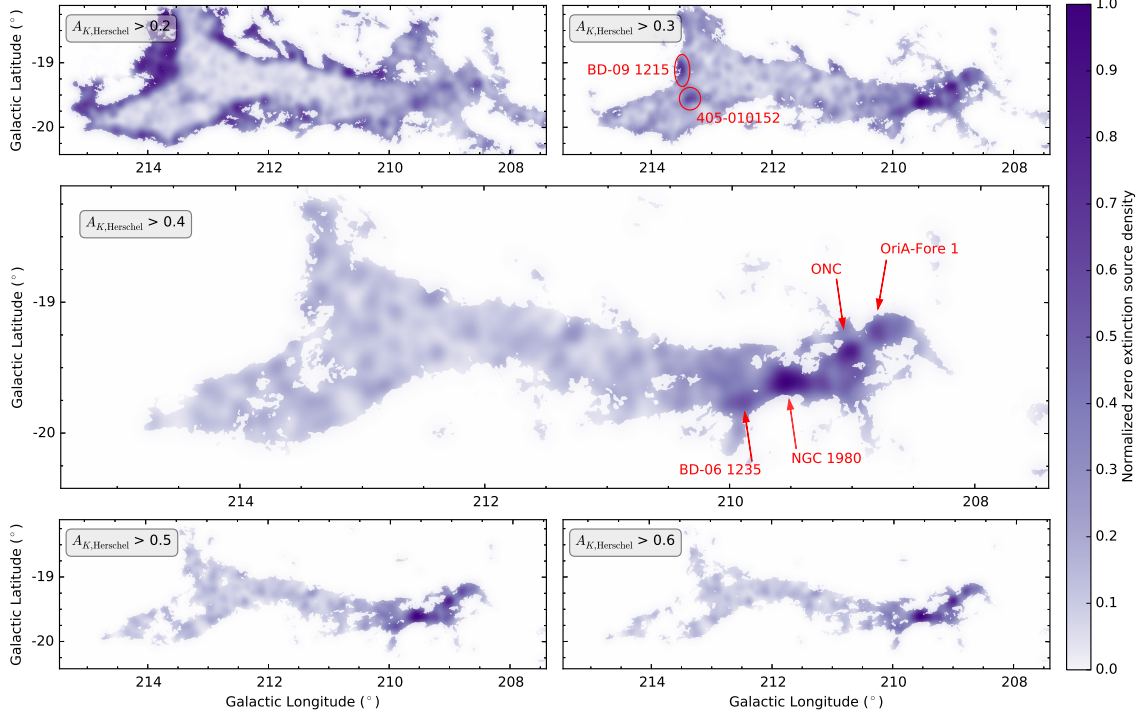
*Ad 3.* The Orion A molecular cloud covers large parts of our survey and therefore many sources in the background will exhibit non-negligible NIR excess. Also, stars embedded within the cloud will naturally show redder colors (compare Fig. 11). Schlafly et al. (2015) do not find significant amounts of dust up to 300 pc in this region, therefore all stars showing NIR color-excess should lie in or behind the cloud. For statistical comparisons we therefore needed to estimate the NIR excess for each source to get intrinsic  $K_S$ -band magnitudes. To this end we used the NICER method (Lombardi & Alves 2001) which de-projects the measured colors using both measurement errors and the color distribution of sources in the CF to derive line-of-sight extinctions. This method, together with its extension, NICEST, described in Lombardi (2009), has found many successful applications showing its robustness (see, e.g., Lombardi et al. 2006; Alves et al. 2014 for examples; and Goodman et al. 2009 for an independent comparison of different column density tracers). We note here that NICEST has no effect on extinction measurements for point sources since the method only attempts to correct for cloud substructure in the subsequent construction of extinction maps. The reliable determination of color-excesses with NICER depends not only on the photometric quality, but also on the intrinsic color distribution of the sources as measured from the CF. Typically for stars in the NIR, this distribution is very narrow (compare Fig. 11), however, when also including galaxies the estimate of the color-excesses can be significantly

biased<sup>11</sup> (Foster et al. 2008). This, however, is no problem here since galaxies were not included in the determination of excess source counts in Orion A, assured by a subsequent magnitude cut at  $K_S = 13$  mag (see step 4). Also, the method itself allows negative values for color-excesses. These are sources lying below the adopted color zero point and comprise hot stars with low foreground extinction. For this application of NICER, we used the extinction coefficients from Indebetouw et al. (2005). A comparison of the raw Orion A data and its dereddened KLF is shown in the top righthand plot of Fig. 14.

*Ad 4.* We estimated the excess of young stars over the Galactic field by calculating the differences between (a) the dereddened Orion A KLF and the CF KLF and (b) the Orion A KLF only for point-like sources and the Besançon KLF. For a distance modulus of  $\mu = 8.1$  mag and an age of 1 Myr, we find the assumed hydrogen burning limit of  $0.08 M_\odot$  at  $K_S \approx 13$  mag (Baraffe et al. 1998, 2002). Both methods clearly show an excess that is also readily visible in the histograms in the two lower panels of Fig. 14. The Orion A KLF shows an excess of  $\sim 2900$  sources compared to the CF and  $\sim 3200$  sources compared to the scaled Besançon KLF. To test for the statistical significance of these results, we determined the error of the excess assuming Poisson statistics for the histograms. In both cases the errors amount to  $\sim 170$  sources or approximately 5–6%. Including

<sup>11</sup> NICER includes the intrinsic color distribution in the calculation of the errors.





**Fig. 15.** Density maps of zero-extinction sources restricted to an area above a given column density threshold taken from the *Planck-Herschel* map. Kernel densities were evaluated on a grid with 1 arcmin resolution and a Gaussian kernel with 3 arcmin FWHM. Here we used the cloud to effectively block background sources to assess the population size of true foreground stars. For low column density, clearly the background contaminates the sample. Our final threshold of  $A_{K,Herschel}$  is a trade-off between reliable filtering and area coverage. Also note here the preferred clustering of sources towards NGC 1980 and the ONC.

the absolute difference between the two methods, we assume an error of 10% on these estimates for the following analysis. The comparison with the Besançon KLF also produces an excess of bright sources when compared to the scaled CF data. While the difference on the bright end between the VISION KLF and the CF KLF can be explained by low number statistics (smaller on-sky coverage of the CF), the VISION KLF overall is not complete for these bright sources since only clean unsaturated measurements from 2MASS were adopted (which leads to a lack of bright sources when comparing to the Besançon model). This difference, however, only has a negligible effect on our number estimates. For the subsequent population estimates we adopt a KLF excess of 3000 sources up to  $K_S = 13$  mag.

*Ad 5.* The Orion A KLF excess of 3000 sources includes not only the population of Orion A, but also parts of the young populations mainly associated with NGC 1980, which is claimed to be a foreground population not emerging from the cloud (Alves & Bouy 2012; Pillitteri et al. 2013; Bouy et al. 2014). To estimate the number of foreground stars, we used the molecular cloud to effectively block background sources. In principle, sources seen in projection toward Orion A that at the same time show negligible extinction as determined via NICER (i.e.,  $A_{K,NICER} \leq 0$  mag), can only lie in the foreground of the cloud. In Fig. 15 we show zero-extinction source densities<sup>12</sup> for different column density

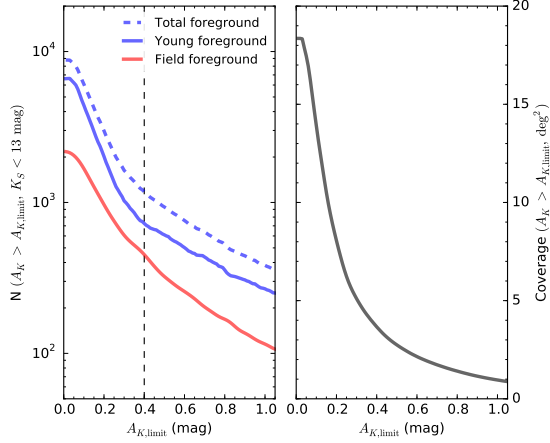
<sup>12</sup> The source densities were estimated with a symmetric 3 arcmin-wide Gaussian kernel.

thresholds as given in the *Herschel-Planck* map from Lombardi et al. (2014). Here we note some key characteristics:

1. For low column densities, background sources contaminate the sample, whereas for large thresholds the background disappears, however at the cost of reduced coverage.
2. At about  $A_{K,Herschel} \gtrsim 0.3-0.4$  mag, the background contamination starts to become negligible because we cannot see steep gradients toward the edges of the covered area.
3. We find significant substructure in the distribution of foreground sources, which are discussed in Sect. 5.2.

Surely, not all the entire foreground population is captured with this method. However, the bulk of the young foreground stars are thought to be associated with NGC 1980, which mostly falls on top of heavily extinguished regions. For  $A_{K,Herschel} > 0.4$  mag, we find a total zero extinction, i.e., a foreground population (galactic field + young stars) of about 1200 sources with  $K_S \leq 13$  mag.

*Ad 6.* To split the estimate from step (5) into young foreground sources and galactic field we compared our findings to the Besançon model of the Galaxy. For the area given by  $A_{K,Herschel} > 0.4$  mag we counted all stars up to the adopted distance of 414 pc. Here we find about 500 stars brighter than  $K_S = 13$  mag. Given a total foreground population of 1200 sources (within our constraints), we therefore estimate the young foreground to comprise about 700 stars below the magnitude limit at  $K_S = 13$  mag within our limited field.



**Fig. 16.** Left: number of statistically derived foreground stars as a function of *Herschel* column density expressed in  $A_K$ . The total foreground was identified as stars seen in projection against areas above a given threshold while showing negligible extinction. The field foreground was extracted from the Besançon model of the galaxy. The difference between those denotes the young foreground stars unrelated to the galactic field. The vertical dashed line indicates the optimal column density threshold for which background contamination is still negligible. Right: the given area on sky as a function of column density.

*Ad 7.* For a column density threshold of  $A_{K,\text{Herschel}} > 0.4$  mag and taking the 3000 KLF excess sources in the KLF together with the 700 sources associated with the young foreground, we estimated ~2300 sources comprising the Orion A stellar population. We consider this number a lower limit for two reasons:

1. Our survey is not complete with respect to protostars. Using *Spitzer* observations covering large parts of Orion A (and also Orion B), Megeath et al. (2012) find a total of 2818 protostars and disk sources as defined in their paper in the region of Orion A. Overall, we find 2751 out of these 2818 (98%) sources to have a detection in at least one band when using a 2 arcsec cross-matching radius; here we are 99.6% complete with respect to disks (10 out of 2446 missing) and 84% complete regarding protostars (53 out of 329 missing). We checked the ten remaining undetected disks in the *Spitzer* data, and about half of them turned out to be misdetections of nebulosity. As a result, we are essentially complete with respect to disks. Requiring a simultaneous detection in  $J$ ,  $H$ , and  $K_S$  (as for our dereddening step), the Megeath et al. (2012) YSO number count detected in our survey decreases by about 10% to 2500 sources.
2. The estimate of 700 young foreground sources with  $K_S < 13$  mag can be biased by a population emerging from Orion A. Such emerging stars would show negligible extinction and thus be included in our foreground sample. An indicator of the magnitude of this potential contamination, together with statistical errors, is derived in Sect. 5.2 and amounts to ~10%. We derive an additional indicator for contamination in our foreground sample by comparing the zero-extinction sample to the YSO catalog of Megeath et al. (2012). We find that about 9% of our total foreground sample are associated with *Spitzer*-identified YSOs within our selection criteria, which matches the above-mentioned error estimate well.

**Table 4.** Identified groups of young foreground populations in this work.

Group name	RA (J2000) (hh:mm)	Dec (J2000) (dd:mm)	Reference
NGC 1980	05:35.4	-05:54.9	1, 2, 3, 4, 5
OriA-Fore 1	05:35.3	-05:10.3	1, 5
BD-06 1235	05:35.07	-06:17.12	1, 5
405-010152 <sup>a</sup>	05:41.93	-09:11.84	1
BD-09 1215	05:43.9	-09:01.99	1

**Notes.** The list below the separator indicates potential groups and require further analysis. The names of these potential groups refer to the brightest star (optical) in the region. <sup>(a)</sup> UCAC4 identifier.

**References.** (1) This work; (2) Bally (2008); (3) Alves & Bouy (2012); (4) Pillitteri et al. (2013); (5) Bouy et al. (2014).

In light of these arguments and including the 10% error indicator of the KLF excess determination, we find an upper limit for the young population toward Orion A of about 3000 YSOs (excluding unresolved binaries).

The most critical problem with this deduction is the dependency of our population estimates on the column density threshold. In Fig. 16 we plot the dependency of the foreground estimates as a function of the chosen threshold in the *Herschel-Planck* column density map. Here, the total foreground refers to step (5), the young foreground and galactic field foreground to step (6). At about  $A_K \sim 0.35$  mag, we see a sudden increase in the slope in the foreground estimates. At about this point, the background contamination becomes non-negligible, and estimates significantly below this level would be biased. This compares well to the appearance of the source density maps in Fig. 15. Scaling our results to the size of the cloud is difficult owing to the anisotropic distribution of young foreground stars (Bouy et al. 2014). Linearly extrapolating to the cloud size would therefore not yield realistic results. We therefore settled for the number estimates obtained with a column density threshold of  $A_{K,\text{Herschel}} = 0.4$  mag. This threshold provides significant shielding from background contamination and at the same time encompasses most of NGC 1980 and therefore the bulk of the young foreground sources.

## 5.2. Foreground population

Figure 15 shows substantial substructure in the distribution of foreground sources. The peak associated with NGC 1980 confirms previous results from Alves & Bouy (2012) and Bouy et al. (2014). The latter study also finds an overdensity near OMC 2/3, called OriA-Fore 1, which is visible in our maps as well. We find one additional overdensity to the southeast of NGC 1980 near HH 322 and HH 323. Here we can also spot a few very bright and blue sources in the VISTA data, where BD-06 1235 is the brightest star in this region. Only a few arcminutes south, Bouy et al. (2014) speculate on a possible foreground population in the vicinity of an overdensity in X-ray sources they called L1641W. The other prominent peak associated with the ONC and roughly centered on the Trapezium cluster does not have a straightforward interpretation. It appears in Alves & Bouy (2012), which uses NIR photometry, but it is barely present in Bouy et al. (2014), which uses mostly optical data. Certainly the error associated with the density estimations in this region is by far the largest across the survey simply because of the extreme surface density of stars toward the Trapezium cluster. If the enhancement is not due to statistical errors alone, the next interpretation would be that the enhancement represents the Trapezium young population emerging from the dusty

molecular cloud. We can estimate the global bias due to statistical errors and potentially emerging stars by comparing the source counts in our zero-extinction sample to the total number of sources in the region near the ONC. Here about 10% of the total VISTA source catalog comprise the zero-extinction subsample. As a third alternative explanation for the ONC enhancement one can consider the possibility of a real, yet undiscovered, foreground population. But in light of the previous scenarios and since no other study found any evidence of such a population, this third explanation seems unlikely. In addition to these significant overdensities in the western parts of Orion A we can identify two new groups to the east when lowering the threshold to  $A_{K_{\text{Herschel}}} > 0.3$ . This map might already show minor background contamination, so we refer to the identified overdensities as potential groups that require further analysis. Both regions are marked in the top right plot of Fig. 15). The labels refer to the brightest stars in the vicinity of the peaks, which are BD-09 1215 and UCAC4 405-010152 (Zacharias et al. 2013). Since SDSS does not cover the eastern parts of Orion A, newly acquired deep optical data supplemented by our NIR survey would greatly aid in the confirmation of these potential new foreground populations.

By comparing the KLF excess with expected Galactic field stars in Sect. 5.1, we found that about 700 sources are related to the young foreground population. When limiting the  $\sim 2100$  foreground sources identified in Bouy et al. (2014) to our constraints (field and magnitude limit), their sample decreases to  $\sim 650$ , which agrees very well with our results.

Bouy et al. (2014) propose a star formation scenario toward Orion A where the foreground populations in this region formed 5–10 Myr ago. Subsequent supernovae then could have triggered episodic star formation leading to the formation of the embedded clusters we see today. With our new results, we can extend this view. Both number estimates of the young foreground from previous studies ( $\sim 2100$ , Bouy et al. 2014) and the embedded population ( $\sim 2300$ ) are incomplete. The number of foreground stars suffers from incomplete data coverage, while our estimate of the embedded population misses some of the deeply embedded sources. The numbers, however, are similar, suggesting that a comparable star formation event took place a few million years ago. The formation of NGC 1980 with its massive stars might even have been similar to what we see in the Orion Nebula today. Furthermore, our data offers the first view of the foreground population toward all parts of Orion A. That we only see a substantial amount of young foreground stars in the direction of the ONC and the integral-shaped filament, where the bulk of all star formation in Orion A occurs today, suggests a causal connection. The star formation events that produced NGC 1980 and the other foreground groups are then not responsible for the formation of the molecular cloud as a whole, but indeed seem to have had an enhancing effect with respect to the formation of new stars on the western parts of the cloud, which is the only part where massive stars are forming in Orion A.

## 6. Summary

The VISTA Orion A survey provides the most detailed view of this massive star-forming region in the NIR yet. In this paper we presented survey strategy, data calibration, catalog generation, the main data products, and first results with this rich data set. Here, we summarize our main results.

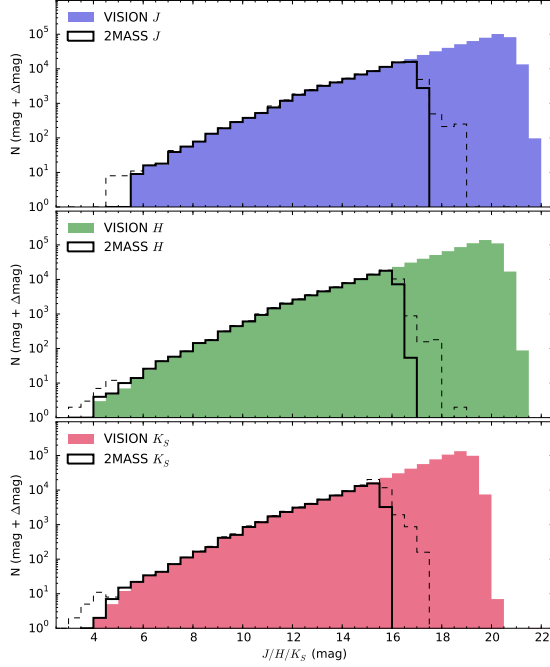
1. Our survey of the Orion A molecular cloud in the NIR bands  $J$ ,  $H$ , and  $K_S$  covered in total  $\sim 18.3 \text{ deg}^2$  on a pixel scale of  $1/3 \text{ arcsec/pix}$ .
2. We implemented independent data reduction procedures that avoid some disadvantages in the standard CASU VIRCAM pipeline. Most notably we improved the resolution on average by about 20% over the pipeline processed data.
3. The generated source catalog contains 799 995 sources, a gain of almost an order of magnitude compared to 2MASS, translating into a gain of three to four magnitudes. The 90% completeness levels (i.e., 90% of the sources are detected with our source extraction) are 20.4, 19.9, and 19.0 mag in  $J$ ,  $H$ , and  $K_S$ , respectively. We also improved depth and coverage of all previously available ONC catalogs in the NIR. In contrast to the pipeline, our photometry is calibrated toward the 2MASS photometric system. The source catalog will be made available through the CDS.
4. In addition to the source catalog, we also provided optimized three-color image data in HEALPix format, also available through the CDS in the future.
5. Cross-matching with the YSO catalog from Megeath et al. (2012) reveals that we are essentially complete (99.6%) with respect to disks and 84% complete regarding protostars as classified in their paper.
6. From these data we identified several notable YSOs associated with characteristic nebulosity. Based on the morphology of this test set, we identified five new YSO candidates.
7. Based on the surface density of extended sources in the catalog we identified ten new galaxy cluster candidates.
8. We estimated the entire young stellar population in Orion A by means of the  $K_S$  band luminosity function. We find lower and upper limits of 2300 and 3000 sources, respectively, which compares well to results from earlier studies.
9. Separated from the young population in Orion A, we can confirm previous results regarding the young foreground population toward Orion A. Here we find the same complex pattern of foreground groups mostly toward the integral-shaped filament, including the Orion nebula. Toward the eastern parts of the cloud we could identify two new potential small foreground groups.
10. Given the asymmetric east-west projected distribution of foreground sources it is unlikely that this population played an important role in assembling the Orion A cloud. Nevertheless, given the good correlation with the enhanced star formation activity in the integral-shaped filament, it is likely that the foreground population is responsible instead for compressing the western part of the cloud via feedback processes (winds, supernovas).

**Acknowledgements.** Stefan Meingast is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute for Astrophysics, University of Vienna. This work has been supported by the Austrian Science Fund (FWF) under project number P 26718-N27. We gratefully acknowledge the referee, Tom Megeath, for carefully reading the manuscript and the useful comments that served to improve both the clarity and quality of this study. This research made use of Montage, funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology. Montage is maintained by the NASA/IPAC Infrared Science Archive; The SIMBAD database, operated at the CDS, Strasbourg, France. "Aladin sky atlas" developed at the CDS, Strasbourg Observatory, France. The VizieR catalogue access tool, CDS, Strasbourg, France. Astropy is a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013). H. Bouy is supported by the Ramón y Cajal fellowship program number RYC-2009-04497 and by the Spanish Grant AYA2012-38897-C02-01.



## References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Alcalá, J. M., Covino, E., & Leccia, S. 2008, Orion Outlying Clouds, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 801
- Ali, B., & Depoy, D. L. 1995, *AJ*, **109**, 709
- Allen, L. E., & Davis, C. J. 2008, Low Mass Star Formation in the Lynds 1641 Molecular Cloud, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 621
- Alves, J., & Bouy, H. 2012, *A&A*, **547**, A97
- Alves, J., Lombardi, M., & Lada, C. J. 2014, *A&A*, **565**, A18
- Alves, J. F. 1998, Ph.D. Thesis, European Southern Observatory
- Andre, P., Ward-Thompson, D., & Barsony, M. 2000, *Protostars and Planets IV*, 59
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bally, J. 2008, Overview of the Orion Complex, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 459
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, *A&A*, **337**, 403
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 2002, *A&A*, **382**, 563
- Bertin, E. 2006, in *Astronomical Data Analysis Software and Systems XV*, eds. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, *ASP Conf. Ser.*, **351**, 112
- Bertin, E. 2009, *Mem. Soc. Astron. It.*, **80**, 422
- Bertin, E. 2010, Astrophysics Source Code Library [[record ascl:1010.068](#)]
- Bertin, E. 2011, in *Astronomical Data Analysis Software and Systems XX*, eds. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, *ASP Conf. Ser.*, **442**, 435
- Bertin, E. 2012, in *Astronomical Data Analysis Software and Systems XXI*, eds. P. Ballester, D. Egret, & N. P. F. Lorente, *ASP Conf. Ser.*, **461**, 263
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in *Astronomical Data Analysis Software and Systems XI*, eds. D. A. Bohlender, D. Durand, & T. H. Handley, *ASP Conf. Ser.*, **281**, 228
- Bessell, M. S., & Brett, J. M. 1988, *PASP*, **100**, 1134
- Blaauw, A. 1964, *ARA&A*, **2**, 213
- Boch, T., & Fernique, P. 2014, in *Astronomical Data Analysis Software and Systems XXIII*, eds. N. Manset & P. Forshay, *ASP Conf. Ser.*, **485**, 277
- Bonifacio, P., Monai, S., & Beers, T. C. 2000, *AJ*, **120**, 2065
- Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, **143**, 33
- Bouy, H., Alves, J., Bertin, E., Sarro, L. M., & Barrado, D. 2014, *A&A*, **564**, A29
- Briceno, C. 2008, The Dispersed Young Population in Orion, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 838
- Calabretta, M. R., & Greisen, E. W. 2002, *A&A*, **395**, 1077
- Carpenter, J. M. 2000, *AJ*, **120**, 3139
- Carpenter, J. M., Snell, R. L., & Schloerb, F. P. 1995, *ApJ*, **450**, 201
- Carpenter, J. M., Heyer, M. H., & Snell, R. L. 2000, *ApJS*, **130**, 381
- Chen, H., & Tokunaga, A. T. 1994, *ApJS*, **90**, 149
- Da Rio, N., Robberto, M., Hillenbrand, L. A., Henning, T., & Stassun, K. G. 2012, *ApJ*, **748**, 14
- Da Rio, N., Tan, J. C., Covey, K. R., et al. 2015, *ApJ*, **818**, 59
- Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, in *SPIE Conf. Ser.*, **6269**
- Davis, C. J., Froebrich, D., Stanke, T., et al. 2009, *A&A*, **496**, 153
- Duchon, C. E. 1979, *J. Applied Meteorology*, **18**, 1016
- Emerson, J., McPherson, A., & Sutherland, W. 2006, *The Messenger*, **126**, 41
- Evans, II, N. J., Dunham, M. M., Jørgensen, J. K., et al. 2009, *ApJS*, **181**, 321
- Fernique, P., Allen, M. G., Boch, T., et al. 2015, *A&A*, **578**, A114
- Fischer, W. J., Megeath, S. T., Ali, B., et al. 2010, *A&A*, **518**, L122
- Foster, J. B., Román-Zúñiga, C. G., Goodman, A. A., Lada, E. A., & Alves, J. 2008, *ApJ*, **674**, 831
- Gälfalk, M., & Olofsson, G. 2008, *A&A*, **489**, 1409
- Genzel, R., Reid, M. J., Moran, J. M., & Downes, D. 1981, *ApJ*, **244**, 884
- Gomez, M., & Lada, C. J. 1998, *AJ*, **115**, 1524
- Gonzalez, O. A., Rejkuba, M., Zoccali, M., Valenti, E., & Minniti, D. 2011, *A&A*, **534**, A3
- Goodman, A. A., Pineda, J. E., & Schnee, S. L. 2009, *ApJ*, **692**, 91
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759
- Gruen, D., Seitz, S., & Bernstein, G. M. 2014, *PASP*, **126**, 158
- Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, *ApJS*, **184**, 18
- Gutermuth, R. A., Pipher, J. L., Megeath, S. T., et al. 2011, *ApJ*, **739**, 84
- Herbig, G. H., & Jones, B. F. 1983, *AJ*, **88**, 1040
- Hillenbrand, L. A., & Hartmann, L. W. 1998, *ApJ*, **492**, 540
- Hirota, T., Bushimata, T., Choi, Y. K., et al. 2007, *PASJ*, **59**, 897
- Hodapp, K.-W. 1994, *ApJS*, **94**, 615
- Indebetouw, R., Mathis, J. S., Bahler, B. L., et al. 2005, *ApJ*, **619**, 931
- Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, in *Optimizing Scientific Return for Astronomy through Information Technologies*, eds. P. J. Quinn & A. Bridger, *SPIE Conf. Ser.*, **5493**, 411
- Janesick, J. R. 2001, *Scientific charge-coupled devices*, Press Monographs (Society of Photo Optical)
- Jeffries, R. D. 2007, *MNRAS*, **376**, 1109
- Lada, C. J. 1985, *ARA&A*, **23**, 267
- Lada, C. J. 1987, in *IAU Symp.* 115, *Star Forming Regions*, eds. M. Peimbert, & J. Jugaku, 1
- Lada, C. J., Muench, A. A., Haisch, Jr., K. E., et al. 2000, *AJ*, **120**, 3162
- Lada, C. J., Muench, A. A., Rathborne, J., Alves, J. F., & Lombardi, M. 2008, *ApJ*, **672**, 410
- Lada, E. A. 1992, *ApJ*, **393**, L25
- Lada, E. A., Depoy, D. L., Evans, II, N. J., & Gatley, I. 1991, *ApJ*, **371**, 171
- Lada, E. A., Evans, II, N. J., & Falgarone, E. 1997, *ApJ*, **488**, 286
- Lang, D. 2014, *AJ*, **147**, 108
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, **379**, 1599
- Lewis, J. R., Irwin, M., & Bunclark, P. 2010, in *Astronomical Data Analysis Software and Systems XIX*, eds. Y. Mizumoto, K.-I. Morita, & M. Ohishi, *ASP Conf. Ser.*, **434**, 91
- Lombardi, M. 2009, *A&A*, **493**, 735
- Lombardi, M., & Alves, J. 2001, *A&A*, **377**, 1023
- Lombardi, M., Alves, J., & Lada, C. J. 2006, *A&A*, **454**, 781
- Lombardi, M., Alves, J., & Lada, C. J. 2011, *A&A*, **535**, A16
- Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, *A&A*, **566**, A45
- Markwardt, C. B. 2009, in *Astronomical Data Analysis Software and Systems XXIII*, eds. D. A. Bohlender, D. Durand, & P. Dowler, *ASP Conf. Ser.*, **411**, 251
- Megeath, S. T., & Wilson, T. L. 1997, *AJ*, **114**, 1106
- Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, *AJ*, **144**, 192
- Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, *A&A*, **474**, 515
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, *New Astron.*, **15**, 433
- Muench, A., Getman, K., Hillenbrand, L., & Preibisch, T. 2008, *Star Formation in the Orion Nebula I: Stellar Content*, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 483
- Muench, A. A., Lada, E. A., & Lada, C. J. 2002, in *Meeting Abstracts, BAAS*, **34**, 1210
- Nakamura, F., Miura, T., Kitamura, Y., et al. 2012, *ApJ*, **746**, 25
- Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, *ApJ*, **278**, L1
- O'Dell, C. R., Muench, A., Smith, N., & Zapata, L. 2008, *Star Formation in the Orion Nebula II: Gas, Dust, Proplyds and Outflows*, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 544
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Machine Learning Res.*, **12**, 2825
- Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, *A&A*, **524**, A42
- Peterson, D. E., & Megeath, S. T. 2008, *The Orion Molecular Cloud 2/3 and NGC 1977 Regions*, in *Handbook of Star Forming Regions*, Vol. I, ed. B. Reipurth, 590
- Phelps, R. L., & Lada, E. A. 1997, *ApJ*, **477**, 176
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, **518**, L1
- Pillitteri, I., Wolk, S. J., eath, S. T., et al. 2013, *ApJ*, **768**, 99
- Popowicz, A., Kurek, A. R., & Filus, Z. 2013, *PASP*, **125**, 1119
- Reipurth, B., & Heathcote, S. 1997, in *Herbig-Haro Flows and the Birth of Stars*, eds. B. Reipurth, & C. Bertout, *IAU Symp.*, **182**, 3
- Reipurth, B., Heathcote, S., Morse, J., Hartigan, P., & Bally, J. 2002, *AJ*, **123**, 362
- Robberto, M., Soderblom, D. R., Scandariato, G., et al. 2010, *AJ*, **139**, 950
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, **409**, 523
- Román-Zúñiga, C. G., Elston, R., Ferreira, B., & Lada, E. A. 2008, *ApJ*, **672**, 861
- Sandstrom, K. M., Peek, J. E. G., Bower, G. C., Bolatto, A. D., & Plambeck, R. L. 2007, *ApJ*, **667**, 1161
- Schlaflay, E. F., Green, G., Finkbeiner, D. P., et al. 2014, *ApJ*, **786**, 29
- Schlaflay, E. F., Green, G., Finkbeiner, D. P., et al. 2015, *ApJ*, **799**, 116
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Soto, M., Barbá, R., Gunthardt, G., et al. 2013, *A&A*, **552**, A101
- Spezzi, L., Petr-Gotzens, M. G., Alcalá, J. M., et al. 2015, *A&A*, **581**, A140
- Strom, K. M., & Strom, S. E. 1993, *ApJ*, **412**, L63
- Strom, K. M., Strom, S. E., & Merrill, K. M. 1993, *ApJ*, **412**, 233
- Stutz, A. M., & Kainulainen, J. 2015, *A&A*, **577**, L6
- Teixeira, P. S., Lada, C. J., Young, E. T., et al. 2006, *ApJ*, **636**, L45
- Tody, D. 1986, in *Instrumentation in astronomy VI*, ed. D. L. Crawford, *SPIE Conf. Ser.*, **627**, 733
- Trumpler, R. J. 1931, *PASP*, **43**, 255
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, **143**, 9
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, **154**, 1
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, *AJ*, **145**, 44



**Fig. A.1.** Luminosity functions (histograms) for the three observed bands with bin widths of 0.5 mag. The solid black lines represent the 2MASS histograms for sources with a quality flag of at least C for the same coverage as the VISTA survey. The dashed lines show the complete 2MASS histograms.

**Table A.1.** Tabulated completeness estimates from our artificial star tests for the entire Orion A survey.

Completeness (%)	<i>J</i> (mag)	<i>H</i> (mag)	<i>K<sub>S</sub></i> (mag)
99	19.06	18.74	18.5
95	20.38	19.87	18.97
90	20.56	20.09	19.11
80	20.74	20.29	19.26
50	21.04	20.58	19.51
10	21.45	20.94	19.85

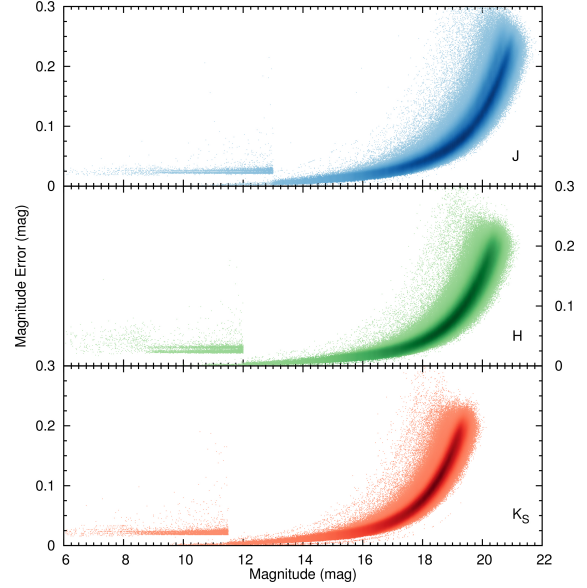
**Notes.** The large gap between 99% and 95% is a consequence of the variable observing conditions.

## Appendix A: Data characteristics

Here, we want to provide additional information on the VISTA Orion A source catalog, in particular on the photometric and astrometric quality of the data reduction, as well as present some quality control parameters.

### A.1. Photometric properties

The luminosity functions for all three bands are displayed in Fig. A.1. Here we show both the complete 2MASS point source catalog histograms, as well as only sources with a quality flag of either A, B, or C since only these were added into the VISTA Orion A catalog. At the bright end we closely match with

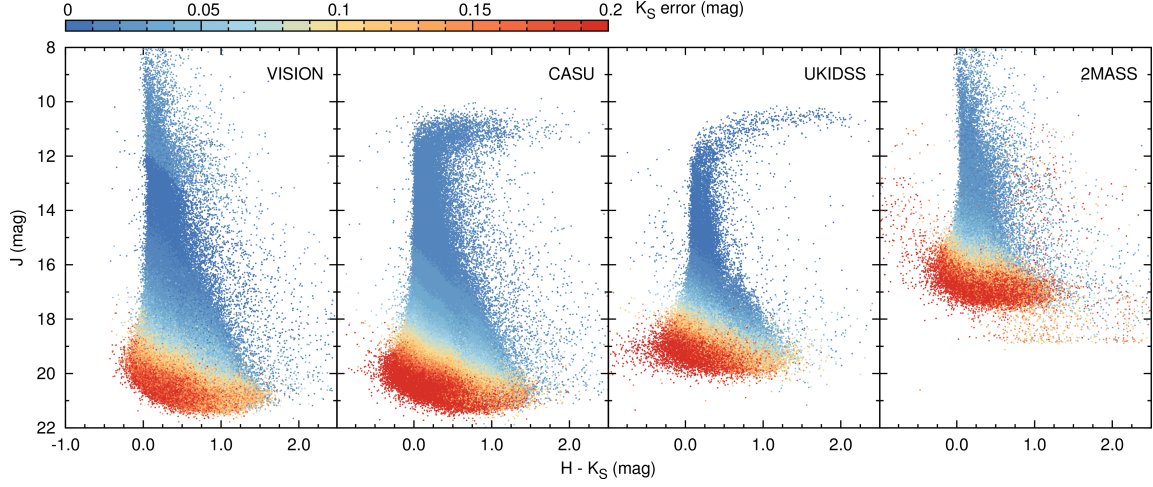


**Fig. A.2.** Magnitudes and their associated errors in our survey. The discontinuity at the bright end is due to the replacement with 2MASS measurements. The shading indicates source density in a  $0.1 \times 0.02$  mag box in this parameter space.

2MASS since essentially all bright sources originate from the reference catalog. The minor discrepancy between the VISTA histogram and the cleaned 2MASS histogram at the bright end mostly comes from the region around the ONC which we cleaned from bad detections by hand. Depending on the band, we gain between three and four magnitudes in dynamic range over the reference catalog.

Magnitudes and their errors are shown in Fig. A.2. The magnitude errors only start to increase significantly around 18 mag in all bands. One can clearly see the discontinuity at the bright end due to the catalog extension with 2MASS. There still are pure VISION sources below the cut-offs since some sources lie above the cleaning threshold in 2MASS (see Sect. 3.7.3 for details), but have a brighter magnitude in our survey. At a given magnitude, the error distribution does not follow a Gaussian, but naturally has a longer tail towards larger errors. This effect is introduced by unequal coverage and if we select those sources with errors larger than the median error, we find that these indeed fall into regions with low effective exposure time. We again note here that these errors serve as lower limits only. In addition, we emphasize that the errors in the public catalog come from two different data sets (VISTA and 2MASS) which must be considered carefully for any application involving them.

Figure A.3 shows the color-magnitude diagrams (*J* vs. *H* − *K<sub>S</sub>*) of our survey in comparison to several other data products including the CASU reduction, UKIDSS (DR10, Lawrence et al. 2007), and the 2MASS point source catalog. The errors of the photometry in *K<sub>S</sub>* are shown in a discrete color-code. We note several things here: (a) our dedicated survey goes much deeper than UKIDSS and 2MASS; (b) VISION covers the largest dynamic range because we replaced the bright end of the luminosity function with 2MASS photometry; (c) The photometric errors also seem to improve over the standard CASU pipeline; (d) our optimized aperture photometry produces a narrower



**Fig. A.3.**  $J$  vs.  $H - K_s$  color-magnitude diagrams generated for four different data sets. Our data features the largest consistent dynamic range, as well as the lowest photometric error overall. Also, due to our optimized aperture photometry, the stellar sequence is slightly narrower at the faintest end. Note here, that UKIDSS and 2MASS data come from unoptimized surveys in this region and naturally do not reach equal depths.

stellar sequence for faint magnitudes (compare e.g. the blue sources at  $J \sim 20$  mag).

The completeness of our survey is expected to be spatially highly variable due to changing observing conditions. A reliable completeness estimate can only be determined from the same data on which source detection and extraction was performed. Thus, any test can not be performed on the stacked pawprints, but must be applied to the Orion A mosaics from which the final source catalogs were generated. For this reason we performed artificial star tests on 10 arcmin wide sub-fields of the full mosaics (the size of a VIRCAM detector amounts  $\sim 11.6 \times 11.6$  arcmin). For each of these fields we applied multiple subsequent processing steps: we (1) constructed a PSF model with PSFEx; (2) performed source extraction and profile fitting with SExtractor to subtract all significant sources; (3) measured the source density; (4) constructed a set of artificial stars from the given PSF (with the calculated stellar density) with magnitudes ranging from 17 to 22.5 mag with Skymaker; (5) performed source extraction on the artificial sources placed on the PSF-subtracted fields; (6) calculated the completeness relative to the input source list of the artificial stars. Steps 4–6 were repeated 50 times for statistical reasons. The final completeness estimate (as a function of magnitude) for each field was calculated as the mean of all iterations. The completeness-magnitude distribution for these sub-fields were very well fitted by a modified logistic function of the form:

$$f(x) = -\frac{1}{1 + e^{-k(x-x_0)}} + 1. \quad (\text{A.1})$$

We combined the results of these individual sub fields to estimate the completeness for the entire survey. The results of this procedure are displayed in Fig. A.4, where we show the completeness as a function of source magnitude as well as its spatial dependency (displayed at the 90% level) and tabulate the results in Table A.1. We note here, that the completeness function for the entire cloud is not well described by the aforementioned function due to the spatial variations. While for individual fields we observe almost 100% completeness up to the point where the function appears to drop, the distributions for the entire survey

show a small continuous decline towards this point. As a consequence we see a large difference in completeness magnitude between e.g. 99% and 95%, even though large parts of the survey are essentially complete at the latter limit.

The spatial variations of the completeness correlate well with the observing conditions (compare with Fig. A.6) and the presence of bright nebulous emission near the ONC. For e.g. tile S6 (east-most tile) in  $H$  band we observe PSF FWHMs exceeding 1 arcsec, and at the same time a noticeably shallower completeness. Tile S3 in  $J$  featured the best observing conditions in our survey where we see an increase in sensitivity of  $\sim 1$  mag for this region.

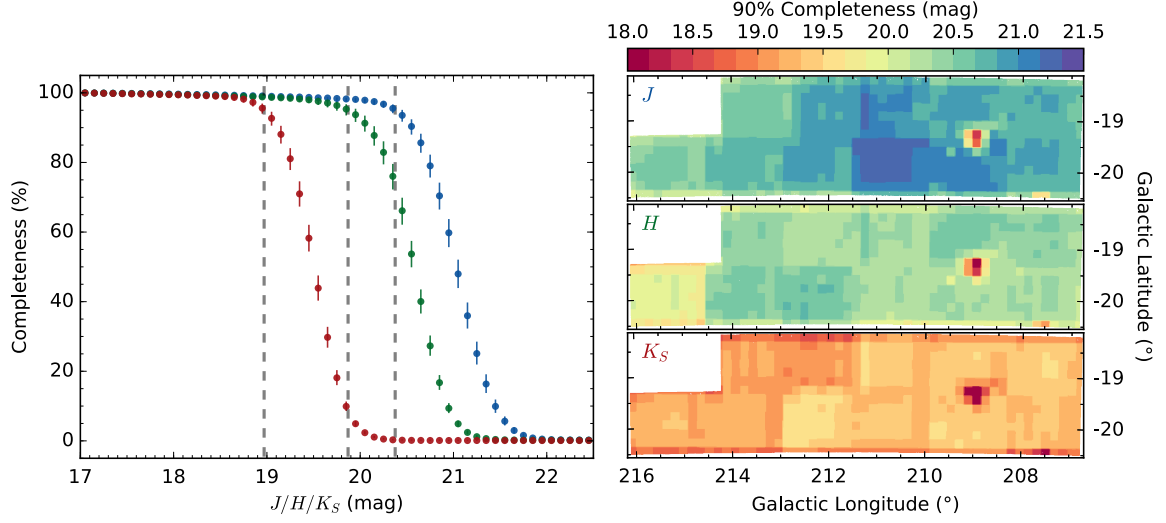
## A.2. Astrometric properties

The overall astrometric calibration of our data was done with Scamp using 2MASS as a reference catalog (see Sect. 3.4 for details). The global error budget resulted in an rms of about 70 mas with respect to reference sources and about 40 mas when considering internal source matches only. The most important factors contributing to this discrepancy are 2MASS S/N limits, general catalog errors, unresolved multiple sources, and unaccounted proper motions. To check for any remaining local systematic errors in VISION we used the final catalog, cross-matched again to 2MASS with a maximum allowed distance of 1 arcsec and subsequently calculated the mean astrometric offsets in boxes of  $15 \times 15$  arcmin. The result is displayed in Fig. A.5 which shows the sum of all offsets in a given box including magnitude and direction. No local systematic trend can be seen. The errors in this figure are generally smaller than the given global error due to the averaging in each given box. When decreasing the box size to below 10 arcmin we still see no systematic trends across the field, but the magnitude of the errors increase as expected.

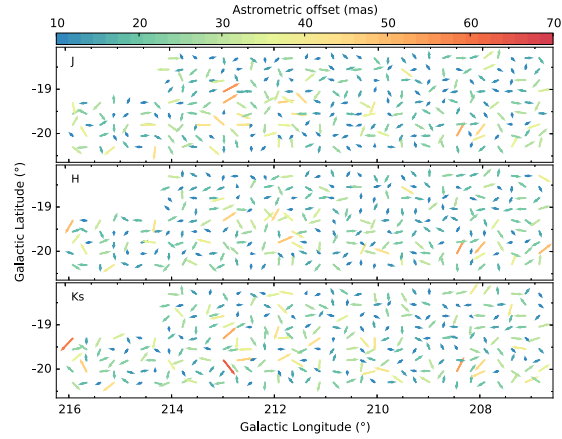
## A.3. Quality control parameters

As part of the data calibration we also generated several quality control parameters for the entire survey. The most important among them are the image quality and survey coverage. Each





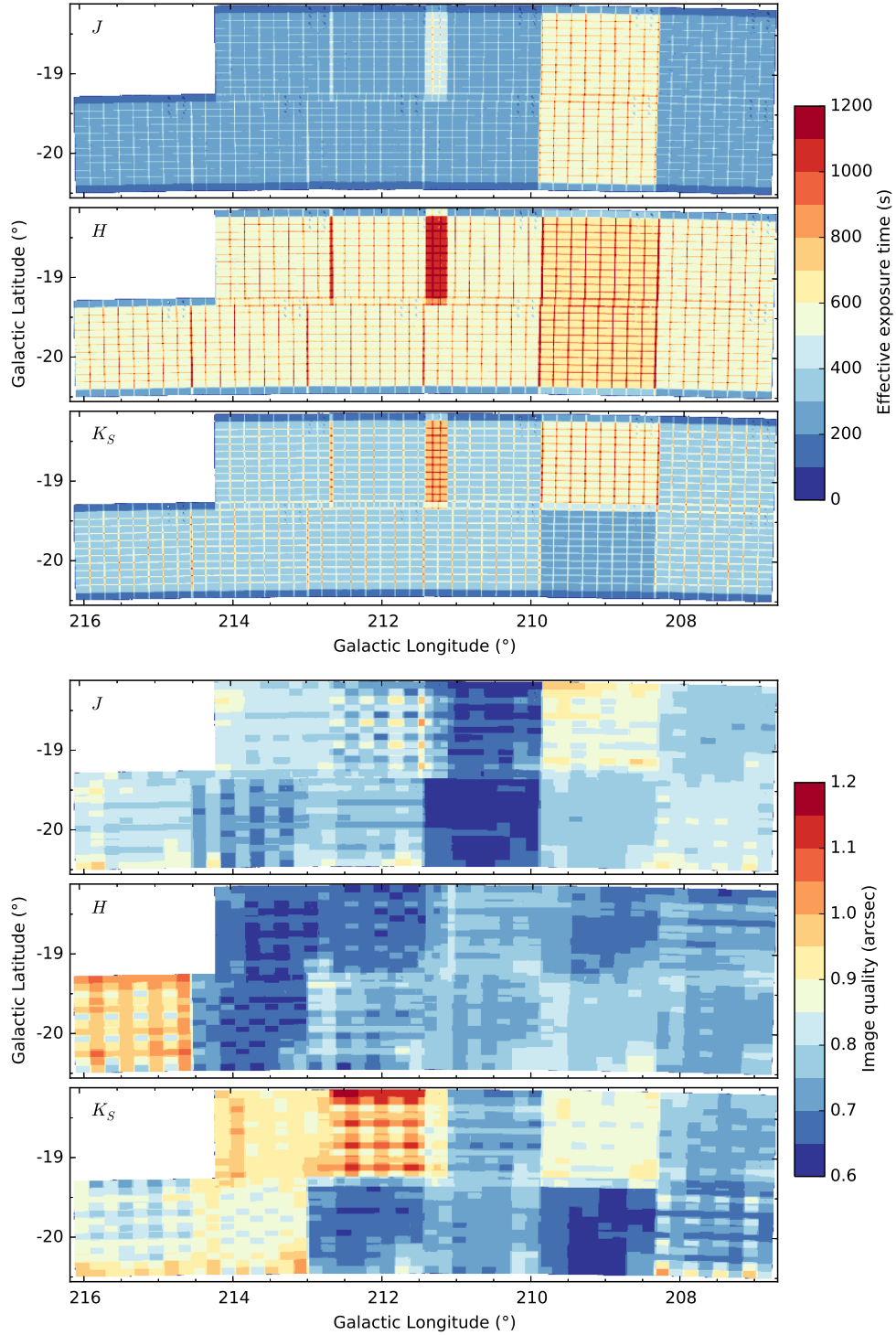
**Fig. A.4.** Completeness estimates for the entire survey in all three observed bands. *Left panel:* the completeness estimate as a function of source magnitude. The blue, green, and red points represent the  $J$ ,  $H$ , and  $K_S$  bands respectively. The errorbars correspond to a 99% bootstrap confidence interval ( $\alpha = 0.01$ ) estimated with  $10^3$  samples. The vertical gray lines correspond to 90% completeness. *Right panel:* spatial dependency of the completeness at the 90% level. Clearly, the bright nebula near the ONC has a negative impact on source extraction. Well visible are the effects of observing conditions on the completeness. Clearly, better seeing (compare e.g.  $J$  band with Fig. A.6) results in more sensitive observations.



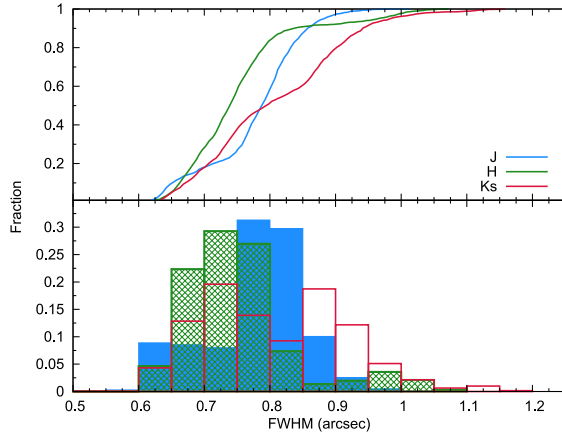
**Fig. A.5.** Mean astrometric offsets relative to 2MASS for the final Orion A source catalog in  $15 \times 15$  arcmin boxes. No systematic trends across the entire field for all three bands are visible. The colors indicate the mean offset in the box which also is linearly proportional to the arrow lengths.

source in the catalog is also supplemented with its associated effective exposure time, frame coverage, observing date (MJD), source FWHM and local seeing conditions. Figure A.6 visualizes effective exposure time and image quality for our survey. The image quality here refers to the measured seeing value for each photometrically stable subset (see Sect. 3.5 for details). In the exposure time maps the overlaps between the detectors and tiles are clearly visible. Note here tile S2 in  $K_S$  for which we only included one of the observed sequences due to the large discrepancy in observing conditions. Fortunately the other sequence features one of the best image qualities of the entire survey and no obvious decrease in the completeness is observed with respect to the other tiles (compare with Fig. A.4). We can also see that the image quality can vary by a few tens of percent even within tiles. This complex structure is simply a consequence of the observing strategy with VISTA where the same position is observed by multiple detectors spread out over the entire length of the OB.

Most data were taken under excellent seeing conditions. The mean seeing, determined as the FWHM from a bright, high S/N subset, was 0.78, 0.75, 0.8 arcsec for Orion A, and 0.69, 0.77, 0.76 for the CF in  $J$ ,  $H$  and  $K_S$ , respectively. Figure A.7 displays histograms of the image quality statistics. The  $K_S$  data show a bi-modal distribution which can (coincidentally) also clearly be seen in Fig. A.6 in a roughly east-west oriented gradient.



**Fig. A.6.** Effective exposure time (*top*) and image quality (FWHM, *bottom*) across the Orion A mosaic. Note the lower effective exposure time for tile S2 in  $K_S$  and the very patchy FWHM structure.



**Fig. A.7.** Seeing statistics for the entire Orion A survey. The *bottom row* shows the normalized histogram for all three bands in bin widths of 0.05 arcsec whereas the top graph displays the cumulative histograms with a refined resolution. Most data have been taken during very good seeing conditions resulting in FWHMs measured on point-sources in the range between 0.7 and 0.85 arcsec for *J* and *H*; only *K<sub>s</sub>* features a significant amount of data with point-source FWHM values  $\geq 0.85$  arcsec.



## Appendix B: Tables

Here we provide supplementary tables for the main text. In addition to Table 3, which shows the published columns of the source catalog, Table B.1 contains a sample of 12 out of the 35 total columns from the final catalog. Table B.2 contains coordinates and photometry for all sources discussed in Sect. 4.2.1.

**Table B.1.** Sample data taken from the VISTA Orion A catalog.

VISION	RAJ2000 (hh:mm:ss)	DEJ2000 (dd:mm:ss)	Class_cog {0, 1}	Class_sex [0, 1]	$K_S$ (mag)	$K_S$ _err (mag)	$K_S$ _mjd	$K_S$ _exptime (s)	$K_S$ _fwhm (arcsec)	$K_S$ _seeing (arcsec)	$K_S$ _coverage
05362896-0409496	05:36:28.96	-04:09:49.6	0	1.0	19.289	0.184	56319.047	400	1.55	0.76	10
05362896-0711209	05:36:28.96	-07:11:20.9	0	1.0	18.396	0.103	56322.035	440	1.85	0.7	11
05362896-0724448	05:36:28.96	-07:24:44.8	1	0.973	17.443	0.044	56322.043	560	0.79	0.73	14
05362896-0652099	05:36:28.96	-06:52:09.9	0	0.324	17.627	0.058	56322.03	400	1.34	0.72	10
05362896-0822224	05:36:28.96	-08:22:22.4	1	0.959	18.868	0.198	56332.07	200	0.88	0.74	5
05362896-0450050	05:36:28.96	-04:50:05.0	0	0.004	19.096	0.174	56320.555	630	1.54	0.86	21
05362896-0555005	05:36:28.96	-05:55:00.5	0	1.0	19.084	0.168	56281.82	450	1.13	0.85	15
05362896-0521375	05:36:28.96	-05:21:37.5	0	1.0	18.583	0.105	56320.555	750	2.37	0.87	25
05362896-0808415	05:36:28.96	-08:08:41.5	1	0.979	17.534	0.059	56332.07	400	0.9	0.83	10
05362896-0411396	05:36:28.96	-04:11:39.6	1	0.61	19.143	0.155	56319.043	480	1.19	0.77	12
05362896-0406475	05:36:28.96	-04:06:47.5	1	0.995	18.981	0.136	56319.04	400	1.16	0.76	10
05362896-0454271	05:36:28.96	-04:54:27.1	1	0.983	17.672	0.059	56320.555	600	0.86	0.85	20
05362896-0334519	05:36:28.96	-03:34:51.9	0	1.0	17.548	0.076	56319.05	200	2.77	0.75	5
05362897-0740077	05:36:28.97	-07:40:07.7	1	0.999	18.011	0.066	56322.035	520	0.85	0.74	13
05362897-0544574	05:36:28.97	-05:44:57.4	0	0.0	18.761	0.127	56320.55	600	1.57	0.87	20
05362897-0354050	05:36:28.97	-03:54:05.0	1	0.983	17.353	0.048	56319.04	360	0.8	0.75	9
05362897-0827114	05:36:28.97	-08:27:11.4	1	0.996	14.922	0.02	56332.066	200	0.78	0.75	5
05362897-0452245	05:36:28.97	-04:52:24.5	1	0.98	17.89	0.07	56320.555	600	0.94	0.85	20
05362897-0528390	05:36:28.97	-05:28:39.0	1	0.978	17.14	0.047	56320.555	600	0.91	0.85	20
05362897-0531284	05:36:28.97	-05:31:28.4	1	0.983	15.278	0.017	56320.555	600	0.88	0.85	20
05362897-0328278	05:36:28.97	-03:28:27.8	1	0.943	18.101	0.11	56319.055	200	1.12	0.76	5
05362898-0438323	05:36:28.98	-04:38:32.3	1	0.069	17.997	0.08	56320.555	600	1.12	0.88	20
05362898-0407127	05:36:28.98	-04:07:12.7	1	0.911	15.942	0.021	56319.043	480	0.8	0.76	12
05362898-0814538	05:36:28.98	-08:14:53.8	0	0.819	18.699	0.115	56332.08	400	1.02	0.72	10
05362898-0449499	05:36:28.98	-04:49:49.9	1	0.989	16.752	0.034	56320.555	660	0.87	0.86	22
05362898-0502207	05:36:28.98	-05:02:20.7	1	0.975	18.191	0.068	56322.04	630	1.13	0.86	21
05362898-0748052	05:36:28.98	-07:48:05.2	1	0.96	17.806	0.052	56327.055	800	0.8	0.75	20
05362898-0434172	05:36:28.98	-04:34:17.2	1	0.947	16.247	0.025	56320.555	600	0.93	0.87	20
05362898-0654349	05:36:28.98	-06:54:34.9	1	0.999	18.537	0.1	56322.03	400	0.82	0.72	10
05362898-0532286	05:36:28.98	-05:32:28.6	1	0.975	18.349	0.105	56320.555	600	1.02	0.85	20
05362898-0327518	05:36:28.98	-03:27:51.8	1	0.994	15.162	0.02	56319.055	200	0.83	0.76	5
05362899-0702345	05:36:28.99	-07:02:34.5	1	1.0	19.029	0.144	56322.043	520	0.92	0.73	13
05362899-0343035	05:36:28.99	-03:43:03.5	1	1.0	19.647	0.199	56319.047	560	0.68	0.77	14
05362899-0810323	05:36:28.99	-08:10:32.3	1	0.968	16.264	0.028	56332.07	400	0.81	0.83	10
05362899-0754268	05:36:28.99	-07:54:26.8	0	1.0	18.765	0.133	56332.086	400	1.19	0.79	10

**Notes.** Only 12 out of the 35 total columns are shown. The full table is available in electronic form at the CDS.

**Table B.2.** Cross identifications, coordinates, and magnitudes for all the objects identified in Sect. 4.2.1.

VISION	Label <sup>a</sup>	Right Ascension (hh:mm:ss)	Declination (dd:mm:ss)	<i>J</i> (mag)	<i>H</i> (mag)	<i>K<sub>S</sub></i> (mag)	Size <sup>b</sup> (arcsec, pc)	ID <sup>c</sup>
Class I								
05381810-0702259	a	05:38:18.10	−07:02:25.9	9.253	6.762	5.154	330, 0.66	670
05402745-0727300	b	05:40:27.45	−07:27:30.0	—	—	9.694	180, 0.36	536
05424707-0817070	c	05:42:47.07	−08:17:07.0	15.998	14.158	11.063	33, 0.07	287
05413419-0835274	d	05:41:34.19	−08:35:27.4	16.232	14.430	12.103	68, 0.14	257
05352985-0626583	e	05:35:29.85	−06:26:58.3	14.957	14.491	12.302	68, 0.14	879
05363034-0432170	f	05:36:30.34	−04:32:17.0	15.496	13.641	12.648	50, 0.10	2748
05410201-0806019	g	05:41:02.01	−08:06:01.9	17.494	14.606	13.118	68, 0.14	362
05412474-0754081	h	05:41:24.74	−07:54:08.1	—	17.745	13.416	60, 0.12	439
05412398-0753421	i	05:41:23.98	−07:53:42.1	—	17.437	13.731	60, 0.12	445
05350554-0551541	j	05:35:05.54	−05:51:54.1	18.628	16.003	14.108	66, 0.13	1165
05344909-0541419	k	05:34:49.09	−05:41:41.9	—	16.780	15.240	53, 0.11	1294
05402095-0756240	l	05:40:20.95	−07:56:24.0	—	—	15.540	57, 0.11	423
05325056-0534424	m	05:32:50.56	−05:34:42.4	19.486	15.958	16.066	45, 0.09	1433
05361721-0638016	n	05:36:17.21	−06:38:01.6	—	17.089	—	50, 0.10	823
—	o	—	—	—	—	—	53, 0.11	1504
Class II								
05362543-0642577	a	05:36:25.43	−06:42:57.7	8.107	6.964	5.947	180, 0.36	796
05384279-0712438	b	05:38:42.79	−07:12:43.8	10.823	9.283	8.124	83, 0.17	618
05384322-0658089	c	05:38:43.22	−06:58:08.9	12.046	9.978	8.617	75, 0.15	726
05404806-0805587	d	05:40:48.06	−08:05:58.7	9.914	9.140	8.751	60, 0.12	364
05404662-0807128	e	05:40:46.62	−08:07:12.8	11.928	10.203	9.333	120, 0.24	351
05353163-0500141	f	05:35:31.63	−05:00:14.1	14.517	11.964	10.129	45, 0.09	2450
05362378-0623113	g	05:36:23.78	−06:23:11.3	15.377	13.225	11.111	60, 0.12	925
05410413-0923194	h	05:41:04.13	−09:23:19.4	13.497	13.033	12.637	27, 0.05	129
05363700-0614579	i	05:36:37.00	−06:14:57.9	16.731	14.379	13.258	45, 0.09	996
05384652-0705375	j	05:38:46.52	−07:05:37.5	16.978	15.933	13.550	45, 0.09	649
Class III								
05350906-0614200	a	05:35:09.06	−06:14:20.0	9.612	9.292	9.188	30, 0.06	125
05375451-0656455	b	05:37:54.51	−06:56:45.5	10.704	9.916	9.697	30, 0.06	1040
05352974-0548450	c	05:35:29.74	−05:48:45.0	10.739	10.067	9.893	30, 0.06	765
05345803-0612238	d	05:34:58.03	−06:12:23.8	10.893	10.163	9.974	30, 0.06	137
05431072-0831500	e	05:43:10.72	−08:31:50.0	10.823	10.297	10.125	30, 0.06	289
New YSO candidates								
05312709-0427593	a	05:31:27.09	−04:27:59.3	13.383	11.086	9.425	135, 0.27	—
05315171-0523082	b	05:31:51.71	−05:23:08.2	12.006	10.492	9.709	90, 0.18	—
05324165-0535461	c	05:32:41.65	−05:35:46.1	17.965	13.797	11.389	45, 0.09	—
05324165-0536115	d	05:32:41.65	−05:36:11.5	—	18.889	14.060	30, 0.06	—
05305155-0410348	e	05:30:51.55	−04:10:34.8	—	—	14.415	45, 0.09	—
05305129-0410322	e	05:30:51.29	−04:10:32.2	—	—	13.835	45, 0.09	—
New galaxy cluster candidates								
—	a	05:40:44	−09:57:56	—	—	—	150, —	—
—	b	05:47:05	−08:55:17	—	—	—	150, —	—
—	c	05:41:51	−09:06:33	—	—	—	150, —	—
—	d	05:39:46	−08:47:38	—	—	—	150, —	—
—	e	05:44:53	−08:03:58	—	—	—	120, —	—
—	f	05:42:36	−06:59:34	—	—	—	200, —	—
—	g	05:40:25	−05:59:36	—	—	—	160, —	—
—	h	05:32:06	−06:05:21	—	—	—	150, —	—
—	i	05:31:08	−05:30:48	—	—	—	160, —	—
—	j	05:30:28	−04:14:42	—	—	—	170, —	—

**Notes.** These objects are also shown in Figs. 12 and 13. <sup>(a)</sup> Refers to the labels in the bottom left corners of the sub-figures in Figs. 12 and 13. <sup>(b)</sup> Size of the postage stamp in Figs. 12 and 13. The physical sizes were calculated with the adopted distance of 414 pc. <sup>(c)</sup> For Class I/II sources the ID refers to the internal numbering of Megeath et al. (2012), for Class III to Pillitteri et al. (2013).





# Estimating extinction

## 3.1 Overview

With the obtained photometric database from the VISTA Orion A survey as presented in the previous chapter, the next goal was to investigate whether established extinction tools can be used together with the latest deep photometric observations. The major concern here was, that these methods are conceptually based on statistical approaches which can introduce significant biases in the extinction measurements under certain circumstances. Initial tests quickly revealed that traditional methods to estimate line-of-sight extinctions are not easily applicable anymore to the new deep photometric data. The cause for this issue is a significant amount of detected extragalactic sources for which statistical properties are not properly taken into account by most previous published methods. These older methods either require prior information on source characteristics, or deliver large statistical errors in the determination of the line-of-sight extinction.

These issues called for the development of a new technique which is capable of determining unbiased extinction measurements, regardless of the sensitivity limit of the observations. This new method, named PNICER in reference to well-established techniques, is based on machine learning principles and is capable of de-reddening sources in arbitrary feature combinations and numbers of dimensions. Furthermore, one major development goal was to make this tool accessible to large parts of the astronomical research community. For this reason, two major additional goals were (a) to offer publicly available, easy-to-install software with a simple interface and (b) to provide suitable performance, even for cases of millions (or even billions) of sources.

In this manuscript, I first present a short literature review on the available techniques and continue to describe their properties and disadvantages when being confronted with deep photometric data. I continue to explain the concepts of the newly developed extinction tool and explain the methodology with a simple, yet practical, example. Finally, I verify the method with an application to typical photometric data where I construct a large extinction map and compare the results to data available from previous studies. I also review the code performance for typical

use-cases and outline the software prerequisites. I note here, that this manuscript is not part of the VISION publication series, because its scope goes beyond the NIR VISION observations of Orion A.

## 3.2 Publication details

**Title:** Estimating extinction using unsupervised machine learning

**Authors:** Stefan Meingast, Marco Lombardi, João Alves

**Status:** Accepted for publication in Astronomy & Astrophysics, February 24th 2017

**Bibliographic reference:** Astronomy & Astrophysics, Volume 601, id.A137, 12 pp.

**DOI:** [10.1051/0004-6361/201630032](https://doi.org/10.1051/0004-6361/201630032)

**Software availability:** <http://smeingast.github.io/PNICER/>

**Own contributions:** Literature research, method development, software implementation and publication, testing and validation, application to VISTA and 2MASS data, preparation of figures and plots, paper writing.

# Estimating extinction using unsupervised machine learning

Stefan Meingast<sup>1</sup>, Marco Lombardi<sup>2</sup>, and João Alves<sup>1</sup>

<sup>1</sup> University of Vienna, Department of Astrophysics, Türkenschanzstrasse 17, 1180 Wien, Austria  
e-mail: [stefan.meingast@univie.ac](mailto:stefan.meingast@univie.ac).

<sup>2</sup> University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

Received 8 November 2016 / Accepted 24 February 2017

## ABSTRACT

Dust extinction is the most robust tracer of the gas distribution in the interstellar medium, but measuring extinction is limited by the systematic uncertainties involved in estimating the intrinsic colors to background stars. In this paper we present a new technique, PNICER, that estimates intrinsic colors and extinction for individual stars using unsupervised machine learning algorithms. This new method aims to be free from any priors with respect to the column density and intrinsic color distribution. It is applicable to any combination of parameters and works in arbitrary numbers of dimensions. Furthermore, it is not restricted to color space. Extinction toward single sources is determined by fitting Gaussian mixture models along the extinction vector to (extinction-free) control field observations. In this way it becomes possible to describe the extinction for observed sources with probability densities, rather than a single value. PNICER effectively eliminates known biases found in similar methods and outperforms them in cases of deep observational data where the number of background galaxies is significant, or when a large number of parameters is used to break degeneracies in the intrinsic color distributions. This new method remains computationally competitive, making it possible to correctly de-redden millions of sources within a matter of seconds. With the ever-increasing number of large-scale high-sensitivity imaging surveys, PNICER offers a fast and reliable way to efficiently calculate extinction for arbitrary parameter combinations without prior information on source characteristics. The PNICER software package also offers access to the well-established NICER technique in a simple unified interface and is capable of building extinction maps including the NICEST correction for cloud substructure. PNICER is offered to the community as an open-source software solution and is entirely written in Python.

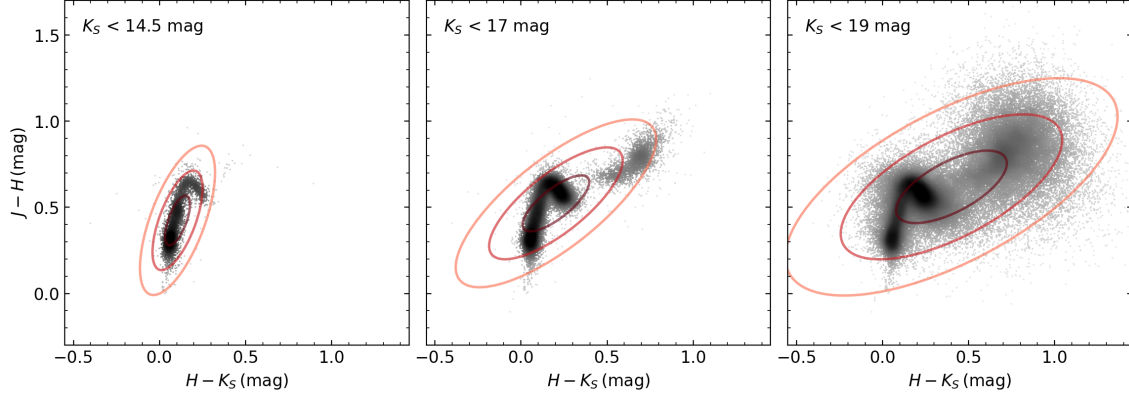
**Key words.** dust, extinction – methods: data analysis – methods: statistical – techniques: miscellaneous

## 1. Introduction

Mapping the gas and dust distribution in the interstellar medium is vital to understand how diffuse clouds evolve into stars and planets, allowing for important insights on the physical mechanisms involved in processes such as cloud assemblage, evolution of dust grains, core formation and collapse, cluster formation, and the role of turbulence and feedback. Traditional techniques to map large-scale column density distributions, relying on optical star counts (e.g., Bok & Cordwell 1973; Cambrésy 1999; Dobashi et al. 2005) are limited to low column-densities and with the advent of near-infrared (NIR) cameras sensitive to wavelengths where clouds become transparent and reddened background stars are detected, new methods exploiting reddening have been developed to systematically study dense gas in the interstellar medium (e.g., Lada et al. 1994; Alves et al. 1998; Lombardi & Alves 2001; Foster et al. 2008; Lombardi 2009; Majewski et al. 2011). The classic methods using star counts can still be applied to NIR observations (e.g., Dobashi 2011; Dobashi et al. 2013) for greater dynamic range, but their general limitations remain and more advanced methods making use of reddening can deliver lower-noise and more robust results for the same data set. Today, the most commonly used techniques to trace column density in the dense interstellar medium rely on: 1) measuring dust thermal emission at mm and far-infrared wavelengths; 2) molecular line emission; or 3) NIR dust extinction.

Each of these techniques has its own strengths but also disadvantages. While mapping the dust thermal emission can provide

large dynamic range and high-resolution maps particularly in regions away from rich stellar backgrounds, the conversion from the measured continuum emission to column-densities is far from trivial as it requires assumptions about dust emissivity and temperature. At least in regions of active star formation the temperature varies widely due to feedback processes from early-type stars. Molecular line emission can become optically thick in dense environments and furthermore relies on local (constant) conversion factors of the measured emission relative to the hydrogen abundance (the so-called X-factor; for a discussion on its variations see e.g., Pineda et al. 2008). The NIR dust extinction method relies on measuring color excesses of sources in the background of molecular clouds. These discrete measurements are then used to reconstruct the smooth column-density distribution (e.g., with Gaussian estimators). Goodman et al. (2009) showed that NIR extinction is relatively bias-free and provides more robust measurements of column density than the other tracing techniques. Ultimately, however, NIR dust extinction measurements are limited by the available number of background sources in the region of interest. In particular this method is limited in regions where even very sensitive observations will not be able to “peer through” high column-densities of ( $A_V \gtrsim 100$  mag) or where naturally fewer background sources are available (e.g., toward the galactic poles). Due to the declining dust opacity toward longer wavelengths one could argue that going beyond the NIR would provide further benefits. This is, however, not the case as beyond  $\sim 5 \mu\text{m}$  dust emission starts to dominate over the abrupt drop in stellar flux of background stars, acting as a bright screen, and more complex dust absorption



**Fig. 1.** Color-color diagram of the NIR data from Meingast et al. (2016) for different magnitudes limits. The red ellipses are the covariance estimates drawn for one, two, and three standard deviations. We find a relatively small distribution width for the typical 2MASS magnitude limit at  $K_S \sim 14.5$  mag in both axes which allows for an efficient description of intrinsic colors via a mean and covariance estimate. By increasing the magnitude limit, galaxies at  $H - K_S \sim 0.7$ ,  $J - H \sim 0.8$  mag significantly broaden the distribution and introduce a large statistical error when estimating intrinsic colors with a single averaged value. Also very well visible is the shift of the mean intrinsic color (i.e., the center of the ellipse) even beyond the M-branch of the stellar main sequence toward galaxies.

and scattering processes play a role. Several empirical and theoretical studies have shown that there is relatively little variation in the NIR extinction law across different environments and variable dust properties (Indebetouw et al. 2005; Flaherty et al. 2007; Ascenso et al. 2013), making it ideal for robust column density measurements (and particularly for the dense gas mass distribution) in the interstellar medium.

## 2. Motivation

In order to derive color excesses for individual sources it is necessary to estimate their intrinsic colors. Color excess occurs as a consequence of absorption and scattering processes when light travels through the interstellar medium and is defined via

$$E(m_1 - m_2) = (m_1 - m_2) - (m_1 - m_2)_0 \quad (1)$$

$$= (m_1 - m_{1,0}) - (m_2 - m_{2,0}) = A_{m_1} - A_{m_2}, \quad (2)$$

where the  $m_i$  describe source magnitudes in different passbands (e.g.,  $H$  and  $K_S$ ). The first term on the right-hand side of Eq. (1) refers to observed colors while  $(m_1 - m_2)_0$  are intrinsic colors and the  $A_m$  define the total extinction in the  $m_i$  passband in magnitudes. Accurate estimates of intrinsic colors are not trivially derived and in principle require detailed knowledge about the characteristics of each source. For example stars need to be distinguished from (unresolved) galaxies and different stellar spectral classes show diverse intrinsic colors (e.g., main sequence and dwarf stars need to be separated from giants). The situation for main sequence stars becomes more relaxed for near, and mid-infrared wavelengths as the spectral energy distribution flattens and thus produces relatively narrow sequences in color-color space. Therefore, inferring a single average intrinsic color for all sources introduces only small statistical and systematic errors as long as this assumption does not break down. Lada et al. (1994) pioneered this technique using the  $H$  and  $K_S$  bands to map the dust distribution throughout IC 5146. Later, Alves et al. (1998) improved the method and named it the NICE technique (for Near-Infrared Color Excess) and applied it in the investigation of the internal structure of the dark cloud Barnard 68

based on color excess measurements made with deep NIR data (Alves et al. 2001). This method was then developed into the multi-band technique NICER (for near-infrared color excess revisited) by Lombardi & Alves (2001) which also offered an extended description of the intrinsic colors by measuring their distribution in an extinction-free nearby control field. Based on a combination of calculated covariance estimates with photometric measurement errors, color-excesses are calculated in a maximum-likelihood approach minimizing the resulting variance. Several studies of nearby giant molecular cloud complexes have used data from the Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006) in combination with the NICER method to study the dense gas mass distribution (e.g., Lombardi et al. 2006, 2008, 2011, Alves et al. 2014). In principle, the NICER method can be generalized and applied to any given set of color combinations as long as the interstellar reddening law at the corresponding wavelengths is well determined.

With the description of intrinsic colors via a Gaussian distribution, characterized by the mean and covariance of the measured colors in an extinction-free control field, a particular problem affects applications of NICER with very deep observations: for 2MASS data the mean  $J-H$  and  $H-K_S$  colors are well determined and show only a relatively small variance. For deeper and more sensitive observations, however, a large number of galaxies enters the color space. The arising issue is illustrated in Fig. 1 where the NIR data of the control field from the Meingast et al. (2016) Orion A observations are displayed at different magnitude cuts. For these data, the completeness limit is found at  $K_S \sim 19$  mag, while for 2MASS data this limit is typically found at  $K_S \sim 14.5$  mag. The covariance estimates of this color combination are displayed as ellipses and are drawn for one, two, and three standard deviations. We find that the 2MASS sensitivity limits conveniently occur at magnitudes where galaxies are not detected in large quantities, resulting in a narrow distribution with standard deviations of  $\sigma_{J-H} = 0.14$  and  $\sigma_{H-K_S} = 0.07$  mag. Increasing the magnitude limit significantly broadens this distribution where for  $K_S < 17$  mag we find  $\sigma_{J-H} = 0.19$  and  $\sigma_{H-K_S} = 0.16$  mag and for  $K_S < 19$  mag  $\sigma_{J-H} = 0.32$  and  $\sigma_{H-K_S} = 0.21$  mag. For these data we therefore find that by



increasing the sensitivity limit by about 5 mag, the width of the distribution in the  $J - H$  vs.  $H - K_S$  color space is tripled.

The variance in the estimated extinction with NICER<sup>1</sup>, however, depends on the size of the ellipse along the extinction vector. We have tested the impact of increasing magnitude limits in the control field on the extinction error in the NICER algorithm by creating artificial photometry without photometric errors. This ensures that the resulting errors are exclusively determined by the covariance of the control field data. For the magnitude limits of  $K_S < \{14.5, 17, 19\}$  mag (as displayed in Fig. 1), we find  $\sigma_{A_K} = \{0.1, 0.15, 0.2\}$  mag. Hence, when the covariance of the control field dominates the error budget (i.e., small photometric errors) the error in the extinction estimates with NICER is doubled when increasing the magnitude limit from 14.5 to 19 mag. Moreover, the mean color (ellipse center) in the rightmost panel in Fig. 1 falls between the stellar M-branch and the galaxy locus and thus the extinction, on average, will be underestimated for stars and overestimated for galaxies. We emphasize here that NICER will still accurately reflect this behavior by returning larger statistical errors. However, the calculated (mean) extinction estimate will be systematically shifted for both stars and galaxies.

In addition to the increased errors when dealing with deep observations, NICER is affected by a bias when estimating color excess in highly extinguished regions. In this case the populations in the science field and the control field will be different from each other as for the high column-density regions intrinsically faint sources (preferentially galaxies) will be shifted beyond the photometric sensitivity limit of the observations. For observations with a given sensitivity limit, the effect on the observed population by applying a given amount of extinction is the same as applying a magnitude cut. Looking at the  $K_S < 19$  mag panel (right-most) in Fig. 1, one can imagine that by applying an extinction of  $A_K = 2$  mag, all sources in a magnitude range from  $K_S = 17$  to 19 mag will be shifted beyond the sensitivity of the survey. Thus, the observed population would be best represented by a magnitude-limited control field. In this particular example, of a sensitivity limit of  $K_S = 19$  mag and an extinction of  $A_K = 2$  mag, the optimal control field would be limited to  $K_S < 17$  mag (the middle panel of Fig. 1). Similarly, for an extinction of  $A_K = 4.5$  mag, the intrinsic colors for the observed population in the science field should be described as given in the left-hand side panel in Fig. 1. In its basic implementation, however, NICER always refers to the same (only sensitivity-limited) intrinsic color distribution, thus not optimally comparing observed populations. In other words, regardless of the amount of extinction, NICER will always compare to intrinsic colors as given in the right-hand side panel of Fig. 1 and therefore compare different populations. We discuss this issue further in Sect. 4.2 where we test NICER with a set of magnitude-limited samples.

As a solution to the problem of increasingly large uncertainties Foster et al. (2008) proposed to use high-resolution NIR observations to discriminate between stars and galaxies based on morphological information. They show that by including separate color excess estimates for galaxies and stars a decrease in the errors of individual pixels in extinction maps can be achieved while at the same time the resolution method (GNICER), however, comes with the handicap that prior knowledge on source characteristics is required, which is often unreliable or not available.

Majewski et al. (2011) implemented a variant of the Lada et al. (1994) NICE technique (Rayleigh-Jeans color excess, RJCE) where the authors use a combination of NIR and mid-infrared bands to minimize the intrinsic color distribution

width for stellar sources. Investigating different stellar populations (dwarfs and giants), the authors find that intrinsic colors for such stars (e.g.,  $H - 4.5 \mu\text{m}$ ) show little degeneracy and can in general be much better constrained than NIR colors. This assumption, also breaks down in the above described case of detecting a significant amount of galaxies. Furthermore, in its basic implementation the method does not combine multiple color-combinations which, in principle, would be trivially achievable using the NICER minimum variance approach.

More recently, Juvela & Montillaud (2016) present an approach to estimate extinction based on discretized intrinsic colors where the estimates are derived with Markov chain Monte Carlo frameworks. For deep NIR observations their method delivers better results than NICER since intrinsic colors are discretized and not described by a single parameter. Their method, however, is extremely computationally expensive and also works best when prior information about the column density distribution is available. The authors also did not investigate the possibility of including more than the three standard NIR bands  $J$ ,  $H$ , and  $K_S$  and only note that such an effort may come at an additional steep increase in computation time. Moreover, the method does not offer the possibility to extend the parameter space beyond photometric colors.

In this manuscript we present a new method, PNICER<sup>1</sup>, to calculate extinction toward point sources. We characterize the extinction with a probability density function (PDF) determined by fitting Gaussian mixture models (GMM) along the extinction vector to extinction-free observations. Subsequently PNICER translates the determined intrinsic parameter PDF into extinction by comparing the distribution to the observed parameters while relying on a defined extinction law. The well-established techniques to construct bias-free column-density maps from these irregularly sampled “pencil-beam” measurements can still be applied when the extinction is discretized via a specific metric (e.g., the expected value or the maximum of this distribution). PNICER is exclusively data-driven, is applicable to any combination of parameter spaces (thus, not restricted to color-space like the methods above), and does not rely on any prior information on column-densities or on synthetic models. In the following discussion we will demonstrate that the method is computationally inexpensive and our implementation is capable of statistically calculating reliable intrinsic colors in multiple dimensions for tens of millions of sources in a matter of seconds. The publicly available PNICER code is purely written in Python and is implemented in such a way as to easily allow adaptation for individual use cases.

To illustrate the PNICER design concept, its algorithms, and for subsequent verification we use data from the Vienna Survey in Orion (Meingast et al. 2016, hereafter referred to as VISION). These data include deep NIR observations of the Orion A molecular cloud, as well as a nearby extinction-free control field. The sources were additionally cross-correlated with the ALLWISE source catalog (Wright et al. 2010; Cutri et al. 2013) to increase the number of available parameters, but was restricted to the first and second WISE bands (hereafter referred to as W1 and W2) at  $3.4 \mu\text{m}$  and  $4.6 \mu\text{m}$ , respectively. For the remainder of this article we use the extinction law as given in Table 1. After describing the essential functionality and design concept of PNICER including all key algorithms in Sect. 3, Sect. 4 will demonstrate the

<sup>1</sup> The  $P$  in PNICER is a reference to the calculated probability densities. We acknowledge that the original name of NICER, NICE revisited, specifically refers to NIR photometry (despite being applicable to any other colors) and that we adopt a similar name only for consistency.

**Table 1.** Extinction law used for the verification and test setup.

Band	$J$	$H$	$K_S$	$W1$	$W2$
$\lambda$ ( $\mu\text{m}$ )	1.25	1.65	2.15	3.37	4.62
$A_J/A_K$	2.5	1.55	1.0	0.74	0.54
Reference	1	1	1	2	2

**References.** (1) Indebetouw et al. (2005); (2) Ascenso, J. (priv. comm.).

improvements and reliability by comparing the new method directly to NICER. In addition we will also briefly discuss the software performance and availability. Section 5 summarizes all key aspects of PNICER. Information on the software structure and its dependencies along with a simplified example are given in Appendix A.

### 3. Method description

One of the main features and strengths of PNICER over other extinction estimators is that the method is easily applicable to any combinations and number of parameters. For instance it is easily possible to combine source colors with apparent brightness information. For this reason we will refer to individual input parameters as features. In practice these features will mostly consist of magnitudes and colors, but in principle PNICER can be used with any parameter as long as the effects of interstellar extinction on the given feature are known.

The main PNICER algorithm can be summarized as follows: for each source for which a line-of-sight extinction is to be calculated an intrinsic feature distribution is derived along the extinction vector in the same feature space of a given extinction-free control field. The data in the control field are fitted with GMMs to construct the PDFs which serve as a probabilistic description of the intrinsic features (e.g., intrinsic colors). Due to the fact that many sources will not have measurements in all available features, PNICER constructs all available combinations and automatically chooses the optimal (minimum variance) result. Put into machine learning terms, PNICER uses the intrinsic feature (e.g., color) distribution from a control field to classify the intrinsic features of an extincted science field. We will now proceed to describe the details of this procedure and each individual processing step. For visual guidance, we provide a detailed three-dimensional example and follow all processing steps for two sources with artificial colors and symmetric errors in Fig. 2 which will be referred to in the following discussion.

#### 3.1. The multidimensional feature space

Our example for this demonstration is limited to the four magnitudes  $J$ ,  $H$ ,  $K_S$ , and  $W1$  and for illustration purposes will further be restricted to only allow the three color features  $J - H$ ,  $H - K_S$ , and  $K_S - W1$  with the corresponding color excesses of 0.95, 0.55, and 0.26 mag normalized to  $A_K$  (see Table 1). We note here that the extinction law is fixed for a single PNICER application and the same law will be applied to all input sources (the extinction law needs to be specified upon runtime). However, in particular in the optical and perhaps also at mid-infrared wavelengths the extinction law is expected to vary with the level of extinction. It is therefore the responsibility of the user to take any such potential variations into account.

The three panels on the left-hand side of Fig. 2 only show the available two-dimensional combinations of the selected color features. To calculate color excesses PNICER, however, also uses

all one-dimensional (univariate) parameter spaces, as well as all available higher dimensional combinations. In the case of our three features a total of seven combinations is available: The univariate parameter spaces ( $J - H$ ), ( $H - K_S$ ), ( $K_S - W1$ ), the two-dimensional combinations ( $J - H$ ,  $H - K_S$ ), ( $J - H$ ,  $K_S - W1$ ), ( $H - K_S$ ,  $K_S - W1$ ) and the three-dimensional case ( $J - H$ ,  $H - K_S$ ,  $K_S - W1$ ). Thus, the left-hand side panels of Fig. 2 can be interpreted as projected views of the three-dimensional combination. If, in this case, one also allows individual passbands as features, a total of 127 combinations would be available within the seven-dimensional feature space. The practical limit of usable dimensions depends on the sampling of the control field data space: more dimensions require a proportionally larger number of sources in the control field to have a statistically well sampled feature space. In our test runs, we successfully evaluated up to nine dimensions (511 combinations).

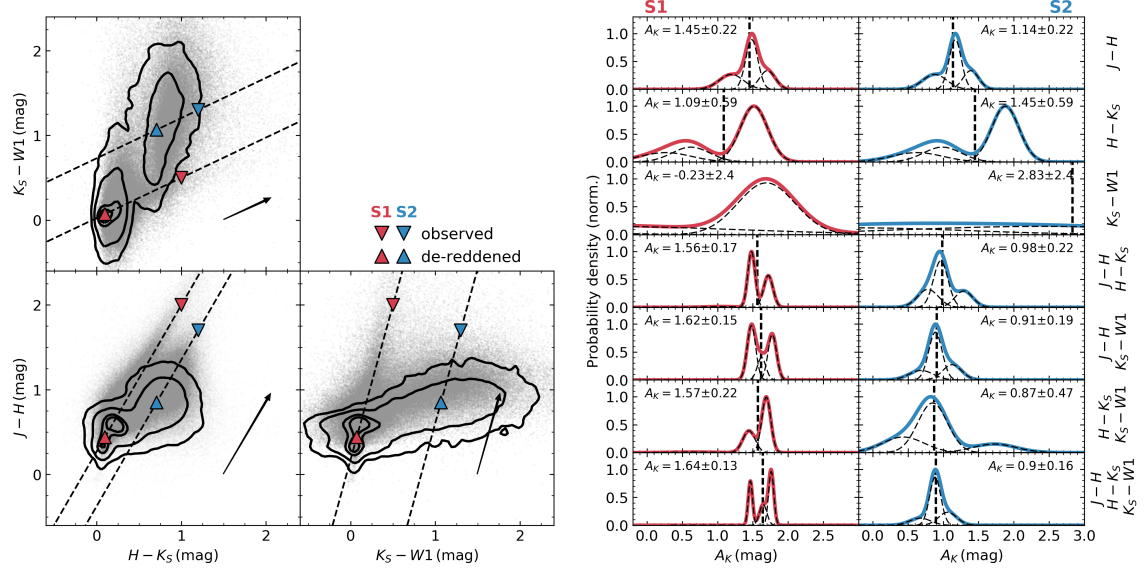
The solid black contours in the left panels of Fig. 2 represent the number density in the control field feature space evaluated with a 0.04 mag wide Epanechnikov kernel, where the levels indicate 0.5%, 3%, 25%, and 50% of the maximum density in the given parameter space. For comparison we also show all sources from the (partly extincted) Orion A VISION data as gray dots in the background. For this demonstration we created artificial sources which we will follow in this example. These are marked in red and blue and are denoted S1 and S2 respectively. Their “observed” colors are  $(J - H)_{S1,S2} = \{2, 1.7\}$ ,  $(H - K_S)_{S1,S2} = \{1, 1.2\}$ , and  $(K_S - W1)_{S1,S2} = \{0.5, 1.3\}$  with a symmetric error of 0.08 mag. These observed colors are marked as triangles with the tip toward the bottom. The black dashed lines are parallels to the extinction vector drawn through the original positions of S1 and S2 and the black arrow corresponds to the effect of 1 mag of extinction in the  $K_S$  band.

Already in this view it becomes apparent that estimates of intrinsic features (as described in the control field feature space) are degenerate along the extinction vector. For many sources the extinction vector will pass through different regions in the intrinsic color distribution. For example when following the extinction vector in the  $(J - H$ ,  $H - K_S)$  space through the observed colors of S1 (red triangle pointing downwards), the vector passes partly through the galaxy colors and then crosses the main sequence for late type stars as well as early type stars. From this feature combination alone it is therefore not entirely clear which intrinsic colors this source should be assigned. For the  $(J - H$ ,  $K_S - W1)$  combination this degeneracy with respect to the galaxy colors seems to be better resolved.

The control field is extinction-free and it is therefore assumed to be an accurate distribution of intrinsic colors (smoothed by the photometric errors). Ideally the data for the control field should have similar completeness limits and comparable errors to accurately reflect the information in the science field. Also, stellar and galactic populations and number densities should be similar in the control field and the science field. These criteria are often met when dealing with data from one set of observations.

#### 3.2. Constructing probability density distributions

To estimate the intrinsic feature probability distribution, PNICER calculates the PDFs along the extinction vector (the dashed lines in the left-hand side panels of Fig. 2) in the control field feature space. Here, the number density of sources is directly interpreted as the probability distribution of intrinsic features. In order to derive this probability density along an arbitrary feature extinction vector in any number of dimensions PNICER initially rotates the



**Fig. 2.** PNICER method concept. *Left:* VISION Orion A data (gray points) and extinction-free control field number density distribution (black solid contours, levels at 0.5%, 3%, 25%, and 50% of the maximum density) for the two-dimensional feature combinations in our test setup. For two test sources S1 (red) and S2 (blue) the extinction is described with the PDF along the reddening vector (black dashed lines) in the control field. *Right:* probability densities functions described by GMMs along the reddening vector for both sources (S1: red, *left column*, S2: blue, *right column*) for all possible combinations of features. Individual combinations are shown in *different rows* and their labels are found at the right-hand side. The individual Gaussian components are also displayed as dashed lines. The annotated extinction estimates refer to the expected value (also marked as vertical dashed lines) of the PDF and its population variance. The PDF with the minimum population variance is chosen as the best approximation for the line-of-sight extinction. In this example the extinction is estimated via the expected value of the PDFs and is marked with dashed vertical black lines.

data space with the given extinction vector until only the first feature component remains non-zero. In other words, in the rotated feature space, the extinction vector has only one non-zero component and is parallel to the first feature axis. We construct the final  $n$ -dimensional rotation matrix via a sequence of applications of

$$V = \hat{u}_1 \otimes \hat{u}_1 + \hat{u}_2 \otimes \hat{u}_2, \quad (3)$$

$$W = \hat{u}_1 \otimes \hat{u}_2 - \hat{u}_2 \otimes \hat{u}_1, \quad (4)$$

$$R = I_n + V [\cos(\alpha) - 1] + W \sin(\alpha), \quad (5)$$

where  $I_n$  is the identity matrix for  $n$  dimensions. Here the rotation matrix  $R$  allows to rotate an  $n$ -dimensional feature space by an angle  $\alpha$  in the plane spanned by the unit vectors  $\hat{u}_1$  and  $\hat{u}_2$  where  $\hat{u}_1 \cdot \hat{u}_2 = 0$  and  $|\hat{u}_1| = |\hat{u}_2| = 1$ . We apply these rotations  $n - 1$  times until only one component remains non-zero. The rotation of the intrinsic feature space allows to directly fit GMMs to the data on a discrete grid since the extinction vector in this space is parallel to the axis spanning the first dimension. The discretization of the grid is typically chosen to oversample the data by a factor of two with respect to the average feature errors. We found that GMMs with typically three components are sufficient to model the density distribution along the extinction vector. The total number of fitted components for the GMM defaults to three, but can be adapted by the user depending on the complexity of the feature space.

It would also be possible to derive the underlying PDF without the assumption that the distribution can be fitted with a limited number of Gaussian functions by directly calculating normalized kernel densities. However, doing so would imply to

define a kernel bandwidth which may artificially broaden the distribution. In our method, all Gaussian functions in the fitted model use independent covariance matrices and thus optimally describe the underlying PDF of the intrinsic feature distribution. Furthermore, by modeling the probability density distributions with GMMs one has to store only a limited number of parameters. This is particularly important when estimating extinction for large ensembles.

### 3.3. Estimating extinction

The process of creating probability density functions is repeated for all possible combinations of features. Using all combinations ensures that always the optimal feature space is selected and even a single feature can provide an extinction measurement<sup>2</sup> for cases where sources do not have measurements in all given features. The final extinction estimate described by a PDF is then chosen from the combination of features which minimizes the population variance. In almost all cases the combination with the largest dimensionality will be selected. Only when an additional feature has significantly larger errors than the other parameters, a reduced feature space may deliver better results. In addition we require at least twenty sources in the control field feature space to be present along the reddening vector, otherwise the extinction estimate would be highly biased by the low number of control field sources and the model fitting process may

<sup>2</sup> We only allow single features in color space, but not in magnitude space as a single magnitude can not be taken as a reliable indicator for extinction.

not converge. This case affects mostly sources with large photometric errors or “uncommon” observed colors such as young stellar objects (YSO) which may not be well represented in the control field feature space. The process of estimating extinction is illustrated in the right-hand side panels of Fig. 2 which show the extracted PDFs (in this case constructed from three independent Gaussian functions) for all seven possible combinations of our test features. These panels also display the calculated extinction when the expected value of the PDF is used (vertical dashed lines; see Sect. 3.3 for details). When using this estimator on just one available feature, PNICER reproduces the NICER results when applied to colors alone. In this case, the expected value of the PDF is equal to the mean of the color distribution. Hence, the data points are projected onto the same intrinsic color. In reference to right-hand side panels in Fig. 2 we note several characteristics here:

- Clearly, some feature combinations are better suitable than others because they show much narrower distributions. For example in the  $J - H$  color alone (topmost sub-plot) all objects share a very similar color, making it ideal for extinction determinations in our case<sup>3</sup>. On the other hand, for example the  $H - K_S$  and  $K_S - W1$  colors only poorly constrain the extinction since the intrinsic feature space shows a very broad distribution. In fact, in this case we can see that all combinations which include a  $J$  magnitude offer superior results.
- For the one-dimensional parameter spaces (top three rows) all colors share the same PDF shape for both sources (for  $K_S - W1$  this is not well visible due to the extreme width of the mixture). We only observe a shift (depending on observed color) in the PDF describing the extinction.
- As expected, the best combination (i.e., smallest variance) is found when all features are available (bottom panels).
- For some combinations we observe a degeneracy in the probability density space. Consider source S1 in the  $(J - H, H - K_S)$  feature space. Moving along the reddening vector we pass the outermost edges of the galaxy locus ( $J - H \sim 1$ ,  $H - K_S \sim 0.4$ ) and then cross both the M sequence dwarf branch ( $J - H \sim 0.6$ ,  $H - K_S \sim 0.2$ ), as well as an enhancement caused by early-type stars ( $J - H \sim 0.3$ ,  $H - K_S \sim 0.1$ ). This is reflected by the asymmetric shape in the density profile. The double-peaked nature of this degeneracy due to the two stellar peaks is also apparent in the  $(J - H, K_S - W1)$  feature space. The expected values for these PDFs place the intrinsic color of the source in the low probability valley between these peaks, demonstrating that such an estimator can be sub-optimal. In this example, even a combination of three colors does not break the degeneracy.
- The source S2 is equally well constrained, though here most distributions (those with  $J$  band) show only one single peak clearly marking the source as a galaxy.
- In the case of estimating the extinction and error with the expected value and population variance of the distribution, we see a continuous improvement when using higher-dimensional feature spaces. This trend is expected to continue when even more features are added.

We again emphasize here that this technique is not limited to color space, but can be applied to any feature so long as its extinction component is known. In fact, optimal results are achieved when combining color and magnitude space since for example, a bright source is unlikely to be a galaxy. Among all

<sup>3</sup> For magnitude-limited samples without galaxies  $H - K_S$  shows a smaller variance compared to  $J - H$ .

available combinations we chose the PDF that shows the smallest distribution width.

We note that the de-reddening process is the same for all sources and the extinction PDFs are all drawn from a given intrinsic color distribution in a control field. Hence, any objects that are not represented in the control field, for example YSOs in star-forming regions, will also use the given intrinsic feature set (e.g., main-sequence stellar colors or galaxy colors) for the extinction estimate. For a correct de-reddening of YSOs it is therefore necessary to use intrinsic features of such sources instead of typical main-sequence or galaxy features. This issue is mitigated to some degree because typical intrinsic NIR YSO colors for classical T Tauri stars (e.g., Meyer et al. 1997) are found to be very similar to intrinsic galaxy colors. Furthermore, for extinction mapping of star-forming molecular clouds, YSOs should be removed beforehand from the input source list as they do not sample the full cloud column-density.

If a single value for the extinction is desired, it is possible to calculate the expected value (maximum probability, or any other meaningful descriptor) of this distribution. The uncertainty in the calculated discrete extinction can be estimated with the population variance of the distribution:

$$\mu_f = \int x f(x) dx \quad (6)$$

$$\text{Var}_f = \int x^2 f(x) dx - \mu^2. \quad (7)$$

Here  $\mu$  refers to the expected value of a probability density function  $f(x)$ . In the case of a one-dimensional Gaussian mixture model, the mean and variance of the mixture can be written in terms of the means, variances, and weights of its components.

$$\mu_{\text{mixture}} = \sum_i \mu_i w_i \quad (8)$$

$$\text{Var}_{\text{mixture}} = \sum_i w_i \sigma_i^2 + \sum_i w_i \mu_i^2 - \left( \sum_i w_i \mu_i \right)^2, \quad (9)$$

where  $\mu_i$  refers to the mean of the  $i$ th Gaussian component, the  $w_i$  are their weights with  $\sum w_i = 1$ , and  $\sigma_i^2$  are the variances.

### 3.4. Creating smooth extinction maps

PNICER derives PDFs for single sources which describe the line-of-sight extinction. To create extinction maps it is therefore necessary to construct the smooth column-density distribution from these irregularly spaced samples. If a single value of the extinction is derived from the PDFs (e.g., the expected value or the maximum probability) one can employ the well-tested approach of the NICER method (Lombardi & Alves 2001). PNICER also comes with built-in fully automatic solutions to create extinction maps with valid world coordinate system projections.

The PNICER software package offers a variety of estimators for the smoothing process including nearest-neighbor, Gaussian, and simple average or median methods. We note here that extinction maps constructed from discrete measurements can suffer from foreground contamination and unresolved cloud substructure which may introduce a bias in the column-density measurement. For more distant clouds foreground stars may represent the majority of detected sources in a pixel of the extinction map. These issues are described in more detail in Lombardi (2009) where the authors also introduce a new technique, NICEST, to minimize this bias. For discretized extinction measurements PNICER optionally also includes this method.



Constructing smooth column-density maps using the full probabilistic description of extinction via GMMs will be the subject of a follow-up study.

#### 4. Method validation

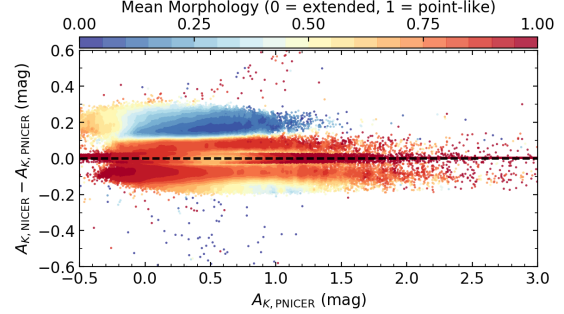
To evaluate the new method we compared results from PNICER to NICER by directly deriving extinction and associated errors via the expected value and the population variance of the PDFs. In this section we first analyze results when both algorithms are applied to the VISION Orion A data. Subsequently we investigate the intrinsic color distribution as measured in the VISION control field and discuss the bias of comparing different observed populations when the science field is extinguished. This part will also highlight the effects of using increasing numbers of parameters. In addition, we also compare wide-field extinction maps calculated with both techniques from 2MASS data. Finally, we also examined the software performance and describe where interested users can access the source code and potentially even contribute to the development. We will not discuss the relation of the estimated errors for both methods, as they are derived in different ways and are therefore not comparable.

##### 4.1. Applying NICER and PNICER to real data

In a first evaluation of the PNICER method, we applied the algorithm together with NICER to the photometric color data of the VISION NIR observations (no magnitudes in parameter space, only  $J, H, K_S$ ). For PNICER we use the expected value of the PDFs and their population variances to directly calculate extinction and errors (see Eqs. (8) and (9)). Taking the expected value of the PDFs makes the methods directly comparable since NICER relies on a similar method (see Sect. 2 for details). The difference in the derived color-excesses is shown in Fig. 3 where the data are color-coded by source morphology. At first glance, the distribution appears to be bimodal and we clearly see that for galaxies (morphology  $\approx 0$ ) on average a much larger difference between the methods is seen when compared to point-like sources. As expected, PNICER delivers a smaller extinction toward galaxies when compared to NICER, because the latter method places the mean intrinsic color between the galaxy and M-sequence locus (compare Fig. 1). At closer examination, however, we observe a more complex structure: there is a distinguished distribution at  $A_{K,NICER} - A_{K,PNICER} = 0$  which is caused by sources having only measurements in two photometric bands. In this case (PNICER has only access to colors) the results of NICER and PNICER are identical. There is also an enhancement of sources toward negative values in  $A_{K,NICER} - A_{K,PNICER}$ . This is caused by sources which are de-reddened beyond the assumed mean of NICER, which in this case, are mostly stellar sources, since, as already mentioned above, the mean of the intrinsic color distribution is found between galaxy and stellar loci.

##### 4.2. The intrinsic color distribution and population bias

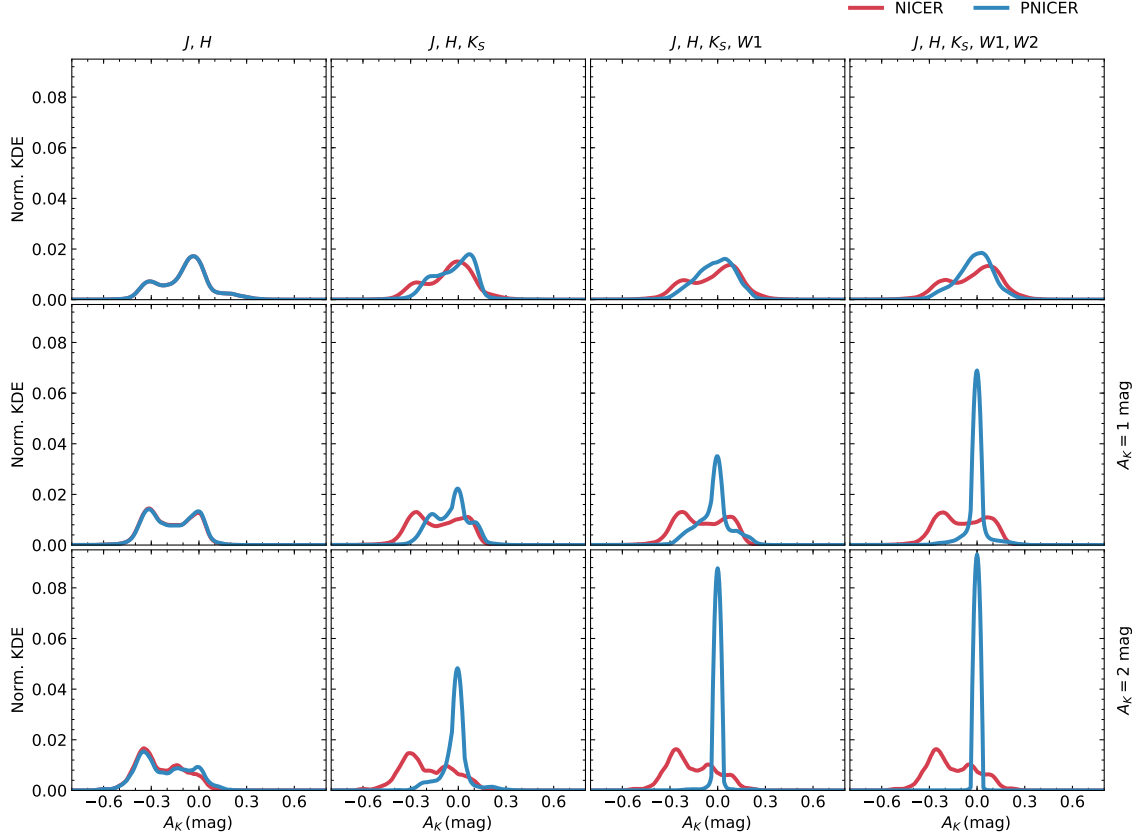
For the PNICER and NICER methods an extinction-free control field typically observed at similar galactic latitudes as that of the science field, by assumption, holds the information for intrinsic features. Therefore, when the techniques are applied to the control field itself, the measured extinctions should ideally be close to 0. We test this hypothesis under a variety of circumstance: we (a) vary the number of available parameters and (b) additionally apply magnitude cuts to the “science” field (for this test



**Fig. 3.** Direct comparison of PNICER and NICER extinction estimates when applied to the NIR VISION observations. The mean morphology was calculated with the SExtractor morphology class of the source catalog in a  $0.05 \times 0.05$  large box in this parameter space. As expected, PNICER delivers smaller extinction estimates for galaxies.

the control field itself) while always using the same (only sensitivity limited) control field data. The idea behind point (b) is to simulate the effect of extinction. While in an extinguished field intrinsically faint sources will be shifted beyond the photometric sensitivity limit (i.e., fewer faint stars and galaxies will be observed), the control field does not suffer from these effects. Furthermore, to increase the dimensionality for these tests we combine the NIR VISION control field photometry with the first two bands of the ALLWISE source catalog. NICER, as usual, is restricted to color information, but to highlight additional differences we allow PNICER to construct the multidimensional feature space from both color and magnitude information simultaneously. Thus, NICER has access to up to four dimensions (colors only), while PNICER has access to nine dimensions (five magnitudes and four colors) at most.

Figure 4 displays the results when applying both the NICER and PNICER algorithms to the VISION control field itself. All panels show kernel densities (histograms, bandwidth = 0.04 mag) for the distributions of the derived extinction (for PNICER again the expected value of the PDF), where the blue lines refer to the expected value of the PDF, where the blue lines refer to PNICER results and the red lines to NICER. The separate columns in the figure refer to different parameter combinations with increasing dimensionality from left to right. In the first column, the analysis is restricted to  $J$  and  $H$  only, while in the last column we show the results when using  $J, H, K_S, W1$ , and  $W2$  photometry. The different rows in this plot matrix refer to different magnitude cuts for the science field (the control field remains untouched). Applying magnitude cuts to the data simulates the effects of extinction on the observed population. Consider an extinction of  $A_K = 1$  mag and our sensitivity limit at  $K_S = 19$  mag. In this case all sources which have intrinsic apparent magnitudes between  $K_S = 18$  and  $19$  mag will shift beyond the detection limit, creating a different observed population with a different feature (e.g., color) distribution. Therefore, limiting our data to  $K_S < 18$  mag (and the other bands according to Table 1) has the same effect as 1 mag of extinction in this band. In Fig. 4 the first row does not apply magnitude cuts, while the second and third row simulate the effects of having an extinction of  $A_K = 1$  and  $2$  mag, respectively. We note here that the calculated extinction for individual sources can become negative when the estimated intrinsic color is “redder” than the observed color for the investigated source. This is possible for unextinguished sources and especially affects



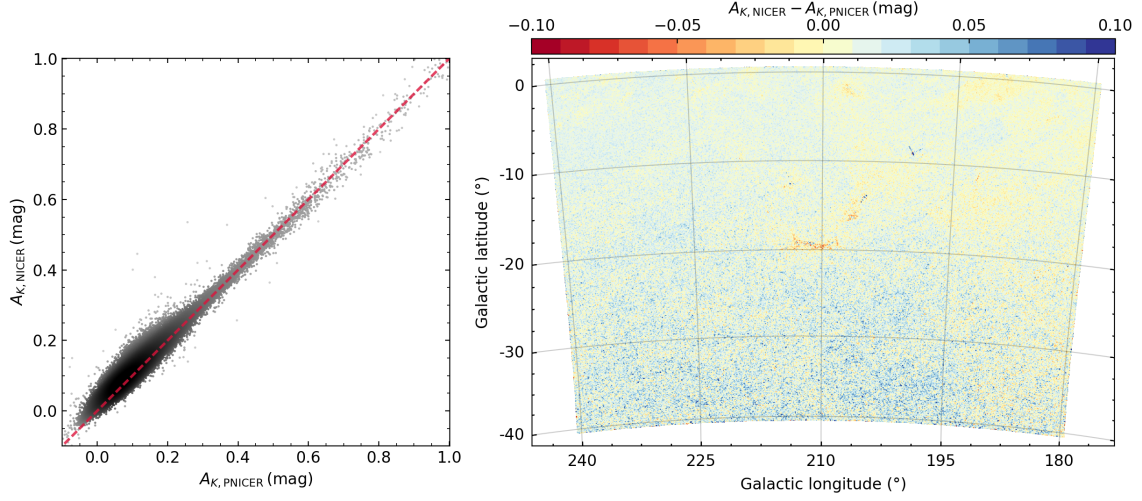
**Fig. 4.** Kernel densities for the derived line-of-sight extinction values for both PNICER (blue) and NICER (red) when applying the algorithms to the VISION control field itself. In this case and since the entire control field is extinction-free, the calculated values should be close to zero. *From left to right* the number of features is incrementally increased, while *from top to bottom* we apply magnitude cuts to the “science” field. The magnitude cuts simulate the effects of having an extinguished science field, resulting in different populations when compared to the extinction-free control field (e.g., galaxies will be shifted beyond the survey sensitivity limit). In the *first row* we use all data, while the *second and third rows* simulate the effect of 1 and 2 mag extinction in the  $K_S$  band (i.e.,  $K_S < 18$  mag and  $< 17$  mag, respectively for our completeness of  $K_S = 19$  mag). Clearly, PNICER performs better for increasingly different populations and number of available features.

early-type stars. As a consequence, the kernel densities extend both to the positive and negative side in this analysis.

For only two features ( $J$  and  $H$ ; leftmost column) we obtain very similar results for both methods across all magnitude cuts. However, for increasingly strict magnitude cuts (more extinction) we can see that the peak in the distribution is shifted toward negative values. This is expected, since in this scenario only stellar sources remain in the science field while the mean color of the full control field is calculated from data including galaxies. In this case both methods are biased since there is not enough information to break degeneracies in the intrinsic feature space. The resulting extinction distributions are almost identical overall, but as it also allows magnitude information, the PNICER extinction distribution can appear slightly different. Here we note again that when considering only one color without magnitude information the results from PNICER and NICER are identical.

When increasing the number of available features to three ( $J, H, K_S$ , second column) we already observe significant differences. While NICER still suffers from the bias of different populations in the science and control field (the peak is again

shifted to negative  $A_K$ ), PNICER starts to perform systematically better. For a simulated extinction of  $A_K = 1$  and 2 mag (i.e.,  $K_S < 18$  and 17 mag) the increased dimensionality helps to break degeneracies in the intrinsic feature space and the extinction distribution shows a prominent peak at  $A_K = 0$  mag. This effect is even more pronounced when including four or five parameters (third and fourth column) where PNICER overall shows systematically better results than NICER. Especially in the case of largely different populations in the science and control field PNICER manages to overcome this issue and delivers far better results. We note here that for the case of similar observed populations (top row), PNICER is only marginally better than NICER (~20–30% narrower distribution width when using five features), but we expect that the remaining degeneracy in intrinsic colors could be better lifted by including one or two more suitable passbands (e.g.,  $Y$  at  $1 \mu\text{m}$ ). Including bands toward optical wavelengths would particularly help in this case, since here the stellar sequences are typically more pronounced than in the NIR allowing PNICER to better separate object types. The practical limit in the number of features here depends on the sensitivity



**Fig. 5.** Comparison of PNICER and NICER when applied to 2MASS data. *Both panels* shows a pixel-by-pixel comparison of extinction maps (5 arcmin resolution) calculated with both methods from the same data. As expected, within the statistical error the results are identical. Nevertheless we observe a small systematic shift toward larger NICER extinctions in low column-density regions. The relatively large deviation toward the center of the map is attributed to the Orion A and B molecular clouds where extinction estimates have even larger variance.

of the observations in the given passbands (higher extinction toward bluer bands) and the sampling of the control field feature space (more dimensions require more sources for accurate sampling). We conclude that NICER is biased in cases where the populations in the science and control field are different, that is, in regions with extinction. PNICER on the other hand starts to break the degeneracy in intrinsic colors when having access to more than one color-feature and performs even better when including more parameters.

#### 4.3. Extinction maps

We also validated the new method’s functionality by comparing wide-field extinction maps created with NICER and PNICER based on 2MASS data only. As in the tests above, we calculated the extinction toward each source as the expected value of the associated PDF for PNICER. Since the data are restricted to the three NIR bands  $J$ ,  $H$ , and  $K_S$  and only include a negligible number of galaxies (if any) the color distributions are relatively narrow and we expect very similar results without considerable improvement in the quality of the extinction map. Nevertheless, this test should demonstrate that under these simple circumstances PNICER works equally well as NICER.

For this purpose we created extinction maps with a resolution of 5 arcmin for approximately the same region as the maps in Lombardi et al. (2011) who studied a  $\sim 40 \times 40 \text{ deg}^2$  field including the Orion, Monoceros R2, Rosette, and Canis Major star forming regions. The control field in this case was chosen as a  $2 \times 2 \text{ deg}^2$  wide sub-region centered on  $l = 233.3$ ,  $b = -19.4$ , the position of the VISION control field. We note here that for such a large region it would be more appropriate to use multiple control fields located at different galactic latitudes to accurately sample the galactic stellar population. This application, however, only serves as a demonstrator and to verify the method. Therefore variations in the field population can safely be neglected.

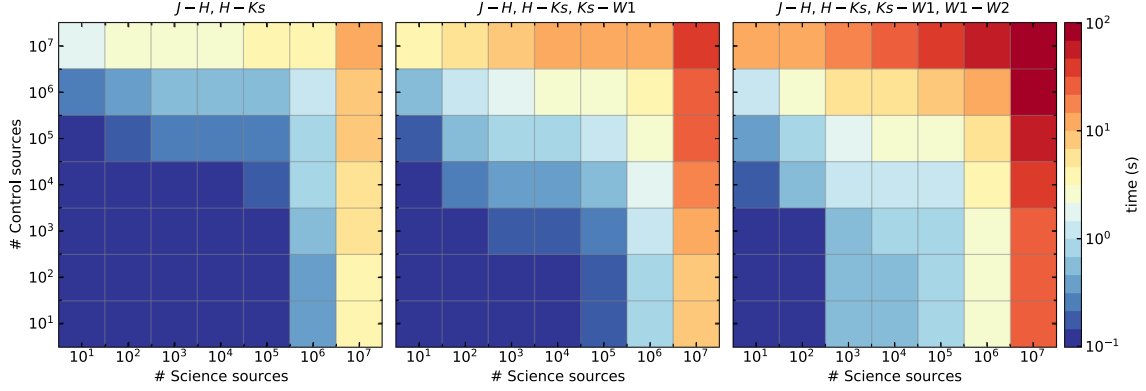
The results of this test are displayed in Fig. 5 where we show pixel-by-pixel comparisons for both maps. The results can be

considered equal within the noise properties of the data, but we also see a small systematic trend of NICER giving slightly larger extinction values for low column-density regions when compared to PNICER. The Orion A molecular cloud (approximately located at the center of the map) is also well visible in this comparison in the right-hand side panel of Fig. 5. In this case we do not interpret this behavior as systematic differences between the methods since the noise levels also significantly increase in this region due to high column-densities and fewer available background sources. This may be caused to some extent by the above discussed population bias. Hence, we conclude that PNICER and NICER work equally well for cases where only the typical NIR bands are considered and galaxies do not contaminate the intrinsic color space.

#### 4.4. Performance

We evaluated the performance of our PNICER implementation by generating several test scenarios with variable numbers of science and control field sources. All performance tests have been conducted on a machine with a 4 GHz CPU (Intel Core™i7-6700 K) with 16 GB RAM. For all tests we used the VISION control field for the intrinsic feature distribution and applied randomly generated extinction to the sources (according to Table 1) to simulate extincted science fields. From this data pool we then randomly drew sets of variable size by (a) varying the number of sources in the science field; (b) varying the number of control field sources; and (c) increasing the number of available features.

The results of these performance tests are visualized in Fig. 6. The individual panels represent different numbers of features (two, three, and four from left to right). Each matrix element in the panels shows the results of a single PNICER run. For example the bottom right elements show the run time results of using  $10^7$  science field sources with a control field that



**Fig. 6.** PNICER performance results. For the tests displayed in the *left-most panel* two features were used ( $J - H$ ,  $H - K_s$ ), for the *panel in the center* three features, and for the *right-most results* four features. The colors of each matrix element display the code execution time for a given setup where the number of sources in the science field and the number of sources in the control field are incrementally increased. While for typical applications for a few thousand to a few ten thousand sources the de-reddening process finishes in a fraction of a second, only for extreme (and probably rare) use cases ( $10^7$  sources) we measure execution times longer than a minute.

only contains ten sources<sup>4</sup>. The most extreme case in the top right uses  $10^7$  sources in both the science and control fields. For small sample sizes ( $\leq 10^4$ ), runtimes are mostly found well below 1 s across all feature combinations. For a VISION-type use case ( $10^6$  science sources,  $10^5$  control field sources and using only two colors) the execution time of PNICER is 0.8 s. Only for very extreme cases of  $10^7$  sources and more than two features, the de-reddening process requires more than a minute. Extrapolating these results to cases using a combination of more than four features, we recommend to keep the sample size below  $10^6$ . With these results we therefore conclude that the PNICER performance is highly competitive and that the application of the method also remains practical in extreme cases.

#### 4.5. Software availability

PNICER is an open-source software and is accessible to all interested users. The Python package and source code are available at <https://github.com/smeingast/PNICER> where also the latest versions will be made available. Future upgrades may include advanced treatment of photometric errors, support for additional metrics beyond photometric measurements, and weighted fitting of the Gaussian mixture models as soon as the necessary libraries are updated or become available.

## 5. Summary

We have presented a new method, PNICER, to derive extinction probability density functions for single sources in arbitrary numbers of dimensions. Our findings can be summarized as follows.

1. The well-established NICER method to calculate line-of-sight extinctions suffers from increasing variance in intrinsic color estimations for deep NIR observations when galaxies enter the color space. As a consequence, the color excess estimates have large statistical errors. For details see Fig. 1. Other methods do not implement satisfying solutions for

this problem since they either rely on additional information (e.g., morphology) or are restricted to few dimensions and are computationally expensive.

2. We introduce a new method, PNICER, which uses unsupervised machine learning tools to calculate extinction toward single sources. To this end we fit the intrinsic feature distribution from an extinction-free control field with Gaussian mixture models. The resulting probability density distribution describes the probability of intrinsic features (e.g., colors) and therefore also extinction. From these distributions the extinction can be estimated with the expected value (or maximum probability), its uncertainty with the PDF variance. Details on the method are visualized in Fig. 2.
3. PNICER is entirely data-driven and does not require prior information of source characteristics or the column-density distribution. The intrinsic feature probabilities are automatically constructed from the control field data with unsupervised algorithms.
4. PNICER works in arbitrary numbers of dimensions and features (e.g., magnitudes or colors) can be combined in any way as long as the corresponding extinction law is known.
5. We investigated the effects of the intrinsic color distribution and compared the PNICER and NICER performance when the observed populations in the extincted science and the extinction-free control field are different (i.e., in regions of significant extinction). We find that NICER is biased when different populations are observed and that PNICER performs significantly better in these cases. To break degeneracies in the intrinsic feature space with PNICER, more than one parameter is required (e.g., two colors). For details see Fig. 4.
6. Using 2MASS data (three NIR bands, no galaxies) PNICER reproduces the NICER extinction mapping results within the statistical errors (Fig. 5).
7. The PNICER software is entirely written in Python and is publicly available at <https://github.com/smeingast/PNICER>. It includes simple interfaces to apply either the PNICER or the NICER method to real data and subsequently construct extinction maps. Furthermore, the algorithm remains computationally competitive for large ensembles, calculating extinction PDFs for millions of sources in a matter of seconds.

<sup>4</sup> The minimum number of 20 sources along the reddening vector has been disabled for this test.



*Acknowledgements.* Stefan Meingast is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute for Astrophysics, University of Vienna. We thank Kai Polsterer for the helpful discussions on Machine Learning and his valuable input regarding the methods presented in this publication. We also thank the anonymous referee for useful comments that helped to improve the quality of this publication. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013).

## References

- Alves, J., Lada, C. J., Lada, E. A., Kenyon, S. J., & Phelps, R. 1998, *ApJ*, **506**, 292
- Alves, J., Lombardi, M., & Lada, C. J. 2014, *A&A*, **565**, A18
- Alves, J. F., Lada, C. J., & Lada, E. A. 2001, *Nature*, **409**, 159
- Ascenso, J., Lada, C. J., Alves, J., Román-Zúñiga, C. G., & Lombardi, M. 2013, *A&A*, **549**, A135
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bok, B. J., & Cordwell, C. S. 1973, in *Molecules in the Galactic Environment*, eds. M. A. Gordon, & L. E. Snyder (New York, NY: John Wiley and Sons), 54
- Cambrésy, L. 1999, *A&A*, **345**, 965
- Cutri, R. M., et al. 2013, *VizieR Online Data Catalog: II/328*
- Dobashi, K. 2011, *PASJ*, **63**, S1
- Dobashi, K., Uehara, H., Kandori, R., et al. 2005, *PASJ*, **57**, S1
- Dobashi, K., Marshall, D. J., Shimoikura, T., & Bernard, J.-P. 2013, *PASJ*, **65**, 31
- Flaherty, K. M., Pipher, J. L., Megeath, S. T., et al. 2007, *ApJ*, **663**, 1069
- Foster, J. B., Román-Zúñiga, C. G., Goodman, A. A., Lada, E. A., & Alves, J. 2008, *ApJ*, **674**, 831
- Goodman, A. A., Pineda, J. E., & Schnee, S. L. 2009, *ApJ*, **692**, 91
- Hunter, J. D. 2007, *Comp. Sci. Eng.*, **9**, 90
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, *ApJ*, **619**, 931
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python* [Online; accessed 2017-02-16]
- Juvela, M., & Montillaud, J. 2016, *A&A*, **585**, A78
- Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, *ApJ*, **429**, 694
- Lombardi, M. 2009, *A&A*, **493**, 735
- Lombardi, M., & Alves, J. 2001, *A&A*, **377**, 1023
- Lombardi, M., Alves, J., & Lada, C. J. 2006, *A&A*, **454**, 781
- Lombardi, M., Lada, C. J., & Alves, J. 2008, *A&A*, **489**, 143
- Lombardi, M., Alves, J., & Lada, C. J. 2011, *A&A*, **535**, A16
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, **739**, 25
- Meingast, S., Alves, J., Mardones, D., et al. 2016, *A&A*, **587**, A153
- Meyer, M. R., Calvet, N., & Hillenbrand, L. A. 1997, *AJ*, **114**, 288
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, *ApJ*, **679**, 481
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comp. Sci. Eng.*, **13**, 22
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868

## Appendix A: Software dependencies, structure, and availability

One of the main method and software design goals for PNICER was to make it accessible and usable for as many people as possible. For this reason we have kept the number of required dependencies at a minimum and we additionally set a high value for the computational performance. The entire software is written in Python<sup>5</sup> and its main dependencies are NumPy (Van Der Walt et al. 2011) and SciPy (Jones et al. 2001) for numerical calculations, Astropy (Astropy Collaboration et al. 2013) for I/O and world coordinate system support, and scikit-learn (Pedregosa et al. 2011) for machine learning tools. In addition, the plotting methods make use of matplotlib (Hunter 2007) and the Astropy affiliated wcsaxes package<sup>6</sup>. All of these packages are easily accessible through the Python Package Index<sup>7</sup>.

For high performance, some functions in PNICER have been parallelized to offer even better results on modern machines. The Python parallelization oftentimes works with straight-forward code-blocks in Python, however, the multiprocessing library of our choice is (at the time of writing this manuscript) not compatible with Unix-based operating systems and Windows. Therefore, at the moment, PNICER can only be used on Unix machines. All essential tests have been successfully performed under macOS 10.12 and Ubuntu 14.04 LTS and we do not foresee any major compatibility issues with future operating system or Python versions. Using our software implementation and its functions only requires to instantiate for example photometric data as *ApparentColors* or *ApparentMagnitudes* objects. Once the instance is created running PNICER or NICER only requires a single line of code. Subsequently running the

discretization and creating an extinction map from the extinction estimates also only require an additional single line of code. All customization and options for running the software are implemented as keyword arguments in the PNICER or NICER call.

Future versions may offer enhanced construction of extinction maps and support for additional feature metrics. In its current form the software only allows photometric data to be instantiated, but in principle any metric can be used and combined with other parameters as long as the extinction can be described in the same way as for magnitudes or colors. Running the algorithms returns *ContinuousExtinction* objects from which discretized *DiscreteExtinction* objects and subsequently extinction maps can be created. In its simplest form a typical PNICER session may look like the following example.

```
# Import PNICER class for apparent colors
from pnicer import ApparentColors as AC

# Instantiate Color objects from observed data
sci = AC(colors=mag_sci, errors=err_sci, ...)
con = AC(colors=mag_con, errors=err_con, ...)

# Calculate extinction PDFs
pdf = sci.pnicer(control=con)

# Discretize PDFs
extinction = pdf.discretize()

# Construct the extinction map
emap = extinction.build_map(bandwidth=0.1)

# Save the map with WCS projection
emap.save_fits(path="/path/to/map.fits")
```

<sup>5</sup> <https://www.python.org>

<sup>6</sup> <https://github.com/astrofrog/wcsaxes>

<sup>7</sup> <https://pypi.python.org/pypi>

# Infrared Extinction in Orion A

## 4.1 Overview

This chapter is devoted to the investigation of infrared dust extinction in Orion A. The work presented here is based on the Orion A NIR photometric database from Chapter 2 and the new PNICER extinction tool introduced in Chapter 3. In contrast to the previous manuscripts, this paper is currently undergoing the reviewing process and is not yet accepted for publication.

I start with a literature review of the current state of research with respect to infrared extinction. Here, I specifically highlight the recent findings of flat MIR extinction toward many different sightlines and also discuss the common assumption of a universal NIR extinction law. Specifically, I stress some contradicting findings on these topics in the literature and argue that many interpretations are difficult to compare due to sample inhomogeneity. In contrast to many other studies in the literature, my work is based on a single isolated molecular cloud, where biases are expected to be minimal (e.g. due to physically separate, but overlapping regions along the line of sight when observing toward the galactic center).

I continue to create a large photometric database, covering wavelengths from  $\sim 1$  to  $25\ \mu\text{m}$ , for which the NIR Orion A VISTA observations constitutes the basis. Together with refined statistical methods, this database serves as the foundation to investigate the infrared extinction law in the cloud. This is done by fitting color-color diagrams with a linear relation which, following the definition of the color excess, can be used as a proxy for the extinction law. For this purpose, I introduce a simplified, but effective framework to derive reliable linear fits with meaningful errors. Reasonable error estimates are necessary since any variations are expected to be only at the percent-level.

Using this framework, I determine an average extinction law for the entire cloud and compare it to other regions and predictions of several dust grain models. Furthermore, using the large number of detected sources from the VISTA survey, it becomes possible, for the first time, to determine if the local environment impacts the dust grain composition. Here, I find a clear signature that regions, which seem to be affected by radiative feedback from the massive cluster stars, show a different

extinction law, when compared to more quiet parts of the cloud. Furthermore, I compare these findings to model predictions to set constraints on the origin of the variable extinction law. However, even the most simple explanation of having different grain sizes (e.g. the radiation from the massive cluster stars could prevent grain growth) can not be confirmed by this comparison.

Finally, using the average extinction law for Orion A, I present a new high resolution extinction map. This new map is then compared to dust optical depth measurements, where I find a clear bias in previous efforts to determine the gas mass distribution from dust emission measurements.

## 4.2 Publication details

**Title:** VISION - Vienna survey in Orion. II. Infrared extinction in Orion A

**Authors:** Stefan Meingast, João Alves, Marco Lombardi

**Status:** Submitted to in Astronomy & Astrophysics, June 19th, 2017

**Own contributions:** Literature research, catalog combination, development and implementation of methods, derivation of extinction law and its variations, interpretation of results, extinction map, *Herschel* calibration and comparison, preparation of figures and plots, paper writing.



# VISION - Vienna Survey in Orion

## II. Infrared extinction in Orion A

Stefan Meingast<sup>1</sup>, João Alves<sup>1</sup>, and Marco Lombardi<sup>2</sup>

<sup>1</sup> Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, 1180 Wien, Austria

<sup>2</sup> University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

Received 19 June 2017 / Accepted ...

### ABSTRACT

We investigate the shape of the extinction curve in the infrared up to  $\sim 25 \mu\text{m}$  for the Orion A star-forming complex. The basis of this work is near-infrared data acquired with the Visual and Infrared Survey Telescope for Astronomy, in combination with Pan-STARRS and mid-infrared *Spitzer* photometry. We obtain color excess ratios for eight passbands by fitting a series of color-color diagrams. The fits are performed using Markov chain Monte Carlo methods, together with a linear model under a Bayesian formalism. The resulting color excess ratios are directly interpreted as a measure of the extinction law. We show that the Orion A molecular cloud is characterized by flat mid-infrared extinction, similar to many other recently studied sightlines. Moreover, we find statistically significant evidence that the extinction law from  $\sim 1 \mu\text{m}$  to at least  $\sim 6 \mu\text{m}$  varies across the cloud. In particular, we find a gradient along galactic longitude, where regions near the Orion Nebula Cluster show a different extinction law compared to L1641 and L1647, the low-mass star-forming sites in the cloud complex. These variations are of the order of only 3% and are most likely caused by the influence of the massive stars on their surrounding medium. While the observed general trends in our measurements are in agreement with model predictions, both well-established and new dust grain models are not able to fully reproduce our infrared extinction curve. We also present a new extinction map featuring a resolution of  $1''$  and revisit the correlation between extinction and dust optical depth. This analysis shows that cloud substructure, which is not sampled by background sources, affects the conversion factor between these two measures. In conclusion, we argue that specific characteristics of the infrared extinction law are still not well understood, but Orion A can serve as an unbiased template for future studies.

**Key words.** ISM: clouds - ISM: structure - dust, extinction - Methods: data analysis - Methods: statistical

### 1. Introduction

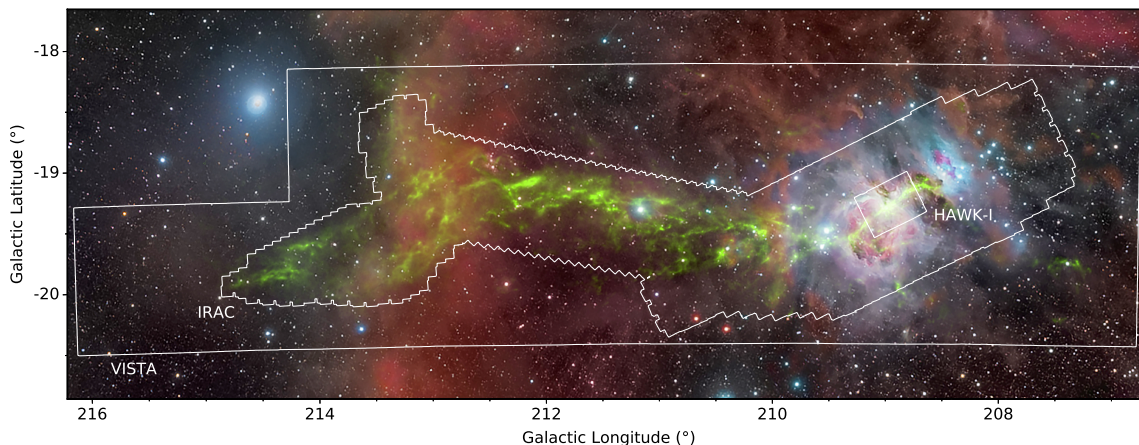
Extinction is the attenuation of electromagnetic radiation by gas and dust. The term extinction combines the effects of absorption and scattering processes which lead to a loss in the measured intensity of a light-emitting source. The strength of this effect is generally much stronger for shorter wavelengths, often obscuring sightlines associated with large amounts of dust. Including the fact that much of the absorbed energy is re-radiated thermally in the infrared, it is generally understood that these processes play a fundamental role in the determination of the radiation field of galaxies at both small and large scales. It therefore has long been recognized that measuring and understanding extinction is crucial for a large variety of astrophysical problems (e.g. Lada et al. 1994). At the smallest scales, measuring extinction allows to constrain dust properties, such as grain sizes and element abundances (e.g. based on models from Weingartner & Draine 2001; Draine 2003). In star-forming regions, dust extinction shields pre-stellar cores from the surrounding UV radiation. On galactic scales, dust converts a large fraction of the absorbed starlight into thermal radiation (Bernstein et al. 2002) and at the largest scales the extra-galactic distance ladder critically depends on our understanding of the extinction law (e.g. Nataf 2015).

The shape of the extinction curve (i.e. how extinction varies with wavelength) has been the topic of many studies in the last few decades. Almost 30 years ago, Cardelli et al. (1989) provided the framework for the characterization of extinction curves

based on the fundamental work by Rieke & Lebofsky (1985), Fitzpatrick & Massa (1986), and Fitzpatrick & Massa (1988). They introduced a parameter family to describe the shape of the extinction curve and furthermore showed that an approximation is possible by using just one quantity:  $R_V \equiv A_V/E_{B-V}$ , with  $A_V$  as the extinction in V band and  $E_{B-V}$  as the color excess in B-V. This parameter is nowadays widely used to describe extinction characteristics in general, and in particular also to probe variations in the extinction law across large spatial scales. To this day, several key characteristics of the extinction curve remain a controversial topic. For example, at UV to optical wavelengths, the extinction curve is relatively featureless with the exception of the prominent bump at  $2175 \text{ \AA}$  of which the origin still remains under debate. (e.g. Fitzpatrick & Massa 1986; Mathis 1994; Bradley et al. 2005).

While in the UV and optical spectral range the extinction law is generally found to vary by a large degree for different sightlines, the near-infrared (NIR) extinction law is widely believed to show little to no variation at all. Even though there are studies reporting different extinction parameters for the three NIR bands J, H, and K<sub>S</sub>, conclusive evidence of variations depending on dust characteristics is still missing. Numerous examples on the determination of the NIR extinction law can be found in the literature. Most of these are based on the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006) where some results indicate variability of the law (e.g. Nishiyama et al. 2006; Zasowski et al. 2009; Gosling et al. 2009; Wang et al. 2013; Nataf et al.

Article number, page 1 of 20



**Fig. 1.** The region under investigation in this work: the Orion A molecular cloud is visualized in green with the dust optical depth map from Lombardi et al. (2014b) blended with optical image data (deepskycolors.com; © Roberto Bernal Andreo). The solid lines indicate the coverage of the various surveys used to investigate the properties of dust extinction in the cloud, where the IRAC contour outlines the region covered by either of the first two IRAC channels. We also make use of MIPS data, for which the covered region generally encompasses the IRAC contour.

[toggle labels](#)

2016), while others do not (e.g. Stead & Hoare 2009; Majaess et al. 2016).

Investigating an often-used power-law parametrization in the NIR ( $A_\lambda \propto \lambda^{-\gamma}$ ), Stead & Hoare (2009) further complicate the situation by claiming, that the value of the power law index is critically dependent on the choice of the filter wavelength. They also continue to show that previously found variations can be explained by systematic errors. Only recently, Wang & Jiang (2014) used APOGEE (Eisenstein et al. 2011) data to select K-type giants scattered across the galactic plane and presented convincing evidence that the NIR extinction law does not vary with color excess. As a result, and in contrast to their previous work (Wang et al. 2013), this result indicates that the NIR extinction law is constant from diffuse to dense environments.

In the mid-infrared (MIR), this picture drastically changes. Here, many different absorption features - the most prominent ones at  $\sim 9.7$  and  $\sim 18 \mu\text{m}$  are caused by silicates - depend on environmental conditions. This has led to a variety of results, where some sightlines are well explained by the standard silicate-graphite model (Draine & Lee 1984; Draine 1989; Weingartner & Draine 2001; Draine 2003), while other regions (including the galactic center) show a particularly flat distribution between 4 and  $8 \mu\text{m}$ . The origin of this flat extinction law between the NIR bands and silicate features is still not well understood (e.g. Wang et al. 2013, 2015; Voshchinnikov et al. 2017, and references therein).

By far the most popular targets for extinction studies in the NIR and MIR spectrum are found in the galactic plane, and in particular also near the galactic center. Sightlines in these directions naturally provide many advantages when studying extinction in the infrared. Not only are there orders of magnitudes more sources compared to regions far away from the plane, but also distinct stellar populations can aid in the determination of the extinction law. Moreover, in recent year, the *Spitzer Space Telescope* (Werner et al. 2004), and in particular the GLIMPSE

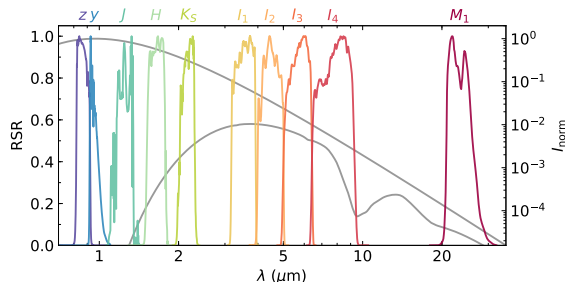
program (Benjamin et al. 2003; Churchwell et al. 2009), have provided a large MIR database of these regions.

There are many examples of studies inferring the MIR extinction law from *Spitzer* data. Most prominently, Indebetouw et al. (2005) used the well-known intrinsic magnitudes of red clump giants in the galactic plane for two different sightlines to determine the absolute extinction law for the NIR  $JHK_s$  and the *Spitzer*/IRAC passbands. Furthermore, the regions near the galactic center have been particularly popular to investigate the origins of the often-observed flat MIR extinction law (e.g. Lutz et al. 1996; Lutz 1999; Nishiyama et al. 2009; Fritz et al. 2011). Variations in the MIR extinction law have been detected in the Coalsack Nebula ( $l \sim 301$ ,  $b \sim -1$ ) by Wang et al. (2013) in the sense of having more pronounced extinction in diffuse regions compared to dense environments. In contrast, Asencio et al. (2013) studied the dense cores B59 and FeSt 1-457 in the Pipe Nebula and do not find statistically significant evidence for a dependency of the MIR extinction law on local gas density.

In light of these often-contradicting findings, it is rather challenging to draw final conclusions about the infrared extinction law and the origin of potential variability. Most of the above mentioned studies concentrate on regions in the galactic plane and use one or another combination of photometric passbands to determine color excess ratios as a proxy for the extinction law. In the galactic plane, however, physical conditions of the interstellar environment are rarely well-known and measured color excesses can be influenced by multiple, physically separate, stacked regions along the line-of-sight. To reach a full understanding of the extinction law and its characteristics in the infrared, it is therefore crucial to provide observational data for regions where environmental conditions are well-known and can be factored into the result.

In the spirit of this series of papers, we will base our investigation on the Orion A giant molecular cloud (Fig. 1; for an overview see the first paper in this series and references therein: Meingast et al. 2016, hereafter Paper I). Orion A is located at a distance of 414 pc (Menten et al. 2007) toward the direction of the galactic anti-center, and far below the plane. Since the region

<sup>1</sup> This manuscript contains some figures with the option to switch between different layers. The implementation uses JavaScript and therefore only works within Adobe® Acrobat®.



**Fig. 2.** The passbands used in our analysis cover the infrared wavelength range from just below  $1\,\mu\text{m}$  to about  $30\,\mu\text{m}$  utilizing the Pan-STARRS ( $z$ ,  $y$ ), VISTA ( $J$ ,  $H$ ,  $K_S$ ; calibrated against 2MASS), and *Spitzer* ( $I_1$ ,  $I_2$ ,  $I_3$ ,  $I_4$ ,  $M_1$ ) photometric systems. The filter curves have each been normalized to their maximum transmission and their relative spectral response (RSR) is shown as a function of wavelength. The two solid gray lines are 3000 K black bodies with a normalized intensity ( $I_{\text{norm}}$ ). The upper line is the unmodified black body, whereas the line at the bottom shows the resulting spectral energy distribution with 5 mag extinction in the  $K_S$  band and using the  $R_V = 3.1$  Draine (2003) extinction law.

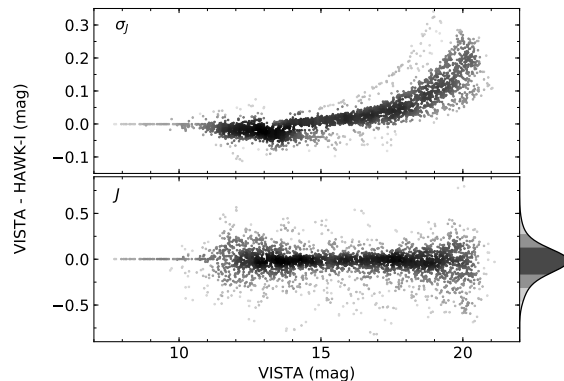
harbors many isolated star-forming events, as well as the massive young Orion Nebula Cluster (ONC), it provides a unique opportunity to study different physical conditions in a self-consistent way. Moreover, the cloud is relatively well isolated from other star-forming regions in the vicinity. This isolation assures that any characteristics with respect to extinction are only originating from the gas and dust associated with one single molecular cloud.

In this manuscript we will investigate the extinction law in the infrared from just below  $1\,\mu\text{m}$  up to about  $25\,\mu\text{m}$ . Based on statistically significant variations of the color excess ratios in several passbands, we will show that the extinction law in the infrared varies with environmental conditions. In particular we will show that the most likely explanation for these variations is the influence of the massive stars in the ONC on the surrounding interstellar medium. Their intense radiative feedback seems to impact the general dust composition and changes the overall extinction law for both NIR and MIR passbands. Comparing our findings to model predictions, we furthermore conclude, that popular dust models do not sufficiently explain the global infrared extinction law in Orion A. Based on these findings we will construct a new extinction map of the region and demonstrate that cross-calibration with *Herschel* dust optical depth data can be biased by cloud substructure that is not sampled by background sources.

To arrive at these conclusions, we first introduce our data sample in Sect. 2 before continuing to give an in-depth overview of the applied methods in this manuscript (Sect. 3). Following this technical overview, we separate our results into two major parts. We first investigate the characteristics of the infrared extinction law in Sect. 4, followed by the results with respect to our new extinction map in Sect. 5. Finally, we summarize our findings in Sect. 6.

## 2. Data

This section will start with an overview of our data sources and a general description of their properties. Subsequently, we will describe how the data sets were matched, filtered, and finally combined into a master catalog.

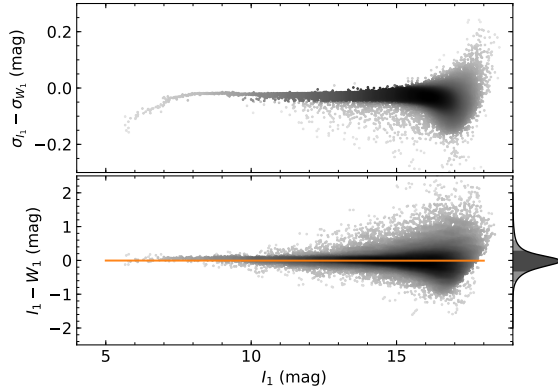


**Fig. 3.** Comparison of VISTA and HAWK-I photometry in  $J$  band. The panel at the top shows the difference in photometric errors as a function of the apparent VISTA magnitude. Similarly, the bottom plot displays the difference in the measured source magnitudes. The histogram to the right shows the distribution of the magnitude difference as a kernel density. Assuming a Gaussian distribution, the differently shaded areas under the curve represent the 1, 2, and 3- $\sigma$  limits.

### 2.1. Overview

To study extinction in Orion A, we include a variety of passbands at infrared wavelengths. To this end, we combined several different sets of photometric data. The bluest bands under investigation are  $z$  and  $y$ , supplied by the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS DR1; Chambers et al. 2016; Magnier et al. 2016; Flewelling et al. 2016), for which the observations cover the entire molecular cloud. The bulk of the NIR photometry is based on our work presented in Paper I. In this study we introduced a large-scale imaging survey of Orion A with the Visible and Infrared Survey Telescope for Astronomy (VISTA; Dalton et al. 2006; Emerson et al. 2006) which encompasses observations in the  $J$ ,  $H$ , and  $K_S$  passbands. The survey covers an on-sky area of  $\sim 18.3\,\text{deg}^2$  with average 90% completeness limits of 20.4, 19.9, and 19.0 mag in  $J$ ,  $H$ , and  $K_S$ , respectively. These data are supplemented by the observations of Drass et al. (2016), who used HAWK-I (Kissler-Patig et al. 2008) to study the ONC and its immediate surroundings. With a coverage of  $\sim 0.17\,\text{deg}^2$ , they only observed a small area of the entire cloud complex. The completeness limits for these data are 18 mag in  $J$  and  $H$ , and 17.5 mag  $K_S$  in the innermost parts of the surveyed area (the estimated VISTA completeness limits in the same area are a few tenths of a magnitude brighter). Our analysis also covers the MIR bands as observed by *Spitzer* in the four channels of the InfraRed Array Camera (IRAC; Fazio et al. 2004) and the first channel of the Multiband Imaging Photometer for *Spitzer* (MIPS; Rieke et al. 2004). These *Spitzer* data for the Orion star-forming region were provided by Megeath et al. (2012) and cover most parts of Orion A which are associated with large gas column-densities ( $\sim 5.86\,\text{deg}^2$  with IRAC and  $\sim 11.4\,\text{deg}^2$  with MIPS). The survey borders of the  $JHK_S$  and IRAC data are shown in Fig. 1.

Since we include a total of 10 different passbands in the infrared we will use a consistent convention when referring to them throughout this manuscript. While for the bands  $z$  through  $K_S$  we will simply use the original filter name, the four IRAC channels will be referred to as  $I_1$ ,  $I_2$ ,  $I_3$ , and  $I_4$ . Furthermore, we only have access to the first MIPS channel, which will be denoted  $M_1$ . We will also make use of the first two *WISE* bands



**Fig. 4.** Correlation between  $W_1$  and  $I_1$ . The top panel shows the difference in photometric errors, while at the bottom the difference in source magnitudes is displayed. The orange line is our linear transformation between the two systems, which maps the two bands in almost a 1:1 relation. Similar to Fig. 3, the histogram to the right represents the kernel density of the magnitude differences.

(Wright et al. 2010) for calibration purposes, which will be denoted  $W_1$  and  $W_2$ . Furthermore, we associate the passbands  $z$  through  $K_S$  ( $\sim 0.9 - 2.2 \mu\text{m}$ ) with the NIR region of the electromagnetic spectrum, while the *Spitzer* channels  $I_1$  through  $M_1$  ( $\sim 3.5 - 24 \mu\text{m}$ ) are placed in the MIR range. An overview of our wavelength coverage is given in Fig. 2, where the relative spectral response (RSR) curves of all passbands are displayed. The figure also shows the typical effects of extinction on a 3000 K black body.

## 2.2. Cross-matching and cleaning

All catalogs were combined with TOPCAT (Taylor 2005) using a  $1''$  cross-matching radius where we only kept sources within the survey boundaries of the VISTA data (see Fig. 1). All detected sources in any band of either the  $JHK_S$  or the MIR catalogs were retained, independent of whether a match in another passband was found. Only for the Pan-STARRS data we required a match in any of the other catalogs due a large number of detections of nebulousity in regions with extended emission. With these criteria, the mean separation between VISTA and *Spitzer* sources was  $0.3''$  and between VISTA and Pan-STARRS sources  $0.14''$ . For sources which were detected in multiple catalogs, we averaged their coordinates.

The  $JHK_S$  catalogs from Paper I and Drass et al. (2016) did not require photometric transformations between each other, as they are both calibrated toward the 2MASS photometric system. To minimize systematic errors, cross-matched sources between the VISTA and HAWK-I data were only retained if the magnitude difference was smaller than 0.5 mag. For sources which are in agreement with this criterion, we calculated average magnitudes weighted by their measurement errors. For a total of 4 340 sources in the HAWK-I catalog, we find 3 827 cross-matches with the VISTA catalog ( $\sim 88\%$ ), thus 513 new sources. About 5% of the matched sources were then removed based on their magnitude differences. Interestingly, most of the new sources are found in a rim around the core of the ONC and are also quite easily distinguishable from the background in the VISTA image data. The fact that they were not detected in these data can be explained by the source detection procedure: While for the

**Table 1.** Basic properties for all photometric bands used in this work, as well as corresponding source counts in our master catalog for both the Orion A region and the control field.

Band	$\lambda_{\text{eff}}^a$ ( $\mu\text{m}$ )	FWHM <sup>a,b</sup> ( $\mu\text{m}$ )	Orion A (#)	CF (#)
$z$	0.87	0.10	291 259	54 730
$y$	0.96	0.06	233 128	40 099
$J$	1.24	0.22	568 116	84 686
$H$	1.66	0.26	743 352	83 192
$K_S$	2.16	0.28	636 388	70 750
$I_1^c$	3.51	0.74	192 934	30 853
$I_2^c$	4.44	1.01	125 951	25 567
$I_3$	5.63	1.39	21 126	-
$I_4$	7.59	2.83	19 324	-
$M_1$	23.21	5.32	5 914	-

**Notes.** <sup>(a)</sup> All parameters for the photometric passbands were obtained via the Spanish Virtual Observatory (<http://svo2.cab.inta-csic.es/theory/fps3/index.php>). <sup>(b)</sup> FWHM of the filter bandpass. <sup>(c)</sup> *Spitzer* photometry for the control field was obtained via linear transformations from *WISE* (Eqs. 1 and 2).

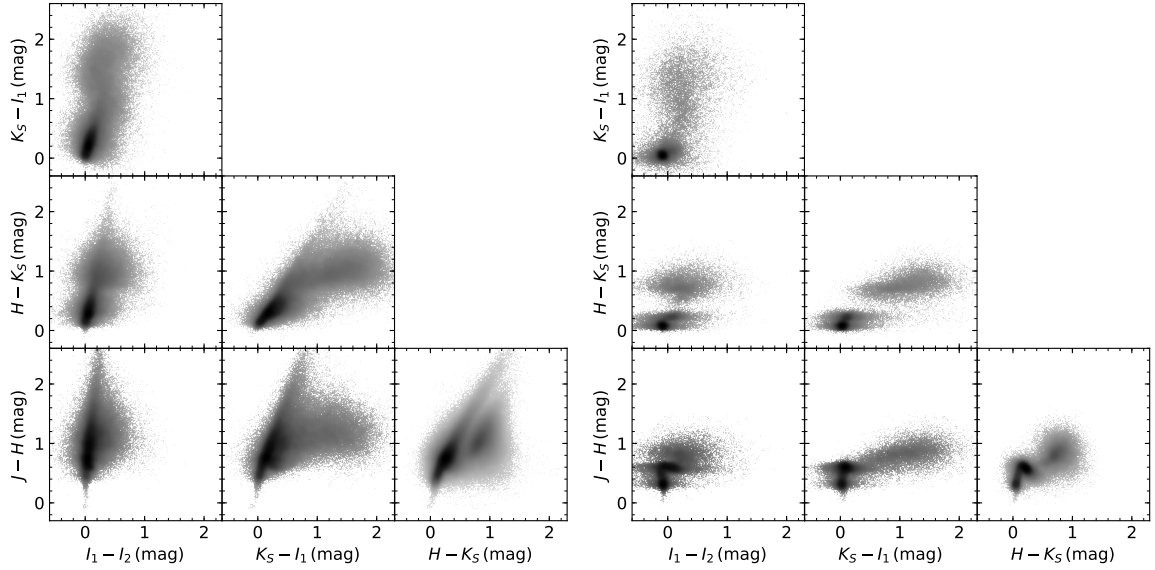
HAWK-I data, the entire mosaic has been searched for sources by eye, for Paper I we only cleaned the very central region ( $11' \times 11'$ ) by hand. This also agrees well with the similar completeness limits stated above. Figure 3 shows a comparison of all matched sources in  $J$  band (the distributions for  $H$  and  $K_S$  look almost identical). In general, we find negligible systematic offsets between the two  $JHK_S$  data sets ( $\Delta\text{mag} \sim -0.02$  mag). The magnitude difference between the two catalogs shows a standard deviation of  $\sim 0.15$  mag across all three bands. The measurement errors, on the other hand, show discontinuities for bright sources due to different criteria on how 2MASS data was incorporated in the catalogs. As expected, for faint sources, the HAWK-I photometry features smaller errors.

As our objective was to explore characteristics of dust extinction in the molecular cloud, we are furthermore primarily interested in background sources as probes of the gas distribution. Therefore, we removed foreground stars near NGC1980, as identified by Bouy et al. 2014, and Young Stellar Objects (YSOs) in the cloud as classified by Megeath et al. (2012) and Großschedl et al. (in prep.). Furthermore, upon creating the final extinction map, we found that several sources produced artifacts on the resulting map. Essentially all of these were associated with very bright stars ( $K_S \lesssim 8$  mag), or their point spread function halos in the VISTA images and therefore were also removed from our master catalog in an iterative procedure.

## 2.3. Control field data

For statistical comparisons of source populations we also required photometry in the same bands for an extinction-free field at approximately the same galactic latitude.  $JHK_S$  photometry for a suitable control field ( $l = 233.3$ ,  $b = -19.4$ ) is available in our data release from Paper I. Pan-STARRS data for the  $z$  and  $y$  bands was acquired and cross-matched in the same way as for the Orion A science field. Only *Spitzer* photometry was not available in this field. Fortunately, the first two channels of IRAC share very similar transmission curves when compared to the two bluest *WISE* bands which enabled us to calculate a simple linear transformation between them. Due to the lower resolution of the *WISE* image data, we relaxed our cross-matching radius to  $2''$  between the NIR sources and *WISE*, but required a de-





**Fig. 5.** A subset of color-color diagrams constructed from our master-catalog. The left-hand panels show data for the Orion A region, while the panels on the right-hand side show the same combinations for the extinction free control field. Evidently, the Orion A field is heavily affected by reddening due to dust in the molecular cloud. The *Spitzer* photometry for the control field was derived with a linear transformation from *WISE* data and therefore shows a larger dispersion when compared to the high-sensitivity data of the *Spitzer* Orion program.

tection with VISTA. In addition we required a photometric error smaller than 0.5 mag in each band. For the fit, we obtained *All-WISE* photometry (Cutri et al. 2013) in the corresponding bands and derived the linear transformation equations using the method later described in Sect. 3.1. Using the profile-fitting photometry in the catalog (columns w1mp, w2mp and their associated errors), we find an almost perfect 1:1 correlation for both bands, for which the transformation equations read

$$I_1 = 0.999 \times W_1 + 0.011 \quad (1)$$

$$I_2 = 1.003 \times W_2 - 0.003. \quad (2)$$

As an example of the well-behaved fit between the two systems, we show the correlation between  $I_1$  and  $W_1$  in Fig. 4 (the correlation between  $I_2$  and  $W_2$  looks almost identical).

#### 2.4. Master catalog

With the criteria listed above we constructed our final master catalogs for both the Orion A, as well as the control field. Number counts and basic passband properties of all bands under consideration are tabulated in Table 1. Furthermore, a selection of color-color combinations is displayed in Fig. 5 for both the Orion A region (left-hand panels) and the control field (right-hand panels). Clearly, the Orion A field is affected by a significant amount of reddening when compared to the extinction-free control field. In almost all diagrams a clear distinction between extragalactic sources and main sequence stars is visible (e.g.  $J - H \sim 1$ ,  $H - K_S \sim 0.8$ ,  $K_S - I_1 \sim 1.2$ ). Another interesting feature in the Orion A diagrams is the clearly discernible changing extinction vector: the slope in the diagrams gets significantly steeper from  $H - K_S$  to  $I_1 - I_2$  indicating decreasing extinction differences in the two bands constituting each color. For example, in the diagrams including  $I_1 - I_2$ , the extinction vector is almost vertical.

A closer look at the control field diagrams, including the transformed *Spitzer* photometry, reveals a non-negligible amount of additional noise compared to the original *Spitzer* Orion A photometry. Most prominently, the stellar main sequence at  $I_1 - I_2 \sim 0$  is significantly broader in the control field. This presents a problem for statistical comparisons. We will examine this issue further in Sect. 5 where these diagrams are used to construct an extinction map. For the remainder of this article, all further analysis is based on this master-catalog and throughout the individual sections we will refer to various subsets on multiple occasions.

### 3. Methods

In this section we will give an in-depth overview of all methods used in this manuscript. In particular, this includes mathematical background on how we fit observed noisy data with linear relations (Sect. 3.1), how to derive an infrared extinction law (Sect. 3.2.1), and how to measure extinction for point-sources (Sect. 3.2.2). Finally, we will also describe how to use these point-source measurements to construct the smooth gas distribution in the ISM in the plane of the sky with extinction mapping procedures (Sect. 3.2.3).

#### 3.1. Linear fitting

Many results derived in this manuscript depend on reliable linear fitting procedures. Moreover, our interpretations and analysis also depend on the statistical significance of the resulting fit. For this purpose, we develop in this section a robust Bayesian framework which takes errors in both dimensions into account and ultimately enables us to derive reliable model parameters. The resulting framework is then employed in combination with a Markov chain Monte Carlo (MCMC) ensemble sampler. Specif-

ically, we define a likelihood function which can be numerically optimized and also can, in principle, take any form, but must be an appropriate choice for the given data characteristics. The MCMC algorithm derives suitable model parameters by sampling the parameter space in a so-called “chain”. Individual positions in the chain are chosen based on the used samplers (e.g. Metropolis; [Mackay 2003](#)) and at each position in the chain, the posterior probability is evaluated. The probabilities at each step constitute a “trace” and these traces allow statistical evaluations for each parameter. Here, we always choose to use the mean and the standard deviation of the traces to describe the model parameters and its uncertainties. For reasons of simplicity and reproducibility we chose the *emcee*<sup>2</sup> MCMC ensemble sampler ([Foreman-Mackey et al. 2013](#); [Hogg et al. 2010](#)). This implementation allows to optimize model parameters with respect to a given pre-defined model likelihood. Details on how *emcee* constructs the chains are given in [Foreman-Mackey et al. \(2013\)](#).

The MCMC approach requires to define a model for the data and an associated likelihood. We start by defining a simple linear relation between generic values on the abscissa and the ordinate in a Euclidean space, which is given by

$$y(x | \alpha, \beta) = \alpha + \beta x \quad (3)$$

where  $\alpha$  is the intercept, and  $\beta$  the slope. Furthermore, we define our generic observed dataset as

$$X = \{x_i, y_i\} \quad (4)$$

where the  $x_i$  and  $y_i$  are the measurements with corresponding errors  $\sigma_{x,i}$  and  $\sigma_{y,i}$ . We also define the unknown true values  $\bar{x}_i$  and  $\bar{y}_i$ , which are connected to the above set via

$$x_i = \bar{x}_i + \epsilon_{x_i} \quad (5)$$

$$y_i = \bar{y}_i + \epsilon_{y_i}. \quad (6)$$

Here we adopt a normal distribution for the errors  $\epsilon$  in both coordinates such that

$$\epsilon_{x_i} = N(0, \sigma_{x_i}^2) \quad (7)$$

$$\epsilon_{y_i} = N(0, \sigma_{y_i}^2). \quad (8)$$

Given these assumptions and our data set, we can derive a model likelihood, for which the dispersions in  $x$  and  $y$  along the linear relation are described by normal distributions, which in turn are characterized by the measurement errors. Thus, the probability for each data point, given  $\alpha$  and  $\beta$  defining a model as in Equ. 3, is

$$P(x_i, y_i | \alpha, \beta, \bar{x}_i, \sigma_{x,i}, \sigma_{y,i}) = \frac{1}{\sqrt{2\pi}\sigma_{y_i}} \exp\left[-\frac{(y_i - \alpha - \beta\bar{x}_i)^2}{2\sigma_{y_i}^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left[-\frac{(x_i - \bar{x}_i)^2}{2\sigma_{x_i}^2}\right], \quad (9)$$

where we already take the linear relation, described by the intercept  $\alpha$  and slope  $\beta$ , into account to eliminate the true values  $\bar{y}_i$ . This definition, however, still includes the unknown true values of  $x_i$ ,  $\bar{x}_i$ . For our application we therefore need to find a way to eliminate them as well. In the following lines we will derive our final model likelihood along which we will also deal with this issue.

<sup>2</sup> Description, documentation, and source code are available at <http://dan.iel.fm/emcee/current/>

The full posterior probability distribution of the model parameters  $\alpha$  and  $\beta$ , given the set of measurements  $X$  can be rewritten to include the true values  $\bar{x}_i$ .

$$P(\alpha, \beta | X, \sigma_{x,i}, \sigma_{y,i}) = \int P(\alpha, \beta, \bar{x}_i | X, \sigma_{x,i}, \sigma_{y,i}) d\bar{x}_i \quad (10)$$

Following Bayes’ theorem, the term in the integral can be written as

$$P(\alpha, \beta, \bar{x}_i | X, \sigma_{x,i}, \sigma_{y,i}) \propto P(X | \alpha, \beta, \sigma_{x,i}, \sigma_{y,i}, \bar{x}_i) P(\alpha, \beta, \bar{x}_i) \quad (11)$$

The right-hand side of this relation splits up into the likelihood function  $P(X | \alpha, \beta, \sigma_{x,i}, \sigma_{y,i}, \bar{x}_i)$  and the so-called prior function  $P(\alpha, \beta, \bar{x}_i)$ . For simplicity, we assume a flat prior on  $\bar{x}_i$ , i.e.  $P(\bar{x}_i) = 1$ . Also, the normalization (the denominator in Bayes’ theorem) can be ignored in our case since it is a constant and would only become important if we attempted to compare different model definitions. Thus, Equ. 10 becomes

$$P(\alpha, \beta | X, \sigma_{x,i}, \sigma_{y,i}) \propto P(\alpha, \beta) \int P(X | \alpha, \beta, \sigma_{x,i}, \sigma_{y,i}, \bar{x}_i) d\bar{x}_i. \quad (12)$$

The likelihood function for the data ensemble  $X$ , given by  $P(X | \alpha, \beta, \sigma_{x,i}, \sigma_{y,i}, \bar{x}_i)$ , can be constructed by multiplying over all  $i$

$$P(X | \alpha, \beta, \sigma_{x,i}, \sigma_{y,i}, \bar{x}_i) = \prod_i P(x_i, y_i | \alpha, \beta, \bar{x}_i, \sigma_{x,i}, \sigma_{y,i}), \quad (13)$$

which, together with Equ. 9 can be inserted into Equ. 12, giving

$$P(\alpha, \beta | X, \sigma_{x,i}, \sigma_{y,i}) \propto P(\alpha, \beta) \times \prod_i \int \frac{1}{\sqrt{2\pi}\sigma_{y_i}} \exp\left[-\frac{(y_i - \alpha - \beta\bar{x}_i)^2}{2\sigma_{y_i}^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left[-\frac{(x_i - \bar{x}_i)^2}{2\sigma_{x_i}^2}\right] d\bar{x}_i. \quad (14)$$

This relation contains an integral over the product of two Gaussians, i.e. a convolution, and thus can be rewritten as a single normal distribution. Furthermore, by rearranging the argument of the exponential term in the first Gaussian function, we are also able to eliminate  $\bar{x}_i$  during this process. The integral in the above equation then evaluates to

$$\int P(X) d\bar{x}_i = \frac{1}{\sqrt{2\pi(\sigma_{y_i}^2 + \beta^2\sigma_{x_i}^2)}} \exp\left[-\frac{(y_i - \alpha - \beta x_i)^2}{2(\sigma_{y_i}^2 + \beta^2\sigma_{x_i}^2)}\right]. \quad (15)$$

Defining

$$\xi_i^2 = \sigma_{y,i}^2 + \beta^2\sigma_{x,i}^2 \quad (16)$$

allows to rewrite our likelihood function which finally reads in logarithmic form

$$\ln P(X | \alpha, \beta, \xi_i) \propto -\frac{1}{2} \sum_i \left[ \ln(2\pi\xi_i^2) + \frac{(y_i - \alpha - \beta x_i)^2}{\xi_i^2} \right]. \quad (17)$$

This likelihood allows to efficiently derive optimized linear fitting parameters, taking measurement errors in both dimensions into account. Interestingly, [Clutton-Brock \(1967\)](#) also investigated this issue and arrive at the same final likelihood definition, using a more traditional approach.

For the posterior probability defined in Equ. 11 we also need the prior function  $P(\alpha, \beta)$ . Here, we chose a uniform prior for

our application and write the logarithmic prior for the intercept  $\alpha$  and the slope  $\beta$  in the form of a piecewise function:

$$\ln P(\alpha, \beta) = \begin{cases} 0 & \alpha_{\text{low}} \leq \alpha \leq \alpha_{\text{high}} \wedge \beta_{\text{low}} \leq \beta \leq \beta_{\text{high}} \\ -\infty & \text{otherwise} \end{cases}, \quad (18)$$

where  $\alpha_{\text{low}}$ ,  $\alpha_{\text{high}}$ ,  $\beta_{\text{low}}$ , and  $\beta_{\text{high}}$  are reasonable choices for the given application. For instance, when fitting color-color diagrams, we know that the intercept must be close to 0. We will, however, be very conservative with the choice of our priors, which mostly are needed for faster convergence of the numerical optimization process. Finally, with the prior given in Equ. 18 and the likelihood given in Equ. 17 we can calculate the posterior probabilities of the model parameters  $\alpha$  and  $\beta$  from Equ. 12 as

$$\ln P(\alpha, \beta | X, \xi_i) \propto \ln P(X | \alpha, \beta, \xi_i) + \ln P(\alpha, \beta) \quad (19)$$

We want to close this section on our linear fitting algorithm by acknowledging that we are fully aware that our choice of a Gaussian distribution is not always strictly true. However, for most problems it is extremely challenging to define accurate likelihood functions. For instance, when fitting color-color diagrams, the intrinsic distribution of stellar colors and the dependence of this distribution on extinction (faint low-mass stars have different colors than bright high-mass stars) produce a highly complex spread in colors along the extinction vector, which is very difficult to model without prior information on the intrinsic source characteristics. Therefore, choosing a gaussian likelihood presents a very good compromise between the actual distribution and neglecting these errors in the first place, as is usually done when facing such problems.

### 3.2. Measuring extinction

In the following sections we discuss our procedures to derive an extinction law from photometric data, to extinction for single sources, and to construct the smooth column-density distribution from these measurements.

#### 3.2.1. Deriving an extinction law

Many previous studies have derived extinction laws for a variety of different sightlines. In general, methods relying on spectroscopic data have proven to be very effective for determining the line-of-sight extinction as a function of wavelength for individual sources. For these, the underlying procedure relies on comparing spectra of the target of interest to an extinction-free source. On the other hand, the availability of photometric surveys, covering large spectral and spatial ranges, has also led to the development of techniques to derive extinction laws for the various available photometric passbands.

In general, these methods can be split into absolute determinations of the extinction law with respect to wavelength, and relative measurements. In the case of absolute determinations, a direct calibrator is required. Noteworthy examples of such calibrations can be found in Indebetouw et al. (2005), Schlafly et al. (2010), and Schlafly & Finkbeiner (2011). Indebetouw et al. (2005) parametrized the apparent magnitudes of red clump giants to determine the interstellar reddening as a function of distance. Schlafly et al. (2010) measured the main-sequence turnoff at the blue edge for various SDSS colors and then fitted this “blue tip” with a reddening-dependent relation, whereas Schlafly & Finkbeiner (2011) compared measured SDSS colors to model-predicted source colors.

In contrast, studies determining only a relative dependence of the extinction law require a zero-point to calibrate their results (e.g. Flaherty et al. 2007; Ascaso et al. 2013). In our case, we will also only determine a relative extinction dependence, mainly because it is very challenging to establish an absolute law for the specific case of the Orion A molecular cloud based on photometry. One can easily imagine that techniques relying on a distinct population of red giant stars can not be applied here due to the location of the cloud near the galactic anti-center and  $\sim 20^\circ$  below the plane. Also, our limited field-of-view which, to a large extent, is heavily affected by dust-extinction, makes statistical determinations of features in color space (e.g. the blue edge) unreliable without prior assumptions on the dust distribution. Another reason for our choice of a relative calibration is that the interpretation of our findings does not depend on a direct calibration.

Here, we will derive a wavelength-dependent extinction law for a variety of photometric passbands based on the source distribution in a set of color-color diagrams. The basis for this method lies in the definition of a source’s color excess

$$E_{m_1-m_2} = (m_1 - m_2) - (m_1 - m_2)_0 \quad (20)$$

$$= (m_1 - m_{1,0}) - (m_2 - m_{2,0}) = A_{m_1} - A_{m_2}, \quad (21)$$

where  $E_{m_1-m_2}$  is the notation for the color excess measured in the passbands  $m_1$  and  $m_2$  (e.g.  $J$  and  $H$ ). This excess is simply the difference between the measured apparent color ( $m_1 - m_2$ ) and the intrinsic color  $(m_1 - m_2)_0$  of the same source, or the difference in extinction in the two passbands ( $A_{m_1} - A_{m_2}$ ). Using two color excesses in differently combined passbands the color excess ratio  $\beta_{m_i}$  is defined as

$$\beta_{m_i} = \frac{E_{J-m_i}}{E_{J-K_S}} = \frac{A_J - A_{m_i}}{A_J - A_{K_S}}, \quad (22)$$

and represents the slope in the diagram plotting the two parameters against each other. Here,  $m_i$  denotes one of the several passbands we investigate in this manuscript and  $A_{m_i}$  refers to the extinction in this particular band. Furthermore, for all applications we will use  $J - K_S$  as the baseline of this parameter. This choice offers both a large range in measured colors ( $\Delta(J - K_S) > 8$  for many areas) and superior number statistics.

Upon extending the total extinction ratio  $A_{m_i}/A_J$  and substituting with a rearranged Equ. 22 we find

$$\frac{A_{m_i}}{A_J} = \frac{A_J - (A_J - A_{m_i})}{A_J} = \beta_{m_i} \left( \frac{A_{K_S}}{A_J} - 1 \right) + 1. \quad (23)$$

Using this equation, we see that the total extinction ratio for any passband can be determined by measuring the slope  $\beta_{m_i}$  and assuming a value of  $A_{K_S}/A_J$ . For the remainder of this manuscript we adopt  $A_J/A_{K_S} = 2.5 \pm 0.15$  ( $A_{K_S}/A_J = 0.4 \pm 0.024$ ) as published by Indebetouw et al. (2005). However, since this conversion necessarily includes further systematic errors, we prefer to make any comparisons directly with the fitted color excess ratio.

A closer examination of the equations 20, 21, and 22 reveals that the determination of  $\beta_{m_i}$  requires knowledge about the intrinsic source colors. It is, however, possible to bypass this requirement. Ignoring differences across the filter bandpasses due to varying spectral energy distributions, all sources in a color-color diagram are pushed in the same direction along the extinction vector in a diagram plotting the apparent colors of the sources. Thus, the slope in the distribution of apparent source colors equals the color excess ratio in Equ. 22.

One caveat of this assumption is tied to the intrinsic source color distribution in color-color diagrams. That is to say that the intrinsic color distribution of the stellar main sequence is not parallel to the extinction vector which can introduce a bias in the measured slope. [Ascenso et al. \(2012\)](#) developed a technique to fit color-color diagrams taking this effect into consideration. However, preliminary tests with our data showed that the application of this method is unreliable in our case because the intrinsic source color distribution rapidly changes for fields with variable extinction. In other words, for large amounts of reddening, only intrinsically bright background sources are visible through the cloud, while for regions with low extinction, the full stellar main sequence along with a set of galaxies are sampled in the diagram. For example, when searching for a spatially variable extinction law, this difference in the sampled source population introduces a bias in the fitting results, because the intrinsic color distribution will be different for the individual fields. In contrast, when dealing with a spatially limited field with a well defined background source population, this method minimizes the bias introduced by the distribution of intrinsic source colors (for an application, see [Ascenso et al. 2013](#)).

In our case, the influence of the intrinsic source color distribution is minimal in the first place, since the fields we investigate show a very large range of apparent source colors. Typically we find  $\Delta(J - K_S) > 8$  mag for our fields, whereas changes in intrinsic stellar colors are mostly limited to  $\Delta(J - K_S) < 1$  mag. Thus, the slope difference between intrinsic source colors and the distribution generated by pushing sources along the extinction vector only leads to a broadening of the source distribution along the extinction vector. Using our linear model defined in Sect. 3.1 this broadening will be taken into account when fitting the various color-color diagrams by introducing a systematic noise term which will later be discussed in the relevant sections.

### 3.2.2. Estimating line-of-sight extinction with photometry

Over the last decades a number of methods have been developed to derive extinction from point-source measurements including the well-established techniques relying on measured stellar colors (e.g., [Lada et al. 1994](#); [Alves et al. 1998](#); [Lombardi & Alves 2001](#); [Majewski et al. 2011](#); [Juvela & Montillaud 2016](#)). In this work we use the recently published PNICER method ([Meingast et al. 2017](#)) to derive the extinction toward point sources using their measured colors with a given extinction law. Briefly summarized, PNICER builds on and extends some of the above mentioned methods. The routine applies machine learning techniques to photometric measurements of extincted sources, in combination with reddening-free control field data, to determine the extinction along the line of sight. In particular, PNICER fits Gaussian Mixture Models along the extinction vector in arbitrary numbers of dimensions to derive probability densities describing the extinction for each source. This process also effectively bypasses often-used simplifications of the reddening-free source distribution. In this way, PNICER circumvents many issues related to the more traditional methods mentioned above when highly sensitive photometric data are used. For example, the bias introduced by comparing the source population of highly extincted regions with extinction-free control field data can be eliminated by choosing a large number of dimensions (see Fig. 4 in [Meingast et al. 2017](#)). PNICER does neither require any prior knowledge of the projected gas density distribution, nor does it rely on any model predictions. The method is strictly data-driven and outperforms the above mentioned techniques when the observed source population includes a significant amount of late-type stars

and galaxies which can introduce a complex pattern in the color distribution of the observed sources. Furthermore, PNICER also is capable of reproducing results derived with the established methods for more simple setups (e.g. when using 2MASS data) and is therefore applicable in a multitude of setups.

### 3.2.3. Extinction mapping

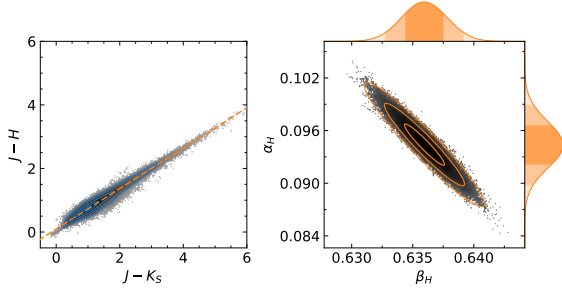
As with determining the line-of-sight extinction toward single sources, there are also a number of mapping techniques available to construct the smooth projected gas surface density from photometric measurements. Traditionally, methods relying on the number density of sources have been used as a proxy for the density-distribution of gas in the interstellar medium (e.g., [Bok & Cordwell 1973](#); [Cambr sy 1999](#); [Dobashi et al. 2005](#)) and these continue to be useful to this day (e.g. [Beitia-Antero & G mez de Castro 2017](#)). Extinction measurements based on stellar colors have also been widely used in the past. Based on the work of [Lada et al. \(1994\)](#), [Lombardi & Alves \(2001\)](#) introduced the multi-band technique NICER which uses a weighted average smoothing process to calculate an extinction map (see also [Alves et al. 2001](#)). The pixel values in such extinction maps are weighted averages derived from line-of-sight “pencil-beam” measurements towards single sources. To combine this method with PNICER, we convert the derived extinction probability densities to discretized values with the expected value of the distribution (see Sect. 3.3 in [Meingast et al. 2017](#)).

Extinction mapping techniques based on discrete measurements of line-of-sight extinctions all suffer from unresolved substructure in the cloud: A given set of stars in a defined region (e.g. a pixel of an extinction map) only provides a limited number of samples of the (presumably smooth) column-density distribution of a molecular cloud. Moreover, the measured background sources have a higher probability to sample the low-column density parts of a cloud because larger extinctions push more background sources beyond the sensitivity limit of the observations. To counter this effect, [Cambr sy et al. \(2002\)](#) introduced a method which combines color-based reddening calculations with star counts. Later, [Dobashi et al. \(2008\)](#) introduced a percentile-based method to minimize this bias and finally [Lombardi \(2009\)](#) published the so-called NICEST method which has already found many applications in recent years (e.g. [Lombardi et al. 2011](#); [Alves et al. 2014](#)). NICEST adds an additional weighting factor -  $10^{\alpha_{\text{NICEST}} k_{m_1} A_{m_1}}$  - to the spatial smoothing process which attempts to correct for the unresolved cloud substructure. Here,  $\alpha_{\text{NICEST}}$  refers to the slope in the expected number counts and the factor  $k_{m_1}$  represents the extinction law in the passband  $m_1$ . When combining higher numbers of dimensions, however, these coefficients are not straight-forward to determine. We will therefore derive the best-fitting value in an iterative procedure.

## 4. The infrared extinction law in Orion A

In this section we present, discuss, and interpret our findings with respect to the infrared extinction law in Orion A. Firstly, we will derive an average extinction law for the entire survey region. Subsequently, we continue to investigate spatial variations in our sample, by separately exploring the effects of the local environment, as well as a potential dependence on gas (column-)density.





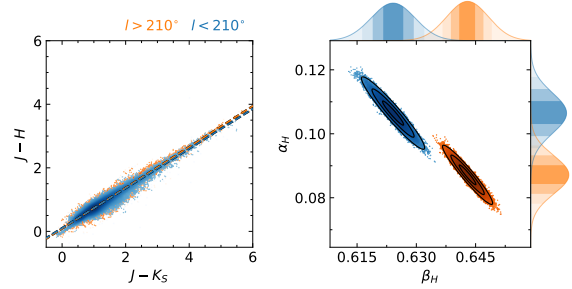
**Fig. 6.** Fitting results for the color excess ratio in  $H$  band. The panel on the left-hand side shows the filtered source distribution in a greyscale scatter plot, where the greyscale is proportional to the number density in this parameter space. The thin blue contours refer to 0.5, 10, 40, and 80 % of the maximum source density. Also plotted is the resulting fit as the orange dashed line. The panel on the right-hand side displays the posterior probability distribution for the intercept  $\alpha_H$  and the slope  $\beta_H$  (the color excess ratio). The ellipses are the 1-, 2-, and 3-sigma covariances of the distribution. Projected to the top and to the right are normalized kernel densities of the resulting distribution for each parameter. The differently shaded filled areas under the curve highlight the 1-, 2-, and 3- $\sigma$  ranges.

#### 4.1. Fitting color-color diagrams

As outlined in Sect. 3.2, we use a series of color-color diagrams with various passband combinations to determine color excess ratios as a proxy for the extinction law. By investigating Fig. 5 in more detail, it becomes apparent that all such combinations show a rather complex distribution in the multivariate color spaces. Since we are interested only in the effects of reddening, ideally only a well defined subset with sources of similar intrinsic colors should be used to fit the relation (e.g. Wang & Jiang 2014). As we do not have thorough source classifications for the entire sample, we start by reducing the amount of contamination with various filtering steps. To this end, subsets for each band have been defined to reduce the dispersion, resulting from intrinsic source colors, as well as the contamination by galaxies, nebulosity, and other spurious detections. Each band was treated separately with the following conditions:

- Only sources with photometric errors smaller than 10 % were considered.
- For the Pan-STARRS bands, each source had to have at least 10 independent detections ( $n_z, n_y > 10$ ). This criterion significantly reduced the amount of nebulous detections.
- For *Spitzer* data we only considered sources brighter than (14, 13, 13) mag in  $(I_1, I_2, I_3)$ . The main purpose here was also to reduce the amount of false detections (for  $I_4$ , the error criterion above was enough to create a clean subset).
- To effectively remove extended sources, which in general have significantly different intrinsic colors compared to stars, we used the flags from the catalog in Paper I (Class\_cog = 1  $\wedge$  Class\_sex > 0.9).
- In addition to the criteria listed above, a few dozen remaining contaminating sources were removed by hand.

Applying the above criteria to our master-catalog resulted in a remarkably clean selection of bright main sequence stars. Nevertheless, this reduced set still contained sources of different intrinsic colors. Intrinsic colors of stars show natural broad sequences, which are, in general, not parallel to the extinction vector. Thus, an additional systematic noise term is present in these data and as



**Fig. 7.** Similar to Fig. 6, we here show the fitting results for  $H$  when splitting the survey region in East-West direction along  $l = 210^\circ$ . Our MCMC fitting method determines a significant difference in the color excess ratios for the two sub-regions, indicating a variable extinction law across the molecular cloud.

a consequence, the dispersion in both axes along the extinction vector is not fully described by the photometric measurement errors anymore. Due to these characteristics, extinguished sources along the reddening vector are affected by a systematic broadening in addition to their photometric errors. For accurate linear fitting, however, this term needs to be accounted for.

To include the above mentioned additional dispersion factor, we measure the natural width of the main sequence in the extinction-free control field for each color. Here, we use the previously defined filtering criteria (the list above) on the control field data and add the variances accounting for the natural broadening to the corresponding squared photometric errors before running the fitting procedure. The variances of our dataset,  $\sigma_{x,i}^2$  and  $\sigma_{y,i}^2$  (Equ. 4), therefore become

$$\sigma_{x,i}^2 = \sigma_{x,i,\text{phot}}^2 + \sigma_{x,i,\text{intrinsic}}^2 \quad (24)$$

$$\sigma_{y,i}^2 = \sigma_{y,i,\text{phot}}^2 + \sigma_{y,i,\text{intrinsic}}^2 \quad (25)$$

where the subscript “phot” refers to the measured errors of the photometry and “intrinsic” to the width of the intrinsic color distribution as measured in the control field. The resulting total variances become rather large due to this additional systematic term, which, however, is well balanced by the large amount of available data points.

In addition to these modifications, we furthermore only include sources satisfying  $J - K_S > 1$  mag in the fit. This value represents the typical red limit in this color for unextinguished late-type main sequence stars and therefore ensures that only sources which are at least minimally affected by extinction are included in the fit. Furthermore, we define the prior function (Equ. 18) with generous limits for the slope and the intercept as

$$\ln P(\alpha, \beta) = \begin{cases} 0 & -5 \leq \alpha \leq 5 \wedge -10 \leq \beta \leq 10 \\ -\infty & \text{otherwise} \end{cases} \quad (26)$$

This definition allows very large ranges in both parameters and at the same time helps to speed up the convergence in the Markov chain.

We note here, that in our case for accurate determinations of the color excess ratios, we also need to allow the intercept  $\alpha_{m_i}$  to be a free parameter in the fit. This is necessary since each color is characterized by a different intrinsic mean, leading to asymmetric offsets in the color-color diagrams. Furthermore, we also note that the actual dispersion along the reddening vector is also a function of extinction: only intrinsically bright background sources (early spectral types) will be visible in regions

**Table 2.** Fitting results for the color excess ratio (slope)  $\beta_{m_i} = E_{J-m_i}/E_{J-K_S}$  and the intercept  $\alpha_{m_i}$  using all data for the entire Orion A molecular cloud. Due to this definition, the values for  $J$  and  $K_S$  are fixed. Tabulated are the average values for the cloud (Sect. 4.2), results from the literature ( $\beta_{m_i,\text{lit}}$ ) including the reference, as well as the fitting results when splitting the survey region at  $l = 210^\circ$  into East and West parts (Sect. 4.3). Also the relative difference in  $\beta_{m_i}$  between East and West regions ( $\Delta\beta$ ) is given. The total extinction ratios relative to the  $K_S$  band and the associated errors have been calculated using Equ. 23 and adopting  $A_J/A_{K_S} = 2.5 \pm 0.15$  mag.

Band	$\langle\alpha_{m_i}\rangle$ (mag)	$\langle\sigma_\alpha\rangle$ (mag)	$\langle\beta_{m_i}\rangle$	$\langle\sigma_\beta\rangle$	$\beta_{m_i,\text{lit}}$	$\langle A_{m_i}/A_{K_S}\rangle$	$\langle\sigma_{A_{m_i}/A_{K_S}}\rangle$	$\beta_{m_i,\text{East}}$	$\sigma_{\beta,\text{East}}$	$\beta_{m_i,\text{West}}$	$\sigma_{\beta,\text{West}}$	$\Delta\beta$ (%)
$z$	-0.306	0.009	-1.371	0.007	-1.560 <sup>1</sup>	4.56	0.29	-1.372	0.009	-1.334	0.012	$2.8 \pm 1.1$
$y$	-0.410	0.008	-0.925	0.007	-0.920 <sup>1</sup>	3.89	0.24	-0.930	0.009	-0.903	0.011	$3.0 \pm 1.6$
$J$	-	-	0.000	-	-	2.50	0.15	0.000	-	0.000	-	-
$H$	0.094	0.002	0.636	0.002	0.640 <sup>2</sup>	1.55	0.10	0.643	0.002	0.624	0.003	$-3.0 \pm 0.6$
$K_S$	-	-	1.000	-	-	1.00	-	1.000	-	1.000	-	-
$I_1$	-0.070	0.014	1.239	0.008	1.243 <sup>3</sup>	0.64	0.09	1.224	0.010	1.264	0.012	$3.2 \pm 1.2$
$I_2$	-0.096	0.023	1.297	0.012	1.307 <sup>3</sup>	0.56	0.09	1.281	0.016	1.321	0.019	$3.0 \pm 1.8$
$I_3$	-0.059	0.023	1.335	0.012	1.331 <sup>3</sup>	0.50	0.09	1.323	0.015	1.358	0.021	$2.6 \pm 1.9$
$I_4$	-0.052	0.021	1.324	0.011	1.329 <sup>3</sup>	0.51	0.09	1.329	0.013	1.319	0.024	$-0.8 \pm 2.1$
$M_S$	-0.172	0.089	1.364	0.044	1.370 <sup>3</sup>	0.45	0.11	1.376	0.069	1.342	0.059	$-2.5 \pm 6.8$

**References.** (1) Schlafly & Finkbeiner (2011); (2) Indebetouw et al. (2005); (3) Flaherty et al. (2007).

of high column-density, thus reducing this systematic dispersion for increasingly redder colors. However, instead of introducing more systematic errors by adding an additional model component, we chose to be conservative and use the same intrinsic color dispersion for all sources in a given color combination.

#### 4.2. The average extinction law

After applying the filtering steps as listed above to the data collection, we first derived an average extinction law for the entire Orion A molecular cloud by fitting all available data in each band. The results of this procedure are (among others) tabulated in Table 2 for each band, where the measured color excess ratios  $\beta_{m_i}$  are also converted to total extinction ratios  $A_{m_i}/A_{K_S}$  using Equ. 23 and  $A_J/A_{K_S} = 2.5 \pm 0.15$  mag adopted from Indebetouw et al. (2005). With this equation it also becomes trivial to convert the measured slopes into total extinction ratios, in case a different zero-point is desired. As a representation for all bands, Fig. 6 shows the fitting results for the  $H$  band. Also displayed are the posterior probability distributions of the color excess ratio  $\beta_H$  (the slope) and the intercept  $\alpha_H$ . In the specific case of  $H$ , we find  $\langle\beta_H\rangle = 0.636 \pm 0.002$ , thus a statistical error well below 1%. Upon converting this value to the total extinction ratio, we also take the errors from Indebetouw et al. (2005) into account. We then find  $A_H/A_{K_S} = 1.55 \pm 0.1$ .

Due to the introduction of this additional error component when converting the fitted slope to total extinction ratios, it is far better to directly compare the color excess ratios  $\beta_{m_i}$  to results available in the literature. For the Pan-STARRS bands  $z$  and  $y$ , we convert the results of Schlafly & Finkbeiner (2011) to  $\beta_{z,\text{lit}} = -1.56$  and  $\beta_{y,\text{lit}} = -0.92$ . Compared to our values of  $\beta_z = -1.371$  and  $\beta_y = -0.925$ , only  $y$  band matches, while for  $z$  we find a clear deviation. This, however, can be caused by the different normalization as Schlafly & Finkbeiner (2011) list the extinction for the other NIR bands in the UKIRT photometric system. In contrast to our data, the UKIRT  $K$  band is slightly different compared to VISTA  $K_S$ .

In the NIR, literature results generally agree very well with our value. Most prominently, the often-cited work by Indebetouw et al. (2005) and Wang & Jiang (2014) both find  $\beta_{H,\text{lit}} = 0.64$ . Other results for this specific value are from Martin & Whittet (1990): 0.63; Nishiyama et al. (2006): 0.64; Stead &

Hoare (2009): 0.65; Wang et al. (2013): 0.65. All of these are in excellent agreement with our findings and reinforce our conviction that the chosen fitting model and the filtering procedure accurately describe the color excess ratios. Also, we emphasize here that the calibration toward total extinction ratios with the results from Indebetouw et al. (2005) is not in conflict with these findings, because we derive the slope  $\beta$  prior to the conversion.

Flaherty et al. (2007) already derived the total extinction ratios for Orion A in the IRAC bands with data from the *Spitzer* Orion program. In contrast to our study, they used the less sensitive 2MASS observations to fit the color excess ratios and therefore have fewer data points for their analysis. Nevertheless, also here, our values are in excellent agreement with their result of  $\beta_{I_1,\text{lit}} = 1.243$ ,  $\beta_{I_2,\text{lit}} = 1.307$ ,  $\beta_{I_3,\text{lit}} = 1.331$ , and  $\beta_{I_4,\text{lit}} = 1.329$ . For  $M_1$ , they did not derive the extinction for Orion A, but find  $\beta_{M_1,\text{lit}} = 1.37$  and 1.32 for Serpens and NGC 2068/2071, respectively, which also agrees extremely well with our findings.

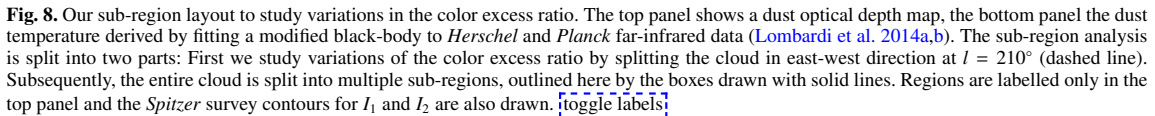
#### 4.3. Spatial variations across the cloud

The results presented in this section are split into two parts. Firstly, we investigate how the extinction law varies across the projected surface of the molecular cloud and, in particular, between regions which are affected by the intense radiation from the massive stars in the ONC. Secondly, we test whether the extinction law is constant when fitting regions with variable gas densities.

##### 4.3.1. Dependence on environment

Among all nearby star-forming regions, the Orion A molecular cloud presents a unique opportunity to study variations in the extinction law across different environments. While the western<sup>3</sup> parts of the cloud are affected by feedback from the hot, ionizing radiation of the massive stars, located in or near the ONC, to the East we find much more quiescent regions with smaller groups of YSOs and isolated star-forming events. Thanks to the supreme sensitivity of the VISTA observations, it becomes possible to systematically study changes in the color excess ratios for indi-

<sup>3</sup> In this manuscript, we always refer to the galactic coordinate frame when using cardinal directions.

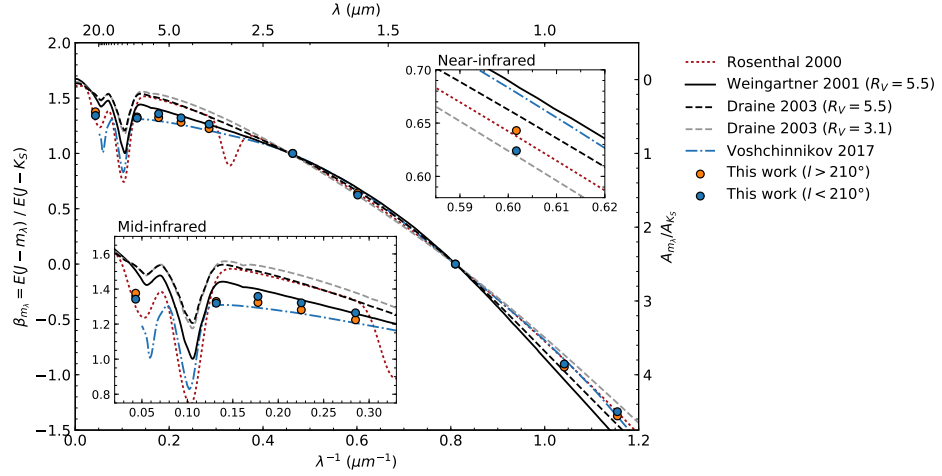


Our selection of the sub-regions is displayed in Fig. 8 for which we repeat the fitting procedure with the same setup as outlined before in Sect. 4.2. In order to investigate spatial variations due to different environments, in a first step we split the survey region in East-West direction into two parts along  $l = 210^\circ$ . This border separates the relatively cold L1641/L1647 regions from the ONC and its surroundings. The results for the color excess ratios for both sub-regions are also tabulated in Table 2 and Fig. 7 visualizes the fitting results for the  $H$  band.

Nevertheless, even though our linear fitting setup is already very conservative regarding statistical and systematic errors, at this point one can still be very skeptical about our findings.

Fig. 9 shows our fitted slopes as a function of wavelength on top of the models from [Weingartner & Draine \(2001\)](#), hereinafter WD01), [Draine \(2003, D03\)](#)<sup>4</sup>, and [Voshchinnikov et al. \(2017, model 25; V17\)](#), as well as the data from [Rosenthal et al. \(2000\)](#). These models represent a mixture of dust particle characteristics and size distributions. The D03 models are based on case A of [Weingartner & Draine \(2001\)](#) which consists of carbonaceous and silicate grains with sizes ranging from a few Å to several  $\mu\text{m}$ . D03 adjusted the size distribution for these models where for  $R_V = 3.1$  the grain abundance was reduced, while for  $R_V = 5.5$  the abundance was increased. The WD01  $R_V = 5.5$  model in Fig. 9 refers to case B in their work case B contains a significant fraction of very large carbonaceous dust grains ( $1 - 10 \mu\text{m}$ ), while case A stops at around  $1 \mu\text{m}$ . The work by V17 is based on laboratory measurements of optical properties of three-layered spheres with the specific aim to understand the widely observed flat MIR extinction law. The comparison to [Rosenthal et al. \(2000\)](#) is especially interesting as they measure an infrared extinction curve toward OMC-1 which lies very close to the ONC.

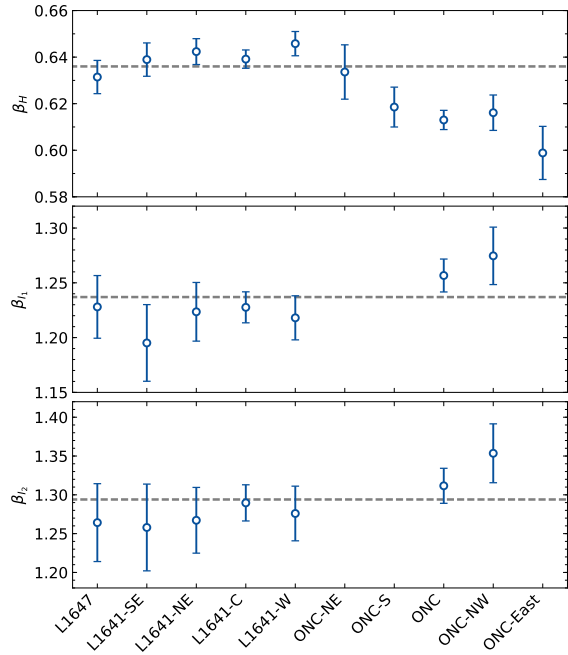
<sup>4</sup> The models from WD01 and D03 are available for download at <https://www.astro.princeton.edu/~draine/dust/dustmix.html>



**Fig. 9.** The infrared extinction curve of Orion A in terms of the color excess ratio  $\beta_{m_i}$  as used in this manuscript. The total extinction ratio  $A_{m_i}/A_{K_S}$  on the right axis of the plot is the conversion of this ratio as given in Equ. 23 with values adopted from Indebetouw et al. (2005). The extinction law as predicted from models by Weingartner & Draine (2001, WD01), Draine (2003, D03), and Voshchinnikov et al. (2017, V17) is displayed with various lines. The red dotted line is the fit of Rosenthal et al. (2000) to their data for OMC-1 near the ONC (including the  $\sim 3.1 \mu\text{m}$  ice feature). The blue and orange circles show our measurements for the East-West split of the survey region.

that for the East-West split, all passbands up to  $I_3$  show a deviation of  $\sim 3\%$ . For the passbands  $z$ ,  $y$ ,  $I_1$ ,  $I_2$ , and  $I_3$  we find a positive difference, while for  $H$  this difference is negative. More specifically, for the region including the ONC and its surroundings ( $l < 210^\circ$ ), we see larger  $\beta_{m_i}$  values for the Pan-STARRS and the *Spitzer* MIR bands, while in  $H$   $\beta_{m_i}$  is smaller. In reference to Fig. 9, we note that the same characteristic is also present in the dust models. Considering for the moment only the models of WD01 and D03, we can see a different behaviour of  $\beta_{m_i}$  when comparing the  $R_V = 3.1$  to the  $R_V = 5.5$  models: While for  $z$ ,  $y$ , and the MIR *Spitzer* bands,  $\beta_{m_i}$  is systematically smaller for  $R_V = 5.5$  compared to the 3.1 models, in  $H$  this trend is reversed (note that due to our color excess ratio definition we have fixed values of  $\beta_{m_i}$  for  $J$  and  $K_S$ . Only for  $H$  band we derive an independent estimate of the slope). Excluding the  $I_4$  and  $M_1$  channel with their much larger uncertainties, we see the exact same behaviour in our fitted color excess ratios. This comparison therefore serves as a critical reinforcing argument with respect to the statistical significance of our linear fits. If our derived values were to be affected by an additional systematic noise component (e.g. biased photometry near the ONC due to extended emission), we would not expect to see such an agreement in the trends, but rather a random distribution.

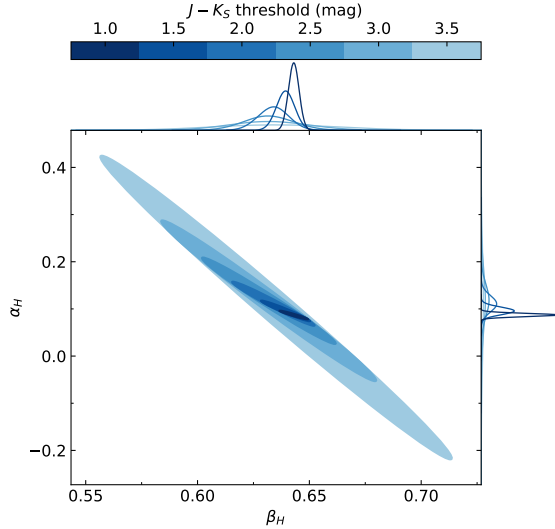
To investigate the origin of these variations in more detail, we divided our dataset into even smaller sub-regions. These definitions are shown in Fig. 8 and are adjusted so that (a) regional variations can be mapped and (b) there were still a reasonably significant amount of sources available. Due to the latter criterion, it was only possible to get reliable fits for  $H$ ,  $I_1$ , and  $I_2$ . The other bands did not have the necessary amount of measured sources to support such an analysis. Also here we repeated the fitting procedure for each sub-region with the same setup as before. The resulting color excess ratios for the three passbands are displayed in Fig. 10. The order of the regions in the figure is organized in such way that the galactic longitude overall decreases from left to right. In  $H$  band there is a clearly visible trend towards smaller  $\beta$  for decreasing longitude. Starting with



**Fig. 10.** Color excess ratios  $\beta_{m_i}$  in  $H$ ,  $I_1$ , and  $I_2$  for the sub-regions of Orion A as defined in Fig. 8. From left to right, overall the galactic longitude of the regions decreases. The dashed horizontal lines represent the averages across the entire cloud. The slopes for L1641/L1647 are systematically larger than the average in  $H$  (smaller for  $I_1$  and  $I_2$ ), while for regions near the ONC, we observe the opposite behaviour.

the ONC-S region and continuing toward the western parts, the values of  $\beta$  are systematically smaller than the average value for





**Fig. 11.** Posterior probability distributions for the color excess ratio  $\beta_H$  and the intercept of the linear fit in the L1641/L1647 region ( $l > 210^\circ$ ) when restricting the sample to various  $J - K_S$  color thresholds. For example, for the  $J - K_S = 1$  mag threshold only source above this limit are fitted. The different thresholds are shown in increasingly dark blue colors and, similar to Fig. 6, the ellipses represent 3-sigma covariances of the parameter distribution. The histograms represent kernel densities for each parameter of the fit.

the cloud (the dashed horizontal line). Furthermore, the values for the L1641 regions are fairly constant and are near or slightly above the average. Within the error bars this is also true for L1647. The regions near the ONC, on the other hand, can be associated with value of  $\beta_H$  smaller than the average (with the exception of ONC-NE).

For  $I_1$  and  $I_2$ , three regions (ONC-NE, ONC-S, and ONC-East) had to be omitted in this analysis, because these were only partly covered by the *Spitzer* survey and did not include the necessary number of sources to produce a reliable fit. Even though the results for these two bands are not as convincing as for  $H$ , the general trend persists. Moreover, in agreement with model predictions, also here  $\beta$  increases for the regions near the ONC, while for  $H$  the color excess ratio decreases. In light of all these findings, we finally conclude, that our results are indeed reliable and significant, which show that the infrared extinction law varies across the cloud. These variations are only of the order of 3%, but this value is consistent across the six passbands  $z$  through  $I_3$ .

#### 4.3.2. Dependence on column-density

Due to external environmental effects on local dust characteristics, it is also reasonable to assume that embedded regions (i.e. regions with large extinction) harbor different dust grains compared to more exposed parts of a molecular cloud. To test this dependency, we repeated the fitting procedure once again, but this time on samples with variable amounts of extinction. As extinction is directly proportional to a source's color, we performed this test by restricting our samples with a series of color thresholds in  $J - K_S$ . Because of our previous result on the variable infrared extinction law in East-West direction across the survey

region, we need to take extra precautions here. Therefore, for an unbiased result, we performed this test only for the sub-region defined by  $l > 210^\circ$  (L1641/1647), which shows a relatively constant behaviour in the color excess ratio. The results of the fit are visualized for  $H$  in Fig. 11, where the posterior probabilities for the intercept and the slope are shown. The distributions for  $I_1$  and  $I_2$  look almost identical with the exception of larger errors.

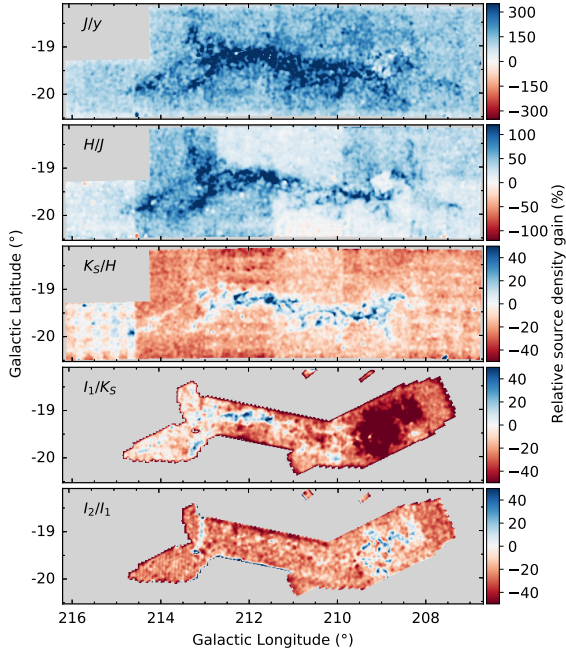
Across all passbands and sub-regions, and within the statistical significance of our fits, we are not able to determine a variation in the color excess ratio as a function of  $J - K_S$  color. However, the regional variations determined above are of the order of only a few percent. The smaller sample sizes with increasing color threshold lead to increasingly large errors in this determination for high extinction thresholds. Therefore, our sample size may not be large enough to make a final assessment on an extinction-dependent color excess ratio in Orion A. Nevertheless, these results agree with the findings of Wang & Jiang (2014) who also concluded that the NIR extinction law does not vary with gas density.

#### 4.4. The shape of the Orion A infrared extinction curve

Studying Fig. 9 in more detail, reveals several interesting findings with respect to the general trend in the infrared extinction. Overall, we find that our fitted color-excess ratios are well traced by the various models. In particular, we see a flat MIR extinction curve, very similar to the ones that have been observed toward the galactic center (e.g. Wang et al. 2015, and references therein). This contrasts to the findings of Rosenthal et al. (2000) for the OMC-1 region. Compared to our measurements of a relatively flat MIR extinction law in the IRAC channels, their data suggests a continued decline in the extinction toward the  $9.7 \mu\text{m}$  feature. The same trend is seen in both D03 models. Our interpretation of this discrepancy is, that the work by Rosenthal et al. (2000) traces only very local conditions within OMC-1, which are not representative for the extinction law on larger scales in this region. However, they also state, that their derived extinction minimum at  $6.5 \mu\text{m}$  is not very well constrained. On the other hand, the WD01  $R_V = 5.5$  (case B) model along with the new results from V17 are a much better match to our results in the MIR.

Agreements between a measured flat MIR extinction and the WD01  $R_V = 5.5$  case B model are often interpreted as an indicator of the presence of very large grain sizes (up to  $10 \mu\text{m}$ ). The comparison to the model by V17, however, shows, that this must not necessarily be the case. The V17 size distribution is based on work by Hirashita & Voshchinnikov (2014), who investigate dust grain growth over several hundred Myrs. Their initial distribution is taken from case A of WD01, thus only featuring smaller grain sizes. Even after a few Myrs, dust particle sizes do not exceed  $1 \mu\text{m}$ , yet the new models of V17 are able to explain the flat MIR extinction curve.

In  $H$ , the situation is reversed. Here, the Rosenthal et al. (2000) and D03  $R_V = 3.1$  models fit much better to our color excess ratios and the models that fit best to our MIR findings (V17 and WD01) are by far the worst matches for this passband. Even worse, for both  $M_1$  and the Pan-STARRS bands  $z$  and  $y$ , yet another set of models seems to reproduce our findings better. It therefore seems, that these dust models are not able to fully reproduce the infrared extinction law from  $\sim 1$  to  $\sim 30 \mu\text{m}$ , but are tuned to fit a narrower range. In conclusion for the overall infrared extinction law in Orion A, we argue that grain sizes are not the only necessary factor to explain variations, but more sophisticated models are required.



**Fig. 12.** Comparison of source densities for consecutive passbands. The individual panels show the ratio of source densities for two particular passbands (e.g.  $J$  and  $y$  in the topmost panel) evaluated with a  $1'$  Epanechnikov kernel. The color scale indicates the gain (or loss) of the redder band over the bluer. Clearly, the  $JHK_S$  VISTA observations show overall the best sampling of sources in the region and even redder bands include more extincted sources in the cloud. Due to the small difference in the extinction between  $I_1$  and  $I_2$ , the source density gain is negligible for  $I_2$ . The apparent large-scale pattern for the second and third panel from the top is caused by the variable completeness of the VISTA observations.

Our analysis also revealed a variable extinction law in East-West direction across the survey region: Toward the East of the ONC we find a systematically different extinction law as in regions near the cluster. Because of the above mentioned caveats when comparing our color excess ratios to dust models, it is almost impossible to arrive at a conclusion with respect to a variation in specific dust characteristics. As disappointing as it is, for the moment we can only conclude that the grain population in general changes across the cloud. Whether this is due to grain sizes, different particle structures, or abundances, we can not say.

We can however, speculate on the origin of these variations. The bottom panel of Fig. 8 highlights the dust temperature derived with black body fits to *Herschel* and *Planck* dust emission measurements. In this figure it is clearly visible that the L1641/L1647 regions have a significantly lower dust temperature (10 – 15 K), while the other parts are apparently heated by the hot stars in the cluster. Even in the ONC-East region, which is spatially clearly separated from the cluster region, it seems that cold dust is surrounded by a layer of much warmer material. In contrast, the L1641W region, which is (in this projected view) closer to the cluster, does not have such an apparent layer of warm dust. It therefore seems, as if L1641 sits in the shadow of the cloud and is shielded from the intense radiation originating in the massive cluster stars. Therefore, the conclusion here

is, that the radiative feedback from the hot stars significantly impacts the dust grain population.

Considering that we found variations in the color excess ratio of the order of only 3%, we furthermore argue that research into variations of the infrared extinction law (in particular for the NIR) should be carried out in environments where systematic errors, or external effects can be controlled to at least some extent. For instance, when determining the extinction law for various sightlines near the galactic center or in the plane, it becomes very difficult to control the influence of physically separate regions along the line of sight, potentially diluting signals of the order of only a few percent. Thus, such studies should ideally be carried out in isolated environments, such as the Orion A molecular cloud.

## 5. Orion A extinction map

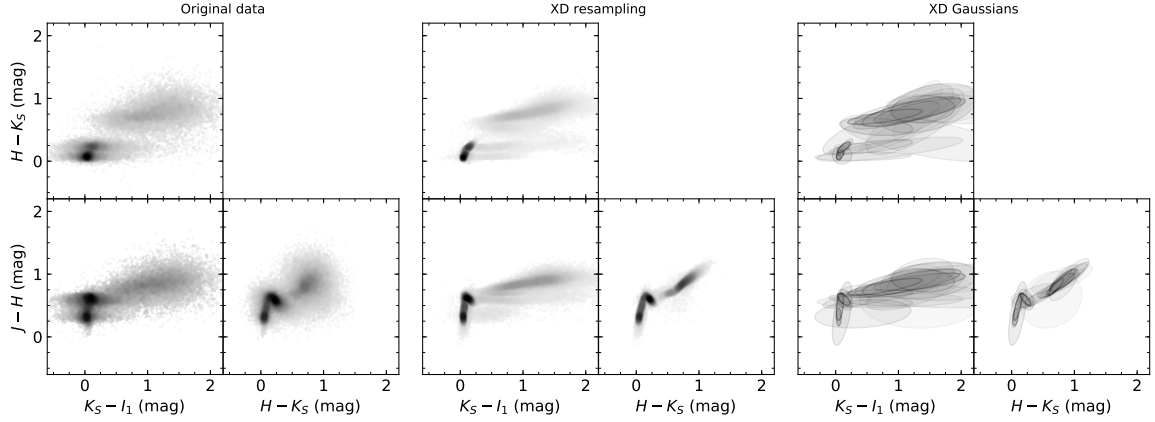
Following the determination and discussion on the extinction law in the previous sections, this section concentrates on creating a new extinction map for Orion A.

### 5.1. Preparations

From our analysis of the variability of the color excess ratio it is not entirely clear which regions exactly are characterized by a given infrared extinction law. We were able to determine a trend in East-West direction which is most likely caused by the feedback from the massive cluster stars. However, drawing an exact border is not possible. For this reason we construct an extinction map for the entire survey region with the average extinction law determined in Sect. 4.2 (see Table 2). To calculate the individual color excesses for each source in our catalog, we use the PNICER method (see Sect. 3.2.2 for details) which is able to combine photometric measurements in arbitrary numbers of dimensions. Usually, including more dimensions (i.e. more passbands) helps to decrease the statistical noise in the extinction calculations. However, the more bands included, the more susceptible the method becomes to systematic errors. For example, in our case the Pan-STARRS photometry is often affected by extended emission, leading to systematic offsets in the measured magnitudes which are not accounted for by the statistical error from the photometry. Furthermore, in the case of our combined photometric master catalog, it may not even be necessary to include some bands due to very different sensitivity limits and the number of available sources in general.

To find an optimal compromise, we investigated the source densities across the survey area for each band. The results for a subset of these is shown in Fig. 12, where the gain in number density is plotted for consecutive passbands. Easily visible is the tremendous source density gain between the Pan-STARRS  $y$  and the VISTA  $H$  band. Not only do the NIR VISTA bands suffer much less from extinction (Fig. 2), but also the survey in general is tuned to observe much fainter objects. Another significant gain in the entire region is achieved when using  $H$  over  $J$  band. For  $K_S$  over  $H$  and  $I_1$  over  $K_S$  we find significantly more sources in the highly extincted regions. From the first to the second IRAC channel there is almost no gain due to the very small difference in extinction in these two bands (see Fig. 9).

In light of these findings, we chose to include only the NIR bands  $J$ ,  $H$ , and  $K_S$ , as well as  $I_1$  in the MIR. We specifically chose to not include  $I_2$  due to the small length of the extinction vector in  $I_1 - I_2$ , which can lead to large systematic errors when determining the color excess with just these two bands. Moreover,  $I_2$  does not add sources to the highly extincted regions



**Fig. 13.** Extreme deconvolution (XD) resampling process of the three-dimensional control field data space ( $J - H$ ,  $H - K_S$ ,  $K_S - I_1$ ). The leftmost set of panels shows the original data distribution as projected views. The three panels in the center show the same set of combinations for the resampled data with  $10^5$  randomly drawn sources. The rightmost panels show the locations of the individual three-dimensional Gaussians. For these, the opacity of the ellipses is an indicator of their respective weight in the model.

compared to  $I_1$ . The Pan-STARRS bands are omitted, because they include significantly fewer sources than the VISTA observations and do not sample the high column-densities in the first place.

PNICER relies on intrinsic color determinations in a control field. As outlined in Sect. 2, we do not have access to suitable *Spitzer* control fields data and therefore use transformed *WISE* photometry (equations 1 and 2). However, as the *WISE* data are associated with larger uncertainties, the intrinsic color distribution is expected to be significantly broader compared to *Spitzer* data. This is readily visible in the comparison of the Orion A data to the control field in Fig. 5. Furthermore, it would also be beneficial to reduce the scatter in the NIR data due to photometric errors for a better description (i.e. narrower sequences) of the intrinsic colors.

For this purpose we fit Gaussian Mixture Models to the density distributions in the three-dimensional color space ( $J - H$ ,  $H - K_S$ ,  $K_S - I_1$ ) incorporating the measurement errors. In Bayesian statistics in astronomy, this process is commonly referred to as extreme deconvolution (Bovy et al. 2011). We use the astroML<sup>5</sup> Python implementation (Vanderplas et al. 2012; Ivezić et al. 2014) of this method and apply it to the control field data in all three dimensions simultaneously. The number of model components (Gaussians) is chosen by minimizing the Bayesian Information Criterion (BIC; Schwarz 1978)

$$\text{BIC} = -2 \ln L + k \ln N, \quad (27)$$

where  $k$  is the number of model parameters,  $N$  the number of sources, and  $L$  the likelihood of the data under the given model. This is necessary because, one common caveat of such methods is that in principle it is possible to artificially increase the likelihood of models by increasing the number of components. The BIC definition attempts to correct for this fact by introducing a penalty term which includes this number ( $k$ ). In our case the BIC showed a very flat distribution starting with about 5 – 10 Gaussians. However, by visually inspecting the fitting results we found that some features of the distribution were only reproduced by relatively high numbers of Gaussians (20 – 30). The

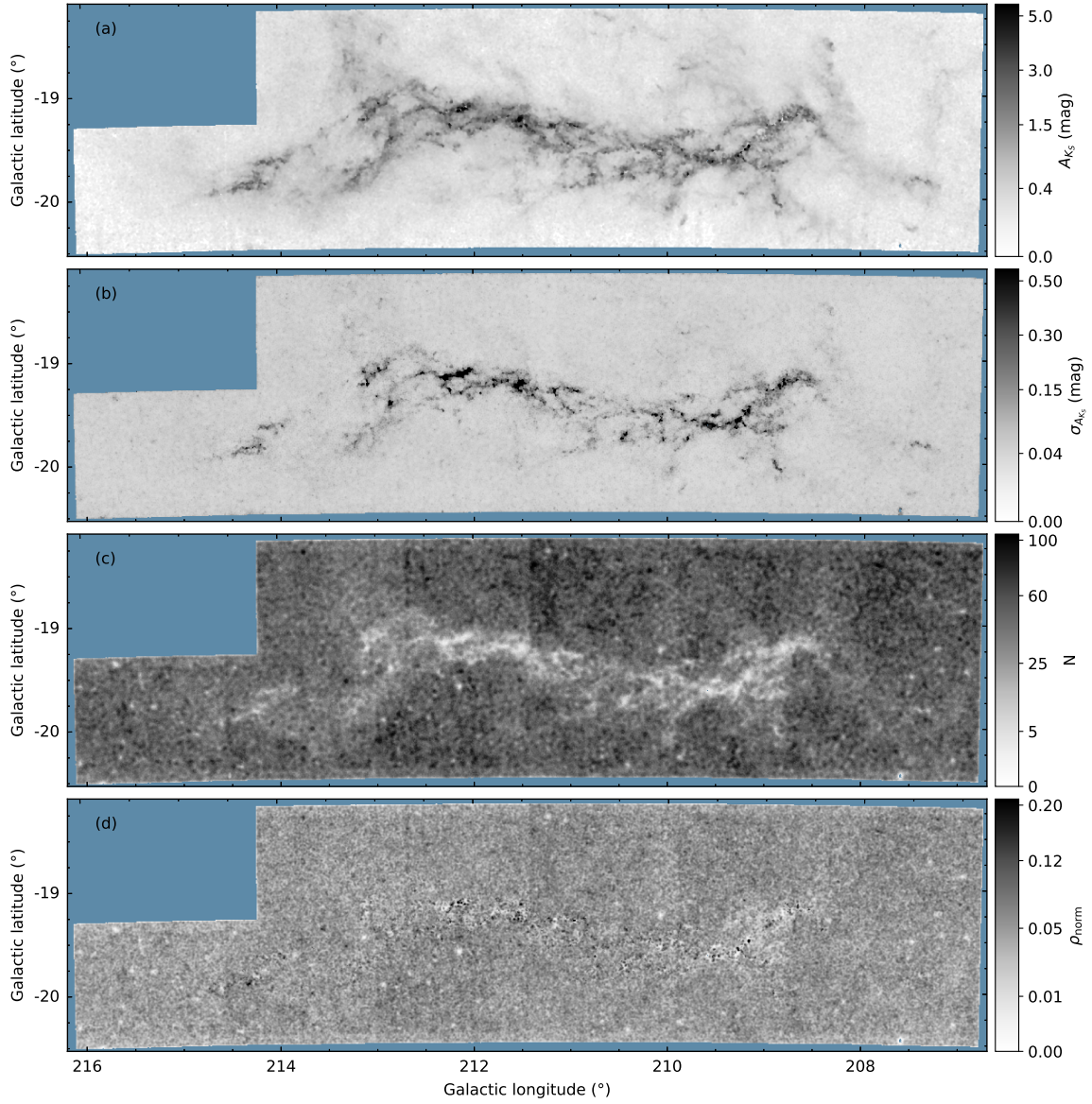
fitting results of our final choice of 25 Gaussians is shown in Fig. 13. The cloned distribution of the XD resampling provides a very clean sample of the color space. Moreover, the resampling process of the data space successfully produced a distribution which reduces the influence of photometric errors in the control field. We used this resampled distribution of the color space to draw  $10^6$  randomly selected sources to construct a model control field. Using PNICER we then calculated line-of-sight extinctions for all sources which had independent measurements in at least two bands.

## 5.2. The extinction map

Before creating an extinction map one further complication was to select the correct density correction scaling factor  $\alpha_{\text{NICEST}}$  (Sect. 3.2.3). This parameter represents the slope of the luminosity function and depends on the used bands. For NIR  $JHK_S$  data this value typically is 1/3 (Lombardi 2009). Here, however, we also include the *Spitzer*  $I_1$  channel which complicates the situation. Instead of trying to empirically derive the correct factor, we created a series of preliminary extinction maps with variable correction factors ( $\alpha_{\text{NICEST}} = [0, 1]$  in steps of  $10^{-2}$ ). The goal of this exercise was to calculate the extinction-corrected source density map and minimize the dispersion therein. The minimum value was found at  $\alpha_{\text{NICEST}} = 0.25$ .

One major disadvantage of extinction mapping with discrete line-of-sight samples is the method's dependency on the number of available background sources. The number density of the sources in highly extinguished regions ultimately restricts the resolution of the map. Here, the case of the Orion A molecular cloud is one of the more difficult regions with respect to extinction mapping. This is because the cloud is located toward the galactic anti-center and also well below the galactic plane, thus offering intrinsically very few background sources. For essentially all other nearby molecular clouds (e.g. the Ophiuchus molecular cloud or the Pipe Nebula) many more background sources are available, since they are projected against the galactic bulge. For these regions it is rather common to achieve much better extinction map resolutions compared to *Herschel* dust emission maps. To determine this limit for the given data, we created a

<sup>5</sup> Source code and description available at <http://www.astroml.org> and <https://github.com/astroML/astroML>.

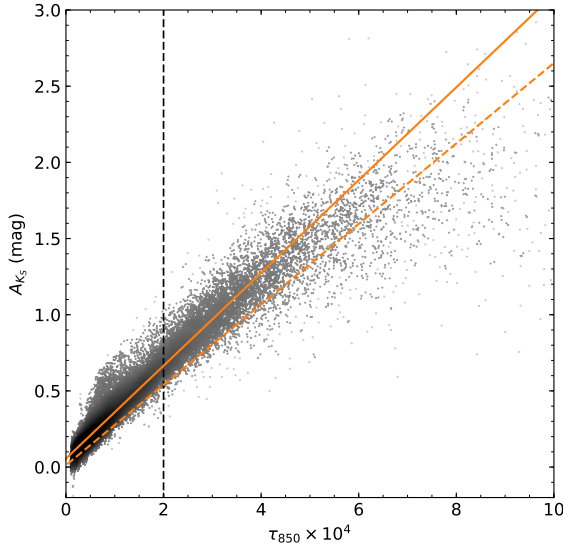


**Fig. 14.** Orion A extinction map constructed with VISTA and *Spitzer* data. At the top, panel (a) shows the extinction map in  $K_S$  magnitudes with a resolution of  $1'$ . Panel (b) displays the associated  $1-\sigma$  errors and (c) the number of sources used for each pixel to calculate the average extinction. The bottom panel (d) shows the source density when adjusted with the Nicest technique, which applies a correction factor to account for a bias due to cloud substructure.

series of extinction maps with a variety of resolutions following the method described in Sect. 3.2.3. We found that the limit in resolution lies at about  $1'$ . With this value only very few pixels could not be sampled in the high density regions of the cloud due to a lack of background sources. Well-sampled pixels are only achieved at lower resolutions of about  $1.5' - 2'$ . We therefore provide a series of extinction maps calculated at different resolutions which will be made available to the community via CDS.

Figure 14 shows the final Orion A extinction map at  $1'$  resolution in units of  $K_S$  band extinction. In particular, we used a gaussian kernel with  $\text{FWHM} = 1'$  which was truncated at 5 standard deviations and a pixel size of  $0.5'$  for sufficient sampling. The figure also includes further statistics for each pixel: the extinction error ( $1-\sigma$ ), the number of sources ( $N$ ), and the corrected sources densities resulting from Nicest ( $\rho$ ). The extinction map offers a large dynamic range at much better resolutions than the previously available map of Lombardi et al.





**Fig. 15.** Relation between *Herschel-Planck* dust optical depth at 850  $\mu\text{m}$  ( $\tau_{850}$ ) and dust extinction in  $K_S$ . The orange solid line is our linear fit with the slope  $\gamma = 3042 \text{ mag}$ . The dashed orange line is the previous fit of Lombardi et al. (2014b,  $\gamma = 2640 \text{ mag}$ ) who used an extinction map based on 2MASS data. They also limited their fit to  $\tau_{850} < 2 \cdot 10^4$ , as indicated by the vertical dashed line, while we use the entire range displayed in this figure. We find a  $\sim 15\%$  larger slope in the relation due to much better sampling of the cloud's substructure.

(2011) and traces the dense gas structure of the cloud very well. The map tracing the number of the used sources ( $N$ ) clearly outlines the regions of high extinction. In contrast, the flat corrected source density map indicates a reliable substructure correction with NICEST. Only very localized prominent features are visible which are caused by parts in the VISTA image and the source catalog where the correction factor is not applicable. This can be either due to extremely bright stars with halos on the images ( $\rho_{\text{norm}} \sim 0$ ), or due to stellar clusters in cloud ( $\rho_{\text{norm}} \gtrsim 0.2$ ). For such regions a density correction based on the expected number of sources not valid and leads to the residual pattern across the density map.

### 5.3. *Herschel-Planck* re-calibration

The final maximum resolution of our extinction map of  $1'$  is still well above the resolution of the dust emission measurements by *Herschel* ( $\sim 36''$ ). Thus, for the case of Orion A, it is practical to use the higher-resolution *Herschel* dust thermal emission measurements and convert them to column-density estimates. This procedure has been performed a number of times in the recent past for various cloud complexes (e.g. Stutz et al. 2010), but here we will use the method introduced by Lombardi et al. (2014b). To convert the far-infrared dust emission to column-density estimates, the authors used *Herschel* and *Planck* data to estimate the dust optical depth at 850  $\mu\text{m}$  ( $\tau_{850}$ ) in the cloud by fitting modified black bodies to each pixel in the various channel maps. In a subsequent step, the resulting map of dust optical depth was converted to extinction (and therefore column-density) by cross-calibration with an extinction map. Here, the assumption was made that the dust optical depth and extinction share a linear

correlation described by

$$A_K = \gamma \tau_{850} + \delta. \quad (28)$$

For a reliable fit, an upper limit in the parameters is necessary, because a linear correlation is expected to be only valid up to the point where the extinction map is unbiased. With typical 2MASS extinction maps this limit can be set to  $\tau_{850} = 2 \cdot 10^4$  (Lombardi et al. 2014b; Zari et al. 2016). With our new extinction map, which features a much larger dynamic range, this limit is expected to be pushed to higher values.

For fitting the relation we also use the MCMC-based method outlined in Sect. 3.1. In this case our prior function (Equ. 18) takes the form

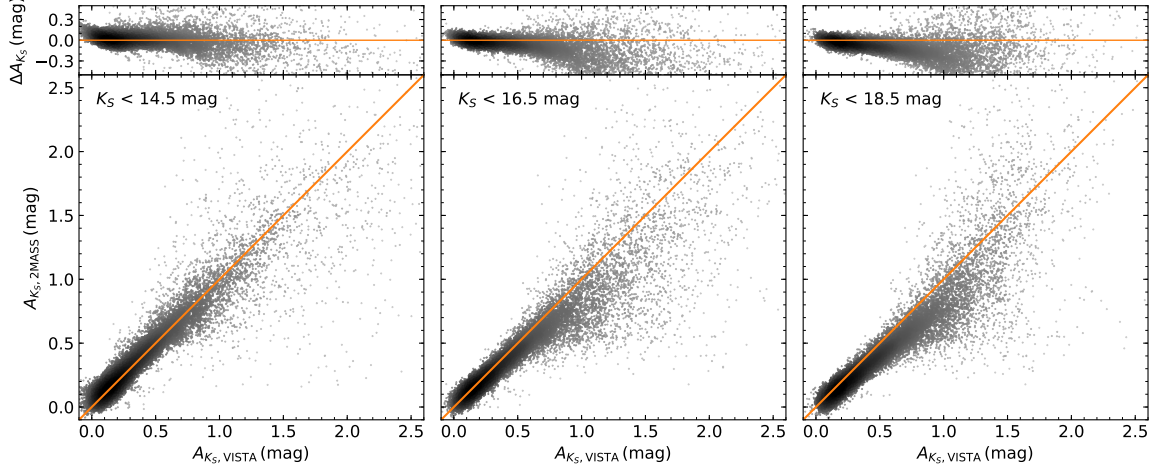
$$\ln P(\delta, \gamma) = \begin{cases} 0 & -1 \leq \delta \leq 1 \wedge 0 \leq \gamma \leq 10^4 \\ -\infty & \text{otherwise} \end{cases}, \quad (29)$$

where now  $\delta$  is the intercept, and  $\gamma$  the slope. In this case, however, we are not able to estimate additional systematic uncertainties in the correlation and rely only on the statistical errors derived for the extinction map and the dust optical depth. For this reason, the error in our fit is severely underestimated.

In Fig. 15 we plot the relation between extinction in the  $K_S$  band against the dust optical depth at 850  $\mu\text{m}$  along with our fitting result. Clearly, the linearity of the relation seems to persist beyond the previously adopted value and we fit the entire range up to  $\tau_{850} = 10^5$ . We find a slope of  $\gamma = 3042 \text{ mag}$ , which is about 15% steeper when compared to the previous fit of Lombardi et al. (2014b), who derived  $\gamma = 2640 \text{ mag}$ . In light of additional unknown systematic errors, we round this value to  $\gamma = 3050 \text{ mag}$ .

We now continue to investigate and discuss the origin of the discrepancy in the correlation of dust optical depth and extinction between the results of Lombardi et al. (2014b) and our fit. A detailed comparison of the extinction maps from 2MASS (Lombardi et al. 2011) and our new map reveals the reason. Figure 16 shows a comparison of the pixel values for  $3'$  extinction maps (the resolution of the 2MASS map) created from 2MASS and VISTA data, respectively. This comparison is shown for three different magnitude limits, in the sense of only including sources brighter than the given limit to create the map. For the typical 2MASS sensitivity limit of 14.5 mag in  $K_S$ , the map values share a 1:1 correlation. However, when including more sources in the VISTA map (2MASS tops out at  $\sim 14.5 \text{ mag}$ ), a clear trend to higher extinctions in each pixel becomes visible. The most likely reason for this is that individual pixels have significant cloud substructure which only gets sampled when enough background sources are available. In the case of the typical 2MASS magnitude limit, only relatively few background sources are available per pixel. Thus, structure with a size smaller than the resolution of the map is partly not being sampled. In contrast, the VISTA data offer an order of magnitude more background sources which significantly increase the sampled area.

The maps, for which we display the pixel values in Fig. 16 have been created using the NICEST substructure correction (Sect. 3.2.3). This correction is in principal designed to minimize such an effect, however, a certain bias will still remain if structures are not sampled by background sources in the first place. It therefore seems likely, that the 2MASS observations were apparently not sensitive enough to provide a full sampling of the cloud substructure. This trend continues for even higher magnitude limits but changes become less significant when approaching the sensitivity limit of the VISTA data ( $K_S \sim 19 \text{ mag}$ ). Still, we can not say



**Fig. 16.** Comparison of the pixel values for extinction maps created with 2MASS and VISTA data (sensitivity limits:  $\sim 14.5$  mag and  $\sim 19$  mag, respectively). Each of the three panel groups shows a different magnitude limit. The leftmost plot shows a comparison for extinction maps where all source brighter than  $14.5$  mag in  $K_S$  are used. The panel in the center and to the right display the same comparison, but with sources brighter than  $16.5$  and  $18.5$  mag in  $K_S$ , respectively. While the bottom figures directly compare the two measures, we plot their difference ( $A_{K_S, 2MASS} - A_{K_S, VISTA}$ ) at the top of each group. The orange lines represent 1:1 correlations between the parameters.

with absolute confidence that the new data sample the substructure up to a point where no more improvements could be found by including even more background sources.

Other values for  $\gamma$  between extinction and dust optical depth in the literature are published for Orion B ( $\sim 3460$  mag; Lombardi et al. 2014b) and for the Perseus cloud complex ( $\sim 3900$  mag; Zari et al. 2016). Also here the authors used 2MASS extinction maps, and both clouds share a similar unfavourable position compared to Orion A with respect to the number of available background sources. Orion B lies at about the same distance as Orion A but is closer to the galactic plane. The Perseus cloud, on the other hand lies significantly closer ( $240$  pc), which reduces the problem with structure not being sampled by background sources. Nevertheless, for these clouds it seems reasonable to assume that they are also affected by this problem to a certain extent. Furthermore, we also can not exclude the possibility that the different calibrations of the *Herschel* dust optical depth measurements are also affecting this relation.

The value of  $\gamma$ , in principle, can be correlated with physical conditions of the dust since it is proportional to the opacity at  $850\mu\text{m}$ , and the extinction coefficient in the  $K_S$  passband ( $\sim 2.2\mu\text{m}$ ). In light of our findings with respect to cloud substructure and the consequential introduced bias in the slope, we conclude, that at the moment it is difficult to make decisive physical interpretations, given a particular value of  $\gamma$ . More studies in regions, where extinction maps convincingly sample the entire cloud structure are necessary to arrive at a better understanding of the slope  $\gamma$ . Similar to what we found with the new VISTA data for Orion A, future observations will provide better insight into this problem. In particular, the VISTA Star Formation Atlas<sup>6</sup> will perform high-sensitivity, wide-field  $JHK_S$  observations of major nearby star-forming complexes. The NIR photometry will enable the construction of high resolution extinction maps,

with well-sampled pixels and therefore will provide further insight into this problem.

## 6. Summary

In this second publication in the context of the Vienna Survey in Orion, we have investigated infrared extinction properties in the Orion A molecular cloud complex. The main results of this paper are as follows.

1. We combined several large-scale photometric databases, covering a wavelength range from just below  $1\mu\text{m}$  to about  $25\mu\text{m}$ . Our master catalog includes NIR data from Pan-STARRS, as well as HAWK-I and the VISTA observations presented in Paper I. The MIR is covered by *Spitzer* IRAC and MIPS photometry. A detailed description of our final catalog is given in Sect. 2.
2. Our results critically depend on the reliability of linear fitting procedures. For this reason, we have developed a two-dimensional Bayesian framework which is used together with a Markov chain Monte Carlo algorithm. Specifically, we have derived a model likelihood which takes measurement errors in both dimensions into account. This framework can be adapted to include more complex models and therefore can be used in a multitude of additional aspects far beyond the general scope of this publication. The final model likelihood is given in Equ. 17, for which a detailed derivation can be found in Sect. 3.1.
3. We derived the infrared extinction law in Orion A by fitting color excess ratios in a variety of color-color diagrams. Here, we first measured the average extinction law (Sect. 4.2), and continued to investigate spatial variations across the molecular cloud complex (Sect. 4.3). Fitting results and the total extinction ratios for all included passbands are available in Tab. 2. The results for the fitting procedures are visualized for the  $H$  passband in Figs. 6 and 7. For the average extinction law we find excellent agreement compared to results in the literature.

<sup>6</sup> VISIONS - The VISTA Star Formation Atlas is an ongoing ESO public survey and is based on this paper series. More information is available at <http://visions.univie.ac.at>.

4. For the passbands up to about  $6\mu\text{m}$  we find statistically significant spatial variations in the color excess ratios of the order of 3%. We investigated the origin of these variations by splitting the survey into distinct sub-regions (Fig. 8). Here, we find a clear signature, that those regions, which seem to be affected by radiative feedback from the massive stars in the ONC, show a different extinction law when compared to other star-forming sites in the cloud (Fig. 10).
5. To support and interpret our results on the spatial variations from the linear fitting procedure, we compared our findings to various model predictions (Fig. 9). Here, we found that the general trends can be reproduced by the models, reinforcing our claim with respect to a variable extinction law. However, a more detailed analysis revealed that these models are not able to fully explain all details of our findings across the entire infrared spectral range.
6. The Orion A MIR extinction law is characteristically flat, similar to many other recently investigated sightlines (Fig. 9). This result contradicts previous investigations in the Orion A region. Furthermore, flat MIR extinction is oftentimes associated with extraordinary large dust grains ( $\sim 10\mu\text{m}$ ). However, a comparison with recently published dust grain models shows that this must not necessarily be the case (Sect. 4.4).
7. Based on the derived average extinction law, we constructed a new extinction map of Orion A with a resolution of  $1'$  (Fig. 14). We compared this map to *Herschel* dust optical depth measurements (Sect. 5.3) and derived a new calibration to convert these data to extinction (Fig. 15). However, we also found alarming evidence, that the calibration of *Herschel* data is likely biased by cloud substructure which is not sampled by background sources (Fig. 16). Another source of systematic uncertainty in this correlation may be the conversion of the dust optical depth from *Herschel* dust emission measurements itself.

Finally, we want to highlight that Orion A presents a unique opportunity to study the extinction law in an unbiased way. Many well-known results in the literature are based on measurements in the galactic plane and specifically toward the galactic center. In these regions, however, it is difficult to interpret the extinction law and its variations, since many physically separate layers of interstellar material can overlap along the line of sight. Therefore, and in light of our findings on the variations in the extinction law, we argue that it is extremely difficult to find such small deviations (3%) in these regions. Moreover, since Orion A hosts the nearest massive star-forming region, it is the ideal location for detailed investigations into dust grain populations and their generic properties under the influence of radiative feedback.

**Acknowledgements.** Stefan Meingast is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute for Astrophysics, University of Vienna. Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO programmes 082.C-0032(A) and 090.C-0797(A). This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013). This research has made use of "Aladin sky atlas" developed at CDS, Strasbourg Observatory, France (Bonnarel et al. 2000). This research has made use of the SVO Filter Profile Service (<http://svo2.cab.inta-csic.es/theory/fps/>) supported from the Spanish MINECO through grant AyA2014-55216. We also acknowledge the various Python packages that were used in the data analysis of this work, including NumPy (Van Der Walt et al. 2011), SciPy (Jones et al. 2001), scikit-learn (Pedregosa et al. 2011), and Matplotlib (Hunter 2007).

## References

- Alves, J., Lada, C. J., Lada, E. A., Kenyon, S. J., & Phelps, R. 1998, *ApJ*, 506, 292
- Alves, J., Lombardi, M., & Lada, C. J. 2014, *A&A*, 565, A18
- Alves, J. F., Lada, C. J., & Lada, E. A. 2001, *Nature*, 409, 159
- Ascenso, J., Lada, C. J., Alves, J., Román-Zúñiga, C. G., & Lombardi, M. 2013, *A&A*, 549, A135
- Ascenso, J., Lombardi, M., Lada, C. J., & Alves, J. 2012, *A&A*, 540, A139
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Beitia-Antero, L. & Gómez de Castro, A. I. 2017, *ArXiv e-prints* [arXiv:1704.06059]
- Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, *PASP*, 115, 953
- Bernstein, R. A., Freedman, W. L., & Madore, B. F. 2002, *ApJ*, 571, 107
- Bok, B. J. & Cordwell, C. S. 1973, in *Molecules in the Galactic Environment*, ed. M. A. Gordon & L. E. Snyder, 54
- Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, 143, 33
- Bouy, H., Alves, J., Bertin, E., Sarro, L. M., & Barrado, D. 2014, *A&A*, 564, A29
- Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, *Ann. Appl. Stat.*, 5, 1657
- Bradley, J., Dai, Z. R., Erni, R., et al. 2005, *Science*, 307, 244
- Cambrésy, L. 1999, *A&A*, 345, 965
- Cambrésy, L., Beichman, C. A., Jarrett, T. H., & Cutri, R. M. 2002, *AJ*, 123, 2559
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, *ArXiv e-prints* [arXiv:1612.05560]
- Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
- Clutton-Brock, M. 1967, *Technometrics*, 9, 261
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, *VizieR Online Data Catalog*, 2328
- Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, in *Proc. SPIE*, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 62690X
- Dobashi, K., Bernard, J.-P., Hughes, A., et al. 2008, *A&A*, 484, 205
- Dobashi, K., Uehara, H., Kandori, R., et al. 2005, *PASJ*, 57, S1
- Draine, B. T. 1989, in *ESA Special Publication*, Vol. 290, *Infrared Spectroscopy in Astronomy*, ed. E. Böhm-Vitense
- Draine, B. T. 2003, *ARA&A*, 41, 241
- Draine, B. T. & Lee, H. M. 1984, *ApJ*, 285, 89
- Drass, H., Haas, M., Chini, R., et al. 2016, *MNRAS*, 461, 1734
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, 142, 72
- Emerson, J., McPherson, A., & Sutherland, W. 2006, *The Messenger*, 126, 41
- Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, 154, 10
- Fitzpatrick, E. L. & Massa, D. 1986, *ApJ*, 307, 286
- Fitzpatrick, E. L. & Massa, D. 1988, *ApJ*, 328, 734
- Flaherty, K. M., Pipher, J. L., Megeath, S. T., et al. 2007, *ApJ*, 663, 1069
- Flwelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2016, *ArXiv e-prints* [arXiv:1612.05243]
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Fritz, T. K., Gillessen, S., Dodds-Eden, K., et al. 2011, *ApJ*, 737, 73
- Gosling, A. J., Bandyopadhyay, R. M., & Blundell, K. M. 2009, *MNRAS*, 394, 2247
- Hirashita, H. & Voshchinnikov, N. V. 2014, *MNRAS*, 437, 1636
- Hogg, D. W., Bovy, J., & Lang, D. 2010, *ArXiv e-prints* [arXiv:1008.4686]
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, *ApJ*, 619, 931
- Ivezić, Z., Connelly, A. J., VanderPlas, J. T., & Gray, A. 2014, *Statistics, Data Mining, and Machine Learning in Astronomy* (Princeton University Press)
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python*, [Online; accessed 2017-04-20]
- Juvela, M. & Montillaud, J. 2016, *A&A*, 585, A78
- Kissler-Patig, M., Pirard, J.-F., Casali, M., et al. 2008, *A&A*, 491, 941
- Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, *ApJ*, 429, 694
- Lombardi, M. 2009, *A&A*, 493, 735
- Lombardi, M. & Alves, J. 2001, *A&A*, 377, 1023
- Lombardi, M., Alves, J., & Lada, C. J. 2011, *A&A*, 535, A16
- Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014a, *A&A*, 568, C1
- Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014b, *A&A*, 566, A45
- Lutz, D. 1999, in *ESA Special Publication*, Vol. 427, *The Universe as Seen by ISO*, ed. P. Cox & M. Kessler, 623
- Lutz, D., Feuchtgruber, H., Genzel, R., et al. 1996, *A&A*, 315, L269
- Mackay, D. J. C. 2003, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press), 640
- Magnier, E. A., Schlafly, E. F., Finkbeiner, D. P., et al. 2016, *ArXiv e-prints* [arXiv:1612.05242]
- Majaess, D., Turner, D., Dékány, I., Minniti, D., & Gieren, W. 2016, *A&A*, 593, A124

- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, 739, 25
- Martin, P. G. & Whittet, D. C. B. 1990, *ApJ*, 357, 113
- Mathis, J. S. 1994, *ApJ*, 422, 176
- Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, *AJ*, 144, 192
- Meingast, S., Alves, J., Mardones, D., et al. 2016, *A&A*, 587, A153
- Meingast, S., Lombardi, M., & Alves, J. 2017, *ArXiv e-prints* [arXiv:1702.08456]
- Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, *A&A*, 474, 515
- Nataf, D. M. 2015, *MNRAS*, 449, 1171
- Nataf, D. M., Gonzalez, O. A., Casagrande, L., et al. 2016, *MNRAS*, 456, 2692
- Nishiyama, S., Nagata, T., Kusakabe, N., et al. 2006, *ApJ*, 638, 839
- Nishiyama, S., Tamura, M., Hatano, H., et al. 2009, *ApJ*, 696, 1407
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Rieke, G. H. & Lebofsky, M. J. 1985, *ApJ*, 288, 618
- Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, 154, 25
- Rosenthal, D., Bertoldi, F., & Drapatz, S. 2000, *A&A*, 356, 705
- Schlafly, E. F. & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlafly, E. F., Finkbeiner, D. P., Schlegel, D. J., et al. 2010, *ApJ*, 725, 1175
- Schwarz, G. 1978, *Annals of Statistics*, 6, 461
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Stead, J. J. & Hoare, M. G. 2009, *MNRAS*, 400, 731
- Stutz, A., Launhardt, R., Linz, H., et al. 2010, *A&A*, 518, L87
- Taylor, M. B. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *ArXiv e-prints* [arXiv:1102.1523]
- Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in *Conference on Intelligent Data Understanding (CIDU)*, 47–54
- Voshchinnikov, N. V., Henning, T., & Il'in, V. B. 2017, *ApJ*, 837, 25
- Wang, S., Gao, J., Jiang, B. W., Li, A., & Chen, Y. 2013, *ApJ*, 773, 30
- Wang, S. & Jiang, B. W. 2014, *ApJ*, 788, L12
- Wang, S., Li, A., & Jiang, B. W. 2015, *ApJ*, 811, 38
- Weingartner, J. C. & Draine, B. T. 2001, *ApJ*, 548, 296
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, 154, 1
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- Zari, E., Lombardi, M., Alves, J., Lada, C. J., & Bouy, H. 2016, *A&A*, 587, A106
- Zasowski, G., Majewski, S. R., Indebetouw, R., et al. 2009, *ApJ*, 707, 510



# Conclusions & Outlook

In the final chapter of this thesis I provide a concise summary of the major results of my thesis work. Moreover, large parts of this section are dedicated to follow-up projects with particular emphasis on how to further improve our understanding of nearby young stellar systems and currently ongoing star formation in molecular clouds within 500 pc.

## 5.1 Summary of results

In each of the three preceding chapters I presented a first author publication with a focus on infrared extinction in the Orion A molecular cloud. Below, I provide a list of the most important results across all three manuscripts. The following summary concentrates on the essential points with respect to the original thesis goals as given in Sect. 1.2.

1. Large-scale NIR ESO VISTA observations have been conducted of the star-forming Orion A molecular cloud complex. This survey covers  $\sim 18 \text{ deg}^2$  in the NIR bands  $J$ ,  $H$ , and  $K_S$  and constitutes the basis of the Vienna Survey in Orion (VISION) project and is furthermore the foundation of all results in this thesis. The telescope and imaging systems are described in Sect. 1.2.1, while a detailed account of the observations themselves is given in Sect. 2.4.
2. Initial investigations of the raw telescope data showed that the data reduction pipeline environment, employed by ESO, delivers less-than-ideal quality of the reduced data products. Most notably, the image quality in the assembled tiles is significantly degraded during reduction, but also biases in the photometry with respect to extinction measurements are introduced. A detailed account of the discovered problems is given in Sect. 2.5.1.
3. To avoid the issues with the original data reduction software, a completely independent pipeline environment has been developed. Details on all individual processing steps are given in Sect. 2.5. This new software package

is currently not publicly available because it is being further developed for follow-up observations at the moment.

4. The new pipeline environment has been used to reduce the ESO VISTA observations to produce the so far most comprehensive database of NIR photometry of the Orion A molecular cloud. An account of data products is given in Sect. 2.6. In addition to the survey data products, I also performed a brief analysis of the YSO population of the cloud (Sect. 2.7) to demonstrate the data quality, where I also identified previously unknown YSO candidates.
5. In Sect. 3.3 and 3.4, I demonstrated that traditional techniques to measure line of sight extinction fail to produce satisfactory results when used together with deep photometric data. The main reason for this issue is the detection of significant amounts of extragalactic objects, which were not taken into account in the design of most previous methods. This called for the development of a new method to calculate line of sight extinction toward observed sources.
6. The new method, named PNICER in reference to the original techniques (NICER, NICEST), is presented in Sect. 3.5, where also the theoretical background and practical examples are given. PNICER requires photometry from multiple passbands, as well as similar observations of a nearby unextincted control field. Interstellar extinction is estimated by fitting Gaussian mixture models along the extinction vector to determine a source's intrinsic photometric properties. Section 3.6 validates PNICER based on a comparison with results from the literature. Overall, PNICER delivers more reliable results than traditional techniques in regions of high column-densities and is capable of de-reddening millions of sources within a matter of seconds. The method is publicly available for download as a Python package and will be the backbone for future extinction measurements with upcoming NIR surveys.
7. The Orion A NIR photometry, together with the PNICER method, was used for an analysis of infrared extinction in the molecular cloud complex. In a first step, I introduced a linear fitting framework to enable robust calculations of the infrared extinction law via color excess ratios (Sect. 4.5.1). By fitting a series of color-color diagrams, I derived the infrared extinction law in the Orion A molecular cloud from 1 to  $\sim 25\ \mu\text{m}$  (Sect. 4.6).

8. The average extinction law for Orion A (Sect. 4.6.2) is generally in very good agreement with results from the literature. The MIR is characterized by flat extinction, similar to many other recently observed sightlines in the Galaxy.
9. By repeating the fitting procedure for several sub-regions of the cloud, I discovered that the infrared extinction law is variable across the cloud (Sect. 4.6.3). The variability of the extinction law persists from 1 to at least  $\sim 6\ \mu\text{m}$  and amounts to only  $\sim 3\%$ . The fact that the extinction law varies also in the NIR is the first clear evidence of such a characteristic. In contrast to my findings, previous results typically use a variety of different sightlines, leading to inhomogeneous samples and contradicting results (see Sect. 4.3 for an account).
10. While the general trend in the variation of the extinction law follows model predictions, a detailed investigation of the physical origin (e.g. different grain sizes) remains difficult (Sect. 4.6.4). This is because the observed trends could not be consistently explained within a single theoretical framework. For example, models with exceptionally large dust grains can reproduce the flat MIR extinction just as well as a smaller grain size distribution from another model which uses different grain structures and compositions. However, the origin of the variations seems to be associated with the massive cluster stars. Regions which apparently are affected by radiative feedback systematically show a different extinction law compared to more quiescent parts of the cloud.
11. Based on the average extinction law in the cloud, I constructed a new high-resolution extinction map of Orion A (Sect. 4.7). This map reproduces the spatial structures observed in dust emission very well and improves previous results in resolution and sensitivity with respect to high column-densities. A detailed comparison of dust extinction to dust optical depth revealed biases in the calibration of column-density maps, which can have a significant impact on both already published and future results. While extinction and optical depth still correlate linearly, a clear trend with sampling becomes visible. It therefore seems that gas mass estimates based on dust emission data can be significantly underestimated.

The above mentioned findings demonstrate that the entire Orion A molecular cloud remains one of the most interesting, yet still puzzling, objects to study the earliest stages of star formation. Follow-up work is necessary to address remaining

questions and investigate newly emerged issues with respect to the findings of my thesis work. In particular, the cluster stars should be confirmed as the origin of the variable infrared extinction law, and also the specific physical differences of the dust grain population across the cloud should be investigated in more detail. One approach here would be to take NIR and MIR spectra of well-characterized background sources and employ a spectroscopic comparison of extincted sources across the cloud complex. One problem, however, remains in the fact that the variations of the infrared extinction law are only at the percent level. A pilot study would be necessary to clarify whether such small differences are even measurable with an analysis based on spectroscopy. A spectroscopic investigation in the MIR would also enable a detailed study of particular extinction features and their variations, such as the  $3.1\ \mu\text{m}$  water ice and the  $9.7\ \mu\text{m}$  silicate feature. Moreover, the posed questions, together with the developed methods presented in this thesis can, in principle, be applied to all other nearby star-forming regions. For example, an investigation of the extinction law and the cross-calibration with dust emission data (via optical depth) in other nearby clouds would provide a solid, homogeneous foundation to further investigate grain properties.

The results of this thesis were only made possible by observing Orion A with state-of-the-art instrumentation which provide high-quality photometric data. Among the nearby molecular clouds, similar observations are currently only available for the Orion B cloud (Spezzi et al. 2015). For this reason, it would be highly desirable to have access to similar photometry with comparable or better quality for other nearby molecular clouds. This specific follow-up work has already been put into action and I dedicate the last section of my thesis to this new project which runs under the acronym VISIONS.

## 5.2 VISIONS - VISTA Star Formation Atlas

While working on my thesis projects, the research group in Vienna, and in particular myself, have developed extensive skills and expertise regarding large-scale imaging observations and precision photometry in the NIR. In 2015, ESO published a call for Letters of Intent for research groups to declare interest in leading new projects, utilizing the VISTA telescope together with VIRCAM for a new generation of public surveys. These public surveys are very large programs which last over several years and usually are based on international collaborations. Their aim is to provide



(publicly available) critical data to the astronomical community to address a wide variety of research topics. The procedures specified by ESO dictated that the project would only be accepted after passing two additional review stages, where first a full proposal was to be submitted, followed by a detailed Survey Management Plan. Only after passing all three stages, the survey would finally be implemented where the entire procedure up to the final decision typically takes more than one year and includes several reviews and iterations by different panels.

With the newly developed experience from the work presented in Chapter 2, a team, consisting of about 30 collaborators and led by the research group in Vienna, successfully applied for ~550 hours of observing time under the program name VISIONS - VISTA Star Formation Atlas. This name also reflects the intimate connection to the Vienna Survey in Orion (VISION). The observations commenced in April 2017 and are scheduled to be completed after three years by April 2020. VISIONS is a direct outcome of my thesis project and will be the main focus of my work over the next few years. In this section, I will first briefly outline my own contributions to the project and then continue to discuss the scientific motivation, the survey's objectives, and the implementation.

### 5.2.1 Own contribution

I was involved in VISIONS from the very beginning. The main motivation to start this project was the successful VISION program, based on observations of the Orion A molecular cloud with VISTA as presented in Chapter 2. Specifically, the developed technical expertise was fundamental to the eventual success of the project. While João Alves (thesis advisor) is the principal investigator of the program, I am part of the survey management team as Co-PI. Specifically, I have formulated some of the final science cases, for which I will also take a leading role in the future when the first reduced data arrive. Moreover, I was responsible for the design of the entire survey strategy and now continue to coordinate the survey team during the period of the observations. Also, I have written major parts of the Phase 1 proposal and essentially the entire Survey Management Plan<sup>1</sup>. Furthermore, I note that much of the presented information in the next sections is also available in either of these two documents to which I have contributed significantly.

---

<sup>1</sup>The VISIONS proposal is currently not publicly available, but will be added to the VISIONS homepage ([visions.univie.ac.at](https://visions.univie.ac.at)) at a later stage. The Survey Management Plan can be downloaded from ESO's public survey website <https://www.eso.org/sci/observing/PublicSurveys.html>

### 5.2.2 Motivation

During the past decade a tremendous wealth of new data has been provided to the astronomical community, most notably from well-known space observatories such as *Spitzer* (Werner et al. 2004), *WISE* (Wright et al. 2010), *Herschel* (Pilbratt et al. 2010), and *Planck* (Tauber et al. 2010; Planck Collaboration et al. 2011a). These missions provided measurements in the MIR and submillimeter spectral ranges where certain stages of the formation of a new star can be observed directly. The VISIONS project represents the complement to these surveys in the NIR, providing deep, wide-field, high image quality (seeing  $\lesssim 1''$ ) ground-based observations. The observed regions encompass the well-known star-forming complexes of Ophiuchus, Lupus, the Pipe Nebula, Corona Australis, Chamaeleon, and Orion for a total on-sky coverage of more than 500 deg<sup>2</sup>.

In general, the survey was designed, on the one hand, to provide state-of-the-art NIR photometry for all nearby star-forming regions, which are accessible from the southern hemisphere, similar to the data presented in this thesis. On the other hand, the observations also aim at opening a new parameter space for all stars in the observed regions: proper motions. In this respect, VISIONS will be fully complementary to upcoming Gaia data releases (Gaia Collaboration et al. 2016). While Gaia will map the 3D distribution and space motions for the nearby young, optically revealed populations with distances to better than 1%, this new public survey will enable measurements of proper motions for embedded sources, as well as low-mass objects. Such sources are difficult to detect (or not visible at all) with Gaia because they are intrinsically faint at optical wavelengths. Thus, VISIONS and Gaia will together allow the construction of 3D maps of the nearby star-forming interstellar medium.

The VISIONS program will monitor large fields around nearby star-forming regions twice per year over the duration of at least three years, providing a total of six position measurements. Together with information from previous VISTA public survey programs, the baseline for measuring proper motions can be extended to about 8 – 10 years for large parts of the observed regions. This combination will allow precision proper motion measurements down to only a few mas/yr (see e.g., Bouy et al. 2013), or equivalent transverse velocities of 0.1 – 0.3 km/s. This information can then be connected with a variety of existing infrared spectroscopic survey programs and will eventually enable the construction of 3D space velocities for the young stellar population. Notable complementary spectroscopic surveys are RAVE (e.g. Steinmetz et al. 2006), APOGEE (e.g. Wilson et al. 2010), or the

Gaia-ESO spectroscopic survey (Gilmore et al. 2012). Moreover, many other science cases have been developed to maximize the efficiency of the survey’s data products, which include a series of fundamental ISM and star formation topics.

VISIONS will also allow to extend the investigation with respect to variations in the NIR extinction law, as demonstrated in the publication presented in Chapter 4 in this thesis. This will have a critical impact on our understanding of interstellar dust and will carry important consequences, for example, for observational cosmology (e.g. Nataf 2015), where often a universal reddening law is assumed. VISIONS will provide a robust and homogeneous framework for the NIR extinction law where many hypotheses can be tested on large scales in a consistent manner. Moreover, by combining submillimeter dust opacity measurements with NIR extinction, it becomes possible to identify regions where the dust population shows signs of grain growth (e.g., Forbrich et al. 2015).

### 5.2.3 Objectives

In addition to the above mentioned science cases of YSO dynamics, dust properties, and NIR extinction, several other science cases have been developed for VISIONS. The most prominent of these are listed below.

1. **IMF.** The VISIONS observations will detect objects down to a few  $M_{\text{Jup}}$  within 500 pc for the typical age of the nearby star-forming regions. Thus, it will be possible to construct the entire stellar and sub-stellar inventory of these regions and to identify the product of star formation in these cloud complexes. This will enable the investigation of regional IMF variations and the impact of massive YSOs on the local star formation.
2. **CMF/IMF relation.** Based on the development presented in Chapters 3 and 4, VISIONS will allow the construction of core mass functions (CMF) for the nearby star-forming regions. In contrast to other popular methods, dust extinction techniques are not affected by systematic errors due to assumptions about dust temperature and emissivity, or molecular depletion (see Sect. 1.1.1). Furthermore, these maps will enable a detailed comparison of dense cores extracted from archival molecular line and dust emission maps to study the physical and chemical properties of the core population. The NIR observations will be complemented by *Spitzer* and *WISE* data in regions of high column-density, similar to the extinction map presented in Sect. 4.7.2.

The extinction maps are expected to reach a resolution of 10-15'' toward the galactic bulge and 1' toward the galactic anti-center (for an example, see the Orion A extinction map displayed in Fig. 4.13). In particular, this analysis will also allow to investigate a connection between the dense gas and the young stellar population in these regions (e.g., [Alves et al. 2007](#)) and compare properties of starless and non-starless cores.

3. **Star formation rates and efficiencies.** Extinction maps, together with calibrated dust emission data ([Lombardi et al. 2014](#)) will become the reference for measuring the distribution of gas mass in the cloud complexes. Together with accurate and mostly complete YSO catalogs, these will enable to establish the Kennicutt-Schmidt relation ([Schmidt 1959](#); [Kennicutt 1998](#)) down to the dense core size for each cloud separately. Moreover, it will become possible to further investigate hypotheses, such as the existence of a star formation threshold ([Lada et al. 2010](#); [Lewis and Lada 2016](#)).
4. **YSO identification and characterization.** The VISTA data reduction techniques presented in Chapter 2 of this thesis have a critical impact on the image quality of the data products. The increase in spatial resolution is essential to better understand the nature of young stars ([Großschedl et al. in prep.](#)). In particular, scattered light around nearby young stars can be a powerful indicator of the nature of these objects, oftentimes also allowing to put constraints on disk inclination angles. Such information can be used to constrain YSO models (with e.g., [Robitaille et al. 2006, 2007](#)).
5. **Molecular cloud structure down to au scales.** The high-resolution extinction maps created from VISIONS data will reach a resolution of 2000 au at a distance of 130 pc. This resolution is more than twice higher than what can be achieved using *Herschel* emission data, which are typically limited to the SPIRE-500 resolution ( $\sim 36''$ ). A view of entire molecular clouds at this unprecedented resolution will reveal smaller structures than ever before. These maps will therefore contribute to our understanding of the internal structure of molecular clouds.
6. **Cluster formation and evolution.** The evolution of stellar clusters is believed to be determined at their earliest age. In particular, the fraction of mass contained in stars relative to the gas mass, and the timescales for gas dissipation are critical parameters which determine their eventual fate. VISIONS will allow to construct a complete dynamical profile of young stellar



groups and clusters in the solar neighbourhood and will therefore enable to constrain model predictions on cluster evolution.

7. **Wide multiple systems.** Wide YSO binaries and higher order multiples (separations  $\gg 100$  au) are difficult to identify and so far there is no comprehensive study of young wide multiple systems of entire star-forming complexes. The seeing-limited VISIONS observations will allow to identify such systems for all observed regions similar to the study of [Kounkel et al. \(2016\)](#). This will allow, for the first time, to study properties of wide multiple systems in various environmental conditions for complete star-forming complexes. Furthermore, it will also become possible to compare properties of such systems at various stages in the lifetime of YSOs.

## 5.2.4 Survey Design & Observing strategy

To achieve the above mentioned scientific goals, the VISIONS survey was divided into three major categories.

- The majority of the observations belong to the *wide* sub-survey. These cover the dispersed young (and low-mass) population in wide fields around all major star-forming complexes accessible from the southern hemisphere: Ophiuchus, Lupus, the Pipe Nebula, Corona Australis, Chamaeleon, and Orion. In addition, these regions will be covered in a total of six epochs over the duration of three years which will facilitate measurements of proper motions.
- *Deep* observations of the parts of the cloud complexes which are associated with large column-densities ( $A_V > 5$  mag). Here, it is critical to have longer effective integration times to not only recover the embedded young stellar population, but also to penetrate the cloud to measure background sources for extinction studies.
- *Control* field observations measure the typical galactic field population (and galaxy contamination) for statistical comparison. For an example of their importance, with respect to statistical analyses, see Chapters 3 and 4 in this thesis.

Figure 5.1 shows the final layout for the *wide* and *deep* survey regions. The *deep* fields for the largest cloud complex, Orion, has already been observed (Chapter 2). While the data for Orion A are part of this thesis, similar observations for the

Orion B molecular cloud have been conducted during VISTA science verification. Table 5.1 lists a simplified observing setup for all scheduled regions, as well as the requested time. In general, a 1' overlap is included for the tile area that is observed twice when adjacent fields are available.

To facilitate an easier visualization of the sub-surveys, we have also created an interactive display of the observed area for VISIONS. For this purpose we offer an easily accessible script which can be loaded in the Aladin Sky Atlas (Bonnarel et al. 2000)<sup>2</sup>. After starting the software, interested users can load the script by simply typing

```
load https://www.univie.ac.at/alveslab/VISIONS/coverage.ajs
```

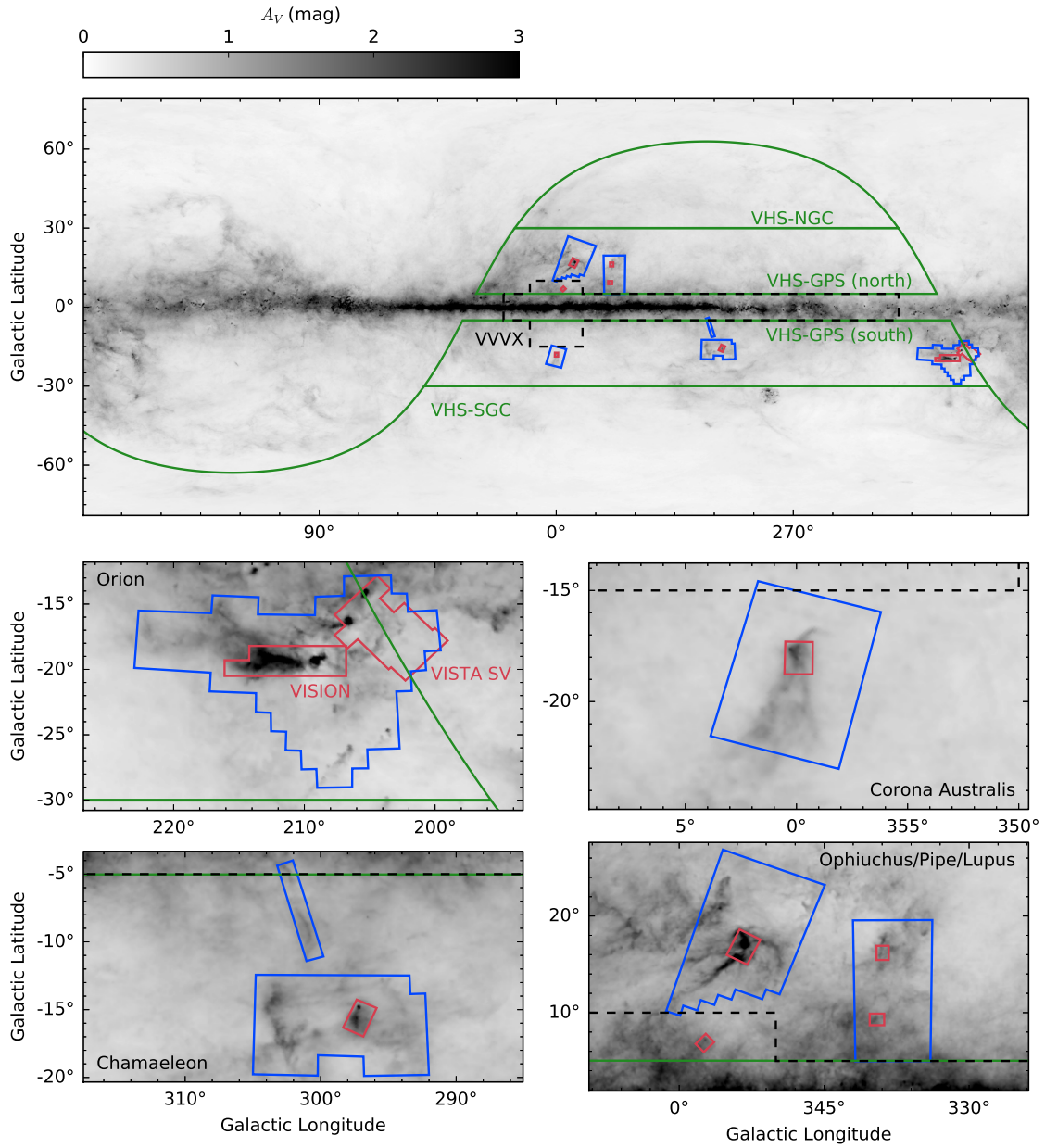
in the command line bar. After waiting a few seconds to load the script, a similar view as in Fig. 5.1 is displayed. Each sub-survey will now be discussed in more detail.

As already mentioned above, the goal of the *wide* sub-survey is to provide large-scale NIR photometry of nearby star-forming regions to facilitate measurements of proper motions. Among our target regions, only the Pipe nebula will not be observed in this mode since a parallel VISTA public survey (VVX<sup>3</sup>) will cover this region in multiple epochs. The survey coverage was chosen on the one hand to cover large parts of the high column-density parts of the clouds. On the other hand, the *wide* observation also allow to observe the more dispersed, possibly optically revealed, but faint, young stellar population.

All regions have already been observed by the previous generation of VISTA public surveys. In particular the VISTA Hemisphere survey, (VHS, McMahon et al. 2013) will provide another epoch of position measurements and will allow to extend the baseline for all data up to 10 years. Moreover, VHS already covered all target regions in single epoch  $J$  and  $K_S$  band observations. For VISIONS, we therefore chose to collect data in the  $H$  band for a complete set of NIR photometry. This will furthermore improve the quality of extinction measurements since here it is especially critical to have access to as many passbands as possible (see chapter 3). Similar to VHS, the *wide* observation will feature an effective integration time of 60 s per pixel, specifically we use  $DIT = 3$  s,  $NDIT = 2$ ,  $NJITTER = 5$  (most pixels are covered twice for a full tile). This setup saturates sources brighter than  $H \approx 11.7$  mag with an expected completeness down to  $H \approx 19$  mag.

<sup>2</sup>Available for download via <http://aladin.u-strasbg.fr>.

<sup>3</sup>Information available at <https://vvvsurvey.org>.



**Figure 5.1.** VISIONS on-sky coverage on top of the [Schlegel et al. \(1998\)](#) all-sky reddening map. Blue solid lines depict the *wide* survey coverage and red lines the *deep* observations. While the top panel shows the global distribution of the target cloud complexes, the bottom four panels are close-up views of the individual observed regions. The Orion A observations (Chapter 2) are marked with the VISION label. Orion B was observed during VISTA science verification (VISTA SV label). Furthermore, the coverage of other public VISTA surveys is shown with green solid lines (VHS, previous program) and black dashed lines (VVVX, current public survey cycle). This figure has been adopted from the VISIONS survey management plan.

**Table 5.1.** VISIONS observation itemized for all sub-surveys. This table has been (partly) adopted from the survey management plan.

Region	Band	Tiles (#)	Area (deg <sup>2</sup> )	Exptime (s)	OB time (hh:mm:ss)	Total time (hh:mm:ss)
<i>Wide</i>						
Ophiuchus	<i>H</i>	93	139.6	60	-	20:43:56
Lupus	<i>H</i>	70	105.1	60	-	15:35:40
Corona Australis	<i>H</i>	25	37.5	60	-	05:35:00
Chamaeleon	<i>H</i>	56	84.1	60	-	12:28:32
Orion	<i>H</i>	114	171.1	60	-	25:23:48
<i>Deep</i>						
Ophiuchus	<i>J</i>	8	6	300	00:51:06	06:48:48
	<i>H</i>	8	6	300	01:00:06	08:00:48
	<i>K<sub>S</sub></i>	8	6	300	01:00:06	08:00:48
Lupus	<i>J</i>	4	3	300	00:51:06	03:24:24
	<i>H</i>	4	3	300	01:00:06	04:00:24
	<i>K<sub>S</sub></i>	4	3	300	01:00:06	04:00:24
Corona Australis	<i>J</i>	2	1.5	300	00:51:06	01:42:12
	<i>H</i>	2	1.5	300	01:00:06	02:00:12
	<i>K<sub>S</sub></i>	2	1.5	300	01:00:06	02:00:12
Chamaeleon	<i>J</i>	4	3	300	00:51:06	03:24:24
	<i>H</i>	4	3	300	01:00:06	04:00:24
	<i>K<sub>S</sub></i>	4	3	300	01:00:06	04:00:24
Pipe	<i>J</i>	1	1.5	600	00:52:24	00:52:24
	<i>H</i>	1	1.5	600	01:10:24	01:10:24
	<i>K<sub>S</sub></i>	1	1.5	600	01:10:24	01:10:24
<i>Control</i>						
Ophiuchus	<i>J</i>	1	1.5	600	00:52:24	00:52:24
	<i>H</i>	1	1.5	600	01:10:24	01:10:24
	<i>K<sub>S</sub></i>	1	1.5	600	01:10:24	01:10:24
Lupus	<i>J</i>	2	3	600	00:52:24	01:44:48
	<i>H</i>	2	3	600	01:10:24	02:20:48
	<i>K<sub>S</sub></i>	2	3	600	01:10:24	02:20:48
Corona Australis	<i>J</i>	1	1.5	600	00:52:24	00:52:24
	<i>H</i>	1	1.5	600	01:10:24	01:10:24
	<i>K<sub>S</sub></i>	1	1.5	600	01:10:24	01:10:24
Chamaeleon	<i>J</i>	1	1.5	600	00:52:24	00:52:24
	<i>H</i>	1	1.5	600	01:10:24	01:10:24
	<i>K<sub>S</sub></i>	1	1.5	600	01:10:24	01:10:24
Pipe	<i>J</i>	1	1.5	600	00:52:24	00:52:24
	<i>H</i>	1	1.5	600	01:10:24	01:10:24
	<i>K<sub>S</sub></i>	1	1.5	600	01:10:24	01:10:24
<b>Total</b>	-	2223	561.4	-	-	<b>552:37:24</b>



The VISIONS *wide* survey will cover a total of  $537.4 \text{ deg}^2$  for 358 individual pointings. Thus, in the six epochs, a total of 2148 tiles need to be observed. The total observing time for all *wide* fields (including optimization due to concatenations) amounts 478h 41m 36s. In contrast to the *wide* sub-survey, the *deep* VISIONS observation will image smaller regions in a single epoch, but with longer integration times. This includes a full NIR set of  $J$ ,  $H$ , and  $K_S$  data for the high column-density parts of the clouds. Here, the two Orion clouds (Orion A and Orion B) have already been observed. Since these clouds are by far the (both physically and in projected view) largest clouds, the *deep* observations constitute only a small fraction of the requested observing time. Specifically, the central parts of Ophiuchus and Corona Australis, Lupus I and Lupus III, the Chamaeleon I and B59 in the Pipe Nebula will be observed.

The observations for the *deep* fields were planned based on the experience from the Orion A observations. The total effective exposure time of 600s per pixel will be achieved with  $\text{DIT} = 5 \text{ s}$ ,  $\text{NDIT} = 10$ ,  $\text{NJITTER} = 6$  in  $J$ , and  $\text{DIT} = 2 \text{ s}$ ,  $\text{NDIT} = 25$ ,  $\text{NJITTER} = 6$  for the  $H$  and  $K_S$  passbands. Some fields with extended emission also include sky offsets. For this setup, saturation occurs at  $J \approx 12.1 \text{ mag}$ ,  $H \approx 11.3 \text{ mag}$ , and  $K_S \approx 10.6 \text{ mag}$ , while the completeness will reach  $J \approx 21.5 \text{ mag}$ ,  $H \approx 20.5 \text{ mag}$ , and  $K_S \approx 19.5 \text{ mag}$ . The total requested time amounts 54h 36m 36s for 10 individual pointings and 57 VISTA tiles. For statistical comparisons, also *control* fields are scheduled to be observed for each region. These, however, are not shown in Fig. 5.1 since they will be positioned only before the scheduled observing semester. They will be placed in regions of similar galactic latitude and low extinction. The observing strategy is the same as for the *deep* fields. The *control* fields require a total observing time of 19h 19m 12s for 18 individual positions.

## 5.2.5 Project schedule

The execution time of VISIONS totals to 552h 37m 24s. This amount splits up into 478h 41m 36s (87%) for the *wide* sub-survey, 54h 36m 36s (10%) for the *deep* observations, and 19h 19m 12s (3%) for the *control* field observations.

The observations of the six epochs for five star-forming regions, scheduled for the *wide* survey, are distributed over three years. To optimize the efficiency, and at the same time maximize the spacing between epochs, the following schedule is planned

- P99/101/103 (April - September): 2/6 epochs CrA, 1/6 epoch for Chamaeleon, Ophiuchus, and Lupus
- P100/102/104 (Oct - March): 2/6 epochs Orion, 1/6 epoch for Chamaeleon, Ophiuchus, and Lupus,

where the P99-P104 refer to ESO observing semesters (April 2017 - March 2020). The *deep* observations are not subject to a strict time schedule. They have, however, more stringent observing conditions. Since all *deep* regions are best visible during the Chilean summer, they are scheduled for odd semesters and at the same time distributed over the entire survey duration. The *control* fields are each associated with individual *deep* fields and are therefore scheduled to be observed in the same semester as their corresponding *deep* field.

In total three data releases are scheduled over the next years. The first release will happen in October 2018, where photometry for the then observed *deep* and *control* field data products will be made available (similar to the Orion A observation presented in Chapter 2). One year later in October 2019, the photometry for the *wide* survey will be made publicly available in addition to the *deep* and *control* fields. The band-merged multi-epoch catalog including all sources will be made available in the third, and last, data release in October 2020. All data releases will also include photometrically and astrometrically calibrated VISTA images (pawprints and tiles), the complete band-merged source catalogs, as well as various quality control units.

## 5.2.6 Data processing

Three different types of data calibrators are required for the processing of the images. Firstly, for the removal of the instrumental signature a variety of calibration frames are necessary. These include dark frames and flat fields in several different sequences to allow for the correction of the dark current, non-linearity, bad pixels. In addition these are also required to calculate the gain and the readout noise of each detector. Secondly, the photometric calibration will be performed with external standard stars via the 2MASS catalog. Here, the same approach as in Chapter 2 will be used to also apply an illumination correction and calibrate the photometric zero-point. Thirdly, the astrometric calibration will mainly be based on the 2MASS catalog, but will be refined with newly available Gaia position measurements. The above mentioned calibration units only require the calibration

frames which are provided by the standard ESO calibration plan for VIRCAM. For this reason, no special requirements have been defined for VISIONS.

The data reduction cascade follows the procedure as detailed in Sect. 2.5 where all modules have been ported to Python and are collected in the stand-alone package *Astropy*<sup>4</sup>. A schematic overview of the data reduction cascade is displayed in Fig. 5.2. Furthermore, as an example for the data calibration, Fig. 5.3 shows the correction for non-linearity, generated by *Astropy* for all 16 VIRCAM detectors. Details on the individual procedures can be found in the first publication presented in this thesis (Sect. 2.5).

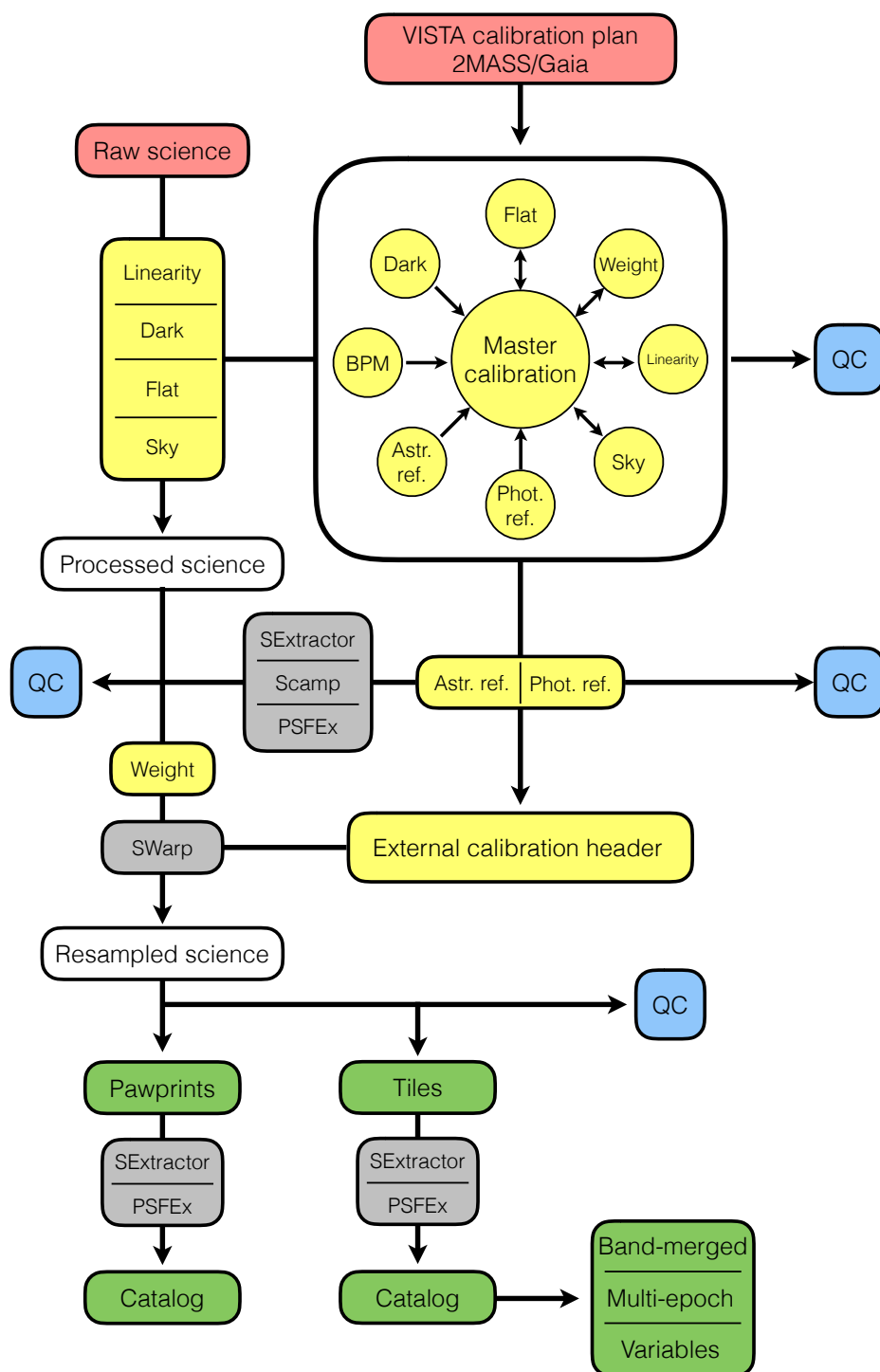
### 5.2.7 Follow-up observations

Like many other public surveys, VISIONS provides an excellent foundation for follow-up observations. The following list includes the most prominent cases for planned future projects.

- In addition to the archival data of the *Herschel* and *Planck* missions, several other data sets on molecular line observations will become available in the future. This includes the only recently completed IRAM 30m project for Orion A (e.g., [Hacar et al. 2017](#)). More observations are planned for the other star-forming regions in the VISIONS sample.
- Existing APOGEE ([Wilson et al. 2010](#)) and future APOGEE-2 ([Majewski et al. 2016](#)) observations will add many more radial velocities to the detected VISIONS sources. This will provide an even more thorough insight into the dynamical properties of young systems. In addition, the future VISTA instrument 4MOST ([de Jong et al. 2012](#)) will be able to provide even more spectroscopic data after VISIONS has been completed.
- The identification of resolved disk-bearing systems will be an excellent baseline for high-resolution follow-up observations with ALMA, SPHERE ([Beuzit et al. 2006](#)), the James Webb Space Telescope ([Gardner et al. 2006](#)), and the ESO Extremely Large Telescope. In particular, the target selection will benefit from information derived from VISIONS, including disk (cavity) orientation and luminosities.

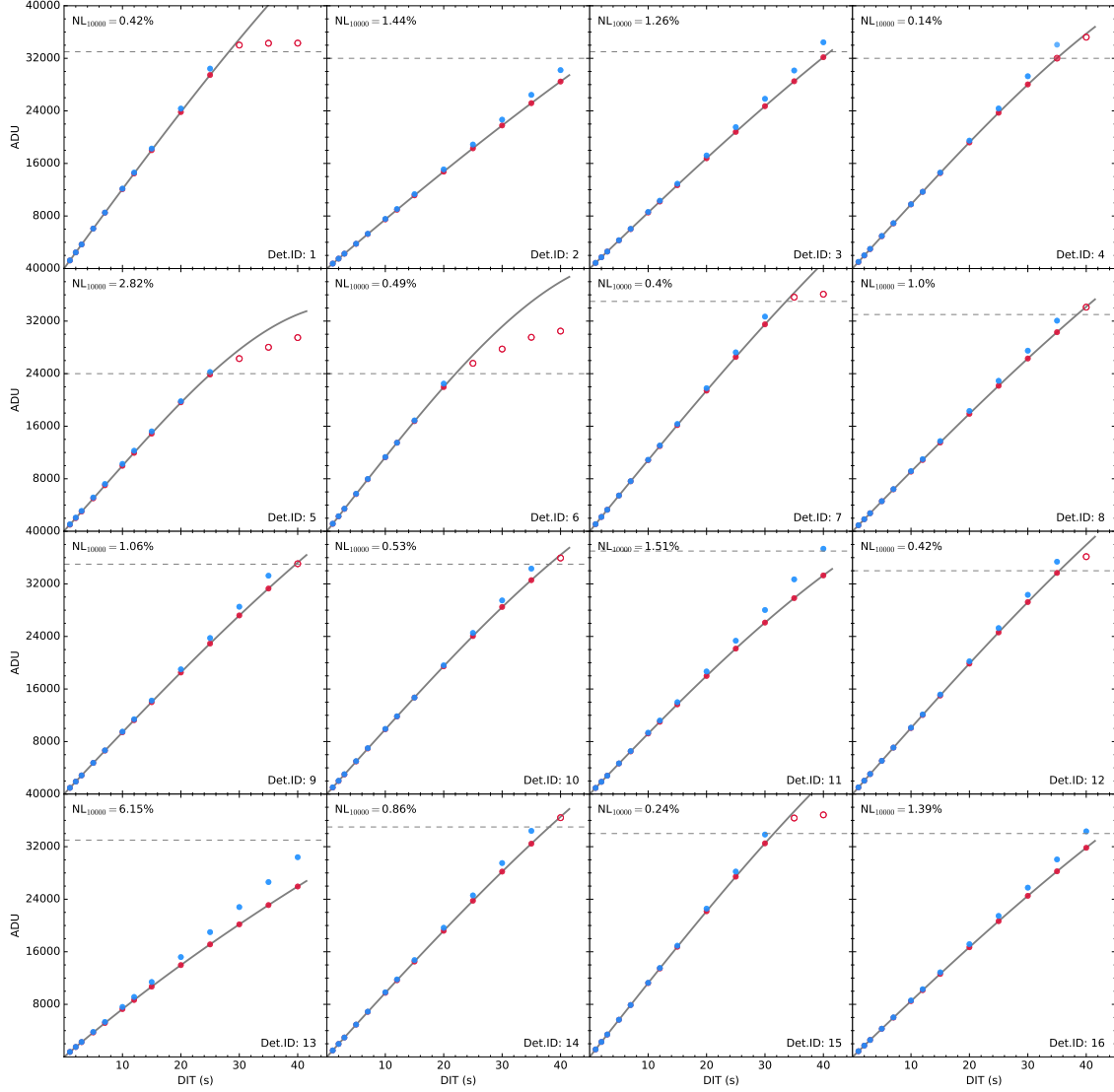
---

<sup>4</sup>The package will be made publicly available on Github in the future at <https://github.com/smeingast/astropy>, when module testing has been completed. At the time of writing, the repository was still private.



**Figure 5.2.** Astropy reduction cascade. Input data is displayed in red, calibration steps in yellow, quality control in blue, intermediate products in white, external software in gray, and data products in green.





**Figure 5.3.** Non-linearity correction for a randomly selected flat-field sequence of VIRCAM from the ESO science archive.

- Similar to the above listed item, newly identified binary and higher-order multiple systems will be a solid foundation for dedicated follow-up with IR space observatories, ALMA, and even AO-supported ELT observations.

## 5.3 Concluding statement

My thesis work on the VISION Orion A project highlights that using new telescope technologies and innovative methods can have a significant impact on our understanding of the process of star formation. This thesis demonstrates that using large survey telescopes is a necessary step forward to reach a full empirical description of star formation in general and molecular clouds and their intrinsic characteristics in particular. By addressing the very specific topic of infrared extinction in molecular clouds, I showed that such new surveys make it necessary to rethink well-established methods and techniques when deriving physical parameters of star-forming regions. In this respect, the VISION Orion A project marks only the beginning of our efforts to use state-of-the-art telescope technology to consistently map and study entire star-forming sites and connect their properties with their surrounding environment. Building on the results of the VISION project in this thesis, follow-up work in the form of the VISIONS collaboration will continue to deliver exciting new results to provide a solid empirical foundation of the physical mechanisms in the nearby star-forming interstellar medium.

# References

- J. Alves and H. Bouy. Orion revisited. I. The massive cluster in front of the Orion nebula cluster. *Astronomy & Astrophysics*, 547:A97, November 2012. doi: 10.1051/0004-6361/201220119. URL <http://adsabs.harvard.edu/abs/2012A%26A...547A..97A>.
- J. Alves, M. Lombardi, and C. J. Lada. The mass function of dense molecular cores and the origin of the IMF. *Astronomy & Astrophysics*, 462:L17–L21, January 2007. doi: 10.1051/0004-6361:20066389. URL <http://adsabs.harvard.edu/abs/2007A%26A...462L..17A>.
- J. F. Alves, C. J. Lada, and E. A. Lada. Internal structure of a cold dark molecular cloud inferred from the extinction of background starlight. *Nature*, 409:159–161, January 2001. URL <http://adsabs.harvard.edu/abs/2001Natur.409..159A>.
- E. E. Barnard. On the dark markings of the sky, with a catalogue of 182 such objects. *Astrophysical Journal*, 49, January 1919. doi: 10.1086/142439. URL <http://adsabs.harvard.edu/abs/1919ApJ....49....1B>.
- N. Bastian, K. R. Covey, and M. R. Meyer. A Universal Stellar Initial Mass Function? A Critical Look at Variations. *Annual Review of Astronomy & Astrophysics*, 48: 339–389, September 2010. doi: 10.1146/annurev-astro-082708-101642. URL <http://adsabs.harvard.edu/abs/2010ARA%26A...48..339B>.
- R. A. Bernstein, W. L. Freedman, and B. F. Madore. The First Detections of the Extragalactic Background Light at 3000, 5500, and 8000 Å. I. Results.

- Astrophysical Journal*, 571:56–84, May 2002. doi: 10.1086/339422. URL <http://adsabs.harvard.edu/abs/2002ApJ...571...56B>.
- J.-L. Beuzit, M. Feldt, K. Dohlen, D. Mouillet, P. Puget, J. Antichi, A. Baruffolo, P. Baudoz, A. Berton, A. Boccaletti, M. Carbillet, J. Charton, R. Claudi, M. Downing, P. Feautrier, E. Fedrigo, T. Fusco, R. Gratton, N. Hubin, M. Kasper, M. Langlois, C. Moutou, L. Mugnier, J. Pragt, P. Rabou, M. Saisse, H. M. Schmid, E. Stadler, M. Turrato, S. Udry, R. Waters, and F. Wildi. SPHERE: A 'Planet Finder' Instrument for the VLT. *The Messenger*, 125, September 2006. URL <http://adsabs.harvard.edu/abs/2006Msngr.125...29B>.
- F. Bonnarel, P. Fernique, O. Bienaymé, D. Egret, F. Genova, M. Louys, F. Ochsenbein, M. Wenger, and J. G. Bartlett. The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. *Astronomy & Astrophysics, Supplement*, 143:33–40, April 2000. doi: 10.1051/aas:2000331. URL <http://adsabs.harvard.edu/abs/2000A%26AS...143...33B>.
- H. Bouy, E. Bertin, E. Moraux, J.-C. Cuillandre, J. Bouvier, D. Barrado, E. Solano, and A. Bayo. Dynamical analysis of nearby clusters. Automated astrometry from the ground: precision proper motions over a wide field. *Astronomy & Astrophysics*, 554:A101, June 2013. doi: 10.1051/0004-6361/201220748. URL <http://adsabs.harvard.edu/abs/2013A%26A...554A.101B>.
- E. Bressert, N. Bastian, R. Gutermuth, S. T. Megeath, L. Allen, N. J. Evans, II, L. M. Rebull, J. Hatchell, D. Johnstone, T. L. Bourke, L. A. Cieza, P. M. Harvey, B. Merin, T. P. Ray, and N. F. H. Tothill. The spatial distribution of star formation in the solar neighbourhood: do all stars form in dense clusters? *Monthly Notices of the Royal Astronomical Society*, 409:L54–L58, November 2010. doi: 10.1111/j.1745-3933.2010.00946.x. URL <http://adsabs.harvard.edu/abs/2010MNRAS.409L...54B>.
- J. A. Cardelli, G. C. Clayton, and J. S. Mathis. The relationship between infrared, optical, and ultraviolet extinction. *Astrophysical Journal*, 345:245–256, October 1989. doi: 10.1086/167900. URL <http://adsabs.harvard.edu/abs/1989ApJ...345...245C>.
- M.-R. L. Cioni, G. Clementini, L. Girardi, R. Guandalini, M. Gullieuszik, B. Miszalski, M.-I. Moretti, V. Ripepi, S. Rubele, G. Bagheri, K. Bekki, N. Cross, W. J. G. de Blok,



- R. de Grijs, J. P. Emerson, C. J. Evans, B. Gibson, E. Gonzales-Solares, M. A. T. Groenewegen, M. Irwin, V. D. Ivanov, J. Lewis, M. Marconi, J.-B. Marquette, C. Mastropietro, B. Moore, R. Napiwotzki, T. Naylor, J. M. Oliveira, M. Read, E. Sutorius, J. T. van Loon, M. I. Wilkinson, and P. R. Wood. The VMC survey. I. Strategy and first data. *Astronomy & Astrophysics*, 527:A116, March 2011. doi: 10.1051/0004-6361/201016137. URL <http://adsabs.harvard.edu/abs/2011A%26A...527A.116C>.
- G. B. Dalton, M. Caldwell, A. K. Ward, M. S. Whalley, G. Woodhouse, R. L. Edeson, P. Clark, S. M. Beard, A. M. Gallie, S. P. Todd, J. M. D. Strachan, N. N. Bezawada, W. J. Sutherland, and J. P. Emerson. The VISTA infrared camera. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6269 of *Proceedings of the SPIE*, page 62690X, June 2006. doi: 10.1117/12.670018. URL <http://adsabs.harvard.edu/abs/2006SPIE.6269E...0XD>.
- R. Davies, J. Schubert, M. Hartl, J. Alves, Y. Cl  net, F. Lang-Bardl, H. Nicklas, J.-U. Pott, R. Ragazzoni, E. Tolstoy, T. Agocs, H. Anwand-Heerwart, S. Barboza, P. Baudoz, R. Bender, P. Bizenberger, A. Boccaletti, W. Boland, P. Bonifacio, F. Briegel, T. Buey, F. Chapron, M. Cohen, O. Czoske, S. Dreizler, R. Falomo, P. Feautrier, N. F  rster Schreiber, E. Gendron, R. Genzel, M. Gl  ck, D. Gratadour, R. Greimel, F. Grupp, M. H  user, M. Haug, J. Hennawi, H. J. Hess, V. H  rmann, R. Hofferbert, U. Hopp, Z. Hubert, D. Ives, W. Kausch, F. Kerber, H. Kravcar, K. Kuijken, F. Lang-Bardl, M. Leitzinger, K. Leschinski, D. Massari, S. Mei, F. Merlin, L. Mohr, A. Monna, F. M  ller, R. Navarro, M. Plattner, N. Przybilla, R. Ramlau, S. Ramsay, T. Ratzka, P. Rhode, J. Richter, H.-W. Rix, G. Rodeghiero, R.-R. Rohloff, G. Rousset, R. Ruddenklau, V. Schaffenroth, J. Schlichter, A. Sevin, R. Stuik, E. Sturm, J. Thomas, N. Tromp, M. Turatto, G. Verdoes-Kleijn, F. Vidal, R. Wagner, M. Wegner, W. Zeilinger, B. Ziegler, and G. Zins. MICADO: first light imager for the E-ELT. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Proceedings of the SPIE*, page 99081Z, August 2016. doi: 10.1117/12.2233047. URL <http://adsabs.harvard.edu/abs/2016SPIE.9908E...1ZD>.
- R. S. de Jong, O. Bellido-Tirado, C. Chiappini,   . Depagne, R. Haynes, D. Johl, O. Schnurr, A. Schwoppe, J. Walcher, F. Dionies, D. Haynes, A. Kelz, F. S. Kitaura, G. Lamer, I. Minchev, V. M  ller, S. E. Nuza, J.-C. Olaya, T. Piffl, E. Popow, M. Steinmetz, U. Ural, M. Williams, R. Winkler, L. Wisotzki, W. R. Ansorge, M. Banerji, E. Gonzalez Solares, M. Irwin, R. C. Kennicutt, D. King, R. G. McMa-

- hon, S. Koposov, I. R. Parry, D. Sun, N. A. Walton, G. Finger, O. Iwert, M. Krumpe, J.-L. Lizon, M. Vincenzo, J.-P. Amans, P. Bonifacio, M. Cohen, P. Francois, P. Jagourel, S. B. Mignot, F. Royer, P. Sartoretti, R. Bender, F. Grupp, H.-J. Hess, F. Lang-Bardl, B. Muschielok, H. Böhringer, T. Boller, A. Bongiorno, M. Brusa, T. Dwelly, A. Merloni, K. Nandra, M. Salvato, J. H. Pragt, R. Navarro, G. Gerlofsma, R. Roelfsema, G. B. Dalton, K. F. Middleton, I. A. Tosh, C. Boeche, E. Caffau, N. Christlieb, E. K. Grebel, C. Hansen, A. Koch, H.-G. Ludwig, A. Quirrenbach, L. Sbordone, W. Seifert, G. Thimm, T. Trifonov, A. Helmi, S. C. Trager, S. Feltzing, A. Korn, and W. Boland. 4MOST: 4-metre multi-object spectroscopic telescope. In *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446 of *Proceedings of the SPIE*, page 84460T, September 2012. doi: 10.1117/12.926239. URL <http://adsabs.harvard.edu/abs/2012SPIE.8446E..0TD>.
- S. Dib. Testing the universality of the IMF with Bayesian statistics: young clusters. *Monthly Notices of the Royal Astronomical Society*, 444:1957–1981, October 2014. doi: 10.1093/mnras/stu1521. URL <http://adsabs.harvard.edu/abs/2014MNRAS.444.1957D>.
- B. T. Draine and A. Li. Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era. *Astrophysical Journal*, 657:810–837, March 2007. doi: 10.1086/511055. URL <http://adsabs.harvard.edu/abs/2007ApJ...657..810D>.
- A. Edge, W. Sutherland, K. Kuijken, S. Driver, R. McMahon, S. Eales, and J. P. Emerson. The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift. *The Messenger*, 154:32–34, December 2013. URL <http://adsabs.harvard.edu/abs/2013Msngr.154...32E>.
- N. J. Evans, II, M. M. Dunham, J. K. Jørgensen, M. L. Enoch, B. Merín, E. F. van Dishoeck, J. M. Alcalá, P. C. Myers, K. R. Stapelfeldt, T. L. Huard, L. E. Allen, P. M. Harvey, T. van Kempen, G. A. Blake, D. W. Koerner, L. G. Mundy, D. L. Padgett, and A. I. Sargent. The Spitzer c2d Legacy Results: Star-Formation Rates and Efficiencies; Evolution and Lifetimes. *Astrophysical Journal, Supplement*, 181:321–350, April 2009. doi: 10.1088/0067-0049/181/2/321. URL <http://adsabs.harvard.edu/abs/2009ApJS..181..321E>.

- J. Forbrich, C. J. Lada, M. Lombardi, C. Román-Zúñiga, and J. Alves. Smoke in the Pipe Nebula: dust emission and grain growth in the starless core FeSt 1-457. *Astronomy & Astrophysics*, 580:A114, August 2015. doi: 10.1051/0004-6361/201425375. URL <http://adsabs.harvard.edu/abs/2015A%26A...580A.114F>.
- Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans, and et al. The Gaia mission. *Astronomy & Astrophysics*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272. URL <http://adsabs.harvard.edu/abs/2016A%26A...595A...1G>.
- M. Galametz, R. C. Kennicutt, M. Albrecht, G. Aniano, L. Armus, F. Bertoldi, D. Calzetti, A. F. Crocker, K. V. Croxall, D. A. Dale, J. Donovan Meyer, B. T. Draine, C. W. Engelbracht, J. L. Hinz, H. Roussel, R. A. Skibba, F. S. Tabatabaei, F. Walter, A. Weiss, C. D. Wilson, and M. G. Wolfire. Mapping the cold dust temperatures and masses of nearby KINGFISH galaxies with Herschel. *Monthly Notices of the Royal Astronomical Society*, 425:763–787, September 2012. doi: 10.1111/j.1365-2966.2012.21667.x. URL <http://adsabs.harvard.edu/abs/2012MNRAS.425..763G>.
- J. P. Gardner, J. C. Mather, M. Clampin, R. Doyon, M. A. Greenhouse, H. B. Hammel, J. B. Hutchings, P. Jakobsen, S. J. Lilly, K. S. Long, J. I. Lunine, M. J. McCaughrean, M. Mountain, J. Nella, G. H. Rieke, M. J. Rieke, H.-W. Rix, E. P. Smith, G. Sonneborn, M. Stiavelli, H. S. Stockman, R. A. Windhorst, and G. S. Wright. The James Webb Space Telescope. *Space Science Reviews*, 123:485–606, April 2006. doi: 10.1007/s11214-006-8315-7. URL <http://adsabs.harvard.edu/abs/2006SSRv...123..485G>.
- G. Gilmore, S. Randich, M. Asplund, J. Binney, P. Bonifacio, J. Drew, S. Feltzing, A. Ferguson, R. Jeffries, G. Micela, and et al. The Gaia-ESO Public Spectroscopic Survey. *The Messenger*, 147:25–31, March 2012. URL <http://adsabs.harvard.edu/abs/2012Msng...147...25G>.
- A. A. Goodman, J. E. Pineda, and S. L. Schnee. The “True” Column Density Distribution in Star-Forming Molecular Clouds. *Astrophysical Journal*, 692:91–103, February 2009. doi: 10.1088/0004-637X/692/1/91. URL <http://adsabs.harvard.edu/abs/2009ApJ...692...91G>.

M. J. Griffin, A. Abergel, A. Abreu, P. A. R. Ade, P. André, J.-L. Augueres, T. Babbedge, Y. Bae, T. Baillie, J.-P. Baluteau, M. J. Barlow, G. Bendo, D. Benielli, J. J. Bock, P. Bonhomme, D. Brisbin, C. Brockley-Blatt, M. Caldwell, C. Cara, N. Castro-Rodriguez, R. Cerulli, P. Chanial, S. Chen, E. Clark, D. L. Clements, L. Clerc, J. Coker, D. Communal, L. Conversi, P. Cox, D. Crumb, C. Cunningham, F. Daly, G. R. Davis, P. de Antoni, J. Delderfield, N. Devin, A. di Giorgio, I. Didschuns, K. Dohlen, M. Donati, A. Dowell, C. D. Dowell, L. Duband, L. Dumaye, R. J. Emery, M. Ferlet, D. Ferrand, J. Fontignie, M. Fox, A. Franceschini, M. Frerking, T. Fulton, J. Garcia, R. Gastaud, W. K. Gear, J. Glenn, A. Goizel, D. K. Griffin, T. Grundy, S. Guest, L. Guillemet, P. C. Hargrave, M. Harwit, P. Hastings, E. Hatziminaoglou, M. Herman, B. Hinde, V. Hristov, M. Huang, P. Imhof, K. J. Isaak, U. Israelsson, R. J. Ivison, D. Jennings, B. Kiernan, K. J. King, A. E. Lange, W. Latter, G. Laurent, P. Laurent, S. J. Leeks, E. Lellouch, L. Levenson, B. Li, J. Li, J. Lilienthal, T. Lim, S. J. Liu, N. Lu, S. Madden, G. Mainetti, P. Marliani, D. McKay, K. Mercier, S. Molinari, H. Morris, H. Moseley, J. Mulder, M. Mur, D. A. Naylor, H. Nguyen, B. O'Halloran, S. Oliver, G. Olofsson, H.-G. Olofsson, R. Orfei, M. J. Page, I. Pain, P. Panuzzo, A. Papageorgiou, G. Parks, P. Parr-Burman, A. Pearce, C. Pearson, I. Pérez-Fournon, F. Pinsard, G. Pisano, J. Podosek, M. Pohlen, E. T. Polehampton, D. Pouliquen, D. Rigopoulou, D. Rizzo, I. G. Roseboom, H. Roussel, M. Rowan-Robinson, B. Rownd, P. Saraceno, M. Sauvage, R. Savage, G. Savini, E. Sawyer, C. Scharnberg, D. Schmitt, N. Schneider, B. Schulz, A. Schwartz, R. Shafer, D. L. Shupe, B. Sibthorpe, S. Sidher, A. Smith, A. J. Smith, D. Smith, L. Spencer, B. Stobie, R. Sudiwala, K. Sukhatme, C. Surace, J. A. Stevens, B. M. Swinyard, M. Trichas, T. Tourette, H. Triou, S. Tseng, C. Tucker, A. Turner, M. Vaccari, I. Valtchanov, L. Vigroux, E. Virique, G. Voellmer, H. Walker, R. Ward, T. Waskett, M. Weilert, R. Wesson, G. J. White, N. Whitehouse, C. D. Wilson, B. Winter, A. L. Woodcraft, G. S. Wright, C. K. Xu, A. Zavagno, M. Zemcov, L. Zhang, and E. Zonca. The Herschel-SPIRE instrument and its in-flight performance. *Astronomy & Astrophysics*, 518:L3, July 2010. doi: 10.1051/0004-6361/201014519. URL <http://adsabs.harvard.edu/abs/2010A%26A...518L...3G>.

T. Güver and F. Özel. The relation between optical extinction and hydrogen column density in the Galaxy. *Monthly Notices of the Royal Astronomical Society*, 400: 2050–2053, December 2009. doi: 10.1111/j.1365-2966.2009.15598.x. URL <http://adsabs.harvard.edu/abs/2009MNRAS.400.2050G>.



A. Hacar, J. Alves, M. Tafalla, and J. R. Goicoechea. Gravitational collapse of the OMC-1 region. *Astronomy & Astrophysics*, 602:L2, June 2017. doi: 10.1051/0004-6361/201730732. URL <http://adsabs.harvard.edu/abs/2017A%26A...602L...2H>.

William Herschel. Account of some observations tending to investigate the construction of the heavens. by william herschel, esq. f. r. s. *Philosophical Transactions of the Royal Society of London*, 74:437–451, 1784. ISSN 02610523. URL <http://www.jstor.org/stable/106602>.

L. A. Hillenbrand and L. W. Hartmann. A Preliminary Study of the Orion Nebula Cluster Structure and Dynamics. *Astrophysical Journal*, 492:540–553, January 1998. doi: 10.1086/305076. URL <http://adsabs.harvard.edu/abs/1998ApJ...492..540H>.

D. Hollenbach and E. E. Salpeter. Surface Recombination of Hydrogen Molecules. *Astrophysical Journal*, 163:155, January 1971. doi: 10.1086/150754. URL <http://adsabs.harvard.edu/abs/1971ApJ...163..155H>.

M. J. Jarvis, D. G. Bonfield, V. A. Bruce, J. E. Geach, K. McAlpine, R. J. McLure, E. González-Solares, M. Irwin, J. Lewis, A. K. Yoldas, S. Andreon, N. J. G. Cross, J. P. Emerson, G. Dalton, J. S. Dunlop, S. T. Hodgkin, F. O. Le, M. Karouzos, K. Meisenheimer, S. Oliver, S. Rawlings, C. Simpson, I. Smail, D. J. B. Smith, M. Sullivan, W. Sutherland, S. V. White, and J. T. L. Zwart. The VISTA Deep Extragalactic Observations (VIDEO) survey. *Monthly Notices of the Royal Astronomical Society*, 428:1281–1295, January 2013. doi: 10.1093/mnras/sts118. URL <http://adsabs.harvard.edu/abs/2013MNRAS.428.1281J>.

J. Kainulainen, H. Beuther, T. Henning, and R. Plume. Probing the evolution of molecular cloud structure. From quiescence to birth. *Astronomy & Astrophysics*, 508:L35–L38, December 2009. doi: 10.1051/0004-6361/200913605. URL <http://adsabs.harvard.edu/abs/2009A%26A...508L...35K>.

R. C. Kennicutt, Jr. The Global Schmidt Law in Star-forming Galaxies. *Astrophysical Journal*, 498:541–552, May 1998. doi: 10.1086/305588. URL <http://adsabs.harvard.edu/abs/1998ApJ...498..541K>.

- M. Kounkel, S. T. Megeath, C. A. Poteet, W. J. Fischer, and L. Hartmann. An HST Survey for 100-1000 au Companions around Young Stellar Objects in the Orion Molecular Clouds: Evidence for Environmentally Dependent Multiplicity. *Astrophysical Journal*, 821:52, April 2016. doi: 10.3847/0004-637X/821/1/52. URL <http://adsabs.harvard.edu/abs/2016ApJ...821...52K>.
- E. Krügel, editor. *The Physics of Interstellar Dust*, volume 10 of *Wiley Praxis Series in Astronomy & Astrophysics*, December 2002. doi: 10.1201/9781420033335. URL <http://adsabs.harvard.edu/abs/2002WSAA...10....K>.
- C. J. Lada and E. A. Lada. Embedded Clusters in Molecular Clouds. *Annual Review of Astronomy & Astrophysics*, 41:57–115, 2003. doi: 10.1146/annurev.astro.41.011802.094844. URL <http://adsabs.harvard.edu/abs/2003ARA%26A...41...57L>.
- C. J. Lada, E. A. Lada, D. P. Clemens, and J. Bally. Dust extinction and molecular gas in the dark cloud IC 5146. *Astrophysical Journal*, 429:694–709, July 1994. doi: 10.1086/174354. URL <http://adsabs.harvard.edu/abs/1994ApJ...429..694L>.
- C. J. Lada, M. Lombardi, and J. F. Alves. On the Star Formation Rates in Molecular Clouds. *Astrophysical Journal*, 724:687–693, November 2010. doi: 10.1088/0004-637X/724/1/687. URL <http://adsabs.harvard.edu/abs/2010ApJ...724..687L>.
- A. Lawrence, S. J. Warren, O. Almaini, A. C. Edge, N. C. Hambly, R. F. Jameson, P. Lucas, M. Casali, A. Adamson, S. Dye, J. P. Emerson, S. Foucaud, P. Hewett, P. Hirst, S. T. Hodgkin, M. J. Irwin, N. Lodieu, R. G. McMahon, C. Simpson, I. Smail, D. Mortlock, and M. Folger. The UKIRT Infrared Deep Sky Survey (UKIDSS). *Monthly Notices of the Royal Astronomical Society*, 379:1599–1617, August 2007. doi: 10.1111/j.1365-2966.2007.12040.x. URL <http://adsabs.harvard.edu/abs/2007MNRAS.379.1599L>.
- K. Leschinski, O. Czoske, R. Köhler, M. Mach, W. Zeilinger, G. Verdoes Kleijn, J. Alves, W. Kausch, and N. Przybilla. SimCADO: an instrument data simulator package for MICADO at the E-ELT. In *Modeling, Systems Engineering, and Project Management for Astronomy VI*, volume 9911 of *Proceedings of the SPIE*, page

- 991124, August 2016. doi: 10.1117/12.2232483. URL <http://adsabs.harvard.edu/abs/2016SPIE.9911E..24L>.
- J. A. Lewis and C. J. Lada. Protostars at Low Extinction in Orion A. *Astrophysical Journal*, 825:91, July 2016. doi: 10.3847/0004-637X/825/2/91. URL <http://adsabs.harvard.edu/abs/2016ApJ...825...91L>.
- M. Lombardi and J. Alves. Mapping the interstellar dust with near-infrared observations: An optimized multi-band technique. *Astronomy & Astrophysics*, 377:1023–1034, October 2001. doi: 10.1051/0004-6361:20011099. URL <http://adsabs.harvard.edu/abs/2001A%26A...377.1023L>.
- M. Lombardi, J. Alves, and C. J. Lada. 2MASS wide field extinction maps. I. The Pipe nebula. *Astronomy & Astrophysics*, 454:781–796, August 2006. doi: 10.1051/0004-6361:20042474. URL <http://adsabs.harvard.edu/abs/2006A%26A...454..781L>.
- M. Lombardi, J. Alves, and C. J. Lada. 2MASS wide field extinction maps. IV. The Orion, Monoceros R2, Rosette, and Canis Major star forming regions. *Astronomy & Astrophysics*, 535:A16, November 2011. doi: 10.1051/0004-6361/201116915. URL <http://adsabs.harvard.edu/abs/2011A%26A...535A..16L>.
- M. Lombardi, H. Bouy, J. Alves, and C. J. Lada. Herschel-Planck dust optical-depth and column-density maps. I. Method description and results for Orion. *Astronomy & Astrophysics*, 566:A45, June 2014. doi: 10.1051/0004-6361/201323293. URL <http://adsabs.harvard.edu/abs/2014A%26A...566A..45L>.
- M. Lombardi, J. Alves, and C. J. Lada. Molecular clouds have power-law probability distribution functions. *Astronomy & Astrophysics*, 576:L1, April 2015. doi: 10.1051/0004-6361/201525650. URL <http://adsabs.harvard.edu/abs/2015A%26A...576L...1L>.
- S. R. Majewski, APOGEE Team, and APOGEE-2 Team. The Apache Point Observatory Galactic Evolution Experiment (APOGEE) and its successor, APOGEE-2. *Astronomische Nachrichten*, 337:863, September 2016. doi: 10.1002/asna.201612387. URL <http://adsabs.harvard.edu/abs/2016AN....337..863M>.

- H. J. McCracken, B. Milvang-Jensen, J. Dunlop, M. Franx, J. P. U. Fynbo, O. Le Fèvre, J. Holt, K. I. Caputi, Y. Goranova, F. Buitrago, J. P. Emerson, W. Freudling, P. Hudelot, C. López-Sanjuan, F. Magnard, Y. Mellier, P. Møller, K. K. Nilsson, W. Sutherland, L. Tasca, and J. Zabl. UltraVISTA: a new ultra-deep near-infrared survey in COSMOS. *Astronomy & Astrophysics*, 544:A156, August 2012. doi: 10.1051/0004-6361/201219507. URL <http://adsabs.harvard.edu/abs/2012A%26A...544A.156M>.
- R. G. McMahon, M. Banerji, E. Gonzalez, S. E. Koposov, V. J. Bejar, N. Lodieu, R. Rebolo, and VHS Collaboration. First Scientific Results from the VISTA Hemisphere Survey (VHS). *The Messenger*, 154:35–37, December 2013. URL <http://adsabs.harvard.edu/abs/2013Msngr.154...35M>.
- S. T. Megeath, R. Gutermuth, J. Muzerolle, E. Kryukova, K. Flaherty, J. L. Hora, L. E. Allen, L. Hartmann, P. C. Myers, J. L. Pipher, J. Stauffer, E. T. Young, and G. G. Fazio. The Spitzer Space Telescope Survey of the Orion A and B Molecular Clouds. I. A Census of Dusty Young Stellar Objects and a Study of Their Mid-infrared Variability. *Astronomical Journal*, 144:192, December 2012. doi: 10.1088/0004-6256/144/6/192. URL <http://adsabs.harvard.edu/abs/2012AJ....144..192M>.
- S. T. Megeath, R. Gutermuth, J. Muzerolle, E. Kryukova, J. L. Hora, L. E. Allen, K. Flaherty, L. Hartmann, P. C. Myers, J. L. Pipher, J. Stauffer, E. T. Young, and G. G. Fazio. The Spitzer Space Telescope Survey of the Orion A and B Molecular Clouds. II. The Spatial Distribution and Demographics of Dusty Young Stellar Objects. *Astronomical Journal*, 151:5, January 2016. doi: 10.3847/0004-6256/151/1/5. URL <http://adsabs.harvard.edu/abs/2016AJ....151....5M>.
- D. Minniti, P. W. Lucas, J. P. Emerson, R. K. Saito, M. Hempel, P. Pietrukowicz, A. V. Ahumada, M. V. Alonso, J. Alonso-Garcia, J. I. Arias, R. M. Bandyopadhyay, R. H. Barbá, B. Barbuy, L. R. Bedin, E. Bica, J. Borissova, L. Bronfman, G. Carraro, M. Catelan, J. J. Clariá, N. Cross, R. de Grijs, I. Dékány, J. E. Drew, C. Fariña, C. Feinstein, E. Fernández Lajús, R. C. Gamen, D. Geisler, W. Gieren, B. Goldman, O. A. Gonzalez, G. Gunthardt, S. Gurovich, N. C. Hambly, M. J. Irwin, V. D. Ivanov, A. Jordán, E. Kerins, K. Kinemuchi, R. Kurtev, M. López-Corredoira, T. Maccarone, N. Masetti, D. Merlo, M. Messineo, I. F. Mirabel, L. Monaco, L. Morelli, N. Padilla, T. Palma, M. C. Parisi, G. Pignata, M. Rejkuba, A. Roman-Lopes, S. E. Sale, M. R.



- Schreiber, A. C. Schröder, M. Smith, L. S. , Jr., M. Soto, M. Tamura, C. Tappert, M. A. Thompson, I. Toledo, M. Zoccali, and G. Pietrzynski. VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way. *New Astronomy*, 15:433–443, July 2010. doi: 10.1016/j.newast.2009.12.002. URL <http://adsabs.harvard.edu/abs/2010NewA...15..433M>.
- F. Motte, P. Andre, and R. Neri. The initial conditions of star formation in the rho Ophiuchi main cloud: wide-field millimeter continuum mapping. *Astronomy & Astrophysics*, 336:150–172, August 1998. URL <http://adsabs.harvard.edu/abs/1998A%26A...336..150M>.
- D. M. Nataf. Uncertainties in the interstellar extinction curve and the Cepheid distance to M101. *Monthly Notices of the Royal Astronomical Society*, 449:1171–1176, May 2015. doi: 10.1093/mnras/stv156. URL <http://adsabs.harvard.edu/abs/2015MNRAS.449.1171N>.
- G. L. Pilbratt, J. R. Riedinger, T. Passvogel, G. Crone, D. Doyle, U. Gageur, A. M. Heras, C. Jewell, L. Metcalfe, S. Ott, and M. Schmidt. Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. *Astronomy & Astrophysics*, 518:L1, July 2010. doi: 10.1051/0004-6361/201014759. URL <http://adsabs.harvard.edu/abs/2010A%26A...518L...1P>.
- J. E. Pineda, P. Caselli, and A. A. Goodman. CO Isotopologues in the Perseus Molecular Cloud Complex: the X-factor and Regional Variations. *Astrophysical Journal*, 679:481–496, May 2008. doi: 10.1086/586883. URL <http://adsabs.harvard.edu/abs/2008ApJ...679..481P>.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, M. Baker, A. Balbi, A. J. Banday, and et al. Planck early results. I. The Planck mission. *Astronomy & Astrophysics*, 536:A1, December 2011a. doi: 10.1051/0004-6361/201116464. URL <http://adsabs.harvard.edu/abs/2011A%26A...536A...1P>.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, and et al. Planck early results. XIX. All-sky temperature and dust optical depth from Planck and IRAS. Constraints on the “dark gas” in our Galaxy. *Astronomy & Astrophysics*, 536:A19,

December 2011b. doi: 10.1051/0004-6361/201116479. URL <http://adsabs.harvard.edu/abs/2011A%26A...536A...19P>.

A. Poglitsch, C. Waelkens, N. Geis, H. Feuchtgruber, B. Vandenbussche, L. Rodríguez, O. Krause, E. Renotte, C. van Hoof, P. Saraceno, J. Cepa, F. Kerschbaum, P. Agnèse, B. Ali, B. Altieri, P. Andreani, J.-L. Augeres, Z. Balog, L. Barl, O. H. Bauer, N. Belbachir, M. Benedettini, N. Billot, O. Boulade, H. Bischof, J. Blommaert, E. Callut, C. Cara, R. Cerulli, D. Cesarsky, A. Contursi, Y. Creten, W. De Meester, V. Doublier, E. Doumayrou, L. Duband, K. Exter, R. Genzel, J.-M. Gillis, U. Grözinger, T. Henning, J. Herreros, R. Huygen, M. Inguscio, G. Jakob, C. Jamar, C. Jean, J. de Jong, R. Katterloher, C. Kiss, U. Klaas, D. Lemke, D. Lutz, S. Madden, B. Marquet, J. Martignac, A. Mazy, P. Merken, F. Montfort, L. Morbidelli, T. Müller, M. Nielbock, K. Okumura, R. Orfei, R. Ottensamer, S. Pezzuto, P. Popesso, J. Putzeys, S. Regibo, V. Reveret, P. Royer, M. Sauvage, J. Schreiber, J. Stegmaier, D. Schmitt, J. Schubert, E. Sturm, M. Thiel, G. Tofani, R. Vavrek, M. Wetzstein, E. Wieprecht, and E. Wiezorrek. The Photodetector Array Camera and Spectrometer (PACS) on the Herschel Space Observatory. *Astronomy & Astrophysics*, 518:L2, July 2010. doi: 10.1051/0004-6361/201014535. URL <http://adsabs.harvard.edu/abs/2010A%26A...518L...2P>.

C. Reina and M. Tarengi. X-ray Absorption and Optical Extinction in Interstellar Space. *Astronomy & Astrophysics*, 26:257, July 1973. URL <http://adsabs.harvard.edu/abs/1973A%26A...26..257R>.

T. P. Robitaille, B. A. Whitney, R. Indebetouw, K. Wood, and P. Denzmore. Interpreting Spectral Energy Distributions from Young Stellar Objects. I. A Grid of 200,000 YSO Model SEDs. *Astrophysical Journal, Supplement*, 167:256–285, December 2006. doi: 10.1086/508424. URL <http://adsabs.harvard.edu/abs/2006ApJS...167..256R>.

T. P. Robitaille, B. A. Whitney, R. Indebetouw, and K. Wood. Interpreting Spectral Energy Distributions from Young Stellar Objects. II. Fitting Observed SEDs Using a Large Grid of Precomputed Models. *Astrophysical Journal, Supplement*, 169: 328–352, April 2007. doi: 10.1086/512039. URL <http://adsabs.harvard.edu/abs/2007ApJS...169..328R>.

- D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophysical Journal*, 500:525–553, June 1998. doi: 10.1086/305772. URL <http://adsabs.harvard.edu/abs/1998ApJ...500..525S>.
- M. Schmidt. The Rate of Star Formation. *Astrophysical Journal*, 129:243, March 1959. doi: 10.1086/146614. URL <http://adsabs.harvard.edu/abs/1959ApJ...129..243S>.
- M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *Astronomical Journal*, 131:1163–1183, February 2006. doi: 10.1086/498708. URL <http://adsabs.harvard.edu/abs/2006AJ...131.1163S>.
- L. Spezzi, M. G. Petr-Gotzens, J. M. Alcalá, J. K. Jørgensen, T. Stanke, M. Lombardi, and J. F. Alves. The VISTA Orion mini-survey: star formation in the Lynds 1630 North cloud. *Astronomy & Astrophysics*, 581:A140, September 2015. doi: 10.1051/0004-6361/201425417. URL <http://adsabs.harvard.edu/abs/2015A%26A...581A.140S>.
- S. W. Stahler and F. Palla. *The Formation of Stars*. 2005. URL <http://adsabs.harvard.edu/abs/2005fost.book.....S>.
- M. Steinmetz, T. Zwitter, A. Siebert, F. G. Watson, K. C. Freeman, U. Munari, R. Campbell, M. Williams, G. M. Seabroke, R. F. G. Wyse, Q. A. Parker, O. Bienaymé, S. Roeser, B. K. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, J. F. Navarro, D. Burton, C. J. P. Cass, J. A. Dawe, K. Fiegert, M. Hartley, K. S. Russell, W. Saunders, H. Enke, J. Bailin, J. Binney, J. Bland-Hawthorn, C. Boeche, W. Dehnen, D. J. Eisenstein, N. W. Evans, M. Fiorucci, J. P. Fulbright, O. Gerhard, U. Jauregi, A. Kelz, L. Mijović, I. Minchev, G. Parmentier, J. Peñarrubia, A. C. Quillen, M. A. Read, G. Ruchti, R.-D. Scholz, A. Siviero, M. C. Smith, R. Sordo, L. Veltz, S. Vidrih, R. von Berlepsch, B. J. Boyle, and E. Schilbach. The Radial Velocity Experiment (RAVE): First Data Release. *Astro-*

- nomical Journal*, 132:1645–1668, October 2006. doi: 10.1086/506564. URL <http://adsabs.harvard.edu/abs/2006AJ....132.1645S>.
- A. Stutz, R. Launhardt, H. Linz, O. Krause, T. Henning, J. Kainulainen, M. Nielbock, J. Steinacker, and P. André. Dust-temperature of an isolated star-forming cloud: Herschel observations of the Bok globule CB244. *Astronomy & Astrophysics*, 518:L87, July 2010. doi: 10.1051/0004-6361/201014537. URL <http://adsabs.harvard.edu/abs/2010A%26A...518L..87S>.
- W. Sutherland, J. Emerson, G. Dalton, E. Atad-Ettinger, S. Beard, R. Bennett, N. Bezawada, A. Born, M. Caldwell, P. Clark, S. Craig, D. Henry, P. Jeffers, B. Little, A. McPherson, J. Murray, M. Stewart, B. Stobie, D. Terrett, K. Ward, M. Whalley, and G. Woodhouse. The Visible and Infrared Survey Telescope for Astronomy (VISTA): Design, technical overview, and performance. *Astronomy & Astrophysics*, 575:A25, March 2015. doi: 10.1051/0004-6361/201424973. URL <http://adsabs.harvard.edu/abs/2015A%26A...575A..25S>.
- J. A. Tauber, N. Mandolesi, J.-L. Puget, T. Banos, M. Bersanelli, F. R. Bouchet, R. C. Butler, J. Charra, G. Crone, J. Dodsworth, and et al. Planck pre-launch status: The Planck mission. *Astronomy & Astrophysics*, 520:A1, September 2010. doi: 10.1051/0004-6361/200912983. URL <http://adsabs.harvard.edu/abs/2010A%26A...520A...1T>.
- R. J. Trumpler. Preliminary results on the distances, dimensions and space distribution of open star clusters. *Lick Observatory Bulletin*, 14:154–188, 1930. doi: 10.5479/ADS/bib/1930LicOB.14.154T. URL <http://adsabs.harvard.edu/abs/1930LicOB..14..154T>.
- M. W. Werner, T. L. Roellig, F. J. Low, G. H. Rieke, M. Rieke, W. F. Hoffmann, E. Young, J. R. Houck, B. Brandl, G. G. Fazio, J. L. Hora, R. D. Gehrz, G. Helou, B. T. Soifer, J. Stauffer, J. Keene, P. Eisenhardt, D. Gallagher, T. N. Gautier, W. Irace, C. R. Lawrence, L. Simmons, J. E. Van Cleve, M. Jura, E. L. Wright, and D. P. Cruikshank. The Spitzer Space Telescope Mission. *Astrophysical Journal, Supplement*, 154:1–9, September 2004. doi: 10.1086/422992. URL <http://adsabs.harvard.edu/abs/2004ApJS..154....1W>.



J. C. Wilson, F. Hearty, M. F. Skrutskie, S. Majewski, R. Schiavon, D. Eisenstein, J. Gunn, B. Blank, C. Henderson, S. Smeed, R. Barkhouser, A. Harding, G. Fitzgerald, T. Stolberg, J. Arns, M. Nelson, S. Brunner, A. Burton, E. Walker, C. Lam, P. Maseman, J. Barr, F. Leger, L. Carey, N. MacDonald, T. Horne, E. Young, G. Rieke, M. Rieke, T. O'Brien, S. Hope, J. Krakula, J. Crane, B. Zhao, M. Carr, C. Harrison, R. Stoll, M. A. Vernieri, J. Holtzman, M. Shetrone, C. Allende-Prieto, J. Johnson, P. Frinchaboy, G. Zasowski, D. Bizyaev, B. Gillespie, and D. Weinberg. The Apache Point Observatory Galactic Evolution Experiment (APOGEE) high-resolution near-infrared multi-object fiber spectrograph. In *Ground-based and Airborne Instrumentation for Astronomy III*, volume 7735 of *Proceedings of the SPIE*, page 77351C, July 2010. doi: 10.1117/12.856708. URL <http://adsabs.harvard.edu/abs/2010SPIE.7735E..1CW>.

M. Wolf. Über den dunklen Nebel NGC 6960. *Astronomische Nachrichten*, 219:109, July 1923. doi: 10.1002/asna.19232190702.

E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astronomical Journal*, 140:1868-1881, December 2010. doi: 10.1088/0004-6256/140/6/1868. URL <http://adsabs.harvard.edu/abs/2010AJ....140.1868W>.



# Appendix



In this Appendix I provide additional information and supplementary material to the main sections of this thesis. The following pages include a German summary of the thesis as required by the University of Vienna, a personal statement, the reprint permission for the published articles in Chapters [2](#) and [3](#) (Chapter [4](#) is still in the peer-review stage), the original ESO proposal for the Orion A VISTA observations, and my Curriculum Vitae.





## A.1 Zusammenfassung

Sternentstehung ist ein wesentlicher Bestandteil des beobachtbaren Universum, reguliert Mechanismen in Galaxien sowohl auf großen, als auch auf kleinen Skalen und hat somit grundlegende Auswirkungen auf deren Entwicklung. Neue Sterne entstehen in riesigen Ansammlungen von Gas und Staub, sogenannten Molekülwolken, über die jedoch bisher nur grundlegende Erkenntnisse gewonnen werden konnten. Um diese Wolken besser zu verstehen, ist es daher notwendig eine umfassende, und vor allem einheitliche Untersuchung dieser Objekte durchzuführen.

Der Großteil unseres Wissens über den Ursprung neuer Sterne stammt aus Beobachtungen nahegelegener Sternentstehungsregionen. Zu diesen zählt die Molekülwolke Orion A, wo Sterne in einer Vielfalt unterschiedlicher Umgebungen entstehen. In dieser Arbeit präsentiere ich Ergebnisse in Bezug auf die gesamte Orion A Molekülwolke, deren Grundlage ein neues, umfassendes Beobachtungsprojekt im nahen Infraroten darstellt. Diese Beobachtungen liefern die bisher präzisen photometrischen Daten dieser Region, wobei deren Genauigkeit die Entwicklung neuer Methoden zur Untersuchung der Gasverteilung notwendig machte.

Die Ergebnisse dieser Arbeit stützen sich auf Extinktion durch interstellaren Staub, ein Effekt bei dem Licht durch Absorptions- und Streuungseffekte abgeschwächt wird. Basierend auf dieser Grundlage lassen sich beispielsweise physikalische Eigenschaften von Molekülwolken berechnen. Dazu gehören unter anderem deren Massenverteilung und Eigenschaften von Staubkörnern. Im Allgemeinen sind Berechnung von Massen, die auf Extinktion beruhen, wesentlich genauer als Methoden, die Staubemission oder Beobachtungen einzelner Moleküllinien verwenden. Zum Zweck der Berechnung der Extinktion präsentiere ich eine neue Methode, die auf Maschinenlernen beruht. Im Gegensatz zu früheren Vorgehensweisen verringert dieser Ansatz systematische Fehler, was vor allem bei hochwertigen photometrischen Daten zu drastisch besserer Genauigkeit führt. Außerdem befasse ich mich mit der seit Jahren ungelösten Frage nach einem allgemein gültigen Extinktionsgesetz im nahen Infraroten und zeige, basierend auf einer statistischen Analyse, dass dieses durchaus variabel ist. Weiters vergleiche ich die errechnete Extinktion mit Staubemissionswerten und mache deutlich, dass zuvor unbekannte systematische Fehler Messungen von Säulendichten beeinträchtigen.

Den Abschluss dieser Arbeit bildet eine Diskussion über zukünftige Projekte, deren Ziele es sind, die hier präsentierten Ergebnisse auch auf andere nahegelegene Wolken zu übertragen. Die in dieser Arbeit vorgestellten Methoden bilden die Grundlage dieser Untersuchungen.



## A.2 Acknowledgements

I will probably remember the time of writing this statement for the rest of my life. It is the moment where my knowledgeable, talented, and incredibly appreciative supervisor told me I could finally go to print this document and submit my doctoral thesis. The last years, beginning with my undergraduate studies at the University of Vienna and leading up to this moment, have been an incredible journey, for which no pile of adjectives is enough to do it justice. It has been a true rollercoaster of feelings, with countless memories burned into my memory. Not all of them are positive, some of them are actually really bad, and many things did not make a whole lot of sense along the way. But for anyone reading this, I can only repeat the wise words of a true pioneer of our generation: "You can't connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future." I will leave it up to you, my dear reader, to find out the origin of these oh-so-true words. I can only say that they have become true also for me and when the dots finally connect, it is among the most wonderful experiences you will have in your life.

Without even trying to explain how all of this ended up with me eventually getting my doctorate, I will continue with my list of honorable mentions. This specifically includes people who are not on the publications, who have continually supported me over the years, and without whom it would not have even been remotely possible to sit here and write these words. Also, it is not easy to find a suitable sequence for such things, so the following order should not be taken too seriously.

The first memory that flashed into my mind when I set the above paragraph was that of me moving into my family's apartment in Vienna in September 2005. In this picture, I walk into my room where I find a freshly made bed with a neat dent in the cushion. It was the moment where it all started, where my parents did not object to the idea of devoting my early life to professional astronomy. Without them none of this could have happened and I will forever be grateful for their never-ending support. I also had one or the other girlfriend in this timeframe who had to endure repeatedly watching Star Wars and action movies from the 1980's (none of them ever watched Star Trek with me, though). They also had to endure this crazy astronomy-rollercoaster, for which I am honestly sorry. Only one of them survived the trip with me, which, I think, is a good sign for the years to come.

Another special place belongs to my fellow students and colleagues, who shared

many things, good and bad, that come with being a PhD candidate in astronomy. I am the first in our party of PhD students in Vienna to finish a doctoral thesis, so I hope I will be able to offer other students advise whenever they need it. My thanks also go to many friends outside astronomy, some of whom I have known for more than a decade, some I only met a few months ago. All of you are part of this thesis, in one way or another.

Working on a dissertation in natural sciences without funding is a concept I will probably never be able to understand, as, I think, it is just a really bad idea. I was lucky because I received a fellowship from the Austrian Academy of Sciences, after fighting for funds for the better part of a year. I want to express my sincere thanks to the Academy for granting me this fellowship, thus setting the very foundation for this work.

Last, but certainly not least, writing a dissertation, of course, requires a thesis advisor. For me this was João Alves, who took me in as an undergraduate student and since then has offered me countless chances to prove myself. I don't have a comparison as to what it means to be a thesis advisor. I can only say that during the last few years I have learned more than I could have ever imagined and I will only say that I made it to the end; that is all that counts.



## Astronomy and Astrophysics

Editor in Chief: T. Forveille

### T. Forveille

Astronomy & Astrophysics  
Observatoire de Paris  
61, avenue de l'Observatoire  
75014 Paris, France

Tel.: 33 0(1) 43 29 05 41  
Fax: 33 0(1) 43 29 05 57  
e-mail: [aanda.paris@obspm.fr](mailto:aanda.paris@obspm.fr)  
Web: <http://www.aanda.org>

merging  
Annales d'Astrophysique  
Arkiv for Astronomi  
Bulletin of the Astronomical Institutes  
of the Netherlands  
Bulletin Astronomique  
Journal des Observateurs  
Zeitschrift für Astrophysik  
Bulletin of the Astronomical Institutes  
of Czechoslovakia

Paris, July 4, 2017

## Reprint Permission

### Material:

Article by Meingast et al. 2016, A&A, 587, A153

Article by Meingast et al. 2017, A&A, 601, A137

### To be used in:

PhD thesis entitled "A Vision of Orion", University of Vienna

### Permission granted to:

Stefan Meingast  
[stefan.meingast@univie.ac.at](mailto:stefan.meingast@univie.ac.at)

I hold copyright on the material referred to above, and hereby grant permission for its use as requested herewith.

The article should be reproduced in the same format as that published in A&A (for example, in an appendix). In particular, the present permission rules do not allow copy-and-pasting parts of the article into the main text of the thesis.

Credit should be given as follows:

Credit: Author, A&A, vol, page, year, reproduced with permission © ESO.



Thierry Forveille  
A&A Editor-in-Chief

Sponsored by Argentina, Austria, Belgium, Brazil, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Poland, Portugal, Slovak Republic, Spain, Sweden, and Switzerland.

Produced and distributed by EDP Sciences for ESO.





# EUROPEAN SOUTHERN OBSERVATORY

Organisation Européenne pour des Recherches Astronomiques dans l'Hémisphère Austral  
Europäische Organisation für astronomische Forschung in der südlichen Hemisphäre

OBSERVING PROGRAMMES OFFICE • Karl-Schwarzschild-Straße 2 • D-85748 Garching bei München • e-mail: [opo@eso.org](mailto:opo@eso.org) • Tel. : +49-89-32 00 64 73

## APPLICATION FOR OBSERVING TIME

PERIOD: **90A**

### Important Notice:

By submitting this proposal, the PI takes full responsibility for the content of the proposal, in particular with regard to the names of CoIs and the agreement to act according to the ESO policy and regulations, should observing time be granted

<b>1. Title</b> ORION DEEP: VISTA NIR imaging of Orion A	<b>Category:</b> <b>C-1</b>																				
<b>2. Abstract / Total Time Requested</b> Total Amount of Time: 0 nights VM, 33 hours SM We propose to carry out a deep survey of the entire giant molecular cloud harboring the closest massive star forming region to Earth. We will focus on the following topics: 1) investigate in a robust manner the relation CMF/IMF (~1000 dense cores), 2) identify and characterize YSOs (via Spitzer-Herschel-WISE-Chandra), 3) compare the IMF of the individual clusters with the composite IMF of the entire cloud, 4) characterize the sub-stellar IMF (down to 2 $M_{Jup}$ unreddened, 1 Myr object), and 5) derive an accurate threshold for star formation, the Star Formation Rate, and map possible variations of these quantities across this giant molecular cloud. The data will be made public and will be useful to other research programs. The feasibility of this project is guaranteed by our analysis of the VISTA Science Verification data on a nearby field. The proposed survey will be the most sensitive NIR survey ever performed on an entire massive star forming molecular cloud.																					
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">3. Run</th> <th style="text-align: left;">Period</th> <th style="text-align: left;">Instrument</th> <th style="text-align: left;">Time</th> <th style="text-align: left;">Month</th> <th style="text-align: left;">Moon</th> <th style="text-align: left;">Seeing</th> <th style="text-align: left;">Sky</th> <th style="text-align: left;">Mode</th> <th style="text-align: left;">Type</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>90</td> <td>VIRCAM</td> <td>33h</td> <td>dec</td> <td>n</td> <td>0.8</td> <td>CLR</td> <td>s</td> <td></td> </tr> </tbody> </table>		3. Run	Period	Instrument	Time	Month	Moon	Seeing	Sky	Mode	Type	A	90	VIRCAM	33h	dec	n	0.8	CLR	s	
3. Run	Period	Instrument	Time	Month	Moon	Seeing	Sky	Mode	Type												
A	90	VIRCAM	33h	dec	n	0.8	CLR	s													
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">4. Number of nights/hours</th> <th style="text-align: left;">Telescope(s)</th> <th style="text-align: left;">Amount of time</th> </tr> </thead> <tbody> <tr> <td colspan="3">a) already awarded to this project:</td> </tr> <tr> <td colspan="3">b) still required to complete this project:</td> </tr> </tbody> </table>		4. Number of nights/hours	Telescope(s)	Amount of time	a) already awarded to this project:			b) still required to complete this project:													
4. Number of nights/hours	Telescope(s)	Amount of time																			
a) already awarded to this project:																					
b) still required to complete this project:																					
<b>5. Special remarks:</b> This legacy-like proposal was tailored to address a variety of star formation topics on which the co-Is are experts. We will make public the reduced data in standard VO format and the extinction map of the entire Orion A complex, enabling further studies not described in this proposal.																					
<b>6. Principal Investigator:</b> D. Mardones, <a href="mailto:diego@das.uchile.cl">diego@das.uchile.cl</a> , CL, Universidad de Chile, Departamento de Astronomia																					
<b>6a. Co-investigators:</b> <table style="width: 100%; border-collapse: collapse;"> <tbody> <tr> <td style="width: 10%;">J.</td> <td style="width: 40%;">Alves</td> <td style="width: 50%;">Universitaet Wien, Institut fuer Astronomie, AT</td> </tr> <tr> <td>C.</td> <td>Alves de Oliveira</td> <td>Centro de Astrobiologia (CSIC-INTA), E</td> </tr> <tr> <td>J.</td> <td>Ascenso</td> <td>ESO Headquarters Garching, ESO</td> </tr> <tr> <td>H.</td> <td>Bouy</td> <td>Centro de Astrobiologia (CSIC-INTA), E</td> </tr> </tbody> </table> <i>Following CoIs moved to the end of the document ...</i>		J.	Alves	Universitaet Wien, Institut fuer Astronomie, AT	C.	Alves de Oliveira	Centro de Astrobiologia (CSIC-INTA), E	J.	Ascenso	ESO Headquarters Garching, ESO	H.	Bouy	Centro de Astrobiologia (CSIC-INTA), E								
J.	Alves	Universitaet Wien, Institut fuer Astronomie, AT																			
C.	Alves de Oliveira	Centro de Astrobiologia (CSIC-INTA), E																			
J.	Ascenso	ESO Headquarters Garching, ESO																			
H.	Bouy	Centro de Astrobiologia (CSIC-INTA), E																			
<b>7. Is this proposal linked to a PhD thesis preparation? State role of PhD student in this project</b> Yes / S. Meingast. The PhD thesis is planned around these data. / starting																					

## 8. Description of the proposed programme

**A – Scientific Rationale:** About 99% of the mass of a molecular cloud is invisible to direct observation, which has been by far the major impediment to the understanding of the process of star formation, since the beginning of star formation studies in the late 1940's. For example, we still do not know, i) what are the initial conditions to star formation, ii) why most stars form in clusters (e.g., Lada & Lada 2003), iii) what is the origin of the Initial Mass Function (IMF) (e.g., Motte et al. 1998) or iv) at what mass does the IMF end (e.g., Alves de Oliveira et al. 2009, Bastian et al. 2010). To address these fundamental questions we propose to carry out the most sensitive survey ever performed on an entire massive star forming molecular cloud (Orion A, the closest to Earth at  $d=414\pm7$  pc, Menten et al. 2007). An absolutely key element needed to successfully address the questions above is a reliable column density map (via dust extinction) because it allows one to correctly unreddened the YSO population, but most importantly, **it provides the most robust representation of the distribution of gas mass in the cloud**. In particular, the advantages of NIR dust extinction as a column density tracer have been discussed independently by Goodman et al. (2009) which performed an unbiased comparison between the three standard density tracer methods, namely, NIR dust extinction (NICER, Lombardi & Alves 2001), dust thermal emission in the mm and far-IR, and molecular line emission. These authors found that dust is a better column density tracer than molecular gas (CO), and that observations of dust extinction provide more robust measurements of column density than observations of dust emission (because of the dependence of the latter on the uncertain knowledge of dust temperatures (T) and dust emissivities ( $\beta$ )). This implies that in a massive star forming cloud such as Orion A, where (at least) T varies widely because of the vast number of embedded YSOs and hot OB stars, dust emission maps alone are severely and fundamentally limited as tracers of cloud mass, in particular at the size of dense cores. At the same time, the proposed deep survey will reveal the stellar and sub-stellar population of the **entire** Orion A cloud, from O-stars down to very low mass brown-dwarfs/planets with masses of a few to several Jupiter masses. While the substellar population of the Orion Nebula Cluster is relatively well understood (e.g., Da Rio et al 2012), the ONC region covers less than 5% of the area of the entire Orion A giant molecular cloud we are proposing to observe (see Figure 1b)). The combination of VISTA data with complementary archival Spitzer–Herschel–Chandra data (already processed by our team), and in particular the new all-sky Wide Field Infrared Survey Explorer (WISE) data covering regions not observed by other satellites, will allow the construction of a benchmark database of YSOs in the closest massive star forming cloud to Earth.

**B – Immediate Objective:** **1) An alternative and more reliable look into the CMF/IMF relation.** We plan to construct the Core Mass Function (CMF) for an estimated  $\sim 1000$  cores via a tracer (dust extinction) that is not affected by changes in T,  $\beta$ , excitation conditions, or depletion. We will use background Spitzer sources located behind high column density regions, where there will be a deficit of H and K-band background sources, as extinction tracers. The map will be compared with the new and similar resolution molecular line maps from FCRAO, archival dust emission maps (SCUBA), as well as with publicly available Herschel maps to study possible variations in dust emissivity  $\beta$  (using a Herschel T map). **2) Improved YSO identification and characterization.** The deeper NIR data is better matched in resolution and sensitivity with the existing Spitzer–Herschel–WISE–Chandra archival data which will allow us to study YSOs in regions too embedded or confused for 2MASS, or altogether missed by Spitzer/Herschel (in bright nebula regions or not covered and accessible via either K-excesses or a combination with all-sky WISE). Scattered light nebula around protostars are very useful for confirming their protostellar nature and determining the inclination of the envelope cavity. **3) Composite versus individual IMF.** The composite IMF of ten  $100 M_{\odot}$  clusters is unlikely to be the same as the IMF of a  $1000 M_{\odot}$  cluster (a  $100 M_{\odot}$  cluster will in all likelihood not have an O-star, for example). So the composite IMF of the entire Orion A cloud (that contains smaller clusters as well as the Orion Nebula Cluster (ONC), e.g., Bally et al. 2008, Allen et al. 2008) will surely differ from the IMF of the ONC. But how exactly? The potential difference between the composite and the individual cluster IMFs in a cloud can be investigated directly and for the first time with the proposed data, and will bring new insights into the current understanding of the field IMF. **4) The IMF down to the brown-dwarf/planet regime** reaching  $2 M_{Jup}$  (for an unreddened, 1 Myr object) for the entire cloud. Identify very low mass objects and brown dwarfs by searching for IR-excess sources in the combined VISTA and archival data. Is there a maximum stellar mass vs. cluster size relation? Is the formation of low mass stars inhibited close to more massive YSOs? **5) The distribution of dense gas vs. the distribution of stars.** Since we will have the best representation of the distribution of mass in the cloud (the extinction map), and the most complete YSO catalog for this cloud, we will be able to derive an accurate threshold for star formation, the star formation rate and efficiency, and map possible variations of these quantities across this giant molecular cloud. We will be able to measure the critical density distribution leading to the formation of clusters and its possible variation with the Orion A cloud.

REFERENCES: Allen, L. et al. 2008, Handbook of SF regions, Bally et al. 2008, Handbook of SF Regions, Bate et al. 2003, MNRAS, 339, 577, Bastian, N. et al. 2010, ARA&A, 48, 339, Bonnell et al. 2001, MNRAS, 323, 785, Da Rio et al. 2012, ApJ, 748, 14 Goodman, A.A. et al. 2009, ApJ, 692, 91, Hennebelle et al. 2008, ApJ, 684, 395, Menten et al. 2007, A&A, 474, 515 Motte, F. et al. 1998, A&A, 336, 150, Kroupa et al. 2003, MNRAS, 346, 369 Lombardi, M. et al. 2001, A&A, 377, 1023, Whitworth et al. 2010, IAU 266, 264



## 8. Description of the proposed programme and attachments

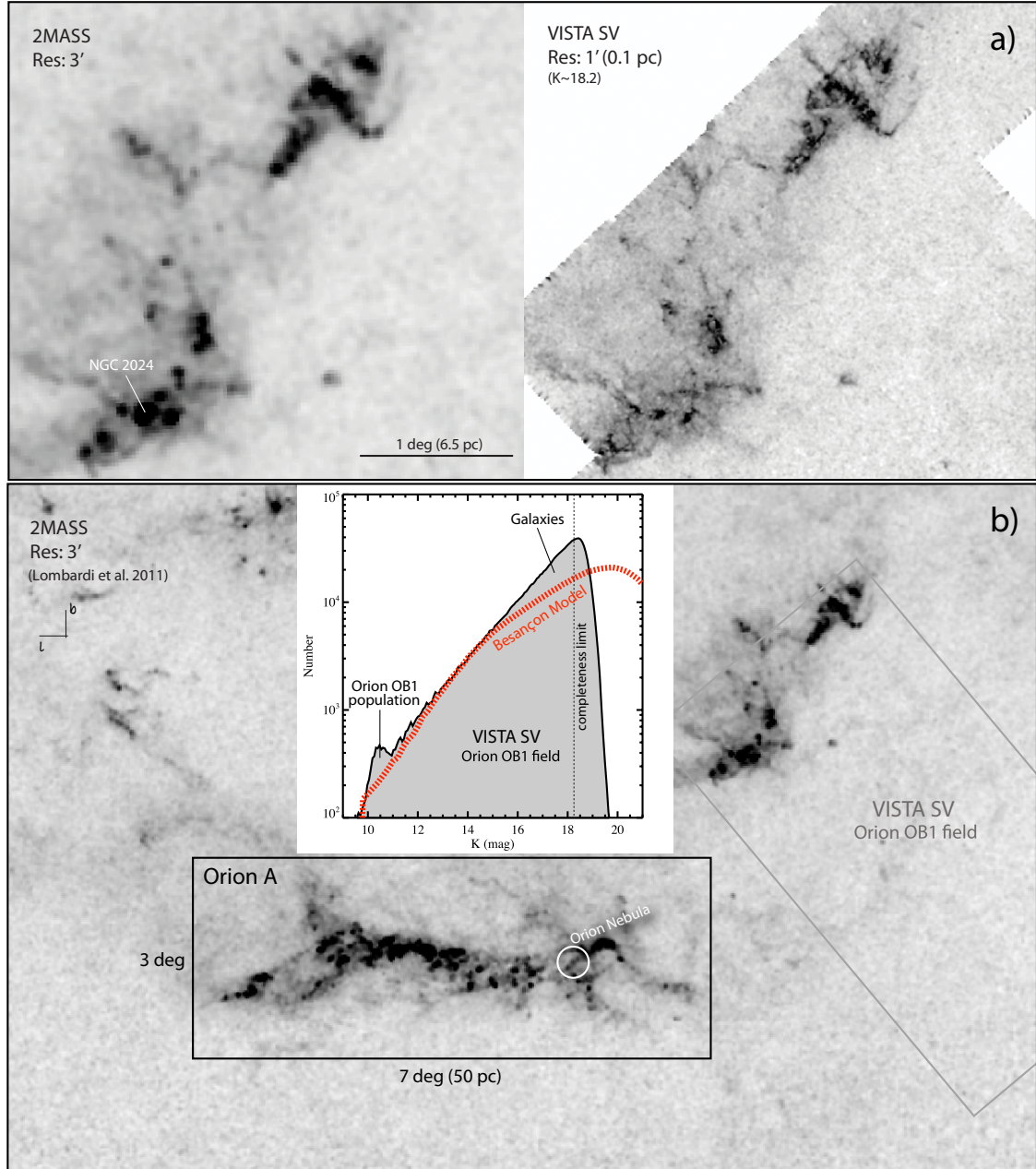


Fig. 1: **a)** Proof of concept: comparison between the 2MASS NICEST (Lombardi 2009) and the VISTA Science Verification (SV) extinction map for the Orion B cloud. **b)** Planned survey of Orion A (solid black rectangle), the complex that harbors the closest massive star formation region to Earth, the Orion Nebula cluster. **The proposed survey will be the most sensitive survey ever performed on an entire massive star forming molecular cloud**, capable of detecting  $\sim 1000$  dense cores and substellar objects down to  $2 M_{Jup}$  (1 Myr, unreddened) across the entire observed area. The K-band Luminosity Function of the VISTA SV survey is also shown. Most of the objects used to construct the VISTA SV extinction map are galaxies (via the GNICER, Foster et al. 2008, ApJ, 674, 831). By going  $\sim 2$  magnitudes deeper than the VISTA SV, i.e.,  $K \sim 20.2$  mag, the proposed survey will allow column density map resolutions of about 0.05-0.08 pc ( $30-45''$ ).

## 9. Justification of requested observing time and observing conditions

**Lunar Phase Justification:** No Moon constraint.

**Time Justification: (including seeing overhead)**

The Orion VISTA survey aims at mapping an area containing the entire giant molecular cloud harboring several star formation events, in particular the closest massive star formation region (Orion Nebula Cluster). The strategy consists of observing  $2 \times 7$  contiguous VISTA tiles covering the required  $\sim 21 \text{ deg}^2$  field shown in Figure 1. Each tile has a size of approximately  $1.0^\circ \times 1.5^\circ$  made by six contiguous VISTA exposures (pawprints). We plan to observe all 15 tiles (14 plus a control field) in the JHKs-bands. With the enormous advantage of having analysed the VISTA SV data, and having proved that an extinction map with a factor of 3 better resolution than 2MASS can be achieved from the SV data alone, (Figure 1 top), we stand on solid ground to estimate integration times. It was the surprising fact that the VISTA SV KLF is an almost perfect power-law down to the completeness limit ( $K_s \sim 18.2 \text{ mag}$ ), **very different from the Galactic model prediction** (see Figure 1 b)), that suggested that this project is indeed feasible due to the rising number of galaxies (that are used to construct the map via the GNICER technique). The most demanding requirement from all science goals is to be able to resolve dense cores (size  $\sim 0.16 \text{ pc}$ ; Lada et al. (2008)), requiring a spatial resolution of  $0.08 \text{ pc}$  ( $45''$ ). Extrapolating from the VISTA SV data, that allows a resolution of  $1'$ , and assuming the power-law displayed by the KLF (Figure 1 center) will not break substantially before  $K_s \sim 20.2 \text{ mag}$  (which is a good assumption as the KLF is dominated by galaxies, e.g., Metcalfe et al. (2006), while the number of stars expected from the Besançon model is rising until  $K_s \sim 19.8 \text{ mag}$ ), then going to  $K_s \sim 20.2$  will allow the detection of  $\sim 5 \times 10^6$  background objects ( $\sim 2 \times 10^6$  more than the VISTA SV data, for the same area), which translates into a resolution improvement of  $\sqrt{5/3}$ , or about 1.3, i.e.,  $0.08 \text{ pc}$  as required ( $45''$ ). This is a worst case scenario as we will use in addition the public Spitzer survey of the cloud which will be able to provide background sources for the critical high column density regions (because extinction is lower at the Spitzer bands.). For the best case scenario, with a favorable estimated number of Spitzer sources at high column density, plus a hybrid technique using both colors and source counts, we could reach  $30''$ , or  $0.05 \text{ pc}$ . The targeted detection limits per filter (in Vega magnitudes:  $J=21.5 \text{ mag}$ ,  $H=21.1 \text{ mag}$ , and  $K_s=20.2 \text{ mag}$ ), driven by the hardest requirement (extinction map resolution) allow the detection an unreddened  $1 \text{ Myr}$  object with a mass of about  $2 M_{Jup}$ , a conservative detection limit for all the other science cases in this proposal.

Using the VISTA ETC we evaluated the exposure time per tile to be  $6 \times (220 + 530 + 390) \text{ sec}$ , for a  $5\text{-}\sigma$  detection. Adding overheads as provided by ESO (Call for Proposals; changing filters, telescope preset and AO), one needs a total time of  $7920 \text{ sec/tile}$ . For the 15 tiles we would then need 33 hours to complete this project.

### 9a. Telescope Justification:

VISTA is a survey telescope and the only one suitable to carry out this project. No other telescope could cover about  $21 \text{ deg}^2$  in the three NIR bands (to the depth we are proposing), in any reasonable amount of time. All data reduction tasks will be carried out by the Cambridge Astronomical Survey Unit (CASU), one of their members is a CoI of this proposal with a scientific interest in the project. They will take responsibility for ensuring a uniform photometric calibration across all bands. The VISTA pipeline (that uses 2MASS photometric calibration) has some problems for regions of high extinction (see Hodgkin et al. 2009, MNRAS, 394, 675 for a discussion). We will use overlaps between the tiles to tie all the tiles onto the same system. This means that some of the survey region needs to include regions of relatively low extinction ( $E(B-V) < 2.0$ ), which it does, to provide a suitable anchor point for the bootstrapping.

### 9b. Observing Mode Justification (visitor or service):

VISTA operates in Service Mode only. If needed, we would be happy to help with the observations.

### 9c. Calibration Request:

Standard Calibration

10. Report on the use of ESO facilities during the last 2 years

J. Alves, J. Ascenso, and A. Moitinho are involved on the ongoing NACO GTO program on the origins of clusters. The final run of this program will take place May 2012. C. Alves de Oliveira programs : 083.C-0092 - Identification of Brown Dwarfs and Planetary Mass Objects in the Rho-Ophiuchus Dark Cloud (published in Alves de Oliveira et al. 2010, A&A, 515, 75). 085.C-0960(B) - Identification of Brown Dwarfs and Planetary Mass Objects in the Rho-Ophiuchus Dark Cloud: All data were analyzed and preliminary results have been presented in scientific meetings. Publication of results is in preparation. Paula Teixeira is currently analyzing data from programs 088.C-0736 and 086.C-0693 on a VISIR Protostellar Binary Survey (PROBIS) in Young Star Forming Regions. Jouni Kainulainen and Alvaro Hacar are involved in 087.C-0583- Streaming subsonic motions along the Musca Filament (APEX-SHeFI, observed, data reduced), & 089.C-0568 - Fragmentation of young filaments within Molecular Clouds (to be observed).

10a. ESO Archive - Are the data requested by this proposal in the ESO Archive (<http://archive.eso.org>)? If so, explain the need for new data.

The data requested in this proposal are not in the ESO Archive. Several well known targets in Orion A, like the Trapezium cluster, can be nevertheless found in the Archive but the data available typically cover a small region around the cluster, less than 1% the area of the survey proposed here. There is some data for the Orion Nebula Cluster, but these data cover less than 5% the area of the survey proposed here. Also, all these archival data came from a variety of ESO telescopes and cameras and cannot be easily added to the uniform survey proposed here.

10b. GTO/Public Survey Duplications:

The proposed target (the giant molecular cloud Orion A) does not appear as a protected target in P90. The entire Southern Hemisphere will be observed at some point by the VHS survey, and at least part of Orion A should be observed by a sub-survey of VHS (VHS-GPS). But **VHS-GPS** is a relatively shallow survey (60 sec per filter,  $K \sim 18$  mag Vega), and with a major drawback for extinction mapping: the survey will be done in the J and K-band only. Not having the H-band will mean that no colors will be available for the (most interesting) high-column density regions in the cloud, effectively degrading the resolution of the extinction map as well as the characterization of the Orion A young stellar population. Part of Orion A will be observed by **UKIDSS Galactic Clusters Survey**, but this is a survey centered on cluster science (Orion Nebula only, not covering the entire cloud), and is about 2 magnitudes shallower than the one proposed here. It is at present about 1/3 finished, according to their web page.

11. Applicant's publications related to the subject of this application during the last 2 years

- Chavarria, L., Mardones, D., Garay, G., Escala, A., Bronfman, L., Lizano, S. 2010, *Four Highly Luminous Massive Star-forming Regions in the Norma Spiral Arm. II. Deep Near-infrared Imaging*, ApJ, 710, 583
- Garay, G. Mardones, D., Bronfman, L., May, J., Chavarria, L., Nyman, L. 2010 *Four Highly Luminous Massive Star-forming Regions in the Norma Spiral Arm. I. Molecular Gas and Dust Observations*, ApJ, 710, 567
- Teixeira, P. S., C. J., Lada, M. Marengo, E. A. Lada 2012. *Spitzer observations of NGC2264: The nature of the disk population*. 2012arXiv1203.3754T
- Petr-Gotzens, M.; Alcalá, J. M.; Briceño, C.; González-Solares, E.; Spezzi, L.; Teixeira, P. et al. 2011 *Science Results from the VISTA Survey of the Orion Star-forming Region*, ESO Messenger, 145, 29
- Lombardi, M. et al., J., 2011. *2MASS wide field extinction maps. IV. The Orion, Mon R2, Rosette, and Canis Major star forming regions*. A&A, 535, 16
- Preibisch, T., Hodgkin, S. et al., 2011. *Near-Infrared properties of the X-ray emitting young stellar objects in the Carina Nebula*. arXiv.org, 1103, 2052
- Román-Zúñiga, C., Alves, J., Lada, C., & Lombardi, M. 2010 *Deep near-infrared survey of the Pipe Nebula. II. Data, Methods, and Dust Extinction Maps*. , ApJ 725, 2232.
- Alves de Oliveira, C. et al., 2010. *The low-mass population of the  $\rho$ -Oph molecular cloud*. A&A, 515, 75
- Lombardi, M., Lada, C.J. & Alves, J., 2010. *2MASS wide field extinction maps. III. The Taurus, Perseus, and California cloud complexes*. A&A, 512, 67

## 12. List of targets proposed in this programme

Run	Target/Field	$\alpha$ (J2000)	$\delta$ (J2000)	ToT	Mag.	Diam.	Additional info	Reference star
A	Orion A	05 35 00	-05 23 00	40.0	20	3×7 deg	Orion A GMC	

Target Notes: See Figure 1.

### 13. Scheduling requirements



#### 14. Instrument configuration

Period	Instrument	Run ID	Parameter	Value or list
90	VIRCAM	A	IMG	ESO filters: JHK

6b. Co-investigators:

*...continued from box 6a.*

M.	Casali	ESO Headquarters Garching,ESO
J.	Forbrich	Universitaet Wien,Institut fuer Astronomie,AT
A.	Hacar	Universitaet Wien,Institut fuer Astronomie,AT
S.	Hodgkin	Institute of Astronomy, University of Cambridge,UK
J.	Kainulainen	Max Planck Institut fuer Astronomie,D
M.	Lombardi	Universita degli Studi di Milano - Bicocca,I
S.	Meingast	Universitaet Wien,Institut fuer Astronomie,AT
A.	Moitinho	(SIM) Laboratrio de Sistemas, Instrumentao e Modelao em Cincias e Tecnologias do Ambiente e do Espao,P
E.	Moraux	Laboratoire d'astrophysique de Grenoble,Observatoire de Grenoble,F
M.	Petr Gotzens	ESO Headquarters Garching,ESO
C.	Román-Zúñiga	UNAM,Instituto de Astronomia,MX
P.	Teixeira	Universitaet Wien,Institut fuer Astronomie,AT





# Stefan Meingast

## *Curriculum Vitae*

### Personal Information

Date of Birth 19.05.1986  
Citizenship Austria  
Address Institute for Astrophysics, University of Vienna  
Türkenschanzstraße 17, 1180 Wien, Austria  
eMail stefan.meingast@univie.ac.at  
Homepage homepage.univie.ac.at/stefan.meingast  
Phone +43 1 427753811  
Mobile +43 650 7130120  
ORCID 0000-0002-0568-5526

### Education

2005 – 2009 **Bachelor studies of Astronomy**, *University of Vienna*.  
Thesis: Astronomical Distance Indicators  
Graduation with distinction  
2009 – 2012 **Master studies of Astronomy**, *University of Vienna*.  
Thesis: Pulsating Stars in NGC 6231  
Graduation with distinction  
since 2014 **Doctoral studies of Astronomy**, *University of Vienna*.  
Thesis: A Vision of Orion  
Planned graduation: Summer 2017

### Experience

#### Professional

2010 – 2011 **Scientific project staff**, *University of Vienna*.  
Project "Excitation of Stellar Pulsations" funded by FWF, Nr.: P20526-N16  
(AP2052611)  
2012 **ESO associate**, *Garching, Germany*.  
N-GLAO Sky Coverage, E-TRE-ESO-257-1321  
[http://homepage.univie.ac.at/stefan.meingast/ELT\\_SkyCoverage.pdf](http://homepage.univie.ac.at/stefan.meingast/ELT_SkyCoverage.pdf)

- 2012 – 2013 **Scientific staff**, *University of Vienna*.  
 2014 – 2016 **Doctoral Fellow**, *Austrian Academy of Sciences, University of Vienna*.

### Observational

- since 2007 More than 200 nights observing experience in time series photometry and spectroscopy of pulsating stars, transits of exoplanets, optical SETI, imaging of emission nebulae and galaxies, and instrumentation using the 80 cm telescope at the Department for Astrophysics, University of Vienna.
- 2011 Hvar Observatory, Croatia. Delaware Asteroseismic Mini Program 4: Photometric monitoring of pulsating white dwarfs.
- 2011 VLT using NACO in visitor mode. Which stars form first? Program-ID: 087.C-0916A
- 2012 VLT using SINFONI in visitor mode. Revealing the Properties of Embedded Clusters beyond the Milky Way. Program-ID: 090.C-0587A
- 2013 VLT using ISAAC in service mode. Which stars form first? Spectroscopic follow-up observations. Program-ID: 290.C-5165A
- 2013 VLT using NACO in visitor mode. Which stars form first? Program-ID: 091.C-0379A
- 2013 Kitt Peak National Observatory using NEWFIRM. Deep NIR Imaging Surveys of the Nearest GMCs.
- 2013 VISTA using VIRCAM. Deep NIR Imaging Surveys of Orion A.
- 2015 ALMA. Fibers and bundles: dissecting the internal substructure of massive filaments.
- 2017 – 2020 VISTA using VIRCAM. Co-PI of the ESO public survey VISIONS - VISTA Star Formation Atlas.

### Teaching

- 2008 **Assistant Tutor**, *Introduction to Astronomy I*, University of Vienna.
- 2010 – 2011 **Tutor**, *Astronomical Laboratory Exercises for Beginners*, University of Vienna.
- 2010 – 2012 **Tutor**, *Practical Observations*, University of Vienna.

### Awards

- 2010 **Performance Scholarship**, *University of Vienna*.  
 Awarded to students for extraordinary achievements and study performance.
- 2012 **Performance Scholarship**, *University of Vienna*.  
 Awarded to students for extraordinary achievements and study performance.
- 2014 **DOC Fellowship**, *Austrian Academy of Sciences*.  
 Doctoral Fellowship spanning three years. Project title: From Dark Clouds to Bright Clusters.



---

## Skills

### Software

Expert	Python, IDL
Advanced	Fortran95, $\LaTeX$ , Adobe Photoshop, Adobe After Effects, Adobe Muse, Office
Basic	HTML, C

### Communication

- 2014 Poster at the Olympian Symposium on Star Formation, Paralia, Greece.
- 2014 Research stay and oral presentation, Brno, Czech Republic.
- 2015 Poster at Orion (Un)plugged, Vienna, Austria.
- 2016 Oral presentation at the European Week of Astronomy and Space Science 2016, Athens, Greece.
- 2016 Poster at Star Formation 2016, Exeter, UK.
- 2016 Poster at Astronomical Data Analysis Software and Systems XXVI, Trieste, Italy.
- 2017 Oral presentation at Star Formation from Cores to Clusters, ESO, Santiago, Chile.

### Languages

German	Mothertongue
English	Fluent

## Publications (chronological)

- 1 G. Handler and **S. Meingast**. New  $\beta$  Cephei stars in the young open cluster NGC 637. *Astronomy & Astrophysics*, 533:A70, September 2011.
- 2 S. N. Longmore, J. Rathborne, N. Bastian, J. Alves, J. Ascenso, J. Bally, L. Testi, A. Longmore, C. Battersby, E. Bressert, C. Purcell, A. Walsh, J. Jackson, J. Foster, S. Molinari, **S. Meingast**, A. Amorim, J. Lima, R. Marques, A. Moitinho, J. Pinhao, J. Rebordao, and F. D. Santos. G0.253 + 0.016: A Molecular Cloud Progenitor of an Arches-like Cluster. *The Astrophysical Journal*, 746:117, February 2012.
- 3 W. R. Leeb, A. Poppe, E. Hammel, J. Alves, M. Brunner, and **S. Meingast**. Single-Photon Technique for the Detection of Periodic Extraterrestrial Laser Pulses full access. *Astrobiology*, 13: 521–535, June 2013.
- 4 **S. Meingast**, G. Handler, and R. R. Shobbrook. Pulsating stars in NGC 6231. Frequency analysis and photometric mode identification near the main-sequence. *Astronomy & Astrophysics*, 559: A108, November 2013.
- 5 W. R. Leeb, J. Alves, **S. Meingast**, and M. Brunner. Simulated low-intensity optical pulsar observation with single-photon detector. *Astronomy & Astrophysics*, 574:A9, February 2015.
- 6 S. Uttenthaler, **S. Meingast**, T. Lebzelter, B. Aringer, R. R. Joyce, K. Hinkle, L. Guzman-Ramirez, and R. Greimel. LX Cygni: A carbon star is born. *Astronomy & Astrophysics*, 585:A145, January 2016.
- 7 **S. Meingast**, J. Alves, D. Mardones, P. S. Teixeira, M. Lombardi, J. Großschedl, J. Ascenso, H. Bouy, J. Forbrich, A. Goodman, A. Hacar, B. Hasenberger, J. Kainulainen, K. Kubiak, C. Lada, E. Lada, A. Moitinho, M. Petr-Gotzens, L. Rodrigues, and C. G. Román-Zúñiga. VISION - Vienna survey in Orion. I. VISTA Orion A Survey. *Astronomy & Astrophysics*, 587:A153, March 2016.
- 8 A. Hacar, J. Alves, J. Forbrich, **S. Meingast**, K. Kubiak, and J. Großschedl. APOGEE strings: A fossil record of the gas kinematic structure. *Astronomy & Astrophysics*, 589:A80, May 2016.
- 9 J. Forbrich, V. M. Rivilla, K. M. Menten, M. J. Reid, C. J. Chandler, U. Rau, S. Bhatnagar, S. J. Wolk, and **S. Meingast**. The Population of Compact Radio Sources in the Orion Nebula Cluster. *The Astrophysical Journal*, 822:93, May 2016.
- 10 B. Hasenberger, J. Forbrich, J. Alves, S. J. Wolk, **S. Meingast**, K. V. Getman, and I. Pillitteri. Gas absorption and dust extinction towards the Orion Nebula Cluster. *Astronomy & Astrophysics*, 593:A7, August 2016.
- 11 K. Kubiak, J. Alves, H. Bouy, L. M. Sarro, J. Ascenso, A. Burkert, J. Forbrich, J. Großschedl, A. Hacar, B. Hasenberger, M. Lombardi, **S. Meingast**, R. Köhler, and P. S. Teixeira. Orion revisited III. The Orion Belt population. *Astronomy & Astrophysics*, 598:A124, February 2017.

- 12 E. Paunzen, G. Handler, M. Lendl, B. Baumann, C. Rab, **S. Meingast**, M. Rode-Paunzen, M. Netopil, V. Antoci, L. Zhu, M. Zejda, H. Bozic. Search for variables in six Galactic open clusters. *New Astronomy*, 52:133, April 2017.
- 13 **S. Meingast**, M. Lombardi, and J. Alves. Estimating extinction using unsupervised machine learning. *Astronomy & Astrophysics*, 601:A137, May 2017.
- 14 **S. Meingast**, J. Alves, and M. Lombardi. VISION - Vienna survey in Orion. II. Infrared extinction in Orion A. Submitted to *Astronomy & Astrophysics*, June 2017.