



**HAL**  
open science

# Detection and classification of type Ia supernovae for cosmology in the complete data set of SNLS

Anais Möller

► **To cite this version:**

Anais Möller. Detection and classification of type Ia supernovae for cosmology in the complete data set of SNLS. *Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]*. Université Paris Diderot (Paris 7), 2015. English. NNT: . tel-01456470

**HAL Id: tel-01456470**

**<https://theses.hal.science/tel-01456470>**

Submitted on 5 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ SORBONNE PARIS CITÉ



**Irfu - CEA Saclay**  
Institut de recherche  
sur les lois fondamentales  
de l'Univers

Thèse préparée à l'UNIVERSITÉ PARIS DIDEROT

École doctorale STEP'UP - ED N°560

CEA Saclay Irfu/SPP Cosmologie

## Détection et classification des supernovae de type Ia pour la cosmologie dans l'ensemble des données de SNLS

par

**Anais Möller**

présentée et soutenue publiquement le 28 septembre 2015

Thèse de doctorat de Physique de l'Univers

dirigée par Vanina Ruhlmann-Kleider

devant un jury composé de:

<b>LIDMAN Chris</b>	Rapporteur
Astronome (Australian National Observatory)	
<b>MAGNEVILLE Christophe</b>	Rapporteur
Physicien senior (CEA Saclay Irfu/SPP)	
<b>GRENIER Isabelle</b>	Président du jury
Professeur (Paris Diderot)	
<b>LE GUILLOU Laurent</b>	Membre
Maître de conférences (UPMC/LPNHE)	
<b>PAIN Reynald</b>	Membre
Directeur de recherche (LPNHE-CNRS)	
<b>STARCK Jean-Luc</b>	Membre
Directeur de recherche (CEA Saclay Irfu/SPP)	
<b>RUHLMANN-KLEIDER Vanina</b>	Directrice de thèse
Directeur de recherche (CEA Saclay Irfu/SPP)	

UNIVERSITÉ SORBONNE PARIS CITÉ



UNIVERSITÉ PARIS DIDEROT

École doctorale STEP'UP - ED N°560

CEA Saclay Irfu/SPP Cosmologie

# Detection and classification of type Ia supernovae for cosmology in the complete data set of SNLS

by

**Anais Möller**

presented on September 28th, 2015

PhD on physics

supervised by Vanina Ruhlmann-Kleider

Jury:

<b>LIDMAN Chris</b>	Referee
Astronomer (Australian National Observatory)	
<b>MAGNEVILLE Christophe</b>	Referee
Senior physicist (CEA Saclay Irfu/SPP)	
<b>GRENIER Isabelle</b>	President of the jury
Professor (Paris Diderot)	
<b>LE GUILLOU Laurent</b>	Member
Maître de conférences (UPMC/LPNHE)	
<b>PAIN Reynald</b>	Member
Research director (LPNHE-CNRS)	
<b>STARCK Jean-Luc</b>	Member
Research director (CEA Saclay Irfu/SAP )	
<b>RUHLMANN-KLEIDER Vanina</b>	PhD supervisor
Research director (CEA Saclay Irfu/SPP)	

*“El universo (que otros llaman la Biblioteca) se compone de un número indefinido, y tal vez infinito, de galerías hexagonales, con vastos pozos de ventilación en el medio, cercados por barandas bajísimas. [...] Quizá me engañen la vejez y el temor, pero sospecho que la especie humana -la única- está por extinguirse y que la Biblioteca perdurará: iluminada, solitaria, infinita, perfectamente inmóvil, armada de volúmenes preciosos, inútil, incorruptible, secreta.”*

Jorge Luis Borges, Ficciones (1941)

# *Abstract*

## **Detection and classification of type Ia supernovae for cosmology in the complete data set of SNLS**

by Anais MÖLLER

In this work, I present improvements on the detection of transient events and the classification of supernovae (SNe) using supernova photometric redshifts in the SNLS deferred analysis. Detection of transient events can provide numerous false detections, while the photometric classification of type Ia SNe is usually contaminated by other types of SNe. Reducing the number of false detections and the misclassified SNe while maintaining the type Ia SN sample are important issues for both present and future surveys.

In order to reduce the artifacts that provide false detections, I developed a subtracted image stack treatment to reduce the number of non SN-like events using morphological component analysis. This technique exploits the morphological diversity of objects to be detected to extract the signal of interest. At the level of our subtraction stacks, SN-like events are rather circular objects while most spurious detections exhibit different shapes. SNIa Monte-Carlo (MC) generated images were used to study detection efficiency and coordinate resolution. When tested on SNLS 3-year data this procedure decreases the number of detections by a factor of two, while losing only 10% of SN-like events, almost all faint ones. MC results show that SNIa detection efficiency is equivalent to that of the original method for bright events, while the coordinate resolution is slightly improved.

The deferred pipeline uses only photometric information to classify SNe. The advantages of a photometric sample include larger number of events classified as type Ia, larger redshift coverage and no need for spectroscopic observations. I present here a new classification using photometric supernova redshifts optimized by a machine learning classification strategy. This algorithm provides redshifts for all events with a better average precision and lower catastrophic errors than the host galaxy photometric redshift catalogue used in the SNLS3 analysis. I optimized the selection strategy using machine learning techniques such as BDTs which increases efficiency and purity of the SNIa sample. This new photometric SN-redshift classification provides a type Ia SN sample with a contamination of less than 10% according to Monte-Carlo studies.

# *Resumé*

## **Détection et classification des supernovae de type Ia pour la cosmologie dans l'ensemble des données de SNLS**

par Anais MÖLLER

Dans cette thèse, je présente des améliorations sur la détection d'événements transitoires et la classification des supernovae (SNe) en utilisant les redshifts photométriques de supernova dans l'analyse diféré de SNLS. La détection des événements transitoires peut fournir de nombreuses fausses détections, tandis que la classification photométrique des SNe de type Ia est généralement contaminé par d'autres types de supernovae. Réduire le nombre de fausses détections et les SNe malclassés tout en maintenant l'échantillon du type Ia sont des questions importantes pour les investigations présentes et futures.

Afin de réduire les artefacts qui fournissent des fausses détections, j'ai développé un traitement pour les stacks des images soustraites pour réduire le nombre d'événements qui ne se ressemblent a SNe en utilisant Morphological Component Analysis. Cette technique exploite la diversité morphologique des objets à détecter pour extraire le signal d'intérêt. Au niveau de nos piles de soustraction, événements comme SN sont plutôt circulaires alors que la plupart des détections parasites présentent des formes différentes. Images des SNe Ia générées par Monte-Carlo (MC) ont été utilisés pour étudier l'efficacité de la détection et la résolution des coordonees. Lors d'un essai sur le données SNLS des 3 ans cette procédure diminue le nombre de détections par un facteur de deux, tout en ne perdant que 10 % d'événements qui se ressemblent au SNe, presque tous faibles. Résultats des MC montrent que l'efficacité de détection SNIa est équivalente à celle de la méthode originale pour les événements lumineux, tandis que la résolution de coordonnées est légèrement améliorée.

L'analyse diféré utilise uniquement des informations photométriques pour classer les supernovae. Les avantages d'un échantillon photométrique comprennent plus grand nombre d'événements classés comme de type Ia, la couverture de redshift plus grande et pas besoin d'observations spectroscopiques. Je présente ici une nouvelle classification en utilisant redshifts photométriques de supernovae optimisées par une stratégie de classification avec machine learning. Cet algorithme fournit des décalages vers le rouge pour tous les événements avec une meilleure précision moyenne et inférieure erreurs catastrophiques que l'analyse avec photométrie redshifts de la galaxie hôte avec SNLS3. J'ai optimisé la stratégie de sélection à l'aide des techniques de machine learning comme BDTs que augmente l'efficacité et la pureté de l'échantillon SNIa. Cette nouvelle classification photométrique SN-redshift fournit un échantillon des SNe type Ia avec une contamination de moins de 10 % selon les études de Monte-Carlo.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Resumé</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xi</b>
<b>Abbreviations</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Physical Universe: cosmology and acceleration of expansion</b>	<b>4</b>
2.1 Rise of General Relativity . . . . .	5
2.2 Symmetry: Cosmological Principle . . . . .	7
2.3 Space: Manifolds and metric . . . . .	7
2.4 Dynamic universe: Friedmann-Lemaître-Robertson-Walker Metric . . . . .	9
2.5 Expanding universe . . . . .	10
2.6 Universe's content . . . . .	14
2.6.1 Matter: baryons and dark matter . . . . .	14
2.6.2 Radiation . . . . .	15
2.6.3 Dark energy as a homogeneous fluid . . . . .	16
2.7 Friedmann equations . . . . .	17
2.8 $\Lambda$ CDM . . . . .	18
<b>3 Cosmological Observable: Supernovae of type Ia</b>	<b>21</b>
3.1 Supernovae . . . . .	21
3.1.1 Empirical classification . . . . .	22
3.2 Supernovae of type Ia . . . . .	23
3.2.1 Spectroscopic properties . . . . .	24
3.2.2 Photometric properties . . . . .	25
3.2.3 Standardizing SNeIa . . . . .	28
3.2.4 SN Ia rates . . . . .	30
3.2.5 Proposed mechanisms . . . . .	31
3.2.5.1 Single degenerate model (SD) . . . . .	32
3.2.5.2 Double degenerate model (DD) . . . . .	33

3.3	Distance measurements with SNe Ia	33
<b>4</b>	<b>SuperNova Legacy Survey (SNLS)</b>	<b>36</b>
4.1	The instrument	36
4.2	SNLS survey	37
4.3	The pipelines	41
<b>5</b>	<b>SNLS deferred photometric analysis</b>	<b>44</b>
5.1	Image Processing	45
5.1.1	Astrometry and resampling	45
5.1.2	Subtractions	46
5.2	Detection of transient events	49
5.2.1	Lunation stacks	50
5.2.2	Detection catalogues	50
5.2.3	Detection efficiency for SNLS3	51
5.3	Selection of SN-like events	53
5.3.1	Light curve reconstruction	53
5.3.2	SN-like cuts	55
5.4	SN classification: type Ia selection	57
5.4.1	SALT2 and photometric host-galaxy redshift	58
5.4.2	Light curve sampling requirements	59
5.4.3	Using SALT2 parameters for SNIa selection	59
5.5	5-year analysis changes	63
5.6	Summary	64
<b>6</b>	<b>A brief introduction to signal processing and sparsity</b>	<b>66</b>
6.1	Basic concepts: sparsity, atom, dictionary and scales	67
6.2	Morphological diversity	69
6.3	Morphological Component Analysis	70
6.4	Dictionaries	71
6.4.1	From Fourier to wavelet	71
6.4.2	Two-dimensional wavelets	73
<b>7</b>	<b>Improving detection of transient events</b>	<b>77</b>
7.1	Subtracted image stacks defects	78
7.2	Disentangling real transient events from artifacts	82
7.2.1	Choice of dictionaries	82
7.2.2	First treatment: removal of main artifacts	84
7.2.3	Second treatment: signal extraction with varying noise	88
7.3	New Detection strategy	90
7.3.1	Validating a detection	93
7.3.2	Assigning coordinates	94
7.3.3	Improving coordinate resolution	95
7.4	Results with SNLS3 data	97
7.5	MC studies	98
7.5.1	Corrections: Isolation	98
7.5.2	Corrections: Redshift and SNIa rate distributions	99
7.5.3	MC coordinate resolution and efficiency	102



7.6	SNLS and beyond...	105
7.7	Summary	106
<b>8</b>	<b>Classification of SNe Ia using photometric redshifts</b>	<b>108</b>
8.1	Machine learning and boosted decision trees	109
8.1.1	Principles	109
8.1.2	TMVA	112
8.2	Considerations for our application	112
8.2.1	Samples used in the BDT set-up and analysis	114
8.2.2	BDT analysis set-up	116
8.2.2.1	Finding the appropriate hyperparameters	116
8.2.2.2	Choosing a set of classification variables	118
8.2.2.3	Choice of the BDT response threshold	119
8.3	Classification of SNe Ia using photometric redshifts	122
8.4	Host-galaxy photometric redshift analysis (zgal)	123
8.4.1	zgal + SALT2 + sequential cuts: SNLS3	124
8.4.2	zgal + SALT2 + BDT analysis	125
8.5	SN photometric redshift analysis (zpho)	132
8.5.1	Classification using SALT2 twice (zpho+SALT2)	134
8.5.1.1	zpho + SALT2 + sequential cuts analysis	134
8.5.1.2	zpho + SALT2 + BDT analysis	136
8.5.2	zpho + general fit + BDT analysis	141
8.6	Summary	148
<b>9</b>	<b>Conclusions</b>	<b>152</b>
<b>A</b>	<b>Resumé du travail en francais</b>	<b>154</b>
	<b>Bibliography</b>	<b>161</b>
	<b>Acknowledgements</b>	<b>171</b>

# List of Figures

2.1	Galaxy distribution from 2dFGRS and SDSS . . . . .	8
2.2	Original diagram by Hubble in 1929 . . . . .	11
2.3	Comparisson $H_0$ measurements by Planck 2015. . . . .	11
2.4	Hubble diagram from the Supernova Cosmology Project . . . . .	16
2.5	Effect dark energy models on $H$ and $d_L$ . . . . .	19
2.6	$\Omega_\Lambda$ vs. $\Omega_m$ confidence contours of the JLA analysis . . . . .	20
3.1	SNe spectra and light curves . . . . .	23
3.2	Image of a galaxy with a SNIa . . . . .	23
3.3	SNIa spectrum for SN1981b . . . . .	24
3.4	SNIa light curve from SNLS . . . . .	26
3.5	Calan Tololo light curves with correction . . . . .	26
3.6	SNe Ia properties correlations as seen by SNLS . . . . .	29
3.7	Stretch and color of SNLS SNIa light curves as a function of their host galaxy stellar mass . . . . .	29
3.8	Observed SN Ia volumetric rate evolution with fitted parameters for the two parameter model . . . . .	31
4.1	CFHT dome . . . . .	37
4.2	The MegaPrime instrument . . . . .	37
4.3	MegaCam imager . . . . .	38
4.4	MegaCam filter transmission curves . . . . .	39
4.5	Full sky map with position of the deep and wide fields of CFHTLS . . . . .	39
4.6	MegaCam effective passbands at the center of the focal plane . . . . .	40
4.7	Light curves for SNLS SN Ia candidates . . . . .	41
4.8	JLA Hubble diagram . . . . .	43
5.1	Example of using subtraction for detecting SNe Ia . . . . .	47
5.2	Reference image for field D1 . . . . .	48
5.3	Image subtraction principle with a PSF adjustment . . . . .	49
5.4	Sketch: detection of transient events as in SNLS3 for each virtual CCD . . . . .	51
5.5	Detection efficiency for SNe Ia as a function of the generated $i_M$ peak magnitude in all survey fields. . . . .	53
5.6	Difference in the reduced $\chi^2$ between $i_M$ light curve fits by a constant and by an SN-like shape . . . . .	55
5.7	SN-like selection cuts applied on SNLS3 . . . . .	57
5.8	SALT2 parameters plot for color and stretch . . . . .	61
5.9	Color-magnitude diagrams . . . . .	62
5.10	Effect of the SNIa selection cuts . . . . .	63

5.11	Distribution of host-galaxy photometric redshifts for photometrically selected SNe Ia in SNLS3 . . . . .	65
6.1	Example of signals in time and Fourier space . . . . .	68
6.2	Basis functions for STFT and wavelet transform . . . . .	72
6.3	NGC2997 starlet decomposition . . . . .	73
6.4	Simulated Hubble Space Telescope image of a distant cluster of galaxies . . . . .	74
6.5	Example of curvelet . . . . .	74
6.6	Ridgelet example . . . . .	75
6.7	Flowchart of image decomposition in the discrete ridgelet dictionary . . . . .	75
6.8	Decomposition flowchart of an image in the first-generation discrete curvelet dictionary . . . . .	76
7.1	Reference image for field D1 . . . . .	79
7.2	Detection map for field D4 in SNLS3 . . . . .	79
7.3	Different defects on the subtracted image stacks on large scale . . . . .	80
7.4	Different defects on the subtracted image stacks on small scale . . . . .	80
7.5	Image stack for field D4, ccd 00 and lunation 10 . . . . .	81
7.6	SN 04D1c with its host-galaxy and after subtraction . . . . .	83
7.7	Typical atoms from the dictionaries used in the MCA algorithm . . . . .	83
7.8	MCA decomposition of a optical ghost . . . . .	85
7.9	MCA Decomposition of a SN Ia event . . . . .	85
7.10	Varying noise: pixel content in a 50x50 pixels box at different locations in a subtracted image stack . . . . .	86
7.11	MCA Decomposition of a SN Ia event where part of the signal leaks into the residuals . . . . .	87
7.12	Image stack before and after first treatment . . . . .	88
7.13	Varying noise after first treatment . . . . .	89
7.14	SNIa in the original subtracted image stack and after both treatments . . . . .	91
7.15	Image stack for field D4 original, after first treatment and after two-step . . . . .	91
7.16	New detection strategy schema . . . . .	92
7.17	A SNIa event in the original subtracted image stacks and after cleaning . . . . .	93
7.18	Number of pixels required to validate a detection: # lost events vs. # detections . . . . .	94
7.19	Detection maps for field D4 in SNLS3 . . . . .	96
7.20	Detection efficiency versus isolation distance for an early version of the new detection procedure compared with the standard detection procedure. . . . .	99
7.21	Redshift distribution of SNe Ia generated by MC in field D1. . . . .	100
7.22	Comoving volume distribution for our MC studies . . . . .	101
7.23	Redshift distribution for MC after rate and comoving volume weights are applied. Arbitrary normalization. . . . .	101
7.24	Efficiency of detection as a function of the generated peak magnitude in $i_M$ . . . . .	102
7.25	Coordinate resolution vs. number of stack years. . . . .	104
7.26	Profile of the coordinate resolution against redshift . . . . .	105
8.1	Classification example: classical approach vs. BDT . . . . .	110
8.2	Graphic representation of a simple decision tree . . . . .	111

8.3	Distribution of the SNIa and core-collapse after SN-like cuts as a function of generated redshift	115
8.4	Samples involved in BDT training and classification	116
8.5	Example ROC curve	117
8.6	Overtraining test	118
8.7	Variable ranking TMVA	118
8.8	Example of a global efficiency versus purity	119
8.9	Example of a contamination-classification efficiency ratio versus BDT response plot	120
8.10	zgal classification: SNLS3 purity of classified SNIa sample and contamination as a function of host-galaxy redshift	125
8.11	zgal classification: SNLS3 global SNIa efficiency as a function of generated redshift	126
8.12	zgal + SALT2 + BDT: AUC score against different choices of BDT hyperparameters	127
8.13	zgal + SALT2 + BDT: example of overtraining with hyperparameters	127
8.14	zgal + SALT2 + BDT: ranking of variables	128
8.15	zgal + SALT2 + BDT: signal and background vs. BDT response	128
8.16	zgal + SALT2 + BDT: study of the signal shoulder in BDT distribution	129
8.17	zgal + SALT2 + BDT: global efficiency vs. purity	130
8.18	zgal + SALT2 + BDT: global efficiency vs. gz	131
8.19	zgal + SALT2 + BDT: purity and contamination	132
8.20	zgal + SALT2 + BDT: SNLS3 data as a function of BDT response	133
8.21	zpho + SALT2 + sequential cuts: parameters plot for color and stretch	135
8.22	zpho + SALT2 + sequential cuts: color-magnitude diagrams	136
8.23	zpho + SALT2 + BDT: signal and background vs. BDT response	137
8.24	zpho + SALT2 + BDT: study of the signal shoulder in BDT distribution	138
8.25	zpho + SALT2 + BDT: global efficiency vs. purity	139
8.26	zpho + SALT2 + BDT: purity and contamination vs. zpho	140
8.27	zpho + SALT2 + BDT: SNLS3 data as a function of BDT response	141
8.28	zpho + SALT2 + BDT: global efficiency	142
8.29	zpho + SALT2 + BDT: study of efficiency low	142
8.30	zpho + general fit + BDT: signal and background vs. BDT response	143
8.31	zpho + general fit + BDT: global efficiency vs. purity	144
8.32	zpho + general fit + BDT: purity and contamination	145
8.33	zpho + general fit + BDT: global efficiency vs. gz	146
8.34	zpho + general fit + BDT: BDT response for SNLS3 data	146
8.35	zpho + general fit + BDT: SNLS3 data vs. BDT response 1	147
8.36	zpho + general fit + BDT: SNLS3 data vs. BDT response 2	148
8.37	Comparison global efficiencies against gz	151
A.1	JLA Hubble diagram	155
A.2	Color-magnitude diagrams	156
A.3	Pile de soustraction D4, ccd 00 et lunation 10 avec étoile brillante et “optical ghost” dans A.3a et le plan de détection avec la méthode originale de SNLS3 A.3b. De nombreuses detections peuvent être attribuées à des artefacts.	158
A.4	Efficiency of detection as a function of the generated peak magnitude in $i_M$	159

---

A.5 comparison des efficiacités globales vs. redshift generée . . . . . 159

# List of Tables

4.1	MegaCam technical specifications . . . . .	38
5.1	Number of detections in lunation and final catalogues for SNLS3 data. . . . .	51
7.1	Number of detections and SN-like events for the original and new procedures applied on SNLS3 data. . . . .	98
7.2	Coordinate resolutions and corresponding magnitude bias of SNIa detection original and new procedures . . . . .	103
8.1	Number of simulated light curve events that were selected as SN-like candidates . . . . .	115
8.2	Volumetric SN rates derived from SNLS data . . . . .	120
8.3	Classification analyses . . . . .	123
8.4	zgal classification: SNLS3 statistics . . . . .	124
8.5	zgal + SALT2 + BDT: statistics . . . . .	131
8.6	zpho + SALT2 + sequential cuts: statistics using sequential cuts . . . . .	136
8.7	zpho + SALT2 + BDT: statistics . . . . .	140
8.8	zpho + general fit + BDT: statistics . . . . .	145
8.9	classification summary: statistics . . . . .	150
A.1	Résolution des coordonnées en pixels (limite supérieure) avec le biais en magnitude correspondant. Pour plusieurs années de données. . . . .	157
A.2	Nombre des détections et événements qui se ressemblent au SNe dans les données de SNLS3 pour l'ancienne et la nouvelle procedure. . . . .	157

# Abbreviations

<b>BAO</b>	<b>B</b> aryon <b>A</b> coustic <b>O</b> scillations
<b>BDT</b>	<b>B</b> oosted <b>D</b> ecision <b>T</b> ree
<b>CC</b>	<b>C</b> ore <b>C</b> ollapse supernovae
<b>CFHT</b>	<b>C</b> anada <b>F</b> rance <b>H</b> awaii <b>T</b> elescope
<b>CMB</b>	<b>C</b> osmic <b>M</b> icrowave <b>B</b> ackground
<b>EFE</b>	<b>E</b> instein <b>F</b> ield <b>E</b> quations
<b>GR</b>	<b>G</b> eneral <b>R</b> elativity
<b>MC</b>	<b>M</b> onte <b>C</b> arlo
<b>SFR</b>	<b>S</b> tar <b>F</b> ormation <b>R</b> atio
<b>SN(e)</b>	<b>S</b> uper <b>N</b> ova(e)
<b>SN(e)Ia</b>	<b>S</b> uper <b>N</b> ova(e) type <b>Ia</b>
<b>SNLS</b>	<b>S</b> upernova <b>L</b> egacy <b>S</b> urvey
<b>SNLS3</b>	<b>S</b> upernova <b>L</b> egacy <b>S</b> urvey <b>3</b> year of data
<b>SNLS5</b>	<b>S</b> upernova <b>L</b> egacy <b>S</b> urvey <b>5</b> year of data
<i>gz</i>	Simulation generated redshift
<i>zgal</i>	Host-galaxy photometric redshift
<i>zpho</i>	Supernova photometric redshift
<i>m()</i> gal	magnitudes in the $(i_M, r_M, g_M, z_M)$ filters for the <i>zgal</i> analysis
<i>m()</i> pho	magnitudes in the $(i_M, r_M, g_M, z_M)$ filters for the <i>zpho</i> analysis
<i>c()</i> gal	chi squared for the $(i_M, r_M, g_M, z_M)$ filters for the <i>zgal</i> analysis
<i>c()</i> pho	chi squared for $(i_M, r_M, g_M, z_M)$ filters for the <i>zpho</i> analysis

# Chapter 1

## Introduction

At the beginning of 20th century, our Universe was thought to be static, everlasting and small. In 1915, when Einstein introduced his theory of General Relativity, he believed that the Universe was static. Around 1920, measurements by Slipher and Hubble pointed that the Universe was bigger than the Milky Way and it was dynamic. Cosmology was born. In 100 years, our conception of the Universe has changed from a static and visible matter universe, to a universe which is expanding and is composed of visible matter, dark matter, radiation and dark energy.

Dark energy emerged as an explanation for the accelerated expansion of the Universe revealed by studies of distant Supernovae of type Ia (SNe Ia) at the end of the XXth century. Until now the nature of this component is unknown and a large scientific effort is being held to characterize it.

Supernova studies provide a robust measurement of the expansion of the Universe in the form of a Hubble diagram, luminosity distance as a function of redshift. SNe Ia possess rest-frame light curves which are observed to be similar. Moreover, the photons from these SNe Ia are affected mainly by the traveled distance, not depending on the clustering of matter, providing us with "standard candles".

This work is based on data from the Supernova Legacy Survey (SNLS), a second generation experiment designed for detecting and measuring precisely SNe at high redshift, in a range between 0.2 and 1, which is the interesting range for studying the expansion of the Universe. SNLS collected 5 years of data from which only 3 have been completely analyzed. The complete 5-year analysis is still ongoing in 2015.



I worked on the deferred photometric pipeline of SNLS set up in Saclay. As its name indicates, it is an independent pipeline and only uses photometric information to detect and classify SNe Ia.

In the era of large future surveys, spectroscopic resources will be limited for candidate follow-up and classification, which makes photometric pipelines interesting to study. The SNLS deferred photometric pipeline provides a 100% photometric sample. The advantages of a photometric sample include larger number of events classified as type Ia, larger redshift coverage and no need for spectroscopic observations.

My work was based on improving the detection of transient events and the classification of supernovae using supernova photometric redshifts in the SNLS deferred analysis. This was done in the view of a complete 5-year SNLS photometric analysis. It must be highlighted that the classification of SNe Ia using supernova photometric redshifts is unprecedented.

The first step for detecting SNe events is to make a sample of transient events to be later classified. Detection using only photometry with difference images in one filter, where a reference image is subtracted, provides a good approach. However, difference images are filled with various artifacts from instrumental defects and incomplete subtraction of permanent objects. Disentangling real transient events from artifacts becomes an important requirement especially for photometric only pipelines. This is also of interest for future surveys which will process large amounts of data, such as LSST which expects to detect one million SNe per year.

In the first part of this work, I will present a new approach for improving transient event detection based on morphological component analysis for difference image stacks in the SNLS deferred processing. Our goal is to obtain a reduction of the number of detections while limiting the loss of SNe Ia in the detection sample. We exploit the different morphologies of objects in the stacks to separate transient objects from artifacts.

In the second part of this work, I will present the photometric classification of a SN-like sample into type Ia and core-collapse supernovae. I explore a new classification using photometric supernova redshifts and different classification strategies, from sequential cuts to machine learning techniques.

---

The improvements and new methods in this work were made in the view of a complete 5-year SNLS deferred photometric analysis that will provide a 100% photometric SNIa sample at high redshift of the order of 1,000 events.

## Chapter 2

# Physical Universe: cosmology and acceleration of expansion

*“Cosmology sits at the crossroads between theoretical physics and astronomy.”*

*Jean-Philippe Uzan*

Cosmology, as the study of the Universe’s evolution and contents, is a field where there is constant interaction between theory and observations. Based on physical laws, theoretical physics tries to describe nature’s components and their interactions. Meanwhile, astronomical observations require introducing hypotheses in order to conciliate phenomena with physical theories. In this chapter we will introduce both the theoretical part of a cosmological model and observations that have contributed to modify these models.

To study the evolution of the Universe we need first to construct a cosmological model. Basically, what we want is to have a set of equations describing the components of the Universe and how they interact, and to solve them for a particular case which is our Universe. The elements needed for constructing such a cosmological model are:

- i. a gravitational theory,
- ii. symmetry hypotheses for the large scales,
- iii. a hypothesis on the topology of the Universe or its global structure,
- iv. a description of the components of the Universe and their non-gravitational interactions.

Our current cosmological model is the  $\Lambda$ CDM model which assumes the gravitational theory to be General Relativity which will be introduced in Section § 2.1. For the large scale hypothesis we assume the Copernican or Cosmological Principle presented in Section § 2.2. In Section § 2.3 we will introduce the space we live in, its topology, where it is assumed that spatial sections are simply connected. The components of our Universe are treated in Section § 2.6, which are the Standard Model of particle physics plus, added due to observational evidence, dark matter and dark energy. Finally, we write the equations that describe the dynamics of our Universe in Section § 2.7.

## 2.1 Rise of General Relativity

To construct physical laws the first question that arises is: does every observer throughout the Universe see the same phenomena? This question has motivated a postulate on which every physical theory is based:

*The postulate of relativity: “The laws of physics are invariant in all inertial reference frames.”*

In the 18th century Newton proposed his theory of gravitation where he defined special reference frames, called “inertial frames”. To go from one inertial frame to another one the Galilean transformation was used. Following the relativity postulate:

*The principle of Galilean Relativity: the laws of motion under Galilean transformations are invariant.*

One century later, Maxwell presented his empirical theory of electrodynamics. A novelty was that the statement that the velocity of light was constant. However, this theory did not satisfy the principle of Galilean Relativity.

With these new laws, Einstein proposed in the early 1900’s to replace Galilean transformations with Lorentz transformations. The latter left Maxwell equations invariant. However, Newton’s gravitation theory did not fulfill this invariance. Einstein proposed a modification of the laws of motion to make them Lorentz invariant. He proposed:

*The principle of special relativity: all physical equations must be invariant under Lorentz transformations.*

Under this principle he formulated the theory of special relativity which postulates that for all observers, the speed of light in a vacuum is the same.

Then, in 1907 Einstein introduced one of the cornerstones of his theory:

*The principle of equivalence of gravitation and inertia: “at a very space-time point in an arbitrary gravitational field it is possible to choose a locally inertial coordinate system such that, within a sufficiently small region of the point in question, the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation” (pg. 68 [1]).*

Finally, in 1916 Einstein published “The Foundation of the General Theory of Relativity” where he described what is now our gravitation theory. The main principles behind **General Relativity** (GR) are:

- the inertial and gravitational masses are equal,
- the inertial and gravitational forces are equivalent,
- special relativity laws without gravitation are valid in a local inertial frame.

General Relativity relates the presence of energy to distortions in the space-time metric. This theory provides the following Einstein equations:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G T_{\mu\nu}}{c^4}, \quad (2.1)$$

where  $G$  is Newton’s gravitation constant,  $g_{\mu\nu}$  the metric,  $T_{\mu\nu}$  is the stress-energy tensor,  $R_{\mu\nu}$  the Ricci tensor,  $R$  the Ricci scalar and  $\Lambda$  the cosmological constant.

### **A word about the cosmological constant:**

At the beginning of the 20th century, the accepted view was that we lived in a static Universe. To account for this Einstein introduced a cosmological constant in his equations. However, observations by Hubble in 1929 showed that the Universe was expanding and there was no need for a cosmological constant that made the Universe static. A

cosmological constant was introduced again at the end of the 20th century to account for the accelerated expansion of the Universe. The latter is part of the current cosmological model.

Einstein principles set up the basis for GR which is our current **gravitational theory**. Solutions for Einstein's equations can be found assuming some symmetries. In the following, our symmetry hypotheses are presented.

## 2.2 Symmetry: Cosmological Principle

The main hypothesis on our theoretical cosmology is the “cosmological principle”:

*The Universe appears the same to all observers at any epoch regardless their individual location.*

It implies that the Universe is spatially homogeneous and isotropic. This is a very strong principle and one of the cornerstones of our cosmological model, it even determines the structure of regions that cannot be observed. Since there are structures and astrophysical objects such as stars and galaxies in our Universe, this principle is considered as valid statistically (viewed on a large enough scale, that is above  $\approx 100\text{Mpc}$ ). The cosmological principle is linked to the “Copernican principle” that states that we do not live in a special place or center of the Universe.

Statistical isotropy and homogeneity have been observationally supported by, for example, measurements on the Cosmic Microwave Background (CMB) anisotropies [2] and the large scale galaxy distribution. Two maps of the latter are shown in Figure 2.1.

## 2.3 Space: Manifolds and metric

The global structure of our Universe, our topology, is described by the way we define distances. The first assumption is that spatial sections are simply connected.

The topological space we live on is a 4-dimensional manifold that resembles Euclidean space near each point. To define distances between two events on this manifold we

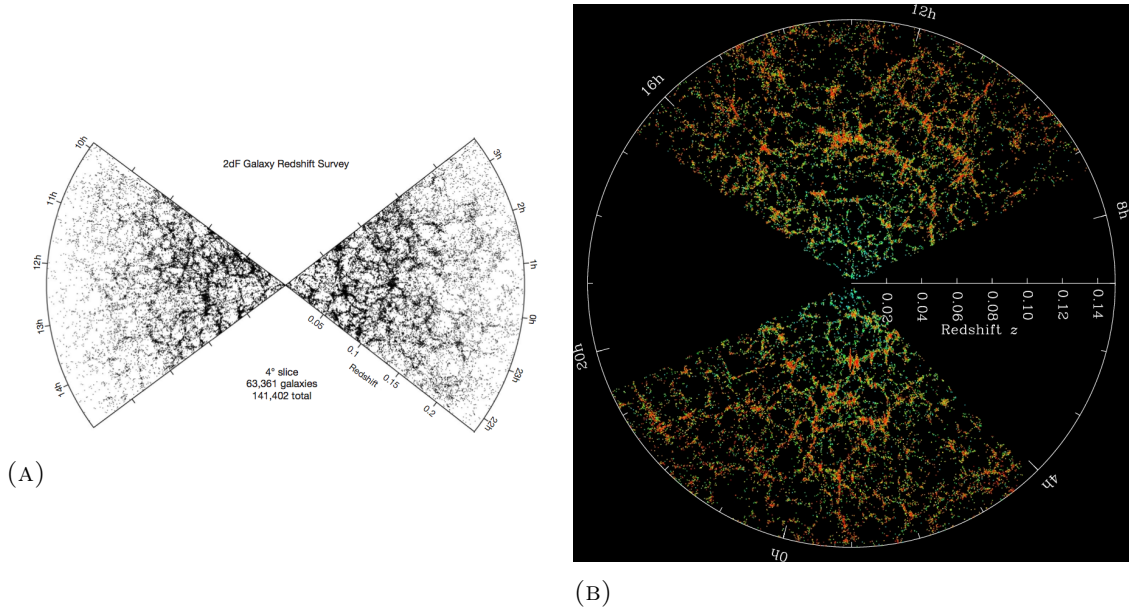


FIGURE 2.1 Galaxy distribution from 2dFGRS [3] and SDSS [4]. In 2.1b galaxies are colored by their star ages, the redder the older stars. The homogeneity and isotropy at large scales are evidenced through these distributions.

require a coordinate system and a metric tensor. An event is described on a coordinate system by a 4-vector with time and space components,  $x = (x^0, x^1, x^2, x^3)$ . The metric tensor  $g_{\mu\nu}(x)$ , which allows to transform coordinates, is symmetrical by construction. The interval between two events  $x$  and  $x + dx$ , also called proper time, is given by:

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu . \quad (2.2)$$

The metric tensor can be dependent on the position where it is considered. In particular, space can be curved and be locally Euclidean but not globally. The metric also allows to incorporate gravity, not as an external force but curving or distorting space-time. In spherical coordinates the proper time can be written as:

$$ds^2 = dt^2 - a(t)^2 (d\chi^2 + f^2(\chi)d\Omega^2) , \quad (2.3)$$

where  $d\Omega^2 = (d\theta^2 + \sin^2(\theta)d\phi^2)$  and

$$f(\chi) = \begin{cases} \sin(\chi) & \text{3-sphere with positive curvature} \\ \chi & \text{plane with zero curvature} \\ \sinh(\chi) & \text{3-hyperbola with negative curvature} \end{cases} \quad (2.4)$$

and  $a(t)$  is the scale factor that accounts for the Universe expansion. These three choices for  $f(\chi)$  represent three types of homogeneous and isotropic universes. The 3-sphere represents a closed universe with a finite volume while the 3-hyperbola is an open universe with infinite volume.

Currently observational measurements show that we live in a flat universe. Data from the Cosmic Microwave Background (CMB) by Wilkinson Microwave Anisotropy Probe in 2003 [5],[6] and then by Planck in 2015 [2] combined with BAO data are consistent with spatial flatness.

## 2.4 Dynamic universe: Friedmann-Lemaître-Robertson-Walker Metric

To describe the dynamical universe we can derive the Friedmann-Lemaître-Robertson-Walker metric using the metric on Equation 2.3 and the cosmological principle. The cosmological principle requires homogeneity and isotropy at every time, then the metric can vary with time. With the change of variable  $r = f(\chi)$  in equation 2.4 we obtain:

$$ds^2 = dt^2 - a(t)^2 \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right), \quad (2.5)$$

with  $k$  of values  $+1$  for a 3-sphere,  $-1$  for an hyperbola and  $0$  for a plane. This is the Friedmann-Lemaître-Robertson-Walker metric.

The scale factor  $a(t)$  sets the scale of the geometry of space. To quantify the change in the scale factor we define the Hubble parameter:

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (2.6)$$

The current Hubble constant is denoted  $H_0$  and can be written using Equation 2.6 with the current time as  $t_0$ . The convention is that for  $t_0$ ,  $a_0 = 1$ .



## 2.5 Expanding universe

The Universe has three possible states according to the sign of  $\dot{a}$ . If  $\dot{a} > 0$  the Universe is expanding, if  $\dot{a} = 0$  it is static and if  $\dot{a} < 0$  it is contracting.

At the beginning of the 20th century the Universe was thought to be static. However, with the galaxy distance measurements by Hubble in 1929 the idea of an expanding and changing universe was born. He measured the distances to “extra-galactic nebulae”, that is galaxies, outside our Milky Way and compared them to their redshift [7]. At that time, he associated the redshift to velocities through the Doppler formula. He found a linear relation between these distances and velocities seen in Figure 2.2. This linear relation is known as the Hubble law and is given by the expression:

$$v = H_0 d . \tag{2.7}$$

The measurement by Hubble in 1929 yielded an approximate value of  $H_0 = 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$  which was largely overestimated due to bad distance calibration. Note that Hubble’s measurements were within a maximum redshift of  $z = 1000 \text{ km s}^{-1} / c \approx 0.003$ . The value of  $H_0$  has been estimated using different physical measurements with some disagreements between determined values. Direct measurements by the Hubble Space Telescope provided a measurement of  $H_0 = 73.8 \pm 2.4 \text{ km/s/Mpc}$  [8]. Indirect measurements, or constraints in the  $\Lambda\text{CDM}$  model, from combining data from WMAP9, SPT, ACT and BAO gave a value of  $H_0 = (69.6 \pm 0.7) \text{ km/s/Mpc}$  [9], while Planck 2015 data provided a value of  $H_0 = (67.6 \pm 0.6) \text{ km/s/Mpc}$  [2]. Comparisons between some  $H_0$  determinations can be seen in Figure 2.3.

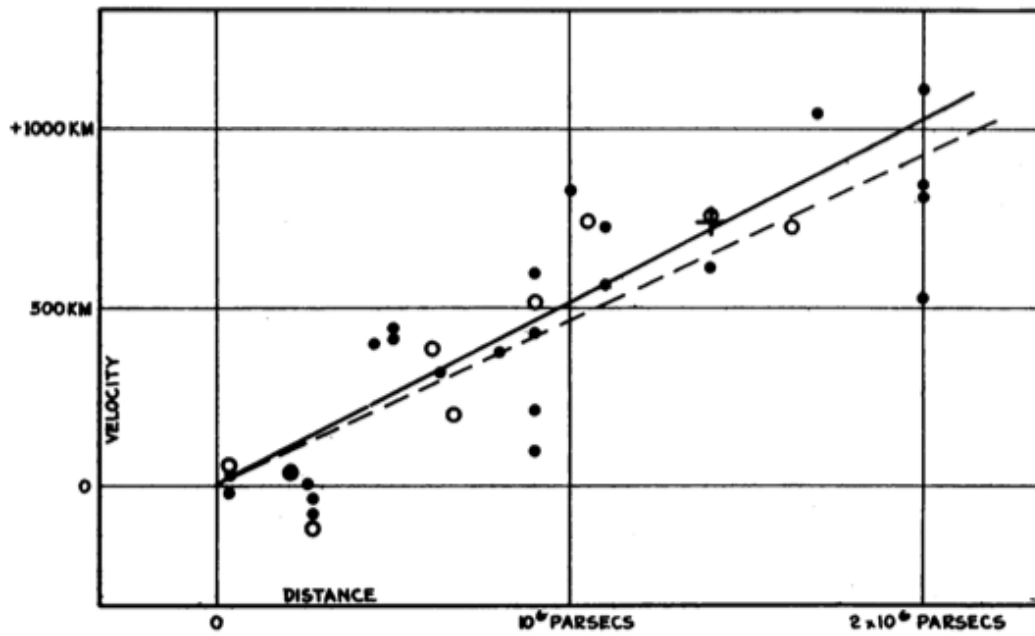


FIGURE 2.2 Original diagram by Hubble in 1929 showing velocity versus distance for galaxies. Due to bad distance calibration these measurements led to an overestimated Hubble constant. The original caption was: “Velocity-Distance Relation among Extra-Galactic Nebulae. Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually.” [7]

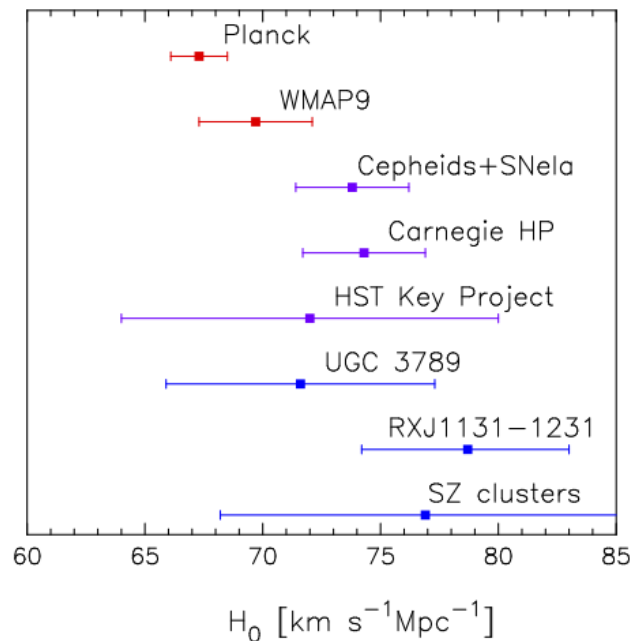


FIGURE 2.3 Comparison of  $H_0$  measurements from different data samples and techniques. Errors are  $1\sigma$  estimates. [2]

For higher redshifts as well, the most direct evidence of the expansion of the Universe is given by the Hubble diagram. While measuring redshifts can be quite straight forward, determining distances can be a challenge. Standard candles, objects with the same intrinsic brightness, provide a way of measuring distances using what is called *luminosity distances*. Examples of standard candles are Cepheid variable stars which intrinsic brightness is related to their period and type Ia Supernovae which are the objects used in this work and described in Chapter § 3. Brief definitions of cosmological redshift and luminosity distance are presented in the following.

### **Cosmological Redshift**

In an expanding universe, the traveling light frequency is modified by the changing metric. This change is called *redshift* and can be defined as:

$$1 + z = \frac{\nu_e}{\nu_o} , \quad (2.8)$$

where  $\nu_e$  is the frequency of light emitted and  $\nu_o$  is the observed one. Taking the FLRW metric in Eq. 2.5, assuming that the scale factor changes negligibly during a single period and that the redshift is only due to the Universe's expansion:

$$1 + z = \frac{a_o}{a_e} , \quad (2.9)$$

where  $a_o$  is the scale at the observation time (usually taken as  $a_o = 1$ ) and  $a_e$  the scale at the emission time. Note that there is no dependence on the evolution of the scale factor at intermediate times. The redshift depends only on the scale factor at the time of emission and observation, it is a direct probe of the ratio of the scale factor at different times.

**Luminosity distance**

For an object with intrinsic luminosity  $L$ , the observed flux  $F$  can be expressed as:

$$F = \frac{L}{4\pi d^2}, \quad (2.10)$$

where the intrinsic luminosity (assuming the photons to have the same energy) is given by:

$$L = \frac{N\nu}{dt}, \quad (2.11)$$

with  $N$  is the number of photons,  $\nu$  their frequency and  $dt$  the infinitesimal emission time span.

If the Universe was static, the  $d$  in Equation 2.10 would represent the radius of the sphere for which the object with luminosity  $L$  would give a flux  $F$ . However, we know that the Universe is not static and this equation should be transformed to a comoving shell. We can write the flux observed as a function of the comoving spherical shell radius,  $f(\chi)$  as:

$$F = \frac{L_o}{4\pi f(\chi)^2}, \quad (2.12)$$

where  $L_o$  is the observed luminosity. The observed luminosity can be related to the emitted (or intrinsic) luminosity  $L$  taking into account the frequency change for a photon by the changing metric in Equation 2.8 and the time dilation  $dt_o = (1 + z)dt$ . The observed luminosity  $L_o$  can be expressed as:

$$L_o = \frac{L}{(1 + z)^2} \quad (2.13)$$

Then, it is useful to define the luminosity distance as:

$$d_L = (1 + z)f(\chi), \quad (2.14)$$

where  $f(\chi) = \chi$  if we are in a flat universe as in equation 2.4. Where the flux can be expressed as:

$$F = \frac{L}{4\pi d_L^2}. \quad (2.15)$$

Luminosity distance depends on the energy content of the Universe and this dependence will be seen in Section § 2.7.

## 2.6 Universe's content

In this section I will present our Universe's different components of matter and energy, equivalent due to the well known  $E = mc^2$  equation.

We can apply the first law of thermodynamics to our dynamic Universe of total energy density  $\rho$ , with pressure  $p$ . This gives the variation of energy  $dU$  of a system for an adiabatic change of volume  $dV$ :

$$dU = -pdV . \quad (2.16)$$

Rewriting 2.16 as a function of the scale factor and expanding the first derivative, we obtain the equation of energy conservation in our expanding Universe:

$$\frac{d\rho}{dt} = -3H(\rho + p) . \quad (2.17)$$

Note that for each component in the Universe this equation should be individually treated. Each content can be simplified treating it as an effective fluid and relating the energy density of the fluid to its pressure through its state equation

$$p = \omega\rho , \quad (2.18)$$

where  $\omega$  is the equation of state parameter.

Two big categories of contents can be defined, relativistic (radiation) and non-relativistic matter which will be introduced in the following. However, recent evidence points towards a third type of content which is due to the so called **cosmological constant**. This is **dark energy** that will be introduced as a third category.

### 2.6.1 Matter: baryons and dark matter

Non-relativistic matter has no pressure and is denoted with the subscript  $m$ . From Equation 2.18, this means that  $\omega = 0$ . The energy density of matter is dominated by its mass energy. Then, the energy conservation equation becomes:

$$\frac{d\rho_m}{dt} = -3H\rho_m , \quad (2.19)$$

when integrated there is an inverse proportionality between its density and the scale factor as

$$\rho_m \propto a^{-3} . \quad (2.20)$$

Originally, non-relativistic matter included only ordinary (**baryonic**) matter. However, observations in the early 20th century by Zwicky and Smith [10] of the velocities of galaxies in the Coma and Virgo clusters provided the first hints for another type of non-relativistic matter. At that time, estimations of the total mass required to gravitationally bind the galaxies were two orders of magnitude above Hubble’s estimation of the galaxy mass. This led Zwicky to postulate the existence of an “invisible” matter that interacts gravitationally. This matter is called **dark** because it doesn’t interact with electromagnetic radiation. Probes like the CM, BAO and SNe also provide information about a larger content of matter in the Universe than the one accounted by baryons [2] [11]. However, no direct or indirect detection of dark matter has been accomplished till the present date. Currently, this matter is thought to be **cold**, meaning that the velocities of the particles are too small to erase structure formation in the early Universe. In the current cosmological model non-relativistic matter is composed of **baryons** and **cold dark matter** (CDM).

### 2.6.2 Radiation

Particles which have velocities close to the speed of light  $c$  are considered as relativistic matter or radiation. Radiation, denoted with the subscript  $r$ , is described by the equation of state  $\rho_r = 3p_r$ . The energy conservation equation for radiation is:

$$\frac{d\rho_r}{dt} = -4H\rho_r , \quad (2.21)$$

integrating

$$\rho_r \propto a^{-4} . \quad (2.22)$$

Radiation undergoes a loss of energy due to the expansion of the Universe but also an additional loss proportional to  $a^{-1}$ .

In the current model radiation is accounted for by **photons** and **neutrinos**.

### 2.6.3 Dark energy as a homogeneous fluid

The effect of the cosmological constant in Einstein’s equations can be reproduced by an homogeneous fluid with pressure  $p_\Lambda = -\rho_\Lambda$  and

$$\rho_\Lambda = \frac{\Lambda c^4}{8\pi G}, \quad (2.23)$$

where  $c$  is usually taken to be 1 for simplification.

Dark energy is considered as an “exotic negative-pressure fluid that provides the impetus for cosmic acceleration.” [12]. The introduction of such a fluid was motivated by the accelerated expansion of the Universe. In 1998, two teams lead by Perlmutter, Riess and Schmidt [13–15] found that the Universe was expanding in an accelerated way that did not agree with a Universe composed only by matter (baryonic and dark matter) and radiation under GR. The Hubble diagram from measurements by *Perlmutter et al.* can be seen in Figure 2.4.

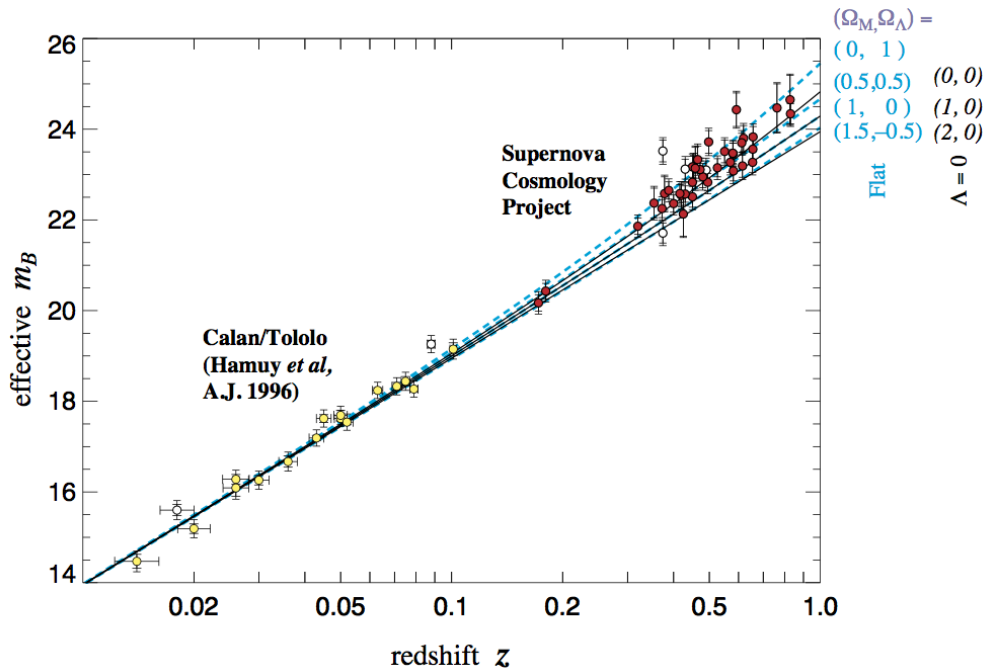


FIGURE 2.4 Hubble diagram showing data from the Supernova Cosmology Project with multiple model lines to compare with data. Data require the addition of a cosmological constant to General Relativity equations. [13]

## 2.7 Friedmann equations

To express the dynamics of our cosmological model we have Friedmann's Equations. Taking the time component on Einstein field equation (Eq. 2.1)  $\mu = \nu = 0$  we obtain, for our flat universe, the **first Friedmann equation**:

$$H^2 = \frac{8\pi G}{3} \left( \rho + \frac{\Lambda}{8\pi G} \right), \quad (2.24)$$

where here  $\rho = \rho_{matter} + \rho_{radiation}$ .

The **second Friedmann equation** can be obtained taking the trace of Einstein's equations (Eq. 2.1)

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3p) + \frac{\Lambda}{3}, \quad (2.25)$$

where  $p = p_{matter} + p_{radiation}$ .

For both Friedmann equations we can introduce the cosmological constant as an energy density  $\rho_\Lambda = \Lambda/8\pi G$ . Then we can define a *total energy density*  $\rho_{total} = \sum \rho_i$  and a *total energy density*  $p_{total} = \sum p_i$  where  $i$  stands for all components in the Universe.

We can define the *critical energy density* as the current energy density in our Universe:

$$\rho_c(t) = \frac{3H_0^2}{8\pi G}. \quad (2.26)$$

For each component  $i$ , the normalized density parameter  $\Omega_i$  is defined as:

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_c(t)}, \quad (2.27)$$

when  $t = t_0$  we have  $\Omega_i^0$ .

As a function of the normalized density parameters  $\Omega_i$ , the first Friedmann equation (Eq. 2.24) can be written as:

$$\left( \frac{H}{H_0} \right)^2 = \Omega_k a^{-2} + \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda, \quad (2.28)$$



where  $\Omega_i$ ,  $a$  and  $H$  are functions of time. Using Equation 2.9 we obtain:

$$\left(\frac{H}{H_0}\right)^2 = \Omega_k(1+z)^2 + \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_\Lambda, \quad (2.29)$$

at  $t = t_0$  this equation becomes

$$\Omega_{r,0} + \Omega_{m,0} + \Omega_{\Lambda,0} = 1 - \Omega_{k,0}, \quad (2.30)$$

where only 3 of the 4  $\Omega$ 's are independent. For our flat Universe we can also find that the total energy density is equal to the critical density

$$\Omega_{r,0} + \Omega_{m,0} + \Omega_{\Lambda,0} = 1 \Rightarrow \rho_{r,0} + \rho_{m,0} + \rho_{\Lambda,0} = \rho_c. \quad (2.31)$$

Using Friedmann equations, luminosity distances in Eq. 2.14 can be expressed as a function of the Hubble parameter which depends on the content of our Universe

$$d_L = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{H(z')/H_0} \quad (2.32)$$

for a flat universe.

This is a very important result since it allows to constrain cosmological parameters,  $\Omega$ s, using an observable, the luminosity distance. Studying the luminosity distance of a *standard candle* as a function of redshift, through a Hubble diagram, the different cosmological models can be tested (see Figure 2.5).

## 2.8 $\Lambda$ CDM

This chapter provides all the ingredients of our current cosmological model, the so called  **$\Lambda$ CDM**. I introduced our current gravitational theory, large scale symmetry hypotheses, a hypothesis on the topology of the Universe and a description of the components of the Universe.

Observations provide data to constrain the model's parameters, the so-called cosmological parameters. Observations are consistent with a spatially flat Universe ( $\Omega_k < 0.005$  [2]) containing : baryonic matter, dark matter, radiation and dark energy. However, the nature of either dark matter or dark energy is unknown motivating research on the area.

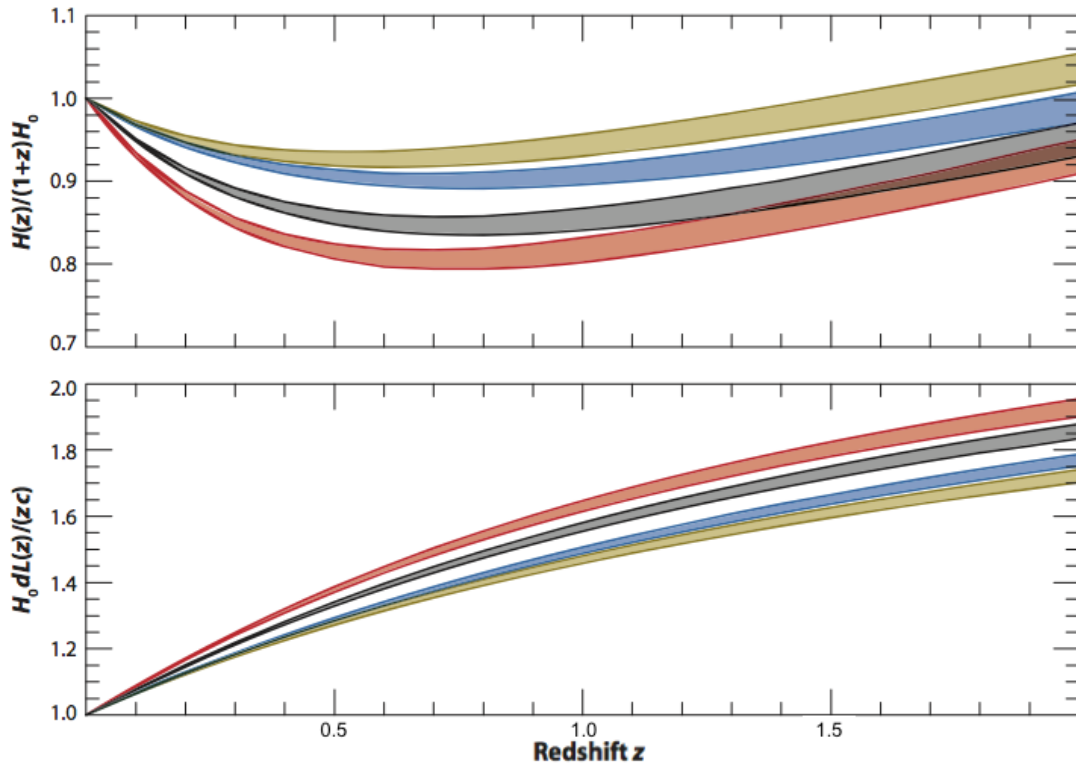


FIGURE 2.5 Examples of different dark energy models with respect the expansion history  $H$  and luminosity distance  $d_L$  as a function of redshift. Dark energy models with different equation of state parameters are shown in red ( $\omega = -1.2$ ), grey ( $\omega = -1$ ) and blue ( $\omega = -1.8$ ). The brown curve is for an alternative gravity theory called Dvali-Gabadadze-Porrati [16]. All models have the same matter density and assume spatial flatness. Uncertainties in the non relativistic matter density are indicated through the curve's thickness [12].

A recent work using SNLS, SDSS-II, HST and several SNIa nearby data (JLA sample) by *Betoule et al.* shows the  $\Lambda$ CDM confidence contours in  $\Omega_\Lambda$  and  $\Omega_m$  for SNe Ia and other probes, see Figure 2.6.

$\Lambda$ CDM is the simplest model that agrees with observations from the accelerated expansion of the universe (e.g. using type Ia supernovae), CMB and large scale structure in the distribution of galaxies. It is the standard model of cosmology and the physical context of my PhD work.

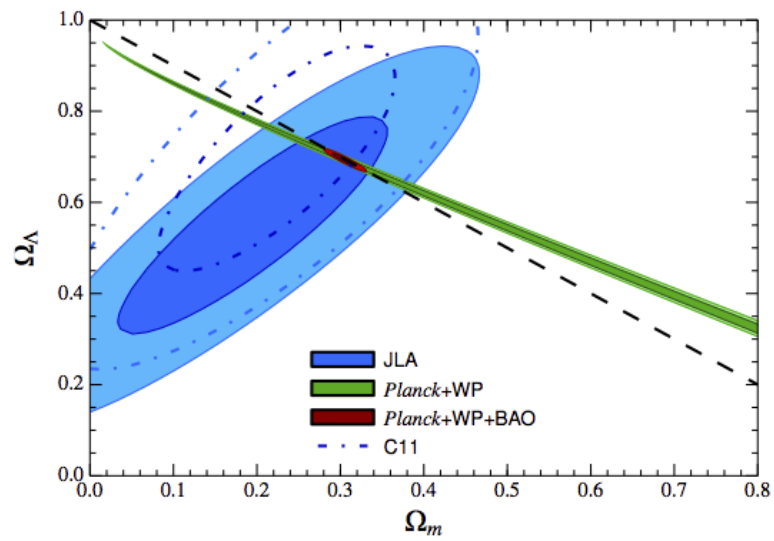


FIGURE 2.6  $\Omega_\Lambda$  vs.  $\Omega_m$  confidence contours (68% and 95%) of the JLA analysis [11]. Contours are shown for type Ia supernovae (JLA with new calibration and all data, C11 the previous compilation by *Conley et al.* [17]), CMB temperature data (Planck), CMB polarization data (WP=WMAP) and Baryon Acoustic Oscillation data (BAO). The dashed line corresponds to a flat universe.

## Chapter 3

# Cosmological Observable: Supernovae of type Ia

To study the expansion of the Universe we need first to be able to measure distances. For this, we require very homogeneous objects with known absolute magnitude, *standard candles*, or objects whose absolute brightness is correlated with other observables, *standardizable objects*. They must be very bright objects that can be detected with current technologies up to high redshift. Such objects are Supernovae of type Ia (SNeIa) which will be described in this Chapter.

Supernovae are defined and classified empirically in Section § 3.1 since their physical mechanisms are still uncertain. Then, in Section § 3.2 I describe SNeIa, their spectral and photometric properties, why they are *standardizable*, their volumetric rate measurement and model, and some of the proposed physical mechanisms behind them. Last, the relation between SN Ia luminosity and distance is shown in Section § 3.3.

### 3.1 Supernovae

Supernovae (SNe) are very luminous stellar explosions that can last several weeks. They have been long observed and studied, but their origins remain an open question. However, their observed properties are well known and can be used to classify them.

The empirical classification of SNe is based both on their spectroscopic and photometric properties. Spectroscopic properties are obtained from the absorption lines on the SNe spectra. Photometric properties are defined through a light curve, which is the variation of the measured flux with respect of time. Light curves can be drawn for one or many broadband filters.

### 3.1.1 Empirical classification

Classification by **spectroscopy** of SNe is done around maximum light. This is due to the evolution of SN spectra which makes easier to observe some features in this time period and to the fact that for distant SNe (redshift above 0.3) the spectrum signal-over-noise ratio is acceptable only around maximum light.

There are two main observational categories of SNe based on the presence or not of hydrogen in their spectra. SNe of type II show hydrogen lines while type I lack them. Different subtypes are found for each category. Further details on the SN spectral classification can be found in [18].

- Type I : no hydrogen lines.
  - Ia : presents silicium lines and no helium. They will be thoroughly described in Section § 3.2.
  - Ib : contains a line of helium.
  - Ic : lacks lines of both helium and silicon.
- Type II : hydrogen lines present
  - II<sub>n</sub>: nominated by narrow lines of hydrogen.
  - II<sub>b</sub>: presents signatures of hydrogen in spectra which fade and are replaced with helium features.
  - II-P and II-L : no particular spectral features, P and L stand for plateau and linear referring to the shape of the post maximum light curve.

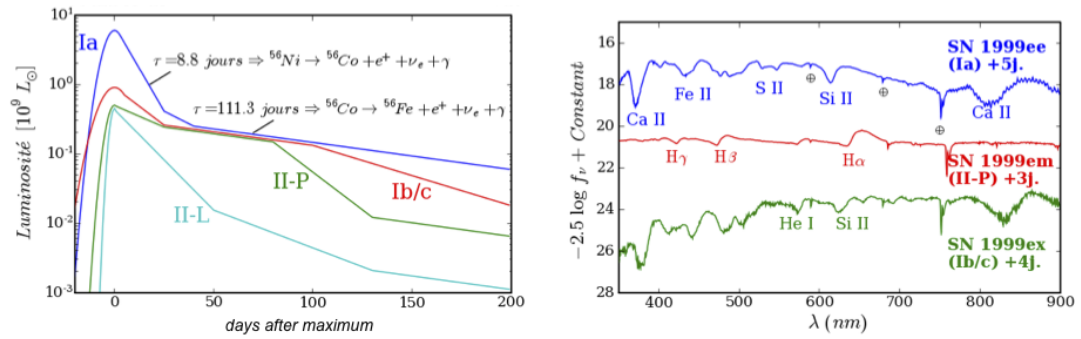


FIGURE 3.1 Left, schematic light curves of SNe Ia, Ib/c, II-P and II-L. Right, spectral lines that allow to classify these types of SNe, from 1999 observations. [19]

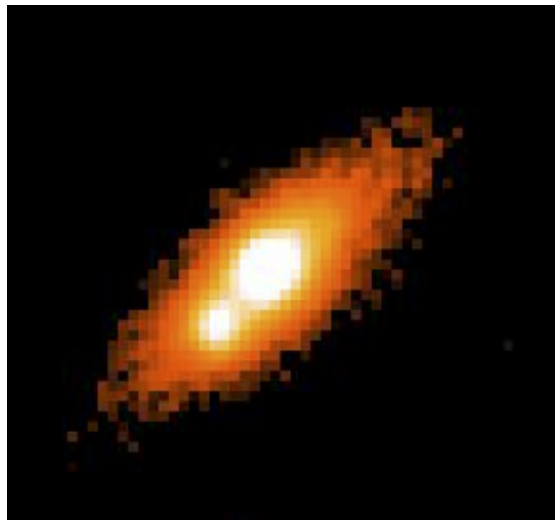


FIGURE 3.2 Galaxy with SNIa *04D1dc* light delivered after explosion on the left side of its center.

## 3.2 Supernovae of type Ia

Type Ia SNe have very homogeneous spectral and photometric properties. Therefore their light curves and spectra are rather reproducible. They are transient events which last circa 60 days and their maximum luminosity is close to their host galaxy luminosity. In Figure 3.2 a galaxy with a SNLS supernova is shown.

SNe Ia have quite homogeneous properties. However there is a dispersion of around 0.3 in magnitude for SNe Ia at the same redshift. We will show in subsection § 3.2.3 that this variability can be reduced by around a factor of two when using measurable features in the SNIa light curves. There are also peculiar SNe Ia that can be identified

by spectral properties and can be either sub or super luminous relative to the bulk of the peak magnitude distribution.

### 3.2.1 Spectroscopic properties

As introduced in Section § 3.1, type Ia SNe possess no hydrogen nor helium lines. Their spectra present lines of intermediate mass elements such as silicium, calcium, magnesium iron and sulfur. Their spectral features, although homogeneous, show some diversity.

Normal SNe Ia like SN1996X [20] and SN1994D [21] have clear spectral features due to Si II, Ca II, S II, O I and Mg II around maximum brightness and then develop lines of Fe II. An example of a normal SNIa spectrum can be seen in Figure 3.3. Peculiar events have a different spectral evolution with some extra features (*e.g.* stronger lines for some elements or extra lines). More details can be found in [22].

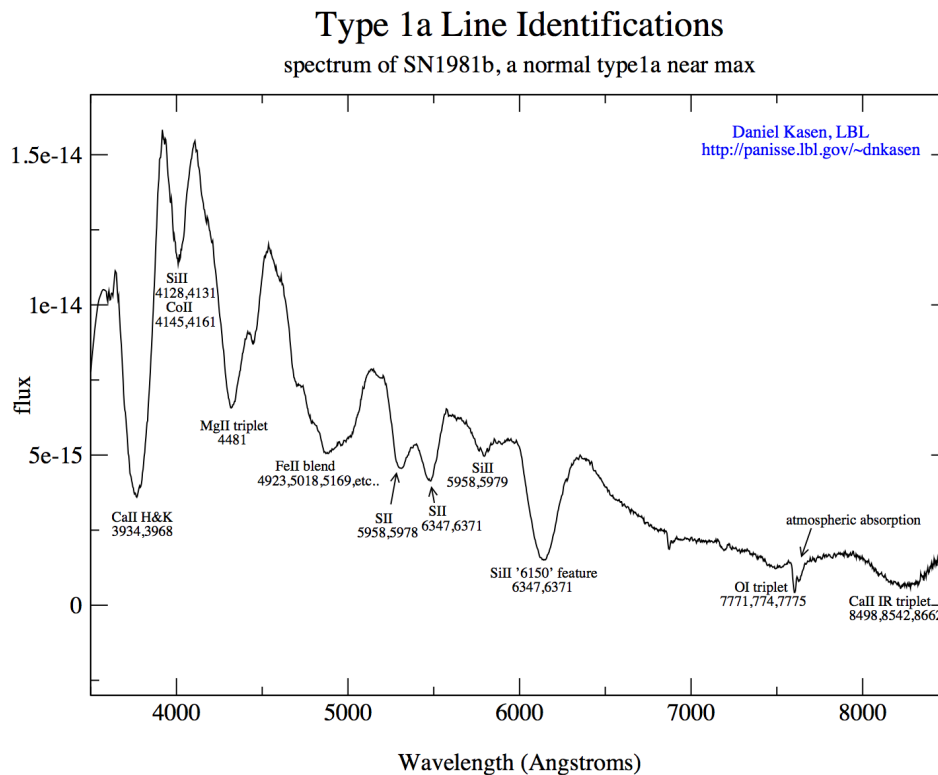


FIGURE 3.3 SNIa spectrum for SN1981b [23].

The characteristics of SNIa spectra are:

- Helium and hydrogen absence.
- Silicon: absorption line of Si II at  $4130\text{\AA}$ . A second Si II line at  $6100\text{\AA}$  has varying depth for some peculiar SNe Ia.
- Sulfur: a “W” shape can be seen due to the doublet of S II at  $5649\text{\AA}$ .
- Calcium: a doublet due to Ca II at  $3934\text{\AA}$  and  $3968\text{\AA}$  is present in all SNe Ia but also in other SNe. In some peculiar SNe Ia this doublet is very weak.

### 3.2.2 Photometric properties

Type Ia supernovae have a varying luminosity that can be roughly modeled through the ballistic expansion of a sphere. In that model the light curve luminosity is powered by the radioactive decay of  $Ni^{56}$  and varies according to the opacity of the ejecta to the decay products. This could explain that the luminosity of a supernova varies in two main phases. At the beginning of the explosion, the emitted light is captured by the star’s high matter density. Then, luminosity increases during circa 15 days (in the object’s rest frame) when density decreases. This period is called the *rising time*. The maximum flux can be as high as  $4 * 10^9$  solar luminosities in the blue band (this band is centered on  $4450\text{\AA}$  and has a bandwidth of  $940\text{\AA}$ ). Then, the luminosity starts slowly to decrease during one month or more. A light curve of an SNIa as detected by SNLS, where these two phases are seen, is shown in Figure 3.4.

Note that, for observations in the red and infrared filters, observed SNIa light curves have a second small peak around 20 days after blue band maximum.

SNe Ia are not standard candles since there is a variability in magnitude for objects at the same redshift. However, we will see that this variability is linked to other observables and thus can be, to some extent, corrected for. I will introduce magnitudes in the following.



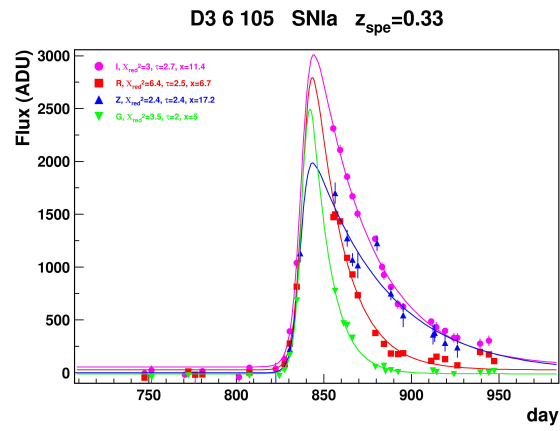


FIGURE 3.4 SNIa light curve from SNLS. Measurements are done in 4 different broad-band filters and are indicated by points, squares and triangles. The lines correspond to fits to measurements. A rising time is clearly seen between days 820 and 850 and then a diming part until day 950.

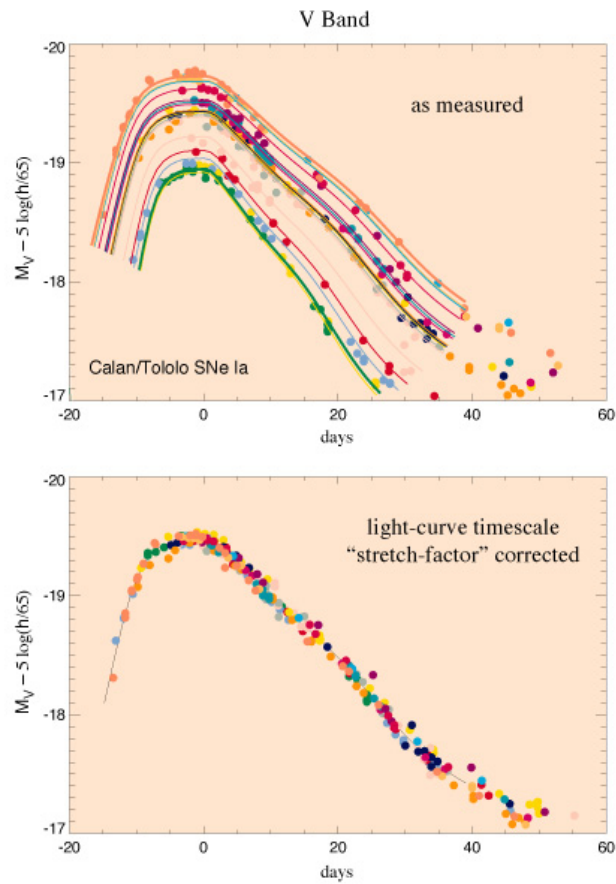


FIGURE 3.5 27 SNIa light curves from the Calan Tololo survey in the V band. On top, observed dispersion of absolute magnitudes in observed SNeIa. Below, the corrected curves [24].

### Magnitudes

The **observed magnitude**,  $m$ , of an object is defined by:

$$m = -2.5 \log_{10} \left( \frac{F}{F_{ref}} \right), \quad (3.1)$$

where  $F$  is the measured flux and  $F_{ref}$  a reference flux. Objects that are more luminous have smaller magnitudes.

Measured magnitudes can be expressed as a function of the luminosity distance  $d_L$  in Equation 2.32 as:

$$m = -2.5 \log_{10} \left[ L \left( \frac{10pc}{d_L} \right)^2 \right] + C, \quad (3.2)$$

where  $L$  is the intrinsic object luminosity and  $C$  a constant.

Magnitudes can be expressed in color bands which are defined according to the survey or standard. In particular, the bands in the *UBVRI* standard are:

- Ultraviolet band  $U$ :  $\lambda_{eff} = 360nm$  and  $\Delta\lambda = 50nm$ .
- Blue band  $B$ :  $\lambda_{eff} = 430nm$  and  $\Delta\lambda = 72nm$ .
- Green band  $V$ :  $\lambda_{eff} = 550nm$  and  $\Delta\lambda = 86nm$ .
- Red band  $R$ :  $\lambda_{eff} = 650nm$  and  $\Delta\lambda = 133nm$ .
- Infrared band  $I$ :  $\lambda_{eff} = 820nm$  and  $\Delta\lambda = 140nm$ .

One can define a **rest-frame B magnitude** (at peak luminosity for our SNe Ia) as:

$$m_B^*(z) = -2.5 \log_{10} \left[ L \left( \frac{10pc}{d_L} \right)^2 \right] + C = 5 \log_{10} \frac{d_L(z)}{10pc} + M_B, \quad (3.3)$$

where  $M_B$  is the absolute B-band magnitude of the supernova. If all type Ia supernovae were identical, this  $M_B$  (for all SNeIa) and  $m_B^*$  (for same redshift SNeIa) should be the same for these objects. However, we know there is an intrinsic absolute magnitude dispersion to be accounted for. This will be addressed in the following.

### 3.2.3 Standardizing SNeIa

SNeIa are not *standard candles*, however they can be *standardized* (estimate a common intrinsic luminosity or absolute magnitude  $M_B$ ). For this, we can use correlations between photometric observables, or variability of the light curve, and the intrinsic luminosity. The two main correlations are: the width-luminosity relation and the brighter-bluer relation.

The variability of the light curve can be expressed using its **stretch** and **color**. Chromatic differences are expressed through color, while stretch corresponds to the time span of the curve. Light curves can be modeled using a luminosity parameter, a decline rate parameter and a single color parameter. This parametrization will be treated in Section § 3.3.

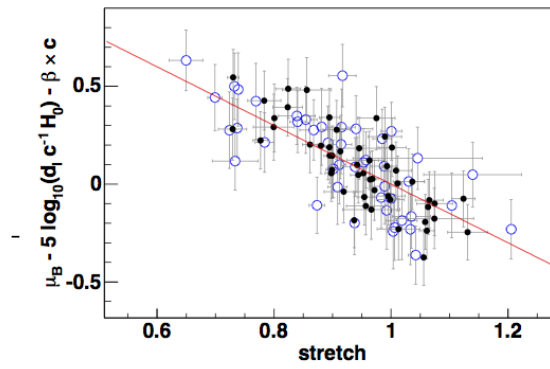
The width-luminosity relation (also known as **brighter-slower**) shows that brighter SNe Ia have broader light curves than fainter ones [25]. The equivalent is that brighter SNe Ia have a slower decline rate than fainter ones (Figure 3.6a).

The **brighter-bluer** relation provides a correlation between color and SNe Ia luminosity (Figure 3.6b). Apart from the intrinsic color, some authors attribute part of this correlation to the reddening caused by dust in the host galaxies [26].

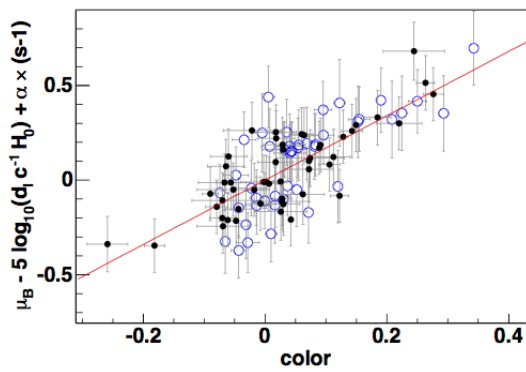
Using both correlations we are able to standardize SNe Ia and to use them as distance indicators, as we will see in Section § 3.3. However, there is a remaining scatter in their absolute peak magnitudes up to  $\approx 0.15$  mag [28, 29].

Other correlations may be useful to reduce the remaining absolute magnitude dispersion. For example, a correlation between SNIa stretch and their host-galaxy mass was found in [30] using SNLS data as can be seen Figure 3.7. Also, it was proposed by [30] to split the SNe Ia sample in two according to host-galaxy mass, due to a seen offset in Hubble residuals for the two samples.

The search for other correlations to reduce the absolute magnitude dispersion is an ongoing research field.



(A) Brighter-slower correlation. Residue on the Hubble diagram for events after being corrected for the brighter-bluer relation.



(B) Brighter-bluer correlation. Residue on the Hubble diagram for events after being corrected for the brighter-slower relation

FIGURE 3.6 SNe Ia properties correlations as seen by SNLS. Blue circles stand for nearby events and black for distant ones [27].

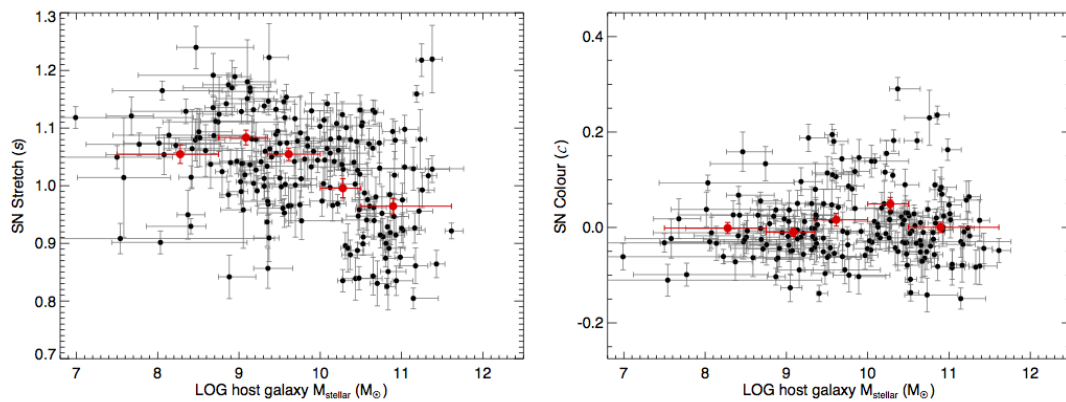


FIGURE 3.7 Stretch and color of SNLS SNIa light curves as a function of their host galaxy stellar mass. The weighted mean for stretch and color are shown as red points [30].

### 3.2.4 SN Ia rates

The SNIa rate is the number of SNe Ia occurring in a space region during a time frame or equivalently in a given volume of the Universe. Measuring this rate and fitting it to a model provides useful information that can constrain the physical mechanisms of SNe Ia and their progenitor models.

The first attempt to measure the SN rate was done by Zwicky (1934-38): “The average frequency of occurrence of supernovae is about one supernova per extra-galactic nebula per six hundred years” for the local volume [31–33]. Recent surveys as SNLS, SDSS, VIMOS, HST have also made effort to measure the SNIa and core-collapse rates.

A very simple model for the SNIa rate was presented by *Pritchett et al.* [34] which we will use to approximate the rate in Chapters § 7 and § 8. The authors argue that the rate of SN Ia explosions is  $\approx 1\%$  of the stellar death rate, independent of star formation history. The SNIa rate is delayed respect to the star formation rate and is expressed as a function of  $(1+z)^\alpha$  where  $\alpha = 2$  [35].

A second model called the two-component model, agrees very well with data at different redshifts. Proposed by *Scannapieco et al.* [36], this model for SNeIa rate contains a component dependent on the star formation rate (SFR) and another dependent on the host galaxy mass:

$$r_V(t) = AM_\star(t) + B\dot{M}_\star(t), \quad (3.4)$$

where  $M_\star$  is the stellar mass of the host galaxy,  $\dot{M}_\star$  is the SFR,  $A$  and  $B$  are constants that can be fit on data and are respectively given by SNe Ia per year per unit mass and SNe Ia per year per unit star formation [36].

High-redshift SNe Ia rates and properties were measured as a function of the stellar mass and star formation in their host galaxies using SNLS data by *Sullivan et al.* [37]. Also for SNLS data, *Neill et al.* in [38] computed the rate for a redshift around 0.5 providing the  $A$  and  $B$  parameters in Equation 3.2.4. The authors in [39] later studied the rate evolution for a larger sample of SNe Ia for three redshift bins between  $0.2 < z < 0.75$ . Figure 3.8 shows the measured SNIa volumetric rate as measured by different experiments, among which SNLS. The measurements are compared with the fit from the two component model.

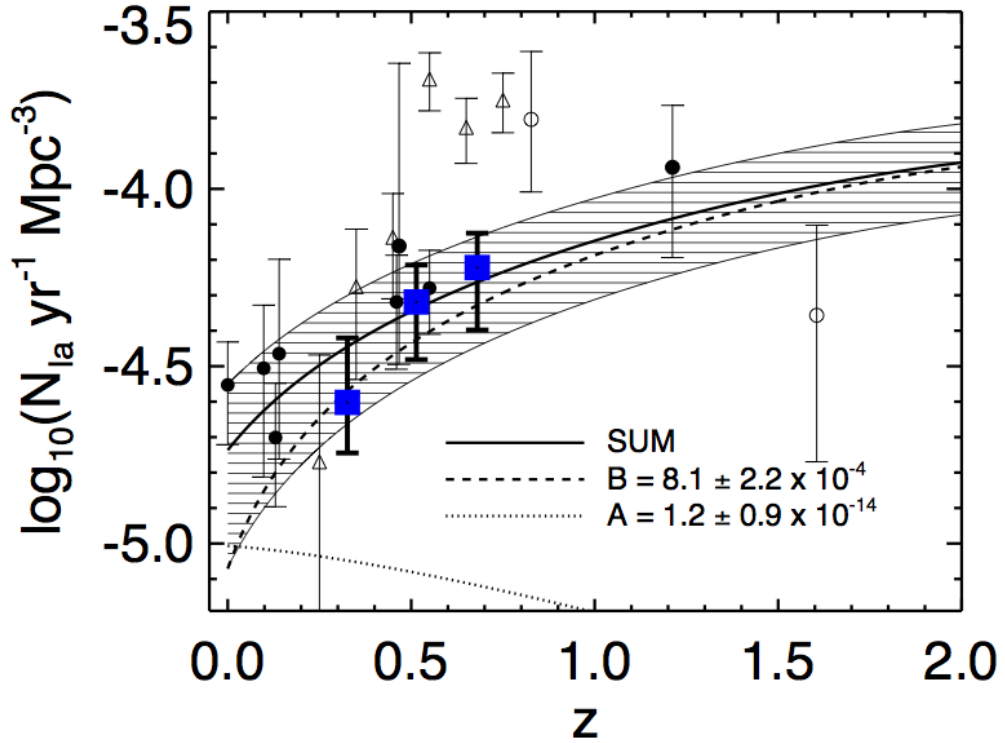


FIGURE 3.8 Observed SN Ia volumetric rate evolution with fitted parameters for model in Equation 3.2.4. SNLS rates are shown as filled blue squares for three redshift bins. Other rates are shown for previously published SN Ia rates derived from samples primarily confirmed by spectroscopy (filled circles), with only 50% spectroscopic confirmation (open circles) and only confirmation through photometric methods (open triangles) [39].

The two-component model implies that at higher redshift the SNIa rate is dominated by star formation. Data at high redshift ( $z > 1$ ) has higher systematics which prevents an accurate SNIa rate to be measured and compared to the model. Until present time, the two-component model remains the one that fits most data at high and low redshifts.

### 3.2.5 Proposed mechanisms

As previously mentioned, the possible mechanisms to explain SNe Ia are still debated. SNe Ia are widely accepted to be thermonuclear explosions of a carbon-oxygen white dwarf (CO WD). The explosion is produced when the CO WD approaches the mass limit of Chandrasekhar. This is the mass at which the electron degeneracy pressure in a WD (equivalent to thermal pressure in a main sequence star) is not enough to resist gravitational collapse. A typical CO WD has a mass of up to  $1.2M_{\odot}$  and to acquire the Chandrasekhar mass,  $1.4M_{\odot}$ , must accrete matter from another object [40].

Since no SNIa progenitor system has been observed until now, based on the explosion's observables there are two main options for the companion object, either another CO WD or a non-degenerate stellar companion (e.g. a main sequence star, a red giant). The model with two WD is called **double degenerate** scenario (DD) while the one with a WD and a companion star is called **single degenerate** (SD) scenario.

CO WD explosions are coherent with observations of SNIa spectra and light curves due to:

- Similar event luminosities: explained by the similar critical mass required to trigger the stellar explosion.
- Light curve timescale: consistent with the half life of  $Ni^{56}$  which is the end of the carbon-oxygen fusion chain.
- Other elements found in the spectra: silicium, sulfur, calcium are products of the carbon-oxygen fusion.
- Lack of helium and hydrogen lines: few stellar objects lack hydrogen and helium, among which CO WD.

### 3.2.5.1 Single degenerate model (SD)

This model proposes a binary system consisting of one white dwarf and another less compact star. The WD accretes mass from its companion until it approaches the Chandrasekhar mass where the Fermi degeneracy pressure can not sustain the non-rotating matter. Some fine tuning is required since the accretion flow has to be self-regulated to avoid self ignition of the material or a red-giant expansion of the WD [40]. When the C-O fusion threshold is reached, the temperature of the star diverges and it explodes. The bulk of the star goes through a runaway fusion reaction that converts C and O into intermediate and heavy mass elements.

This scenario is questioned by astronomical observations and modeling due to:

- Required fine tuning of the self-regulation accretion flow.
- Lack of relevant evidence of spectral lines due to the accreted material: some of the material of the companion star should be burnt during the explosion. For

example, a  $H_\alpha$  emission in spectra should be observed after the explosion but until now no positive detection has been found [41]. However, for some events sodium D absorption lines have been found in SNIa spectra favoring this model [40].

- No observation of the companion: most explosion models indicate that the companion star should survive. In the case of the Tycho supernova remnant, which is 400 years old, no object has been clearly identified as shown in [42]. For the recent and nearby supernova SN 2011fr, no progenitor was detected by *Li et al.* [43]. However it is argued that the progenitor may be too faint to be detected.

### 3.2.5.2 Double degenerate model (DD)

This model was motivated by the shortcomings of the SD one. It proposes that a binary system of two WDs merge after losing angular momentum and energy into gravitational waves. Since WDs are very compact objects, general relativity corrections to the orbits are important. There are two possible outcomes for this system, either the merger (result of the WDs merging) is over the Chandrasekhar mass and explodes, or a more massive WD accretes material from its lighter companion, approaches the Chandrasekhar limit and explodes.

Some studies argue that the merging of two WDs would evolve towards an accretion-induced collapse and not to a SNIa [44][45]. Other studies advocate that under certain conditions a SNIa would be possible [46] [47].

Measurements of SNIa rates are believed by some authors to be consistent with the DD model as the dominant formation channel for long delay times in [48].

It must be highlighted that the agreement on a SNIa progenitor model is still under debate.

## 3.3 Distance measurements with SNe Ia

In previous sections we have seen that SNe Ia can be *standardized* to a common absolute magnitude thanks to the brighter-slower and brighter-bluer relations. We can parametrize a SNIa light curve as a function of: a **luminosity** parameter, a **color** parameter and a decline rate parameter as the **stretch**.



We also know that the apparent rest-frame  $B$  magnitude can be expressed as a function of the luminosity distance and the absolute magnitude as in Equation 3.3. To account for the brighter-slower and brighter-bluer relations we can define the **distance modulus** as  $\mu_B$ :

$$\mu_B = m_B^* - M_B + \alpha x_1 - \beta C \quad (3.5)$$

where  $M_B$  is the absolute magnitude of any SNIa,  $m_B^*$  is the apparent magnitude defined in Equation 3.3,  $x_1$  is the stretch parameter and  $c$  is the color. The last three parameters are derived from a fit to the observed SNIa light curve.  $M_B$ ,  $\alpha$  and  $\beta$  can be constrained during cosmological fits. Absolute luminosity to host-galaxy mass correlations can be taken into account in Equation 3.5 by allowing two different  $M_B$  for masses above and below  $10^{10} M_\odot$ .

*Guy et al.* in SALT 2 <sup>1</sup> constructed an empirical model of SNIa light curves which allows to measure distance moduli of these objects [29]. It is a model of the expected SN flux which varies according to wavelength, decline-rate (equivalent to stretch) and color. SALT2 was trained and tested on hundreds of well measured nearby and distant SNIa spectra and light curves.

Using light curves in several passbands and a redshift assignment, SALT2 extracts for each SN its intrinsic properties. It provides:

- the rest-frame B magnitude at peak  $m_B^*$ ,
- the color parameter  $c$  defined as the difference between the color in the  $B$  and  $V$  bands ( $B - V$ ) and the average color  $B - V$  of SNe Ia,
- the stretch parameter  $x_1$ ,
- the date of maximum luminosity in the B band.

SALT2 also allows to derive photometric redshifts of distant Type Ia supernovae as shown by *Palanque-Delabrouille et al.* in [49]. We mention this for the moment, and we will take advantage of this application later in Chapter § 8.

Summarizing, we are now able to perform distance measurements using type Ia SNe thanks to SALT2 that takes into account the correlations of an SNIa observed properties

---

<sup>1</sup>SALT stands for “Spectral Adaptive Lightcurve Template” and 2 represents the version described in [29].

with its distance modulus. Such measurements ultimately allow to constrain cosmological parameters thanks to the relation between luminosity distance and the energy-matter content of the Universe in Equation [2.32](#).

## Chapter 4

# SuperNova Legacy Survey (SNLS)

In Chapter § 3, we introduced SNe Ia as standard candles and a suitable tool for the study of the accelerated expansion of the Universe. We now need to detect, classify and extract information from these objects. For this we need to perform high-quality sky observations with a proper survey strategy.

The objective of this chapter is to present an overview of the SuperNova Legacy Survey (SNLS). First, the instrument used for acquiring the survey images is described in Section § 4.1. Then in Section § 4.2, we introduce SNLS as the survey designed for detecting hundreds of type Ia SNe using a specific observation strategy. Finally, we present the two different and independent analyses sets for processing SNLS data, standard and photometric, in Section § 4.3.

### 4.1 The instrument

The Canada-France-Hawaii observatory, shown in Figure 4.1, is located at the summit of Mauna Kea, Hawaii, at 4,200 *m*. This location provides excellent observing conditions for the 3.6 m telescope hosted in the observatory (CFHT).

The wide-field optical imaging facility at CFHT is called MegaPrime/MegaCam (Figure 4.2). The imager, MegaCam [52], was developed by CEA/Irfu and at its inauguration in



FIGURE 4.1 CFHT dome at the summit of Mauna Kea, Hawaii [50]



FIGURE 4.2 The MegaPrime instrument [51]

2003 was the world's largest CCD camera with 340 million pixels. The camera consists of a mosaic of 36 CCDs laid out in 4 rows with 9 columns each as can be seen in Figure 4.3. Electronics were designed in order to minimize the readout time by using two amplifiers for each CCD. With this camera a large portion of the sky, almost 1 square degree, can be imaged with a resolution of 0.187 arcsecond per pixel. The camera is cooled down to  $-120^{\circ}\text{C}$  to reduce thermal noise and obtain a good detection efficiency. For observations five broadband filters are available, denoted  $g_M$ ,  $r_M$ ,  $i_M$ ,  $z_M$  and  $u^*$ , which span a wavelength range of 300 to 1000 nm as can be seen in Figure 4.4. Other MegaCam technical specifications are shown in Table 4.1.

## 4.2 SNLS survey

The SNLS is an international collaboration with a primary goal of measuring the equation of state of Dark Energy with SNe Ia with an accuracy of 5% on  $\omega$ . The program was

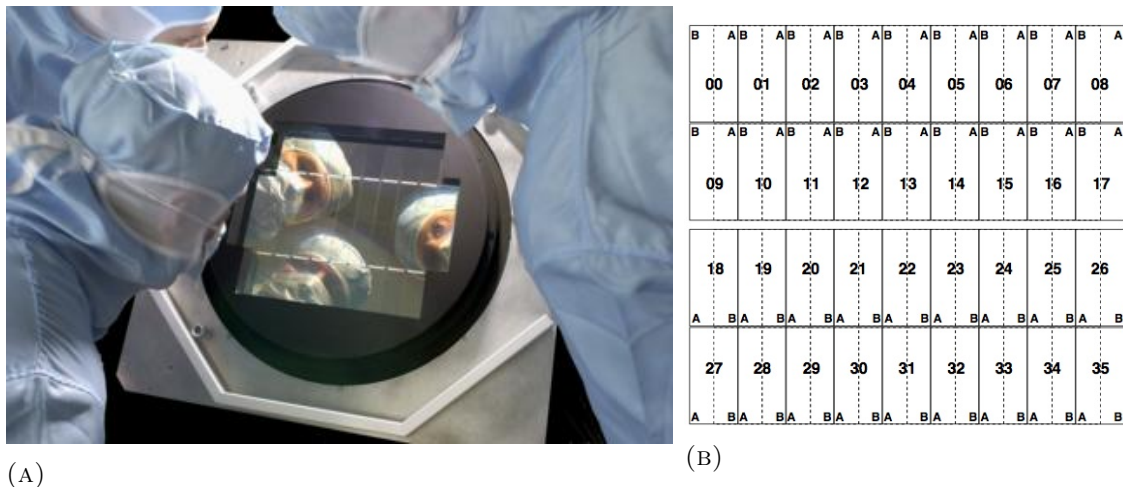


FIGURE 4.3 (A) Photo of the MegaCam imager [53] and (B) the corresponding numbering of the CCD mosaic. In the second image, letters A and B on each CCD correspond to the 2 amplifiers used during readout [54]. North is at the top, East to the left.

CCD number	36
CCD size	2,048 x 4,612 pixels
Pixel size	13.5 $\mu\text{m}$
Pixel angular size	0.187''
Field of view	0.96 deg x 0.94 deg
Readout time	$\approx$ 35 seconds
Readout noise	$\leq$ 5 electrons
Wavelength range	350-1000 nm

TABLE 4.1 MegaCam technical specifications [55].

designed for detecting hundreds of type Ia supernovae in a redshift range between 0.2 and 1. It is a second generation survey focused on more precise measurements than previous experiences. Other second generation surveys include ESSENCE [57] and SDSS-II Supernova Survey projects [58]. Advantages of SNLS include a very good time sampling and the use of four passbands while some surveys only used two.

SNLS is part of the CFHT Legacy Survey (CFHTLS) which included three different surveys: the “very wide”, the “wide” and the “deep”. As their names show, each survey was deeper and narrower than the one before. Over 2,300 hours of the telescope time were allotted to the CFHTLS during a 5 year period, from 2003 to 2008. This corresponds to about 450 nights of observations.

SNLS images come from the CFHTLS “deep” survey, which devoted around 202 nights observing four different patches of one-squared-degree in the sky. These fields D1,D2,D3 and D4 were located high in the galactic plane (Figure 4.5) to minimize Milky Way

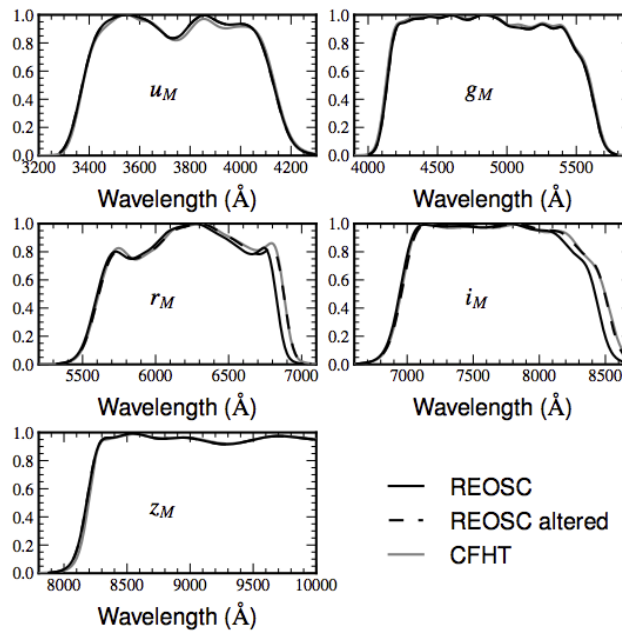


FIGURE 4.4 MegaCam filter transmission curves measured 15 cm from the center. REOSC measurements were done in 2002 by the manufacturer and CFHT measurements were done in 2006. A widening of the passbands is seen for the  $i$  and  $r$  bands [56].

extinction, to avoid extremely bright stars and to have an overlap with other surveys to provide complementary data, such as the SDSS survey. Each field is observable around 7 months in Hawaii and two fields are available simultaneously at any period of the year.

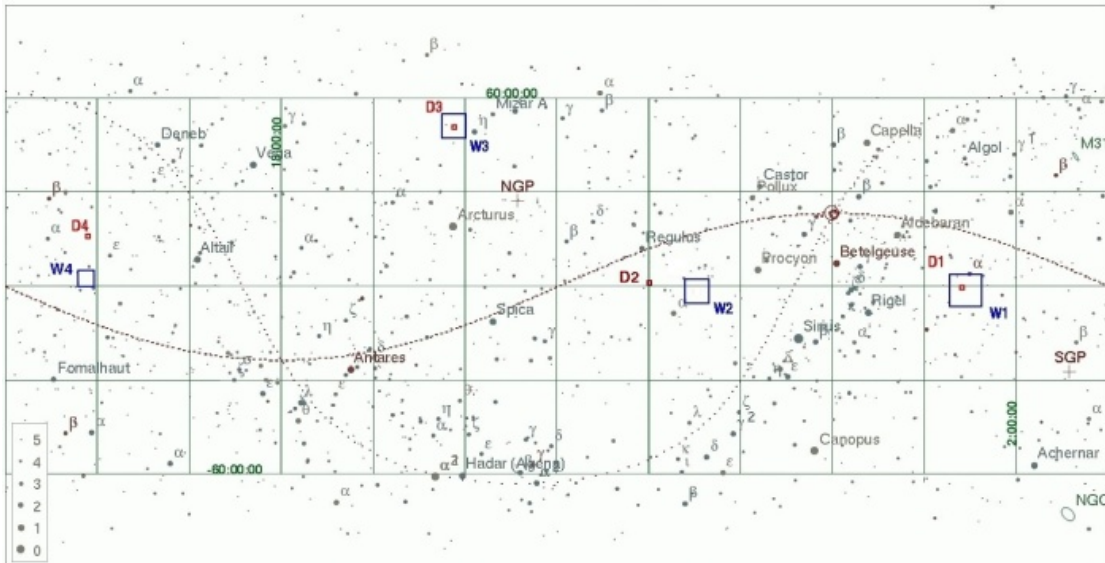


FIGURE 4.5 Full sky map with position of the deep and wide fields of CFHTLS. The ecliptic plane is denoted as the dense red dotted line, the Galactic plane is the other dotted line. NGP denotes the North Galactic Pole [59].

From all the filters available on CFHT, only four were used in the SNLS analysis since the  $u^*$  contribution is minimal to high redshift supernovae. The transmission bands of the  $i_M$ ,  $r_M$ ,  $g_M$  and  $z_M$  filters (hereafter i,r,g,z bands for simplicity) can be seen in Figure 4.6 where they are compared to SDSS filters.

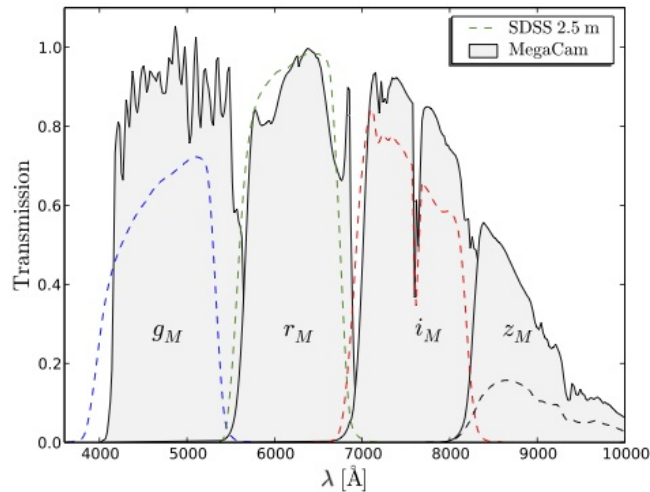


FIGURE 4.6 MegaCam effective passbands at the center of the focal plane in solid lines compared to SDSS 2.5 m effective passbands in dashed lines [60].

SNLS is a rolling search survey. In order to find transient events, SNLS targets the same field every 3-4 days throughout 5 to 7 consecutive lunations, where a lunation is the period of three weeks around a new moon. This allows to track simultaneous events as can be seen in Figure 4.7. Since two fields are simultaneously observable at any given time of the year, every other night the field is changed. For each night with SNLS observations, several exposures in the  $i_M$  and  $r_M$  bands are taken (typically 5 to 7). Plus several exposures in either  $g_M$  or  $z_M$  band.

All images from MegaCam destined for the SNLS analysis were pre-processed at the CFHT using the Elixir pipeline [62]. In this way, many instrumental defects can be disentangled from science data. The pre-processing includes:

- Flat-fielding: for all CCDs in the mosaic the same zero point is obtained by mapping non uniformities in the photometric response of each CCD.
- Subtraction of fringes for the i and z band images.
- Assigning flags for defective pixels and cosmic rays.

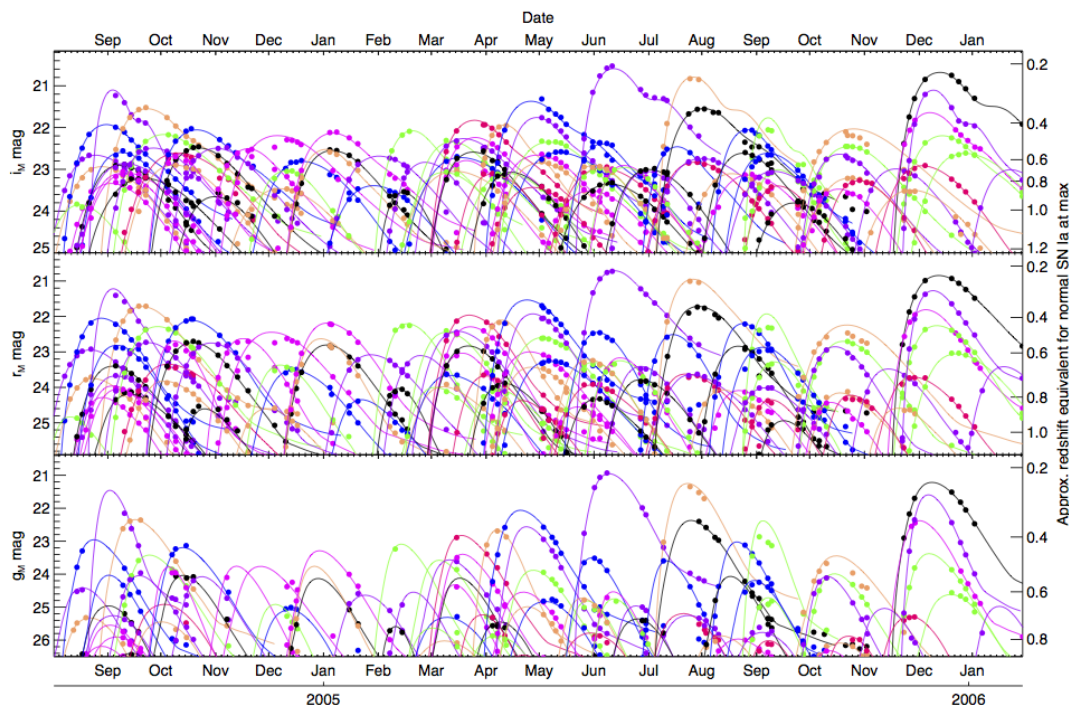


FIGURE 4.7 Light curves for SNLS SN Ia candidates. Three filters are shown here:  $i$  (top),  $r$  (middle), and  $g$  (bottom). Light curve fits (continuous lines) are drawn over observed data (points) [61].

### 4.3 The pipelines

Two independent analysis pipelines process SNLS data. The first, the standard processing relies on the spectroscopic follow-up of detected SN candidates [27]. My work has been centered in the second, the photometric processing that will be thoroughly discussed in the next Chapter § 5. The latter does not require spectroscopic resources and allows to obtain other types of measurements as core-collapse supernovae rate.

#### Standard analysis

This processing is based on real-time detection of transient events and then spectroscopic follow-up for both classifying and analyzing SNe Ia. Spectroscopy provides high precision measurements for studying the expansion of the universe but requires sizable time resources.

The processing goes as follows. First, transient events are detected. Pre-processed images from each night of observation are scanned for any sign of varying flux objects.



Two independent transient events detection pipelines were set up, one based on the Alard algorithm [63] and the other one based on a non-parametric approach. Candidates are found comparing an image with a stack of earlier images and searching for excess of flux compatible with a point source. The data processing showed that the overlap of candidates between the two pipelines was over 90%. More details can be found in *Astier et al.* [27].

Second, since spectroscopic resources are limited, a ranking of promising transient events is done. Detections that are matched to known periodical events such as Active Galactic Nuclei (AGNs) and variable stars are discarded. Photometric information from images is used to fit a light curve model to partial light curves. Information from this fit allows to better classify promising events but also to filter out events such as core-collapse supernovae which, as we have seen in Chapter § 3, have distinguishable light curves in most cases.

Third, candidates are sent to spectroscopic follow up for confirmation of SNe Ia type and redshift determination. Four 8-10 m diameter-telescopes participated in this step, the VLT in Chile, the Gemini Telescopes in Chile and Hawaii and the Keck Telescopes in Hawaii. These telescopes provide very precise redshifts, with accuracy of 0.1% to 0.01%. It must be highlighted that the time necessary for spectroscopic follow-up is well above that required at CFHT for photometry.

For the first 3 years of data from the SNLS (here forth SNLS3) the standard analysis found 252 spectroscopically confirmed SNe Ia with redshifts between 0.15 and 1.1 [64]. Combining this data with low- $z$  surveys, SDSS and high- $z$  as HST yielded the best current Hubble diagram to date which can be seen in Figure A.1. This high confidence result was due to the large statistics collected by SNLS and the high quality of the data selected to enter the Hubble Diagram.

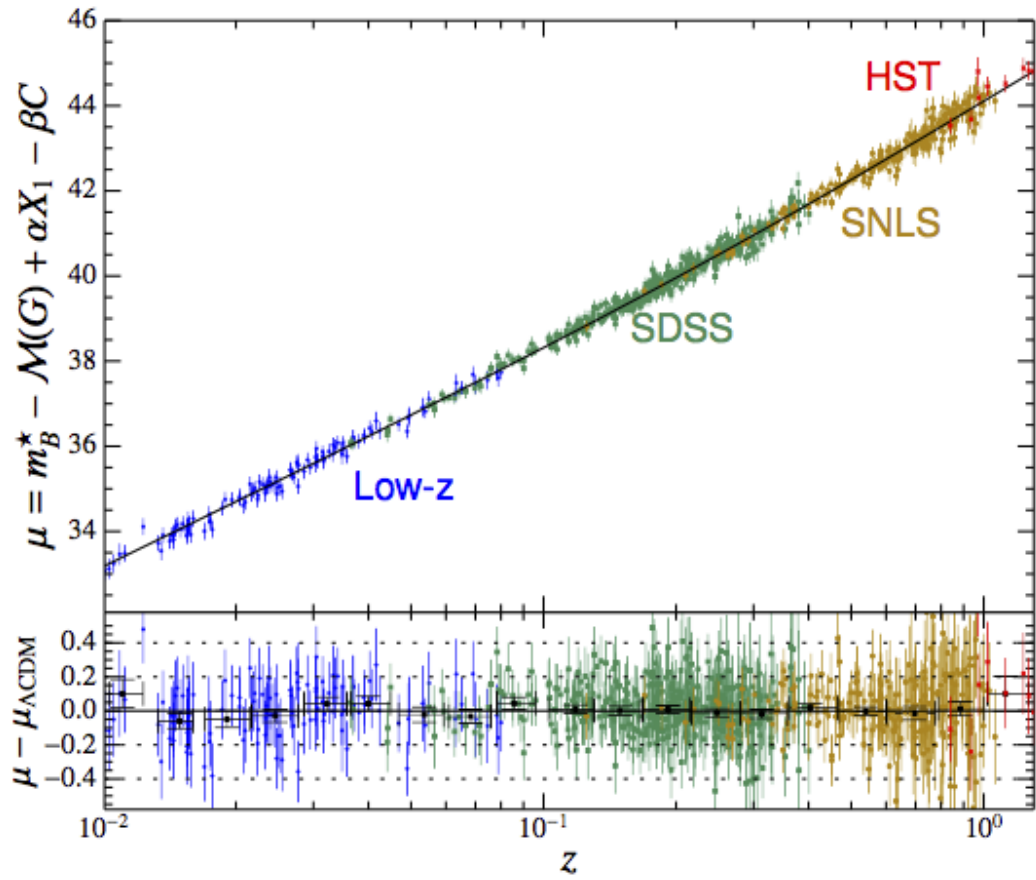


FIGURE 4.8 Hubble diagram using 472 SNeIa from SNLS3 and other low- and high-redshift surveys from JLA with SALT2 parametrization [65]. The black curve shows a  $\Lambda$ CDM model fitted to data.  $\mu$  is given by Equation 3.5.  $\Lambda$ CDM parameters from this analysis include  $\Omega_m = 0.303 \pm 0.012$ ,  $\Omega_\Lambda = -1.027 \pm 0.055$  and  $H_0 = 68.50 \pm 1.27$ .

## Chapter 5

# SNLS deferred photometric analysis

In this chapter I will introduce the deferred photometric analysis developed by the Saclay SNLS group on which this work is based. The aim of my PhD was to optimize this photometric pipeline set up for SNLS3 in view of SNLS5.

The photometric analysis detects and selects type Ia supernovae using only photometry. It is independent of the previously presented real time pipeline (Section § 4.3). Therefore, the image treatment, detection strategy and classification in this chapter are exclusive of the deferred photometric analysis.

Our pipeline has two very interesting characteristics: first, it is **based only on photometry** and second, it is **deferred**. The latter allows to optimize detection and selection strategies at all times. The former means that we are able to detect and select SNe Ia candidates without the use of any spectroscopic resource [66]. In the era of large future surveys, spectroscopic resources will be limited for candidate follow-up and classification which makes photometric pipelines particularly interesting to study. Other advantages of a photometric pipeline include the possibility of detecting other type of SNe such as core-collapse. For example, a study of the core-collapse SNe rate was published using the data from this deferred pipeline in *Bazin et al.* [35].

This chapter will describe the SNLS3 pipeline. It will focus on how to go from survey images to our sample of type Ia supernovae as done in *G. Bazin's PhD thesis* [67] and

*Bazin et al.* [66]. The starting point is the available data, pre-processed images from CFHT, which are processed as shown in Section § 5.1. SNe are transient events, hence a natural first step is to detect all objects with varying flux as explained in Section § 5.2. Since transient events include, but are not only SNe, in Section § 5.3 we present a selection of those transient events which are more likely to be SNe. Once the SN sample is assured, a final classification and selection must be done to obtain our desired object sample, type Ia SNe, explained in Section § 5.4.

To conclude, I will point out changes on this pipeline that are being implemented for the SNLS5 analysis in Section § 5.5. My work on the optimization of the photometric pipeline, presented in this Chapter, will be detailed in Chapters § 7 and § 8.

## 5.1 Image Processing

### 5.1.1 Astrometry and resampling

At this stage, survey images have already been pre-processed at CFHT as shown in Section § 4.2. This was the last step shared with other SNLS analysis pipelines. Now, for our pipeline, we need to align the pixels of our images into the same spatial frame. Basically we want to project our images into a common pixel grid. To this end, we first need to define the transformation to be applied and then, we must do the actual image alignment.

The transformation, so-called astrometric solution, is computed for the images of each SNLS field. This is done using TERAPIX calibration tool SCAMP [68] that computes the projection parameters from a catalogue of known objects, namely the USNO-B1.0 catalogue in our pipeline [69]. SCAMP computes these transformations taking into account possible rotations, translations and deformations between images.

Alignment resolutions are provided by SCAMP comparing coordinates of bright objects. First, an “absolute” resolution is computed between the positions of objects in the external catalogue and positions in the aligned images. Second, an “internal” resolution is found that shows the coordinate dispersion among all aligned images. It must be highlighted that a threshold on matching distances between reference coordinates and

retrieved image coordinates is imposed when computing resolutions. The default is  $2''$  in SCAMP.

For SNLS3 the “external” resolution given by SCAMP is slightly above  $0.4''$ . For any field, the internal resolution between images is found to be non homogeneous and field-dependent. Note that to obtain a better alignment resolution it is highly recommendable to use astrometric exposures from wide fields when available. Such is the case for fields D1 and D3 which are encompassed by the CFHTLS wide fields W1 and W3 respectively. When comparing “internal” resolutions, fields D2 and D4 obtained roughly  $\approx 0.2''$  while for D1 and D3 resolutions were  $\approx 0.1''$ .

The alignment of images, which includes resampling, is done using the TERAPIX tool SWarp [70]. SWarp takes the computed transformation by SCAMP and projects the images on the astrometric frame doing, if needed, pixel interpolation.

Once images are aligned, a cleaning procedure can be performed to tackle defects induced by this step. For example, dead pixel lines and borders of CCD usually have a pixel content of 0 ADU but when resampling this content is altered and pixels around these regions can acquire very high and extreme values. The cleaning uses a median filter to tackle empty zones and a Laplace filter for the borders of dead pixel lines, see *G. Bazin’s PhD thesis* for more details [67].

Now, all images are in the same grid of pixels and the search for transient events can start.

### 5.1.2 Subtractions

Supernovae are transient events and as such can be detected thanks to their varying flux. Our pipeline takes advantage of this by using subtractions to detect them. The idea is to compare “current” images to a “reference” image of the same sky area. In Figure 5.1 a SNIa explosion is clearly seen in a galaxy and the subtraction procedure is illustrated.

First, we need to produce reference images, for each field and filter. The same reference images were used throughout the processing and were constructed using the SWarp tool. Around 20 images with the best photometric quality, based on seeing and absorption,

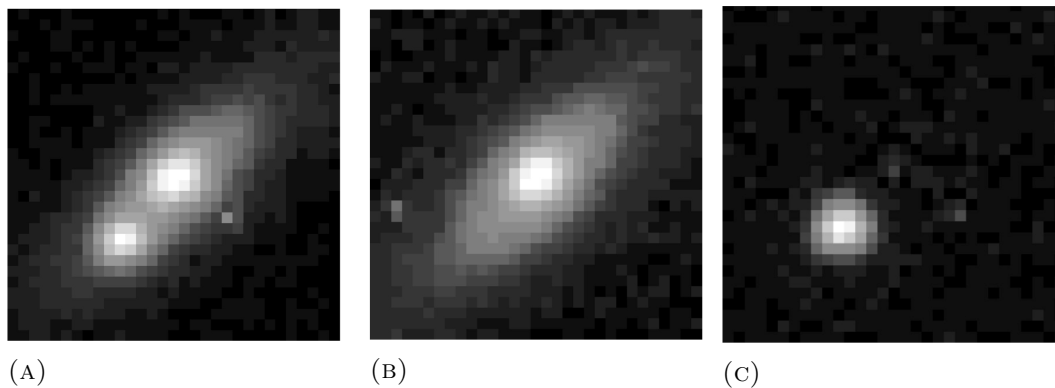


FIGURE 5.1 Example of using subtraction for detecting SNe Ia: (5.1a) Explosion of an SNIa (04D1dc) in a galaxy. The image of the galaxy before explosion (5.1b) is subtracted from the current image 5.1a to obtain an image of our transient object (5.1c). The area shown here represents a very small fraction of the current image.

were chosen from the first and second season depending on the field. This set of images were co-added to obtain a better signal-to-noise ratio but also to ensure complete field coverage. They were also cleaned for resampling defects. An example of a reference image can be seen in Figure 5.2.

Subtractions were done using the TRITON package [71] based on the algorithm by Alard and Lupton [72]. To compare a current image to a reference image, it is necessary to adapt the PSF of the reference image to the one of the current image. For each image a determination of the sky background and a convolution kernel was performed. The convolution kernel translates different PSFs: the shape adapts the seeing between images, while the norm of kernel allows to adapt the flux scale between the two images. The importance of adapting PSFs for subtractions is illustrated in Figure 5.3.

Both the sky background and the convolution kernel were computed independently on eight non-overlapping tiles for each CCD. The number of tiles was determined taking into account two criteria: the optimization of subtraction for spatial variations of the background and kernel which requires small tiles and the necessity of a sufficiently large number of bright events on a tile to compute the convolution kernel. These criteria constrain the size, and therefore the number, of the tiles.

Bright objects were selected in both reference and current image tiles by applying a detection threshold at  $2\sigma$  w.r.t. the sky background using the TERAPIX tool SExtractor [73]. This utility extracts coordinates for detected objects as well other information as flux and background estimations. The convolution kernel and sky background were then

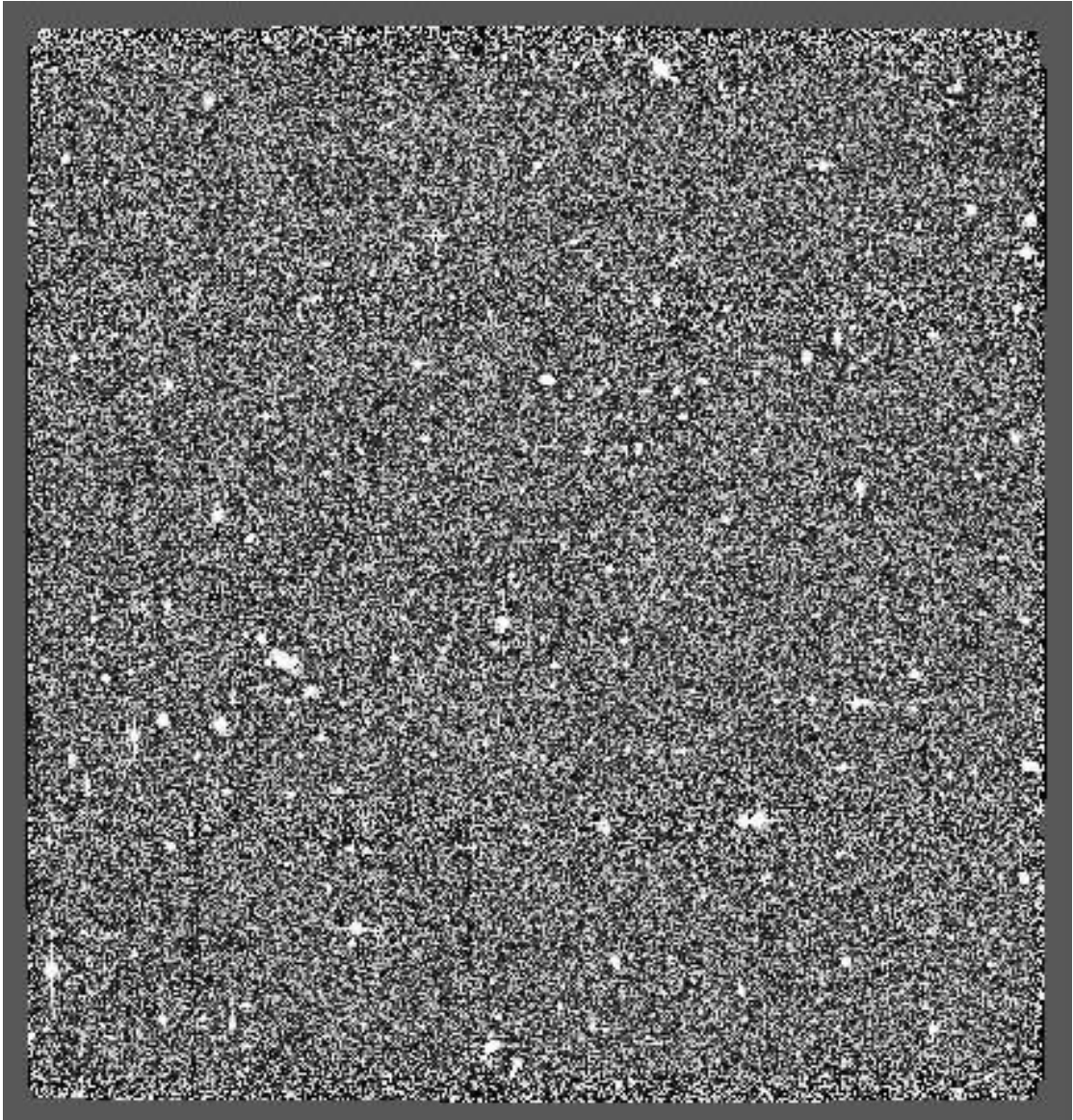


FIGURE 5.2 Reference image for field D1.

fitted in  $59 \times 59$  pixel regions centered on each bright object present in both the reference and current images. These objects were also required to be neither saturated nor too close to a tile boundary. Regions around objects that had more than 20% overlap were discarded. This procedure yielded around 100 object per tile at the average seeing of 0.7 arcseconds.

Once the sky background and the convolution kernel were determined, subtractions were performed for each tile and filter. The use of the convolution kernel ensures that the subtracted image flux scale is scaled to that of the reference. A subtraction is considered valid only if the integral of the convolution kernel is above a filter-dependent threshold. In the SNLS3 processing the rate of lost subtractions was only around a few percent.

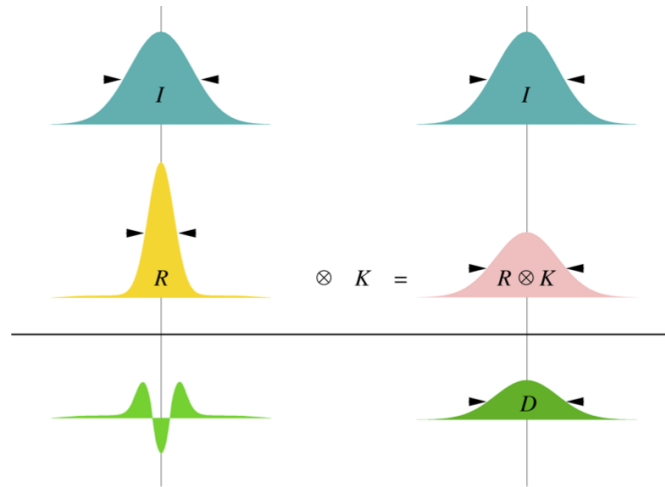


FIGURE 5.3 Image subtraction principle with a PSF adjustment. The reference image (R) is to be subtracted from the current image (I). However, if a pixel by pixel subtraction is done the resulting image is the one on the lower left side which is clearly not correct. If the reference is convolved with a kernel (K) that transforms the reference PSF to the current one, the subtraction is correct. [67]

As an approach for searching transient events, image subtraction is a quite common choice. Although the idea behind it is simple, this procedure is more complex than it appears due to variations on image quality and artifacts present in both reference and current images in spite of the previously mentioned cleaning. We note this for the moment and we will address it in Chapter § 7.

## 5.2 Detection of transient events

The next step is to detect transient events in subtracted images applying a signal-to-noise threshold. This is done using the TERAPIX tool SExtractor [73]. Distant SNe in SNLS have their maximum flux in the  $i_M$  band, then the detection of transient events is done in this filter. Since images are aligned, we assume that the coordinate of an object found in the  $i_M$  images will be the same in other filters.

We want to detect distant supernovae ( $m_i \approx 24$ ). If we apply a signal-to-noise ratio on subtracted images, the threshold should be set low in order to recover faint events (e.g.  $1.2\sigma$  [67]). However, requiring such a low threshold yields a large number of spurious detections. Moreover, doing a detection map for each of the roughly  $2 * 10^6$  subtracted images in SNLS3 requires handling a very large number of simultaneous maps at different times.



### 5.2.1 Luration stacks

In order to increase the signal-over-noise ratio and to reduce the number of spurious detections (since many of these occur at the same position in different images), a strategy based on stacking was developed by *Bazin* in his PhD [67].

The idea is to co-add  $N$  images that cover the same patch of the sky. The theoretical increase in signal-to-noise ratio is of order  $\sqrt{N}$ . However, since noise between subtracted images is correlated due to resampling and the use of the same reference image in all subtractions, this is only approximate.

Virtual CCDs represent common sky areas. These were defined by dividing each field reference image in 36, the number of real CCDs in the mosaic. It must be noted that, due to different pointings between exposures, the virtual CCDs are not aligned with the same real CCDs. From now on, we will only refer to virtual CCDs.

The stacking, or co-addition, was done for subtracted images that belong both to the same sky area and to the same observation period. This period was defined around a new moon, hence the name *luration stacks*. On average, 30 images entered each stack. These stacks provided an improvement on S/N ratio of approximately a factor 5, for SNLS3, according to *Bazin* [67].

### 5.2.2 Detection catalogues

Transient event detection catalogues can now be constructed for each luration stack. They are called *luration catalogues*. This is done using SExtractor. It must be noted that since we are dealing with subtracted image stacks values of flux and background are only indicative to signal and noise levels. For each *luration stack*, a detection required at least 4 pixels with a signal of more than  $2.5\sigma$  w.r.t. sky background. In order to prevent extracting one position for two different close-by objects deblending was also required.

Since the same reference images are kept throughout the processing, some of the supernovae have part of their signal included in the reference and when doing subtraction a negative flux variation may appear. Therefore, both negative and positive flux variations were detected.

	Lunation catalogues	Final catalogues
Number of detections	975,563	302,987

TABLE 5.1 Number of detections in lunation and final catalogues for SNLS3 data.

For SNLS3, in each field around 30 *lunation catalogues* were constructed. A large number of spurious detections were found in different lunation catalogues at close-by positions (2-3 pixels) (e.g. due to bright star residuals which can be recurrent in time). Therefore, the construction of a final CCD detection map was proposed to reduce the number of spurious detections to be further treated.

To do so, *final detection catalogues* were obtained by merging all *lunation catalogues* for each CCD and converting the result into an image where each detection was replaced by a Gaussian of height 1 (to give the same weight to all detections) and width 1 (motivated by the small distance between spurious detections). This image was processed with SExtractor selecting only pixels with a content above a value of 0.01 and deblending objects. In this way, any object detected on several lunations at the same position (within a pixel) gave only one detection, with a position averaged over all lunation stacks.

The result on the number of detections in the *final detection catalogues* compared to the one of all *lunation catalogues* can be seen in Table 5.1. There is a clear reduction of a factor three on the number of detected transient events. A sketch of the complete transient event detection procedure (as done for SNLS3) can be seen in Figure 5.4.



FIGURE 5.4 Detection of transient events as in SNLS3 for each virtual CCD. Boxes with continuous lines represent an image (*fits*) and boxes with dashed lines represent catalogues (*ascii*).

### 5.2.3 Detection efficiency for SNLS3

The detection efficiency of this detection procedure was studied in *Bazin et al.* [66]. The authors used Monte-Carlo generated artificial images produced by *Ripoche* with SNe Ia

added for the D1 field in the  $i_M$  filter [74].

The simulation was done in four stages which are summarized in the following:

- **Generating SNIa coordinates and redshifts:**  
Host galaxies were identified from deep image stacks of the CFHT-LS Deep Fields [75]. Using a two-dimensional gaussian for modeling the galaxies, SN positions within their hosts were randomly generated up to a distance of  $2\sigma$  from the host galaxy centers. This method provided compatible simulated and observed SN-host galaxy angular distance distributions. The redshift assigned to each SN was that of its host galaxy and was restricted to the range between 0.2 and 1.2. In the D1 field 216,000 SNe were generated over the 3 years of SNLS3.
- **Simulating SNIa light curves based on a cosmological model:**  
For each SN the  $i_M$  light curve was simulated according to the SN properties (redshift, color, stretch) using SALT2 (introduced in Section § 3.3). The chosen cosmological parameters were:  $\Omega_m \approx 0.25$ ,  $\Omega_\Lambda \approx 0.70$  and  $H_0 = 67.9$ .
- **Adding SNe to SNLS images:**  
Using the result from the previous step, the generated SN flux as deduced from the light curve at each observation date was added to the corresponding raw image on the assigned position in the host galaxy.
- **The MC images were then processed by the Saclay deferred pipeline as real images.**  
This includes detecting transient events in the subtracted MC image stacks.

Detection efficiency was defined as the fraction of recovered simulated supernovae at the end of the processing. The efficiency as a function of the generated SN peak magnitude in  $i_M$  is given in Figure 5.5. It is consistent with the one expected for a magnitude-limited survey such as SNLS. The efficiency is nearly magnitude independent with  $\epsilon_{max} \approx 0.97$  up to  $m_{0i} = 23.5$  in the SNLS magnitude system<sup>1</sup>, for the D1 field. Using an efficiency model, the efficiency was determined in the other survey fields, D2, D3 and D4 which had slightly different observing conditions.

<sup>1</sup>The SNLS magnitude system uses star BD +17 4708 as a flux standard.

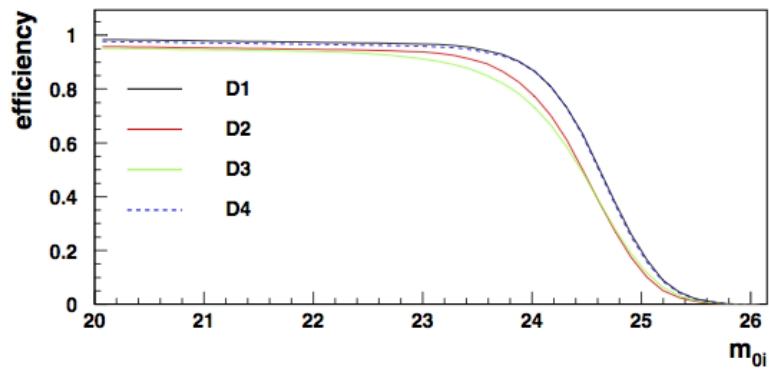


FIGURE 5.5 Detection efficiency for SNe Ia as a function of the generated  $i_M$  peak magnitude in all survey fields.

## 5.3 Selection of SN-like events

### 5.3.1 Light curve reconstruction

To select those transient events that were more SN-like, we took the *final detection catalogue* and constructed four-filter light curves for each candidate. Fluxes are a crucial part on this analysis since their measurements must be properly calibrated and filtered for residual effects.

#### Calibration

Light curves were constructed from individual subtracted images. The  $i_M$  position obtained in the *final detection catalogue* was imposed for all photometric measurements on all filters. At each detected position we applied TRITON differential photometry with PSF fitting.

The PSF was computed for each exposure and CCD. Bright events were detected on each resampled image using SExtractor. Non saturated stars were used to define one PSF per CCD by averaging star profiles computed in a seeing-dependent box.

In order to express all fluxes in the same common scale (the one of the reference image), fluxes were normalized by the integral of the convolution kernel used previously during subtractions.

The measured fluxes were calibrated using SNLS tertiary standards as defined in *Regnault et al.* [76]. This was done by measuring the normalized fluxes of the tertiary standards directly on current images with the same photometry algorithm and PSF as

for transient events. Using the magnitudes of the tertiary standards, the tertiary fluxes were converted into zero-points for each exposure and CCD. Zero-points that were above  $5\sigma$  with respect of the median value on each CCD were rejected.

### Filtering

To ensure the quality of our flux measurements some filtering must be done.

1. Detections due to saturated stars, satellite trails and residuals of bad pixels or cosmic rays were identified using SExtractor as bright objects with a  $3\sigma$  threshold with respect to sky background. Photometric points whose PSF overlapped the spatial extension of these objects were discarded.
2. Fluxes obtained under bad seeing conditions, above  $1.2''$ , were eliminated.
3. Fluxes measured in nights where all exposures were determined to be of bad quality, that is grade C from the TERAPIX visual image quality control, were also eliminated.
4. Points with flux errors either too low or too high, with respect to the expected flux uncertainty due to sky background were rejected.
5. Only fluxes within  $3\sigma$  from the night flux median value were kept.
6. Only nights with at least two exposures that met the above criteria were kept.

### Light curves

Mean light curves in each filter were defined by taking weighted averaged flux measurements within the same night. A common baseline was computed and subtracted to account for a possible contribution of variable objects in the reference image. For SNLS3 reconstructed fluxes were compared to the ones of spectroscopically identified SNe Ia. *Bazin et al.* [66] showed that the flux reconstruction had a resolution of a few percent and uncertainties below 2% on the absolute flux scale. They concluded that this is accurate enough to set up a photometric detection.

### 5.3.2 SN-like cuts

At this stage we want to use light curves to select all types of SNe while rejecting spurious objects. To define cuts, both spectroscopically identified events and synthetic SN Ia light curves were used.

The selection was done in four steps:

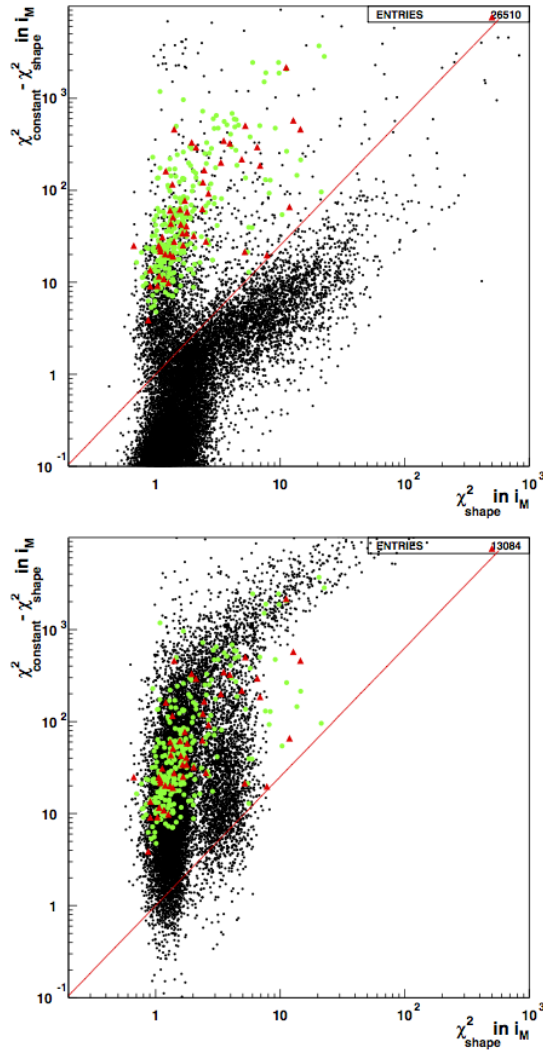


FIGURE 5.6 Difference in the reduced  $\chi^2$  between  $i_M$  light curve fits by a constant and by an SN-like shape. All cuts preceding that illustrated here have been applied. Black dots represent SNLS3 data (on top) and synthetic SN Ia events (bottom). In both plots, green squares are spectroscopic SNe Ia and red triangles are core-collapse SNe. Events above the curve are selected by the analysis. [66].

#### 1. Significant flux variations:

- For each light curve and filter significant flux variations were searched for. A variation was said to start with a photometric point of positive flux and significance above  $1\sigma$ . It ended when a point of negative flux had a significance above  $1\sigma$ , or when two successive points of positive fluxes had significances below  $1\sigma$  or at season ends.
- The most significant variations in the  $i_M$  and  $r_M$  light curves were required to contain at least three points each and to have their maximum flux dates within 50 days from one another.

2. **SN-like variation:** main variation (most significant variation) was required to have a shape compatible to the one of SN-like events. The main variation in each filter ( $k$ ) was fitted independently with the phenomenological shape:

$$f^k = A^k \frac{\exp -(t - t_0^k)/\tau_{fall}^k}{1 + \exp -(t - t_0^k)/\tau_{rise}^k} + c^k, \quad (5.1)$$

where  $A^k$  sets the normalization of the variable signal,  $\tau_{fall}^k$  ( $\tau_{rise}^k$ ) define the fall (rise) time of the variation,  $t_0^k$  is related to the date of maximum as  $t_{max}^k = t_0^k + \tau_{rise}^k \ln(\tau_{fall}^k/\tau_{rise}^k - 1)$  and  $c^k$  is constant. These parameters allow to make cuts on data.

- To reduce contamination by long-term variable objects or by random fluctuations, points away from the main variation were required to be compatible with a constant flux. This was done computing an off-variation  $\chi^2$  defined as:

$$\chi_{off}^2 = \frac{1}{\sum_k N_k} \sum_k \sum_{j \notin var} \left( \frac{F_k^j - c^k}{\sigma_j^k} \right)^2, \quad (5.2)$$

where  $F_j$  is the measured flux at point  $j$  in the light curve with an error  $\sigma_j$ . A cut was applied on this  $\chi_{off}^2$  as a function of the maximum flux observed in  $i_M$ .

- Consistency with the shape in Equation 5.1 was required in the  $i_M$  filter. We expect a significant signal on this filter for all redshift and types of SNe. This cut is illustrated in Figure 5.6.

3. **Star rejection:** Variable or bright stars still contaminate the sample. To avoid these events, known stars were rejected as well as events whose  $i_M$  flux varied

more than 70% between the day of maximum flux and the two closest epochs in the same season.

4. **Sampling requirements:** light curves were required to have sufficient temporal sampling. In the  $i_M$  and  $r_M$  filters at least one epoch before and one after maximum were required in an interval ranging from 30 days before the date of maximum flux in each filter up to 60 days after that date. Finally the dates of maximum,  $t_{max}^i$  and  $t_{max}^r$  from Equation 5.1 were required to be within 50 days of each other.

Cut	Events	Ia	CC	$\epsilon(\text{Ia})$
Detection in $i_M$	295 683	278	55	0.96
Variations in $i_M$ and $r_M$	193 104	277	55	0.94
Consistent $t_{max}$ in $i_M$ and $r_M$	52 284	276	53	0.93
Cut in $\chi_{off}^2$	26 510	276	52	0.93
$\chi_{constant}^2 - \chi_{shape}^2$ cut in $i_M$	2474	276	52	0.93
Star veto	2165	275	52	0.92
Sampling cuts in $i_M, r_M$	1674	253	43	0.79
Sampling cuts in $g_M, z_M$	1571	246	42	0.73
Consistent $t_{max}^{fit}$ in $i_M$ and $r_M$	1483	246	42	0.73

FIGURE 5.7 SN-like selection cuts applied on SNLS3. The first column shows the number of events that pass the cut. The second and third columns correspond to spectroscopically confirmed type Ia and core-collapse SNe respectively. The last column indicates the efficiency of the cuts for bright SNe Ia, derived from synthetic SN Ia light curves at low magnitudes ( $m_{0i} < 23$ ) [66].

The effect of the SN selection cuts on SNLS3 data is shown in Figure 5.7. The reduction by two orders of magnitude of the events, with respect to the small decrease in the spectroscopically identified samples, points towards the presence of a large number of spurious detections. This is obvious in the first three cuts shown in Figure 5.7 where flux variations are required to be physical. To avoid reconstructing light curves of such a large number of events, improvements on the transient event detection must be done and this will be addressed on Chapter § 7.

## 5.4 SN classification: type Ia selection

Our goal is to obtain a sample of type Ia supernovae. At this point we have a sample that contains both SNe Ia and core-collapse (CC) SNe and we must disentangle them. To do so, four-band light curves are fitted simultaneously with SALT2 (Spectral Adaptive



Lightcurve Template) [29] and events are classified using the variables obtained in this fit.

In this Section we will show how to select type Ia SNe from the SN-like sample. However, a selection of core-collapse SNe is also possible which was used to publish a core-collapse rate in *Bazin et al.* [35].

#### 5.4.1 SALT2 and photometric host-galaxy redshift

In order to fit light curves with SALT2 a redshift for each event must be provided. In the SNLS3 analysis this redshift was assigned using **host-galaxy photometric redshifts**.

The selected SN-like events were matched with galaxies in the catalogue published by *Ilbert* [75]. This catalogue provides photometric redshifts for more than 520,000 galaxies with an AB magnitude brighter than 25 in  $i_M$ . The match was considered valid if the nearest galaxy to the event was found within a distance of  $5r_{gal}$ . The effective galaxy radius  $r_{gal}$  was defined as the half-width of the galaxy in the direction of the event and was derived from SExtractor applied on our reference images. Faint galaxies that could not have their size measured were assigned a size of 1 pixel and were considered only if no bright galaxy was found in the vicinity of the event. In SNLS3, when comparing host-galaxy redshifts to spectroscopic ones, some host-galaxy redshifts departed by more than  $5\sigma$  from the spectroscopic ones, that is  $\Delta z/(1+z) > 0.15$ . This can be attributed either to incorrect SN-galaxy matching or to incorrect galaxy photometric redshifts. According to *Ilbert et al.* 3.7% of galaxies with magnitudes below 24 in the  $i_M^{AB}$  are expected to have  $\Delta z/(1+z) > 0.15$ .

Redshift assignments are of great importance since an incorrect assignment can generate significant systematic biases in the classification and selection of SNe Ia. Incorrect assignments mostly occur at redshifts  $z > 0.6$  as shown in *Bazin et al.* [66]. The redshift assignment efficiency for SNLS3 was found to be 83%.

Each light curve was then given to SALT2 with its assigned redshift. Assuming that the event is a SNIa, SALT2 derives the day of maximum light  $t_{max}^B$ , the color  $C$ , the stretch-related parameter  $X_1$  and the rest frame B-band peak magnitude  $m_B^*$ . This fit converged for 98% of SNLS3 SN-like candidates with an assigned host redshift.

### 5.4.2 Light curve sampling requirements

To ensure that the parameters obtained with SALT2 were meaningful, some quality cuts were performed. To do so, the event rest-frame was defined as:

$$\tau = \frac{t_{obs} - t_{max}^B}{1 + z_{gal}} \quad (5.3)$$

and this variable can be used to make meaningful cuts on sampling in the following way:

- To obtain a reasonable estimate of  $t_{max}$ , and thus of the peak magnitude, at least one measurement was required in the range  $-10 < \tau < +5$  days.
- For a reasonable shape evaluation at least one measurement was required in the range  $+5 < \tau < +20$  days.
- For Ia/CC discrimination (when using color diagrams for establishing cuts as we will do in the following) at least one color among  $(g - i)$ ,  $(r - z)$  and  $(i - z)$  was required to have at least one measurement in each band in the range  $-10 < \tau < +35$ .

For SNLS3, these cuts selected 1152 events from which 203 were spectroscopic SNe Ia and 35 spectroscopic core collapse SNe [66].

### 5.4.3 Using SALT2 parameters for SNIa selection

We have now the parameters from the SALT2 fit on the four-band light curves. These parameters include the magnitude at the date of B-band maximum light for each of the bands, the  $\chi^2$  of the fit in each band used in the fit, the color  $C$  and the stretch-related parameter  $X_1$ . A set of complementary cuts may now be defined to disentangle type Ia from remaining core collapse events. These cuts were set up using the SNLS3 spectroscopic samples and the SNIa light curve simulation as a guidance.

#### $\chi^2$ requirements

Poorly fitted light curves can be rejected using reduced  $\chi^2$  cuts. The following criteria were required to ensure proper fitting:

- For the  $g$  filter  $\chi^2 < 10$ .

- For  $r, i, z$  filters  $\chi^2 < 8$ .
- For the overall fit of all bands  $\chi^2 < 6$ .

For SNLS3 these cuts discarded 9 spectroscopic core-collapse events, all of SN II type. However, 7 spectroscopic SNe Ia were also discarded. Two of these had an incorrect host-galaxy redshift assignment while the others had noisy light curves.

Depending on their redshift, SNIa measurements have significant signal in all or a sub-sample of color bands. SALT2 takes this into account by fitting only the relevant bands:

- $g_M$  is used only when  $z < 0.68$ .
- $z_M$  is used only for  $z > 0.26$ .

Outside these redshifts, type Ia SNe are expected to have no significant signal in the corresponding bands. This information can be useful to reject other types of SNe when comparing the flux in the unfitted bands to that corresponding to the SALT2 fit. Since selection cuts to be applied will be looser in cases where not all bands are fitted, we compensated by requiring tight cuts when using unfitted bands. A  $\chi^2$  is computed for each unfitted band and it was required to be less than 6 in  $g_M$  and less than 3 in  $z_M$ .

For SNLS3, these cuts eliminated 4 spectroscopic CC and 10 SNe Ia. All eliminated SNe Ia had an incorrect host-galaxy redshift assignment.

### Stretch and color

A constraint between the fitted  $X_1$  and color  $C$  parameters was implemented as illustrated in Figure 5.8a. It was required:

$$\left(\frac{X_1 - 0.2}{1.4}\right)^2 + \left(\frac{C}{0.35}\right)^2 < 1 \quad (5.4)$$

The upper bound in  $X_1$  discards long duration events among which spectroscopic SNe II. This cut only removed a negligible fraction (1%) of synthetic SNe Ia. Core collapse had in average a color of  $\langle C \rangle = 0.3$ , then the color cut  $C < 0.35$  rejected nine spectroscopic CC. The color cut is helpful against SNIa events with incorrect redshift assignment. This is illustrated in Figure 5.8b for those SNLS3 events that had both host-galaxy and

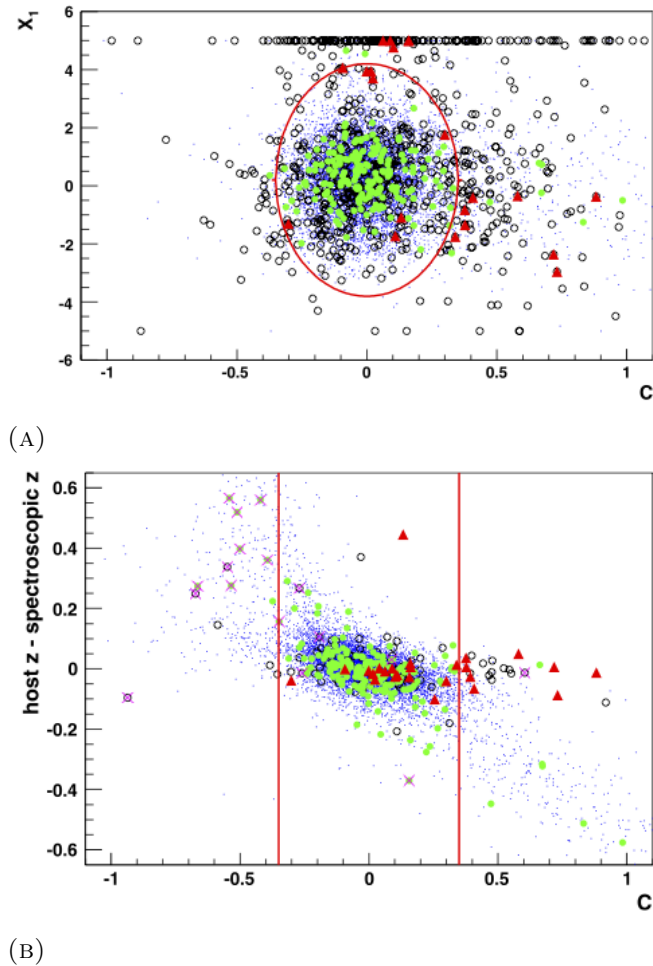


FIGURE 5.8 SALT2 parameters plot for color and stretch.  $X_1$  vs.  $C$  is shown in 5.8a and illustrate the way the  $X_1$ - $C$  cut was chosen (shown as a red ellipse). As an example of the effect of the color cut, the difference between redshift assignment as a function of  $C$  is seen in 5.8b with red lines showing the extreme value of color cuts. In both plots synthetic SNe Ia (blue dots) and data (other symbols) are shown after sampling and  $\chi^2$  constraints. Spectroscopically identified SNe Ia are represented with green circles and spectroscopically identified core collapse events with red triangles. For Figure 5.8b pink crosses show events rejected by  $\chi^2$  cuts in unfitted bands [66].

spectroscopic redshifts.

### Color-magnitude diagrams

The last criteria for rejecting core collapse events were based on color-magnitude diagrams. These can be seen in Figure A.2. Events entered these diagrams only when the filters of interest were used in the fit.

- Figure A.2a contributed only in the range  $0.16 < z < 0.68$ .
- Figure A.2b contributed only in the range  $0.26 < z < 1.15$ .

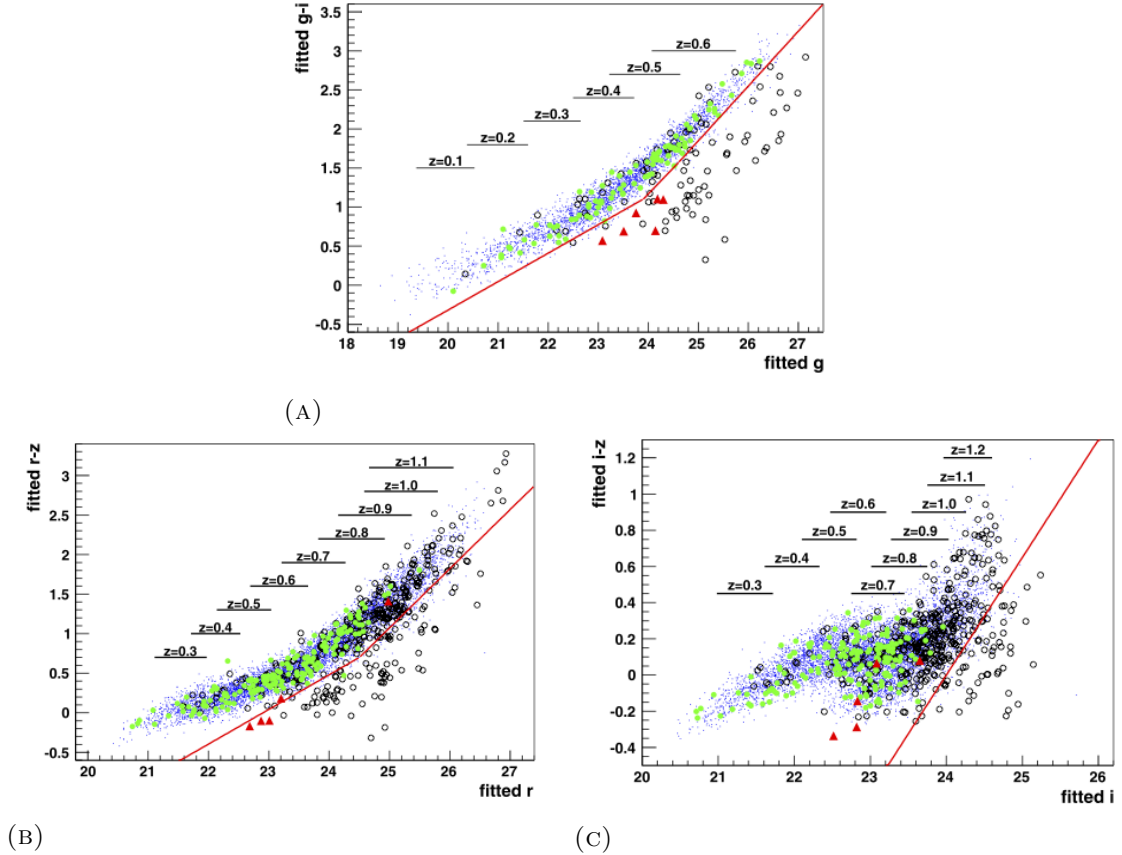


FIGURE 5.9 Color-magnitude diagrams using SALT2 fitted magnitudes for events that passed all cuts up to color and stretch cuts. Spectroscopically identified SNeIa are represented with green circles, core collapse events with red triangles and synthetic SNe Ia as blue dots. Cuts are shown using red lines. Magnitude ranges ( $1\sigma$ ) are indicated for 0.1 redshift bins centered in the given value (vertical location of segments is arbitrary). [66].

- Figure A.2c contributed only in the range  $z > 0.26$ .

Since core collapse events have more  $g$  than SNe Ia, the  $(g - i)$  vs.  $g$  diagram proved helpful in eliminating these.

It must be noted that all cuts were necessary to disentangle type Ia from core collapse SNe. Results for SNLS3 data are presented in Figure 5.10. Using the MC, the core-collapse contamination in the photometric sample was found to be of 4%. The average SNIa efficiency in the selection was 67% using well sampled light curves of bright SNe Ia events as given by *Bazin et al.* [66]. To emphasize the effect of redshift assignment over the average efficiency, if all events were assigned a host galaxy redshift the global efficiency for bright events would be 80%.

Cut	Events	Ia	CC	$\epsilon(\text{Ia})$
SN selection	1483	246	42	0.97
Host redshift available	1233	208	38	0.80
Light curve sampling	1152	203	35	0.75
$\chi^2$	951	186	22	0.68
$X_1$ and colour	596	176	7	0.67
$g - i$ vs. $g$	539	176	1*	0.67
$r - z$ vs. $r$	490	175	1*	0.67
$i - z$ vs. $z$	485	175	1*	0.67

FIGURE 5.10 Effect of the SNIa selection cuts on the data (second column), spectroscopic SNeIa (third column) and core-collapse SNe (fourth column) and efficiency for synthetic bright SNIa ( $m_{0i} < 23$ ) with well sampled light curves [66]. The asterix indicates an event that was later revised and found to have the spectrum of an unrelated object extracted which was identified as a core-collapse SN. After reprocessing it was shown to be a SNIa.

## 5.5 5-year analysis changes

The pipeline as presented in this Chapter was the one used for processing SNLS3. In the view of SNLS5, some changes have been already implemented by *Vanina Ruhlmann-Kleider* at various stages in the pipeline before detection. Changes after detection are under work.

First, during the last SNLS year the  $i_M$  filter was damaged and a new filter commissioned ( $i_{M2}$ ). During the time when no  $i_M$  filter was at CFHT, our detection pipeline used the  $r_M$  filter in the luration stacks to ensure continuity. Filter changes required computing reference images for the new  $i_{M2}$  filter in order to perform subtractions.

Second, the astrometric solution was revised. A two step-procedure was implemented to improve the alignment resolution. The first step consisted on creating an internal reference catalogue for a subset of  $i_M$ ,  $r_M$  and  $i_{M2}$  images of the whole survey and the astrometric exposures. This catalogue was calibrated with respect the USNO-B1.0 using a matching radius of  $2''$  as in the SNLS3 processing. The second step was to compute an astrometric solution in each filter using the internal catalogue, science exposures of years 4 and 5 and astrometric exposures. The matching radius for this second step was set to  $0.5''$ . The two step procedure improves internal resolution being at most  $0.04''$ . Also, within a field, filter's resolutions have now a good uniformity.

Third, star photometry and zero points were recomputed after the more precise calibration presented in *Betoule et al.* [77]. Calibration and grid uniformity corrections were

applied. Non-linearities in calibrated fluxes were found to be below 1.2 mmag on the whole range of star fluxes, to be compared to 3 mmag in the SNLS3 processing.

Further changes are described in the next chapters and are based on my own work.

## 5.6 Summary

In this chapter I presented an overview of the SNLS deferred photometric pipeline for selecting SNe Ia as applied to SNLS3 data by *Bazin et al.* [66].

The advantages of this pipeline are based on the possibility of selecting SNe Ia using no spectroscopy. This allows a larger sample of SNe Ia to be selected over a larger redshift span (Figure 5.11). However, spectra remains essential to achieve high level accuracy for cosmological studies and as such this pipeline does not replaces the real-time one.

Another advantage of the SNLS deferred photometric pipeline is that it includes a selection of SN-like events. This allows the study of non type Ia SNe. For example, the rate of core-collapse was measured using this sample in *Bazin et al.* [35].

However, there is still room for improving this pipeline which I addressed during my PhD.

First, SNLS3 processing showed that transient event detections were dominated by spurious objects, mostly from imperfectly subtracted objects such as bright stars, resampling defects and masks. Processing light curves of such a large number of detections knowing that 80% of those will be rejected by the early steps of the scientific analysis (see Figure 5.7) and do not contribute to science results represents a waste of time. In Chapter § 7 I will present my work to improve the detection of transient events using signal processing tools briefly introduced in Chapter § 6.

Second, I worked on a new classification of SNe Ia. The SNLS3 approach depends on the matching of an event to a host galaxy and its photometric redshift, which had an assignment efficiency of 83%. An alternative is to use the algorithm presented in *Palanque-Delabrouille et al.* [49] which determines SNIa photometric redshifts directly from photometry. This is described in Chapter § 8 where I take advantage of machine learning techniques.

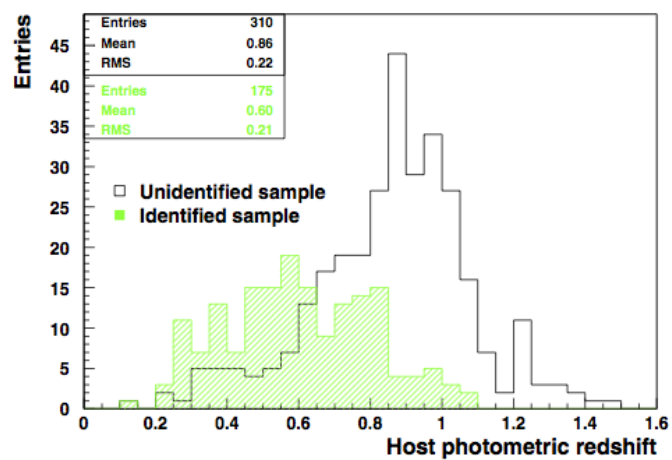


FIGURE 5.11 Distribution of host-galaxy photometric redshifts for photometrically selected SNe Ia in the SNLS3 data. Events with (without) spectroscopic identification are shown in green (black) [66].



## Chapter 6

# A brief introduction to signal processing and sparsity

Using signal processing tools, I found a way to reduce the number of spurious objects detected as transient events by cleaning the subtracted image stacks. In this Chapter, I will briefly introduce some of the signal processing concepts and the Morphological Component Analysis (MCA) algorithm. This is a reference chapter while the application to SNLS data is presented in Chapter § 7.

A fundamental problem in signal and image processing is to be able to decompose an image into its building blocks or superposed contributions from different sources. This problem can be seen as a linear *inverse problem*: to recover a signal from measured data that contains both the interesting signal, other signals and noise.

For my application I used the MCA algorithm by *Starck et al.* [78]. This framework assumes that an observed image can be described as the sum of several components, each exhibiting a distinct morphology. To actually perform this separation, MCA relies on the theory of sparse representation of signals.

This Chapter is heavily based on the book *Sparse Image and Signal Processing* by *J.-L. Stark, F. Murtagh and J. Fadili* [79] and *A wavelet tour of signal processing* by *Stephane Mallat* [80] together with some supporting publications.

Basic signal-processing concepts are defined in Section § 6.1. Then, I introduce the framework that allows treating images with different morphologies in Section § 6.2. The

latter is the basis of the MCA algorithm presented in Section § 6.3. The MCA algorithm relies on a choice of dictionaries to study different morphologies. The construction of dictionaries is introduced in Section § 6.4.

## 6.1 Basic concepts: sparsity, atom, dictionary and scales

Any signal,  $x$ , can be represented in different finite-dimensional spaces or domains (e.g. time domain, Fourier domain, wavelet domain) as a vector

$$x = [x[1], x[2], \dots, x[N]] \quad (6.1)$$

where signal coefficients exhibit different properties depending on the chosen domain.

One property of particular interest is the so called **sparsity** of the coefficients, i.e. the property that only a small number of coefficients are non zero. If a signal is not sparse, it may be *sparsified* using an appropriate domain transformation. As a general rule, the coefficients of the signal will be sparse when the basis functions of the domain (so called *atoms*) are very similar to the signal itself.

Formally, an **atom** is an elementary signal-representing template. Example of atoms are sinusoids, gaussians and diracs. A collection of ordered atoms of a given domain is called a **dictionary**. A dictionary  $\Phi = [\phi_1, \dots, \phi_T]$  is a  $N \times T$  matrix whose columns are the atoms,  $\phi$ , and  $T$  is the number of atoms. A dictionary is ordered according to a variable or index which is dictionary dependent, for example: frequency for the Fourier dictionary and position for the Dirac dictionary.

In particular, the Fourier dictionary is composed by atoms that are sinusoids. This dictionary is useful for the study of stationary signals or stationary periodic functions. This type of signals have statistical properties that do not change over time, the functions repeat themselves once per period, without modification. For example, a sine signal can be sparsely represented using this dictionary. For this signal, the choice of the Fourier dictionary is optimal from a sparsity standpoint since all information is contained in a single coefficient. An illustration can be found in Figure 6.1.

The Fourier transform studies the signal as a whole, it is a global analysis. However, we are interested in studying signals that are localized and have different sizes. For this we

can choose dictionaries that allow local analysis. Instead of ordering those dictionary through frequencies, as in the case of the Fourier dictionary, it is useful to organize it through **scales**. A scale represents the typical size of the atom. We will come back to this in Section 6.4.

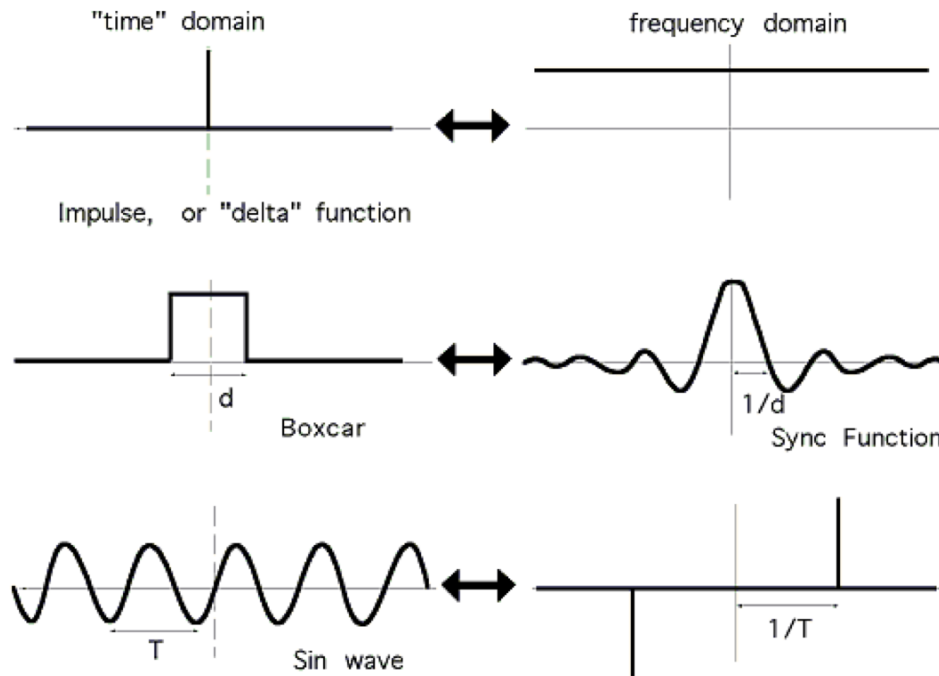


FIGURE 6.1 Example of different signals represented in the time domain (left) and in the Fourier domain (right).

In general, let us denote  $\alpha$  the coefficients of a signal  $x$  in a *dictionary*  $\Phi$ :

$$x = \Phi\alpha = \sum_i \phi_i \alpha_i, \quad (6.2)$$

If  $x$  is sparse in dictionary  $\Phi$  then only a small number of coefficients in  $\alpha$  are non zero.

Given a dictionary, two linear operations must be distinguished:

- Analysis: Associates each signal  $x$  with a vector of coefficients  $\alpha$  on that dictionary. Mathematically this means  $\alpha = \Phi^*x$ .
- Synthesis: Take the coefficients and reconstruct the signal  $x$  by superimposing the atoms of the dictionary  $x = \Phi\alpha$ .

These are different operations which not necessarily have a unique result.

In our case our *signal* of interest are images. Studied images are usually not strictly sparse in a representation as in the previous example. However, they may be *compressible* or *weakly sparse* which means one can neglect most coefficients while remaining with only a small fraction of them without much loss.

This sparsity property is extremely desirable as it can be used as a very powerful prior in the regularization of a wide range of inverse problems. Some applications to astronomy and astrophysics include denoising [81], deconvolution [82], blind source separation for CMB analysis [83], weak gravitational lensing [84].

## 6.2 Morphological diversity

Usually, images are composed by structures with different morphologies. Then, a single dictionary does not represent well the objects in it. However, we can assume that the given image is a linear mixture of several source signals of a more coherent origin. A framework of data modeling that allows to have a sparse representation while using different dictionaries to decompose an image is called **morphological diversity**. There, the image can be modeled as the sum of  $K$  morphologically different components as:

$$X = \sum_{k=1}^K x_k, \quad (6.3)$$

where each  $x_k$  is called *morphological component* and is assumed to be sparse in a given dictionary  $\Phi_k$ .

Given image  $X$  defined in (6.3) as the sum of  $K$  different morphological components, let us introduce  $K$  different dictionaries  $\Phi_k$ , each adapted only to the particular morphology of component  $x_k$  i.e. such that the  $\alpha_k$  coefficients,  $\{\alpha_{ki}\}$ , of  $x_k$  in  $\Phi_k$  are sparse but not the coefficients of  $x_l$  for  $l \neq k$ . The latter means that different dictionaries do not represent the others behaviors or are very inefficient at doing so. Such dictionaries are also called incoherent.

The idea is to perform the separation in the different morphological components through finding an optimal set of coefficients  $\alpha_k$  maximizing the sparsity of the decomposition of each component in the corresponding dictionary.

### 6.3 Morphological Component Analysis

The Morphological Component Analysis (MCA) algorithm has been proposed by *Starck et al.* [85] and uses the morphological diversity of the different features in data. Observed images,  $Y$ , are assumed to be a combination of signals (morphologically different components as in Equation 6.3),  $X$ , plus some noise,  $N$ :

$$Y = X + N, \quad (6.4)$$

MCA is a practical way to perform the decomposition of a signal in different overcomplete dictionaries. Overcomplete dictionaries contain more atoms than their dimension ( $T > N$ ), thus are redundant, and have been proven to lead to more significant representations allowing us to better extract information from images. However, the signal decomposition in these dictionaries is not unique. Therefore, finding the proper coefficients for a sparse representation is a non-trivial problem which is handled directly by the algorithm.

When applying MCA a major role is played by the choice of dictionaries for the decomposition. They must be mutually incoherent dictionaries that are well suited to represent the images' morphologies. Both dictionaries and associated scales must be chosen before applying the MCA algorithm.

The MCA algorithm performs the signal decomposition as the solution of an  $\ell_1$  minimization problem, where  $\ell_1$  denotes the  $\ell_1$ -norm. The  $\ell_1$ -norm promotes the sparsity of the decomposition of each component [86].

The decomposition algorithm solves iteratively the following optimization problem (analysis):

$$\min_{\mathbf{x}_1, \dots, \mathbf{x}_K} \sum_{k=1}^K \|\Phi^* \mathbf{x}_k\|_1 \quad \text{such that} \quad \left\| Y - \sum_{k=1}^K \mathbf{x}_k \right\|_2 \leq \sigma, \quad (6.5)$$

where  $\Phi^* \mathbf{x}_k = \alpha_k$  and  $\sigma$  is the standard deviation of the noise contaminating the data, assumed to be stationary and Gaussian distributed. At convergence, each morphological component is obtained as:

$$x_k = \Phi_k \alpha_k \quad (6.6)$$

This component reconstruction (synthesis) can be restricted to a sub-sample of  $\{\alpha_{ki}\}$ , for example to some size scales in a given dictionary.

## 6.4 Dictionaries

In this section I will briefly present the idea behind dictionaries, how they can be constructed and some applications. Although a very interesting topic, the mathematical rigor required to treat them in depth is beyond the scope of my work.

A dictionary is a family of atoms spanning the space in which lives the signal. A dictionary is equivalent to constructing a basis in a domain such as the Fourier one (dictionary from sinusoids or the basis of Fourier transformations).

Here, I will sketch how to go from a dictionary composed by one-dimensional global transformations to dictionaries which are localized, contain scale information and can analyze 2-dimensional images. The latter are the ones used in the MCA algorithm to treat our subtracted image stacks.

### 6.4.1 From Fourier to wavelet

I already introduced the **Fourier** dictionary. It is well known that for a given function  $f(t)$  its transform into Fourier space is:

$$\hat{f}(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt} dt \quad (6.7)$$

where the basis of the Fourier dictionary is made by  $e^{iwt}$  which are the so-called “sinusoidal atoms”.

The Fourier analysis is based on global information while we want to study local patterns. In the Fourier case, we have functions defined as a function of time which are transformed and analyzed in the frequency domain. To be able to study both frequency and time we need to “localize” the transformation. For this we can use a sliding Gaussian window.

We introduce the **short-time windowed Fourier transform** (STFT) around time point  $\tau$  as:

$$\hat{f}(\tau, w) = \int_{-\infty}^{\infty} f(t)g(t - \tau)e^{-iwt} dt \quad (6.8)$$

where  $g(t - \tau)$  is the window function for a given  $\tau$ . We can see that the STFT of a signal is the inner product of the signal with an element of the set of basis functions  $k_{\tau,w}(t) = g(t - \tau)e^{iwt}$  (windowed sinusoids), which vary over frequency  $w$  and time  $\tau$ .

An equivalent transformation to the Fourier one that will be used in the following is the **wavelet** transformation. The wavelet atoms can be seen as a new basis defined with *window size*  $a$ , inversely proportional to  $w$ , and a *positional parameter*  $b$  equivalent to the previous  $\tau$ . The wavelet atoms  $\psi_{a,b}$  are defined as:

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right) \quad (6.9)$$

where  $\psi^*$  is the complex conjugate of  $\psi$ . In general, a dictionary of wavelets consists of the translations and dilations of a single real or complex valued function  $\psi$ , called the *analyzing wavelet* (also known as the *basic wavelet* or *mother wavelet*). Translation is the shift  $\psi(t)$  to  $\psi(t - b)$ . Dilation of a function means a contraction or spread out of the function, for example  $\frac{1}{\sqrt{a}}\psi(\frac{t}{a})$  is a dilation of  $\psi$ .

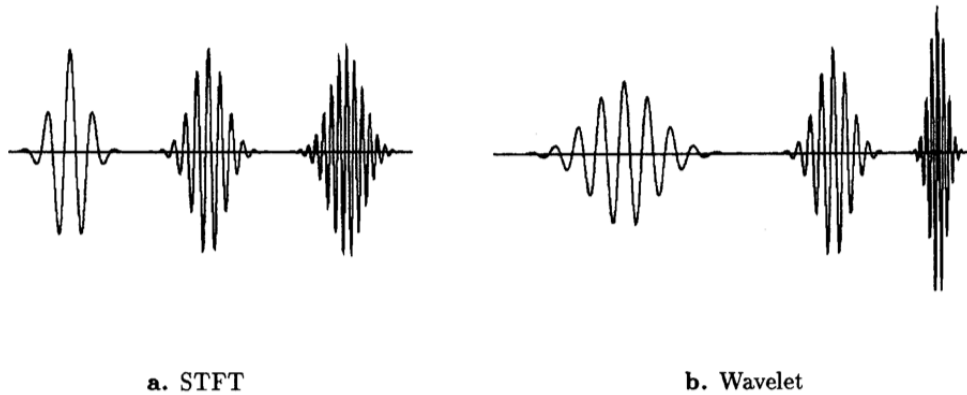


FIGURE 6.2 Basis functions for STFT and wavelet transform. The STFT (a) places a varying number of oscillations under the same envelope. The wavelet (b) keeps the same number of oscillations and scales the amplitude.

The associated **continuous wavelet transform** (CWT) is:

$$Wf(a, b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \quad (6.10)$$

where  $\psi_{a,b}(t)$  are the wavelet atoms defined in 6.9. The wavelet dictionary is ordered by the  $a, b$  parameters and represents one-dimensional signals by time-scale images in  $(a, b)$ . As opposed to windowed Fourier atoms, wavelets have a time-frequency resolution that changes, this can be seen in Figure 6.2.

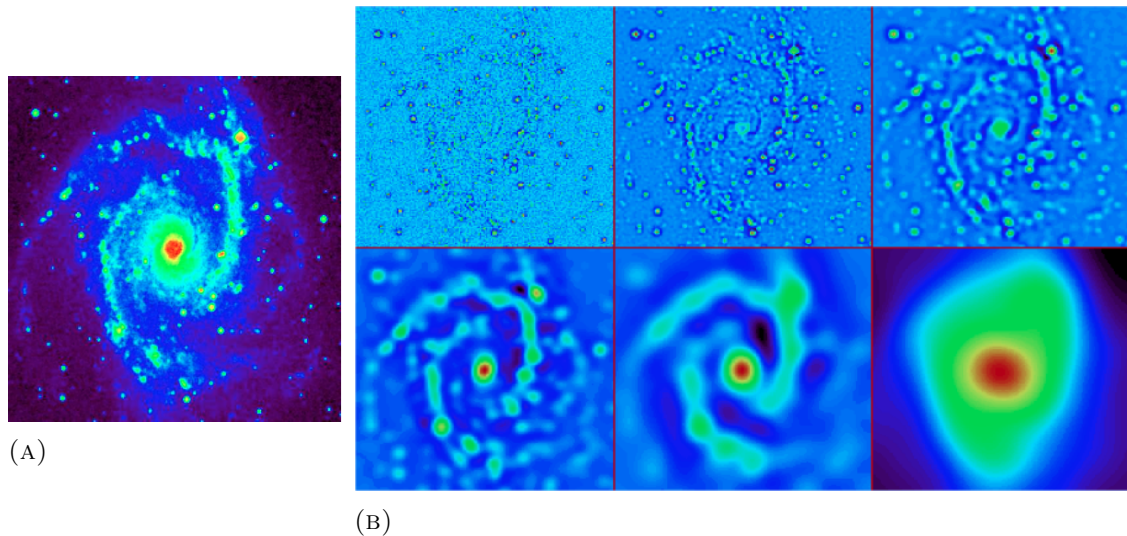


FIGURE 6.3 (A) Image of galaxy NGC2997, (B) starlet transform for different scales of NGC2997. The co-addition of these six images reproduces exactly the original image [87].

It must be noted that, to implement reconstructions based on dictionaries made from continuous transformations an infinite number of computations is required. It is then necessary to have **discrete transformations** which are constructed by tiling the frequency space. Performing this discretization is far from trivial and I will not elaborate on it. However, all following transformations are taken to be discrete.

#### 6.4.2 Two-dimensional wavelets

Until now we have spoken of one-dimensional signals. However, we want to use dictionaries to decompose an image which is not one-dimensional. A two-dimensional wavelet may be constructed directly on the two-dimensional plane, while other functions may be constructed by combining different wavelets and functions using the tensor product (e.g. directional wavelets obtained by rotating a single mother wavelet).

In the astronomical domain, the **starlet** transform is well known. It is also called isotropic undecimated wavelet transform (IUWT) and its directly constructed in two-dimensions. It is very well adapted to the detection of isotropic features (e.g. stars, galaxies). An example of a decomposition with such a transform can be seen in Figure 6.3. The application of this transform to reconstruct a signal with noise as described in Equation 6.4 is seen in Figure 6.4.



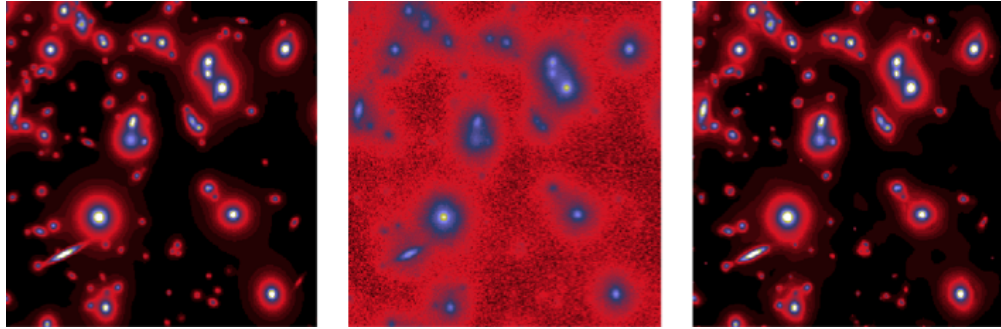


FIGURE 6.4 Simulated Hubble Space Telescope image of a distant cluster of galaxies. (a) original image, (b) aberrated image with noise, (c) wavelet reconstruction [87].

Other generalizations of the wavelet transform are the ridgelet and curvelet transforms. They incorporate alignment information and they can describe non-isotropic elements. They are obtained by rotating, dilating, and translating elementary waveforms. **Curvelets** include radial and angular windows which provide elliptical-like transforms as can be seen in Figure 6.5. **Ridgelets** are extended transformations that include angular orientation. They represent well linear-like morphologies (e.g. dead pixel columns) as can be seen in Figure 6.6.

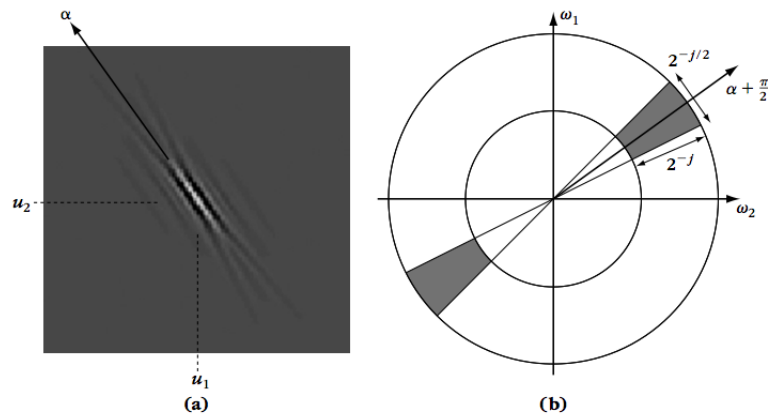


FIGURE 6.5 Example of curvelet (a). The rotation is expressed through  $\alpha$ ,  $\mathbf{u} = (u_1, u_2)$  represents the translation while  $j$  is the scale. (b) shows the so-called frequency support which is the product (wedge) of a radial window with an angular window [80].

The complexity of the MCA algorithm and its dictionaries goes much deeper than introduced in this Chapter. For example, decomposing an image with the discrete ridgelet and curvelet transforms include multiple steps as can be seen in Figures 6.7 and 6.8. I choose not to enter into details which, although of importance in the signal-processing

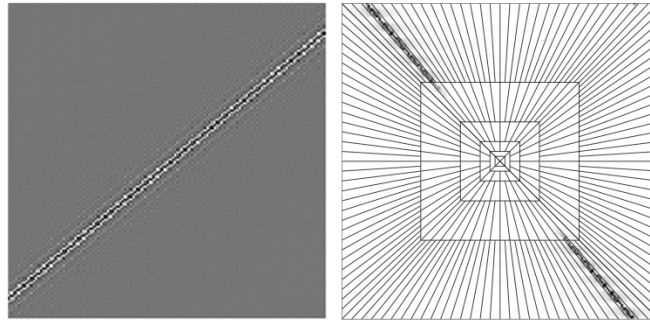


FIGURE 6.6 Ridgelet example obtained from an implementation called fast slant stack (left). The frequency support is shown in the right image [79].

field, are not vital for the understanding of my application. Such details are explained in depth in references [79] and [80].

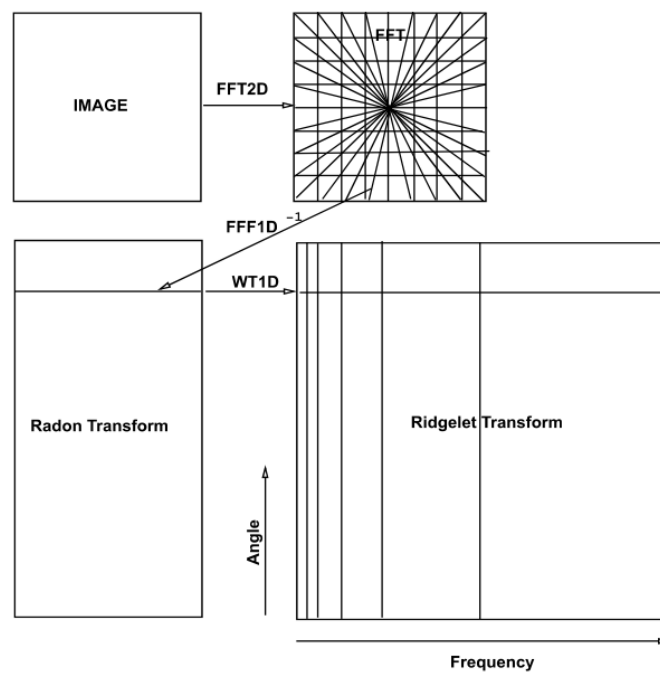


FIGURE 6.7 Flowchart of image decomposition in the discrete ridgelet dictionary. The process includes Fourier (FFT), wavelet (WT) and Radon transforms [79].

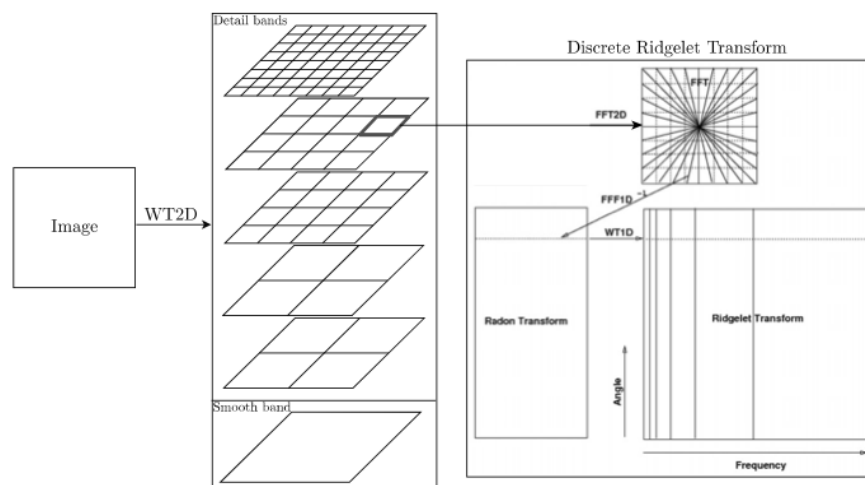


FIGURE 6.8 Decomposition flowchart of an image in the first-generation discrete curvelet dictionary. The process starts with a 2-D wavelet (WT2D) decomposition in different bands or scales, and then the ridgelet transform is applied. The latter includes internal Fourier (FFT), wavelet (WT) and Radon transforms [79].

## Chapter 7

# Improving detection of transient events

*Based on the work done with François Lanusse and Jean-Luc Starck from SAp, CEA.*

*Published in JCAP 04(2015)041 [88] and <http://arxiv.org/abs/1501.02110>*

In this Chapter I will present my work to improve the detection of transient events in the deferred photometric pipeline of SNLS for the analysis of 5-year data sample. The feasibility of detecting SNe Ia in this SNLS pipeline was already shown in Chapter 5.

The motivation behind this work is that transient event candidates in the SNLS 3-year analysis were dominated by spurious objects due to bad subtractions. Allowing these artifacts to be detected leads not only to a waste of resources but also to possible signal coordinate contamination. Our goal is to reduce the number of transient event detections while maintaining the sample of SN-like events.

This said, two main approaches can be used: changing the detection strategy or eliminating those spurious objects before detection, directly on the image stacks. In his PhD work, *Bazin* studied thoroughly the detection strategy for the deferred photometric pipeline and optimized it [89]. Therefore, I decided to address artifacts at the level of the subtracted image stacks (*lunation stacks*) which were defined in § 5.2.1. A two-step procedure using morphological component analysis algorithms was proposed in order to clean image stacks. After cleaning, I reviewed the detection procedure and adapted it

to the cleaned image stacks. An important issue for the photometric analysis, namely coordinate resolution, was also addressed.

In this Chapter, I present the problematic and common defects that yield spurious detections in Section § 7.1. In Section § 7.2, I explain how we chose dictionaries for separating transient events from artifacts and how we implemented the two step procedure, first to remove the main artifacts (§ 7.2.2) and then, to handle the non-stationary noise present in our image stacks (§ 7.2.3). Finally, a new strategy for extracting signal coordinates from our cleaned image stacks with a good resolution is introduced in Section § 7.3.

The performance of our treatment was studied using Monte Carlo (MC) artificial images. However, the available MC had to be modified for obtaining a realistic redshift (§ 7.5.2) and volume distributions of SNe Ia (§ 7.5.1). MC results for both detection efficiency and coordinate resolution are shown in Subsection § 7.5.3, while those for SNLS3 data can be found in Section § 7.4.

## 7.1 Subtracted image stacks defects

Detection of transient events is done (as shown in Chapter § 5) using subtracted image stacks in the  $i_M$  filter. Reference images are constructed for each field from a set of best quality images which are coadded. Each image of the survey has the reference image subtracted using determination of the sky background and a convolution kernel which allows the subtraction to be adapted to different observing conditions. Although it is a very standard procedure in the pipeline, it is not perfect and many defects can be present after subtraction.

We may say that there are mainly two types of artifacts. Those that can't be subtracted and those that are due to imperfect subtraction. While the latter may be reduced through subtraction optimization, the former can not. For example, bright stars may produce an optical ghost that can't be subtracted and can be present in either or both the reference and current images (e.g. Figure 7.1) leading to spurious detections ( Figure 7.2). Some surveys mask such image areas regardless that some SNe Ia may be present in the vicinity.

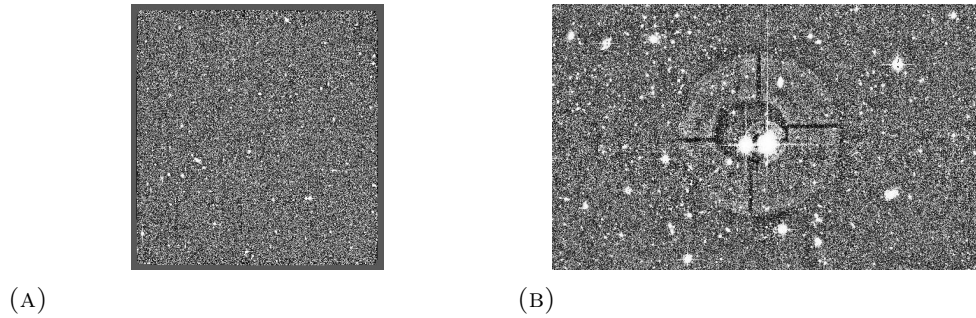


FIGURE 7.1 Reference image for field D1 7.1a and a zoom of this same image 7.1b where two saturated stars are clearly seen, with their diffraction lines and the optical ghost.

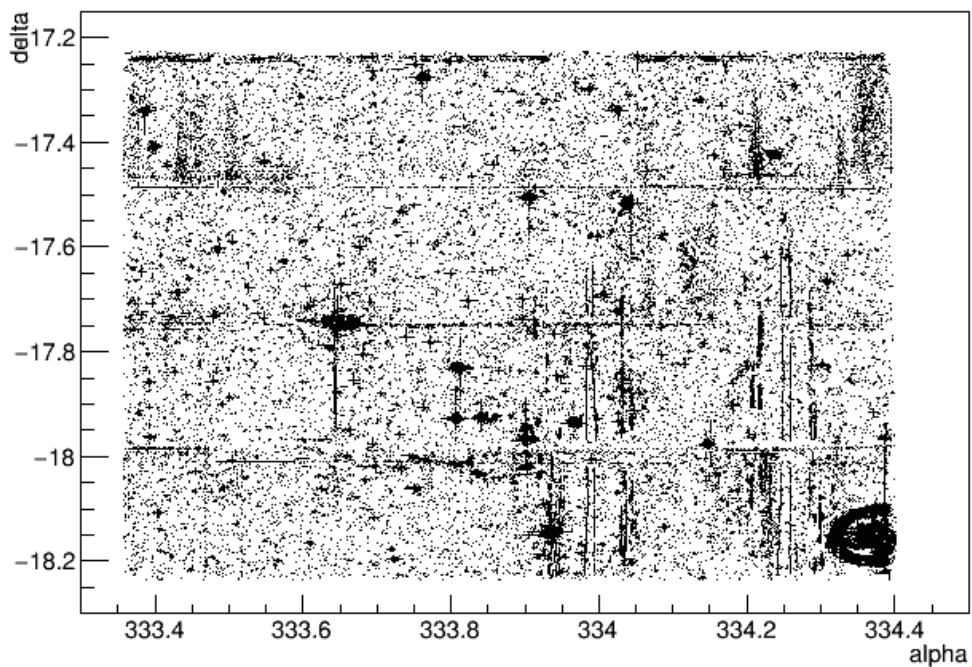


FIGURE 7.2 Detection map for field D4 in SNLS3. In this map, 90,971 detections are represented. At the bottom right an optical ghost defect is clearly seen.

Our goal is not to study the origin of these artifacts but to reduce their impact on the detection. Using SNLS3 D4 subtracted image stacks I searched for the most common defects that yielded spurious detections in the detection maps.

The most common defects come from:

- Imperfect subtraction of bright stars (e.g. Figure 7.3a)
  - Bright star residuals: part of the star's light remains on the subtracted images.

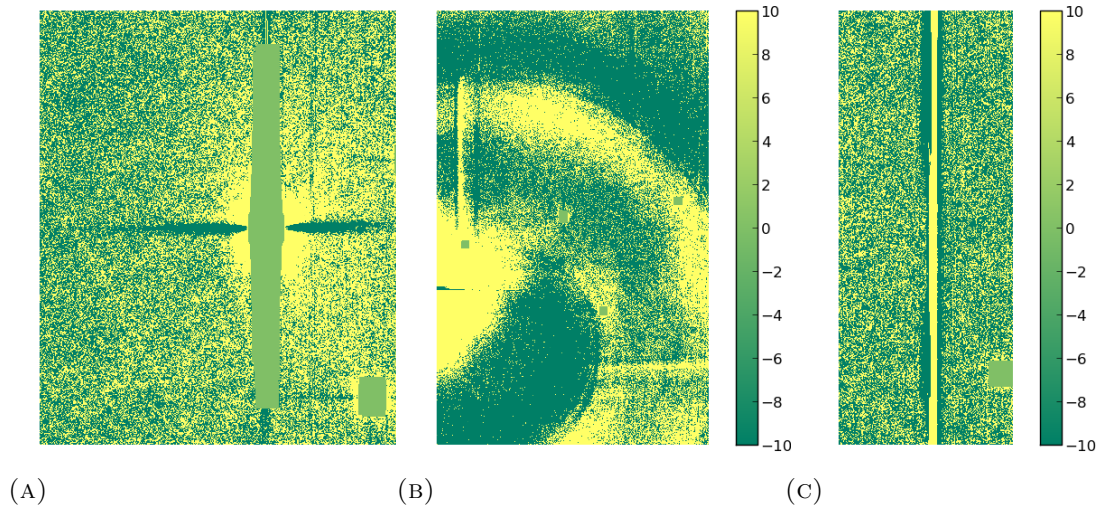


FIGURE 7.3 Different defects on the subtracted image stacks that yield spurious detections on large scale: (a) shows a saturated star with some areas masked by subtraction, (b) a saturated star plus an optical ghost and (c) defects from sampling and dead pixel lines (artificial color scale).

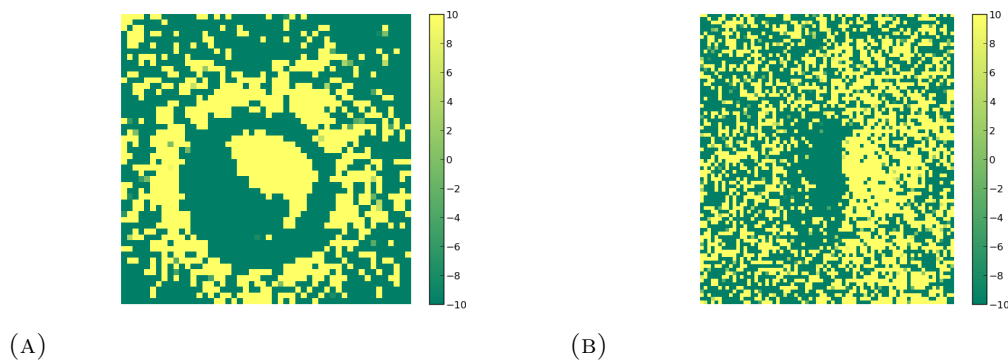


FIGURE 7.4 Defects on the subtracted image stacks that yield spurious detections on small scale: (a) and (b) dipoles from imperfect galaxy subtraction. These are adjacent positive and negative areas on the stacks (artificial color scale).

- Masks: due to saturation there is a maximum value on each pixel content of an image. When a saturated star is subtracted some pixels may have the same value in both the reference and the current image. This produces a region on the subtracted image which has pixels with zero-value.
- Optical ghost caused by reflections from the optical surfaces of the lenses in the camera around a bright stars (e.g. Figure 7.3b)
- Resampling defects and dead pixel lines (e.g. Figure 7.3c)
- Imperfect galaxy subtraction, called dipoles since they are adjacent positive and negative content areas on the stacks. They cover a smaller area than previous defects (see Figure 7.4).

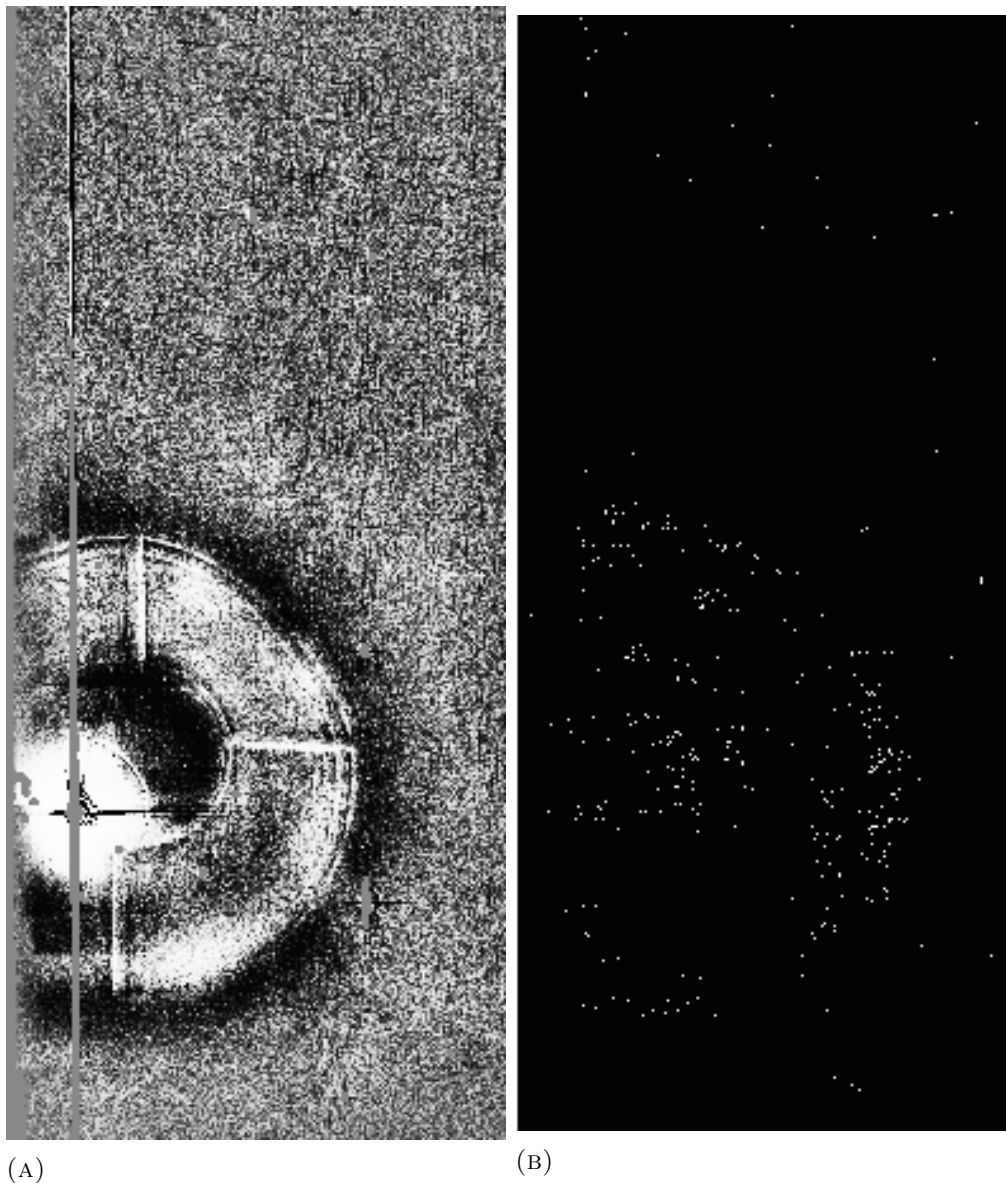


FIGURE 7.5 Image stack for field D4, ccd 00 and lunation 10 with a bright star and optical ghost in [A.3a](#) and its corresponding detection map in the SNLS3 deferred pipeline [A.3b](#). Multiple detections can be attributed to spurious objects.

These defects contribute to multiple spurious detections. For example, in [Figure 7.5](#), a saturated star with optical ghost can be seen in the image stack leading to numerous detections in the map.



## 7.2 Disentangling real transient events from artifacts

Morphological component analysis allows to disentangle artifacts from other signals and can be adapted to treat subtracted image stacks in SNLS. First, the dictionaries are chosen which characterize signal and artifacts distinctively at different size scales. Then, I present a two-step treatment designed to extract interesting SN-like signals and eliminate artifacts.

A sub-sample of SNLS3 data was used to characterize artifacts, see Section § 7.1, and to assign the algorithm parameter values. Field D4 was chosen since it is a summer field with very good observing conditions and a large number of transient detections and events classified as SN-like candidates. These candidates were defined when applying the SN-like cuts in Section § 5.3.2. In SNLS3 field D4 out of 90,971 detections 362 events were selected as SN-like candidates among which 74 were later classified as SNe Ia.

### 7.2.1 Choice of dictionaries

The aim of the filtering approach presented in this section is to separate the signal of interest (SN-like events) from a complex background. The latter is constituted by noise, defects that cannot be subtracted (e.g. Figure 7.3) and features from imperfect subtractions (e.g. Figure 7.4).

A detection strategy based on applying a signal-over-noise cut does not reject many artifacts and yields a large number of spurious detections. Our aim is to leverage additional morphological information to separate the signal of interest from artifacts and noise, by exploiting their stark contrast in both shape and scale.

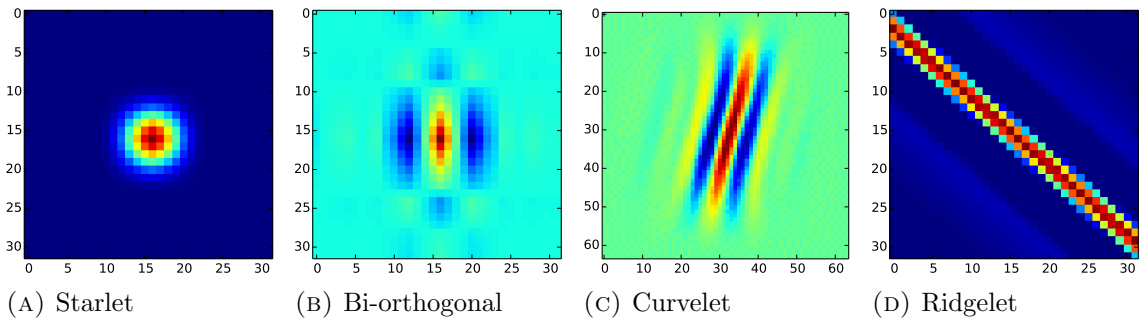
The MCA algorithm separates images into a number of morphological components, using the sparsity level of each component in appropriate dictionaries as a discriminant. Therefore, in the case of SNLS data, it is important to select, on one hand, a dictionary adapted to the morphology of the signal of interest and, on the other hand, additional dictionaries adapted to the artifacts we want to reject. More information on available dictionaries can be found in Chapter 6.



(A) SN 04D1c with its host-galaxy

(B) SN 04D1c in subtracted image.

FIGURE 7.6 SN 04D1c with its host-galaxy and after subtraction.



(A) Starlet

(B) Bi-orthogonal

(C) Curvelet

(D) Ridgelet

FIGURE 7.7 Typical atoms from the dictionaries used in the MCA algorithm. (a) starlet atom representing circular-like signals, (b) bi-orthogonal wavelets for dipole features, (c) curvelets for elliptical signals and (d) ridgelets representing line features (artificial color scale).

Dictionaries can be used to probe features of different sizes. The latter are usually represented by discrete scales  $j \in \llbracket 0, N \rrbracket$ , where the number of scales,  $N$ , can be chosen. Typically, atoms of these transforms at a scale  $j$  have a characteristic size in pixels of  $2^j$ , starting with the finest resolution with details at the pixel scale for  $j = 0$ . The advantage of choosing different scales for each dictionary is that we are able to separate small scale signals from large and small scale defects.

### Signal

SN-like signals are small scale ( $r \lesssim 10$  pixels) circular type shaped objects (specially on subtracted images, see Figure 7.6). A wavelet based dictionary is suited to this kind of morphology. We choose the starlet dictionary since it is composed of isotropic atoms, especially efficient for representing either positive or negative structures such as our SN candidates. An example of a starlet atom is presented on Figure 7.7a.

### Artifacts

For the small scale artifacts (usual size a couple of tenths of pixels) presented in Figure 7.4 we adopt a bi-orthogonal wavelet dictionary (Figure 7.7b). These artifacts result from

improper subtraction of galaxies which lead to characteristic dipole features. The bi-orthogonal dictionary has the advantage of representing such features more efficiently than the starlet, enabling us to discriminate these artifacts from the signal.

For large scale curved or line artifacts (spanning more than 100 pixels) such as the ones in Figure 7.3, we adopt curvelet and ridgelet dictionaries. The curvelet dictionary is composed of localized, elongated atoms which are known to provide a sparse representation for curved features, see Figure 7.7c. The ridgelet atoms are line of different widths and orientations (see Figure 7.7d) which are perfect to represent the second type of artifacts.

### 7.2.2 First treatment: removal of main artifacts

Until now, I have just described the different morphologies of our signal and artifacts as well as the dictionaries that better represent them. Now, we can make use of a robust tool that can decompose our subtracted image stacks efficiently, that is the MCA algorithm by *J.L. Starck* in [85]. This algorithm allows one to choose a set of dictionaries and parameters to decompose an image. The latter were introduced in Chapter 6.

First, using SN-like objects and artifacts from the D4 SNLS3 test-sample, I chose which scales and thresholds values were suitable for each dictionary. This was done by decomposing the signal or artifact using the algorithm and varying only one or two parameters while searching the best decomposition.

- Thresholds : for a dictionary, the higher the threshold, the most adapted to the dictionary an object must be to be assigned to it <sup>1</sup>. After multiple trials and errors the choice was set to 5.
- Scales for each dictionary:
  - Starlet: three scales were required to account for small scale objects. SNe are not punctual but extended objects, so I chose to eliminate the first scale  $2^0$ .
  - Bi-orthogonal wavelets, ridgelets, and curvelets: five scales ( $j = 1, \dots, 5$ ) to account for small and large scale artifacts.

---

<sup>1</sup>This threshold is related to the initial  $\sigma$  required in Equation 6.5

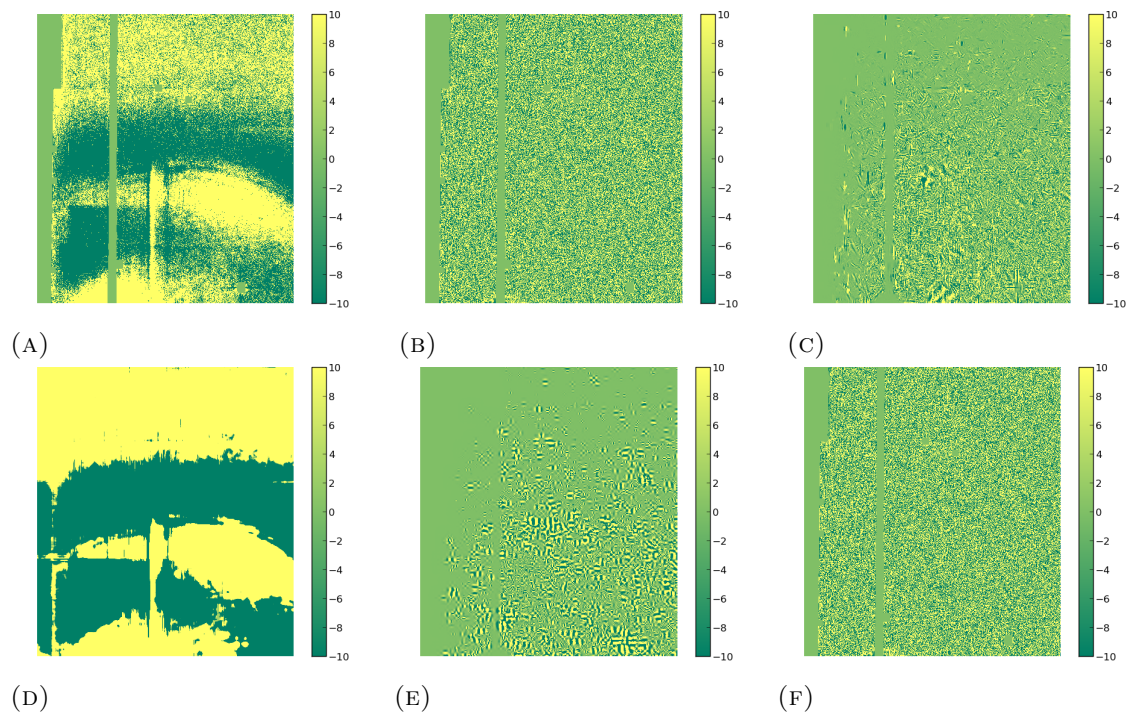


FIGURE 7.8 MCA decomposition of an optical ghost (large scale): (a) the artifact in the original subtracted image stack, (b) the starlet component, (c) the curvelet component, (d) the ridgelet decomposition which represents very well the artifact, (e) the wavelet component and (f) the residuals left after the decomposition (artificial color scale).

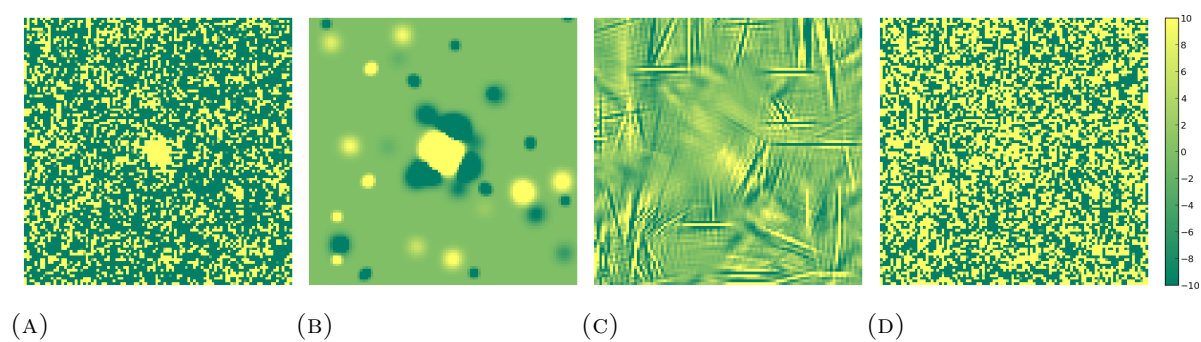


FIGURE 7.9 MCA Decomposition of a SN Ia event: (a) shows the original subtracted image stack centered on the SN event (yellow spot), (b) the starlet component, where the SN signal (yellow spot) is surrounded by remaining galaxy residuals (green spots), (c) the curvelet component and (d) the residuals left after the decomposition (artificial color scale).

Second, algorithm options related to the nature of our images, subtracted stacks, were imposed. For example, both isotropic positive and negative signals were decomposed since some SNe may have part of their flux included in the references, in which case subtraction yields a negative valued signal. Also, when subtracting saturated pixels (which have the threshold value), regions with zero values were created which are similar to masks which must be indicated to the algorithm.

Other parameters in the algorithm were chosen from a trade-off between reducing the total number of detections and keeping most of the SN-like objects in the D4 test sample. Another consideration was the computation time. Example of this was the number of required iterations, a compromise between number of iterations and computation time was achieved with 30 iterations for the decomposition.<sup>2</sup>

The transforms associated to the algorithm's dictionaries, especially that of the curvelet, do not scale well with the image size and too much CPU time and memory would be required for a SNLS subtracted image stack of 2176 by 4912 pixels. Therefore, we tiled them, both to reduce time and memory resources and to allow parallel processing. For reference, one SNLS subtracted image stack divided in 8 tiles requires on average 6 days of HS06 CPU time and 500 Mb of virtual memory to be treated.

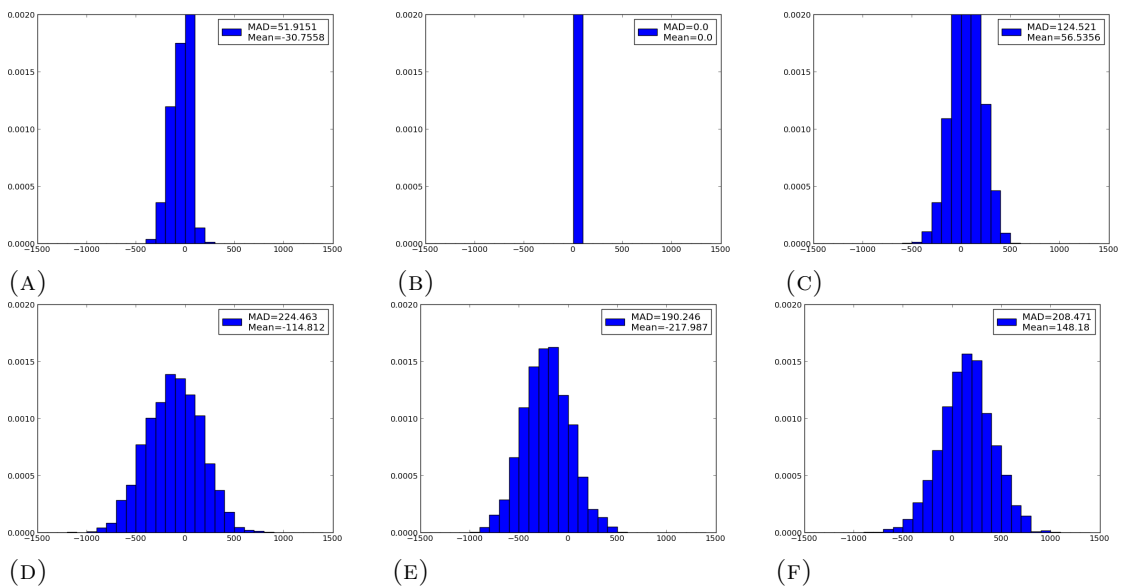


FIGURE 7.10 Varying noise: pixel content in a 50x50 pixels box at different locations in a subtracted image stack before treatment. Histograms are normalized and axis scales are kept fixed to illustrate variation from one location to the other.

<sup>2</sup>Other parameters include norm decrease and type of thresholding (hard or soft).

The first treatment’s algorithm provides images in each dictionary. In our case, we have images in the starlet, wavelet, curvelet and ridgelet dictionaries plus an image containing the residuals. The latter contains mostly noise and other features not properly assigned to any dictionary. Examples of artifacts and signal decompositions can be seen, for a defect, in Figure 7.8 and, for a SN, in Figure 7.9.

The algorithm assumes a stationary and Gaussian noise in the input images which is not the case for our subtracted image stacks (coaddition of subtracted images spanning several weeks of observations). Figure 7.10 shows different regions of a subtracted image stack where it is clearly seen that the noise shape is not gaussian and has varying spread. As a consequence, some SN signals were not properly decomposed and were partially in the residuals, e.g. Figure 7.11.

Thus, at the end of the first treatment the interesting signal may be both in the starlet component and in the residuals of our first treatment decomposition (e.g. components (b) and (d) of Figure 7.9). In the following, I will call “output” of the first treatment the combination of the starlet component and the residuals. An example of such an output can be seen in Figure 7.12 where the reduction of artifacts is clearly seen. An example of the varying noise still present after the first treatment is shown in Figure 7.13.

It must be noted that some of the signal can be also decomposed in the wavelet dictionary but this dictionary is very efficient at decomposing dipoles which are artifacts I want to eliminate. Therefore, combining starlet, wavelet dictionaries and residuals is not a good option.

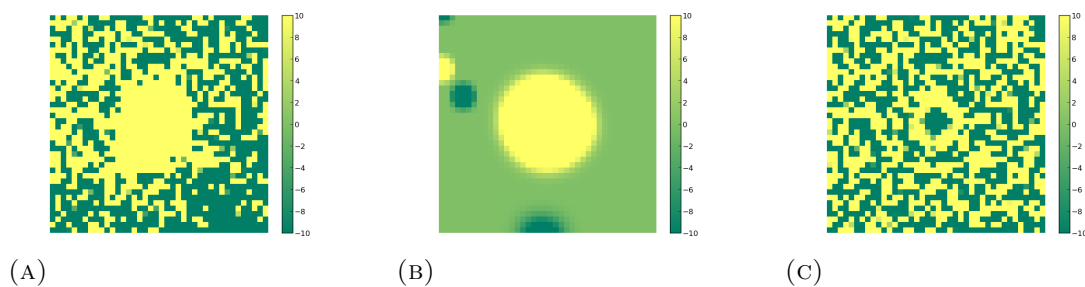


FIGURE 7.11 MCA Decomposition of a SN Ia event where part of the signal leaks into the residuals. (a) shows the original subtracted image stack, (b) the starlet component and (c) the residuals after decomposition, containing part of the signal.

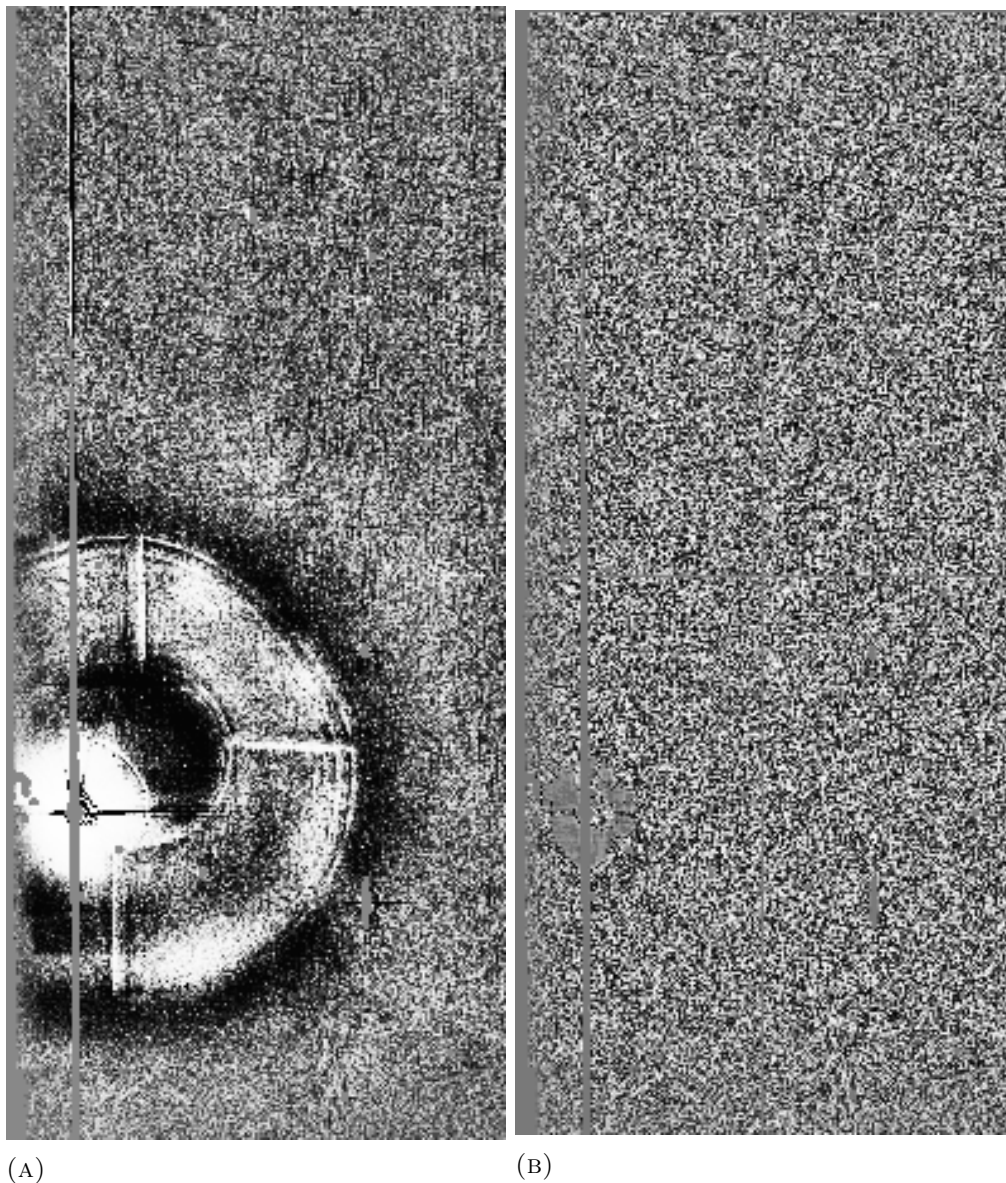


FIGURE 7.12 Image stack for field D4, ccd 00 and lunation 10 before (a) and after (b) first treatment. The optical ghost and part of the saturated star are eliminated in (b).

### 7.2.3 Second treatment: signal extraction with varying noise

Now, many of the artifacts on our subtracted image stacks are eliminated but there are still some spurious artifacts and noise. To handle the non-stationary noise, it was necessary to develop another utility based on the algorithm in [79]. It handles that type of noise and exploits further morphological decomposition for extracting circular-like signals. Once again, the method's parameters were tested using the SNLS3 D4 subsample.

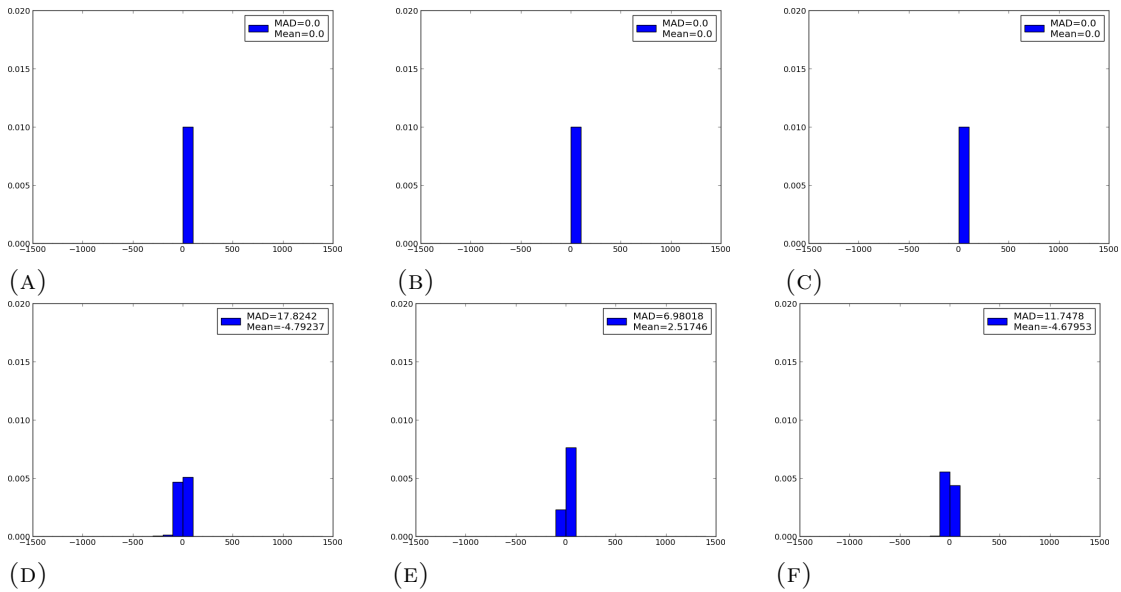


FIGURE 7.13 Varying noise after first treatment: pixel content in a 50x50 pixels box at different locations (same as in Figure 7.10). Histograms are normalized and axis scales are kept fixed to illustrate variation from one location to the other.

Non-stationary noise requires varying the decomposition's threshold for a given dictionary depending on the position of the analyzed pixel. Such a feature can be easily implemented in the starlet dictionary since it can handle actual noise maps (not the case for other dictionaries like curvelets). The latter were computed using a sliding window and a median absolute deviation estimator (MAD) to estimate noise dispersion. The size of the box was chosen to be larger than an individual signal-like object. The procedure was as follows: for each pixel in the noise map, a box of 50 x 50 pixels was used to calculate the MAD and this value was assigned to that pixel. Then the box was slid into the next pixel and the same procedure applied until a complete coverage was done.

As in the first treatment, since some of the SNe may have part of their flux included in the references, both positive and negative signals were treated. The nature of the first treatment output differs from the one of subtracted image stacks, which implied a specific optimization of the decomposition algorithm parameters.

As previously, a D4 sub-sample was chosen to select visually some parameters of the decomposition. Once again, the first scale of the decomposition was not taken into account. I chose 7 scales since SN-like signals are bigger in the starlet reconstruction than in the original image stack.



*The dipole issue:* in spite of the first treatment, at this stage there were still some dipoles present in our cleaned image stacks. Several attempts were done to eliminate them. Since dipoles are simply negative and positive adjacent areas, a first approach was to select areas of positive or negative pixels, see if they were next to each other and eliminate them. Several attempts were tried to set up criteria defining dipole artifacts such as:

- Size: requiring sizes to be above and/or below a certain value.
- Absolute maximal value: requiring that the areas had extreme pixel values.
- Mix of the above criteria.

When applying constrains using these criteria there was a big loss of SN-like events in the sample. Therefore, I decided to optimize the algorithm parameters without any extra criteria to deal with dipoles.

The second treatment utility does not require tiling images since only the starlet dictionary is used. One SNLS image output from the first treatment takes on average 3 hours of HS06 CPU time and 100 Mb of virtual memory to be processed.

All signals present in the output image can be considered as morphologically compatible with circular-like objects. An example of a decomposition of a SNIa event can be seen in Figure 7.14, while an example of a cleaned subtracted image stack and its evolution during the two-step procedure can be seen in Figure 7.15.

Now, I must note that this is a transformed and cleaned image where objects may be larger and displaced slightly from their original position (e.g. size for SNIa example in Figure 7.14). This fact will be very important for the detection procedure explained in the next section.

### 7.3 New Detection strategy

A detection strategy includes both extracting events from an image and reconstructing their coordinates. Event extraction depends on the image and its characteristics, e.g. its local noise information. The TERAPIX tool SExtractor [73] was used for the whole

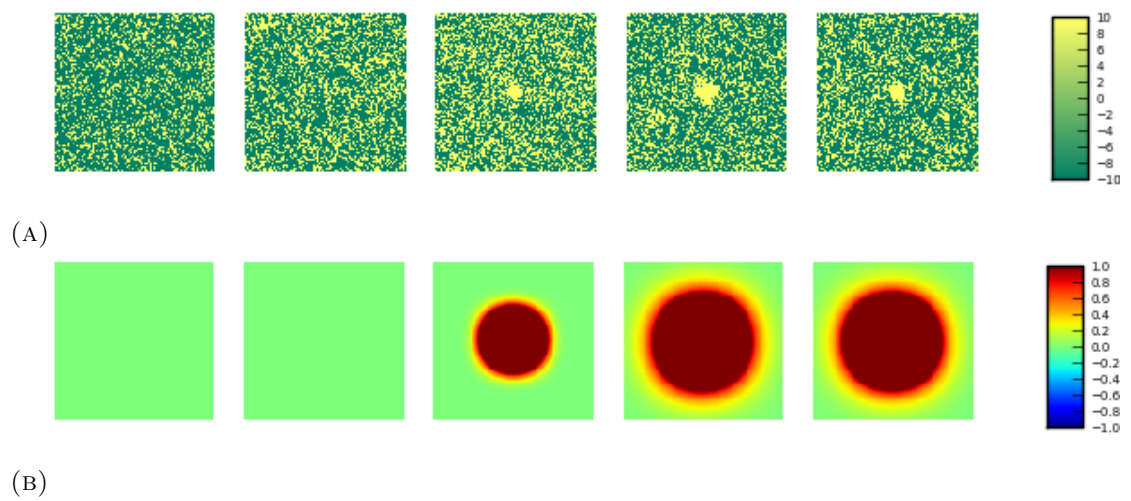


FIGURE 7.14 A SNIa event shown in different lunations around maximum light in the original subtracted image stack (a) and after both treatments of the cleaning procedure (b).

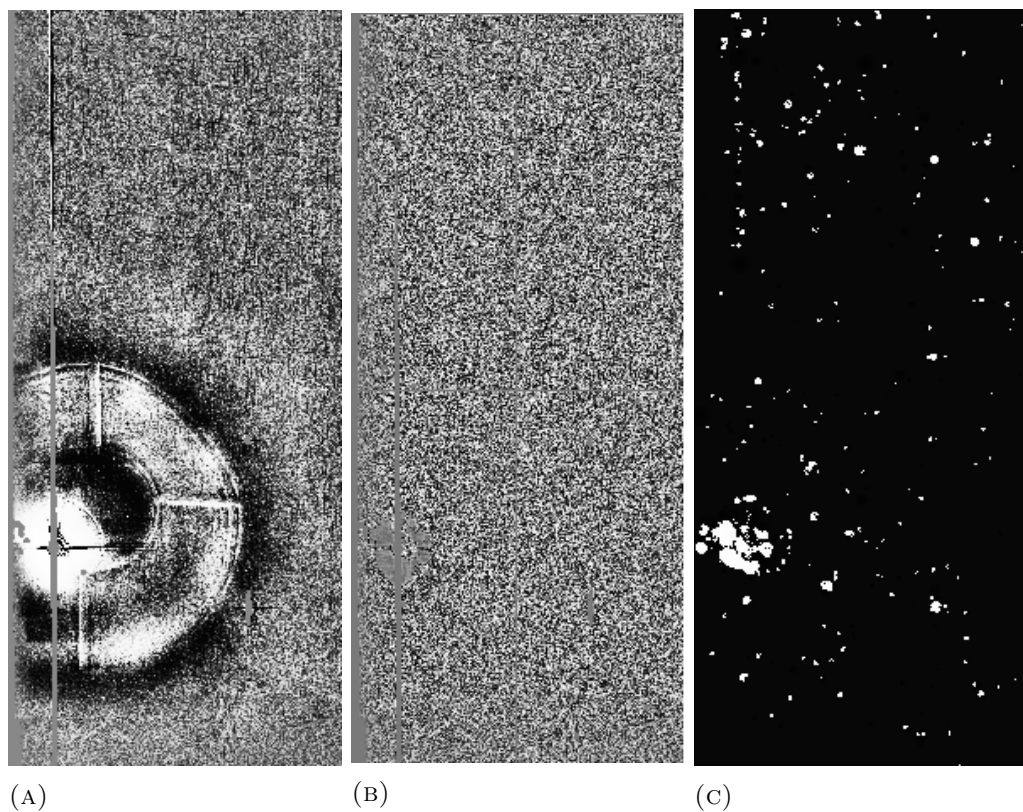


FIGURE 7.15 Image stack for field D4, ccd 00 and lunation 10. Original subtracted image stack is seen in (a), output of the first treatment in (b) and cleaned stack after second treatment in (c).

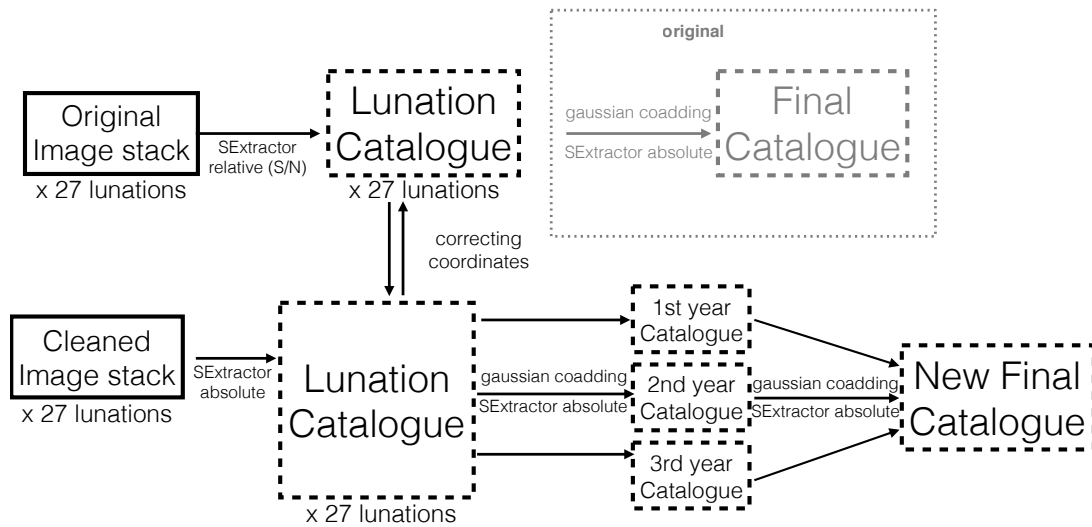


FIGURE 7.16 New detection strategy schema. Dotted lines represent catalogue ASCII files while continuous lines stand for images.

detection strategy both in the original procedure and the new procedure, adjusting its parameters accordingly.

*Original detection strategy (see Chapter § 5):*

In the previous work by *Bazin et al.* [66], the detection strategy consisted in constructing lunation catalogues using SExtractor with deblending, requiring for each detection at least 4 pixels with a flux of more than  $2.5\sigma$  w.r.t. sky background. A final detection catalogue was obtained by merging all lunation catalogues obtained in three years and converting the result into an image where each detection was replaced by a Gaussian of height and width of 1. This image was processed with SExtractor selecting only pixels with a content above a value of 0.01 and deblending objects. This was described on Figure 5.4 and on the top part of figure 7.16.

Our two-step treatment outputs do not have the same properties as the original subtracted image stacks. The noise has been removed. As transformed images they have less objects but original object coordinates cannot be determined accurately from them. I thus proposed a new detection strategy (see Figure 7.16 bottom) which also addresses the degradation of coordinate resolution when using several years of data.

The new detection strategy can be divided in three steps as described with more detail in the next sections. First, detections are validated using the cleaned lunation stacks.

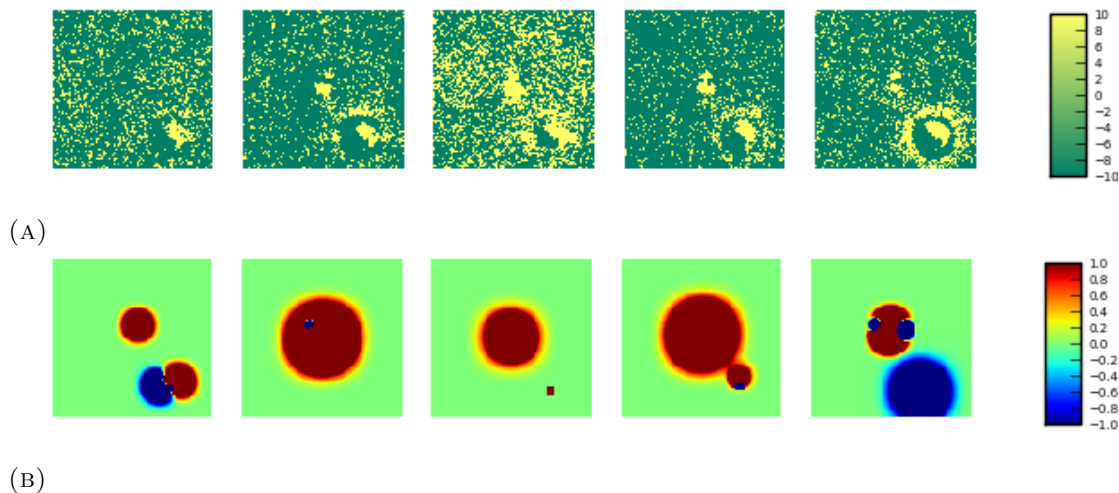


FIGURE 7.17 A SNIa event (center of the image) with galaxy residuals, shown in different lunations around maximum light in the original subtracted image stacks (a) and after cleaning (b). In the original stacks, galaxy residuals are present in all lunations. The cleaning removes them in some cases (artificial color scale).

Second, coordinates are assigned to these detections at the level of lunation catalogues. Third, I propose a new way of stacking when considering many years of data.

### 7.3.1 Validating a detection

Lunation catalogs are constructed from our cleaned subtraction stacks using SExtractor. Deblending is imposed in order to separate adjacent objects. The values of the SExtractor parameters were tuned using the SNLS3 test sample. They resulted from a trade-off between the reduction of the total number of detections and the number of SN-like objects detected on the cleaned image stacks.

Since cleaned subtraction stacks have been already denoised, the pixel value threshold can be set to an absolute, and low, value (no signal-over-noise ratio can be computed). Hence, pixels are required to have a minimum value of 1 to be considered in the detection. It must be noted that the pixel content of the cleaned stacks has no relation with the flux value of the original subtracted image stacks.

Also, interesting objects were larger in the transformed images. Varying the number of required pixels above threshold to detect an object in our test sample (see Figure 7.18), I decided to require at least 200 pixels with a signal value above one to confirm an object. This number of required pixels allows an important reduction on the number of detections to be achieved, while preserving SN-like objects in the sample.

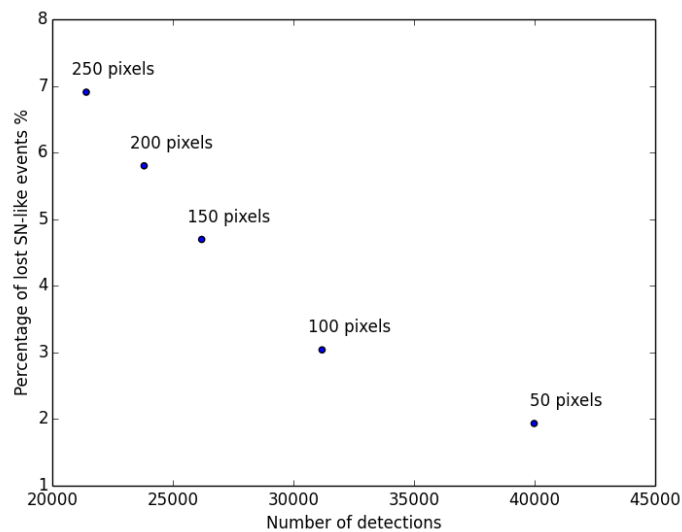


FIGURE 7.18 Percentage of SN-like events that were lost in the D4 SNLS3 test sample with respect to the number of detections for different choices of the minimum value of pixels required to validate a detection (before setting up the complete new detection strategy).

### 7.3.2 Assigning coordinates

Once detections are validated I want to determine precisely their coordinates. This is not possible directly in cleaned image stacks since objects can be slightly displaced or deformed. However, detections from our cleaned stacks can be matched to coordinates in the original procedure catalogues.

It is a very simple process. To each object detected in a lunation I assign the coordinates of the closest detection in the same lunation catalogue of the original procedure. In this way I maintain the reduced number of candidates while having precise coordinates.

It must be noted that when comparing this new strategy with the old one, no new events are detected. I will show in the following sections that the number of detections is strongly reduced for both data (Section § 7.4) and MC studies (Section § 7.5) while keeping good detection efficiency for SNe Ia (Section § 7.5).

### 7.3.3 Improving coordinate resolution

In the original procedure coordinates were averaged over all lunations in the survey period (e.g. 3-year survey). This can degrade signal coordinate resolution due to close-by spurious detections. The latter are not always completely removed by cleaning as seen in Figure 7.17b. When adding data from several seasons, the coordinate resolution degradation becomes more important.

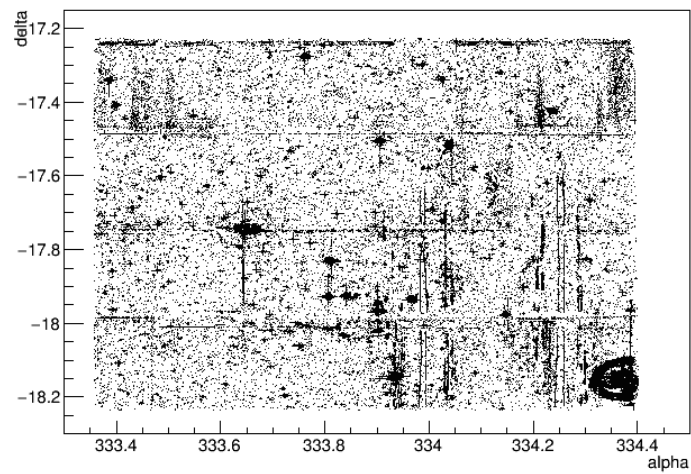
Physically, real SN Ia events can be present in at most three adjacent lunation catalogs but not over several seasons. We can take advantage of this by only averaging lunations which may have the same SN-like objects.

Hence, to address this in the new procedure we:

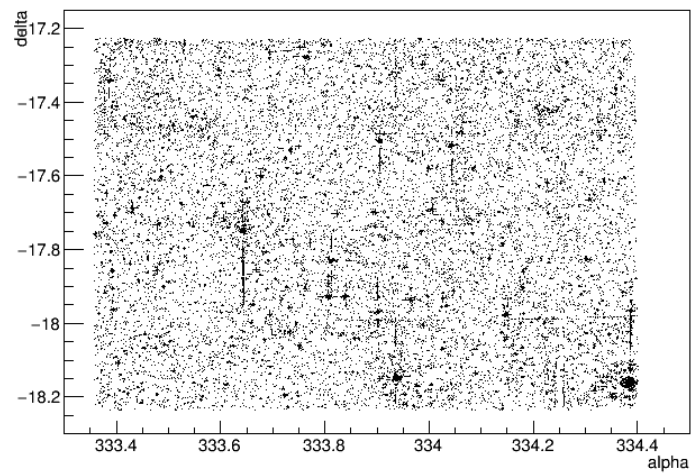
1. *Build a catalogue for each season:* as for the final catalogue in the original procedure that is merging lunation catalogues obtained in a season and converting the result into an image where each detection is replaced by a Gaussian of height and width of 1. Then, this image is processed with SExtractor selecting only pixels with a content above a value of 0.01 and deblending objects.
2. *Build a final catalogue:* merging all season catalogues and once again converting this into an image with 1x1 Gaussians. Then SExtractor selects pixels with content higher than 0.01 in the same way as before.

In this way, coordinate averaging is done first for a season where a transient object can be present and then detections are added from other seasons. It is equivalent to assigning a weight for a given detection taking into account that a SN will be detected only during one season.

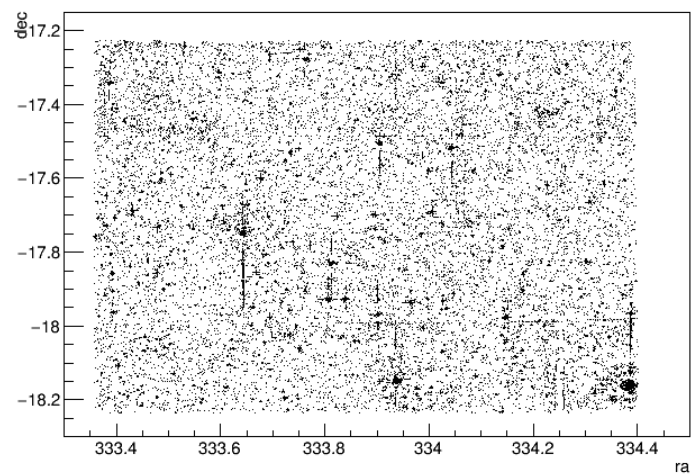
For the D4 SNLS3 test-sample the detection map using the new procedure is cleaner compared to the original procedure map as can be seen in Figure 7.19. The advantage of using season stacks can not be appreciated in such a map but it is seen in coordinate resolution. The latter will be addressed in Section 7.5 using Monte-Carlo studies.



(A) Original procedure: 90,971 detections.



(B) New procedure without season stacks.



(C) New procedure: 40,575 detections with clear reduction of detections due to large scale defects.

FIGURE 7.19 Detection maps for field D4 in SNLS3.

## 7.4 Results with SNLS3 data

SNLS3 data motivated this work since the difference between detected transient events and SN-like events was of two-orders of magnitude. Now, we are able to see the impact of our new procedure in these data separated by fields which is summarized in Table 7.1.

First, let us present field D4 since it was the one used for tuning both algorithm and detection parameters. Using the original procedure 90,971 detections were found from which 362 events were extracted as SN-like objects as described in *Bazin et al* [66]. After our processing, the number of detections is reduced to 40,575. This represents more than a factor of 2 reduction on the number of candidates to be further processed. Loss of SN-like candidates is less than 5% and all lost events are faint (observed magnitude at peak in  $i_M > 24.2$ ). These faint events are usually not well measured, have a large magnitude uncertainty, and therefore are not suitable for further cosmological analysis. It is known that, for previous photometric pipeline SNLS studies, the magnitude uncertainty for events with magnitude above 23.8 is larger than 0.1 *mag* which is the upper limit for uncertainties in the SNLS spectroscopic cosmological analysis.

For all fields, the reduction of the number of detections is similar. The loss of SN-like events is less than 5% in D3 and 15% in D1 and D2. It must be noted that D1 and D2 are the fall and winter fields which have less suitable weather conditions than D3 and D4. All lost events are faint with the exception of one medium brightness event in D1 which is lost during our new detection procedure. This event is found on the output images of the two-step treatment but the number of pixels above threshold is smaller than our criteria to validate a detection.

In summary, when applied to real SNLS3 data, 10% of SN-like events are lost while the number of detections is reduced by more than a factor of two. Almost all lost events are faint with the exception of one medium brightness event which is lost in the detection step.



Field	Old procedure		New procedure	
	# detections	# SN-like	# detections	# SN-like
D1	76,806	444	34,314	382
D2	64,763	300	28,627	258
D3	70,447	377	29,292	359
D4	90,971	362	40,575	346
All	302,987	1,483	127,808	1,345

TABLE 7.1 Number of detections and SN-like events for the original and new procedures applied on SNLS3 data.

## 7.5 MC studies

As in the SNLS3 analysis presented in Section 5.2.3, the performance of our treatment was studied using Monte Carlo (MC) artificial images in the  $i_M$  filter for the D1 field .

The simulation was done in four stages summarized in Chapter § 5. In this case, subtracted MC image stacks were processed by our optimized pipeline.

The MC was used for determining detection efficiency and coordinate resolution for our SNe Ia. In order to avoid biases in these measurements, I had to adapt the MC. For example, the cleaning algorithms do not respond well in high signal density areas and close-by SNe are physically very rare, therefore I proposed to apply an isolation criterion in order not to bias detection efficiency. For obtaining a realistic coordinate resolution I also took into account a correction of the SN redshift and rate distributions.

### 7.5.1 Corrections: Isolation

SNe in the simulation are randomly distributed over a galaxy catalogue. Then it is possible to obtain SNe very close by despite the requirement of a minimal distance of 0.7 arcseconds between simulated supernovae.

In real images, statistically superposed SNe are very rare but also close-by SNe. I wanted to test our detection procedure in realistic subtracted image stacks. Moreover, high density areas are not properly handled by our cleaning algorithms. Thus I decided to put a tighter isolation on the SNe that enter the measurements.

This was done by selecting those MC SNe that were not in dense regions. A good indicator of dense regions is the distance between two objects. To determine the isolation

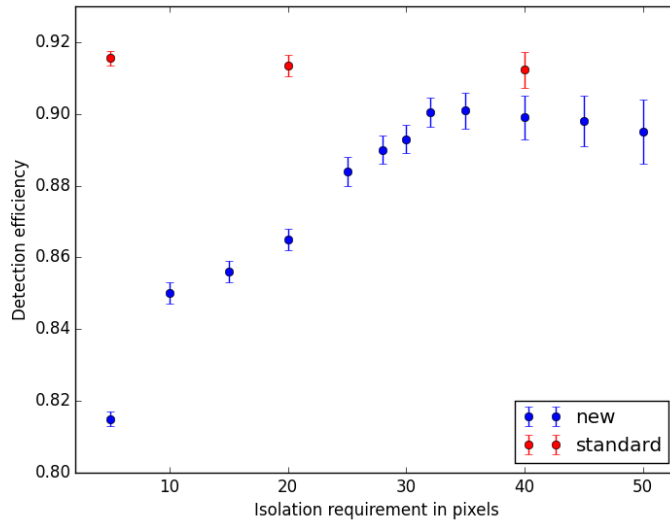


FIGURE 7.20 Detection efficiency versus isolation distance for an early version of the new detection procedure compared with the standard detection procedure.

criterion needed, I monitored the detection efficiency for one year MC while varying the minimum distance required, in pixels, between two SNe.

Figure 7.20 shows the detection efficiency of an early version of the new detection procedure as a function of the imposed minimal distance between SNe. Detection efficiency increases with higher isolation, which is expected from a better performance of the cleaning algorithm until it reaches a plateau. I chose an isolation of 30 pixels to be on the high-efficiency plateau while keeping enough data. In order to have an equivalent in the standard procedure I also isolated events but this time only 20 pixels were necessary to be in the maximum-efficiency plateau.

### 7.5.2 Corrections: Redshift and SNIa rate distributions

In the D1 field 216,000 SNe were generated by Ripoche. The random distribution over a galaxy catalogue resulted in a redshift distribution of simulated SNe which is physically unrealistic as shown in Figure 7.21. However, the redshift distribution should be coherent with a comoving volume distribution and a typical SNIa rate.

To account for this distribution, I applied a weight on the results of our MC studies depending on the SN redshifts. The weights take into account both comoving volume and SNIa rate distributions.

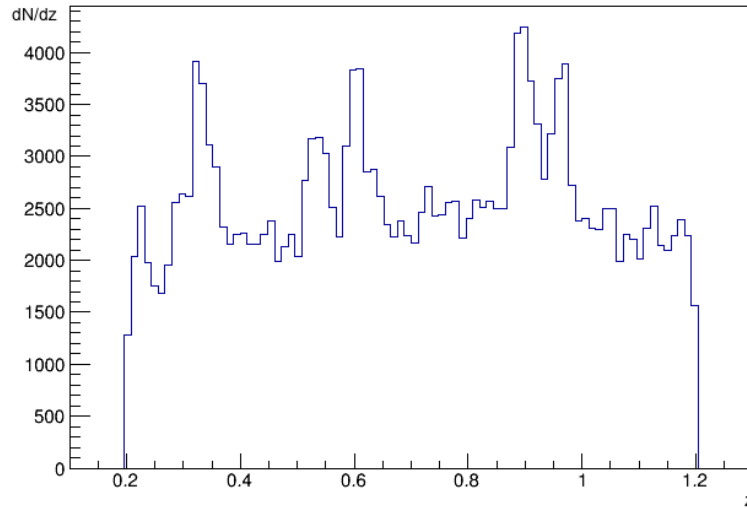


FIGURE 7.21 Redshift distribution of SNe Ia generated by MC in field D1.

The comoving volume in our flat Universe can be expressed as:

$$V = \frac{4\pi}{3} \left( \frac{c}{H_0} \int_0^z \frac{dz'}{H(z)/H_0} \right)^3, \quad (7.1)$$

where  $H(z)/H_0$  can be taken from equation 2.28 using the  $a$  dependence in  $z$  from equation 2.9. I am interested in a weight by redshift bin. Then, I want to express the differential comoving volume as a function of redshift. Since it is a weight, I may ignore all constants and just take into account the redshift dependence. Then:

$$dV_c \propto \left( \frac{dz}{H(z)/H_0} \right) \left( \int_0^z \frac{dz'}{H(z)/H_0} \right)^2. \quad (7.2)$$

For our study I approximated  $\Omega_m \approx 0.3$  and  $\Omega_\Lambda \approx 0.7$ . The resulting comoving volume distribution is shown in Figure 7.22.

Using the simple SNIa rate by Pritchett presented in Section § 3.2.4 and combining it with the comoving volume for 100 bins in the simulation redshift range I obtained the necessary weights for our analysis. The effect of weights on the redshift distribution can be seen in Figure 7.23.

In this sense, I can study the efficiency of detection and coordinate resolution of our new treatment using a weight for a correct SNe Ia distribution on redshift.

With corrections taken into account, I determined the coordinate resolution and efficiency for the new and old procedures.

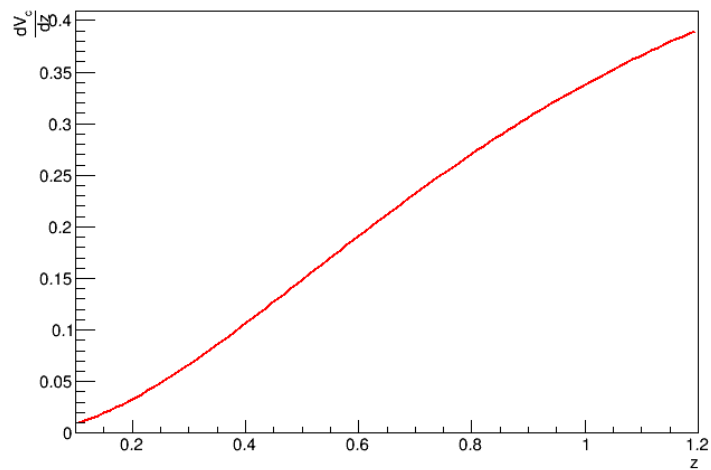


FIGURE 7.22 Comoving volume distribution for our MC studies

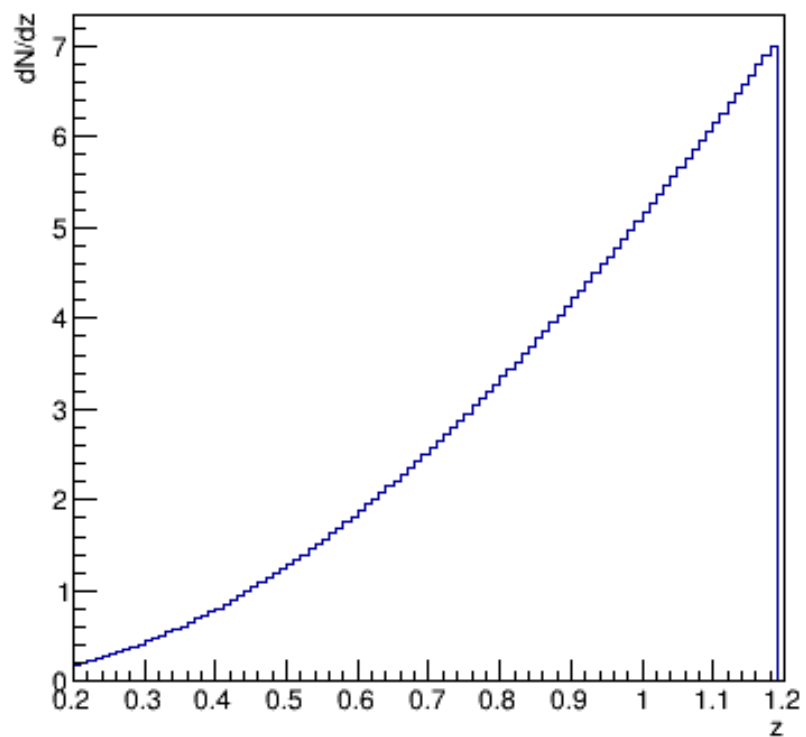


FIGURE 7.23 Redshift distribution for MC after rate and comoving volume weights are applied. Arbitrary normalization.

### 7.5.3 MC coordinate resolution and efficiency

#### Detection efficiency

Detection efficiency is defined as the fraction of simulated supernovae recovered at detection. For both the original and the new procedures I computed the detection efficiency for one year of simulated SNe Ia as a function of the generated SN peak magnitude in the  $i_M$  band as can be seen in Figure A.4. The efficiency is nearly magnitude independent up to  $m_{0i} = 23.5$  and then steeply declines at faint magnitudes. This is typical of magnitude-limited surveys such as SNLS.

When compared to the old procedure, the new procedure corresponds to a loss of 0.5% in the plateau efficiency and a 0.2 downward shift of the magnitude corresponding to 0.50 efficiency. The small reduction for events at higher magnitudes can be explained since signal separation is not perfect and some SN-like signal may not be properly transformed. Note that after the new procedure the efficiency behavior as a function of magnitude is close to the original one.

This MC result will allow us to correct the Malmquist bias of the 5-year photometric sample to be derived from the new procedure in order to perform a cosmological analysis.

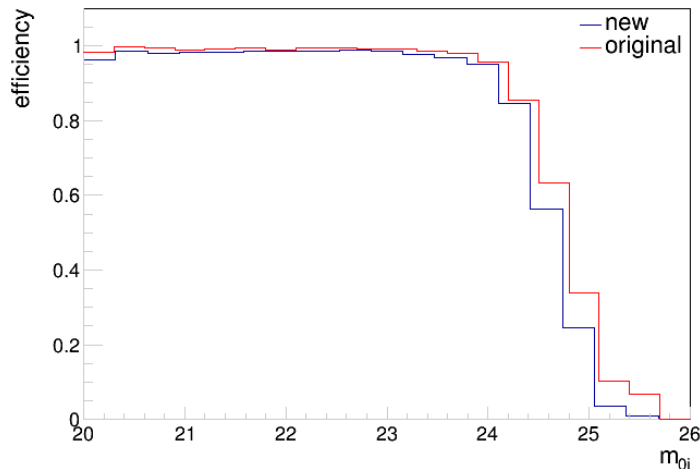


FIGURE 7.24 Efficiency of detection as a function of the generated peak magnitude in  $i_M$ . The new procedure (blue line) is compared to the original one (red line).

#### Coordinate resolution

Coordinate resolution is given by the RMS of the distance between the coordinates at

	Old procedure		New procedure			
			No season stacks		With season stacks	
Stack	coordinate resolution $\pm 0.002$	magnitude bias $\pm 0.0002$	coordinate resolution $\pm 0.002$	magnitude bias $\pm 0.0002$	coordinate resolution $\pm 0.002$	magnitude bias $\pm 0.0002$
1-year	0.709	0.0334	0.698	0.0324	0.698	0.0324
3-year	0.725	0.0349	0.715	0.0340	0.710	0.0335
5-year	0.741	0.0365	0.731	0.0355	0.726	0.0350

TABLE 7.2 Coordinate resolutions (pixels) and corresponding magnitude bias of SNIa detection original, new procedure with no season stacks or complete new procedure with season stacks: for year 3 MC data (1-year stack), adding two additional years of data (3-year stack) and adding 4 additional years of data (5-year stack). Uncertainties shown here are from the statistics of generated SNIa.

generation and at detection. I studied resolutions for the corrected MC for both new and old procedures as can be seen in Table A.1.

The changes on resolution are due to both the cleaning and the choice of season stacks. This can be clearly seen by the improvement on coordinate resolution due to reducing spurious detections with our cleaning procedure (Table A.1, column 2). There is a further gain thanks to the modified detection strategy (Table A.1, column 3).

For one year simulated SNe Ia, the coordinate resolution was slightly improved with the new procedure. The new resolution was found to be  $0.698 \pm 0.002$  pixels to be compared to  $0.709 \pm 0.002$  pixels in the original pipeline.

Since SNLS is a 5-year survey and the goal of this new procedure is to improve detection for the 5-year analysis, I also studied the effect of adding other years of survey data (without simulated SN signal). For this, I constructed catalogues with two or 4 additional years of pure real data.

First, a degradation of coordinate resolution is seen when adding additional data years. This is expected since spurious detections can be in the vicinity of a SNIa detection and they may displace SNe coordinates when being stacked. This degradation is stronger when no season stacks are used.

As expected, for our new procedure a degradation of coordinate resolution is found. However, the new procedure handles better many years of data than the original one. This can be clearly seen in Figure 7.25. For a 5-year detection of transient events the coordinate resolution using the new procedure is  $0.726 \pm 0.002$  which is to be compared with  $0.741 \pm 0.001$  when remaining with the original one. With this procedure, we are

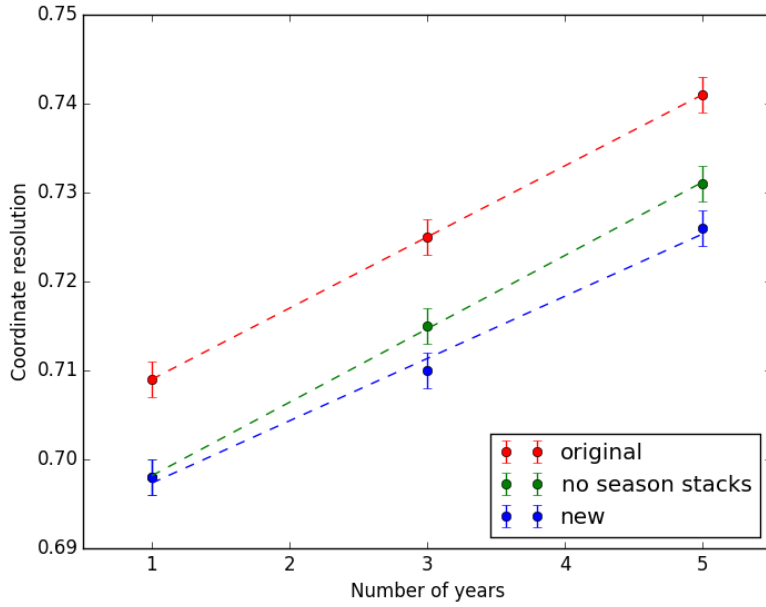


FIGURE 7.25 Coordinate resolution vs. number of stack years.

able to maintain a coordinate resolution for a 5-year processing equivalent to the one found for the 3-year processing by *Bazin et al.* [66] which was the aim in this study.

To check for a possible bias in the coordinate resolution with respect of the redshift, I plotted the distance profile as a function of redshift (Figure 7.26). If the bin with RMS larger than  $0.8\pi ix$  is eliminated, a linear fit provides a flat distribution. However, if points outside the main bulk are eliminated in the fit, there is a slight tendency of higher distance RMS at higher redshifts.

In SNLS, SNIa fluxes are measured in order to obtain luminosity distances. Coordinate precision is crucial to obtain correct fluxes in PSF photometry. If flux and position of a faint object are measured from the same data, position measurement inaccuracy leads to underestimated fluxes because position shifts yield on average smaller PSF fluxes. This result can be found in appendix B of SNLS-3 year sample by *Guy et al.* [64].

Using this appendix, I computed an indicative (upper limit) magnitude bias corresponding to our coordinate resolution. Which is given by:

$$m_{bias} = \frac{d_{RMS}^2}{4\sigma_{seeing}} \quad (7.3)$$

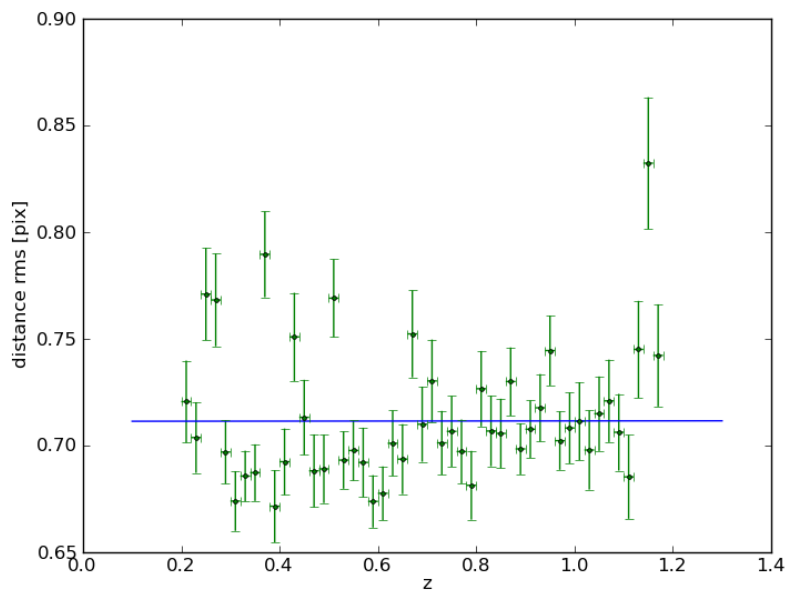


FIGURE 7.26 Profile of the coordinate resolution against redshift. Line fit yielded  $0.0001686x + 0.7114$  when eliminating the bin with RMS larger than 0.8.

where  $d_{RMS}$  is the distance RMS given in pixels,  $\sigma_s = 0.35'' = 1.94 \text{ pixels}$  the RMS of the PSF for the best seeing in SNLS3.

Thanks to the improved coordinate resolution, the new procedure applied on 5-year stacks has similar performance to the old procedure applied on 3-year stacks. The latter was found accurate enough for photometric typing for the SNLS3 in *Bazin et al.* [66].

As a summary, according to MC studies our new procedure efficiently detects transient events while improving coordinate resolution for both 3 and 5 years set of data.

## 7.6 SNLS and beyond...

Morphological component analysis has proven to be a useful approach for cleaning subtracted image stacks such as the ones in the SNLS deferred processing. The choice of algorithm was based on the availability of a robust tool that could decompose our subtracted image stacks efficiently and within our CPU and time resources. For adapting the algorithms I had to use a two-step procedure which was not so clear at the beginning. In the first step several dictionaries and scales were necessary to eliminate the various



artifacts. Note that many of these defects came from the fact that this pipeline uses subtracted image stacks that usually have many residuals. For the second algorithm the goal was to handle non-stationary noise (typical from stacks) in addition to SN-like signals which provided a natural choice of the starlet dictionary for the decomposition. Finally, the choice of algorithm parameters (e.g. number of iterations) was heavily dependent on the efficiency and purity to be achieved and on required computing resources.

Besides improving the subtraction algorithm itself, eliminating artifacts at the level of subtracted images instead of stacks can provide a higher reduction of the number of detections. This could be valid for our deferred pipeline but it should be applied at the beginning of the survey. For implementing such methods, a thorough analysis must be done of the trade-off between gain on signal extraction and removal of artifacts with respect to the high computational and time costs of processing using dictionary decomposition.

Future surveys like LSST may detect around ten thousands SNe Ia a year [90], which is two orders of magnitude higher than in SNLS. Extrapolating what I experienced in the deferred processing of SNLS, the anticipated number of detections in LSST may be as high as  $10^7$  per year which is too large to process. To reduce the number of candidate transient events to process further, cleaning images with a fast multi-resolution method can be of interest. But due to the huge number of detections, additional multi-band and temporal information will be necessary. The above arguments are valid for both real-time and deferred processings, which will both face too large numbers of detections to process. Differences between the two approaches would affect the choice of cleaning algorithms and selection criteria based on multi-band and temporal information.

## 7.7 Summary

I improved the transient event detection procedure using morphological component analysis algorithms based on sparsity. I also reviewed the detection strategy, not only to take into account the output of our cleaning process, but also to reduce the impact of many years of data on coordinate resolution and efficiency.

Morphological component analysis has proven to be a useful approach for cleaning our subtracted image stacks. From my experience, the precise nature of the input images was

---

a key point when choosing and adapting this type of algorithms. The choice of algorithm was based on the availability of a robust tool that could decompose our subtracted image stacks efficiently and within our CPU and time resources.

In summary, this new procedure eliminates a large fraction of artifacts and extracts SN-like objects. It achieves the goal of obtaining a reduction of the number of detections while limiting the loss of SNe Ia.

## Chapter 8

# Classification of SNe Ia using photometric redshifts

In this Chapter I present a new SN Ia classification based on photometric redshifts determined directly from light curves using the algorithm by *Palanque-Delabrouille et al.* [49]. In the SNLS3 pipeline, classification was done using assigned host-galaxy photometric redshifts from an external catalogue. However, such a procedure had an assignment efficiency of 83%. An advantage of this new classification is that SN photometric redshifts can be obtained for all SN Ia events.

Classification, or typing, is performed over the sample of SN-like events obtained after detecting transient events, constructing their light curves and performing SN-like cuts. Classification requires a redshift determination and a SNIa light curve model fit (as in the case of the SNLS3 analysis described in Section § 5.4 hereafter referred to as SNLS3) or a general SN model fit. Using the redshifts and fit variables, it is then possible to disentangle core-collapse and type Ia SNe. In this section we will address core-collapse supernovae as “background” events, although they are an interesting sample by itself, since our goal is to obtain a pure type Ia SN sample.

This chapter is structured as follows. In Section § 8.1, I introduce machine learning techniques, the classification method (Boosted Decision Trees), the TMVA toolkit and some general considerations in Section 8.2. Then, I present in Section 8.3 the possible ways of classifying type Ia SNe in a purely photometric analysis. A first option is presented in Section § 8.4.1 where I review the data classified using the SNLS3 approach

with host-galaxy redshifts and sequential cuts on SALT2 fit variables. Then, I show the feasibility of doing a BDT classification using the same variables as the former approach in Section § 8.4.2. Finally, I implement the BDT SNIa photometric redshift classification in Section § 8.5 using two possible light curve fits.

## 8.1 Machine learning and boosted decision trees

### 8.1.1 Principles

Our goal is to classify different types of SNe and particularly, select type Ia SNe. This can be reduced into a problem of classifying signal and background events. For this, we can take advantage of a subfield of computer science called **machine learning**. The latter encompasses algorithms that can learn from and make predictions on data.

Tools for learning can be classified as supervised and unsupervised, we will center on **supervised learning** for our application. Supervised learning methods usually work in two steps called training and classification. First, known events are fed into the algorithm which learns their characteristics, this is called *training*. Second, unidentified data is provided and the algorithm classifies it based on what it learned during the training phase, this is called *classification*.

**Boosted Decision Trees** (BDTs) are supervised classification methods. They offer many advantages since they are simple to understand and interpret, perform well with large data sets, are well adapted to classify high-dimensional data where boundaries are non-linear and, above all, they can reduce a N-dimensional space of variables into a 1-dimensional space where the variable is the so-called BDT response. An illustration of the advantage of decision trees against a classical approach can be seen for a two-dimensional variable-space in Figure 8.1.

A **decision tree** makes successive rectangular cuts in the parameter space to classify data. It is a process which has many similarities with human decision-making and is very easy to understand. A graphical example of a decision tree can be seen in Figure 8.2. A typical tree begins with a so-called “root-node” which includes all events or data. Binary splits separate the data into subsamples (“leaf nodes”) which at the end of the tree are given a probability of being classified as signal or background. The maximum number of

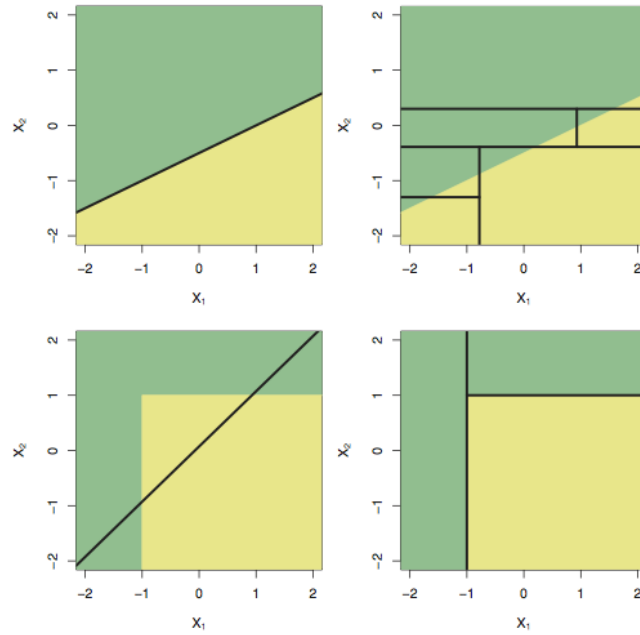


FIGURE 8.1 Top row: a classification example in 2-D where the decision boundary is linear (indicated by color regions). A classical approach is to assume a linear boundary. In this case, such a boundary will outperform a decision tree illustrated on the top right. Bottom row: when the decision boundary is non-linear, a decision tree (bottom right) is more successful than a linear model (bottom left) [91]. For multidimensional cases the boundaries become more complex and decision tree approach is usually more performant.

nodes (or subsamples) is called *maximum depth* and is fixed beforehand. At each split, the algorithm determines the variable  $x_i$  that gives the best separation to discriminate between signal and background (in terms of classification error). The algorithm also determines the cut value  $c_i$  that separates the sample into two complementary subsets  $\{x_i > c_i\}$  and  $\{x_i < c_i\}$ . When splitting data into different “leaf nodes”, to avoid separating data into very small subsamples, a *minimum node size* can be set.

If the tree is too large it will make very specific cuts for each case and might not generalize well, this is called *overtraining*. To avoid overtraining, it is important to separate the training sample (which has known types) in two parts, one for the actual training of the algorithm, and the second to test that the generalization is valid in an independent sample.

When using a simple decision tree, some events may be misclassified. To reduce the number of these events and to obtain better results, we can use a technique called **boosting**. The idea is to combine “weak” classifiers, in this case trees, to get better results. First, a simple decision tree is considered which misclassifies some events. The

events are reweighted to put more emphasis on poorly classified events. Then, with the reweighted data, the training is redone. This is an iterative procedure where the final classification is averaged over all decision trees. There are many boosting techniques, but I centered on AdaBoost which is a largely used boosting method.

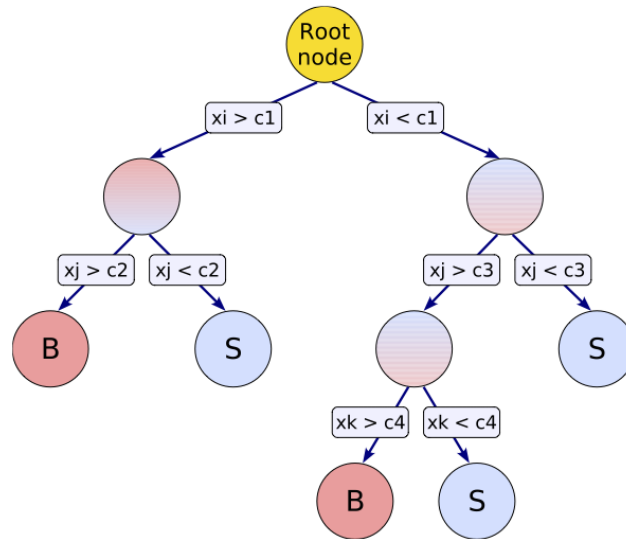


FIGURE 8.2 Graphic representation of a simple decision tree. The root node contains an event with all its information and is our starting point. A sequence of binary splits using variables  $x_i, x_j, x_k$  is applied to data. At each split, the variable that gives the best separation is used to discriminate between signal and background. The end nodes are labelled "S" and "B" for signal and background classification [92].

In order to train BDTs, one has to choose the algorithm's **hyperparameters**. These include the *maximum depth* of a tree, the *number of trees* used for boosting and the *minimum size of a node*. Choices can be done using overtraining tests and applying classification to independent known samples and studying their statistical evolution. Moreover, data used for training (simulations) and application (data) must be preprocessed to eliminate data overflows that may yield missclassifications.

Supervised learning methods can be implemented using different toolkits such as Python's scikitlearn and ROOT's TMVA. Although I performed tests with both, my classification is based on the TMVA toolkit. Possible extensions of this work include exploring other packages as scikitlearn which can mix several supervised learning methods and can optimize hyperparameters in an automatic way.

Further information about multivariate analyses can be found in *The elements of statistical learning* by T.Hastie, R. Tibshirani and J.Friedman [93] and *Introduction to statistical learning* by G. James, D. Witten, T. Hastie, R. Tibshirani [91].

### 8.1.2 TMVA

The Toolkit for Multivariate Analysis (TMVA) provides an environment for the application of multivariate classification [92]. It is based on ROOT and written in C++ which provides fast processing. In this toolkit, supervised learning algorithms such as BDTs are available.

For our BDT analysis, we have the same structure as any supervised learning method. The two-step procedure is implemented as follows:

- Training: signal and background samples are given to the multivariate methods. The samples can be split randomly in half to perform training and crosschecks. An automatic utility provides plots such as: background rejection vs. signal efficiency and classifier output distributions for both training and testing samples which are part of an overtraining check.
- Application or classification: the trained method is used for the classification of data with unknown signal and background decomposition. The output consists of a BDT response variable which takes a value between minus one and one. If the value is close to minus one, it is more likely to be a background event and if it is closer to one it has higher probability of being a signal event. A cut on this variable can later be set to obtain the sample of selected events.

All results in this work were done with ROOT version v5.34.23 and TMVA Version 4.2.0.

## 8.2 Considerations for our application

We have introduced supervised learning and the classification utility TMVA. We will now set-up the generalities for our analyses. Results will be presented in Section 8.3.

Starting from a SN sample, our goal is to obtain a pure SN Ia sample. We consider that our SN-like sample contains only core-collapse events and type Ia SNe. We must

now recall that SNIa observed light curves vary depending of the object’s redshift (e.g. a low redshift core-collapse SN can resemble a SNIa at high redshift). Therefore we are interested only on type Ia SNe which have a correct redshift assignment and consider core-collapse events and type Ia SNe with bad redshift as background.

In our different analyses, SNe Ia with bad redshift are due to incorrect SN-galaxy associations or wrong host-galaxy redshift in the catalogue or incorrect SN photometric redshift (the last two are considered as catastrophic redshifts).

The following samples were used to set up the analyses and to assess their performance:

- **Light curves simulations:** provide large and known samples of core-collapse SNe and SNe Ia. We used the same simulation as the one used for determining SNLS3 global efficiency in Figure 5.10. This simulation is independent of the one used in Chapter § 7 for transient event detection efficiency. It consists of simulated light curves for both type Ia and core-collapse SNe. Synthetic SN Ia light curves were produced assuming a flat  $\Lambda$ CDM cosmology using SALT2. Core-collapse light curves were simulated using the general SN fit introduced in Equation 5.1 and spectroscopic core-collapse events for modeling parameters. The core-collapse sample was divided in “type II” for those long lasting events as type II-P and “type Ibc” for short lasting events. Further details can be found in *Bazin et al.* [66].

In order to obtain an equivalent sample as the SN-like sample in data, simulated events were introduced into the pipeline, after the detection step and submitted to SN-like cuts. The number of events used for our training and testing is given as “selected” events in Table 8.1.

- **The SNLS3 sample:** contains subsamples of known spectroscopic and photometric Ia and core-collapse SNe. The photometric samples of core-collapse and type Ia SNe were determined in dedicated analyses described in *Bazin et al.* in [35] and [66] respectively. In this work SNLS3 refers to the SNIa SNLS3 photometric analysis.



### 8.2.1 Samples used in the BDT set-up and analysis

A BDT analysis performance can be evaluated when classifying a known sample. However, the BDT algorithm needs first to be trained with an independent known sample of signal and background events. In this subsection I present the different training and classification sample options which are summarized in Figure 8.4.

- **Training sample:** two independent simulation samples were randomly created, half of the simulation was used for training and the other half for testing. Before performing the classification, the distributions of the SNIa signal and core-collapse background can be seen in Figure 8.3 for half simulation sample. Each training sample was divided in “signal” and “background” events as following:
  - Signal: type Ia SN light curves with good photometric redshift assignment (either host-galaxy or SN) when compared to generated redshift  $\Delta z/(1+z) < 0.1$ .
  - Background: type IIp and Ibc SNe (core-collapse events) as well as type Ia SNe with bad redshift assignment when compared to generated redshift  $\Delta z/(1+z) > 0.15$ . The latter encompasses the so-called catastrophic redshifts.

To separate those type Ia SNe with good and bad redshift assignments, I decided to perform cuts that clearly separate good and bad redshifts. I considered those SNe Ia with redshift determination  $0.1 < \Delta z/(1+z) < 0.15$  as not clearly belonging to either sample.

- **Classification sample:** the data to be classified. For the final application of this work it will be the SNLS5 data, but for tests performed during my PhD I used two samples:
  - Half of the simulation: half of the simulation used for purity, efficiency and contamination estimations.

- The SNLS3 sample: It is a limited sample of spectroscopic and photometric SNe Ia and core-collapse but is useful for comparing the SNLS3 original analysis with BDT based ones and is very sensitive to BDT hyperparameter changes. SNe Ia with “bad redshift” are those where the assigned redshift (host-galaxy or SN) differs from the spectroscopic redshift as  $\Delta z/(1+z) > 0.15$ .

sample	SNe Ia total	SNe IIp	SNe Ibc
original	20,000	20,000	20,000
selected	10,522	2,321	2,750

TABLE 8.1 Number of simulated events that were selected as SN-like candidates. Note that the detection step is not included here. My goal is to train my BDT to classify events that can be present in the SN-like sample independently of the detection efficiency. In this way my training is more sensitive to type Ia faint events and I can also profit of having a larger sample. Requiring detection reduces the number of simulated events by an extra 10% for each sample. The small fraction of core-collapse SNe in the selected sample is explained by the faintness of these events which are simulated following the volumetric distribution of the survey (selection requires a significant flux variation).

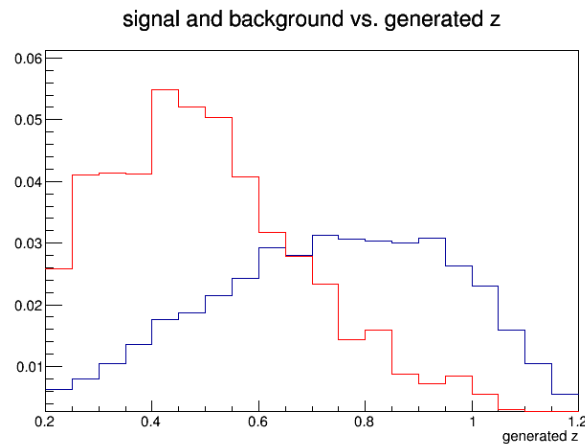


FIGURE 8.3 Distribution of the SNIa signal (blue line) and core-collapse background (red line) events after SN-like and quality cuts (as defined in Section § 5.4.2) as a function of generated redshift ( $gz$ ) for half of the simulation. Not normalized.

For all these samples, summarized in Figure 8.4, light curve sampling requirements and quality cuts, as presented in Subsection § 5.4.2, were applied for all analyses (host-galaxy or photometric redshift).

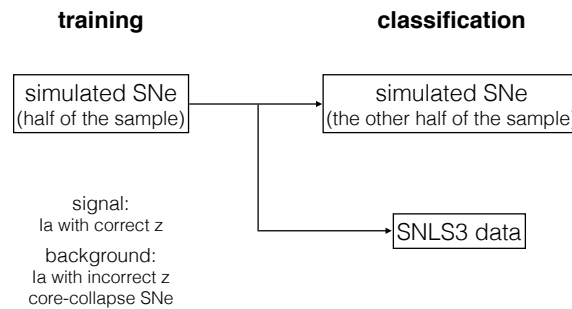


FIGURE 8.4 Samples involved in BDT training and classification

## 8.2.2 BDT analysis set-up

Now that the different sample options for our BDT analyses have been presented, we can address the set-up of a BDT analysis. We need to select hyperparameters and variables that will be used in a classification, as well determining our BDT response threshold for classification. For this, we vary hyperparameters or input variables, train the algorithm and assess the impact on the classifications of known samples such as the ones previously presented and summarized in Figure 8.4.

### 8.2.2.1 Finding the appropriate hyperparameters

I varied each hyperparameter in a broad range of possible values and determined their impact in our classification sample. I chose those parameters that had a better performance.

For the simulation based tests, we need a metric that judges the prediction of our binary problem (events are either signal or background). Such a metric is the **AUC metric** which is commonly used as an evaluation method for dichotomic classifications. AUC stands for Area Under Curve, where the curve is the **ROC curve** (Receiver Operating Characteristic). The ROC curve illustrates the performance of a binary classifier by plotting the true positive rate (signal efficiency) against the false positive rate (background efficiency) as in Figure 8.5. While the ROC curve represents the performance

of a model in two-dimensions, the AUC simplifies this into an scalar. A perfect model would score an AUC of 1.

It must be noted that the choice of hyperparameters must come from the best AUC score available when the algorithm is not overtrained. To verify that no overtraining is done, TMVA provides an automatic plot that divides half of the training sample and uses half for training and half for testing. If both distributions coincide, there is no overtraining. When the training and testing distributions do not coincide, like in Figure 8.6, there is overtraining, lack of classifier generalization.

The SNLS3 sample also proved to be useful to verify the tendency on statistics when varying hyperparameters. However it is a restricted sample and must be used with caution.

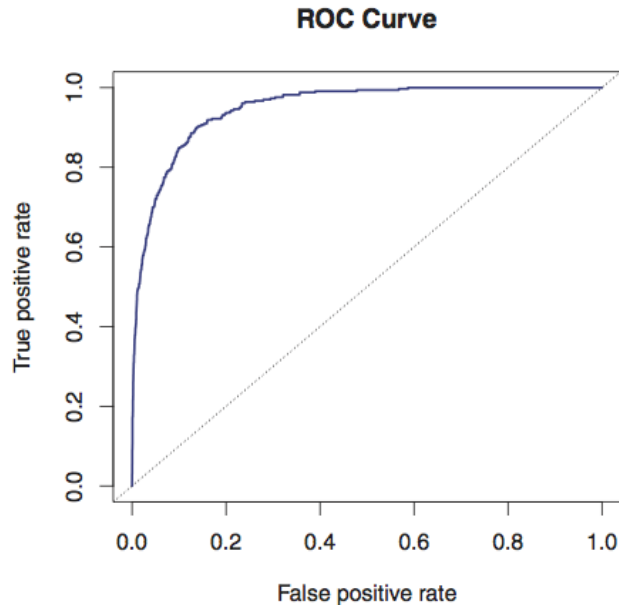


FIGURE 8.5 In blue, an example of ROC curve (true positive rate against false positive rate). The dotted line represents random guess, good classifiers are on the left side of this line [91].

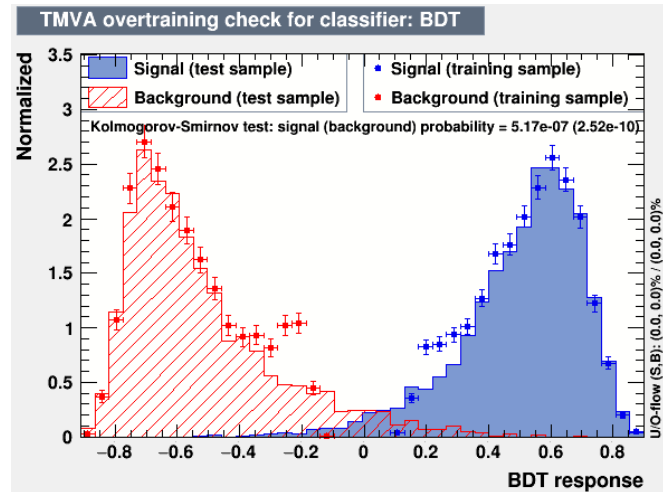


FIGURE 8.6 Overtraining test of TMVA. Signal (background) events are in blue (red): points are training sample and shadowed regions test samples. In this case we have overtraining since both distributions do not coincide.

### 8.2.2.2 Choosing a set of classification variables

We must take into account that not all available variables are useful for our classification. An advantage of the TMVA implementation is that it provides a ranking of the most discriminant variables for a given set of hyperparameters. An example of such a ranking for the host-galaxy photometric redshift analysis is seen in Figure 8.7. For those variables poorly ranked we can consider eliminating them and observing the improvement, or not, of our classifier.

```

: Ranking input variables (method specific)...
: Ranking result (top variable is best ranked)
:
:-----
: Rank : Variable      : Variable Importance
:-----
:      1 : zgal           : 1.423e-01
:      2 : cgal           : 1.368e-01
:      3 : mzgal          : 1.216e-01
:      4 : x1gal          : 1.207e-01
:      5 : czgal          : 7.625e-02
:      6 : migal          : 7.363e-02
:      7 : ctgal          : 5.984e-02
:      8 : cggal          : 5.069e-02
:      9 : crgal          : 4.987e-02
:     10 : cigal          : 4.795e-02
:     11 : mrgal-mzgal    : 4.078e-02
:     12 : mggal-migal    : 4.044e-02
:     13 : mrgal          : 3.910e-02
:-----

```

FIGURE 8.7 Example of variable ranking in TMVA for the host-galaxy redshift analysis.

### 8.2.2.3 Choice of the BDT response threshold

Finally, once the choice of hyperparameters and variables is done, we can classify our SN events in different types. Our N-dimensional problem of **classifying SNe is now reduced to a 1-dimensional problem, represented by the BDT response**. The choice of a BDT threshold was done taking into account mainly the global SNIa efficiency and the purity of the SNIa sample. Two plots were specially useful for choosing a BDT response cut: one giving the global efficiency (including detection and classification) versus purity of the SNIa sample at different BDT cut values and one providing the ratio contamination-classification efficiency as a function of the BDT response.

My main indicator is the **efficiency-purity plot** since our goal is to obtain a large SNIa sample with good purity. An example of such a plot can be seen in Figure 8.8. The second plot I mentioned (shown in Figure 8.9), is very useful to find a cut when no complete separation of signal and background can be obtained. However it must be read with care when that is not the case. In our case, such a separation is obtained which is given by a zero value on the ratio of contamination-classification efficiency. Before the complete separation (at low BDT response values) there is a slow decrease for the ratio as a function of the BDT response which shows the approach toward the tail of the contamination distribution. The ratio tends to infinity for BDT response values corresponding to no signal left.

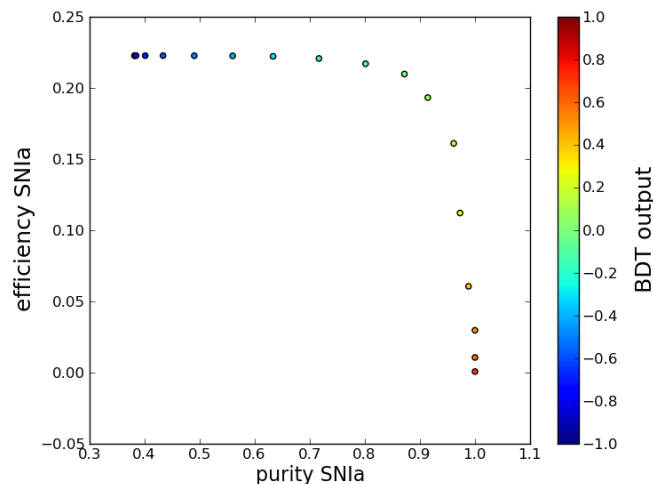


FIGURE 8.8 Example of a global efficiency versus purity for different choices of the BDT response cut.

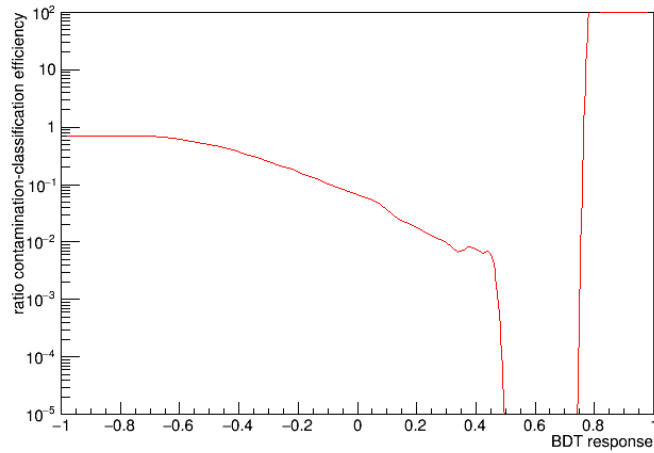


FIGURE 8.9 Example of a contamination-classification efficiency ratio versus BDT response plot. This plot would be particularly useful if no complete separation of signal and background would be possible. Here the zeros in the ratio represent the complete signal-background separation. However such a separation has as a price the loss of efficiency which can be clearly seen in Figure 8.8.

The exact definition of efficiency and purities I used in this work are summarized in the following.

### SN rates and proportions

For computing purity, rates for both type Ia and core-collapse SNe must be taken into account. In Section § 3.2.4 we mentioned that the SNIa rate can be expressed as a function of  $(1+z)^\alpha$  where  $\alpha = 2$  [34]. Core-collapse SNe are expected to reflect the SFR and therefore should have  $\alpha = 3.6$  [35]. In Table 8.2, measured values from SNLS data for SN volumetric rates at different measured redshifts  $z_m$  can be found.

SN type	$\langle z_m \rangle$	$r_V [10^{-4} h_{70}^{-1} \text{yr}^{-1} \text{Mpc}^{-3}]$	$\alpha$
Ia	0.47	0.42	2
CC	0.3	1.63	3.6

TABLE 8.2 SN volumetric rates derived from SNLS data from *Bazin et al.* [35] and *Neill et al.* [38].  $r_V$  is given for the  $z_m$  in each case.

Although all core-collapse SNe are given the same rate, the proportion between short- and long-term events differ. Long-term events as type IIp SNe (in this work called IIp) have a contribution of 30% while short-term events, type Ibc and those type II not plateau (in this work called Ibc) dominate with 70% as seen in SNLS data [35].

### Purity and efficiency

The **purity** of our classified SNIa sample (or the degree of contamination by core-collapse events mostly) can be estimated using simulated light curve events.

First, let's define a *normalized number of events* for any type of SNe which take into account the SN rate, the survey volumetric distribution and the initial distribution of simulated events:

$$N = \sum_{z \text{ bins}} r_V(z_m) \frac{(1+z)^\alpha}{(1+\langle z_m \rangle)^\alpha} \frac{\Delta T}{1+z} V(z) \frac{n(z)}{N_0(z)}, \quad (8.1)$$

where  $\frac{(1+z)^\alpha}{(1+\langle z_m \rangle)^\alpha}$  converts the rate at the redshift of the current bin,  $\Delta T$  accounts for the observation period and  $1/(1+z)$  converts observation time into rest-frame time,  $V(z)$  is the volumetric distribution as shown in 7.5.2,  $n(z)$  is the distribution of events that are classified,  $N_0(z)$  is the distribution of simulated events before any cut. The latter includes the volumetric rate, therefore we can simplify the previous equation into

$$N = \sum_{z \text{ bins}} r_V(z_m) \frac{(1+z)^\alpha}{(1+\langle z_m \rangle)^\alpha} \frac{\Delta T}{1+z} \frac{n(z)}{N_0}, \quad (8.2)$$

where  $N_0$  is the number of simulated events as given in Table 8.1 in the “original” row. Finally, the purity of our SN Ia sample is given by:

$$P_{Ia} = \frac{N_{Ia}^{true}}{N_{Ia}^{true} + N_{Ia}^{bad\ z} + N_{Ibc} + N_{IIp}} = \frac{N_{Ia}^{true}}{N_{total}}. \quad (8.3)$$

The same procedure can be applied to obtain contamination from each core-collapse species and from type Ia SNe with bad redshift assignment.

In this work we will cite “**global efficiency**” as the efficiency of detection, pre-selection and classification. Global efficiency can be expressed as a function of the redshift  $z$  as:

$$\epsilon_{Ia} = \frac{\sum_{z \text{ bins}} r_V(z_m) \frac{(1+z)^\alpha}{(1+\langle z_m \rangle)^\alpha} \frac{\Delta T}{1+z} \frac{n_{Ia}(z)}{N_{0Ia}(z)}}{\sum_{z \text{ bins}} r_V(z_m) \frac{(1+z)^\alpha}{(1+\langle z_m \rangle)^\alpha} \frac{\Delta T}{1+z}}, \quad (8.4)$$

where  $n_{Ia}(z)$  are all the SNe Ia which are classified after being detected and pre-selected and  $N_{0Ia}(z)$  are the SNe Ia which are originally simulated.

For both efficiency and purity we assumed binomial errors on  $n(z)$  and propagated those errors in purity and efficiency formulas 8.4 and 8.3.



Now that I have introduced the technical details of a classification using BDTs we can move towards a complete classification analysis.

### 8.3 Classification of SNe Ia using photometric redshifts

In order to classify SNe Ia we require: a redshift determination, a light curve model fit and a strategy for disentangling core-collapse and type Ia SNe. In the following I will present different possible classifications depending on the choice of these elements.

First, as we are in a purely photometric pipeline we require **photometric redshifts**. Two types of photometric redshifts are used in this work:

- Host-galaxy photometric redshift (*zgal*): selected SN-like events are matched with the photometric redshift catalogue introduced in Section 5.4.1.
- SN photometric redshift (*zpho*): redshifts can be estimated directly from SN light curves using the algorithm by *Palanque-Delabrouille et al.* [49]. This procedure uses the SALT2 light curve fitter to determine the redshift of Type Ia supernovae. For the SNLS3 sample they found an average precision of  $\sigma_{\Delta z/1+z} = 0.022$  up to redshift of one and catastrophic errors under 1.4% for the SNLS3 sample and only 0.4% when restricting the test sample to spectroscopically confirmed Type Ia supernovae.

It must be highlighted that SNIa classification using SN photometric redshifts is an unprecedented work.

Second, we require a **light curve model fit** to extract properties of our SNe. This fit can be a general SN fit or a type Ia SN fit.

- SALT2: introduced in Section 3.3, SALT2 is a type Ia SN light curve fitter. It requires a redshift input in order to extract SNIa intrinsic properties. It provides magnitudes and reduced  $\chi^2$ 's for each fitted filter, as well as the total fit  $\chi^2$  and the SN color and stretch.
- General SN fit: it fits a SN phenomenological shape to the provided light curve. The shape function was previously presented in Equation 5.1.

Third, once we extracted variables that characterize our light curves we can **classify SNe**. For this, we must set up a strategy such as:

- Sequential cuts: this is a set of complementary cuts on the light curve parameters. They are defined by studying the variable distributions and magnitude-color diagrams for spectroscopic data and simulations.
- BDT classification: such a machine learning method can be used to reduce our N-dimensional problem (N=number of variables that characterize the curve) into a 1-dimensional problem (choice of a BDT response cut).

My final goal is to present a new SN Ia classification based on photometric redshifts determined directly from light curves using the algorithm by *Palanque-Delabrouille et al.* However, I will show different possible analyses in order to check consistency and compare results. A summary of the presented analyses can be seen in Table 8.3.

analysis	redshift	fitter	classification
<a href="#">8.4.1</a>	host-galaxy (zgal)	SALT2	sequential cuts
<a href="#">8.4.2</a>	host-galaxy (zgal)	SALT2	BDT
<a href="#">8.5.1.1</a>	SN (zpho)	SALT2	sequential cuts
<a href="#">8.5.1.2</a>	SN (zpho)	SALT2	BDT
<a href="#">8.5.2</a>	SN (zpho)	general	BDT

TABLE 8.3 Classification analyses presented in this work.

## 8.4 Host-galaxy photometric redshift analysis (zgal)

Host-galaxy redshifts have been previously used to classify type Ia SNe successfully. In Section § 5.4 I described how classification was performed using this redshift, SALT2 and sequential cuts. In this section I reproduce these results and introduce a different strategy, BDT, for a host-galaxy and SALT2 analysis.

Before classifying, as in the SNLS3 photometric analysis, we rejected in the classification sample SALT2 fits that failed and applied quality and  $\chi^2$  requirements.

### 8.4.1 zgal + SALT2 + sequential cuts: SNLS3

For consistency, I reprocessed statistics for the SNLS3 analysis. All results are in agreement with published results [66] with the exception of the core-collapse contamination. This triggered a reevaluation of results and we found a divergence due to an error in the normalization procedure. Once corrected, the core-collapse contamination was reevaluated to be  $5 \pm 1\%$  for the original analysis. Results for my computation are shown in Table 8.4. The residual difference may be attributed to host-extinction included in the original analysis (to obtain the observed rate), while my computation takes the rate corrected for extinction.

Simulation			
purity SNe Ia	contamination		efficiency SNIa
	bad redshift SNe Ia	core-collapse	
$94.4 \pm 0.5\%$	$0.65 \pm 0.08\%$	$4.9 \pm 0.5\%$	$29.9 \pm 0.3\%$

SNLS3 data			
# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
486	175	0	0

TABLE 8.4 SNLS3 analysis (zgal + SALT2 + sequential cuts): statistics for SNLS3 data and purities, efficiencies from simulation. The indicated SNIa efficiency is the global efficiency and includes weights and efficiency of host-galaxy assignment. The core-collapse photometric sample was determined by [35].

The purity and contamination distributions as a function of host-galaxy redshift can be seen in Figure 8.10. The contamination of SNe Ia with bad redshift assignment increases slightly with redshift but remains small (below 0.5%) whatever the redshift. The core-collapse contamination is dominated by non-plateau events and decreases with redshift. The contamination statistics is very low and therefore we can't draw any conclusion about the trends, those results are consistent with a flat distribution (within error bars).

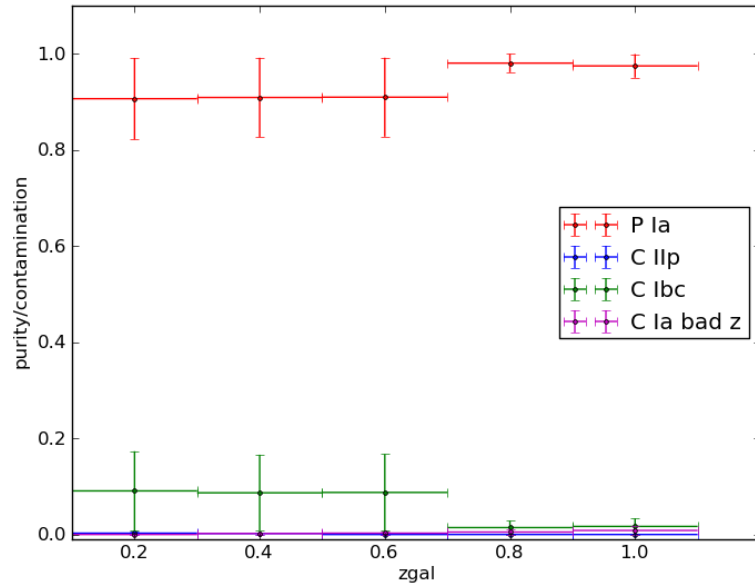


FIGURE 8.10 SNLS3 analysis ( $z_{\text{gal}}$  + SALT2 + sequential cuts): purity of classified SNIa sample (red) and contamination (other colors) as a function of host-galaxy redshift ( $z_{\text{gal}}$ ). Weights are applied for volumetric distribution and SN rates.

The global efficiency as a function of generated redshift is shown in Figure 8.11. This efficiency takes into account the whole pipeline, including detection, redshift assignments and classification. It is fully consistent with what was published in [66].

#### 8.4.2 $z_{\text{gal}}$ + SALT2 + BDT analysis

In this section I will introduce a host-galaxy redshift classification using BDTs and based on SALT2 fit variables. This analysis will serve as an introduction of a BDT analysis and will also validate its implementation when comparing it to the SNLS3 sequential cuts analysis just described.

For this analysis, data were selected using the previously mentioned quality criteria.

First, we need to set-up the BDT classification. The first step is to choose hyperparameters such as maximum depth, number of trees and minimum node size.

Simulation tests provided different AUC scores when varying one by one the parameters, as shown in Figure 8.12. However, these plots are only indicative since overtraining indicators are not shown. The best choice of hyperparameters comes from the best

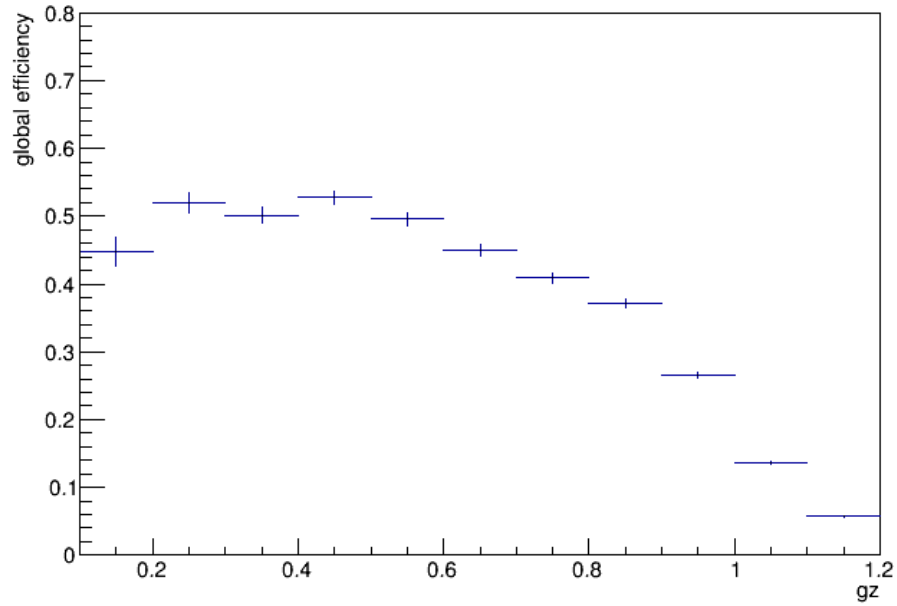


FIGURE 8.11 SNLS3 analysis (zgal + SALT2 + sequential cuts): global SNIa efficiency (including detection and classification) as a function of generated redshift (gz).

AUC score that can be obtained without overtraining. For this I regularly checked overtraining plots.

Although the AUC score increases for large number of trees (Figure 8.12), when the number of trees was high (above 100), the classification was overtrained. 50 trees was a good choice since it avoids overtraining but provides a good classification. The minimum node size was indicated as percentage of the sample that must pass through a node to be valid, I varied it between 0.5% and 1%, being 0.5% my final choice. The maximum depth with the highest performance without overtraining was found to be 5. Overtraining control plots for different hyperparameter choices can be seen in Figure 8.13.

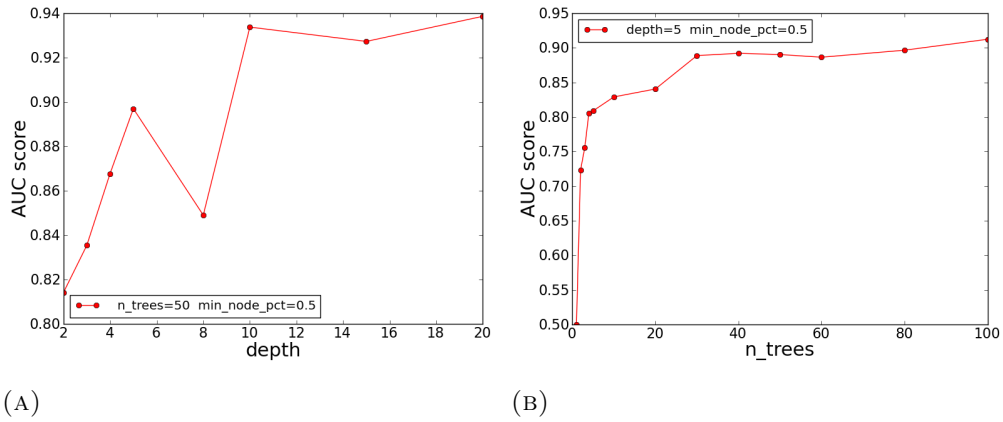


FIGURE 8.12 *zgal* + SALT2 + BDT analysis: AUC score against different choices of BDT hyperparameters. For subfigure 8.12a, above a maximum depth of 10 overtraining occurred. For subfigure 8.12b, overtraining occurred close to 100 trees. From these plots and overtraining tests, I chose a maximum depth of 5 and 50 trees for the BDT host-galaxy analysis.

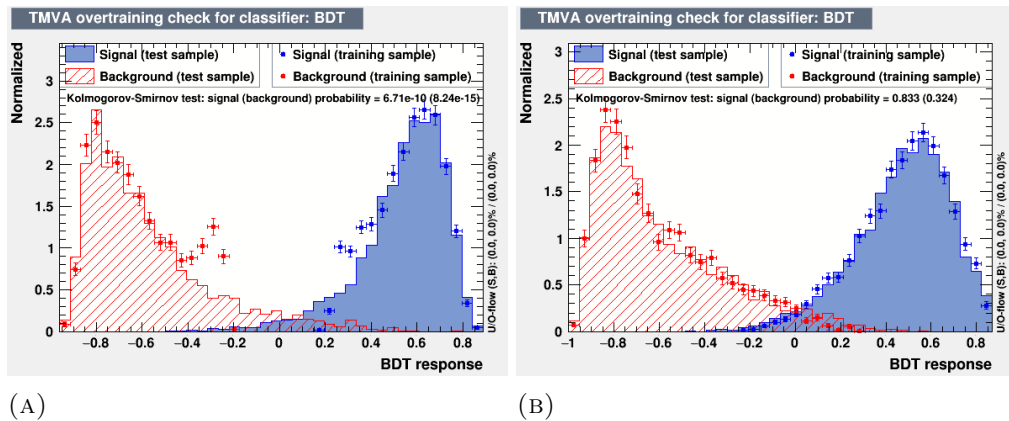


FIGURE 8.13 *zgal* + SALT2 + BDT analysis: Overtraining tests for different hyperparameters choices (see Section 8.1). Taking fixed minimum node size 0.5% but varying number of trees and maximum depth. In 8.13a overtraining is seen for  $n_{trees} = 100$  and  $max\ depth = 10$ . In 8.13b the chosen parameters  $n_{trees} = 50$  and  $max\ depth = 5$  show no overtraining.

The second step is to choose those variables that can be useful for classifying SNe. SALT2 provides magnitudes and reduced  $\chi^2$  for each filter, the total fit  $\chi^2$ , color and stretch. We used these variables together with the host-galaxy redshift and, inspired by the SNLS3 analysis, added “colors” which are given by the magnitude differences on filters ( $g - i$ ), ( $r - z$ ) and ( $i - z$ ). When testing the algorithm the color ( $i - z$ ) did not provide a good discrimination and was eliminated from the final analysis. Another variable that was poorly ranked was the magnitude in the  $g_M$  filter which I also eliminated from the analysis. This can be seen in Figure 8.14.

: Ranking input variables (method specific)...			: Ranking input variables (method specific)...		
: Ranking result (top variable is best ranked)			: Ranking result (top variable is best ranked)		
: Rank	: Variable	: Variable Importance	: Rank	: Variable	: Variable Importance
: 1	: zgal	: 1.222e-01	: 1	: zgal	: 1.406e-01
: 2	: mzgl	: 1.009e-01	: 2	: mzgl	: 1.197e-01
: 3	: x1gal	: 9.461e-02	: 3	: cgal	: 1.051e-01
: 4	: cgal	: 9.031e-02	: 4	: x1gal	: 1.050e-01
: 5	: migal	: 8.125e-02	: 5	: migal	: 7.468e-02
: 6	: ctgal	: 6.393e-02	: 6	: czgal	: 6.424e-02
: 7	: czgal	: 5.983e-02	: 7	: ctgal	: 6.316e-02
: 8	: crgal	: 5.701e-02	: 8	: cigal	: 6.087e-02
: 9	: mrgal	: 5.358e-02	: 9	: crgal	: 4.667e-02
: 10	: cigal	: 5.301e-02	: 10	: mrgal-mzgl	: 4.637e-02
: 11	: mrgal-mzgl	: 4.660e-02	: 11	: cggal	: 4.626e-02
: 12	: cggal	: 4.551e-02	: 12	: mrgal	: 4.557e-02
: 13	: mrgal-mzgl	: 4.460e-02	: 13	: mrgal-mzgl	: 4.239e-02
: 14	: migal-mzgl	: 4.384e-02	: 14	: migal-mzgl	: 3.929e-02
: 15	: mrgal	: 4.273e-02			

(A)

(B)

FIGURE 8.14 *zgal* + SALT2 + BDT analysis: ranking of variables for chosen hyperparameters  $min\ node\ size = 0.5\%$ ,  $ntrees = 50$  and  $max\ depth = 5$  (see Section 8.1). “gal” stands for host-galaxy redshift analysis and “zgal” host-galaxy redshift. SALT2 parameters are given by: “cgal” color, “x1gal” stretch, “m(i,g,r,z)” magnitudes and “c(i,g,r,z)” chi squared in  $i_M, g_M, r_M, z_M$  filters respectively. In 8.14a the whole set of SALT2 variables plus redshift is shown. In 8.14b the ranking is seen for all variables except the  $g_M$  magnitude.

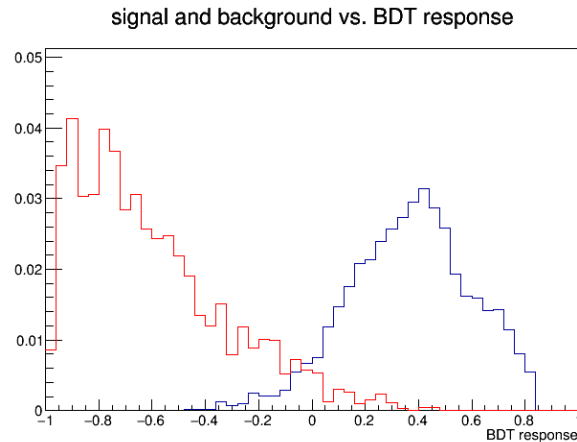
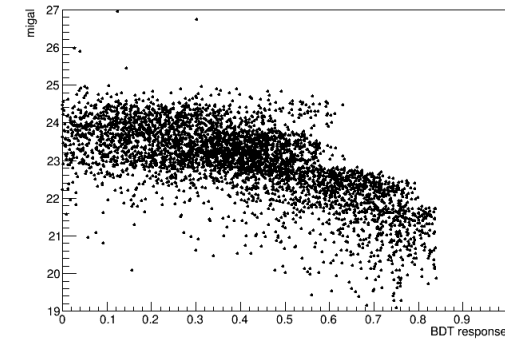


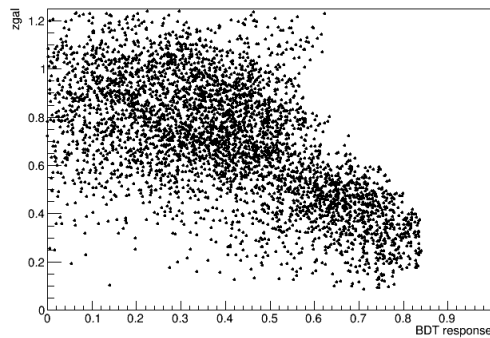
FIGURE 8.15 *zgal* + SALT2 + BDT analysis: signal ( blue line ) and background ( red line ) simulation distributions with respect to the BDT response. A complete separation of the two samples is possible for BDT response  $> 0.5$  but with a poor efficiency. Distributions are not normalized.

The classification was performed with the above set of hyperparameters and variables. The problem was reduced to the choice of a BDT response threshold. The simulation test-sample allows to verify that a separation between signal and background can be obtained when cutting on the BDT response. Figure 8.15 shows the BDT response distribution for signal and background events. A clear discrimination can be seen.

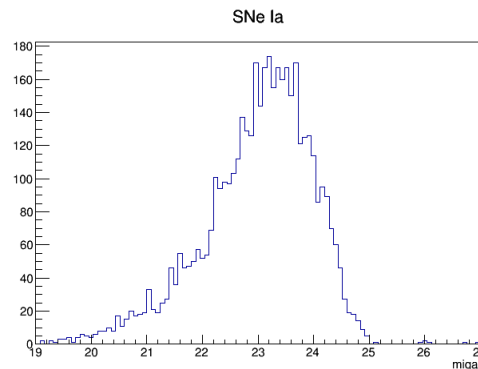
Closely observing Figure 8.15 we see that the signal distribution has a small shoulder at high BDT response. These events are low-redshift ones but with not very high magnitudes in the  $i_M$  filter (therefore bright events) within the bulk of the magnitude distribution for SNe Ia as shown in 8.16.



(A)



(B)



(C)

FIGURE 8.16  $z_{gal}$  + SALT2 + BDT analysis: study of the signal shoulder in Figure 8.15. For  $BDT\ response > .6$  events are shown to be bright in the  $i_M$  filter 8.16a and at close host-galaxy photometric redshift 8.16b.  $i_M$  magnitudes of these events are in the tail of the magnitude distribution for SNIa events 8.16c.

The choice of a BDT threshold was done taking into account the global SNIa efficiency and the purity of the SNIa sample using Figure 8.17. A compromise between the desired



SN Ia purity and the global efficiency had to be taken, the natural place to search for this was near the inflection point between good efficiency and high purity. My choice was to keep events above 0.1 BDT response.

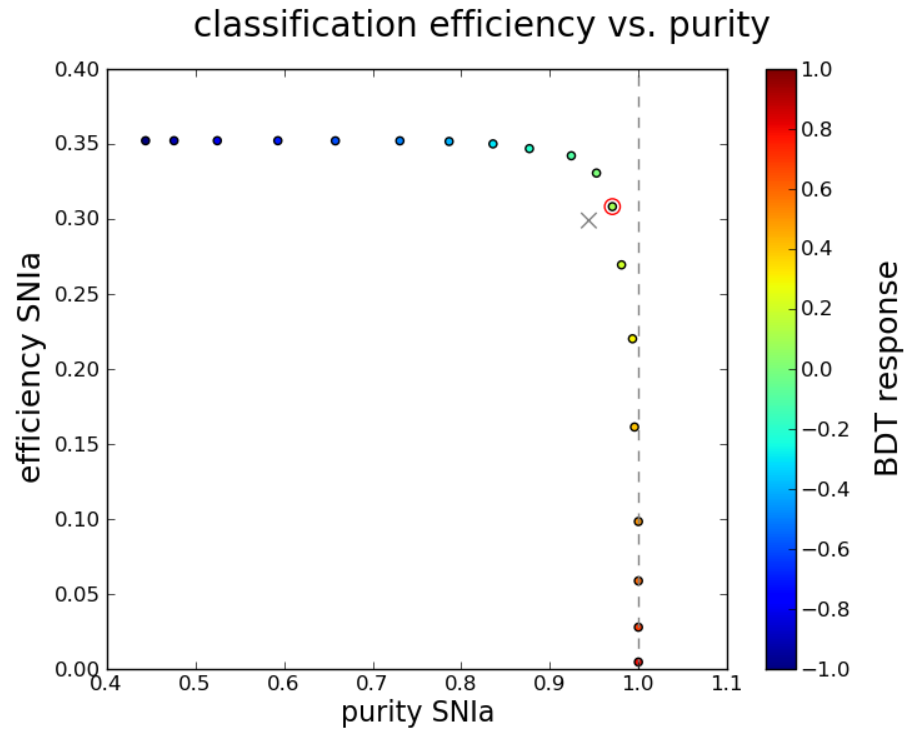


FIGURE 8.17 *zgal* + SALT2 + BDT analysis: Global efficiency of SN Ia versus global purity of SN Ia classified sample. Colors indicate the value of the BDT response. The red circle shows the chosen BDT cut value (0.1). A dashed line indicates SN Ia purity of 1. The gray cross indicates the efficiency and purity values for the SNLS3 sequential cuts analysis.

Cutting at BDT response  $> 0.1$  I obtained the simulation and SNLS3 sample statistics indicated in Table 8.5. The type Ia sample purity and the global efficiency are slightly higher than the ones in the SNLS3 analysis.

Since the global efficiency of this analysis remains close to the one of the sequential cuts SNLS3 analysis, it is surprising that the number of the photometrically selected events is increased by  $\approx 15\%$ . This fact is still under analysis.

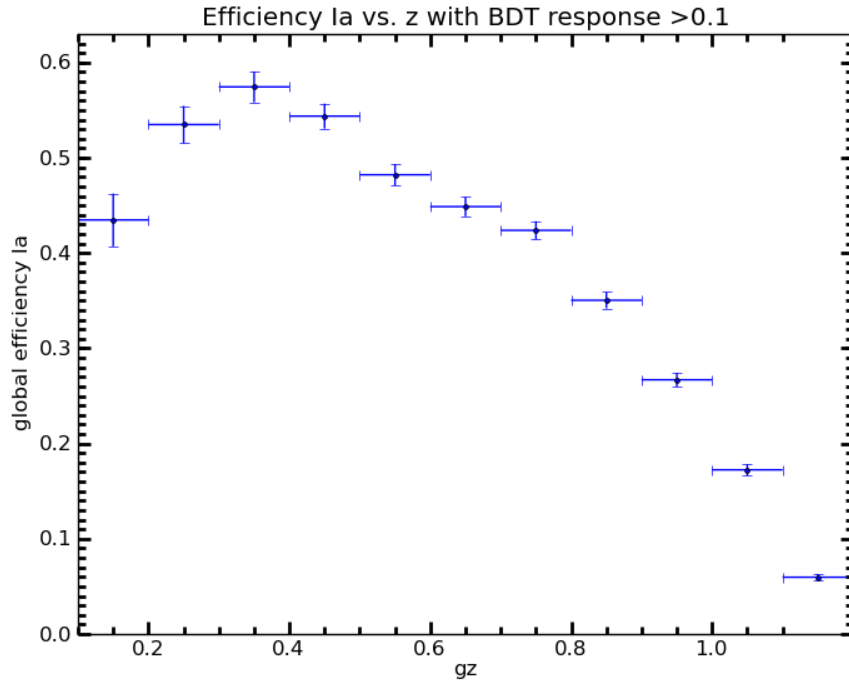


FIGURE 8.18  $z_{\text{gal}} + \text{SALT2} + \text{BDT}$  analysis: Global efficiency as a function of generated redshift ( $gz$ ) for BDT cut 0.1.

#### Simulation

purity SNe Ia	contamination		efficiency SNIa
	bad redshift SNe Ia	core-collapse	
$96.6 \pm 0.6\%$	$0.4 \pm 0.1\%$	$2.9 \pm 0.6\%$	$30.8 \pm 0.8\%$

#### SNLS3 data

# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
524	175	0	0

TABLE 8.5  $z_{\text{gal}} + \text{SALT2} + \text{BDT}$  analysis: Statistics for SNLS3 data and simulation purities for a BDT classification with  $n_{\text{trees}} = 50$ ,  $\text{max depth} = 5$  and  $\text{minimal node size} = 0.5\%$ . Only keeping events above  $\text{BDT response} > 0.1$ . The core-collapse photometric sample was determined by [35].

The global efficiency as a function of generated redshift is shown in Figure 8.18. It is clearly seen that our efficiency is more bell-shaped than in the SNLS3 sequential cut analysis (Figure 8.11). The increase of efficiency in the range  $0.2 < gz < 0.6$  from a maximum of  $\approx 0.5$  (sequential cuts) to a peak of  $\approx 0.6$ , is consistent with the global efficiency increase seen in Table 8.5.

The purity of the SNIa sample as a function of the host-galaxy redshift is shown in Figure 8.19. The SNIa purity is quite flat at all redshifts. Core-collapse contamination remains small.

For the SNLS3 sample, Figure 8.20, illustrates the distribution of known SN (spectroscopic and photometric) samples versus BDT response after classification. This subsample distribution agrees with the expected one from the simulation results, validating our procedure.

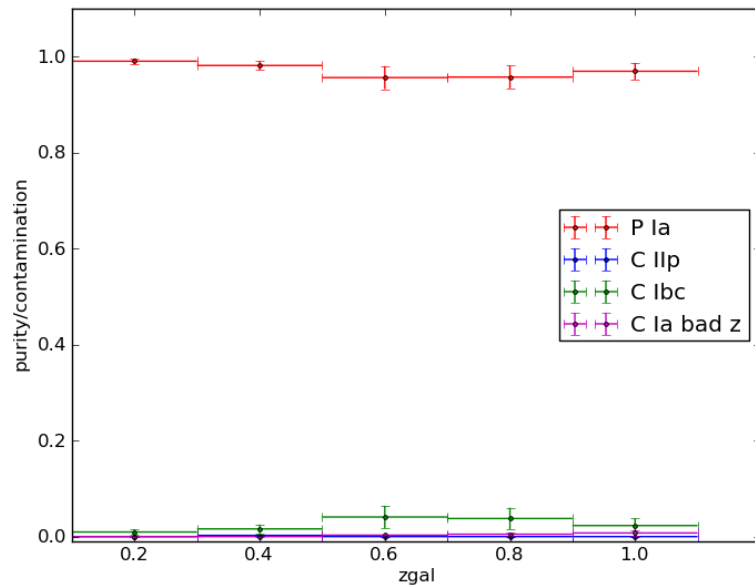


FIGURE 8.19  $z_{\text{gal}} + \text{SALT2} + \text{BDT}$  analysis: Purity of classified SNIa sample (red) and contamination by core-collapse events (other colors) versus host-galaxy redshift ( $z_{\text{gal}}$ ). Weights are applied for volumetric distribution and SN rates.

In summary, it is possible to obtain a high purity and efficiency classification of type Ia SNe using BDTs. Since redshift assignment has an efficiency of only 83%, we propose to set up another analysis based on SN photometric redshifts.

## 8.5 SN photometric redshift analysis (zpho)

I implemented a new classification using photometric redshifts directly obtained from SNIa light curves. The redshift was obtained using the algorithm by *Palanque-Delabrouille et al.* [49] which uses the SALT2 light curve fitter to determine the redshift of Type Ia supernovae.

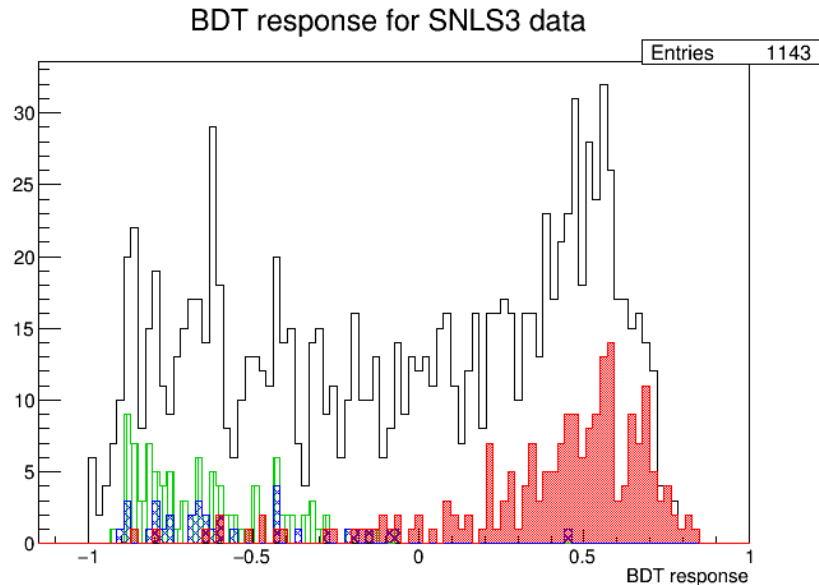


FIGURE 8.20 *zgal* + SALT2 + BDT analysis: SNLS3 data as a function of BDT response. The complete sample is represented with a gray line, while the type Ia SNe are in red and the core-collapse events in blue (crossed area) and green (vertical lines), for photometric and spectroscopic samples respectively.

My goal was, using SN photometric redshifts, to classify type Ia SNe with good efficiency and purity. The redshift plays a crucial part on the classification but also variables that characterize the SN light curves. For the latter, two choices were possible: using a second time SALT2 fitter to obtain variables as in the previous analysis or using a general light-curve fitter such as the one used for SN-like cuts in Equation 5.1.

The first choice, taking SALT2 variables implies using the same fitter for redshift determination and light curve fitting (although with different options). This is a dangerous choice. I will address it in the following subsection to illustrate the results but no robust classification should take twice the same fitter and make the hypothesis (also twice) that events are SNIa-like.

The chosen method was to use a general fit since it provides information from an empirical SN light curve shape. This method has the advantage that it doesn't require a redshift for each event in order to perform the fit. Also, it provides measurements of physical features for all types of events.

The data were required to fulfill the quality and sampling cuts from 5.4.2. To ensure proper photometric redshift determination we required the following for each event:

- Photometric redshift was required to be above 0.1 as expected in our survey. This allowed unreliable SN photometric redshifts to be removed.
- $\chi^2$  of redshift determination  $< 4$ .
- Total  $\chi^2$  of redshift determination  $< 7$ .
- Color  $\chi^2$  of redshift determination  $< 7$ .
- Stretch  $\chi^2$  of redshift determination  $< 3$

Finally, it must be highlighted that parameters for the photometric redshift classification are highly correlated. An analysis based on sequential cuts as in SNLS3 would provide high contamination (shown using SALT2 variables in subsection 8.5.1.1). This is a strong motivation for implementing BDT classification which addresses well a classification with highly correlated variables.

### 8.5.1 Classification using SALT2 twice (zpho+SALT2)

The photometric redshift was determined with the previously mentioned algorithm which uses SALT2. For each event, this redshift, together with the event 4-band light curves were given again to SALT2 (with different options). From this fitting a set of parameters such as magnitude,  $\chi^2$  of the fit, stretch and color were obtained.

The analyses in this section are presented for completeness but must be taken with caution since we assume twice that the events are SNIa-like when using SALT2 twice.

#### 8.5.1.1 zpho + SALT2 + sequential cuts analysis

First, let's explore the space parameter of the variables obtained. Figures 8.21 and 8.22 show the same variable plots as in the host-galaxy redshift analysis in Section § 5.4. The distribution of objects in the parameter space is clearly harder to disentangle than in the previous analysis. The core-collapse events have migrated towards the distribution bulks. This is due to fitting the redshift under the assumption that events are SNe Ia.

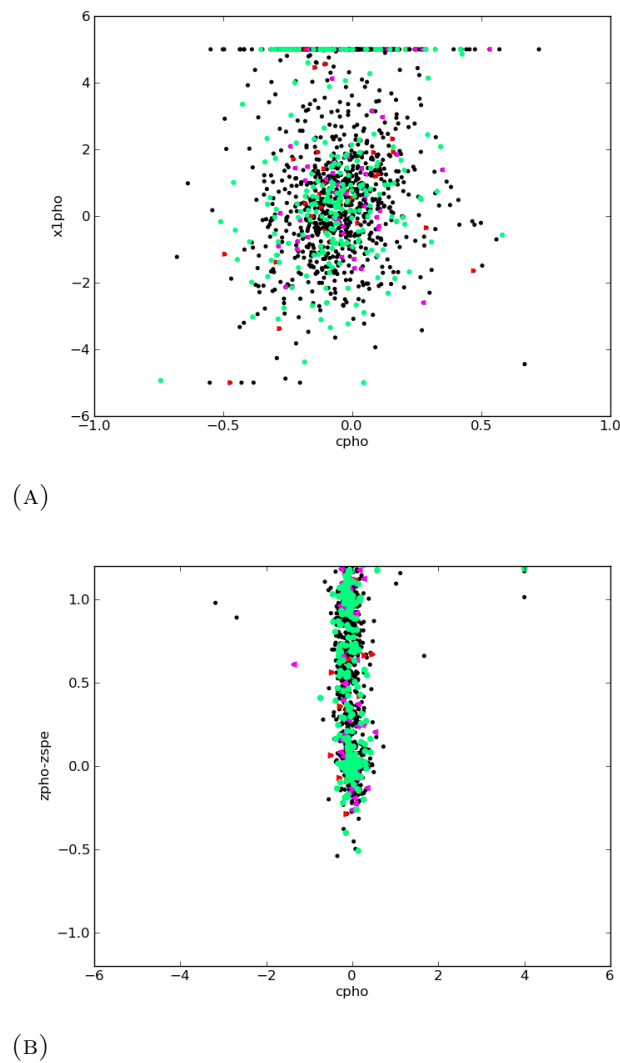


FIGURE 8.21  $z_{p1ho} + \text{SALT2} + \text{sequential cuts}$ : parameters plot for color and stretch.  $X_1$  vs.  $C$  is shown in 8.21a, in SNLS3 this plot motivated a cut based on an ellipse. The difference between redshift assignment as a function of  $C$  is seen in 8.21b. It is clear that a cut in color would be particularly inefficient to tag incorrect redshift determination or core-collapse events. In both plots data are shown (as black points) after sampling and  $\chi^2$  constraints. Spectroscopically identified SNeIa are represented with green circles and spectroscopic (photometric) core collapse events with red (pink) triangles.

As in SNLS3 analysis we can try to apply sequential cuts. Taking the sequential cuts from Section § 5.4 we obtain the results shown in Table 8.6. If we were to continue with such a sequential cut classification, these cuts should be optimized. However, given the obtained contamination and the clear lack of separation between variables in color-magnitude plots, I don't believe that results could achieve a high degree of purity. I show these results here only as reference and motivation to implement other classification techniques.

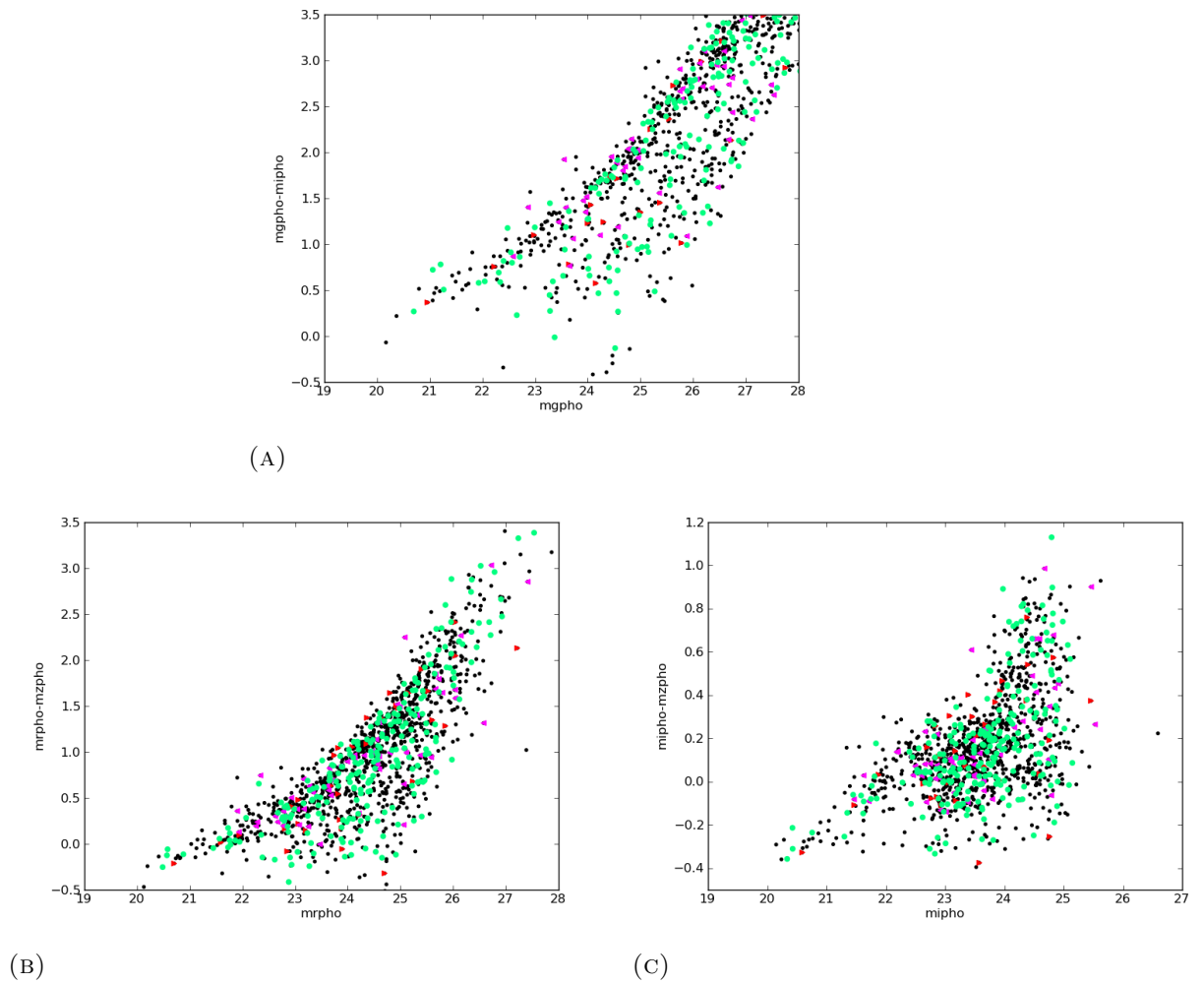


FIGURE 8.22  $z_{\text{pho}} + \text{SALT2} + \text{sequential cuts}$ : color-magnitude diagrams using SALT2 fitted magnitudes for events that passed all cuts up to color and stretch cuts. Spectroscopically identified SNe Ia are represented with green circles, core collapse events (photometric CC) in red (pink).

SNLS3 data

# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
676	189	6	13

TABLE 8.6  $z_{\text{pho}} + \text{SALT2} + \text{sequential cuts}$ : statistics using sequential cuts to classify. The core-collapse photometric sample was determined by [35].

### 8.5.1.2 $z_{\text{pho}} + \text{SALT2} + \text{BDT}$ analysis

Now that I have stated the difficulties of applying sequential cuts to our photometric redshift analysis using SALT2 variables, we can explore classifying events with BDTs.

The hyperparameter space was explored as in the previous BDT case (host-galaxy classification). We obtained as best parameters  $n\text{trees} = 100$  (note that in this analysis there is no overtraining with this number of trees),  $\text{max depth} = 5$  for a given tree and a  $\text{min node size} = 0.5\%$ .

Using these hyperparameters, a classification can be done that disentangles the distributions of signal and background as can be seen in Figure 8.23. When compared to the signal-background distribution in the host-galaxy redshift BDT analysis (Figure 8.15) we see that when using photometric redshifts the two samples are closer and harder to disentangle using this BDT SALT2 classification. This may be attributed to the fact that we use SALT2 twice to extract first redshift and then light curve parameters.

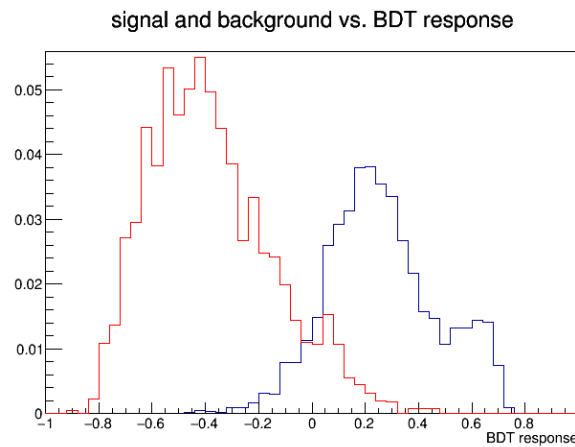


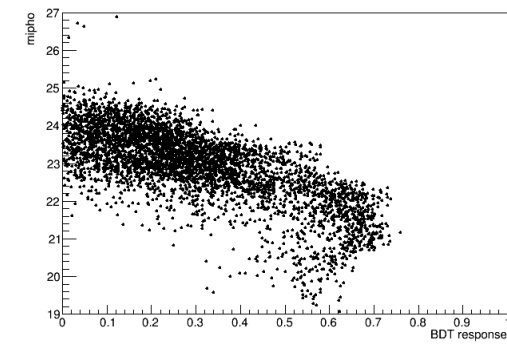
FIGURE 8.23  $z\text{pho} + \text{SALT2} + \text{BDT}$  analysis: signal (blue line) and background (red line) distributions with respect to the BDT response. A complete separation of the two samples is possible for BDT response  $> 0.5$  but with low efficiency. Distributions are not normalized.

The study of the signal shoulder in 8.23 is illustrated in Figure 8.24. Events in the shoulder are objects of relative low magnitude and at redshift  $< 0.6$ . Note that in particular, very bright events  $\text{mihal} < 21$  mostly appear in this shoulder.

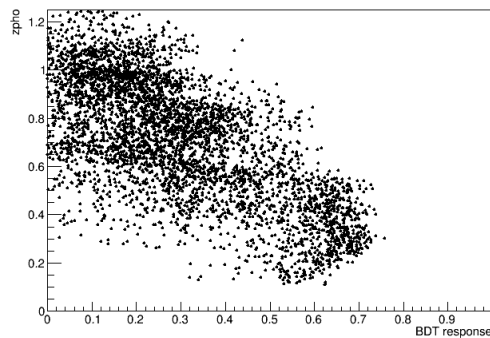
Once again, to determine the value of our BDT response cut we had to take into account both SNIa efficiency and purity of the sample. In Figure 8.25 the efficiency and purity is shown with indications of the BDT response. The chosen cut was to keep events above a BDT response of 0.2.

For the simulation and SNLS3 data, this classification provided the global purity and efficiency shown in Table 8.7. The contamination is lower than that in the SNLS3

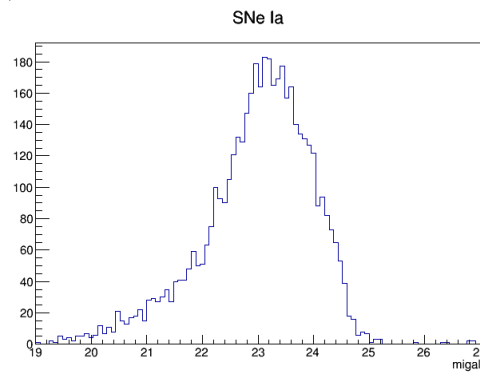




(A)



(B)



(C)

FIGURE 8.24  $z_{pho}$  + SALT2 + BDT analysis: study of the signal shoulder in Figure 8.23. For  $BDT\ response > .55$  events are shown not to be of high magnitude in the  $i_M$  filter (maximum  $\approx 23mag$ ) 8.24a and at SN photometric redshift below 0.6 8.24b. The  $i_M$  magnitudes of these events are in the tail of the magnitude distribution for SNIa events 8.24c. The magnitudes show less dispersion than those in the  $z_{gal}$  shoulder in 8.16.

sequential cut analysis however, the price to pay lies in the SNIa efficiency as it is clear in Figure 8.25.

When comparing with the SNLS3 analysis, the number of spectroscopic core-collapse events is higher, 2 events, than in the original analysis. I must highlight that the

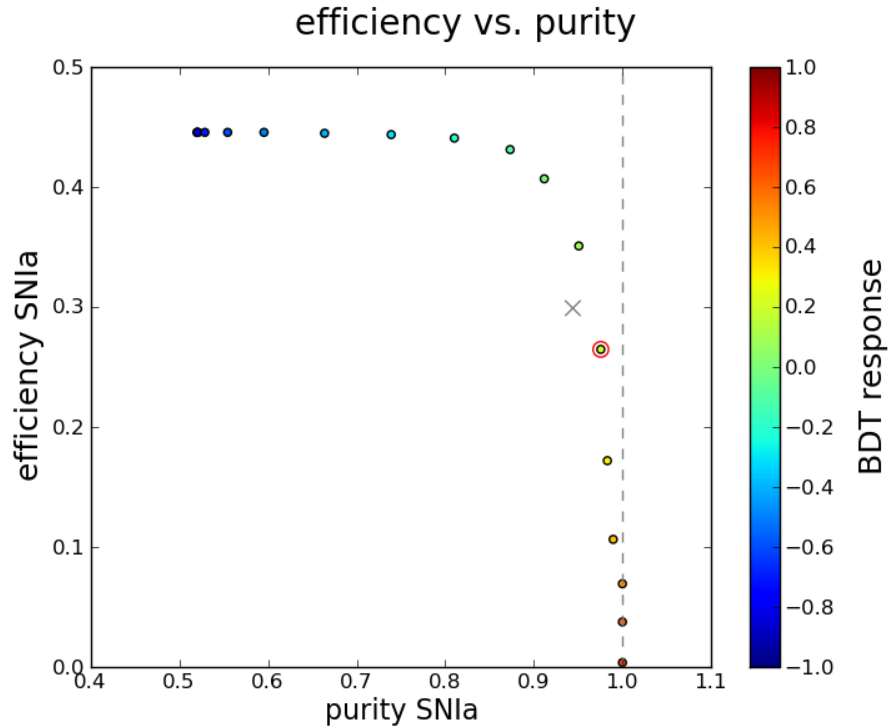


FIGURE 8.25 *zpho* + SALT2 + BDT analysis: Global efficiency of SNIa against global purity of SNIa classified sample. Colors indicate the value of the BDT response. The chosen value for the BDT cut (0.2) is circled in red. Dashed line represents purity = 1. The gray cross represents the efficiency and purity values for the host-galaxy redshift sequential cut analysis SNLS3.

spectroscopic core-collapse sample is not a complete sample and has very low statistics, then these results are not necessarily conclusive of a higher contamination.

The number of spectroscopic SNe Ia in our sample is higher (+23 events representing an increase of 13%) which agrees with higher efficiency at low redshift (see Figure 8.28). The number of events in the photometric sample is 3% higher than in the host-galaxy redshift sequential cuts analysis which is in slight disagreement with the somewhat lower efficiency.

Simulation			
purity SNe Ia	contamination		efficiency SNIa
	bad redshift SNe Ia	core-collapse	
$96.0 \pm 0.8\%$	$0.09 \pm 0.06\%$	$3.9 \pm 0.8\%$	$26.5 \pm 0.8\%$

SNLS3 data			
# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
499	198	2	0

TABLE 8.7 *zpho* + SALT2 + BDT analysis: Statistics for SNLS3 data and simulation purities for a BDT classification with 100 trees, max depth=5 and minimal node size = 0.5%. Only keeping events above BDT response variable 0.2. The core-collapse photometric sample was determined by [35].

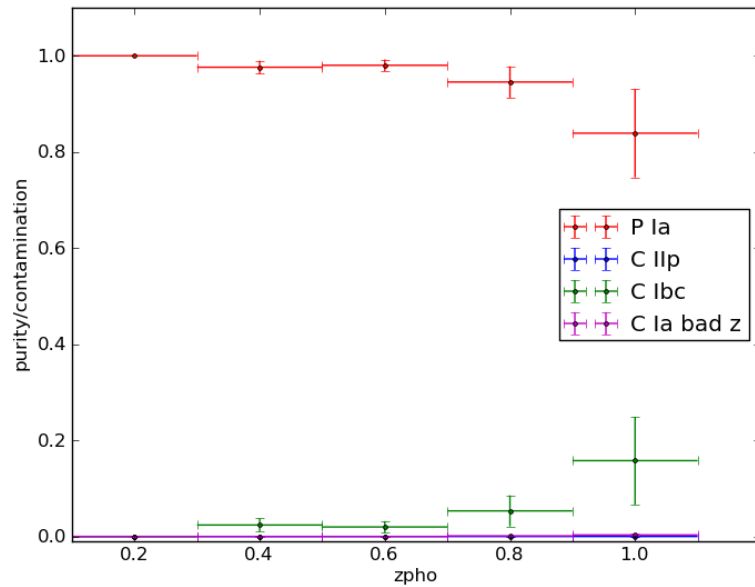


FIGURE 8.26 *zpho* + SALT2 + BDT analysis: Purity of classified SNIa sample (red) and contamination by core-collapse (other colors) versus SN photometric redshift. Weights are applied for volumetric distribution and SN rates.

Purity and contamination as a function of the photometric redshift are shown in Figure 8.26. The purity has a tendency to decrease at higher redshift due to a higher core-collapse contamination. When comparing the purity of this analysis with that of the host-galaxy redshift analysis, the contamination of core collapse is higher at large redshifts for both sequential cuts and BDT analysis using SALT2 twice. This effect is still under study.

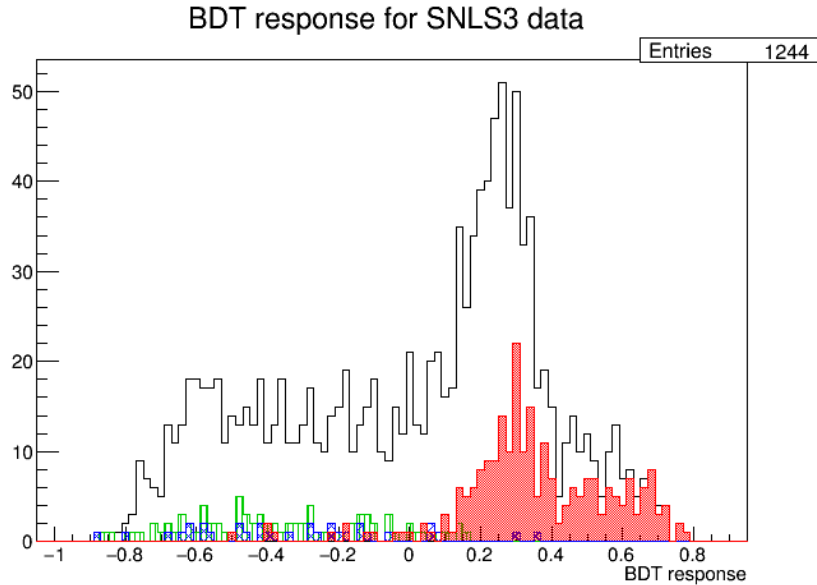


FIGURE 8.27  $z_{\text{pho}} + \text{SALT2} + \text{BDT}$  analysis: SNLS3 data as a function of BDT response. The complete sample is represented with a gray line, while the type Ia SNe are in red and the core-collapse in blue (crossed area) and green (vertical lines), for photometric and spectroscopic samples respectively.

In this analysis, obtaining a pure SNIa sample is a challenge. For SNLS3 data the core-collapse background is more extended in the BDT space as can be seen in Figure 8.27. We see also that to obtain a higher purity sample, the global efficiency is quite reduced. This analysis is affected by the recurrent assumption that we are dealing with SNe Ia, when the case is that we have a sample of SNe and we want to classify them in different types.

The efficiency evolution as a function of simulation generated redshift is shown in Figure 8.28. It is, as the host-galaxy redshift BDT analysis, bell-shaped (with high efficiency at low redshifts) and it decreases fast at high redshifts. The jumpiness of the different bins may be attributed to the varying dispersion of the SN photometric redshift determination for type Ia SNe. For example at  $g_z \approx 0.6$  there is a lot of incorrect assignments as can be seen in Figure 8.29.

### 8.5.2 $z_{\text{pho}} + \text{general fit} + \text{BDT}$ analysis

In this section I will present my classification using SN photometric redshifts and a general SN fit. I used the fit in Equation 5.1 since it fits well all types of SNe which allows extracting variables to differentiate them.

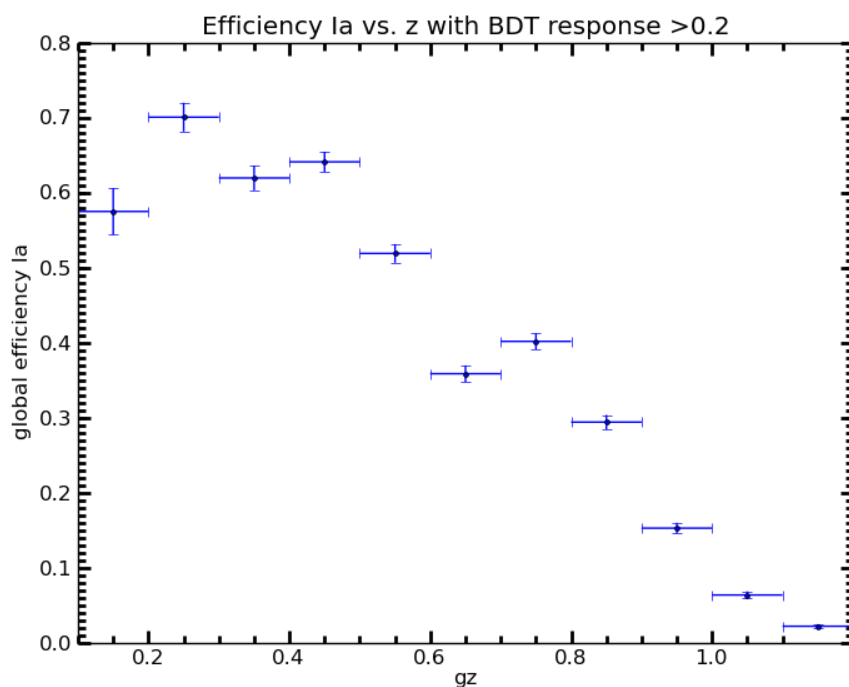


FIGURE 8.28  $z_{\text{pho}} + \text{SALT2} + \text{BDT}$  analysis: Global efficiency of SNIa classification as a function of generated redshift ( $g_z$ ).

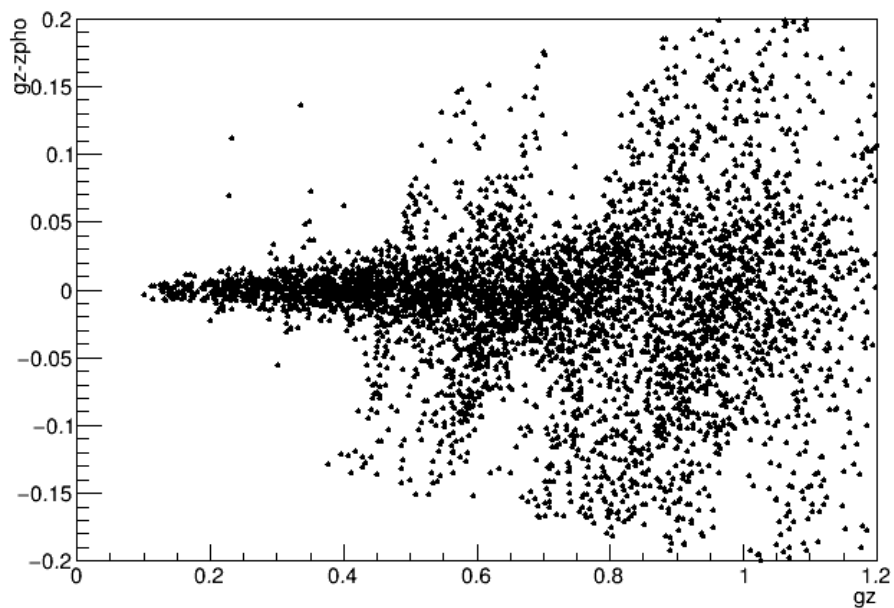


FIGURE 8.29  $z_{\text{pho}} + \text{SALT2} + \text{BDT}$  analysis: dispersion of SN photometric redshift as a function of generated redshifts ( $g_z$ ) for the simulated type Ia SNe. The algorithm converges at small generated redshifts and then dispersion grows with some oscillations.

To select the data, the quality cuts were the same as in Subsection 5.4.2 and the photometric redshift requirements indicated at the beginning of this section. The algorithm’s hyperparameter selection resulted on  $n\text{trees} = 100$  (note that for this analysis there was no overtraining with this number of trees), 0.5% minimum node size and  $\text{max depth} = 5$  for each tree.

The variables provided by the general fit are, for each filter: amplitude, fall and rise times, the estimated date of maximum and a constant. Since both the constant and the time of maximum do not provide relevant information for classification I chose not to take them into account.

Now, we can start training the algorithm with the selected data and the chosen hyperparameters for our BDTs. If we use half of the simulation to train and half to test, we obtain the signal and core collapse distributions as a function of BDT response shown in Figure 8.30. When comparing with the photometric redshift analysis with SALT2 (Figure 8.23) we see that the maximum of signal and background distributions are further away. We have then a stronger classifier than the one presented in the previous section (with SALT2 twice).

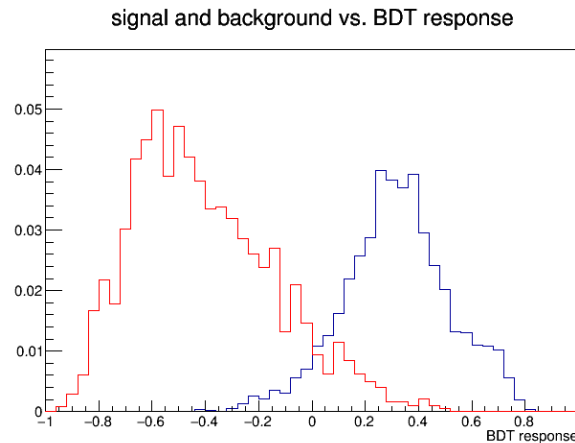


FIGURE 8.30  $z_{\text{pho}} + \text{general fit} + \text{BDT}$  analysis: signal ( blue line ) and background (red line) distributions with respect to the BDT response. A complete separation of the two samples is possible for BDT response  $> 0.5$  but with low efficiency. Distributions are not normalized.

To choose the cut to be applied in the BDT response variable, we examine the global efficiency and purity of SNIa in Figure 8.31. I chose to cut at BDT response  $> 0.2$  to obtain a good purity and efficiency.

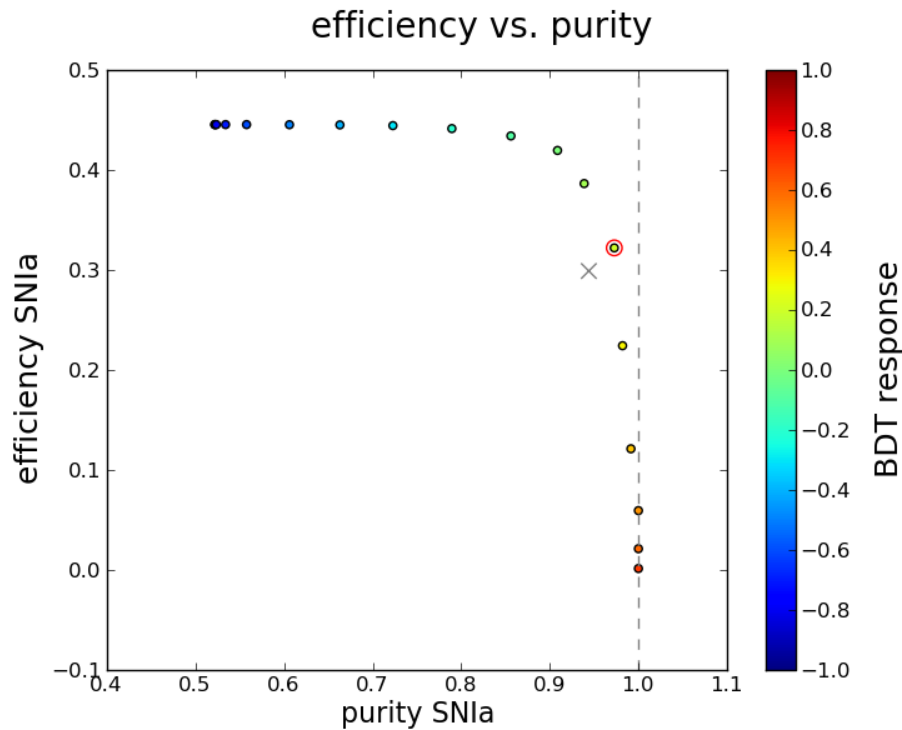


FIGURE 8.31 *zpho* + general fit + BDT analysis: Global efficiency of SNIa against global purity of SNIa classified sample. Colors indicate the value of the BDT response. The chosen BDT response cut (0.2) is indicated with a red circle. The dashed line is for purity equal to one. The gray cross indicate values for the SNLS3 host-galaxy redshift sequential cut analysis.

Purity against redshift is shown in Figure 8.32. We see that the purity is almost uniform in redshift (within error bars). The core-collapse contamination slightly increases with redshift but remains below 5%.

The global purity can be seen in Table 8.8 and is found to be higher than the one obtained using host-galaxy redshifts and sequential cuts (SNLS3 analysis). We also have a higher global efficiency with respect to all preceding analyses, including the SNLS3 host-galaxy sequential cut one.

Finally I show the evolution of global efficiency as a function of redshift for a BDT cut of 0.2 in Figure 8.33. As before, the evolution is bell-shaped and there is an increase of efficiency at low redshifts. The decrease is softer at higher redshifts when compared to the SN photometric redshift with SALT2 twice. This is in agreement with the high global efficiency shown in Table 8.8.

We obtain an equivalent number of photometric SNe Ia in this analysis as in the SNLS3

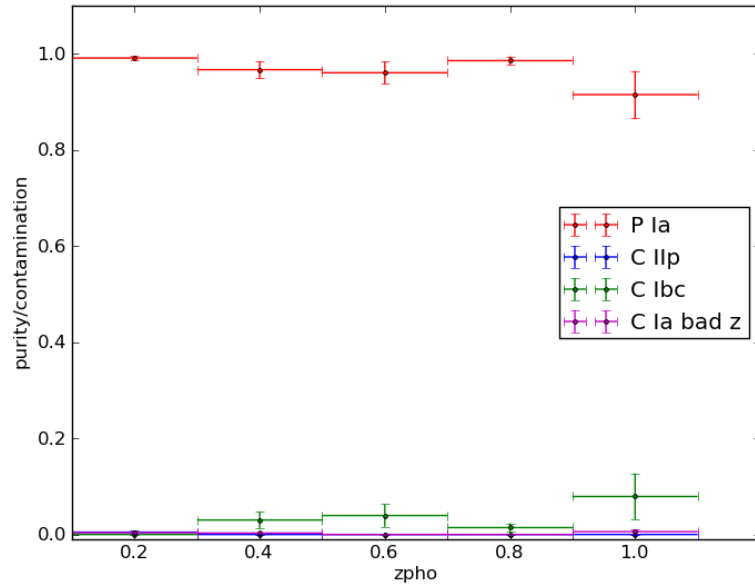


FIGURE 8.32  $z_{\text{pho}}$  + general fit + BDT analysis: Purity of classified SNIa sample (red) and contamination by core-collapse (other colors) against photometric redshifts ( $z_{\text{pho}}$ ). Weights are applied for volumetric distribution and SN rates.

sequential cuts one. However, the number of spectroscopically confirmed SNe Ia is higher in our sample. This is in agreement with the higher efficiency at low redshift seen in Figure 8.33. From the SNLS3 core-collapse events, there is one event in the border of our cut and one other event well above the cut as can be seen in Figure 8.34. Here, we see that the core-collapse background is more spread than in the host-galaxy BDT analysis. Once again, the SNLS3 sample is a limited one and due to its low statistics is not in contradiction with the low core-collapse contamination found using simulations.

#### Simulation

purity SNe Ia	contamination		efficiency SNIa
	bad redshift SNe Ia	core-collapse	
$95.8 \pm 0.8\%$	$0.29 \pm 0.08\%$	$3.9 \pm 0.7\%$	$32.2 \pm 0.8\%$

#### SNLS3 data

# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
482	187	1	1

TABLE 8.8  $z_{\text{pho}}$  + general fit + BDT analysis: Statistics for SNLS3 data and simulation purities for a BDT classification with 100 trees, max depth=5 and minimal node size = 0.5%. Only keeping events above BDT response variable 0.2. The core-collapse photometric sample was determined in an independent analysis [35].

Finally, to illustrate the power of the BDTs analysis I show 2D-distributions of variables (that we could have cut on to perform classification) for known types of SNLS3 data



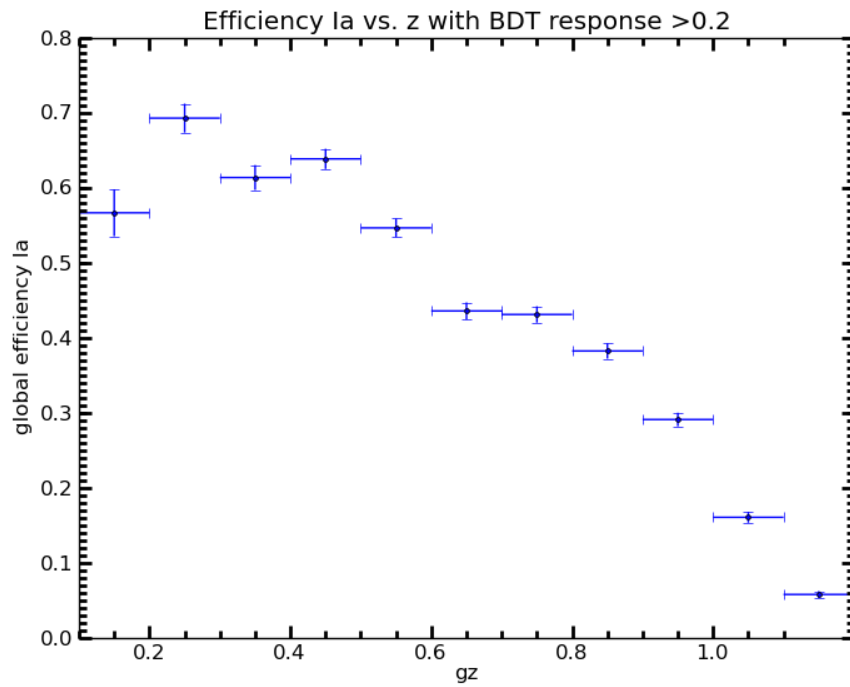


FIGURE 8.33 *zpho* + general fit + BDT analysis: Global efficiency of SNIa classification as a function of generated redshift (*gz*).

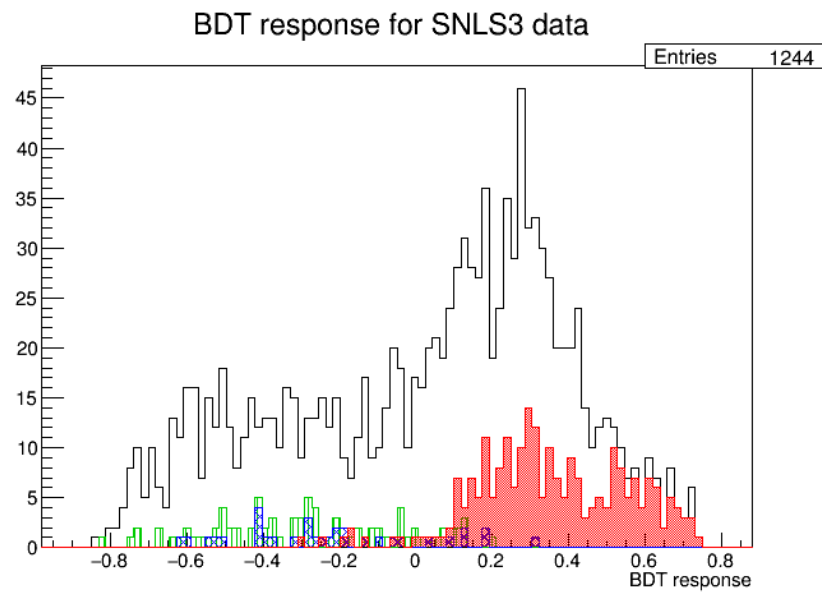


FIGURE 8.34 *zpho* + general fit + BDT analysis: BDT response for SNLS3 data. The complete sample is represented with a gray line, while the type Ia SNe are in red and the core-collapse in blue and green, for photometric and spectroscopic samples respectively.

and the BDT response for this analysis in Figures 8.35 and 8.36.

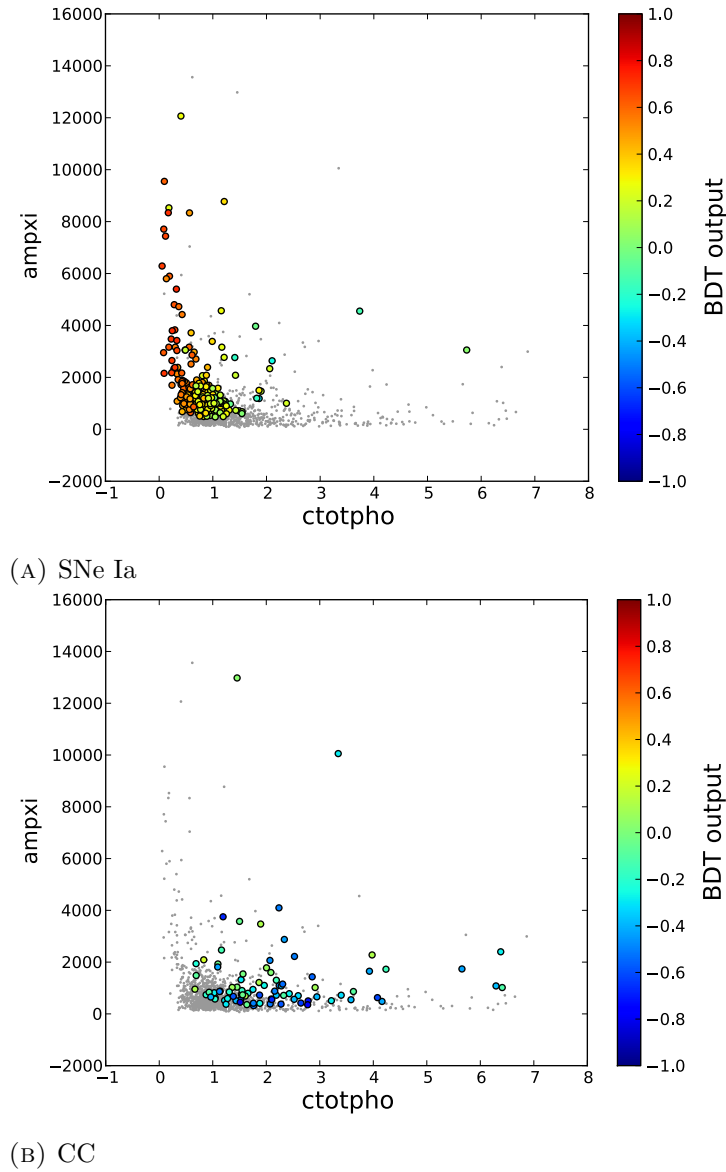


FIGURE 8.35  $z_{\text{pho}} + \text{general fit} + \text{BDT}$  analysis: Example of the BDT response for known events in the SNLS3 sample. In these parameter spaces, there is no obvious cut to disentangle SNe Ia from core-collapse (CC) SNe. 8.35a (8.35b) SNe Ia (CC) fitted amplitude in  $i_M$  versus total  $\chi^2$  of photometric redshift fit. Gray points indicate SNLS3 data, colored points the species with their BDT response score. The BDT response tends toward -1 for background like events, while for signal-like events (SNe Ia) it approaches one.

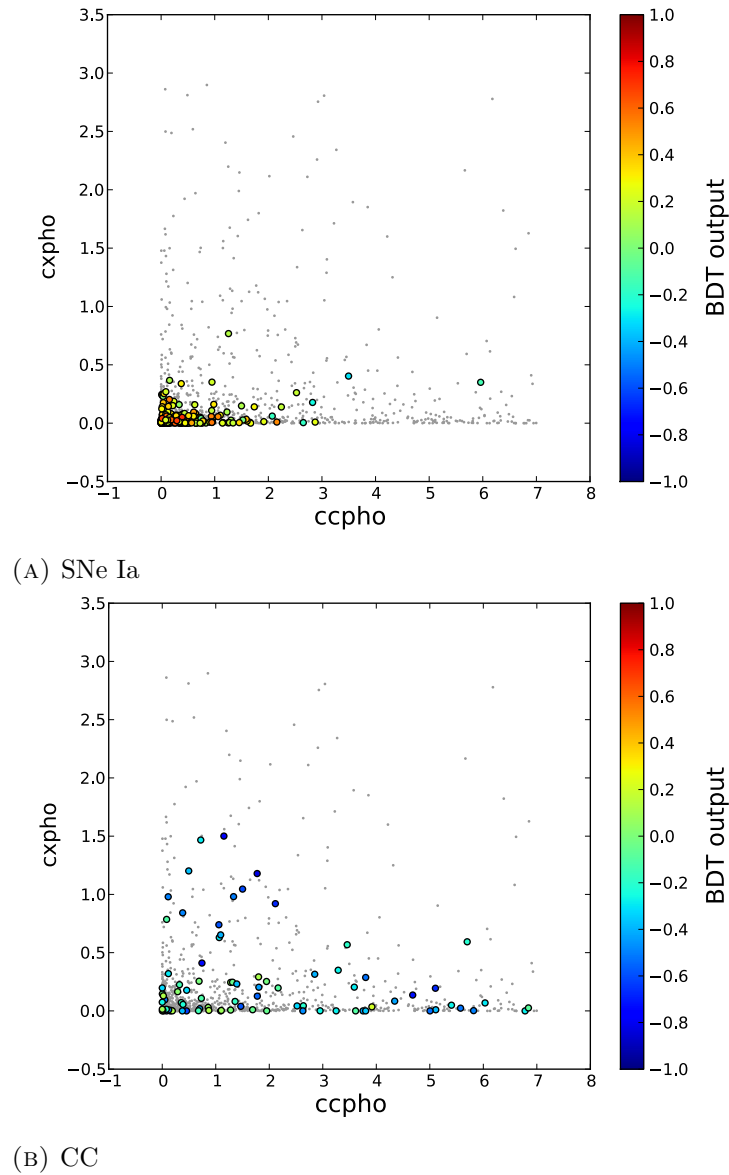


FIGURE 8.36  $z_{pho} + \text{general fit} + \text{BDT}$  analysis: Example of the BDT response for known events in the SNLS3 sample. In these parameter spaces, there is no obvious cut to disentangle SNe Ia from core-collapse (CC) SNe. Stretch  $\chi^2$  of photometric redshift versus color  $\chi^2$  of photometric redshift for SNe Ia 8.36a (CC 8.36b). Gray points indicate SNLS3 data, colored points the species with their BDT response score. The BDT response tends toward -1 for background like events, while for signal-like events (SNe Ia) it approaches one.

## 8.6 Summary

I successfully implemented a classification of type Ia SNe based on SN photometric redshifts. The advantages of such an analysis include that all events have an assigned redshift and that there is no dependence on an external catalogue of host galaxy redshifts.

Classification of events is a high-dimensional problem and as such, it can be addressed using machine learning techniques. I showed that for the host-galaxy redshift analysis, the classification could be improved to obtain slightly higher efficiency and purity of the SNIa sample using BDTs. This is of relevance since one of the main goals of a photometric pipeline is to obtain large statistics over a broad redshift extension without the need of spectroscopy.

For the SN photometric redshift analysis I encountered two main questions. First, which fit to use to extract information from the light curves and, second, how to handle highly correlated variables. The latter was very well addressed by supervised learning methods. Regarding the fit to be used, once SALT2 is used for obtaining the photometric redshift, an assumption about the nature of the object is already done: it is assumed to be a SNIa. Then, it became natural to use a general fit that could extract information to distinguish core-collapse events from type Ia ones. In this Chapter I presented both analyses for completeness.

In general, the global efficiency of the SN photometric redshift analysis is higher. This is partially due to the fact that we have an assigned redshift for each of our selected events. This was not the case for the host-galaxy redshift analysis, which had an assignment efficiency of 83%. Also, the use of supervised learning algorithms such as BDTs provide good results in high dimensional problems such as ours.

The sample obtained using BDTs and the general fit proved of high purity (95.8%) and higher than that of the SNLS3 data. The number of photometric candidates in the SN photometric redshift analysis is equivalent to the one in the SNLS3 host-galaxy redshift with sequential cuts. However, the number of spectroscopic events in our sample increased by 7% with respect of the original analysis. This is in agreement with our higher efficiency at low redshift (Fig. 8.37). A summary of purity, contamination, efficiency and statistics is given in Table 8.9.

It must be highlighted that the core-collapse simulation has very low statistics at the SN-like sample level (the starting point of our classification). Although we expect a low number of core-collapse events at this stage, it would be advisable to have a smoother distribution to be used as test. This is a limitation of the analysis presented here.

Moreover, supervised learning methods, as BDTs, rely on signal and background samples to train the algorithm. I provided core-collapse SNe and SNe Ia with bad redshift assignment as backgrounds for training although there may be other background sources that were not eliminated using SN-like cuts. To correct for this, other backgrounds must be identified which will require to complete SN-like cuts. In the case some background survive we will need to model them to generate simulated event samples for training with large statistics.

I remind here, that this classification (and all my work presented in this PhD) has been centered on a future application of these methods to the SNLS5 analysis.

Simulation				
analysis	purity SNe Ia	contamination		efficiency SNIa
		SNe Ia bad z	core-collapse	
zgal+SALT2+sequential(SNLS3)	$94.4 \pm 0.5\%$	$0.65 \pm 0.08\%$	$4.9 \pm 0.5\%$	$29.9 \pm 0.3\%$
zgal+SALT2+BDT	$96.6 \pm 0.6\%$	$0.4 \pm 0.1\%$	$2.9 \pm 0.6\%$	$30.8 \pm 0.8\%$
zpho+SALT2+BDT	$96.0 \pm 0.8\%$	$0.09 \pm 0.06\%$	$3.9 \pm 0.8\%$	$26.5 \pm 0.8\%$
zpho+general fit+BDT	$95.8 \pm 0.8\%$	$0.29 \pm 0.08\%$	$3.9 \pm 0.7\%$	$32.2 \pm 0.8\%$

SNLS3 data				
	# events	# spec. SNe Ia	# spec. CC	# photo. CC
zgal+SALT2+sequential(SNLS3)	486	176	0	0
zgal+SALT2+BDT	524	175	0	0
zpho+SALT2+BDT	499	198	2	0
zpho+general fit+BDT	482	187	1	1

TABLE 8.9 Classification summary: statistics of different methods. The core-collapse photometric sample was determined in an independent photometric analysis [35].

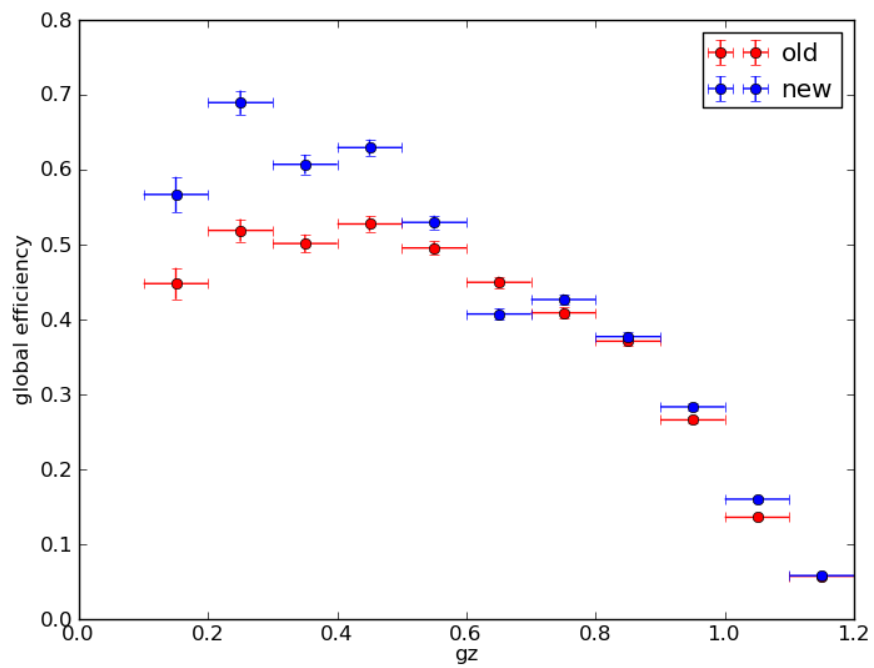


FIGURE 8.37 Comparison of global efficiencies for simulation using the SNLS3 method (red points) and the proposed SN photometric classification (blue points) against generated redshift.

## Chapter 9

# Conclusions

In this PhD work, I presented improvements on the detection of transient events and the classification of supernovae using supernova photometric redshifts in the SNLS deferred analysis.

The deferred photometric pipeline of SNLS provided me a unique opportunity to study a complete and dedicated pipeline for a 100% photometric SNIa sample. In the view of the limitations of spectroscopic resources, obtaining such a reliable and large sample is of importance for SNIa studies.

The first part of my work consisted on improving the detection of transient events which is the first step for detecting supernovae. Transient events are detected in difference images which are usually filled with various artifacts from instrumental defects and incomplete subtraction of permanent objects. In order to eliminate those artifacts and extract interesting events, I proposed a treatment to reduce the number of non SN-like events using morphological component analysis. A new detection strategy was also set up which handles better many years of data and improves coordinate resolution.

Using this new procedure I reached a reduction of spurious events of more than a factor two while maintaining the interesting SN-like sample. Coordinate resolution was shown to be sufficient for a photometric classification. However, there is big room for improving this coordinate resolution. Possibilities to reduce the impact of stacking signal coordinates and coordinates from spurious detections include extracting coordinates at the level of a 3-week lunation stack or extracting them directly from the subtracted images.

The former is at the moment under study, while I leave the later as a possibility since it requires thorough study.

Morphological component analysis algorithms have shown to be very powerful for reducing artifacts in our subtracted image stacks. However, possible improvements can come from eliminating artifacts at the level of subtracted images instead of stacks. This could be valid for our deferred pipeline but it should be applied at the beginning of the survey. Other possibilities for deferred and real-time transient event detection pipelines include cleaning images with a fast multi-resolution method and exploiting additional multi-band and temporal information.

The second part of my work dealt with classifying SN-like events into type Ia and core-collapse SNe. I presented here a new method of classification using supernova photometric redshifts. These redshifts can be obtained for all SN-like events and do not rely on an external catalogue. To exploit all available information I proposed to use machine learning algorithms, in particular BDT.

I presented different classification options depending on the redshift, light curve fitter and strategy used. In general, the global efficiency and purity of the photometric redshift analysis is higher than host-galaxy redshift analyses. We obtained a contamination of the sample below 5% for this redshift analysis using BDTs. This represents an important result since we are improving the quality of our 100% photometric sample using novel algorithms.

This classification work can be improved by simulating larger samples of core-collapse supernovae in order to assess the contamination levels with more accuracy. To take advantage of our machine learning techniques we must study thoroughly the SN-like sample to identify the residual non-SN backgrounds in order either to reject them or to model them and include them in the training of the BDT algorithm. This is a requirement before we can apply the BDT classification to SNLS5 data.



# Appendix A

## Resumé du travail en francais

De 2003 à 2008, le SuperNova Legacy Survey (SNLS) a détecté un millier de supernovae (SNe) distantes auprès du télescope CFH (Canda-France-Hawaii). L'analyse des trois premières années de données a conduit à des publications majeures pour la cosmologie : l'accélération de l'expansion de l'Univers est confirmée à un niveau de confiance supérieur à 99.9% par les seules mesures sur les supernovae de type Ia (Fig. [A.1](#)). Ce résultat s'appuie sur la statistique importante recueillie par SNLS et sur un contrôle poussé des systématiques. Les données de SNLS constituent à ce jour le meilleur échantillon pour les études cosmologiques à base de SNe.

Traditionnellement, les supernovae sont détectées par photométrie, grâce à l'observation de leurs courbes de lumière (mesure du flux lumineux de la SN au cours du temps), puis elles font l'objet d'observations spectrales pour mesurer leur décalage spectral et déterminer leur type, soit thermonucléaire (type Ia), le seul utile pour la cosmologie, soit gravitationnel (types Ib, Ic ou II) qui constitue un fond à éliminer.

Cette stratégie n'est pas envisageable pour les projets futurs dont les centaines de milliers de SNe ne pourront toutes être observées en spectroscopie par manque de temps. Déjà, aujourd'hui, 50% seulement des SNe potentielles peuvent être étudiées par spectroscopie. Il faut donc envisager une procédure qui sélectionne les supernovae, identifie leur type et mesure leur décalage spectral uniquement à l'aide de leurs courbes de lumière (sélection dite photométrique), ce que propose cette thèse.

Au sein de SNLS, le groupe de l'Irfu/SPP est responsable de la sélection photométrique des supernovae. Appliquée aux 3 premières années de données, cette analyse a abouti

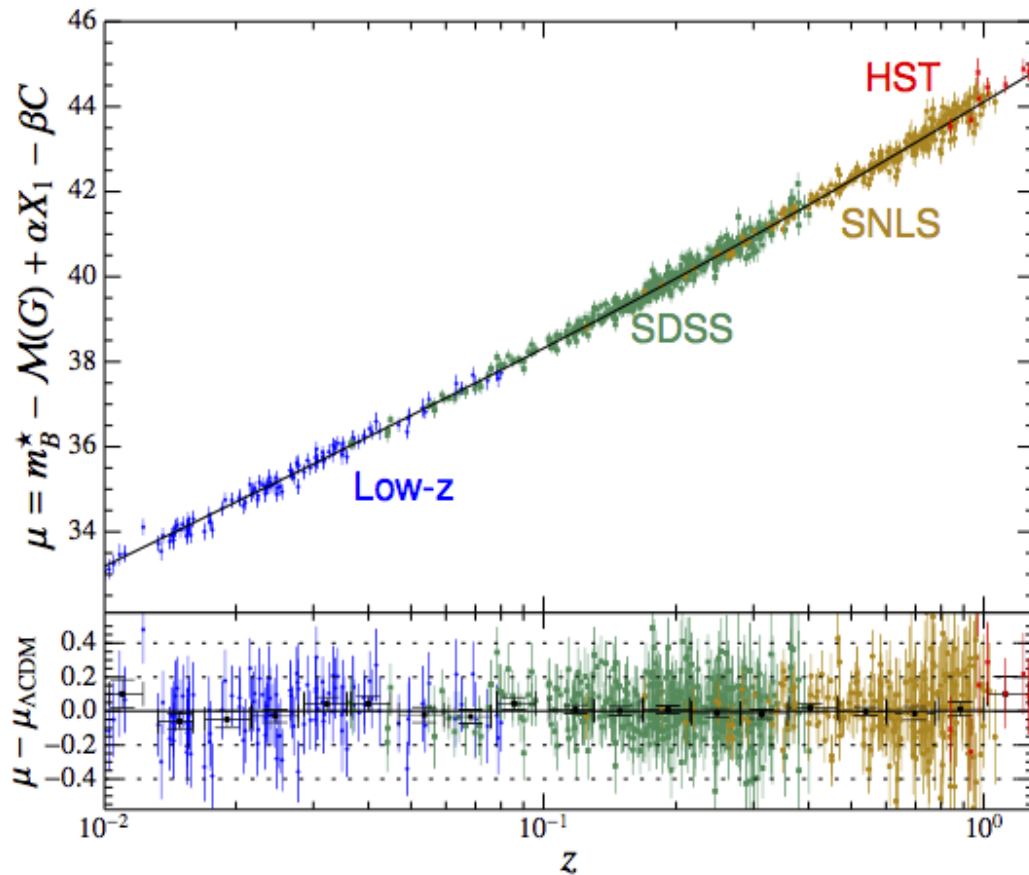


FIGURE A.1 JLA diagramme de Hubble avec 472 SNeIa de SNLS3 et des autres expériences de supernovae. [65]. Modèle  $\Lambda$ CDM représenté en noir.  $\mu$  est donnée dans l'équation 3.5.  $\Omega_m = 0.303 \pm 0.012, \Omega_\Lambda = -1.027 \pm 0.055$  et  $H_0 = 68.50 \pm 1.27$ .

à la sélection d'un lot de 485 SNe Ia situées dans des galaxies hôtes identifiées, de décalage spectral connu ( la sélection est basée sur des coupures sequentielles illustrées dans Fig. A.2). L'efficacité de la méthode est de 80% et la contamination du lot en SNe gravitationnelles, évaluée par simulation rapide, est de 4%. Par rapport à l'analyse traditionnelle, cette sélection permet d'enrichir de 50% le lot de supernovae de type Ia, essentiellement à haut décalage spectral.

Plusieurs sujets ont été abordés grâce à cet échantillon, notamment un algorithme de reconstruction directe du décalage spectral des SNe Ia d'après leurs courbes de lumière. Ces résultats ont fait l'objet de trois publications de SNLS dont le groupe du SPP est premier auteur. Ce travail est part d'une prochaine publication portant sur l'utilisation de ce lot photométrique pour la cosmologie.

Mon travail porte sur des améliorations sur la détection d'événements transitoires et la classification des supernovae (SNe) en utilisant les redshifts photométriques de supernova

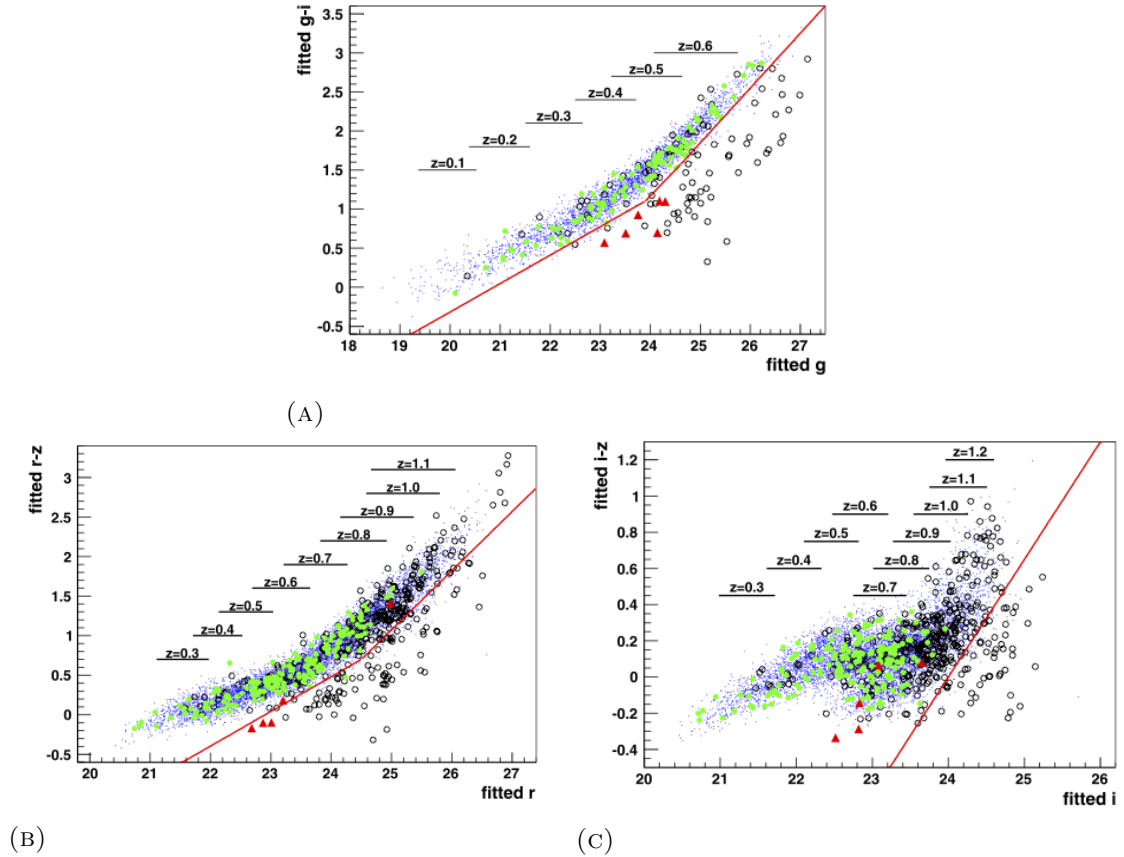


FIGURE A.2 Diagrammes couleur-magnitude (SALT2) pour les événements qui passent les premières coupures de sélection. Données SNLS3 en noir, événements spectroscopiques type Ia (vert), core-collapse (rouge) et simulations en bleu. [66].

dans l'analyse diffère de SNLS. Ce travail va être appliqué dans le lot complet de données de SNLS (5 ans) et sur l'utilisation de ces SNe Ia pour la cosmologie. Ce travail précède la publication finale de SNLS sur le sujet.

La détection des événements transitoires peut fournir de nombreuses fausses détections, tandis que la classification photométrique des SNe de type Ia est généralement contaminée par d'autres types de supernovae. Réduire le nombre de fausses détections et les SNe mal classées tout en préservant l'échantillon du type Ia sont des questions importantes pour les investigations présentes et futures.

Afin de réduire les artefacts qui fournissent des fausses détections (Fig. A.3), j'ai développé un traitement pour les piles de soustractions pour réduire le nombre d'événements qui ne ressemblent pas à des SNe en utilisant l'algorithme Morphological Component Analysis. Cette technique exploite la diversité morphologique des objets à détecter pour extraire le signal d'intérêt. Au niveau de nos piles de soustraction, événements comme

Stack	Ancienne procédure		Nouvelle procédure			
	coordinate resolution $\pm 0.002$	magnitude biais $\pm 0.0002$	No season stacks		With season stacks	
			coordinate resolution $\pm 0.002$	magnitude biais $\pm 0.0002$	coordinate resolution $\pm 0.002$	magnitude biais $\pm 0.0002$
1 an	0.709	0.0334	0.698	0.0324	0.698	0.0324
3 ans	0.725	0.0349	0.715	0.0340	0.710	0.0335
5 ans	0.741	0.0365	0.731	0.0355	0.726	0.0350

TABLE A.1 Résolution des coordonnées en pixels (limite supérieure) avec le biais en magnitude correspondant. Pour plusieurs années de données.

Ancienne procédure		Nouvelle procédure	
# detections	# SN-like	# detections	# SN-like
302,987	1,483	127,808	1,345

TABLE A.2 Nombre des détections et événements qui se ressemblent au SNe dans les données de SNLS3 pour l'ancienne et la nouvelle procedure.

SN sont plutôt circulaires alors que la plupart des détections parasites présentent des formes différentes.

Des images de SNe Ia générées par Monte-Carlo (MC) ont été utilisées pour étudier l'efficacité de la détection (Fig. A.4) et la résolution des coordonnées (Table A.1). Lors d'un essai sur le données SNLS des 3 années cette procédure diminue le nombre de détections par un facteur de deux, tout en ne perdant que 10 % d'événements qui ressemblent à des SNe, presque tous tenus (Table A.2). Les résultats des MC montrent que l'efficacité de détection SNIa est équivalente à celle de la méthode originale pour les événements lumineux, tandis que la résolution des coordonnées est légèrement améliorée.

Par rapport aux analyses publiées sur les 3 premières années, j'ai généralisé la sélection photométrique des SNe Ia au cas difficile - mais plus proche de la problématique des projets futurs - où le décalage spectral des supernovae n'est pas connu grâce à la galaxie-hôte mais est déterminé à partir des courbes de lumière des supernovae elles-mêmes, tout en contrôlant de manière plus précise la contamination de ce lot en SNe gravitationnelles.

L'analyse différée utilise uniquement des informations photométriques pour classer les supernovae. Les avantages d'un échantillon photométrique comprennent plus grand nombre d'événements classés comme de type Ia, la couverture de redshift plus grande et pas besoin d'observations spectroscopiques.

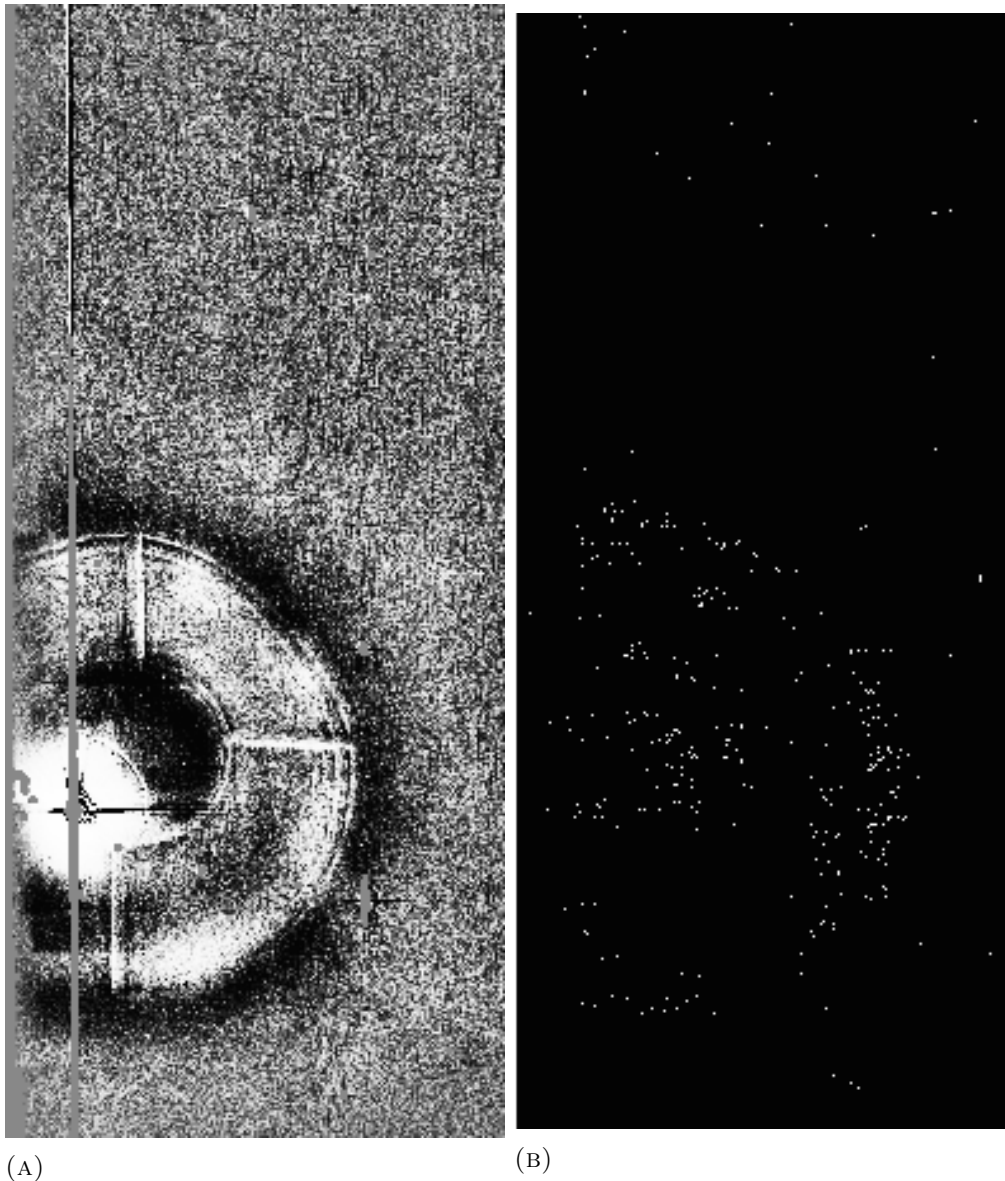


FIGURE A.3 Pile de soustraction D4, ccd 00 et lunation 10 avec étoile brillante et “optical ghost” dans A.3a et le plan de détection avec la méthode originale de SNLS3 A.3b. De nombreuses detections peuvent être attribuées à des artefacts.

Je présente ici une nouvelle classification en utilisant des redshifts photométriques de supernovae optimisée par une stratégie de classification avec machine learning. Cet algorithme fournit des décalages vers le rouge pour tous les événements avec une meilleure précision moyenne et inférieure erreurs catastrophiques que l’analyse avec photométrie redshifts de la galaxie hôte avec SNLS3. J’ai optimisé la stratégie de sélection à l’aide des techniques de machine learning comme les BDTs qui augmentent l’efficacité et la pureté de l’échantillon SNIa. Cette nouvelle classification photométrique SN-redshift fournit un échantillon des SNe type Ia avec une contamination de moins de 10 % selon

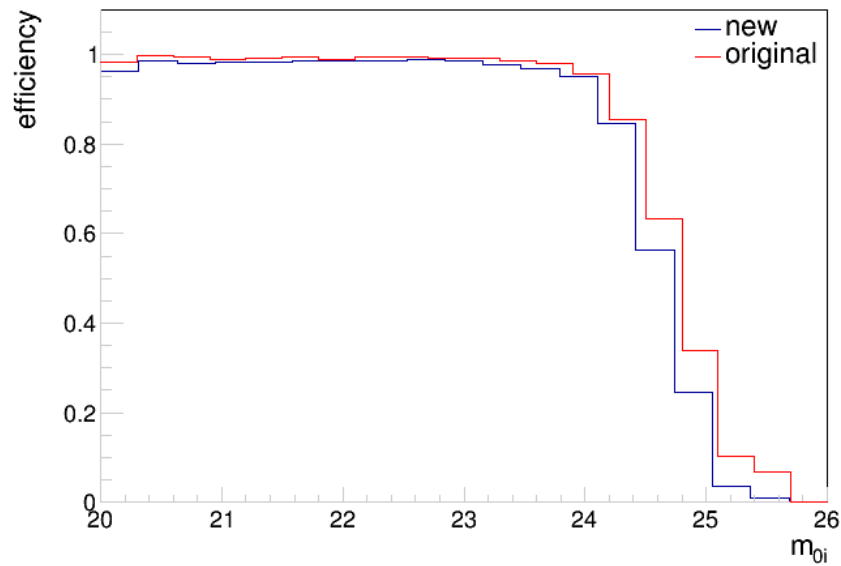


FIGURE A.4 Efficacité de détection en fonction de la magnitude maximale générée dans le filtre  $i_M$ . L'ancienne procédure en rouge est à comparer avec la nouvelle en bleu.

les études de Monte-Carlo. L'efficacité à bas redshifts pour cette méthode est meilleure que celle de l'analyse SNLS3 (Fig. A.5).

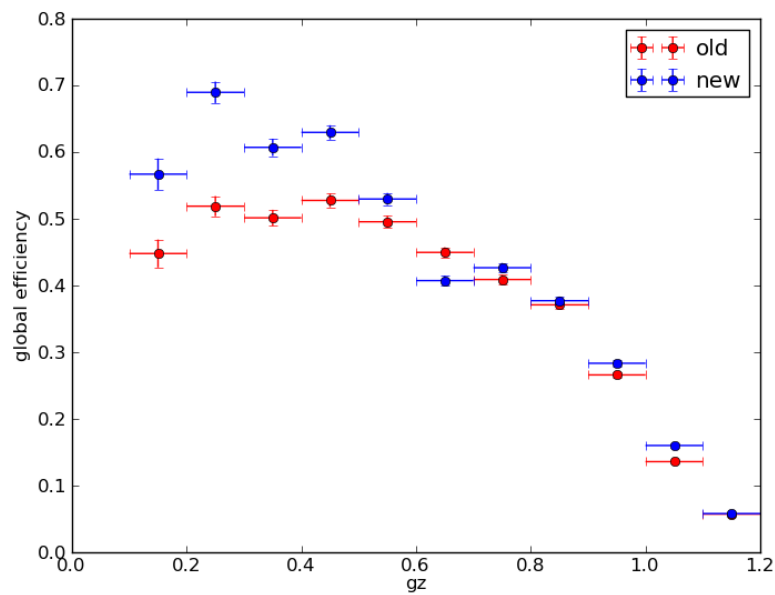


FIGURE A.5 Comparaison des efficacités globales en fonction des redshifts générées pour la procédure de SNLS3 (rouge) et la nouvelle méthode utilisant des redshifts photométriques déterminées à partir des courbes de lumière des supernovae elles-mêmes.

L'étape suivante, avec les outils prêts et les critères testés sur les données des 3 années,

---

est d'appliquer l'analyse à l'ensemble des données de SNLS et le lot photométrique obtenu sera utilisé pour déterminer les paramètres cosmologiques (équation d'état de l'énergie noire et densité de matière) et évaluer les systématiques propres à la sélection photométrique (contamination, reconstruction photométrique du décalage spectral).

# Bibliography

- [1] Steven Weinberg. *Gravitation and Cosmology*. John Wiley and Sons, US, 1st edition, 1972. ISBN 0-471-92567-5.
- [2] P.A.R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. 2015.
- [3] John A. Peacock, Shaun Cole, Peder Norberg, Carlton M. Baugh, Joss Bland-Hawthorn, et al. A Measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410:169–173, 2001. doi: 10.1038/35065528.
- [4] Sdss galaxy map. [http://classic.sdss.org/includes/sideimages/sdss\\_pie2.html](http://classic.sdss.org/includes/sideimages/sdss_pie2.html), . Accessed: 2015-03-13.
- [5] D.N. Spergel et al. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys.J.Suppl.*, 148: 175–194, 2003. doi: 10.1086/377226.
- [6] E. Komatsu et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys.J.Suppl.*, 192:18, 2011. doi: 10.1088/0067-0049/192/2/18.
- [7] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proc.Nat.Acad.Sci.*, 15:168–173, 1929. doi: 10.1073/pnas.15.3.168.
- [8] Adam G. Riess, Lucas Macri, Stefano Casertano, Hubert Lampeitl, Henry C. Ferguson, et al. A 3Space Telescope and Wide Field Camera 3. *Astrophys.J.*, 730:119, 2011. doi: 10.1088/0004-637X/732/2/129,10.1088/0004-637X/730/2/119.
- [9] C.L. Bennett, D. Larson, J.L. Weiland, and G. Hinshaw. The 1% Concordance Hubble Constant. *Astrophys.J.*, 794:135, 2014. doi: 10.1088/0004-637X/794/2/135.



- [10] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv.Phys.Acta*, 6: 110–127, 1933.
- [11] M. Betoule et al. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astron.Astrophys.*, 568:A22, 2014. doi: 10.1051/0004-6361/201423413.
- [12] Robert R. Caldwell and Marc Kamionkowski. The Physics of Cosmic Acceleration. *Ann.Rev.Nucl.Part.Sci.*, 59:397–429, 2009. doi: 10.1146/annurev-nucl-010709-151330.
- [13] S. Perlmutter et al. Measurements of Omega and Lambda from 42 high redshift supernovae. *Astrophys.J.*, 517:565–586, 1999. doi: 10.1086/307221.
- [14] Brian P. Schmidt et al. The High Z supernova search: Measuring cosmic deceleration and global curvature of the universe using type Ia supernovae. *Astrophys.J.*, 507:46–63, 1998. doi: 10.1086/306308.
- [15] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron.J.*, 116:1009–1038, 1998. doi: 10.1086/300499.
- [16] G.R. Dvali, Gregory Gabadadze, and Massimo Porrati. 4-D gravity on a brane in 5-D Minkowski space. *Phys.Lett.*, B485:208–214, 2000. doi: 10.1016/S0370-2693(00)00669-9.
- [17] A. Conley, J. Guy, M. Sullivan, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchett, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, M. J. Hudson, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, J. D. Neill, S. Perlmutter, P. Ripoche, N. Suzuki, and E. S. Walker. Supernova Constraints and Systematic Uncertainties from the First Three Years of the Supernova Legacy Survey. *APJS*, 192:1, January 2011. doi: 10.1088/0067-0049/192/1/1.
- [18] Alexei V. Filippenko. Optical spectra of supernovae. *Ann.Rev.Astron.Astrophys.*, 35:309–355, 1997. doi: 10.1146/annurev.astro.35.1.309.

- [19] Phd thesis baumont. universite paris diderot 7. <https://tel.archives-ouvertes.fr/tel-00196934>, . Accessed: 2015-01-19.
- [20] M.E. Salvo, E. Cappellaro, P.A. Mazzali, S. Benetti, I.J. Danziger, et al. The template type ia supernova 1996x. *Mon.Not.Roy.Astron.Soc.*, 321:254, 2001. doi: 10.1046/j.1365-8711.2001.03995.x.
- [21] F. Patat, S. Benetti, E. Cappellaro, I. J. Danziger, M. della Valle, P. A. Mazzali, and M. Turatto. The type IA supernova 1994D in NGC 4526: the early phases. *MNRAS*, 278:111–124, January 1996.
- [22] David Branch, E. Baron, and David J. Jeffery. Optical spectra of supernovae. 2001.
- [23] Type 1a line identifications. . Accessed: 2015-01-20.
- [24] Type ia lowredshift lc. . Accessed: 2015-01-20.
- [25] M.M. Phillips. The absolute magnitudes of Type IA supernovae. *Astrophys.J.*, 413: L105–L108, 1993. doi: 10.1086/186970.
- [26] R. Tripp and D. Branch. Determination of the hubble constant using a two-parameter luminosity correction for type ia supernovae. *Astrophys.J.*, 525:209–214, 1999. doi: 10.1086/307883.
- [27] P Astier, J Guy, N Regnault, R Pain, E Aubourg, D Balam, S Basa, R G Carlberg, S Fabbro, D Fouchez, I M Hook, D A Howell, H Lafoux, J D Neill, N Palanque-Delabrouille, K Perrett, C J Pritchett, J Rich, M Sullivan, R Taillet, G Aldering, P Antilogus, V Arsenijevic, C Balland, S Baumont, J Bronder, H Courtois, R S Ellis, M Filiol, A C Goncalves, A Goobar, D Guide, D Hardin, V Lisset, C Lidman, R McMahon, M Mouchet, A Mourao, S Perlmutter, P Riposte, C Tao, and N Walton. The Supernova Legacy Survey: Measurement of  $\Omega_M$ ,  $\Omega_{\Lambda}$  and  $w$  from the First Year Data Set. *Astronomy & Astrophysics*, 447:31–48, October 2006.
- [28] Saurabh Jha, Adam G. Riess, and Robert P. Kirshner. Improved Distances to Type Ia Supernovae with Multicolor Light Curve Shapes: MLCS2k2. *Astrophys.J.*, 659: 122–148, 2007. doi: 10.1086/512054.

- [29] Julien Guy et al. SALT2: Using distant supernovae to improve the use of Type Ia supernovae as distance indicators. *Astron.Astrophys.*, 466:11–21, 2007. doi: 10.1051/0004-6361:20066930.
- [30] M. Sullivan et al. The Dependence of Type Ia Supernova Luminosities on their Host Galaxies. *Mon.Not.Roy.Astron.Soc.*, 406:782–802, 2010.
- [31] W BAADE and F ZWICKY. On Super-novae. In *Proceedings of the National Academy of Sciences of the United States of America*, pages 254–259, May 1934.
- [32] F ZWICKY. Some Results of the Search for Super-Novae. *Physical Review*, 53(1): 1019–1020, June 1938.
- [33] J. Melinder, T. Dahlen, L. Mencia Trinchant, G. Ostlin, S. Mattila, et al. The Rate of Supernovae at Redshift 0.1-1.0 - the Stockholm VIMOS Supernova Survey IV. *Astron.Astrophys.*, 545:A96, 2012. doi: 10.1051/0004-6361/201219364.
- [34] Christopher J. Pritchett, D. Andrew Howell, and Mark Sullivan. The Progenitors of Type Ia Supernovae. *Astrophys.J.*, 683:L25, 2008. doi: 10.1086/591314.
- [35] G. Bazin et al. The Core-collapse rate from the Supernova Legacy Survey. *Astron.Astrophys.*, 499:653, 2009. doi: 10.1051/0004-6361/200911847.
- [36] Evan Scannapieco and Lars Bildsten. The Type IA supernova rate. *Astrophys.J.*, 629:L85–L88, 2005. doi: 10.1086/452632.
- [37] Mark Sullivan et al. Rates and properties of type Ia supernovae as a function of mass and star-formation in their host galaxies. *Astrophys.J.*, 648:868–883, 2006. doi: 10.1086/506137.
- [38] James D. Neill et al. The Type Ia Supernova Rate at  $z$  0.5 from the Supernova Legacy Survey. *Astron.J.*, 132:1126–1145, 2006. doi: 10.1086/505532.
- [39] James D. Neill, M. Sullivan, D. Balam, C.J. Pritchett, D.A. Howell, et al. The Supernova Type Ia Rate Evolution with SNLS. *AIP Conf.Proc.*, 924:421–424, 2007. doi: 10.1063/1.2774890.
- [40] Dan Maoz and Filippo Mannucci. Type-Ia supernova rates and the progenitor problem, a review. *Publ.Astron.Soc.Austral.*, 29:447, 2012. doi: 10.1071/AS11052.

- [41] Douglas C. Leonard. Constraining the Type Ia Supernova Progenitor: The Search for Hydrogen in Nebular Spectra. *Astrophys.J.*, 670:1275, 2007. doi: 10.1086/522367.
- [42] Bradley E. Schaefer and Ashley Pagnotta. The center of the Type Ia supernova remnant SNR 0509-67.5 is empty of any ex-companion star to  $M_V = +8.4$ . 2012.
- [43] Jifeng Liu, Rosanne Di Stefano, Tao Wang, and Maxwell Moe. On the Nature of the Progenitor of the Type Ia SN2011fe in M101. *Astrophys.J.*, 749:141, 2012. doi: 10.1088/0004-637X/749/2/141.
- [44] Ken J. Shen, Lars Bildsten, Daniel Kasen, and Eliot Quataert. The Long-Term Evolution of Double White Dwarf Mergers. *Astrophys. J.*, 748:35, 2012. doi: 10.1088/0004-637X/748/1/35.
- [45] S. Darbha, B. D. Metzger, E. Quataert, D. Kasen, P. Nugent, and R. Thomas. Nickel-Rich Outflows Produced by the Accretion-Induced Collapse of White Dwarfs: Lightcurves and Spectra. *Mon. Not. Roy. Astron. Soc.*, 409:846, 2010. doi: 10.1111/j.1365-2966.2010.17353.x.
- [46] Marten H. van Kerkwijk, Philip Chang, and Stephen Justham. Sub-Chandrasekhar White Dwarf Mergers as the Progenitors of Type Ia Supernovae. *Astrophys. J.*, 722:L157–L161, 2010. doi: 10.1088/2041-8205/722/2/L157.
- [47] James Guillochon, Marius Dan, Enrico Ramirez-Ruiz, and Stephan Rosswog. Surface Detonations in Double Degenerate Binary Systems Triggered by Accretion Stream Instabilities. *Astrophys. J.*, 709:L64, 2010. doi: 10.1088/2041-8205/709/1/L64.
- [48] J. S. W. Claeys, O. R. Pols, R. G. Izzard, J. Vink, and F. W. M. Verbunt. Theoretical uncertainties of the Type Ia supernova rate. *Astronomy and Astrophysics*, 563:A83, March 2014. doi: 10.1051/0004-6361/201322714.
- [49] N. Palanque-Delabrouille, V. Ruhlmann-Kleider, S. Pascal, J. Rich, J. Guy, et al. Photometric redshifts for supernovae Ia in the Supernova Legacy Survey. *Astron.Astrophys.*, 514:A63, 2010. doi: 10.1051/0004-6361/200913283.
- [50] Cfht dome. <http://www.cfht.hawaii.edu/en/news/CFHT30/>, . Accessed: 2014-12-29.

- [51] Cfht megaprime. <http://www.cfht.hawaii.edu/Instruments/Imaging/MegaPrime/megaprimecomponents.html>, . Accessed: 2014-12-29.
- [52] Olivier Boulade, Xavier Charlot, P. Abbon, Stephan Aune, Pierre Borgeaud, Pierre-Henri Carton, M. Carty, J. Da Costa, H. Deschamps, D. Desforge, Dominique Eppelle, Pascal Gallais, L. Gosset, Remy Granelli, Michel Gros, Jean de Kat, Denis Loiseau, J. . Ritou, Jean Y. Rousse, Pierre Starzynski, Nicolas Vignal, and Laurent G. Vigroux. Megacam: the new canada-france-hawaii telescope wide-field imaging camera. *Proc. SPIE*, 4841:72–81, 2003. doi: 10.1117/12.459890. URL <http://dx.doi.org/10.1117/12.459890>.
- [53] Megacam. [http://irfu.cea.fr/Sap/Phoce/Vie\\_des\\_labos/Ast/ast\\_technique.php?id\\_ast=2298&onglet=1306](http://irfu.cea.fr/Sap/Phoce/Vie_des_labos/Ast/ast_technique.php?id_ast=2298&onglet=1306), . Accessed: 2014-12-29.
- [54] Arrangement and original numbering scheme of the ccd mosaic. a and b correspond to the 2 amplifiers used during readout. <http://terapix.iap.fr/cplt/T0007/doc/T0007-doc.html>, . Accessed: 2014-12-29.
- [55] Megacam specifications. <http://www.cfht.hawaii.edu/Instruments/Imaging/MegaCam/generalinformation.html>. Accessed: 2015-01-03.
- [56] M Betoule, J Marriner, N Regnault, J C Cuillandre, P Astier, J Guy, C Balland, P El Hage, D Hardin, R Kessler, L Le Guillou, J Mosher, R Pain, P F Rocci, M Sako, and K Schahmaneche. Improved Photometric Calibration of the SNLS and the SDSS Supernova Surveys. *arXiv.org*, December 2012.
- [57] Gajus Miknaitis, G. Pignata, A. Rest, W.M. Wood-Vasey, S. Blondin, et al. The ESSENCE Supernova Survey: Survey Optimization, Observations, and Supernova Photometry. *Astrophys.J.*, 666:674–693, 2007. doi: 10.1086/519986.
- [58] Masao Sako et al. The Data Release of the Sloan Digital Sky Survey-II Supernova Survey. 2014.
- [59] Megacam fields. <http://www.cfht.hawaii.edu/Science/CFHLS/cfhlsdeepwidefields.html>, . Accessed: 2015-01-03.
- [60] N Regnault, A Conley, J Guy, M Sullivan, J C Cuillandre, P Astier, C Balland, S Basa, R G Carlberg, D Fouchez, D Hardin, I M Hook, D A Howell, R Pain,

- K Perrett, and C J Pritchett. Photometric calibration of the Supernova Legacy Survey fields. *Astronomy & Astrophysics*, 506(2):999–1042, November 2009.
- [61] K. Perrett et al. Real-time Analysis and Selection Biases in the Supernova Legacy Survey. *Astron.J.*, 140:518–532, 2010. doi: 10.1088/0004-6256/140/2/518.
- [62] E.A. Magnier and J.-C. Cuillandre. . *PASP*, 116(449), 2004.
- [63] C. Alard and R.H. Lupton. . *ApJ*, 503(325), 1998.
- [64] J Guy, M Sullivan, A Conley, N Regnault, P Astier, C Balland, S Basa, R G Carlberg, D Fouchez, D Hardin, I M Hook, D A Howell, R Pain, N Palanque-Delabrouille, K M Perrett, C J Pritchett, J Rich, V Ruhlmann-Kleider, D Balam, S Baumont, R S Ellis, S Fabbro, H K Fakhouri, N Fourmanoit, S Gonzalez-Gaitan, M L Graham, E Hsiao, T Kronborg, C Lidman, A M Mourao, S Perlmutter, P Ripoche, N Suzuki, and E S Walker. The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *Astronomy & Astrophysics*, 523, 2010.
- [65] M Betoule, R Kessler, J Guy, J Mosher, D Hardin, R Biswas, P Astier, P El Hage, M Konig, S Kuhlmann, J Marriner, R Pain, N Regnault, C Balland, B A Bassett, P J Brown, H Campbell, R G Carlberg, F Cellier-Holzem, D Cinabro, A Conley, C B D’Andrea, D L DePoy, M Doi, R S Ellis, S Fabbro, A V Filippenko, R J Foley, J A Frieman, D Fouchez, L Galbany, A Goobar, R R Gupta, G J Hill, R Hlozek, C J Hogan, I M Hook, D A Howell, S W Jha, L Le Guillou, G Leloudas, C Lidman, J L Marshall, A Möller, A M Mourao, J Neveu, R Nichol, M D Olmstead, N Palanque-Delabrouille, S Perlmutter, J L Prieto, C J Pritchett, M Richmond, A G Riess, V Ruhlmann-Kleider, M Sako, K Schahmaneche, D P Schneider, M Smith, J Sollerman, M Sullivan, N A Walton, and C J Wheeler. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astronomy & Astrophysics*, 568, January 2014.
- [66] G Bazin, V Ruhlmann-Kleider, N Palanque-Delabrouille, J Rich, E Aubourg, P Astier, C Balland, S Basa, R G Carlberg, A Conley, D Fouchez, J Guy, D Hardin, I M Hook, D A Howell, R Pain, K Perrett, C J Pritchett, N Regnault, M Sullivan, N Fourmanoit, S Gonzalez-Gaitan, C Lidman, S Perlmutter, P Ripoche, and E S

- Walker. Photometric selection of Type Ia supernovae in the Supernova Legacy Survey. *Astronomy & Astrophysics*, 534:A43, September 2011.
- [67] Gurvan Bazin. *PhD thesis: Analyse différée des données du SuperNova Legacy Survey*. Theses, Université Paris Diderot - Paris VII, September 2008.
- [68] E. Bertin. Automatic Astrometric and Photometric Calibration with SCAMP. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 112, July 2006.
- [69] D. G. Monet, S. E. Levine, B. Canzian, H. D. Ables, A. R. Bird, C. C. Dahn, H. H. Guetter, H. C. Harris, A. A. Henden, S. K. Leggett, H. F. Levison, C. B. Luginbuhl, J. Martini, A. K. B. Monet, J. A. Munn, J. R. Pier, A. R. Rhodes, B. Riepe, S. Sell, R. C. Stone, F. J. Vrba, R. L. Walker, G. Westerhout, R. J. Brucato, I. N. Reid, W. Schoening, M. Hartley, M. A. Read, and S. B. Tritton. The USNO-B Catalog. , 125:984–993, February 2003. doi: 10.1086/345888.
- [70] E. Bertin, Y. Mellier, M. Radovich, G. Missonnier, P. Didelon, and B. Morin. The TERAPIX Pipeline. In D. A. Bohlender, D. Durand, and T. H. Handley, editors, *Astronomical Data Analysis Software and Systems XI*, volume 281 of *Astronomical Society of the Pacific Conference Series*, page 228, 2002.
- [71] Laurent Le Guillou. *PhD thesis: Recherche de matière noire galactique par détection de microlentilles gravitationnelles en photométrie différentielle*. Theses, Université Pierre et Marie Curie - Paris VI, September 2003. URL <https://tel.archives-ouvertes.fr/tel-00003907>. Jury: Eric Aubourg (directeur de thèse) Jacques Chauveau (président) Yannick Giraud-Héraud (rapporteur) Yannick Mellier (rapporteur) Joseph Silk Daniel Vignaud.
- [72] C. Alard and R.H. Lupton. A Method for optimal image subtraction. *Astrophys.J.*, 503:325, 1998. doi: 10.1086/305984.
- [73] E Bertin and S Arnouts. SExtractor: Software for source extraction. *Astronomy & Astrophysics Supplement Series*, 117(2):393–404, June 1996.
- [74] P Ripoche. *PhD thesis, Université de Marseille*. 2007.

- [75] O Ilbert, S Arnouts, and H.J. McCracken. Ilbert. *Astronomy & Astrophysics*, 457 (841I), September 2006.
- [76] N. Regnault et al. Photometric Calibration of the Supernova Legacy Survey Fields. *Astron.Astrophys.*, 506:999–1042, 2009. doi: 10.1051/0004-6361/200912446.
- [77] M. Betoule et al. Improved Photometric Calibration of the SNLS and the SDSS Supernova Surveys. *Astron.Astrophys.*, 552:A124, 2013. doi: 10.1051/0004-6361/201220610.
- [78] Jean-Luc Starck, Michael Elad, and David Donoho. Redundant multiscale transforms and their application for morphological component separation. *Adv.Imaging Electron Phys.*, 2004.
- [79] JL Starck, F Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge Univ. Press, 1998.
- [80] Stphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008. ISBN 0123743702, 9780123743701.
- [81] Simon Beckouche, Jean-Luc Starck, and Jalal Fadili. Astronomical Image Denoising Using Dictionary Learning. *Astron.Astrophys.*, 556:A132, 2013. doi: 10.1051/0004-6361/201220752.
- [82] Jeremy Schmitt, Jean-Luc Starck, Jean-Marc Casandjian, Jalal Fadili, and Isabelle Grenier. Multichannel Poisson denoising and deconvolution on the sphere : Application to the Fermi Gamma Ray Space Telescope. *Astron.Astrophys.*, 546:A114, 2012. doi: 10.1051/0004-6361/201118234.
- [83] J. Bobin, J.-L. Starck, F. Sureau, and S. Basak. Sparse component separation for accurate CMB map estimation. *Astron.Astrophys.*, 550:73, 2013. doi: 10.1051/0004-6361/201219781.
- [84] A. Leonard, F. Lanusse, and J.-L. Starck. GLIMPSE: Accurate 3D weak lensing reconstructions using sparsity. *Mon.Not.Roy.Astron.Soc.*, 440:1281, 2014. doi: 10.1093/mnras/stu273.
- [85] J.-L. Starck, M. Elad, and D.L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans.Image Process.*, 14:1570–1582, 2005. doi: 10.1109/TIP.2005.852206.



- 
- [86] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, New York, NY, USA, 2010. ISBN 0521119138, 9780521119139.
- [87] Jean-Luc Starck and Jerome Bobin. *Astronomical Data Analysis and Sparsity: from Wavelets to Compressed Sensing*. 2009.
- [88] A. Möller, V. Ruhlmann-Kleider, F. Lanusse, J. Neveu, N. Palanque-Delabrouille, and J.-L. Starck. Snia detection in the snls photometric analysis using morphological component analysis. *Journal of Cosmology and Astroparticle Physics*, 2015(04):041, 2015. URL <http://stacks.iop.org/1475-7516/2015/i=04/a=041>.
- [89] G Bazin. *PhD thesis*. Université Paris Diderot, 2008.
- [90] Christopher M. Carroll, Eric Gawiser, Peter L. Kurczynski, Rachel A. Bailey, Rahul Biswas, et al. Improving the LSST dithering pattern and cadence for dark energy studies. *Proc.SPIE Int.Soc.Opt.Eng.*, 9149:91490C, 2014. doi: 10.1117/12.2057267.
- [91] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370, 9781461471370.
- [92] Andreas Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, et al. TMVA - Toolkit for Multivariate Data Analysis. *PoS, ACAT:040*, 2007.
- [93] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

## *Acknowledgements*

After three years of a PhD, there are many people to thank, many shared memories and a lot of support.

First, I would like to thank my supervisor Vanina Ruhlmann-Kleider for guiding me through this work. Discussing in English made easier the comprehension of all the subtleties of the pipeline and all the processes behind. Thanks for accepting, at the end, all my daring ideas. We encountered many obstacles but we always figured a way out.

Thanks to the people of SPP and the cosmology group. Special thanks to Christophe M. that helped me at the beginning with programming bugs, Jean-Baptiste and Laurent C. for attending my rehearsal, Nathalie and Jim from the SNLS group. Also, the PhD and post-docs from the group with whom I shared many interesting science and non-science discussions Jeremy, Timothée, Graziano, Pierre, Helion and many others. Special thanks to those in the Service who smiled and had always a positive attitude: Marie-Cecile, Martine O. and Fouzia.

My first steps doing science in France have to be thanked to José Ocariz por darme el empujón para venir a Paris. Thanks to the people of the Atlas group in LPNHE and to my LPNHE friends, specially Guillaume Lefebvre for your support.

Of course my “co-buros” who saw and shared the ups and downs of my day-to-day. Merci Romain Kukla pour reciter Racine avec Horacio et pour la compagnie creole! Pour rapporter la provence à Paris et le tropical “caliente”! Homero Martínez, por los cafecitos y por todas las “soirées” en París.

To my friends from the master NPAC. It has been 4 amazing years of friendship and support! Merci Cyril Becot et Francois Lanusse pour les randos et le ski avec plein des dangers. Alexis Helou, Marta Spinella, Vivien Scottez, Camille Moretto pour les soirees GoT et les petites voyages. Vincent Fischer pour les repas NPAC les mercredis et le support dans le labo. Et encore merci Francois pour notre travail ensemble!

Of course, I have to thank those Latin American friends who have become my family abroad. Thanks for reminding me of the “real life” and celebrating holidays together. To la “familia”: Gustavo González, Apsara Sánchez, Elena Mylonas y Laura Gabriela González. Of course: los Ticos, Homero Martínez and the girls of Quedada. Camila Rangel, Rebeca Ribeiro and Reina Camacho your friendship helped me a lot through the tough moments and we shared amazing times together! Thanks for all the talks, advises and travels together. Also, thanks to Thibault for all the good memories and support.

Thanks to my friends that came to the defense from many places in the world: Camila, Harvey, Timothée, Elena and Homero. And to those closer by that came even though they are far from physics!

Ce trois années étaient bien enrichies par mes activités dans l'association voile du CEA. Ça était une expérience magnifique naviguer avec vous et j'ai hâte de retourner sur Neutron Rapide comme skipper un jour!

Gracias a la Universidad Simón Bolívar por la educación de calidad que me brindó. Es uno de los mejores pedazos de Venezuela que llevo conmigo. Agradezco mucho a los profesores que me transmitieron su pasión por la ciencia.

Por supuesto quisiera agradecer a mi familia que, aunque no siempre entendieron por qué me cambié a física, estuvieron allí estos años y me apoyaron. Gracias por incentivar mi curiosidad desde chiquita. A mi mamá, a mi papá y a mis hermanos gracias por cocinar todas las especialidades venezolanas del "pot". Especiales "supernovas" a Bruno Isidro e Isa por hacer bromas con el vocabulario científico y construir frases sin sentido que me hacen reír.