Centro Brasileiro de Pesquisas Físicas

Gabriel da Silva Moreira Teixeira

From Large-Scale Structure to Time-Domain Astronomy: Combining Artificial Intelligence and Astrophysical Analysis

Rio de Janeiro

2026

"FROM LARGE-SCALE STRUCTURE TO TIME-DORMAIN ASTRONOMY: COMBINING ARTIFICIAL INTELLIGENCE AND ASTROPHYSICAL ANALYSIS"

GABRIEL DA SILVA MOREIRA TEIXEIRA

Tese de Doutorado em Física apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência Tecnologia e Inovação. Fazendo parte da banca examinadora os seguintes professores:

Clécio Roque de Bom - Orientador/CBPF

gov.br Documento assinado digitalmente
MARIANA PENNA LIMA VITENTI
Data: 19/05/2026 14:13:25-0300
Verifique em https://validar.iti.gov.br

Mariana Penna Lima Vitenti – UnB

Arianna Cortesi - UFRJ

gov.br Documento assinado digitalmente
REINALDO ROBERTO ROSA
Data: 19/06/2026 11:19:12-0300
Verifique em https://validar.iti.gov.br

Reinaldo Roberto Rosa - INPE

Karín Menendez-Delmestre – CBPF

Rio de Janeiro, 14 de maio de 2026.

Gabriel da Silva Moreira Teixeira

From Large-Scale Structure to Time-Domain Astronomy: Combining Artificial Intelligence and Astrophysical Analysis

Versão final

Tese de Doutorado.

Orientador: Clécio Roque de Bom.

Rio de Janeiro
2026

À minha mãe, Claudia, e à memória de meu pai, Eroslanzio.

Agradecimentos

Enfim escrevendo uma das seções mais difíceis desse documento, não pela complexidade do conteúdo em si, mas sim pelo fato de que qualquer coisa que eu escreva aqui não passará de uma mera aproximação, uma tentativa falha de de fato expressar a gratidão que realmente sinto. Essa tese marca o fim de uma longa jornada, jornada esta que só foi possível graças ao apoio de muitos, e a estes gostaria de expressar minha gratidão.

Em primeiro lugar – afinal, ao pensar em motivos para ser grato, ela é a primeira pessoa que vem à minha mente –, gostaria de agradecer à minha mãe, Claudia. A ela não seria exagero dizer "obrigado por tudo", pois a ela tudo devo. Foi quem sempre lutou pela minha educação e a de meus irmãos, foi quem sempre fez de tudo para que trilhássemos o caminho do conhecimento com o mínimo de preocupações possível, foi quem, acima de tudo, me ensinou o que é ter prazer em aprender. Obrigado, mãe, pois é por você, graças a você, e em grande parte para você que cheguei até aqui.

Infelizmente, ele não está mais entre nós para ler. Fiz questão de que ele soubesse disso em vida, sempre que pude, mas, mesmo assim, gostaria de deixar expressa a minha gratidão a meu pai, Eroslanzio. A ele agradeço por, juntamente com a minha mãe, não ter poupado esforços quando o assunto era cuidado, e em particular educação. Agradeço a ele por ter me ensinado, acima de tudo, o significado de gentileza e amor. Agradeço por ter sido o maior exemplo de homem que já vi. E foi por ter seguido esse exemplo que pude chegar aqui. Obrigado, pai e mãe, por me mostrarem o caminho, por sempre me apoiarem em tudo que eu fizesse, por sempre acreditarem, por me ensinarem o que é amor. Não há título acadêmico que me faça sentir maior orgulho do que o orgulho que tenho de ser filho de Claudia e Eroslanzio.

Agradeço aos meus irmãos Robson, Davi e Maria Eduarda (Tereza) por sempre me apoiarem. Obrigado por estarem com nossos pais, por cuidarem deles em momentos – alguns deles muito difíceis – em que eu não pude estar, por conta dos estudos. Obrigado por sempre me receberem de braços abertos, com muito amor, e por sempre acreditarem em mim e me apoiarem. Ao mais velho, Robson, obrigado por ser exemplo. Aos mais novos, Davi e Maria, tenho muito orgulho das pessoas que vocês estão se tornando; obrigado por serem aqueles que me fazem olhar com cuidado para os caminhos que trilho e para o exemplo que me torno para vocês.

Agradeço à minha família por me mostrar o significado de superação, por me ensinarem que o lugar de onde viemos e as dificuldades que passamos de forma alguma definem nosso futuro. Obrigado pelo exemplo de cuidado e união, obrigado por sempre me apoiarem nessa jornada, por me darem a certeza de que sempre terei um lugar para voltar. Obrigado aos meus padrinhos, Eroslácio e Maria. Obrigado aos meus tios, Amarildo, Claudete e Mariza. Obrigado aos meus primos Alexandre, Carina, Carlos Henrique, Daniela, Gisele, Giorgia, Isabela, Lívia, Maira e Sabrina. Obrigado aos mais novos, Daniel, Gustavo, Heitor, Laura, Luiz Gustavo, Renan e Zayla por me lembrarem do quão

intrigante é o novo, o desconhecido ao olhar de uma criança, por evocarem a criança curiosa que há em mim, lembrando do fascínio pelo mundo e pela natureza que me fez seguir o caminho da ciência.

Obrigado à minha família de Campo Mourão, por toda a torcida, apoio, suporte e cuidado que tiveram para comigo e com meus pais e irmãos. Obrigado Márcia, Marcelly, Keminha, Hevellyn, Hemilly, Joel e Edna.

Agradeço também àqueles que me guiaram no caminho do conhecimento, aos professores, mestres e mentores que tive ao longo da trajetória. Todos desempenharam papel crucial na minha formação, mas gostaria de agradecer especialmente alguns que marcaram a minha vida, servindo de exemplo e fazendo eu me apaixonar pela docencia. Portanto que fique expressa minha gratidão a Cristiane Alves, Elaine Bobeda, Rosangela Lannes, Paulo Farias, William Oliveira, Sandra Amato, Simone Coutinho, Thiago Hartz, Rodrigo Sacramento e Carlos Farina.

Among those who taught me, some played a fundamental role in mentoring me throughout my PhD. I thank Charles Kilpatrick for introducing me to the wonder world of Type II supernovae, for teaching me the fundamentals of supernova and transient science, for all the beers and pizzas in Chicago, and for his guidance throughout the analysis of SN 2022acko.

Agradeço a Emille Ishida por ter me proporcionado a experiência mais transformadora da minha formação, por ter me mostrado uma visão da carreira que me espera que vai muito além de técnicas: envolve respeito e empatia, não só com os outros, mas principalmente conosco mesmos e com os mais novos. Obrigado, Emille, por todas as discussões sobre métodos estatísticos, por toda a ajuda com as aplicações para pós-doutorado, e por ter me proporcionado tantas oportunidades de entrar em contato com instituições, pessoas e culturas diferentes – obrigado por ser a tamanha referência e inspiração que você é para mim.

E, claro, não poderia deixar de mencionar o mentor máximo, Clécio de Bom. Obrigado, Clécio, por todos esses anos de orientação, por ter me formado o cientista que sou, por ter me proporcionado oportunidades que mudaram minha vida, por ter me ensinado toda a base do que sei de astronomia, por sua paciência ao longo desses anos, por todo o empenho em tornar nosso ambiente de trabalho o mais confortável possível, e por estar sempre pensando nas melhores condições e oportunidades para nós. Enfim, a lista poderia continuar *ad eternum*, mas, acima de tudo, agradeço por ter acreditado em mim, simples assim. Agradeço por, no momento em que eu estava prestes a pedir para largar a IC, você ter dito que via algo em mim e ter corrido atrás de uma bolsa. Espero que esta tese represente parte desse algo que você viu. A você, serei eternamente grato.

Dado que um dos intuitos desta tese é demonstrar certa proficiência em determinadas habilidades, já começo por aqui ressaltando algo em que sou muito bom: escolher amigos. Gostaria de agradecer a todos aqueles que posso chamar de amigos.

Agradeço, primeiramente, à minha melhor amiga, que por coincidência também é minha namorada, Juliana. Muito obrigado por me aturar todos esses anos, por me mostrar todos os dias o que é o amor, pela paciência comigo (particularmente durante os momentos de alto estresse acadêmico), e pelo companheirismo. Obrigado por ser você. Eu te amo!

Agradeço aos meus amigos de longa data, Claudiana, João Pedro e Rafaela, por me aturarem e sempre me motivarem desde que tínhamos 10∼13 anos de idade, e por estarem comigo até aqui. Agradeço aos meus amigos da turma do ensino médio, Beatriz Chagas, Gabriela, Gabrielle, Joyce, Manu e Richard por toda a parceria e tolerância durante nosso tempo juntos (e até hoje), por sempre acreditarem em mim, por torcerem pela minha trajetória e pelo esforço para nos encontrarmos como turma anualmente. Agradeço ao Guilherme por muito me ter ensinado sobre a vida, por ser o irmão que é pra mim, por sempre me receber com muita alegria, pelos encontros, reflexões e aprendizados que temos juntos até os dias de hoje. Agradeço a Glauco e João, estes que às vezes até esquecemos que foi no ensino médio que nos conhecemos – obrigado por todas as noites online de jogos, fofocas, desabafos e amparo ao longo de todos esses anos. Agradeço aos meus queridos amigos da República SNC, Abraão, Denise, João Paulo, Juliana, Karina, Lorena e Thais, que antes mesmo de eu saber que queria física foram base para a minha formação como indivíduo. Sou grato por todos os momentos que tivemos e por sempre estarem aqui por mim, independentemente do intervalo de tempo que passemos longe por conta das responsabilidades adultas. Dentre esses também está um dos irmãos que tive a sorte de encontrar pela vida, Jorge Luiz. Obrigado pelos anos de amizade regados de música, companheirismo e suporte.

Agradeço aos Amísicos Julia, Maria, Stalin, Viviane e Zedro, por terem feito da faculdade a melhor experiência possível, obrigado por serem exemplos para mim durante esse início de vida acadêmica e por continuarem sendo até hoje – e obrigado por toda a ajuda e paciência no início da faculdade quando enfrentei muitas dificuldades com o conteúdo. Não poderia deixar de agradecer à minha madrinha acadêmica Ursula, por todas as noites no zap fofocando desde o momento em que nos conhecemos, por todas as calls online que foram cruciais para a manutenção da sanidade mental, em particular durante a pandemia, por ser meu duo bot lane do terror pra vida, por estar sempre torcendo por mim, por ter me acolhido como afilhado acadêmico e por ter exercido o papel de melhor madrinha possível – obrigado por ser mai besto frinto. Obrigado ao meu amigo Yohan, por ter feito da minha estadia na França a melhor possível, por todas as noites de jogatina e reflexões sobre a vida, pela ajuda em Métodos II, e por todas as discussões acaloradas sobre assuntos nada relevantes que engrandecem o espírito. Obrigado aos meus amigos da pós-graduação, Alan, Isaque, Joaquim, Laís, Luiza, Mari, Sobrero, Valadão e Vanessa por todo o apoio de perto, todos os cafés, todos os rolês, e todas as conversas – obrigado por terem feito do meu doutorado o melhor possível, pela convivência com as

melhores pessoas possíveis. Obrigado, Edine e Eva, por serem as melhores companhias possíveis para se morar junto, por toda a torcida mútua pelo sucesso um do outro e por toda a parceria que nos uniu ao longo de 2025 – sou muitíssimo grato pela nossa amizade. Agradeço também a minha amiga Tefa, por todas as conversas motivacionais, reflexivas, e pelas muitas risadas que acompanham qualquer interação nossa.

Obrigado, Phelipe, por todos os rage chats, todos os ARAMs, por ouvir, sem exagero, todos os problemas pelos quais passei, por todas as conversas, todos os conselhos, todo o companheirismo, pelo golden year de 2025, enfim, obrigado por ser meu amigo e por ter feito desse o dream doc. E por fim, obrigado João Matheus, pela sua simples existência, por todas as ratiações, pela sua irmandade, por ser um lugar de refúgio e descanso em todos os momentos, obrigado por viver a vida comigo, obrigado por tudo, e saiba que depois de nós, é nós de novo.

Agradeço aos meus amigos e colaboradores de pesquisa. Primeiramente, agradeço aos que compõem (ou compuseram durante meu doutoramento) o grupo LAB-IA. Agradeço à Luciana por ter me ensinado as bases práticas de IA, por me ajudar a depurar meus primeiros códigos e por estar sempre disponível quando surgiam problemas. Ao André, pela amizade e parceria ao longo dessa árdua jornada que é o doutorado, pelas inúmeras conversas, desabafos e risadas, pelo trabalho na redução de dados e pela colaboração nos artigos de dark sirens e da SN 2022acko. À Viviane, pelos ensinamentos sobre dark sirens e cosmologia. Ao Bernardo, pela ajuda em IA e em tópicos gerais de astrofísica. Ao Luidhy, pelo apoio no desenvolvimento do artigo do DELVE. Ao Juan, pelo companheirismo como colega de sala, pelos memoráveis sanduíches de manteiga de amendoim, e pelos ensinamentos sobre colisão de buracos negros. A Marcos, Paulo e Thais, pelo trabalho contínuo na infraestrutura e na administração do laboratório. Por fim, agradeço a Mariana, Naomi e Pedro pelo apoio e pelas valiosas discussões científicas.

Agradeço também aos colaboradores do S-PLUS, em particular a Erik Lima, que me ensinou os fundamentos de redshifts fotométricos quando eu era apenas um aluno de iniciação científica, e continua me ajudando nessa área até hoje. Sou grato a Claudia Mendes de Oliveira e Laerte Sodré Jr. pelo acolhimento na colaboração e pelo incentivo constante; obrigado, Cláudia, pelo exemplo de liderança e por cultivar um ambiente tão acolhedor para jovens cientistas como o S-PLUS. Agradeço à Liliane pela parceiria em todas as iniciativas conjuntas como professores em escolas de IA/astrofísica e pelas conversas sempre muito produtivas sobre machine learning.

I thank the DELVE collaboration for their support in the development of the photo-z paper. I also thank the CHANCES collaboration for their patience with my limited background in galaxy clusters and galaxy evolution, and for teaching me so much about these topics. I am especially grateful to Hugo Méndez-Hernández, Ciria Lima, and Yara Jaffé for the many meetings and discussions that led to one of the works I am most proud of. I thank the Fink Broker community for all the exchanges and interactions

during meetings, which greatly contributed to the development of my projects. I also thank Marine Hebert, Emmanuel Gangler, Fabrice Jammes, Julien Peloton, and Anais Möller for being so welcoming and supportive to me in Fink.

Agradeço imensamente a todos que compõem o Centro Brasileiro de Pesquisas Físicas. Obrigado aos funcionários da administração, comunicação, limpeza, segurança e suporte técnico, pois é graças a estes que o CBPF pode oferecer uma infraestrutura de excelência e um ambiente de trabalho adequado para todos – algo que foi crucial para meu desenvolvimento. Obrigado a toda equipe da COEDU pelo suporte ao longo do doutorado, e por todos os financiamentos de viagem ao longo desses anos.

Agradeço em especial a Denise Coutinho, Sônia Ribeiro, Guilherme Vieira, Roberto Sarthour, Flávio Garcia por terem me integrado ao time de divulgação científica – parte fundamental da minha formação como cientista e comunicador científico. Obrigado por todo o apoio, pelas conversas acompanhadas de cerveja e churrasquinho no lava-jato, e pelas tantas oportunidades que me deram de viajar pelo Brasil divulgando ciência.

Quem me conhece sabe que a música exerce um papel mais que fundamental na minha vida. A música me acompanha desde o início de tudo e, definitivamente, tem grande influência na forma como vejo o mundo. Gostaria de expressar minha gratidão a essa forma de arte chamada música.

Não me delongarei agradecendo aos muitos estilos e artistas que me servem de referência, mas não poderia deixar de fazer menção ao REP nacional, que, para mim, é muito mais do que apenas um estilo de música. O REP, desde muito novo, me traz consciência do que é esperado como futuro para pessoas como eu, na idade que tenho; me trouxe – e ainda me traz – consciência sobre a relevância de ocupar os lugares que ocupo, e sobre o quanto as oportunidades que tive devem ser valorizadas.

Para além da música, o REP, para mim, é um lembrete de estar sempre olhando para aqueles que vêm de onde eu vim, para os que são como eu sou, para que as oportunidades que tive dependam cada vez menos de sorte e cheguem a essas pessoas. O REP me lembra que estar me doutorando agora não é só por mim, é pelos meus, é pelo meu bairro, Pedra de Guaratiba, e é para que eu chegue cada vez mais perto de contribuir para uma ciência brasileira mais inclusiva, diversa e democrática.

Que fique expressa também a minha gratidão a todos aqueles que lutaram pelo direito à educação para a população marginalizada e pelo sistema de cotas, que me permitiu ingressar na universidade. Obrigado àqueles que, por meio de sua luta, me permitiram estar aqui.

Agradeço a minha psicóloga, Iasmim Souza, por todo o cuidado ao longo desses anos de acompanhamento. Foi muito graças a você que surtos foram minimizados e alguns até evitados. Obrigado por todo o suporte nesse longo caminho que é o autodescobrimento.

Agradeço às agências de fomento do meu doutorado, CNPq (140212/2022-1; 200929/2025-9) e FAPERJ (02.432/2024), por proverem o subsídio necessário para me manter ao longo

desses anos de intensa dedicação à pesquisa.

"Não vai pra grupo, não, a cena é triste
Vamos estudar, respeitar o pai e a mãe e
viver, viver!
Essa é a cena"
Artigo 157 – Racionais MC's

"Somos poeira de estrela, castelos de vento
Por aí, vagando no espaço-tempo"
Poeira de estrela – Maruson

Resumo

Com o rápido aumento do volume de dados na astronomia moderna, métodos escaláveis para o processamento e a extração de informações físicas a partir de observações tornaram-se essenciais. Nesse contexto, a inteligência artificial tem emergido como uma poderosa abordagem orientada por dados. Nesta tese, desenvolvemos e aplicamos tais métodos a dois problemas distintos em astrofísica observacional: a estimação de redshifts fotométricos em levantamentos de campo amplo e a caracterização física de progenitores de supernovas por meio de observações multicomprimento de onda. Para isso, a tese está organizada em duas partes principais.

A primeira parte foca na estimação de redshifts fotométricos para o levantamento DECam Local Volume Exploration Survey. Desenvolvemos um arcabouço de aprendizado profundo que combina uma Rede Neural Recorrente com uma Mixture Density Network para inferir funções de densidade de probabilidade completas para os redshifts de galáxias ao longo da cobertura de $\sim 20{,}000\,\mathrm{deg}^2$ do levantamento. Nosso método apresenta desempenho competitivo em comparação com abordagens bem estabelecidas na literatura. Além disso, demonstramos o impacto dos redshifts fotométricos obtidos no contexto de ondas gravitacionais associadas a sirenes escuras, e em estudos de aglomerados de galáxias.

A segunda parte apresenta uma análise abrangente da supernova do tipo II-P SN 2022acko em NGC 1300, baseada em observações no ultravioleta, óptico e infravermelho, complementadas por espectroscopia em fases nebulares e imageamento pré-explosão. Utilizando curvas de luz em fases iniciais, espectroscopia nebular e dados de arquivo pré-explosão , restringimos as propriedades do sistema progenitor. Nossos resultados indicam um progenitor supergigante vermelha, com estimativas indiretas sugerindo uma massa inicial no intervalo de 10–15 $M_\odot$, enquanto restrições diretas a partir de imagens pré-explosão favorecem uma massa menor de $\sim 7.5\,M_\odot$. Investigamos essa tensão examinando suposições de modelagem, cenários de evolução binária e incertezas sistemáticas na calibração de fluxo, evidenciando limitações nas metodologias atuais. Por fim, apresentamos um arcabouço baseado em inteligência artificial para a extração de parâmetros físicos a partir de curvas de luz transientes esparsas no contexto do Vera Rubin Observatory Legacy Survey of Space and Time. Embora ainda em estágios iniciais de desenvolvimento, o método já fornece informações valiosas sobre as distribuições subjacentes de parâmetros e mostra-se promissor como ferramenta para inferência física robusta na era Rubin.

Palavras-chave: aprendizado profundo, aprendizado de máquina, funções de densidade de distribuições de probabilidade, redshifts fotométricos, levantamentos de grande área, supernovas, evolução estelar, astronomia transiente

Abstract

With the rapidly increasing volume of data in modern astronomy, scalable methods for processing and extracting physical information from observations have become essential. In this context, artificial intelligence has emerged as a powerful data-driven approach. In this thesis, we develop and apply such methods to two distinct problems in observational astrophysics: the estimation of photometric redshifts in wide-field surveys and the physical characterization of supernova progenitors through multiwavelength observations. To this end, the thesis is organized into two main parts.

The first part focuses on the estimation of photometric redshifts for the DECam Local Volume Exploration Survey. We develop a deep learning framework that combines a Recurrent Neural Network with a Mixture Density Network to infer full probability density functions for galaxy redshifts across the $\sim$ 20,000 deg^2 footprint of the survey. Our method achieves competitive performance compared to well-established methods in the literature. We further demonstrate the impact of the produced photometric redshifts on the fields of gravitational-wave dark sirens and galaxy cluster science.

The second part presents a comprehensive analysis of the Type II-P supernova SN 2022acko in NGC 1300, based on ultraviolet, optical, and infrared observations, complemented by late-time spectroscopy and pre-explosion imaging. Using early-time light curves, nebular spectroscopy, and archival data, we constrain the properties of the progenitor system. Our results indicate a red supergiant progenitor, with indirect estimates suggesting an initial mass in the range 10–15 $M_\odot$, while direct pre-explosion constraints favor a lower mass of $\sim$ 7.5 $M_\odot$. We investigate this tension by examining modeling assumptions, binary evolution scenarios, and systematic uncertainties in flux calibration, highlighting limitations in current methodologies. Finally, we present an artificial intelligence-based framework for extracting physical parameters from sparse transient light curves in the context of the Vera Rubin Observatory Legacy Survey of Space and Time. Although still in the early stages of development, the method already provides valuable insights into underlying parameter distributions and shows promise as a tool for robust physical inference in the Rubin era.

Keywords: deep learning, machine learning, probability distributions, photometric redshifts, wide-field surveys, supernovae, stellar evolution, transient astronomy

Contents

List of Figures

List of Tables

List of Acronyms

4MOST *4-metre Multi-Object Spectroscopic Telescope*

`Adam` Adaptive Moment Estimation

ACS Advanced Camera for Surveys

AI *Artificial Intelligence*

ATLAS Asteroid Terrestrial-impact Last Alert System

BAO Baryon Acoustic Oscillations

BASS *Beijing–Arizona Sky Survey*

BBH binary black hole

BNN *Bayesian Neural Networks*

BPASS Binary Population and Spectral Synthesis

BPZ Bayesian Photometric Redshifts

CCSNe core-collapse supernovae

CfA Center for Astrophysics

CGM circumgalactic medium

CHANCES *Chilean Cluster Galaxy Evolution Survey*

CMB Cosmic Microwave Background

CMD color–magnitude diagram

CNN *Convolutional Neural Networks*

CSM circumstellar material

CTIO Cerro Tololo Inter-American Observatory

DECaLS *Dark Energy Camera Legacy Survey*

DECam *Dark Energy Camera*

DELVE *DECam Local Volume Exploration Survey*

DELVE-DR2 DELVE second data release

DES *Dark Energy Survey*

DESI *Dark Energy Spectroscopic Instrument*

DESI-EDR *Dark Energy Spectroscopic Instrument Early Data Release*

DL *Deep Learning*

DT *Decision Tree*

ELBO *Evidence Lower Bound*

EM electromagnetic

GeMS Gemini Multi-conjugate Adaptive Optics System

GPU *Graphics Processing Unit*

GSAOI Gemini South Adaptive Optics Imager

GW gravitational wave

HDR *Highest Density Region*

HEASARC *High Energy Astrophysics Science Archive Research Center*

HMC Hamiltonian Monte Carlo

HST Hubble Space Telescope

IAU International Astronomical Union

IRAC Infrared Array Camera

LAB-IA Laboratory of Artificial Intelligence for Physics

LMU *Legendre Memory Units*

LRIS Low Resolution Imaging Spectrometer

LS *DESI Imaging Legacy Sirvey*

LS-DR10 *Legacy Surveys Data Release 10*

LSS Large-Scale Structure

LSST *Legacy Survey of Space and Time*

LSTM *Long Short-Term Memory*

LVK LIGO-Virgo-KAGRA

MCMC Markov Chain Monte Carlo

MDN *Mixture Density Networks*

MESA Modules for Experiments in Stellar Astrophysics

ML *Machine Learning*

MLP *Multi Layer Perceptron*

MSE *Mean Squared Error*

MzLS *Mayall z-band Legacy Survey*

NLL *Negative Log Likelihood*

NLP *Natrual Language Processing*

NLTE Non-Local Thermodynamic Equilibrium

NLTERT non-local thermodynamic equilibrium radiative transfer

NN *Neural Network*

PCA *Principal Component Analysis*

PDF *Probability Density Function*

PILCA Physics-Informed Machine Learning Light Curve Analyzer

PIT *Probability Integral Transform*

PROMPT Panchromatic Robotic Optical Monitoring and Polarimetry Telescopes

PSF *Point Spread Function*

RF *Random Forest*

RNN *Recurrent Neural Networks*

RSG red supergiants

S-PLUS Southern Photometric Local Universe Survey

SC Shock Cooling

SDSS *Sloan Digital Sky Survey*

SED *Spectral Energy Distribution*

SGD *Stochastic Gradient Descent*

SHSC *Southern Hemisphere Spectrocopic Redshift Compilation*

SIMBAD Set of Identifications, Measurements, and Bibliography for Astronomical Data

SNe II Type II supernovae

SNe II-P Type II-P supernovae

SNR *Signal to Noise Ratio*

Spitzer Spitzer Space Telescope

STEP S-PLUS Transient Extension Program

TF template fitting

UVOT Ultraviolet/Optical Telescope

VizieR *VizieR Catalog Service*

WFC3 Wide Field Camera 3

WFPC2 Wide Field Planetary Camera 2

Presentation

This Ph.D. thesis presents part of the work developed during my doctoral studies at the Centro Brasileiro de Pesquisas Físicas (CBPF). Over the past years, I have worked across different areas of astrophysics and cosmology, with my main contributions spanning two broad domains: Large-Scale Structure (LSS) and Time-Domain Astronomy. Within these areas, I led two peer-reviewed papers, one in each field.

The first paper, which represents the main project of the first half of my Ph.D., focuses on the development of photometric redshifts for the DECam Local Volume Exploration Survey (DELVE) using probabilistic neural networks. My involvement with photometric redshifts began during my undergraduate internship at CBPF, when I worked on redshift estimation for the Southern Photometric Local Universe Survey (S-PLUS) – an early experience that shaped my technical and methodological foundations in the field. Subsequently, my supervisor and I became members of the DELVE collaboration, where I further applied and expanded these skills in the context of a large, ongoing survey.

Alongside this main effort, I continued working on redshift-related projects for other surveys, adapting methods similar to those developed for DELVE to larger DECam-based catalogs, which were later used for dark standard sirens analysis and galaxy cluster target selection for spectroscopic measurements.

In parallel, about halfway through my Ph.D., I became involved in more observational activities within the group through the S-PLUS Transient Extension Program (STEP). This marked my entry into time-domain astronomy, with a strong focus on transient characterization.

My participation in STEP culminated in an in-depth analysis of a single event, the Type II supernova SN 2022acko. This work led to the second paper I authored, in which we characterize both the supernova and its progenitor system using a diverse dataset that includes pre-explosion imaging, early- and late-time light-curve photometry (including original data from STEP), and nebular spectroscopy. This project also motivated new ideas on applying deep learning to time-domain astronomy, particularly the development of physics-informed neural networks for fitting light curves to differentiable physical models.

Given the interdisciplinary nature of my research, I chose to separate this thesis into two parts. Part I focuses on the development of photometric redshifts for DELVE, published in Teixeira et al. (2024), along with a concise – but not too concise – overview of the deep learning foundations needed to follow this thesis, and a description of extensions of this work to other DECam-based surveys and their impact on different areas of astrophysics. Part II is devoted to time-domain astronomy and details the multi-wavelength, multi-epoch characterization of the nearby Type-II supernova SN 2022acko published in Teixeira et al. (2026), from pre-explosion photometry to the nebular phase spectra, as well as our perspectives on an ongoing effort to develop a physics-informed machine learning

framework for transient characterization.

In Part I, as it corresponds to my first paper – written by a very inexperienced Ph.D. student – I expand upon, and in some parts completely rewrote, most of the content. In Part II, particularly Chapter 8, the text is largely based on the published paper, with more detailed descriptions of a few methods only; I also removed some analyses that I did not consider crucial to the narrative.

Presentations and disclaimers being made, let us get to the science.

Part I

Photometric Redshifts for the DELVE Survey

1 Introduction

In the context of special relativity, when a light source moves relative to an observer, the light detected by the observer has a different wavelength than that emitted in the source rest frame. This effect is known as the relativistic Doppler effect, and becomes significant as the source velocity approaches the speed of light c (Rindler, 2006). With λ being the detected wavelength in the observer frame, λ_0 being the rest-frame wavelength of the light emitted by the source, and v_r being the radial component of the velocity of the source with respect to the observer, the relativistic Doppler effect can be written as

$$\lambda = \lambda_0 \sqrt{\frac{1 + v_r/c}{1 - v_r/c}}. \tag{1.1}$$

The relative *shift* of λ_0 is commonly denoted by z, and takes the form

$$z = \frac{\lambda - \lambda_0}{\lambda_0}. \tag{1.2}$$

When $z < 0$, the light source is approaching the observer, and z is referred to as *blueshift*. When $z > 0$, the source is receding from the observer, and z is called *redshift*. Combining Equations 1.1 and 1.2, one can explicitly relate the (blue) redshift to the radial velocity of the source as

$$1 + z = \sqrt{\frac{1 + v_r/c}{1 - v_r/c}}. \tag{1.3}$$

This measurement plays a critical role in astronomy. In the early 20th century, astronomers used the relations described above to estimate the radial velocities of galaxies by measuring the Doppler shift of their spectral lines (Carroll and Ostlie, 2017). Pioneering measurements by Vesto Slipher (Slipher, 1913, 1915, 1917) revealed that most galaxies exhibited redshifted spectra, indicating that they were receding from the observer. This result was unexpected, as the prevailing assumption was that galaxy motions should be randomly distributed.

This picture changed with Edwin Hubble (Hubble, 1929), who measured distances to nearby galaxies using Cepheid variable stars as standard candles, following the period-luminosity relation established by Leavitt and Pickering 1912. Combining these distance estimates with radial velocity measurements, Hubble (1929) demonstrated that the recession velocity of the observed galaxies increases proportionally with distance. This observational evidence converged with the theoretical frameworks independently developed by Friedmann (1922), who applied Einstein's field equations of general relativity to

a homogeneous and isotropic universe and showed that they naturally predict a dynamic universe, and by Lemaître (1927), who independently reached the same conclusion and additionally derived a linear velocity-distance relation – both in contrast to the static universe interpretation of Einstein (1917). Together, these contributions culminated in the Hubble-Lemaître law,

$$v = H_0\, d, \tag{1.4}$$

where v is the recession velocity of the galaxy, d is its distance from the observer, and H_0 is the Hubble constant[1]. This result was revolutionary, providing for the first time observational evidence that the Universe is expanding. In an expanding universe, the observed redshift of galaxies is not a consequence of the Doppler effect from motion through space, but arises from the expansion of spacetime itself, which stretches the wavelengths of photons as they travel through the Universe, giving rise to a cosmological redshift (Carroll and Ostlie, 2017). For small recession velocities ($z \ll 1$), the cosmological redshift reduces to the approximation $v \approx cz$, upon which the Hubble-Lemaître law can be written as

$$cz \approx H_0\, d, \tag{1.5}$$

providing a direct mapping from redshift to distance.

The general redshift-distance relation, valid at higher redshifts, depends on the expansion history of the Universe. Equation 1.5 holds in the regime $z \ll 1$ because the expansion rate can be approximated as constant and equal to H_0. In general, however, the expansion rate of the Universe is not constant throughout its history, but is itself a function of redshift, $H(z)$, which depends on the assumed cosmological model.

The ΛCDM model, the standard cosmological framework most consistent with current observations (Perlmutter et al., 1999; Riess et al., 1998), assumes a spatially flat Universe whose energy budget is dominated by a cosmological constant Λ – interpreted as the dark energy driving the accelerated expansion – and cold dark matter, a pressureless non-relativistic component that drives structure formation (Planck Collaboration et al., 2020). Within this framework, the redshift-distance relation is given by

$$d_C = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \tag{1.6}$$

where d_C is the comoving distance and $E(z) \equiv H(z)/H_0$ is the dimensionless Hubble parameter, encoding the contributions of the different energy components of the Universe to $H(z)$.

[1] In his original formulation, Hubble expressed the relation as $v = Kd$, referring to K simply as a proportionality constant. The term "Hubble constant" was introduced later, and is now commonly denoted as H_0. Notably, Lemaître (1927) had independently derived an equivalent relation two years prior to Hubble (1929) observations, a contribution formally recognised by the International Astronomical Union (IAU) in 2018, which led to the renaming of the "Hubble law" to the "Hubble-Lemaître law" (Kragh, 2018)

By the late 20th century, redshift measurements had proven their relevance to our understanding of the Universe, and large community efforts emerged to provide redshift catalogues for increasingly large samples of galaxies. Pioneering initiatives such as the Center for Astrophysics (CfA) Redshift Survey (Huchra et al., 1983; Davis et al., 1982) opened the way for unprecedented advances in observational cosmology, later followed by larger efforts such asthe *Sloan Digital Sky Survey* (SDSS) (York et al., 2000) and the 2dF Galaxy Redshift Survey (Colless et al., 2001), which together mapped the positions of millions of galaxies across large fractions of the sky. These surveys opened the way for unprecedented advances in observational cosmology, most notably the study of the Large-Scale Structure (LSS) of the Universe – the web-like distribution of galaxies, filaments, voids, and clusters on scales of tens to hundreds of megaparsecs (Bond et al., 1996).

Among the many surveys dedicated to measuring redshifts and mapping the LSS, the aforementioned SDSS (York et al., 2000) has been one of the most influential. Using a dedicated 2.5 m wide-angle optical telescope at Apache Point Observatory, the SDSS imaged more than 200 million galaxies in five photometric bands while obtaining spectra for roughly 2 million of those (Beck et al., 2016). The survey achieved a median spectroscopic redshift of $z \approx 0.1$ for its main galaxy sample, with coverage extending to $z \approx 0.7$ for luminous red galaxies (Eisenstein et al., 2011).

Today, one of the most remarkable redshift surveys is *Dark Energy Spectroscopic Instrument* (DESI). From its first year of observations already, DESI measured Baryon Acoustic Oscillations (BAO) (Bassett and Hlozek, 2010) using over 5.7 million unique galaxy and quasar redshifts spanning $0.1 < z < 2.1$ (Adame et al., 2025a), achieving a combined BAO precision of $\sim$0.52% across six redshift bins. These results, combined with Cosmic Microwave Background (CMB) (Durrer, 2015) data, provide tight constraints on cosmological parameters including the Hubble constant and the dark energy equation of state (Adame et al., 2025b).

The surveys described so far are dedicated to precise redshift assessment obtained from spectroscopy. Briefly, spectroscopic redshifts (spec-z, $z_{\rm spec}$) are estimated by applying Equation 1.2 to calibrated spectra, identifying the shift of prominent spectral lines – such as Hα, the Ca II H&K doublet, or the [O II] doublet – relative to their known rest-frame wavelengths. However, acquiring spec-zs is both time-intensive and resource-demanding, as strong spectral features must be clearly identified (D'Isanto and Polsterer, 2018). For instance, using a single slit on a 4 m-class telescope can require tens of minutes per object at $r \gtrsim 20$ (Bolton et al., 2012; Dawson et al., 2012). For these reasons, obtaining $z_{\rm spec}$ for the full population of photometrically detected galaxies is infeasible: the SDSS, for example, obtained spectra for only $\sim$1% of the galaxies it detected in imaging.

Nevertheless, studies of the LSS of the universe – as well as investigations of cosmic shear and the distribution of dark matter – demand redshift estimates for volumes of galaxies far larger than spectroscopic surveys can provide (Chang et al., 2013). An

alternative approach arises from the use of photometry: instead of measuring redshifts directly from spectra, one maps an empirical relation between the observed broad-band magnitudes and the redshift. Redshifts obtained this way are called photometric redshifts (photo-z, $z_{\rm phot}$).

Estimating photo-zs can be approached through two main families of methods: template fitting (TF) and *Machine Learning* (ML). TF methods model the *Spectral Energy Distribution* (SED) of an object by fitting it to a library of galaxy templates – which can be synthetic or empirical – at different redshifts (Benítez, 2000; Crenshaw and Connolly, 2020; Ilbert et al., 2006; Brammer et al., 2008). This approach requires well-calibrated, unbiased templates together with explicit assumptions for each observed galaxy, such as dust extinction and galaxy type. Furthermore, the available template set fundamentally constrains its applicability to galaxy types and redshift ranges represented within the library.

ML, and particularly *Deep Learning* (DL) (overview in Chapter 2), has emerged as a promising avenue for estimating $z_{\rm phot}$ (e.g., Newman and Gruen, 2022; Duncan, 2022; Schuldt et al., 2021; Pasquet et al., 2019), primarily due to its capacity to learn intricate patterns within large datasets. Additionally, *Graphics Processing Unit* (GPU) parallelization makes ML methods computationally efficient at scale. It is a viable alternative to TF, since it learns a direct mapping from photometry to redshift, bypassing the need for physical assumptions and avoiding errors from unrepresentative templates (Lima et al., 2022) – still, ML-based methods are also constrained by the representativeness of their training data, which are typically much larger in volume than template libraries (Salvato et al., 2019).

Early implementations of ML techniques for photo-z estimation were limited to point estimates of $z_{\rm phot}$ (Henghes et al., 2021), which cannot adequately represent the full uncertainty in the measurement. Beyond simple error characterization, the complete *Probability Density Function* (PDF) of the photo-z $p(z)$ carries additional statistical information that is lost when only a point estimate is retained. The relevance of $p(z)$ has been demonstrated across several fields. A few examples are: weak lensing tomography, where redshift distributions must be accurately characterised to recover unbiased shear power spectra (e.g., Mandelbaum et al., 2008; Hu and Tegmark, 1999); dark siren analyses, where $p(z)$ information improves the statistical association between gravitational-wave events from binary black hole mergers and their potential host galaxies (Alfradique et al., 2024; Bom et al., 2024); and galaxy cluster membership, where $p(z)$ information improves the discrimination of cluster members from foreground and background galaxies (Sifón et al., 2025; Méndez-Hernández et al., 2026).

In certain scenarios, PDFs from ML algorithms may exhibit inconsistencies when interpreted, necessitating careful inspection or re-calibration to ensure their reliability (D'Isanto and Polsterer, 2018; Dey et al., 2022). Common calibration problems arise

from the (under-)overconfidence of PDFs. For example, if a model is trained to achieve the best predicted mean over the PDF of z_{phot}, it may result in unbiased values for z_{phot}; however, this does not directly reflect the reliability of the PDFs, since one could obtain an excessively broad or narrow symmetric PDF centred close to the target value. To ensure that the PDFs are meaningful, we need to assess them using diagnostic tools such as the Probability Integral Transform (Polsterer et al., 2016) and the coverage test (Zhao et al., 2021), which reveal trends towards (under-)overconfidence of the model, as well as possible biases.

Although several modern ML algorithms can produce PDFs (Mucesh et al., 2021; Schmidt et al., 2020; Desprez et al., 2020; Sadeh et al., 2016), the use of more recent probabilistic DL techniques across different network architectures has not yet been fully explored. Recent efforts by Ren et al. (2026) introduced `nflow-z`, a photo-z estimation method based on normalising flows, an advanced DL technique that explicitly models the photo-z PDFs as a sequence of bijective transformations conditioned by the magnitudes and colors. In their work the authors demonstrate its effectiveness across multiple datasets, outperforming several state-of-the-art algorithms, with particularly strong results for faint galaxies with sparse spectroscopic training data. In Zhou et al. (2022), the authors explore *Bayesian Neural Networks* (BNN)s to estimate photo-z PDFs from both galaxy flux and image data for the China Space Station Telescope, using Bayesian *Multi Layer Perceptron* (MLP) and Bayesian *Convolutional Neural Networks* (CNN) architectures, respectively, as well as a hybrid network combining both inputs via *transfer learning* (Zhuang et al., 2020). The authors demonstrate that BNNs can yield reliable PDFs with competitive photo-z accuracy, however at the cost of increased computational complexity relative to traditional neural networks. Closely related to the present thesis, Lima et al. (2022) combined *Mixture Density Networks* (MDN) and MLP to estimate photo-z PDFs for Southern Photometric Local Universe Survey (S-PLUS), outperforming well-established methods in the literature such as Bayesian Photometric Redshifts (BPZ) and `GPz` (Almosallam et al., 2015). The resulting PDFs account for the redshift-color degeneracy through their multi-peaked nature, and since they take the form of closed analytical expressions – linear combinations of Gaussians – the full PDF information is readily accessible in the final catalogues. This work was further extended by combining BNN with MDN, leading to the latest photo-z catalogue of S-PLUS (Lima, 2025).

In this first part of the thesis, we present the development of photo-zs for the *DECam Local Volume Exploration Survey* (DELVE) survey, a deep optical imaging program designed to map the southern sky and enable studies of the Milky Way, nearby galaxies, and LSS. Our aim is to explore novel DL techniques for photo-z estimation and to assess the reliability of DELVE photometry for this purpose. We estimate redshifts, and their respective PDFs using a combination of *Recurrent Neural Networks* (RNN) and MDN.

Chapter 2 provides a concise introduction to neural networks, with particular focus

on the architectures employed in this thesis. In Chapter 3 we describe the main datasets used in this work and provide a detailed description of the DELVE survey. Chapter 4 details the photo-z methods we explored, which are then validated and discussed in Chapter 5. We present the main applications of this work in Chapter 6 and offer concluding remarks in Chapter 7.

2 Neural Networks Overview

The "Artificial Intelligence" in the title of thesis may sound somewhat overpromising. This is because the use of *Artificial Intelligence* (AI) in this work is restricted only to a subset of techniques within the field, namely DL. What, then, distinguishes these two terms? And how does ML fit into this context? In this chapter, we briefly address these questions and provide a concise overview of the fundamental principles of ML and DL, as well as the main algorithms employed throughout this thesis.

2.1 | The Concept

Among many definitions, AI can be described as "the art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990). I choose this specific definition of AI because it does not mention computers explicitly. In fact, some of the earliest AI systems were designed to mimic human behavior long before the existence of modern computers. A classic example dates back to 1739, with Jacques de Vaucanson's *automatas*: machines capable of reproducing human actions, such as playing tambourines or flutes (Kang, 2011).

Although the concept of intelligent machines is at least hundreds of years old, the term AI was officially coined in 1956 at the Dartmouth Conference (McCarthy et al., 2006), marking the establishment of AI as a distinct research field. This conference brought together key figures, such as John McCarthy[1] and Claude Shannon[2], who would become leading contributors to major breakthroughs in the field in the following decades (Russell and Norvig, 2010).

ML was introduced even before the official coining of the term AI. In 1950, Alan Turing[3] introduced a method to assess whether a computer system could be considered intelligent, now known as the Turing Test (Turing, 1950). The simplest version of the Turing Test – originally called "the imitation game" – lists several capabilities that a machine should exhibit to be considered intelligent: natural language processing, knowledge

[1] John McCarthy (1927–2011) was a computer scientist who is widely considered the father of artificial intelligence; he organized the Dartmouth Conference and contributed fundamental concepts such as the Lisp programming language.

[2] Claude Shannon (1916–2001) was an electrical engineer and mathematician, known as the father of information theory; his work on information, computation, and logic influenced the foundations of AI.

[3] Alan Turing (1912–1954), British mathematician and logician, widely considered athe father of computer science.

representation, automated reasoning, and machine learning. In this context, ML refers to the ability to detect patterns in data and extrapolate from them (Russell and Norvig, 2010).

Even without a precise naming date, after the Dartmouth Conference several early works in AI were already laying the foundations of ML (Rosenblatt, 1958; Samuel, 2000; Breiman et al., 1984; Russell and Norvig, 2010). Mitchell (1997) defines ML as "the study of computer algorithms that improve automatically through experience." By that time, ML had already become a scientifically recognized family of statistical methods. Most ML algorithms consist of producing an output (answer) y for a given input (question) x. The input x typically represents observed data, and the output y is a property to be inferred from x. When the mapping from x to y is too complex to be explicitly programmed or analytically tracked, ML provides a valuable alternative by learning the mapping intrinsically from the observed data, or, following the definition of Mitchell (1997), from experience.

2.2 | Machine Learning

There are three main paradigms of learning in ML: supervised learning, unsupervised learning, and reinforcement learning (Goodfellow et al., 2016; Russell and Norvig, 2010).

- *Supervised learning* refers to problems where examples of inputs x are paired with their corresponding outputs (labels) y obtained from reliable observations. This framework is widely adopted when a large amount of input data is available, but only a subset has associated labels. The goal is to learn a mapping from x to y that generalizes to new, unseen inputs. Its broadly used for regression and classification tasks .

- *Unsupervised learning* occurs when only the input data x is available, and the objective is to discover patterns or structures directly from the data. A classical example is clustering, where the algorithm automatically groups similar data points to reveal underlying patterns or relationships (Yin et al., 2024).

- *Reinforcement learning* addresses problems in which an agent interacts with an environment and learns to take actions that maximize a cumulative reward. Unlike supervised learning, explicit input-output pairs are not provided. Instead, the agent learns from feedback in the form of rewards or penalties, improving its strategy through trial and error. This paradigm is particularly useful for dynamic decision-

making tasks, such as game playing, robotics, or autonomous control (Sutton and Barto, 2018).

The methods used in this part of the thesis primarily concern supervised learning (the reader may refer to Goodfellow et al. (2016) for a comprehensive overview of the different learning paradigms). In this paradigm, we generally consider a dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\}$, where $\mathbf{X}$ is an $N \times D_{\text{in}}$ matrix containing N examples of D_{in}-dimensional input vectors $\mathbf{x} \in \mathbb{X}$, and, analogously, $\mathbf{Y}$ is an $N \times D_{\text{out}}$ matrix of corresponding outputs $\mathbf{y} \in \mathbb{Y}$. Here we assume matrices for simplicity; however, the following discussions can be extended to tensors, as in the case of images.

Given a dataset $\mathcal{D}$, ML methods aim to learn the function f that maps the measurements in $\mathbb{X}$ to properties in $\mathbb{Y}$. This function is likely unknown, or very hard to compute. ML methods, then only approximate f by learning $\tilde{f} : \mathbb{X} \to \mathbb{Y}$, parametrized by a set of parameters $\boldsymbol{\theta}$, such that for a given example $\mathbf{x}$, the output $\mathbf{y}$ is approximate by

$$\tilde{f}_{\boldsymbol{\theta}}(\mathbf{x}) \approx f(\mathbf{x}) = \mathbf{y}. \tag{2.1}$$

The learning process consists of finding the set of parameters $\boldsymbol{\theta}$ that best approximates Eq. 2.1, enabling the application of $\tilde{f}$ to unlabeled data.

2.2.1 | Random Forest

One of the most successful ML methods is the *Decision Tree (DT)*. They were originally build to address classification problems. The simplest family, the *Boolean DTs*, provides a clear example to explain the concept. In this type of classification, you have *features* that lead to a *decision* (Breiman et al., 1984; Russell and Norvig, 2010), and both the features and the final output are boolean.

For instance, if Ph.D students want to decide whether to register for a conference, they might analyze important features such as:

"Will them need financial support from the organization?"

"Is the conference outside their institution's country?"

"Does their advisor have funds to support the travel?"

The DT is a logical sequence of decisions that guides you to the final answer. The trees themselves are constructed from a given example dataset $\mathcal{D}$. In this context, the input vectors $\mathbf{x}_i = \mathbf{X}_{i,:}$ represent the features, with $\mathbf{X}_{i,:}$ being the vector given by the i-th row (column) of the matrix $\mathbf{X}$. In the binary classification case, the output reduces to a vector $\mathbf{y} = (y_1, y_2, ..., y_N)$, where N is the number of examples and $\mathbf{X}_{i,j},\ y_i \in \{\text{False}, \text{True}\}$.

The components of a DT and their role during training are:

- *The root*: contains all the data at the start of the tree.
- *The nodes*: perform feature tests, spliting the data depending on the value of a given feature.
- *The branches*: connect nodes and indicate the outcome of a feature test.
- *The leaves*: contain the final answer of a given decision path.

We call *training* the process of presenting the examples in $\mathcal{D}$ to build the model before applying it to unseen data. Figure 2.1 shows an example of the DT training process for a dataset with 6 examples. In this case, the branches test the value of the binary features A_i $(1 \leq i \leq D_{in})$ and split the examples according to the outcome of the test. When all examples in a given branch of the split have the same output, a leaf node is created corresponding to this output. If the examples contain mixed classes, the algorithm continues recursively, selecting another feature to test. In this example, the algorithm tested A_1 after A_2.

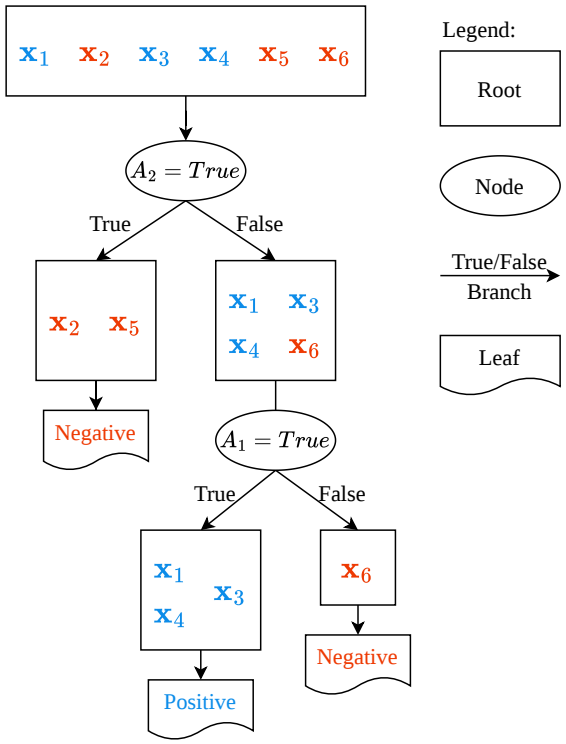

Figure 2.1: Example of a DT constructed from a dataset with 6 examples.

This resulting tree does not imply that the dataset contains only two features; rather, it indicates that these two features were sufficient to classify all examples in this

particular dataset. The algorithm always starts from the *most important feature* down to the less important ones, where *importance* is based by the choice of the *split criterion* adopted. The split criterion is the metric used to evaluate the quality of a candidate split at each node, such as Gini impurity or entropy for classification, and variance reduction for regression (Russell and Norvig, 2010).

This example relies on a very specific case of binary classification with binary features. However, DTs can be applied to a much wider range of problems, including those involving continuous or categorical features and outputs. For instance, if we want to extend this example into a regression problem with continuous features, the changes are roughly:

- *Nodes*: Instead of binary tests, each node performs a conditional test based on a threshold for a given feature. For example, instead of checking whether $A_i = \text{True}$, the node would test $A_i < \text{threshold}$. The choice of the threshold at each node depends on the split criterion.

- *Leaves*: Once the path reaches a leaf, instead of returning the most common class, the leaf predicts the *average* of the y_i values associated with the input vectors $\mathbf{x}_i$ that reached that leaf.

Recall that, in this example, there may be additional features that influence the classification, but the limited number of examples provided to the model is insufficient to capture their effect. This illustrates a key limitation of simple DTs: they are highly sensitive to the training data. Small changes in the training examples can lead to substantially different trees. This implies poor generalization to unseen data (Russell and Norvig, 2010).

Formalized by Breiman (2001), the *Random Forest* (RF) is a DT-based ML algorithm designed to address key weaknesses of traditional DTs. The method is rooted in the principle of *ensemble learning*, according to which the combined prediction of multiple models is typically more accurate and more stable than the prediction of a single model (Russell and Norvig, 2010).

A RF consists of a collection of DTs. Each tree is trained on a subset of the original dataset $\mathcal{D}$, and the final prediction is obtained by combining the individual predictions of all trees in the ensemble – commonly by taking their mean in regression problems. The process of sampling uniformly from $\mathcal{D}$ while allowing repetitions is known as *bootstrapping* (Efron and Tibshirani, 1994), whereas the process of combining the predictions of multiple models is referred to as *aggregation*. The combination of these two procedures defines *bagging*, which constitutes the foundation of RFs.

In its original formulation, the RF introduces an additional source of randomness by selecting, at each tree node, a random subset of the available features to consider when

making splits. Together, bootstrapping and random feature selection significantly reduce the sensitivity of the model to fluctuations in the training data, thereby lowering variance and mitigating the instability that characterizes individual DTs.

DTs, RFs, and *all* the ML/DL algorithms we will visit in this chapter depends on initial parameters to define the general structure of the resulting model. For example, we have to initially define the number of trees in our RF, or the maximum number of leaves in each DT. These parameters are called *hyperparameters* and impact directly on the efficiency and interpretability of your model. Different sets of hyperparameters in the same algorithm leads to different models. In the case of RFs, the main hyperparameters are shown in Table 2.1.

Table 2.1: Main hyperparameters of a Random Forest.

Parameter	**Symbol**	**Type**
Number of trees	N_{trees}	integer
Maximum depth	D_{max}	integer or None
Minimum samples per split	$N_{\text{min}}^{\text{split}}$	integer
Minimum samples per leaf	$N_{\text{min}}^{\text{leaf}}$	integer
Number of features per split	D_{split}	integer or fraction
Bootstrap	–	boolean
Split criterion	–	categorical

This special focus on RFs is motivated by their wide range of applications in astronomy, particularly in photo-z estimation. In Carliles et al. (2010) and Carliles et al. (2008) the authors use RF to estimate photo-z values for SDSS (York et al., 2000) galaxies, as well as to model the associated uncertainties. Henghes et al. (2021) compare RF with other ML-based photo-z estimators and show that RF achieves the lowest *Mean Squared Error* (MSE), although its performance is comparable to that of the other methods. Later in this thesis, we compare RF-based photo-z estimation with the other algorithms proposed in this work.

2.3 | Deep Learning

The discussion on RF so far highlights one of the main limitations of traditional ML: feature engineering. Taking a single DT as an example, its performance heavily depends on how well the algorithm defines the importance of the input features. In many practical cases, the raw observed features (e.g. pixel values, magnitudes, sensor readings) are not directly related to the desired outputs; instead, properties derived from correlations among these features can be more informative – ratios, differences. For many ML methods to succeed, pre-processing and careful adjustment of the input representation

is necessary, often requiring *human expertise* or dimensionality reduction methods such as *Principal Component Analysis* (PCA) (Shlens, 2014).

When the dimensionality of the data is very high, extracting a sufficiently informative representation for traditional ML algorithms becomes difficult. In Rosenblatt (1958), the *perceptron* was introduced, a simple algorithm capable of binary classification that later served as a building block for more complex models. Combining perceptrons in multiple layers led to the development of MLP, which provided an alternative for classification and regression tasks. However, training MLPs with many layers remained challenging for decades due to computational limitations.

Around 2006, it was demonstrated that a *deep* enough network of perceptrons could automatically learn hierarchical representations from input data (Hinton et al., 2006; Bengio et al., 2006; Ranzato et al., 2006). This realization coincided with the rapid increase in computational power, particularly through GPU, enabling practical applications of DL. These technological improvements triggered the modern boom of AI, opening a plethora of opportunities to exploit the power of DL techniques for high-complexity tasks.

It is worth noting that DL, as originally conceived, can be regarded as a subset of ML, inheriting many of its paradigms and formalism. What distinguishes DL from traditional ML methods is its ability to *automatically learn hierarchical representations* from high-dimensional data. This distinction relates not only to the depth of the networks but also to the complexity of the algorithms themselves.

Although many ML algorithms perform well for photo-z estimation, DL has emerged as a promising alternative, particularly in the context of uncertainty estimation(Lima et al., 2022; Ren et al., 2026; D'Isanto and Polsterer, 2018). In this thesis, we investigate DL approaches for this task, benchmarking their performance against traditional reference photo-z methods, and also develop a DL-based approach for transient characterization in Part II.

2.3.1 | Multi-Layer Perceptron

To introduce the main algorithms behind DL, we consider the very first DL *architecture* as an example: the MLP[4]. Recalling the discussion in § 2.2, in the realm of supervised learning our goal is to construct a function $\tilde{f}(\mathbf{X})$ that best approximates $\mathbf{Y}$ while generalizing to unseen $\mathbf{x}$. The function $\tilde{f}$ is parametrized by a set of parameters $\boldsymbol{\theta}$.

[4]The original perceptrons, introduced by Rosenblatt (1958), were actually different from this formulation. Initially designed for classification, they had step functions as outputs, therefore failing to generalize to non-linear mappings. MLPs, while established as a ML algorithm since their early formulation, are considered here in the context of the first deep networks architectures that emerged with the DL revolution of the 2000s.

In the case of RF regression, $\boldsymbol{\theta}$ corresponds to the number of nodes and leaves, as well as the splitting thresholds within each node.

In the case of an MLP, the function $\tilde{f}$ is parametrized by the network *weights* and *biases*. Instead ofsplitting the data as its done in tree-based methods, the network *processes* the input by performing linear transformations followed by nonlinear *activation functions*. These transformations propagate through the layers, producing outputs at each neuron, until the final output layer is reached. Figure 2.2 schematizes the algorithm flow.

The input vector $\mathbf{x} = (x_1, x_2, ..., x_{D_{\text{in}}})$ corresponds to the input layer. It undergoes a linear transformation weighted by the network parameters, producing an intermediate vector $\mathbf{z}$. Each element of $\mathbf{z}$ is then passed through an activation function, producing the neuron outputs. In a single hidden layer network with H neurons, this produces $\mathbf{z} = (z_1, z_2, ..., z_H)$, which is subsequently mapped to the output vector $\mathbf{y} = (y_1, y_2, ..., y_{D_{\text{out}}})$.

More generally, for a network with $L > 1$ hidden layers, we denote the output of the i-th hidden layer as $\mathbf{z}_i = (z_{i,1}, z_{i,2}, ..., z_{i,H_i})$, with $1 \leq i \leq L$. At this point, it may be clear how easily the notation can become overwhelming, even for a network with just a few layers. Therefore, we will restrict most examples to a single hidden layer.

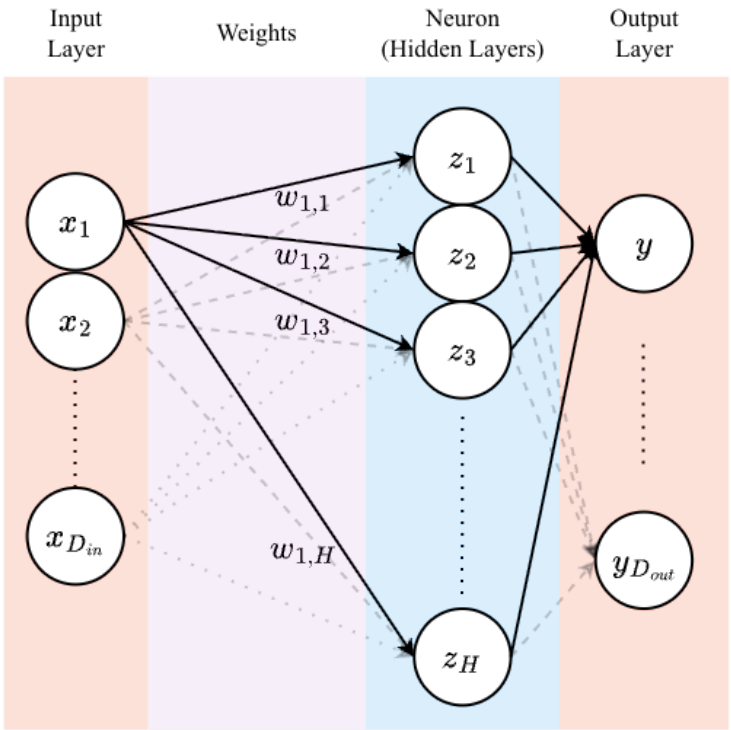

Figure 2.2: Schematic diagram of a MLP with a single hidden layer. The input vector $\mathbf{x}$ *is mapped to the output vector* $\mathbf{y}$ *through a hidden layer of neurons, each applying a linear transformation followed by a nonlinear activation function.*

2.3.1.1 The Neuron

To better understand the process, let us examine the main unit of an MLP, the *artificial neuron*[5]. Figure 2.3 shows a schematic diagram.

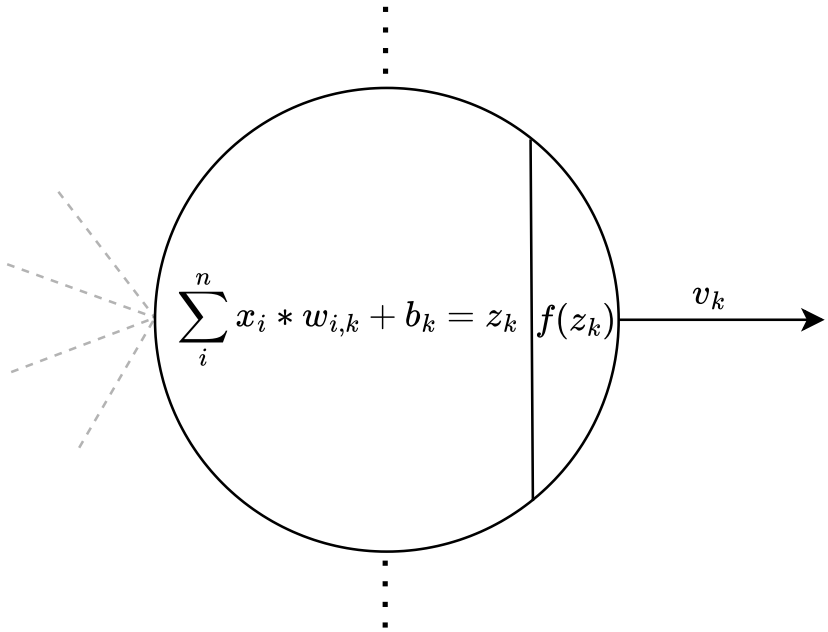

Figure 2.3: Schematic diagram of an artificial neuron. The inputs $x_1, x_2, \ldots, x_{D_{\text{in}}}$ are combined through a weighted sum with a bias term, and passed through a nonlinear activation function σ to produce the neuron output z.

Each neuron computes a weighted sum of all outputs from the previous layer (or the input vector in the first hidden layer) and adds a constant term, called the bias, to allow more flexible affine transformations. To enable the network to represent non-linear mappings, we apply an *activation function* σ, which determines how the neuron's output z is transformed and propagated to the next layer.

After propagating the inputs through the hidden layers, the network produces an output vector $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{D_{\text{out}}})$, where each component is computed as

$$\hat{y}_i = \sigma_{\text{out}}\left(\sum_j w_{j,i}\,\sigma(z_j) + b_i^{\text{out}}\right), \tag{2.2}$$

with $w_{j,i}$ denoting the weight connecting hidden neuron j to output neuron i, b_i^{out} the output bias, and σ_{out} the activation function applied at the output layer. In the simplest case of a single hidden layer (Figure 2.3), the hidden neuron outputs are given by $z_j = \sum_k w_{k,j}x_k + b_j$ and transformed by the hidden activation function σ. Hereafter, we will denote by σ a general activation function, except when it is explicitly assigned to the sigmoid function.

[5]Originally perceptron. Inspired by the natural neuron (McCulloch and Pitts, 1943).

2.3.1.2 The activation functions

The choice of activation function has a significant impact on network performance. It controls the flow of information through the network, enhances the representational capacity of the model, and shapes the range of possible outputs. Figure 2.4 shows four common activation functions. Among them, the ReLU function is generally preferred for hidden layers due to its simplicity and lower computational cost during training (Glorot et al., 2011).

For the output layer, the activation function should be chosen according to the task. For binary classification, the *sigmoid* function is appropriate, whereas for multi-class classification, the *softmax*[6] function is typically used. For regression tasks such as single-point photometric redshift estimation, a *linear* activation is generally suitable. Alternatively, *ReLU* may be used if negative predictions are physically meaningless or undesired.

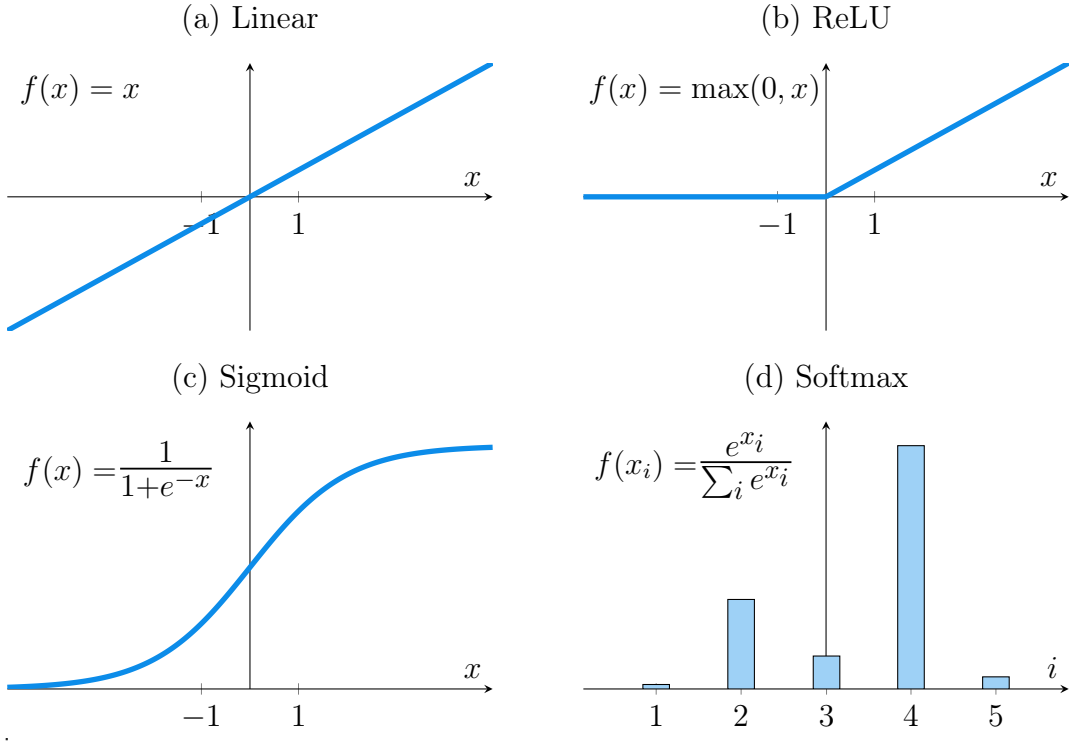

Figure 2.4: Activation functions: Linear, ReLU, Sigmoid, and Softmax.

[6]Unlike other activation functions, softmax operates on the entire layer simultaneously, normalizing the neuron outputs into a probability distribution.

2.3.2 | The Training

2.3.2.1 The Loss Function

The learning consists of adjusting the network parameters to minimize the discrepancy between the predicted outputs and the true labels. The *loss function*[7], denoted by $\mathcal{L}$, quantifies this discrepancy.

The specific form of $\mathcal{L}$ depends on the nature of the problem. For regression tasks, a common choice is the MSE. For a sample of N outputs, the MSE is defined as

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{2.3}$$

For classification tasks, an appropriate choice is the cross-entropy loss, which originates from information theory (Shannon, 1948) and quantifies the divergence between the true and predicted class distributions. If the output is one of m classes and y is represented as a one-hot vector (1 for the correct class, 0 otherwise), the cross-entropy loss for a single example is

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{m} y_i \log(\hat{y}_i), \tag{2.4}$$

where, in this case, $\hat{y}_i$ is a probability assigned to the i-th class – typically coming from a softmax-activated output layer. Minimizing $\mathcal{L}_{\text{CE}}$ is equivalent to maximizing the log-likelihood of the true labels given the predicted distribution.

These are standard loss functions for many supervised learning tasks. In this part of the thesis, the developed models predict full probability distributions over the redshift space rather than a single point estimate of the photometric redshift. For this purpose, we adopt the *Negative Log Likelihood* (NLL) loss. If the network output is a probability distribution function $\hat{p}(z)$, the loss for a single observation is

$$\mathcal{L}_{\text{NLL}}(\hat{p}, y) = -\log \hat{p}(y). \tag{2.5}$$

This encourages the model to assign high probability density at the true redshift y, and is minimised when $\hat{p}(y)$ is maximised. Note that, unlike the cross-entropy loss, the NLL does not require $\mathbf{y}$ to be a distribution – it evaluates the predicted density at a single observed value, making it the natural loss function for continuous probabilistic regression tasks.

[7] This function acctually attends for many different names: cost function, objective function, criterion, error function.

2.3.2.2 The Optimization

Once the data flow through the network and the objective function is well defined, an essential question remains: how do we optimize the model? The optimization algorithm is a core component of DL. Its goal is to find the optimal set of parameters $\boldsymbol{\theta}^*$ that minimizes the chosen loss function $\mathcal{L}$:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\tilde{f}_{\boldsymbol{\theta}}(\mathbf{X}), \mathbf{Y}). \tag{2.6}$$

In practice, exact minimization is rarely possible due to the high dimensionality and non-convexity of $\mathcal{L}$. As a result, the optimization of $\mathcal{L}$ is performed using iterative methods. In DL, these are typically gradient-based. Given a fixed dataset $\mathcal{D}$, the loss function $\mathcal{L}$ depends only on the network parameters. We can therefore define $\mathcal{L} : \mathbb{R}^n \to \mathbb{R},$, where n denotes the number of trainable parameters of the network[8].

The foundational gradient-based optimization method in DL is *gradient descent* (Goodfellow et al., 2016). In this approach, the parameters are updated iteratively by moving in the direction of steepest decrease of the loss function. Consider the weight matrix $\mathbf{W}$ of a given hidden layer of the network. At iteration t, the parameter update rule is given by

$$\mathbf{W}^{t+1} = \mathbf{W}^t - \eta \nabla_{\mathbf{W}} \mathcal{L}|_t, \tag{2.7}$$

where $\nabla_{\mathbf{W}} \mathcal{L}$ denotes the gradient of the loss with respect to the parameters in $\mathbf{W}$, and η is the *learning rate*, which controls the step size of each update.

Let us consider an example network, schematized in Figure 2.5, with a scalar input x, a scalar output $\hat{y}$, and target label y. For simplicity, this is a single-layer, single-neuron network, and we assume a sigmoid activation function σ, with derivative $\sigma'(x) = \sigma(x)[1 - \sigma(x)]$ for both the hidden and output layers. The hidden neuron transforms the input as $z_1 = xw_1 + b_1$, and the output layer transforms this as $z_2 = w_2\sigma(z_1) + b_2$, so that $\hat{y} = \sigma(z_2)$.

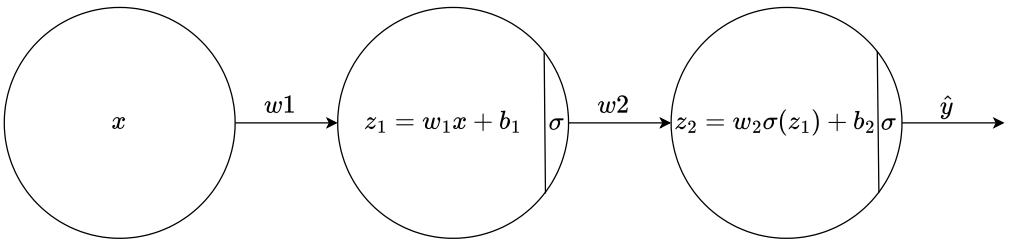

Figure 2.5: Schematic representation of a single-neuron, single-layer neural network with sigmoid activation and scalar inputs and outputs.

[8]The number of parameters n is often used as an indicator of the network *complexity*.

Following the gradient descent method and assuming a MSE loss function (Equation 2.3), the update expressions for the weights and biases are:

$$w_1^{t+1} = w_1^t - \eta \frac{\partial \mathcal{L}}{\partial y} \frac{\partial y}{\partial z_2} \frac{\partial z_2}{\partial z_1} \frac{\partial z_1}{\partial w_1}; \tag{2.8}$$

$$= \{w_1 - \eta 2(\hat{y} - y)\sigma'(z_2)w_2\sigma'(z_1)x\}|_t. \tag{2.9}$$

And following the same reasoning, we would have

$$b_1 = \{b_1 - \eta 2(\hat{y} - y)\sigma'(z_2)w_2\sigma'(z_1)\}|_t; \tag{2.10}$$

$$w_2^{t+1} = \{w_2 - \eta 2(\hat{y} - y)\sigma'(z_2)\sigma(z_1)|_t; \tag{2.11}$$

$$b_2^{t+1} = \{b_2 - \eta 2(\hat{y} - y)\sigma'(z_2)|_t. \tag{2.12}$$

This example illustrates a key concept: the further a layer is from the output, the larger the number of terms (and therefore the computational cost) in the gradient. However, due to a property of the chain rule, most of the terms used to update the last layer are also reused when updating earlier layers. This motivates the definition of

$$\delta^{(2)} = \frac{\partial \mathcal{L}}{\partial y} \frac{\partial y}{\partial z_2} = 2(\hat{y} - y)\sigma'(z_2),$$

which is called the *error*, in the sense that it measures the sensitivity of the loss with respect to the output, and does not explicitly contain information about the weights. Similarly, we can define

$$\delta^{(1)} = \frac{\partial \mathcal{L}}{\partial y} \frac{\partial y}{\partial z_2} \frac{\partial z_2}{\partial z_1} = \delta^{(2)} w_2 \sigma'(z_1).$$

The quantity $\delta^{(l)}$ measures the sensitivity of the loss up to layer l, so that the gradient can be interpreted as the sensitivity of the loss to the information multiplied by the information itself. It is therefore more computationally efficient to start the gradient computation and perform weight updates from the last layer to the first, *backpropagating* the gradient information through the layers. This algorithm, which is central to enabling the learning process in deep learning, is called *backpropagation* (Goodfellow et al., 2016; Russell and Norvig, 2010).

To conclude our toy example, using the definition of error, the weights are updated

as follows (hereafter omitting the evaluation at time t)

$$\begin{aligned} w_1^{t+1} &= w_1^t - \eta\, \delta^{(1)} x, \\ w_2^{t+1} &= w_2^t - \eta\, \delta^{(2)} \sigma(z_1), \\ b_1^{t+1} &= b_1^t - \eta\, \delta^{(1)}, \\ b_2^{t+1} &= b_2^t - \eta\, \delta^{(2)}. \end{aligned} \tag{2.13}$$

In a more general formulation with arbitrary number of layers a weight matrix $\mathbf{W}^{(l)}$ corresponding to the layer l would be updated following

$$\mathbf{W}_{t+1}^{(l)} = \mathbf{W}_t^{(l)} - \eta \boldsymbol{\delta}^{(l)} \sigma(\mathbf{z}^{(l-1)})^\top, \tag{2.14}$$

where the term $\boldsymbol{\delta}^{(l)} \sigma(\mathbf{z}^{(l-1)})^\top$ is the outer product between the activated output of the layer $l-1$ and $\boldsymbol{\delta}^{(l)}$ is the error vector that can be obtained recursively from the subsequent layers by

$$\boldsymbol{\delta}^{(l)} = \left(\mathbf{W}^{(l+1)}\right)^\top \boldsymbol{\delta}^{(l+1)} \odot \sigma'(\mathbf{z}^{(l)}), \tag{2.15}$$

where $\odot$ represents elementwise multiplication between two vectors.[9]

The learning phase of a deep learning method is therefore strongly dependent on the gradients of the network. When the network is very deep or highly complex, this dependence can lead to two well-known problems: *vanishing gradients* and *exploding gradients* (Goodfellow et al., 2016; Russell and Norvig, 2010).

Vanishing gradients occur when the gradients become extremely small as they are backpropagated through many layers, effectively preventing the weights in early layers from being updated and slowing or even halting learning. In contrast, *exploding gradients* happen when the gradients grow exponentially during backpropagation, which can cause unstable updates and numerical overflow.

Depending on how the backpropagation algorithm is implemented in terms of computational efficiency, some of the numerical challenges encountered during training can be mitigated. The algorithms responsible for performing backpropagation, thus enabling the learning of a neural network, are called *optimizers*. Each optimizer has its own characteristics. For instance, the first widely used optimizer, *Stochastic Gradient Descent* (SGD), updates the network weights directly along the negative gradient of the loss. Currently, one of the most popular optimizers is the Adaptive Moment Estimation (`Adam`) (Kingma and Ba, 2017), which combines momentum with adaptive per-parameter learning rates, making it robust to issues such as vanishing and exploding gradients. `Adam` is our choice of optimizer for all the models developed in this work.

[9] Equations 2.14 and 2.15 illustrate how the backpropagation algorithm reduces to matrix operations, which became computationally feasible with advances in graphics hardware in the early 2000s, contributing to the rise of DL.

2.3.2.3 Data Separation

Let us now recall our old friend $\mathcal{D}$. So far, we know that the dataset $\mathcal{D}$ is used during the training process, and evaluating the network's performance on labeled data allows us to extrapolate outputs to unseen data. However, the training process does not always optimize the network using the entire available dataset, nor does it process all the data at the same time.

The first concept we need to understand is that, in supervised learning, we typically split the dataset $\mathcal{D}$ into three main subsets in order to build a ready-to-use network. These subsets are:

- The *training set*: used to fit the network parameters by minimizing the loss function.
- The *validation set*: used to monitor the network's performance during training and to tune hyperparameters.
- The *test set*: used to evaluate the final performance of the trained network on unseen data, providing an unbiased estimate of generalization.

At this point, the concept of a training set[10] may sound familiar, since all the discussion regarding network learning has implicitly relied on it. However, an important component that has not yet been addressed is the validation set.

To motivate the role of the validation set, we must first understand two key scenarios that can arise in terms of the network's generalization capability: *underfitting* and *overfitting*. Figure 2.6 (*left*) qualitatively illustrates the intuition behind these scenarios. Overfitting occurs when the network is complex enough to memorize the patterns within the training set. In this case, the network achieves very high accuracy on the training data, but fails to generalize to unseen points.

On the other hand, underfitting, illustrated in Fig. 2.6(*right*), occurs when the network model is too simple to capture the underlying mapping. The predicted line may be close to the training data, but it does not represent the main behavior and will also fail to generalize.

The validation data is used to assess the network's generalization performance during training. At each iteration, we evaluate the loss function on both the training and validation sets. If the losses diverge over the iterations, this indicates that the model may be underfitting or overfitting. It is worth emphasizing that the validation data is not used to update the network parameters; it serves exclusively for monitoring.

The iterations described above have a specific name. Each time the model processes the entire training dataset, it is counted as an *epoch*. The training process consists of

[10] We will interchangeably use the terms *training set*, *training data*, and *training sample*, as well as the corresponding terms for the validation and test sets.

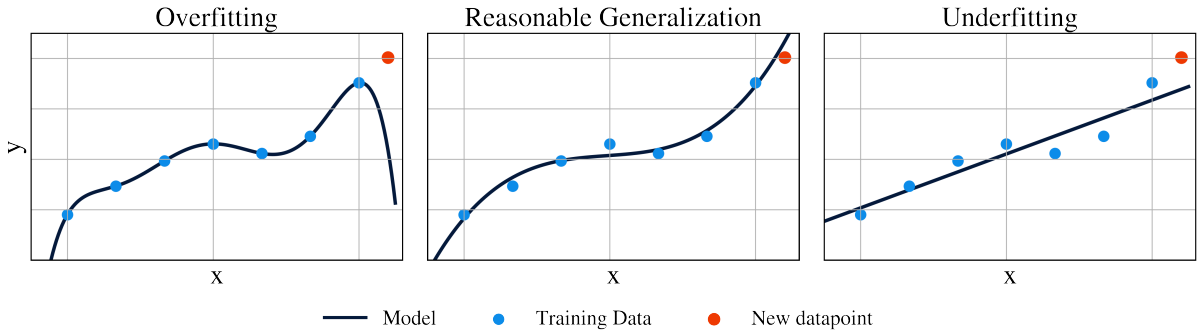

Figure 2.6: Illustration of generalization regimes in supervised learning. left panel: overfitting. center panel: reasonable generalization. right panel: underfitting.

repeatedly presenting the examples from the training set over many epochs, until the loss function converges.

Depending on the size of the training dataset, it may be computationally expensive to process all the data at once. This is because, during training, the optimizer must store the outputs of all neurons for each input in order to perform backpropagation, which can lead to unmanageable memory requirements. Moreover, updating the weights only once per epoch may result in very slow learning process.

To address this, the data is presented to the network in *batches*, which are smaller subdivisions of the training set. Many optimizers, such as `Adam`, compute the loss and update the network parameters after each batch. The *batch size* is therefore an important hyperparameter in model development. The optimal batch size depends on both the computational resources available and the nature of the task. Very large batches can increase training time, overload the `RAM`, and may fail to capture nuances in the loss landscape, leading to suboptimal learning. Conversely, very small batches can produce noisy updates, potentially slowing convergence and reducing the stability during the learning.

One of the most effective ways to visualize the training process of a model is by recording the loss values at each epoch and plotting the resulting *loss curve*[11]. Figure 2.7 illustrates different scenarios of loss curves corresponding to common training issues.

2.3.2.4 Regularization

So far, we have seen a variety of problems that can occur during the training of a *Neural Network* (NN), such as overfitting, (vanishing or) exploding gradients, underfitting, slow training, among others. The set of techniques used to prevent, or at least mitigate, the effects of these problems with the aim of improving the model's generalization is called *regularization*.

[11] Also known as training curve; however, this term can also refer to cases where the monitored metric is not necessarily the loss.

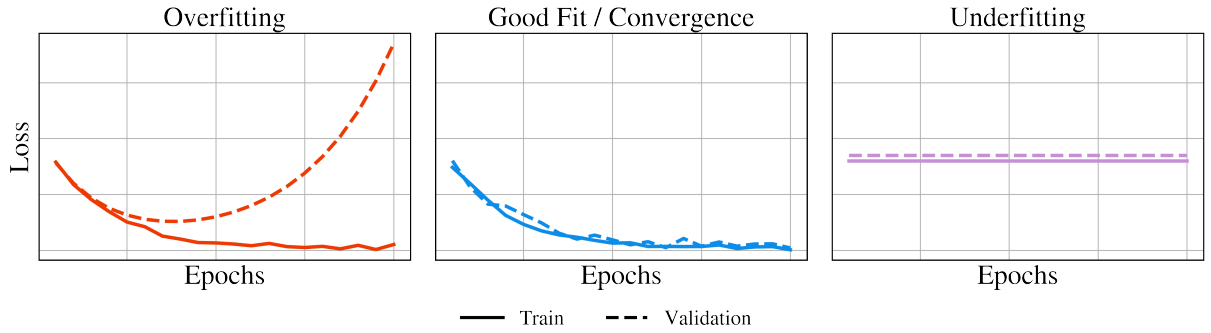

Figure 2.7: Examples of loss curves during training. (left): overfitting where the training loss continues to decrease while the validation loss increases. (center): a well-behaved training process where training and validation losses decrease and converge. (right) underfitting where both training and validation losses remain high.

For example, one possible definition of overfitting in the context of neural networks states that the models are *"memorizing properties of the training set that do not serve them well on the test set"* (Goodfellow et al., 2016). This generally occurs when the size of the network is too large relative to the number of data points in the training sample. There is a balance between the size of the network and its ability to recognize patterns, which means that making the network smaller does not always solve overfitting. Sometimes, the network needs to be large enough to learn meaningful pattern representations, even when the amount of data is limited. To address the issue of co-adaptive neurons, a widely used technique is *dropout* (Srivastava et al., 2014a). Dropout roughly consists of setting the activation of neurons in a given layer to zero[12] with a probability p at each batch – $p = 0.5$ is a common choice for hidden layers. This prevents individual neurons from adapting too specifically to the training data, forcing the network to extract representations from the population of neurons rather than from individual examples.

Another common practice in regularization involves normalizing the numerical inputs of the data as well as the network weights. For example, the $L1$ and $L2$ regularization approaches consist of introducing a term in the loss function that penalizes the network when the weights become too large, which can lead to exploding gradients. These methods also contribute to *smoothing* the model, i.e., reducing the sensitivity of the network to small variations in the input.

One technique that also contributes to preventing exploding gradients – although this is not its main purpose – is *batch normalization*. As the name suggests, this technique normalizes the input of each layer for every batch, usually using standard normalization by setting the mean of the distribution to zero and the standard deviation to one. Batch normalization primarily impacts the learning process, smoothing it and reducing sensitivity to the initial weights of the network. Beyond batch normalization, it is also common

[12] In standard implementations, dropout sets a fraction of activations to zero during training. Variants exist, such as `GaussianDropout`, which applies multiplicative random noise to the activations instead of zeroing them (Abadi et al., 2015; Srivastava et al., 2014b).

practice to normalize the input distribution of the training dataset (and, in some cases, the outputs) to reduce strong dependencies on early layers and further facilitate the training process.

Even when these techniques are applied, the network can begin overfitting after a certain number of epochs. Instead of retraining the model from scratch, we can retain the version of the model corresponding to the point at which the training and validation losses were still converging. This approach is called *early stopping*. Although it does not directly regularize the model or the data, it prevents the deployment of an overfitted model, preserving the highest possible generalization performance.

2.3.2.5 Cross Validation

If $\mathcal{D}$ is sufficiently small, the separation between training and validation data may lead to a biased learning curve. This means that the model may appear to converge well for that specific choice of training and validation split. In such cases, when applying the model to unlabeled data that are not closely related to the training/validation domain, the results may be unreliable. One way to address this issue is to evaluate the training convergence across different combinations of training and validation samples. This technique is called *cross-validation* (Breiman and Spector, 1992; Stone, 1974).

Figure 2.8 illustrates one of the most commonly used cross-validation algorithms, also employed in this work, namely K-Fold cross validation. The procedure starts by splitting $\mathcal{D}$ into training and test sets. The training set is then further divided into K equal folds. We train K separate models with identical architectures and hyperparameters, each using one of the K folds as the validation set while the remaining folds form the training data. If the model is robust with respect to the data, and the dataset is sufficiently representative (i.e., no small random sample differs significantly from the overall distribution), the training curves will be similar. Conversely, if there is high variance among the training curves, this indicates that the model is sensitive to the specific splits used.

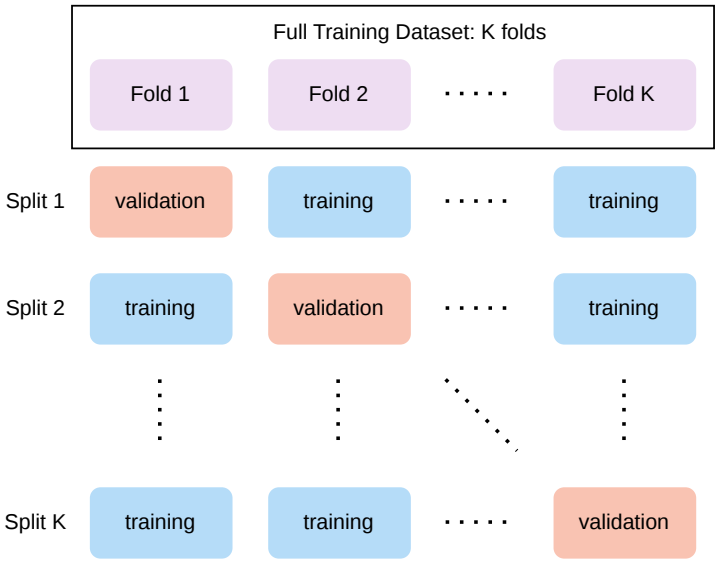

Figure 2.8: Illustration of the K-fold cross-validation procedure. The dataset is divided into K folds. A separate model is trained from scratch for each train/validation split, in which a different fold is used as the validation set while the remaining folds form the training set.

2.4 | The Neural Net Zoo

So far, we have covered most of the key basic concepts of deep learning. For simplicity, all examples and derivations were based on the simplest type of NN, the MLP. Although MLPs have a broad range of potential applications, they are naturally limited. For instance, the input layer can only receive one-dimensional vectors. When dealing with images, we have to *flatten* them (representing the image as a vector containing all pixel values), which prevents traditional MLPs from effectively capturing complex spatial correlations.

A class of models that has played a central role in the development of modern computer vision are the CNN (LeCun et al., 1989). These networks process images by convolving windows of the input across learnable filters. The resulting feature maps within the network retain a spatial structure, allowing them to capture important spatial patterns. Although CNNs have been crucial for advances in AI and are widely used in astrophysics (Bom et al., 2023; Pasquet et al., 2019; Monsalves et al., 2024), they are not employed in this work. For a more detailed study, we refer the reader to Goodfellow et al. (2016).

Another example arises when processing time series, particularly in the context of *Natrual Language Processing* (NLP). Enabling computers to understand the meaning of words has been a long-standing challenge since the foundations of AI. The problem with using MLP is that, due to their structure, the network would treat the input words in a

sentence as independent features. This is fundamentally incorrect, as the context – and therefore the meaning of the sentence – depends on the order of the words. Analogous to CNN for images, there is a specific family of networks designed to handle sequential data: RNNs (Rumelhart et al., 1986).

The range of possible architectures extends even further when including probabilistic models and, of course, the numerous variations derived from CNN, RNN, and other base architectures. This wide variety of NN architectures, each designed for different types of problems that can potentially be addressed with DL, collectively forms a diverse *Neural Network Zoo* (Leijnen and Veen, 2020).

In this section, we explore different architectures of NN, with particular focus on those employed throughout this thesis. The key DL concepts discussed in the previous sections can be extended to the NN presented here with only minor conceptual adaptations (although the implementation details may change substantially), so we will not go into extensive detail. Our attention will instead be focused on the main structure of the networks and how these structures relate to the problems we aim to solve.

2.4.1 | Recurrent Neural Networks

As introduced previously, RNNs were originally designed to address problems with an inherent sequential structure. These architectures began to emerge in the late 1980s and early 1990s, with foundational contributions such as Rumelhart et al. (1986), Elman (1990), and Elman (1990).

The fundamental difference between RNN and MLP lies in the structure of the inputs. In a standard MLP, the input is a fixed-size vector

$$\mathbf{x} = (x_1, x_2, \ldots, x_n). \tag{2.16}$$

In contrast, in the sequential setting, a single input is a sequence of vectors:

$$\mathbf{x} = \left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\tau)}\right), \tag{2.17}$$

where τ denotes the length of the sequence.

In this context, the order of elements is crucial. Each element of the sequence is indexed by a time step t, and the model processes the sequence while maintaining an internal hidden state that carries information from previous steps. This temporal dependency allows RNNs to capture sequential correlations that standard feedforward MLPs cannot represent.

Figure 2.9 illustrates the simplest architecture of a RNN. Given an input sequence

$\mathbf{x}$, the network computes hidden states $\mathbf{h}^{(t)}$ at each time step t, which depend on the current input $\mathbf{x}^{(t)}$ and the previous hidden state $\mathbf{h}^{(t-1)}$. In this way, the network can capture dependencies across the sequence over time. This architecture also introduces the concept of a network's *memory*, as information from the first elements of the sequence are stored and propagated through the network.

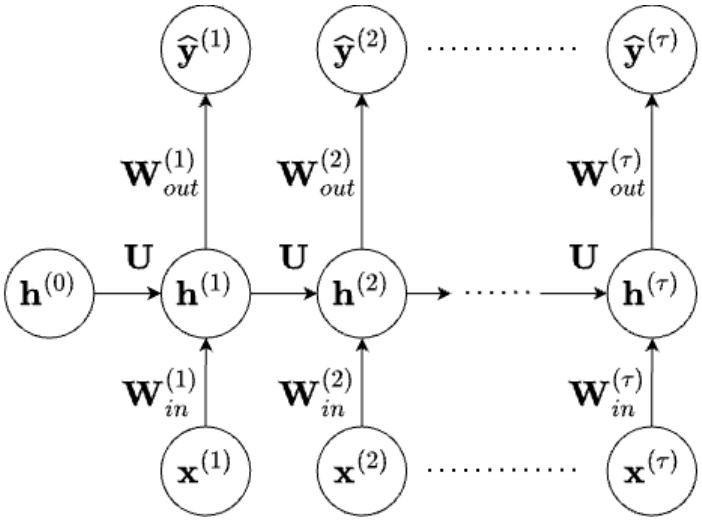

Figure 2.9: Schematic representation of a simple RNN architecture.

At time step t, an activated hidden state $\mathbf{h}^{(t)}$ is given by

$$\mathbf{h}^{(t)} = \sigma\left(\mathbf{U}\mathbf{h}^{(t-1)} + \mathbf{W}_{in}\mathbf{x}^{(t)} + \mathbf{b}\right). \tag{2.18}$$

In this formulation, the matrix $\mathbf{W}$ is the weight matrices that transform the input information, $\mathbf{U}$ refers to the weights that propagate information across time within the hidden layer, $\mathbf{b}$ is the bias vector, and σ a general activation function (usually ReLU for inner layers in vanilla RNN). One of the key concepts behind RNN are the *shared wights*. Here $\mathbf{U}$ and $\mathbf{W}$ are fixed in a given layer, i.e., it does not change across time stamps, therefore the different hidden states shares the same transformation across time, so the network can make predictions over sequences of different lengths τ.

The t-th element of the output sequence $\hat{\mathbf{y}}$ is then defined by

$$\hat{\mathbf{y}}^{(t)} = \sigma\left(\mathbf{W}_{out}\mathbf{h}^{(t)} + \mathbf{b}_{out}\right). \tag{2.19}$$

RNNs elevate learning to a higher level of complexity because the gradient must now be propagated through both layers and time. The *backpropagation through time* (BPTT) algorithm is conceptually similar to the traditional backpropagation discussed in Section 2.3.2.2, with the main difference being that the hidden layers must account for two sources of the gradient: one from the subsequent layers, and another from subsequent time steps. This additional computation makes learning in RNNs more complex and, in some cases, numerically challenging.

2.4.1.1 Legendre Memory Units

The previous example serves to introduce the fundamental components of RNNs. However, when dealing with long sequences, the memory of relevant information from the *past* (early time steps) may gradually degrade over time. Furthermore, because the same weight matrices are repeatedly applied across many time steps, RNNs are particularly susceptible to the vanishing (or exploding) gradient problem, depending on the magnitude of the weights. This effect can severely hinder the learning process.

Around the late 1990s, several methods were proposed to address the main limitations of RNNs. Among them, the architecture that most successfully mitigated the vanishing gradient problem was the *Long Short-Term Memory* (LSTM) (Hochreiter and Schmidhuber, 1997; Gers et al., 2000).

In short, the LSTM processes and propagates short-term information while preserving longer term information across many time steps through two distinct pathways: the hidden states $\mathbf{h}^{(t)}$ and the cell state $\mathbf{s}^{(t)}$. This mechanism is enabled by the use of *gates*, which regulate the flow of information within the cell. There are three main gates in the LSTM cell: the *forget gate* $\mathbf{f}$, *input gate* $\mathbf{i}$, and the *output gate* $\boldsymbol{\phi}$[13]

Figure 2.10 illustrates the information flow in an LSTM cell. The cell state $\mathbf{s}$ carries the long-term memory across time steps, and is modulated by three gating mechanisms driven by the previous hidden state $\mathbf{h}^{(t-1)}$ and the current input $\mathbf{x}^{(t)}$. The forget gate $\mathbf{f}$ controls how much of the previous cell state is retained. The input gate $\mathbf{i}$ determines how much of the new information is added to the cell state. Finally, the output gate controls how much of the updated cell state is exposed as the new hidden state $\mathbf{h}^{(t)}$.

The gates are represented by sigmoid activations σ. Formally speaking, the cell state $\mathbf{s}^{(t)}$ is given by

$$\mathbf{s}^{(t)} = \mathbf{f}^{(t)} \odot \mathbf{s}^{(t-1)} + \mathbf{i}^{(t)} \odot \sigma\left(\mathbf{U}\mathbf{h}^{(t-1)} + \mathbf{W}\mathbf{x}^{(t)} + \mathbf{b}\right), \tag{2.20}$$

where the matrix $\mathbf{U}$ and $\mathbf{W}$ have the same meaning of 2.18. The gate vectors also make use of idependent weight matrix like that. The gate vectors $\mathbf{f}^{(t)}$ and $\mathbf{i}^{(t)}$, as well as $\boldsymbol{\phi}^{(t)}$ have a similar shape. In general, given gate vector $\mathbf{g}^{(t)}$ is given by

$$\mathbf{g}^{(t)} = \sigma\left(\mathbf{U}_{\mathrm{g}}\mathbf{h}^{(t-1)} + \mathbf{W}_{\mathrm{g}}\mathbf{x}^{(t)}\right). \tag{2.21}$$

The weight matrixs of each gate are independent, and are kept the same towards different time steps. Finally, the hidden state $\mathbf{h}^{t}$ is given by

$$\mathbf{h}^{(t)} = \boldsymbol{\phi}^{(t)} \odot tanh(\mathbf{s}^{(t)}). \tag{2.22}$$

[13]The letter o is avoided for the output gate due to its potential ambiguity with other mathematical symbols.

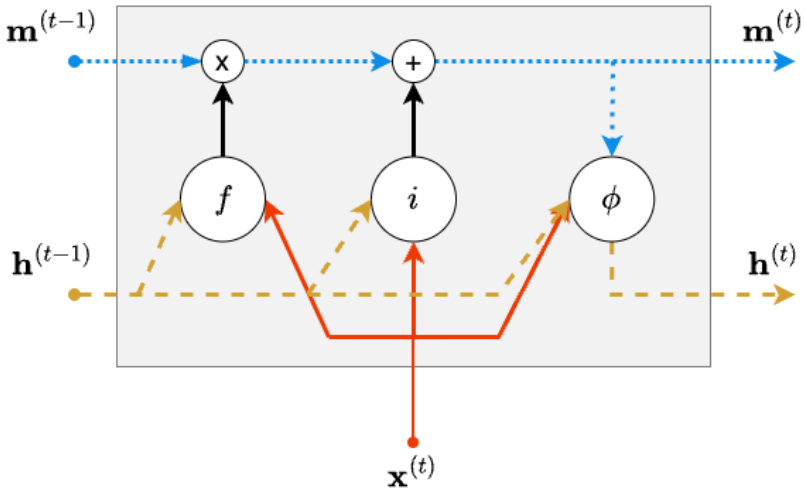

Figure 2.10: Schematic diagram of an LSTM cell. The cell state $\mathbf{s}$ *carries long-term memory across time steps and is modulated by three gates – forget (*$\mathbf{f}$*), input (*$\mathbf{i}$*), and output – which control the flow of information from the previous hidden state* $\mathbf{h}^{(t-1)}$ *and current input* $\mathbf{x}^{(t)}$ *to the updated cell state and new hidden state* $\mathbf{h}^{(t)}$*.*

Understanding the LSTM workflow is crucial to gaining intuition about other long-term memory architectures. Many natural processes evolve in continuous time, and it is unclear how LSTMs perform in such situations, since the memory capacity required can be very large. In other words, continuous-time problems often correspond to sequences that are extremely long (on the order of thousands of time steps) (Arjovsky et al., 2016; Le et al., 2015; Li et al., 2019).

Legendre Memory Units (LMU) (Voelker et al., 2019) are a type of recurrent cell designed to capture long-range dependencies even when the input is nearly continuous. The key idea is to represent the recent history of the input signal within a sliding time window of length θ using an orthogonal basis of Legendre polynomials.

Figure 2.11 illustrates the schematic of an LMU cell. Its structure is somewhat similar to that of an LSTM, but with important differences: there are no gating units, and the memory state vector does not store the input signal directly. Instead, it stores the coefficients of the Legendre polynomial expansion that approximates the input history over the past θ time units.

In this way, u_t represents the input signal, while $\mathbf{m}^{(t)}$ is the memory vector of length d, containing the coefficients of the expansion that allow reconstruction of the history of u over the past θ time units:

$$u(t-\bar{t}) = \sum_{i=0}^{d-1} m_i^{(t)} \mathcal{P}_i\left(\frac{\bar{t}}{\theta}\right), \quad 0 < \bar{t} < \theta, \tag{2.23}$$

where $\mathcal{P}_i$ are the Legendre polynomials of order i. The memory is dynamically updated according to

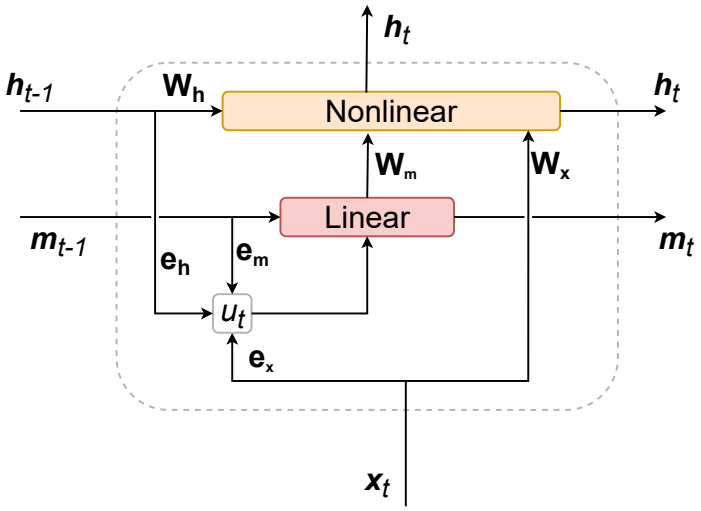

Figure 2.11: Structure of the LMU cell. Adapted from Voelker et al. (2019).

$$\mathbf{m}^{(t)} = \mathbf{A}\,\mathbf{m}^{(t-1)} + \mathbf{B}\,u^{(t)}, \tag{2.24}$$

where $\mathbf{A}$ and $\mathbf{B}$ are non-trainable, analytically derived matrices that satisfy the requirements of this dynamical system (Voelker et al., 2019).

The vectors $\mathbf{e}_\mathrm{h}$, $\mathbf{e}_\mathrm{m}$, and $\mathbf{e}_\mathrm{x}$ are encoding vectors that form the input representation u_t at time t:

$$u^{(t)} = (\mathbf{e}_\mathrm{h})^T\mathbf{h}^{(t-1)} + (\mathbf{e}_\mathrm{m})^T\mathbf{m}^{(t-1)} + (\mathbf{e}_\mathrm{x})^T\mathbf{x}^{(t)}, \tag{2.25}$$

and the hidden states $\mathbf{h}_t$ are computed as

$$\mathbf{h}^{(t)} = \sigma(\mathbf{W}_\mathrm{m}\mathbf{m}^{(t)} + \mathbf{W}_\mathrm{h}\mathbf{h}^{(t-1)} + \mathbf{W}_\mathrm{x}\mathbf{x}^{(t)}, \tag{2.26}$$

where the matrices $\mathbf{W}$ represent the network weights, and σ is a non-linear activation function (typically *tanh*).

Although photometric redshifts problem does not relate with continuous sequence, the LMU composed one of the main methods developed in this thesis. The motivation behind will be discussed in Chapter 4.

2.4.2 | Mixture Density Networks

The DL methods discussed so far are *deterministic*, i.e., after training, a given input $\mathbf{x}$ will correspond to a single unique output $\mathbf{y}$ every time it is processed. When performing regression, this leads to a major issue in modeling the uncertainty of the outputs. For instance, this is one of the main challenges in photometric redshift estimation (Dey et al.,

2022; D'Isanto and Polsterer, 2018).

Many DL models are designed to capture the probabilistic nature of problems by estimating the PDF of $\mathbf{y}$ given $\mathbf{x}$. This can be approached either by making the model *stochastic*, producing a different output for $\mathbf{y}$ each time $\mathbf{x}$ is forwarded, or by modeling the PDF of $\mathbf{y}$ deterministically, where the model outputs the full probability distribution for a given input.

Among the approaches for deterministically estimating PDFs are MDN (Bishop, 1994). MDN combine a neural network with a parametric mixture model (McLachlan and Basford, 1988). Unlike traditional neural networks that predict a single value for the target parameter, an MDN outputs a conditional PDF represented as a linear combination of multiple individual distributions (kernels). This allows for a more complete characterization of predictions, including uncertainty estimation.

The neural network outputs the parameters and weights (called *mixing coefficients*) of each distribution conditioned on the inputs, which are then combined to build the mixture model. In principle, a mixture model can be flexible enough to approximate any arbitrary distribution, making MDNs a powerful tool for regression problems (McLachlan and Basford, 1988; Bishop, 1994).

In Lima et al. (2022) we showed that an MDN using a Gaussian Mixture Model (Reynolds et al., 2009) outperformed both template fiting and traditional NN methods when inferring z_{phot} PDFs with S-PLUS data. In this work, we employ a similar model, where the resulting PDF is given by a linear combination of C Gaussian kernels, with *mixing coefficients* $\{\alpha_k\}$, means $\{\mu_k\}$, and standard deviations $\{\sigma_k\}$:

$$\mathrm{p}(z) = \sum_{i=1}^{C} \alpha_i \mathcal{N}(z|\mu_i, \sigma_i), \tag{2.27}$$

with constraints $\sum_{i=1}^{C} \alpha_i = 1$ and $0 < \alpha_i < 1$. The training is performed by minimizing the log-likelihood loss function defined in Equation 2.5. In the methods we developed, the MDN is combined with LMUs and MLPs to estimate photometric redshifts. The implementation of the MDN adopted here is purely deterministic; however, we acknowledge that hybrid approaches combining deterministic and stochastic components are also possible, as presented in Nakazono et al. (2024) and Lima (2025).

2.4.3 | Bayesian Neural Networks

Unlike their deterministic counterparts (including probabilistic models such as the MDN), stochastic models produce probability distributions by explicitly introducing randomness into the network parametrization – or, in some cases, into the inputs or outputs

(e.g., through bootstrapping).

BNNs (Neal, 1996) form a class of stochastic models grounded in Bayesian inference, in which uncertainty is represented by placing probability distributions over the network parameters rather than treating them as fixed values. As a consequence, for a given input $\mathbf{x}$, a single trained network yields different outputs $\mathbf{y}$.

Within the Bayesian framework, probability is interpreted as a degree of belief[14]. In this formulation, we consider two fundamental sets: a set of hypotheses $\mathcal{H}$ describing possible explanations of a system, and a dataset $\mathcal{D}$ containing observations of that system. The data update our belief about the hypotheses according to Bayes' rule:

$$p(\mathcal{H} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathcal{H})\, p(\mathcal{H})}{p(\mathcal{D})}, \tag{2.28}$$

where $p(\mathcal{H})$ is the prior distribution, representing our belief about the hypotheses before observing any of $\mathcal{D}$. The term $p(\mathcal{D} \mid \mathcal{H})$ is the likelihood, which quantifies how probable the observed data are under a given hypothesis. The denominator $p(\mathcal{D})$, known as the evidence or marginal likelihood, ensures proper normalization and can be written as

$$p(\mathcal{D}) = \int_{\mathbb{H}} p(\mathcal{D} \mid \mathcal{H}')\, p(\mathcal{H}')\, d\mathcal{H}', \tag{2.29}$$

where $\mathbb{H}$ denotes the hypothesis space. Finally, the posterior distribution $p(\mathcal{H} \mid \mathcal{D})$ represents the updated belief about the hypotheses after observing the data. In most Bayesian inference problems, the primary objective is to characterize or approximate the posterior distribution over $\mathcal{H}$.

Let us recall the very early discussion on ML in Section 2.2. In this context, the function $\tilde{f} : \mathbb{X} \to \mathbb{Y}$ would represent hour hypothesis, and our goal is to find the distribution of possible $\tilde{f}$ that best describe the system from which $\mathcal{D}$ was measured. The function $\tilde{f}$ in case of NN is then represented by the weights $\mathbf{w}$ – for this discussion $\mathbf{w}$ represents all the weights of the network. The BNN approach is then to find

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}. \tag{2.30}$$

That way, we can asses the predictions $\mathbf{y}$ by sampling $\tilde{f}_{\mathbf{w}}(\mathbf{x})$. In this context, the likelihood $p(\mathcal{D}|\mathbf{w})$ is defined by the choice of the network architecture and the loss function. For instance, considering a given data example $\{\mathbf{x}, \mathbf{y}\}$ of the network, a gaussian likelihood would then be defined as

$$p(\{\mathbf{x}, \mathbf{y}\}|\mathbf{w}) = \mathcal{N}(\mathbf{y}|\boldsymbol{\mu} = \tilde{f}_{\mathbf{w}}(\mathbf{x}), \boldsymbol{\sigma}), \tag{2.31}$$

[14] In contrast, the *frequentist* interpretation defines probability as the limiting frequency of an event over an infinite number of repeated trials.

where $\boldsymbol{\sigma}$ represents the error in the network prediction.

Two major concerns arise when looking to the other terms: how to choose the prior $p(\mathbf{w})$, and how to compute the evidence $p(\mathcal{D})$. For the former, indeed it is obscure how to get any information about the weights of an arbitrary network before any contact with $\mathcal{D}$. In cases like that, we recognize our ignorance by setting uninformative priors, typically gaussians with small deviation centered in zero. The idea behind that is that with iterations after looking the data, the influence of the prior is minimal ("Goan and Fookes, 2020).

Now, for the latter, this difficulty is not specific to BNN, but arises more generally in Bayesian inference. Computing the evidence integral is computationally expensive and, in most practical settings, intractable due to the high dimensionality of the parameter space $\mathbf{w}$. Moreover, when the prior is weakly informative, a significant portion of the probability mass may be spread over a large region of the parameter space, requiring extensive exploration and further increasing the computational cost ("Goan and Fookes, 2020; Neal, 1996). Then, how do we access the posterior distribution? A solution to this problem, and the core strategy behind the success of BNNs, arrises in the context of *variational inference* (Jordan et al., 1999; Blei et al., 2017). In this context, rather than computing the posterior directly, we aim to approximate it using a tractable parametric distribution $q_\theta(\mathbf{w})$ parameterized by θ. That way, our problem is framed by a minimization of the difference between $q_\theta(\mathbf{w})$ and $p(\mathbf{w}|\mathcal{D})$. This difference is measured by the Kullback-Leiber (KL divergence (Cover and Thomas, 2005)

$$\mathrm{KL}(q_\theta(\mathbf{w})||p(\mathbf{w}|\mathcal{D})) = \int d\mathbf{w} q_\theta(\mathbf{w}) \log \frac{q_\theta(\mathbf{w})}{p(\mathbf{w}|\mathcal{D})} \tag{2.32}$$

$$= \int d\mathbf{w} q_\theta(\mathbf{w}) \left[\log \frac{q_\theta(\mathbf{w})}{p(\mathbf{w})} - \log p(\mathcal{D}|\mathbf{w}) + \log p(D) \right] \tag{2.33}$$

$$= \mathrm{KL}(q_\theta(\mathbf{w})||p(\mathbf{w})) - \mathbb{E}_{\mathbf{w}\sim q_\theta}[\log p(\mathcal{D}|\mathbf{w})] + \log p(\mathcal{D}). \tag{2.34}$$

This means that training the network to best approximate q_θ to the true prior – i.e., adjusting the parameters via gradient descent to minimize the KL divergence between the distributions – does not require computing the evidence. Equation 2.34 then defines what we call the *Evidence Lower Bound (ELBO)* loss function ("Goan and Fookes, 2020; Blei et al., 2017). More explicitly:

$$\mathcal{L}_{\mathrm{ELBO}} = \mathrm{KL}\big(q_\theta(\mathbf{w}) \,\|\, p(\mathbf{w})\big) - \mathbb{E}_{\mathbf{w}\sim q_\theta}\big[\log p(\mathcal{D} \mid \mathbf{w})\big]. \tag{2.35}$$

Summarizing, a simple MLP-based BNN replaces the deterministic network weights with probability distributions. At each iteration, the computations are the same as described in Section 2.3.2, but the weights are sampled from the defined distributions. In

this framework, what is actually optimized are not the weights $\mathbf{w}$ themselves, but the parameters θ of each parametrized distribution q_θ.

We will recall BNNs in the second part of this thesis in the context of variational inference of physical parameters for *supernova* models.

3 Data

The photometric redshifts presented in this thesis were developed for the DELVE. In this chapter, we will go into the main characteristic of the survey, its data aspects, and how this influenced the proposed approaches for the photo-z estimation. We will also visit the construction of the truth table sample, that consits on a large compilation of spectroscopic redshifts.

3.1 | The Spectroscopic Sample

Recalling the discussion in 2.3, the models developed in this work rely on supervised learning. Therefore, a dataset of spectroscopic redshifts is required to construct the training, validation, and test samples for the network.

As described in the published version of this work (Teixeira et al., 2024), we initially combined three major spectroscopic redshift compilations: the spectroscopic sample described in Gschwend et al. (2018), the *Dark Energy Spectroscopic Instrument Early Data Release* (DESI-EDR) catalog (DESI Collaboration et al., 2023), and the *Southern Hemisphere Spectrocopic Redshift Compilation* (SHSC) (Lima, 2024). However, we later verified that the SHSC already incorporates the former two datasets.

The SHSC is a compilation of spectroscopic measurements gathered from several public surveys and databases. The catalog is assembled through automated queries to archival services such as *VizieR Catalog Service* (VizieR) (Ochsenbein et al., 2000), *High Energy Astrophysics Science Archive Research Center* (HEASARC)[1], and SDSS (York et al., 2000), after which the retrieved catalogs are merged into a single spectroscopic dataset. Figure 3.1 shows the sky coverage of the SHSC sample.

The compilation pipeline performs cross-matching between catalogs, applies basic quality checks, and assigns object classifications when available. It also includes an additional cross-match with Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD)[2] to recover classifications for sources lacking this information. In our analysis, we excluded objects classified as stars or transient sources. At the time of the publication of this work, no additional restrictions were imposed on the spectroscopic catalog. However, in more recent versions of the photometric redshift

[1] https://heasarc.gsfc.nasa.gov/
[2] https://simbad.cds.unistra.fr/simbad/

pipeline, we verified that the spectroscopic redshift uncertainty was smaller than 0.0004 for approximately 99% of the catalog.

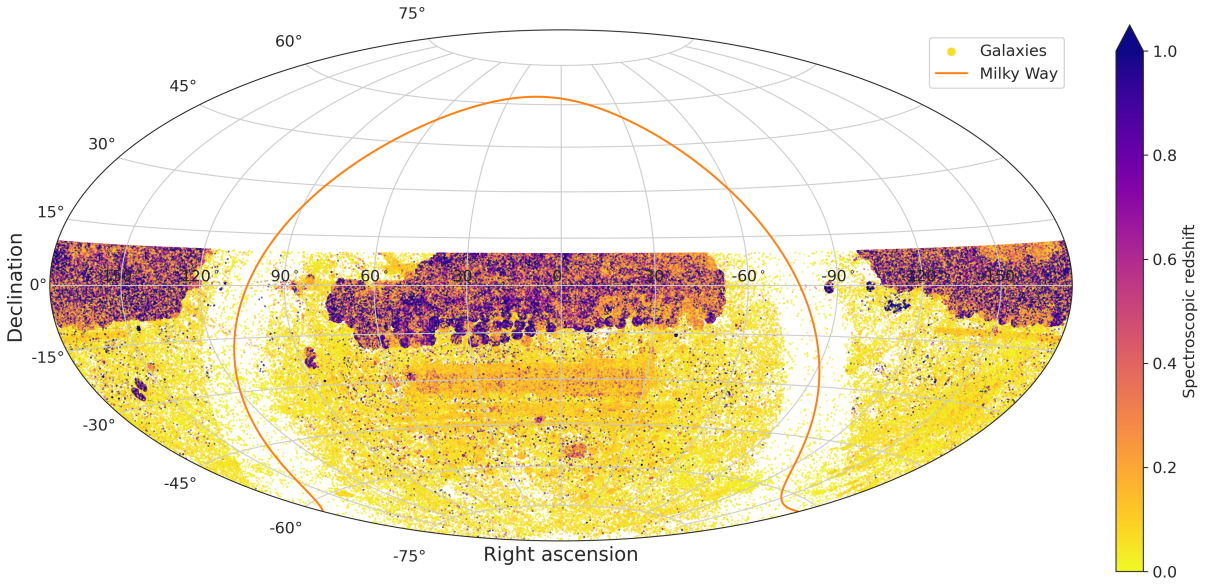

Figure 3.1: Sky coverage of the SHSC spectroscopic compilation used in this work. The figure was retrieved from `https://github.com/ErikVini/`.

3.2 | The DELVE Survey

DELVE is an ongoing survey primarily motivated by the study of satelite, ultra faint galaxies in the Milky Way's Local Group visible from the southern hemisphere (Drlica-Wagner et al., 2021). The survey is divided into three main programs: *DELVE-WIDE*, designed to complement *Dark Energy Camera* (DECam) coverage at high Galactic latitudes and provide accurate observations of ultrafaint nearby galaxies; *DELVE-MC*, which aims to improve our understanding of the population of faint satellites around the Magellanic Clouds; and *DELVE-DEEP*, which targets the surroundings of four nearby galaxies similar to the Magellanic Clouds in terms of stellar mass, enabling the study of low-mass satellite systems around non-massive host galaxies.

In the DELVE second data release (DELVE-DR2), they combined archival data from the *Dark Energy Survey* (DES), the *Dark Energy Camera Legacy Survey* (DECaLS), and more than 270 publicly available community programs (Drlica-Wagner et al., 2022). This release encompasses an area of 21,000 $\deg^2$, of which 17,000 $\deg^2$ have been observed in all four broadband filters (g, r, i, and z). Figure 3.2 shows the coverage of DELVE-DR2 in each of the four bands, as well as the area covered simultaneously in all bands.

The DELVE exposures were processed using the DES Data Management "Final Cut" pipeline (Morganson et al., 2018). Source detection and photometric measure-

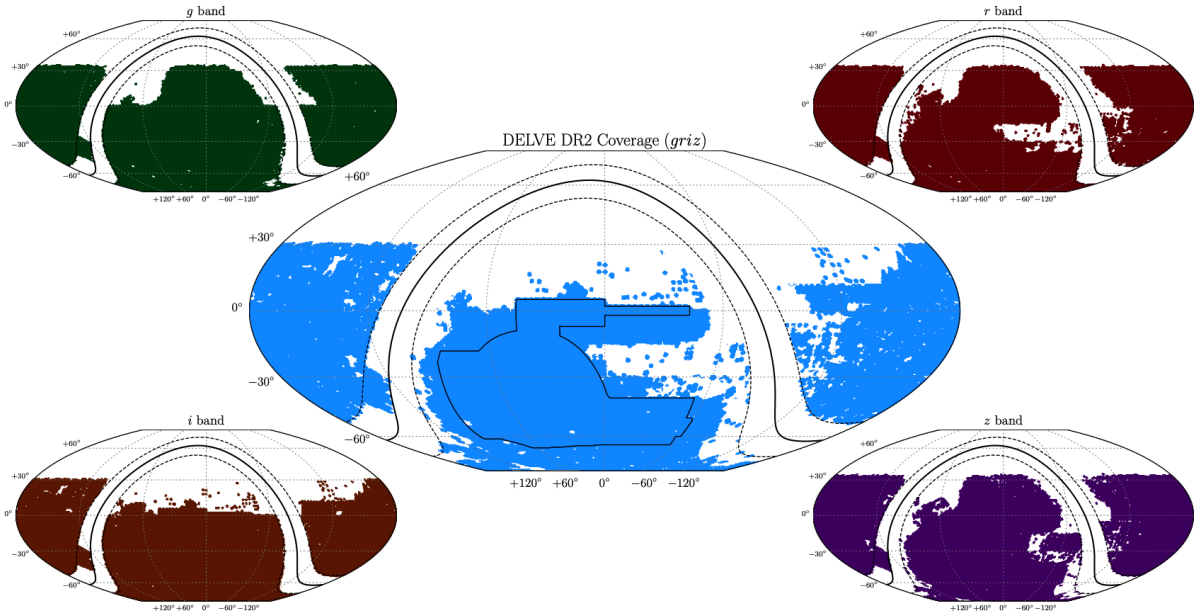

Figure 3.2: DELVE-DR2 survey footprint in the g, r, i, and z bands. Figure from Drlica-Wagner et al. (2022)

ments were performed with SourceExtractor (Bertin and Arnouts, 1996) and PSFEx (Bertin, 2011). For the data included in DELVE-DR2 that overlap with the DES footprint, the photometric zeropoint calibration was carried out using the Forward Global Calibration Method (Burke et al., 2017). Exposures outside the DES footprint were calibrated against the The Asteroid Terrestrial-impact Last Alert System All-Sky Stellar Reference Catalog (Tonry et al., 2018), applying color transformations derived from stellar loci, as described in Drlica-Wagner et al. (2022).

The original DELVE observations are limited to declinations $\lesssim 30°$ and reach 5σ point-source depths of 24.3, 23.9, 23.5, and 22.8 mag in the g, r, i, and z bands, respectively. When considering the combined observations from other DECam surveys included in the release, the catalog is nearly homogeneous to a depth of ~ 22.5 mag. The DELVE-DR2 catalog comprises approximately 2.5 billion unique astronomical sources, of which about 618 million have measurements available in all four photometric bands.

3.2.1 | Motivation

As mentioned previously, DELVE is a survey focused on satelite galaxies. However, having such a wide sky coverage and exceptional depth over such different regions area is crucial for several other fields, including observational cosmology and astrophysics (Anbajagane et al., 2025; Chan et al., 2022; Bom et al., 2024; Sifón et al., 2025; Drlica-Wagner et al., 2018). Up to its DELVE-DR2, the DELVE collaboration had not yet provided

redshift measurements for the large number of galaxies detected in the survey.

In this context, we joined the collaboration with the proposal of developing a novel catalog of photometric redshifts for the approximately 21 000 $\deg^2$ survey footprint. At that time, DL methods were already emerging as powerful tools for photometric redshift estimation, particularly probabilistic approaches capable of providing robust uncertainty estimates (Lima et al., 2022; Dey et al., 2022; Henghes et al., 2021; Henghes et al., 2022; D'Isanto and Polsterer, 2018).

These factors motivated the development of the official photometric redshift catalog for the collaboration[3]. The following sections describe the main steps involved in preparing the data and training the deep learning models.

3.3 | Pre-Processing and Sample Definitions

3.3.1 | MCAT

DELVE-DR2 is an extended catalog that inevitably contains undesired objects or spurious detections To ensure the purity and reliability of the final product, we applied a series of selection cuts to the downloaded catalog. This section describes the general cuts applied to the raw data.

To remove contaminant stars present in the catalog, we adopted the galaxy–star separation criteria used in the DES Year 1 results (Drlica-Wagner et al., 2018). This method combines the `SourceExtractor` parameters `SPREAD_MODEL` and its associated error (`SPREADERR_MODEL`) to construct the object classification flag `MODEST_CLASS`. The `SPREAD_MODEL` parameter measures the discriminant between the local *Point Spread Function* (PSF) model and a circular exponential disk model convolved with the PSF at the position of each source (Desai et al., 2012; Soumagnac et al., 2015).

The `MODEST_CLASS` flag is defined through a set of boolean operations involving the `SPREAD_MODEL` parameter and its uncertainty. In Drlica-Wagner et al. (2018), the reference band adopted for the `MODEST_CLASS` definition is the i band. However, following Drlica-Wagner et al. (2022), we adopt the g band as the reference because of its higher sky coverage and deeper magnitude limit. Table 3.1 presents the definition of each class. We exclude classes 0 and 3, keeping only high-probability galaxies and ambiguous classifications.

In addition, we considered only objects with `FLAGS` $\leqslant 2$ from the SourceExtractor detection process. The `FLAGS` parameter is provided as a column in the DELVE-DR2

[3]Now available at https://datalab.noirlab.edu/data/delve#delve-dr2

Table 3.1: Morphological classification criteria based on the `SPREAD_MODEL` *parameter and its uncertainty. Table adapted from Drlica-Wagner et al. (2018).*

Class ID	Selection Criterion	Class Name
0	Likely Star	$\texttt{SPREAD\_MODEL\_G} + (5/3)\,\texttt{SPREADERR\_MODEL\_G} < -0.002$
1	High-confidence Galaxy	$(\texttt{SPREAD\_MODEL\_G} + (5/3)\,\texttt{SPREADERR\_MODEL\_G} > 0.005)$ AND NOT $(\lvert\texttt{WAVG\_SPREAD\_MODEL\_G}\rvert < 0.002$ AND $\texttt{MAG\_G} < 21.5)$
2	High-confidence Star	$\texttt{SPREAD\_MODEL\_G} + (5/3)\,\texttt{SPREADERR\_MODEL\_G} < 0.002$
3	Ambiguous	$0.002 < \texttt{SPREAD\_MODEL\_G} + (5/3)\,\texttt{SPREADERR\_MODEL\_G} < 0.005$

catalog and encodes potential issues identified during source detection and measurement in each band. This cut ensures that only objects with no severe detection or measurement problems are included in the sample.

Furthermore, the catalog does not provide magnitudes corrected for Galactic extinction. To account for this effect, we subtracted the values in the columns `EXTINCTION` from the corresponding magnitudes present in the column `MAG_AUTO`. The `EXTINCTION` columns were computed using the reddening correction for each band, defined as

$$A_b = R_b \times E(B-V), \tag{3.1}$$

where A_b is the extinction in band b. The fiducial interstellar extinction coefficients adopted were those from DES DR1, namely $R_g = 3.185$, $R_r = 2.140$, $R_i = 1.571$, and $R_z = 1.196$ (Abbott et al., 2018). The $E(B-V)$ values were obtained by interpolating the extinction maps of Schlegel et al. (1998) at the position of each source.

Additionally, we selected objects with $m_g < 23.5$ to ensure greater depth homogeneity across the footprint, where m_g denotes the magnitude measured in the g band. We note that $m_g = 23.5$ is below the nominal magnitude limits of the original DELVE exposures. In this sense, the cut $m_g < 23.5$ can be considered conservative. Although it includes objects close to the survey detection limits, it allows us to maintain a reasonably homogeneous sample while avoiding a significant loss of sources. Objects near this limit are expected to have their redshifts estimated in a regime where the network partially extrapolates the training distribution.

Finally, we excluded objects with *Signal to Noise Ratio* (SNR) < 2. We refer to the remaining catalog after applying all these cuts as **MCAT** (main catalog), which contains approximately $\sim$ 350M sources. MCAT serves as the basis for the subsequent cross-matches with the spectroscopic sample, as well as for constructing the DL training datasets. The final product of this work consists of estimating photometric redshifts for all objects in MCAT (whenever a sufficient number of bands is available).

3.3.2 | SZCAT

We crossmatched MCAT with SHSC using a search radius of 1 arc second. This crossmatch resulted in $\sim$ 2.9 M sources with reliable spec-z measurements available. These 2.9 million objects constitute he spectroscopic catalog called **SZCAT** (spectroscopic redshift catalog). Since we use the MCAT to crossmatch with the spectroscopic data, SZCAT is also constrained by the initial cuts imposed on MCAT. This catalog is used to build the truth table for training, validation, and test stages.

3.3.3 | DL Samples

Although we adopted a conservative cut of $m_g < 23.5$ for MCAT, when training the NN model we aimed to ensure the homogeneity of the dataset across the different data sources. With this in mind, we applied an additional cut to SZCAT requiring $m_{\{g,r,i,z\}} < 22.5$.[4] Combined with the SNR requirements, this selection ensures a high fraction of reliable measurements and produces a dataset largely free from spurious sources, low-quality detections, and extremely faint galaxies.

Additionally, we applied color cuts that follows Drlica-Wagner et al. (2018) to exclude nonphysical objects from SZCAT. We also restricted our z_{spec} interval to avoid to avoid stars and galaxies at very low redshift. A summary of the photometric quality cuts applied to SZCAT is

$$-1 < c_{gr} < 4,$$

$$-1 < c_{ri} < 4,$$

$$-1 < c_{iz} < 4,$$

$$m_{\{g,r,i,z\}} < 22.5,$$

$$0.01 < z_{spec},$$

where $c_{ab} = m_a - m_b$ is the color between the bands a and b. After applying all the data selection criteria, our initial SZCAT sample was reduced to 52% of its original size (see Table 3.2). All training and evaluation of our DL model will be performed using data from this refined dataset, which constitutes the **DLCAT** (Deep Learning Catalog). Figure 3.3 shows the z_{spec} and magnitudes distribution for DLCAT. It is worth noticing that after these cuts, $\sim$ 98% of the objects are limited as $z_{spec} < 1$.

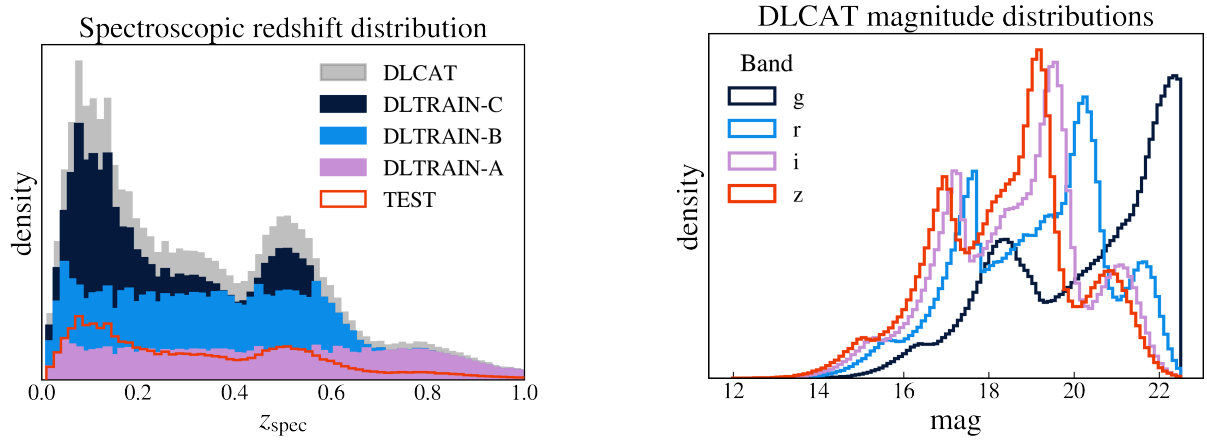

Figure 3.3: left pannel: the spectroscopic redshift distributions for the different training datasets. right pannel: the magnitude distributions in the griz bands for the objects included in DLCAT.

Table 3.2: Description of the different training datasets used to quantify the generalization capacity of our model.

Dataset	n^o objects	Description
MCAT	457383847	All DR2 Objects with mag-g< 23.5, `MODEST_CLASS` $\in \{1, 3\}$ and `FLAGS_G`$\leqslant 2$
SZCAT	2866993	All MCAT objects with spectroscopic correspondence through the described spectroscopic catalogs (after removing duplicates)
DLCAT	1503422	All SZCAT objects remained from the quality cuts in Section 3.3.3
DLTRAIN-A	384305	Uniform distribution of in bins of 0.05 width from $z = 0$ to $z = 0.8$ containing $\approx$ 20000 objects per bin; all non-Test post-selections data available for $z > 0.8$
DLTRAIN-B	857823	Uniform distribution of in bins of 0.05 width from $z = 0$ to $z = 0.6$ containing $\approx$ 60000 objects per bin; all non-Test post-selections data available for $z > 0.6$
DLTRAIN-C	1202597	All non-Test post-selections data available
Test set	300825	Randomly selected 20% of DLCAT before any other selection

The photo-z estimation approach being used here consisted supervised learning (see Section 2.3), where the model is trained by comparing its predictions with known results. Here, the input data corresponding to each object is defined as

$$\mathbf{x}_i = (m_g,\, m_r,\, m_i,\, m_z,\, c_{gr},\, c_{gi},\, c_{gz},\, c_{ri},\, c_{rz},\, c_{iz})\,. \tag{3.2}$$

And the corresponding output of the model is $p(z|\mathbf{x_i})$, such that in the ideal scenario

$$\arg\max_z p(z|\mathbf{x}_i) \approx z_{spec,i}. \tag{3.3}$$

While magnitudes carry information about the overall brightness of galaxies, colors are fundamental for representing the shape of the SED, which is crucial for determining redshifts (Arnouts and Ilbert, 2011). Since we are using DL, one could argue that, because colors are simply differences between magnitudes, the network should be able to learn such relations easily within its hidden layers and extract the underlying information automatically. However, using colors as explicit input features is historically the standard practice for retrieving photometric redshifts with less sophisticated methods, and it remains widely adopted in many modern DL-based approaches (Tagliaferri et al., 2003; Benítez, 2000; Lima et al., 2022; Dey et al., 2022). This practice is primarily motivated by the fact that magnitudes are proportional to the logarithm of the measured fluxes, while colors correspond to ratios between fluxes measured in different bands. As a result, colors are largely independent of survey-specific calibration choices, such as the absolute zero-point flux. A feature importance analysis presented in Lima (2025) shows that colors constitute the majority of the most informative features for estimating photometric redshifts in the case of S-PLUS.

As discussed in Section 2.2, three datasets are required to define the training and evaluation pipeline of a standard supervised DL application: the training, validation, and test sets. The test sample used in this work corresponds to 20% of the DLCAT dataset. The following paragraphs describe the construction of the training and validation datasets. We emphasize that the selection of the test sample was performed prior to any of the procedures described below.

DL models can be sensitive to the distribution of the training data. In general, models perform better on unseen data whose features resemble those encountered during training (Jeon et al., 2022; Kim et al., 2019). Therefore, the distribution of both the input features and the target variables in the training set plays a crucial role in the model's generalization ability. When dealing with supervised learning tasks, it is important to balance the training data as much as possible in order to minimize systematic biases. Figure 3.3 (left pannel) shows the $z_{\rm spec}$ distribution for the test sample, for DLCAT, and

[4]The riz bands are mentioned here only for consistency with the code. In practice, the cut, if applied only in g, would ensure that for $\sim 99\%$ of the objects $m_{riz} < 22.5$ as well.

for all training catalogs described below.

In the context of photometric redshift estimations, a major issue arises from the under-representation of the z_{spec} distribution compared to the all-sky galaxies redshift distribution. The models may exhibit biases toward more frequent redshifts. Many techniques are being developed to mitigate this issue, such as Data Augmentation and the attribution of weights to loss calculations (Moskowitz et al., 2024; Schuldt et al., 2021; Lima et al., 2008; Lima, 2025).

Preliminary tests with such techniques did not show significant improvements within the scope of this work. Nevertheless, as an alternative approach motivated by Zou et al. (2019), we address this issue by exploring three different $spec - z$ distributions for the training dataset. The underlying hypothesis is that a more homogeneous distribution in $spec - z$ should lead to less biased predictions.

The first training distribution (DLTRAIN-A) consists of a homogeneous distribution from spec-z= 0.01 up to spec-z≈ 0.8 in bins of width 0.05, containing approximately 2×10^4 objects per bin.

The second distribution (DLTRAIN-B) follows a similar construction, but remains homogeneous only up to spec-z≈ 0.6, using the same bin width and containing approximately 6×10^4 objects per bin (three times the size of DLTRAIN-A). The main motivation for this configuration is to quantify the impact of increasing the number of objects in the training set while allowing the distribution to become non-uniform at higher redshift.

The third distribution (DLTRAIN-C) includes all DLCAT objects not present in the test sample. This setup aims to empirically assess whether the shape of the training distribution or the total amount of data has a stronger impact on the model bias. Table 3.2 summarizes the main characteristics of these three training datasets.

We will train three identical models (in terms of architecture and hyperparameters), each using one of the DLTRAIN datasets. At this stage, we do not define a dedicated validation set because model evaluation will rely on a cross-validation strategy. The use of cross-validation will be described in detail in Chapter 4.

In Chapter 5, we compare the performance of the three models and show that, overall, DLTRAIN-A provides the most suitable training distribution.

3.3.4 | The incomplete coverage

The DELVE–DR2 galaxy dataset contains approximately 77% of its objects with complete photometric coverage, meaning that observations are available in all four bands (g, r, i, and z). Consequently, a small but significant fraction of the objects in MCAT does not provide the full band coverage required to compute photometric redshifts according

to the input definition in Equation 3.2.

In contrast, only $\sim 26\%$ of the objects in SZCAT lack full coverage: about 2% are missing only the z band, while 24% are missing only the i band. This number of objects is insufficient to construct a well-balanced training dataset for models specifically designed to estimate z_{phot} from incomplete photometric data (e.g., using only *gri* or *grz* bands).

One possible approach would be to combine objects with incomplete and complete photometric coverage in a single training set and allow the model to learn how to handle the missing information.

Instead, to enable redshift estimation for objects observed in at least three bands (i.e., those lacking either the i or z measurement), we trained two additional versions of the MDN model. These were obtained by training the same architecture on DLTRAIN-A while excluding the i or z magnitudes from the input features. Their performance is evaluated in Chapter 5.

4 Methodology

In this chapter we describe the specific DL methods employed to obtain the photometric redshifts for DELVE-DR2. We present the definition of our network architecture, as well as other photo-z methods used for comparison. We also describe the choices of hyperparameters and, finally, the tools and metrics used to evaluate the performance of the network.

4.1 | Neural Network Architecture

In Zuntz et al. (2021), we participated in the 3×2 Tomography Optimization Challenge (TomoChallenge, for short), which aimed to develop methods for galaxy tomography in preparation for the *Legacy Survey of Space and Time* (LSST). In this context, galaxy tomography consists of classifying galaxies into redshift bins rather than estimating continuous redshift values, effectively allowing the study of the Universe in redshift slices (Troxel and Ishak, 2015).

Our group developed several ML and DL approaches for this challenge, from which seven models were submitted. Although we were not among the winning entries, our proposed models performed competitively with many other submissions. This effort resulted in a diverse set of neural network architectures designed for galaxy tomography, a task closely related to photometric redshift estimation. The explored models included LSTMs, MLPs , traditional RNNs, and LMUs, among others.

When we joined the DELVE collaboration (coincidentally during the period of the challenge), our initial experiments on redshift estimation consisted of adapting the networks developed for the TomoChallenge to the photo-z problem. In this first stage, our primary goal was to empirically evaluate which model performed best for point estimates of the redshift, as opposed to probability density function (PDF) predictions. Among the tested architectures, the LMU achieved the lowest mean squared error. We therefore decided to proceed with further developments using the LMU as the core network architecture.

We emphasize the empirical nature of this decision, as there was no strong indication in the literature suggesting that RNN-based architectures are particularly advantageous for photometric redshift estimation. Nevertheless, we considered this an opportunity to explore whether the characteristics of this architecture could provide advantages for

the photo-z task. It is also important to note that, in these preliminary tests, we did not exhaustively optimize the hyperparameters for all candidate models. Therefore, it is possible that other architectures could achieve comparable performance if tuned more extensively.

In parallel, we demonstrated in Lima et al. (2022) that MDNs provide a promising approach for estimating redshift probability distribution functions. Motivated by these results, we combined the LMU architecture with an MDN output layer. The resulting architecture is illustrated in Figure 4.1.

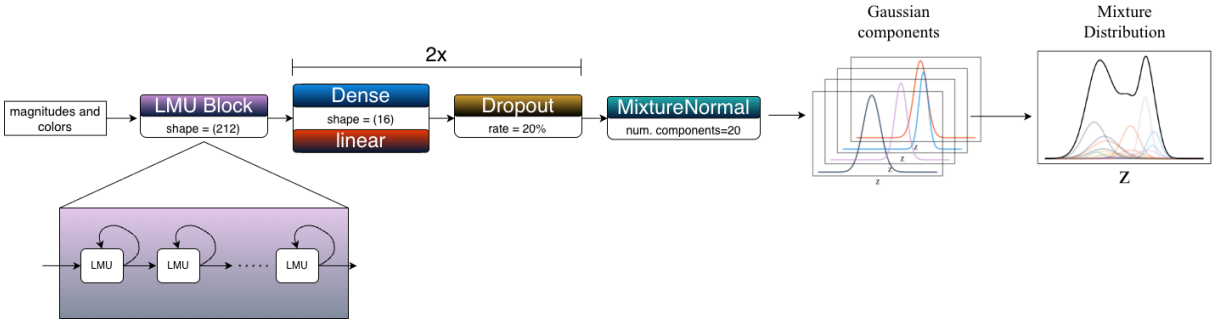

Figure 4.1: Architecture of the proposed network. The input vector of magnitudes and colors is processed sequentially by an LMU layer, whose output is fed into a simple MLP, and then forwarded to a MDN layer.

The input vector described in Section 3.3.3 is first processed by an LMU layer containing 212 units. As discussed in Section 2.4.1, the network interprets the input vector as a sequence, in which each element corresponds to a scalar feature representing either a magnitude or a color, ordered according to Equation 3.2.

The outputs of the LMU layer are then passed to a multilayer perceptron composed of two fully connected layers with 16 neurons each. Both layers include a dropout rate of 20%. The final layer implements a mixture density network that outputs a Gaussian mixture with 20 components, representing the predicted redshift PDF. For each input $\mathbf{x}_i$, the network predicts the vectors $\boldsymbol{\alpha}_i$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\sigma}_i$, corresponding to the mixture weights, means, and standard deviations of the Gaussian components.

The hyperparameters were initially selected empirically, inspired by the configuration adopted in (Lima et al., 2022) for the MDN component and by the best-performing configurations obtained during the tomography challenge for the LMU. Additional hyperparameter searches were conducted using the `grid_search` algorithm and the `AutoKeras` framework; however, no significant improvements in the final performance were found. Table 4.1 summarizes the network architecture and training configuration.

Table 4.1: Architecture and training configuration of the LMU-MDN network used for photometric redshift estimation.

Component	Configuration
Input representation	Magnitudes and colors ordered as in Eq. 3.2
Sequence interpretation	Each element treated as a scalar time step by the LMU
LMU layer	212 units
Fully connected layer 1	16 neurons
Dropout (layer 1)	0.20
Fully connected layer 2	16 neurons
Dropout (layer 2)	0.20
Output layer	MixtureNormal
Number of Gaussian components	20
Loss function	Negative log-likelihood of the Gaussian mixture
Optimizer	`Adam`
Learning rate	10^{-3}
Batch size	*1024*
Number of epochs	*200*

4.1.1 | Training

To avoid numerical issues in the network arising from the different orders of magnitude of the input variables, we normalize the inputs using standard deviation normalization `StandardScaler` Virtanen et al. (2020). In this procedure, the distribution of each input feature x_i in the training set is transformed such that

$$\bar{x}_i = \frac{x_i - \mu_i}{\sigma_i}. \tag{4.1}$$

Here, μ_i and σ_i represent the mean and standard deviation of the feature x_i computed over the entire training sample. For the test data, we apply the same transformation defined in Equation 4.1, using the values of μ_i and σ_i computed from the training sample. Similarly, all objects in the MCAT sample are normalized using the same values of μ_i and σ_i.

As mentioned in Section 2.3.2.2, we adopted the `Adam` optimizer for the training procedure. The optimizer hyperparameters were kept at their default values, except for the learning rate, which was set to 2×10^{-4}. We employed k-fold cross-validation (see Section PPP) during training. All models were trained for 200 epochs with a batch size of 1024.

As discussed in Section 3.3.3 and Section 3.3.4, we trained an independent model for each `DLTRAIN` sample, as well as for different combinations of observed bands. All training runs used the same architecture and hyperparameters described above. The loss function adopted during training is defined in Equation 2.5, and each model was trained for 200 epochs.

Figure 4.2 shows the training and validation loss curves for all configurations. No signs of overfitting or other training instabilities are observed. In fact, the slopes of the loss curves suggest that the models could potentially benefit from additional training epochs. However, preliminary tests showed no significant performance improvement beyond 200 epochs, and therefore we adopted this value as the standard training length.

The decision to use a fixed number of epochs for all models was also motivated by the goal of ensuring a fair comparison, allowing each configuration to undergo (as much as possible) the same training procedure. In principle, a more precise way to extract the best performance from each model would be to allow the training to stop automatically once a given convergence criterion is met. This technique, commonly referred to as *early stopping*, halts the training process when the validation loss ceases to decrease. Nevertheless, we stress again that our tests indicate that extending the training beyond 200 epochs does not lead to significant improvements for any of the models considered here.

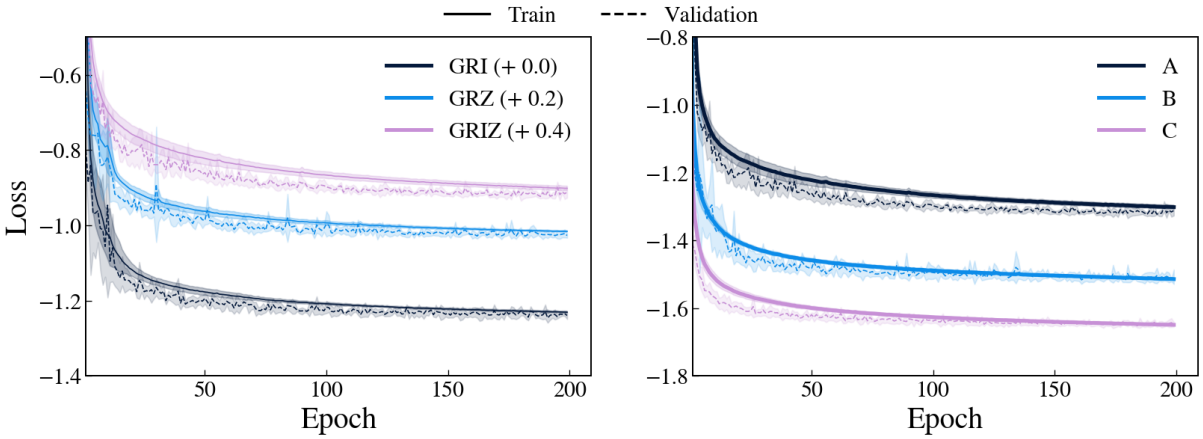

Figure 4.2: Training (solid lines) and validation (dashed lines) loss curves for the different model configurations. The shaded regions indicate the standard deviation across the five cross-validation folds. The left panel compares models trained with different photometric band combinations (`GRI`, `GRZ`, and `GRIZ`), while the right panel shows the results for the different `DLTRAIN` subsets (`A`, `B`, and `C`).

The code necessary to reproduce the models, as well as the main analysis presented in this part of the thesis, is publicly available at https://github.com/gsmteixeira/teixeira24-delve-dr2-phz.

4.1.2 | Computational Resources

The network was implemented using the `TensorFlow 2` and `TensorFlow Probability` libraries[1] (Abadi et al., 2015). The implementation of the LMU layer was obtained from the public repository https://github.com/abr/keras-lmu. All models were trained us-

[1] TensorFlow v2.9.1; TensorFlow Probability v0.17.0; `keras-lmu` v0.5.0.

ing a single NVIDIA RTX 3090 GPU. Under this configuration, the training time was approximately 4 seconds per epoch.

4.2 | Other Methods and Surveys

In this work we are proposing a non-standard DL technique for estimating photo-z. In order to evaluate whether the proposed complexity is worth developing, we compare our method with standard photo-z methods in the literature. This section describes the different techniques used to this purpose.

4.2.1 | Benchmarks

Henghes et al. (2021) presents a systematic analysis of several machine learning methods for estimating photometric redshifts, including the RF and the MLP (see Chapter 2). In this work, they conclude that the RF had the best performance in terms of quality metrics, even though it had the slowest training process.

We adapted the parameterization of the RF and MLP models from the benchmarking ones derived from their work (Table 4.2). The RF model was built in the Scikit-learn API[2] (Pedregosa et al., 2011) and the MLP in Tensorflow (same version used to the MDN). Both RF and MLP were trained using DLTRAIN-A as training dataset. No cross-validation was implemented for the MLP model, and we used 10% of the DLTRAIN-A dataset as a validation set.

[2]sklearn v1.1.2

Table 4.2: Hyperparameters adopted for the Dense Neural Network and Random Forest models.

Parameter	Dense Neural Network	Random Forest
Hidden layers	3	—
Neurons per hidden layer	100	—
Hidden activation	tanh	—
Output neurons	1	—
Output activation	ReLU	—
Loss function	MSE	—
Optimizer	`Adam`	—
Learning rate	2×10^{-4}	—
Batch size	1024	—
Epochs	200	—
Number of trees ($n_{\text{estimators}}$)	—	94
Max features per split	—	4
Minimum samples per leaf	—	8
Minimum samples per split	—	13
Minimum weight fraction per leaf	—	0
Split criterion	—	absolute error
Verbosity	—	1

4.2.2 | The DESI Legacy Imaging Survey

We also compare our photo-z estimates with those provided in the public catalog of the *DESI Imaging Legacy Sirvey* (LS). The LS were assembled from the combination of three imaging surveys: DECaLS, *Beijing–Arizona Sky Survey* (BASS), and *Mayall z-band Legacy Survey* (MzLS). These surveys were primarily designed to provide imaging data for the target selection of the DESI experiment, whose goal is to obtain spectroscopic redshifts for approximately 35 million galaxies and quasars.

In addition to optical photometry, the LS catalogs incorporate mid-infrared observations from the *WISE* and *NEOWISE* missions in the $W1$, $W2$, $W3$, and $W4$ bands, which are particularly important for the identification of quasars and luminous red galaxies.

The *Legacy Surveys Data Release 10* (LS-DR10) is the most recent data release of the LS. It significantly expands the survey footprint by incorporating additional DECam observations from NOIRLab and has substantial overlap with the DELVE survey. The LS-DR10 catalog also provides official photometric redshift estimates (Zhou et al., 2021, 2023). These were derived using a traditional RF approach. In addition to point estimates, the method provides photo-z uncertainties for each source. However, these uncertainty estimates assume Gaussian redshift PDFs and are primarily dominated by the propagation of photometric measurement errors.

4.3 | Evaluation Metrics

4.3.1 | Point-Like Metrics

We refer to as point-like metrics those derived from the direct comparison between a single photometric redshift estimate and the corresponding spectroscopic redshift. In accordance with the probabilistic nature of the model, we define the point estimate of the photometric redshift as the mode of the predicted PDF, i.e.

$$z_{\text{phot}} = \arg\max_{z} p(z). \tag{4.2}$$

We adopt standard metrics from the literature (Brammer et al., 2008; Ilbert et al., 2006; Lima et al., 2022) to evaluate the point-like photo-z estimates:

Bias (Δz): the mean residual between the photometric and spectroscopic redshifts,

$$\Delta z = \langle \delta z \rangle = \langle z_{\text{phot}} - z_{\text{spec}} \rangle \,, \tag{4.3}$$

where $\delta z = z_{\text{phot}} - z_{\text{spec}}$ is the residual.

Normalized Median Absolute Deviation (σ_{NMAD}): a robust estimator of the scatter defined as

$$\sigma_{\text{NMAD}} = 1.48 \times \text{median}\left(\left| \frac{\delta z - \text{median}(\delta z)}{1 + z_{\text{spec}}} \right| \right). \tag{4.4}$$

We adopt the definition used in Brammer et al. (2008) and Lima et al. (2022), which is less sensitive to outliers than the alternatives presented in Ilbert et al. (2006) and Li et al. 2022, for example.

Outlier fraction (η): the fraction of objects satisfying

$$\left| \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right| > 0.15. \tag{4.5}$$

An alternative definition identifies outliers in terms of σ_{NMAD}. In this case, the right-hand side of the inequality in Equation 4.5 would be replaced by $n\sigma_{\text{NMAD}}$ for a given integer n. For instance, Brammer et al. (2008) adopt a threshold of $5 \times \sigma_{\text{NMAD}}$, which corresponds to ≈ 0.15 for our test sample based on the results obtained with the MDN method. We therefore keep the explicit threshold of 0.15.

4.3.2 | PDF Diagnostics

There is little value in using a model that predicts photo-z PDFs if these distributions do not carry a meaningful probabilistic interpretation. Since our goal is to obtain reliable uncertainty estimates for the photometric redshifts, it is essential to verify that the predicted PDFs are properly calibrated. To assess this, we employ three diagnostic metrics.

4.3.2.1 Probability Integral Transform (PIT)

The first diagnostic is the *Probability Integral Transform* (PIT). The PIT distribution has been widely used to evaluate the reliability of photo-z PDF estimations in several studies (e.g., Mucesh et al., 2021; Lima et al., 2022; Newman and Gruen, 2022; Dalmasso et al., 2020; D'Isanto and Polsterer, 2018). For each galaxy, the PIT is defined as the value of the cumulative distribution function (CDF) of the predicted PDF evaluated at the corresponding spectroscopic redshift (Lima et al., 2022; Polsterer et al., 2016; Mucesh et al., 2021), i.e.,

$$\mathrm{PIT} = \int_{-\infty}^{z_{\mathrm{spec}}} p(z)\, dz. \tag{4.6}$$

A schematic representation of this concept is shown in Figure 4.3. The interpretation of the PIT arises from its distribution across a sufficiently large sample of galaxies. As discussed by Mucesh et al. (2021), if the predicted PDFs are well calibrated, the PIT values should follow a uniform distribution in the interval $[0, 1]$. Intuitively, this occurs because, for a correct probabilistic model, the spectroscopic redshift should behave as a random draw from the predicted PDF. Consequently, the CDF evaluated at the true value must be uniformly distributed.

Deviations from uniformity in the PIT histogram reveal systematic behaviors in the predicted PDFs. A concave ("U-shaped") distribution indicates that the PDFs are underdispersed, meaning that they are too narrow. Conversely, a convex distribution suggests overdispersed PDFs that are too broad. In addition, tilted or sloped PIT distributions typically indicate biased redshift estimates (see Polsterer et al., 2016).

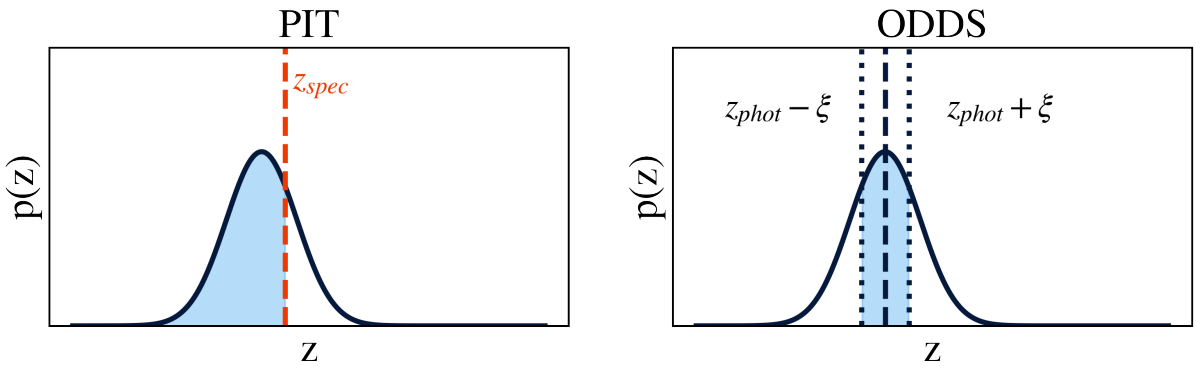

Figure 4.3: Illustration of two common diagnostics used to evaluate photometric redshift PDFs. Left: schematic representation of the PIT, defined as the cumulative probability of the PDF up to the spectroscopic redshift $z_{\rm spec}$. Right: illustration of the ODDS metric, corresponding to the integrated probability within a fixed interval around the most probable redshift $z_{\rm phot}$.

4.3.2.2 Odds

The second diagnostic is the *Odds* parameter, originally introduced in the context of Bayesian photometric redshifts (Benítez, 2000). The Odds measures the reliability of a given PDF by quantifying the probability mass contained within an interval around the most probable redshift. A schematic illustration of this metric is shown in Figure 4.3.

For a given PDF, the Odds is defined as

$$Odds = \int_{z_{\rm peak}-\xi_z}^{z_{\rm peak}+\xi_z} p(z)\, dz, \tag{4.7}$$

where $z_{\rm peak}$ corresponds to the most probable value of the PDF, which in our definition is equivalent to $z_{\rm phot}$ (Lima et al., 2022; Benítez, 2000).

High values of *Odds* indicate that a large fraction of the probability mass is concentrated near $z_{\rm peak}$, corresponding to narrow and well-constrained PDFs. Conversely, low values of *Odds* typically indicate broader distributions, reflecting larger uncertainties in the photometric redshift estimation.

The value adopted for the integration width ξ_z varies in the literature depending on the application. For example, Lima et al. (2022) use $\xi_z = 0.02$ when analyzing photometric redshift deviations reported in Molino et al. (2020) using the BPZ code, while Coe et al. (2006) adopt $\xi_z = 0.06$ for galaxies in the Hubble Ultra Deep Field (Beckwith et al., 2006). In this work, we adopt $\xi_z = 0.06$.

In general, a population of confident and reliable PDFs tends to produce an *Odds* distribution peaked near unity with a smooth shape.

4.3.2.3 Coverage Test

The coverage test (CV-test) is a metric commonly used to assess the reliability of conditional density estimations in terms of *Highest Density Region* (HDR) (Hermans et al., 2022; Zhao et al., 2021). The HDR is defined as the smallest region of the inferred variable space that contains a specified probability mass α, such that all points inside the region have probability density greater than or equal to any point outside it.

The CV-test evaluates how well calibrated the predicted PDFs are under the null hypothesis that a fraction $\hat{r} = \alpha$ $(0 \leqslant \alpha \leqslant 1)$ of the estimated PDFs should contain the spectroscopic redshift $z_{\rm spec}$ within the HDR enclosing a probability mass α.

If, for a given probability mass α, the fraction $\hat{r}$ is lower (grater) than α of the test sample PDFs containing $z_{\rm spec}$ within the corresponding HDR, the predicted PDFs are considered to be overconfident (underconfident). In practice, this means that when evaluating $\hat{r}$ for different values of α across the HDRs of the PDFs in the test sample, the empirical coverage should follow the identity relation $\hat{r} = \alpha$.

Figure 4.4 illustrates this concept with an example of three HDRs, represented by different colored regions. In this example, $z_{\rm spec}$ falls within the 50% HDR. The CV-test evaluates whether this occurs for approximately 50% of the objects in the sample.

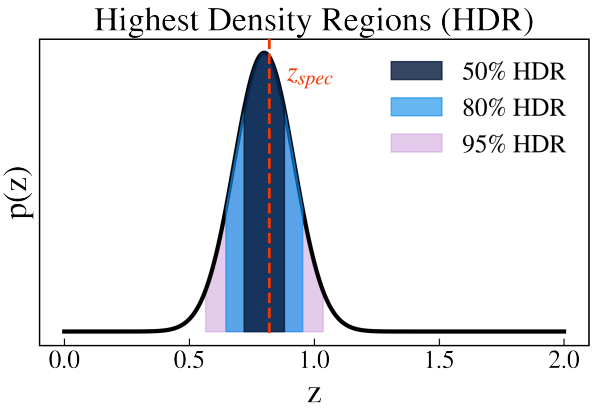

Figure 4.4: Illustration of highest density regions (HDRs) for different probability masses. The shaded areas represent HDRs of increasing probability mass, while the vertical line indicates the true value $z_{\rm spec}$. In this example, the true value lies within the 50% HDR.

This metric is closely related to the PIT distribution. However, the coverage test has an advantage when dealing with multimodal PDFs, since HDRs may consist of disjoint regions. Nevertheless, neither PIT, CV-test, nor Odds distributions alone can guarantee the reliability of the predicted PDFs. Individual diagnostic metrics may appear well behaved while the underlying PDFs remain miscalibrated. Therefore, these diagnostics test should be used and interpreted jointly.

5 Evaluation

In this chapter we present and discuss the photo-z produced by our MDN model. We will base our analysis on the metrics presented in Section 4.3, as well as present comparisons to the other methods and surveys presented in Section 4.2. All the discussed results in this chapter relate to the evaluation of the models on the test set.

5.1 | Point estimates

When comparing point estimates of photo-z, evaluating global metrics over the entire test sample is informative. However, analyzing these metrics as a function of z_{spec} provides deeper insight into systematic trends and biases.

Here, we compute metric curves as a function of z_{spec} using all five folds of the MDN photo-z predictions. The solid lines represent the mean across folds, while the shaded regions indicate the corresponding standard deviation.

We begin by comparing the performance of MDN models trained on the DLTRAIN-A, DLTRAIN-B, and DLTRAIN-C datasets using all available magnitudes from DELVE-DR2. Figure 5.1 shows that the models trained on DLTRAIN-B and DLTRAIN-C exhibit a pronounced bias around $z_{\text{spec}} \approx 0.65$. Overall, the MDN trained on DLTRAIN-A achieves better performance in terms of bias, except in the low-redshift regime ($z_{\text{spec}} < 0.15$), where the DLTRAIN-B and DLTRAIN-C models more closely follow the $\overline{\Delta z} = 0$ relation.

Regarding the uncertainties, there is no significant difference in the scatter (σ_{NMAD}) among the training configurations, except in the region around $z_{\text{spec}} \sim 0.6$. We also observe a slightly higher outlier fraction for the DLTRAIN-A model relative to the others up to $z_{\text{spec}} \approx 0.46$. At higher redshifts, however, the DLTRAIN-A model provides the best overall performance.

The improved performance at low redshift for DLTRAIN-B and DLTRAIN-C can be attributed to the higher density of objects in this regime within these training sets. In contrast, at higher redshifts – where these datasets are less well represented – the MDN trained on DLTRAIN-A outperforms the others, indicating better generalization in sparsely sampled regions of parameter space. For this reason, we hereafter use DLTRAIN-A as the reference training dataset for the MDN. Notably, the publicly available photo-z estimates for DELVE-DR2 were also trained using DLTRAIN-A.

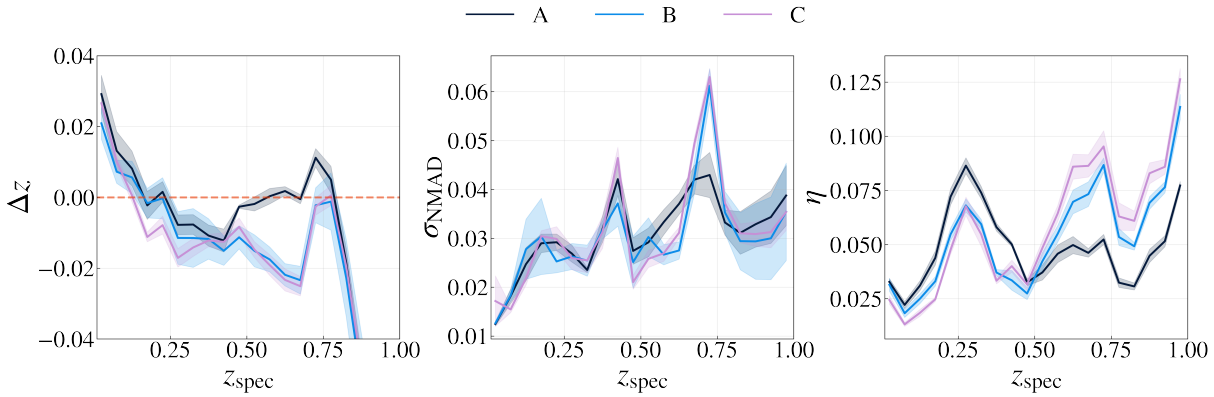

Figure 5.1: Performance metrics as a function of spectroscopic redshift for MDN models trained on the DLTRAIN-A, DLTRAIN-B, and DLTRAIN-C datasets. From left to right, we show the bias (Δz), the normalized median absolute deviation (σ_{NMAD}), and the outlier fraction (η). The curves are computed in bins of z_{spec} using all five cross-validation folds. Solid lines represent the mean across folds, while shaded regions indicate the corresponding standard deviation.

We present the impact of missing photometric bands in Figure 5.2. The MDN model was trained on the DLTRAIN-A dataset using three band configurations: *griz*, *gri*, and *grz*, as described in Section 3.3.4.

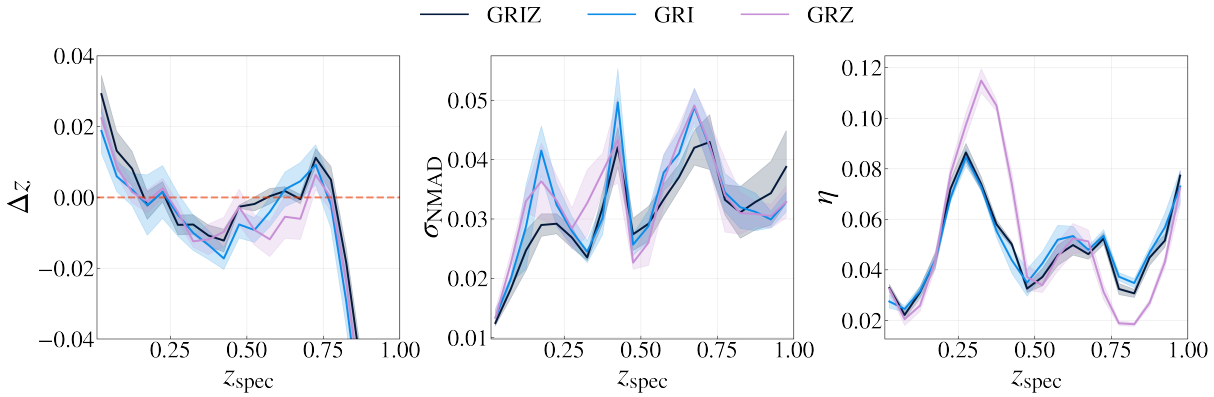

Figure 5.2: Performance metrics as a function of spectroscopic redshift for MDN models trained with different photometric band configurations (griz, gri, and grz) using the DLTRAIN-A dataset. From left to right, we show the bias (Δz), the normalized median absolute deviation (σ_{NMAD}), and the outlier fraction (η). The curves are computed in bins of z_{spec} using all five cross-validation folds. Solid lines represent the mean across folds, while shaded regions indicate the corresponding standard deviation.

All configurations exhibit similar bias overall, except in the low-redshift regime ($z_{\mathrm{spec}} < 0.15$), where models trained with missing bands show a slightly lower bias. The model trained with the full *griz* set generally achieves the lowest scatter (σ_{NMAD}). We note that the curves represent the mean and standard deviation across all folds; therefore, individual folds may achieve slightly better performance.

The *grz* configuration exhibits a localized increase in the outlier fraction around $z_{\mathrm{spec}} \approx 0.3$, but shows fewer outliers at higher redshifts ($z_{\mathrm{spec}} > 0.65$). No significant differences are observed between the *griz* and *gri* configurations in terms of outlier fraction, but in the case of *grz* this fraction is very pronounced.

These trends are also reflected in Figure 5.3, which shows the distribution of $z_{\rm phot}$ versus $z_{\rm spec}$. In the low-redshift regime ($z_{\rm spec} < 0.15$), the *grz* and *gri* models appear to perform slightly better. However, the *grz* configuration shows a noticeable bias in the range $0.15 < z_{\rm spec} < 0.3$, and both reduced-band models exhibit increased bias and dispersion at higher redshifts. A particular feature appear around $z \approx 0.45$ in all the configurations. This seems to be a systematic issue of the MDN model. It is with noticing that this feature appear to be slightly stronger for the *grz* model. This may be related to the fact that at this redshift range one of the most important spectral features, the 4000Åbreak–that comes mostly from the absorption lines of ionized metals–is crossing the band r to i, making the color $r - i$ crucial for the photo-z (Dahlen et al., 2013).

The publicly available photo-z catalog produced by our method includes the columns `MODEL_{GRIZ, GRZ, GRI}`, where a value of 1 indicates which model was used to generate the prediction. Notably, the *gri* and *grz* models are only employed in cases where one of the photometric bands is missing.

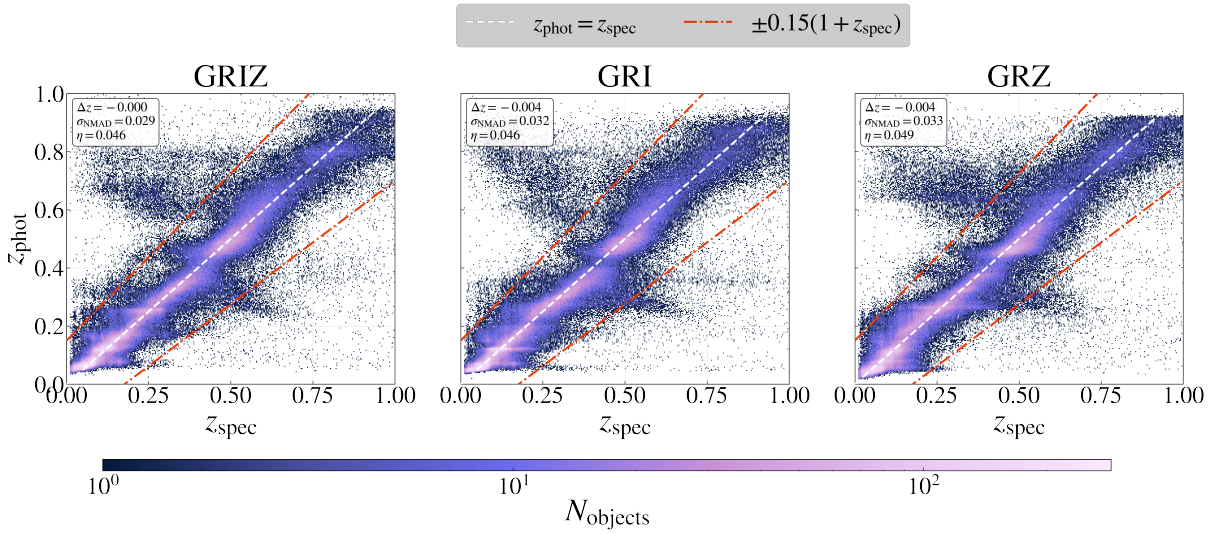

Figure 5.3: Distribution of photometric versus spectroscopic redshifts for MDN models trained with different band configurations using DLTRAIN-A. The color scale represents the density of objects. The metrics shown in each panel correspond to the full test sample.

Figure 5.4 presents our point estimate results for all methods described in Chapter 4. We also computed the point-like metrics using the photometric redshift values available for the same galaxies in the public LS-DR10 (Section 4.2.2). The resulting curves related to the LS-DR10 data are also shown in Figure 5.4.

It is worth mentioning that when crossmatching the test sample with the LS-DR10 catalogs, a small number of objects were missed due to the lack of photo-z measurements in LS-DR10. These objects were removed from the test sample prior to computing the metric curves. This reduction is not statistically significant and does not affect the results.

During our analysis, the trained methods did not exhibit any signs of overfitting or training-related issues. Therefore, we retained a single model for each comparison analysis.

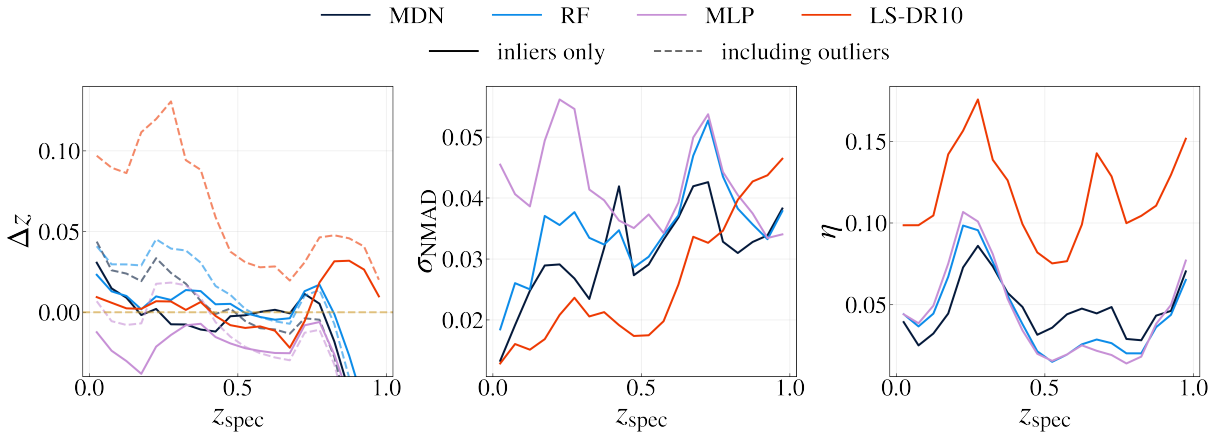

Figure 5.4: Performance metrics as a function of spectroscopic redshift for four different methods evaluated on the test set. The dark blue solid lines show results for DELVE-DR2 using our MDN model. Light blue and violet lines show results for the same data using the RF and MLP methods described in Section 4.2, respectively. Orange lines correspond to the publicly available z_{phot} from the LS-DR10 catalog. In the left panel, dashed lines indicate the bias computed over the full sample within each spectroscopic redshift bin, while solid lines show the bias after excluding objects classified as outliers.

We also removed the error bars from the MDN curve to improve the visualization of the left panel in Figure 5.4.

The left panel of Figure 5.4 presents two bias curves for each method. The solid lines represent the bias as a function of z_{spec} computed after excluding outliers. From these curves, it is clear that the MLP model performs the worst overall. Up to $z_{\mathrm{spec}} \approx 0.46$, the public photo-$z$ from LS-DR10 outperforms the other methods, except for a narrow region around $z_{\mathrm{spec}} \approx 0.2$, where the MDN model more closely approaches zero bias. Beyond $z_{\mathrm{spec}} \approx 0.46$, the MDN method outperforms the others, although the RF model yields comparable biases.

When considering the bias computed over the full sample (i.e., including outliers, excluding only objects without LS-DR10 photo-z), the behavior changes significantly. In this case, the LS-DR10 estimates show the worst performance. The MLP presents a reduced bias, while the MDN model outperforms all other methods across most of the redshift range, with the exception of a narrow region around $z_{\mathrm{spec}} \approx 0.6$, where the RF method is closer to zero bias.

This comparison highlights that, under optimal conditions (i.e., excluding outliers), LS-DR10 and the MDN model achieve similar performance. However, in realistic scenarios – where spectroscopic redshifts are unavailable and no outlier rejection can be applied – the relevant comparison is given by the full-sample (dashed) curves. In this regime, the MDN model exhibits the lowest bias. The apparent improvement of the MLP in this case is misleading, as its curve is artificially driven toward zero by the presence of outliers. A more complete assessment therefore requires considering additional metrics.

Despite exhibiting a higher outlier fraction, the LS-DR10 photo-z estimates yield the lowest σ_{NMAD} as a function of z_{spec}. As discussed in Section 4.3, this metric is robust to

outliers, which explains this behavior. The MDN method outperforms the other models in terms of σ_{NMAD} across most of the redshift range, except for a localized increase around $z_{\mathrm{spec}} \approx 0.46$, which appears to be associated with the systematic feature observed in Figure 5.3.

Regarding the outlier fraction, the LS-DR10 photo-z estimates exhibit the highest fraction among all methods. The MDN model outperforms the RF and MLP methods up to $z_{\mathrm{spec}} \approx 0.35$, but shows an increased outlier rate at higher redshifts. The strong impact of outliers in the LS-DR10 results is evident when comparing the shapes of the corresponding bias and outlier fraction curves.

From this comparison, we observe that despite the high fraction of outliers, the LS-DR10 redshift estimates yield the least dispersed results. Our methods, however, achieve competitive performance while additionally providing full redshift PDFs, which are significantly more informative than point estimates and their associated uncertainties.

In particular, when inspecting the PDFs of objects classified as outliers, we often observe bimodal distributions. In such cases, the primary mode typically corresponds to the outlier photo-z estimate, while a secondary mode lies closer to the spectroscopic redshift. This additional information is crucial for analyses that marginalize over redshift full PDF, rather than relying on single-point estimates.

Such approaches are especially relevant in applications like dark sirens (Alfradique et al., 2024), where the full redshift PDF enters directly into the inference. We explore this application in more detail in Chapter 6.

5.2 | PDF evaluation

As we did in the case of the point-like metrics, we will not only evaluate the global aspect of the PDFs calibration, but also how does it changes with zspec. We start our analysis by examining the PIT and Odds distributions showed in Figure 5.5.

The PIT distribution for the full sample is shown in gray. It exhibits a shallow negative slope, indicating a mild tendency of the MDN model to overestimate the redshift at which the probability density is concentrated. In other words, the predicted PDFs appear to be, on average, slightly shifted toward higher redshifts relative to z_{spec}.

By inspecting the PIT distributions in bins of z_{spec}, we find that this behavior is primarily driven by low-redshift objects. This trend is consistent with the behavior observed in the point-estimate analysis, where the model tends to overestimate redshifts in the regime $z \lesssim 0.2$. A similar effect is also observed in the photo-z PDF estimates from the independent S-PLUS survey, which employs an approach conceptually similar to the one proposed here (Lima, 2025).

At intermediate redshifts, this effect becomes less pronounced and the PIT distribution approaches uniformity. In contrast, at higher redshifts ($z \gtrsim 0.8$), the PIT distribution suggests the opposite trend, with a bias toward underestimating z. This behavior is expected given the relative scarcity of training objects in this regime. This interpretation is consistent with the sharp change in the bias observed in Figure 5.4 **change the figure expanding xlim**.

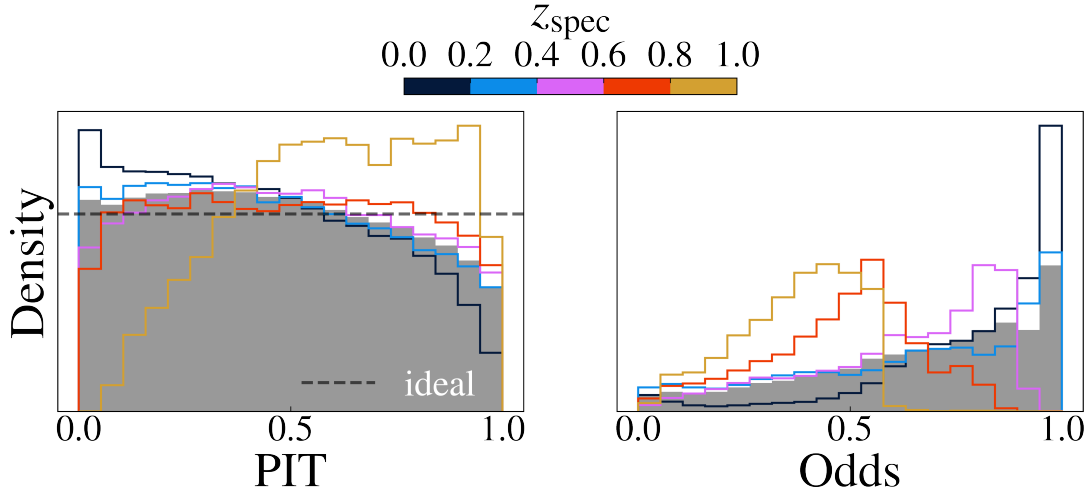

Figure 5.5: PIT (left) and Odds (right) distributions for the MDN model in bins of $z_{\rm spec}$. A uniform PIT distribution indicates perfect calibration, while Odds values close to unity reflect highly informative PDFs.

The Odds distributions show a smooth transition with redshift over the range $0.2 \lesssim z \lesssim 1.0$, with the expected trend of increasing uncertainty toward higher redshifts. For objects at low redshift ($z < 0.2$), however, the Odds distribution exhibits a pronounced peak near unity, indicating highly concentrated PDFs. While this could suggest high precision, it does not necessarily imply accurate calibration. Indeed, in this regime, the PIT distribution deviates from uniformity, reflecting a systematic bias.

Interestingly, as shown in Figure 5.6, the main contribution to the deviation from uniformity in the global PIT distribution arises from objects with low Odds values. This indicates that the observed discrepancy is primarily driven by locally underconfident PDFs (i.e., broader or more complex distributions), rather than by the high-Odds population. In other words, the apparent issue in the PIT distribution is dominated by PDFs with large dispersion or multimodal structure localized at low redshifts.

It is also important to note that, in the low-redshift regime, the Odds integration interval parameter $\xi_z = 0.06$ is relatively large compared to the typical redshift values. As a result, a significant fraction of the total probability can be enclosed within this interval even for moderately broad PDFs. Therefore, high Odds values in this regime do not necessarily imply sharply peaked or overconfident distributions, but rather reflect the relative scale of the integration window with respect to the redshift.

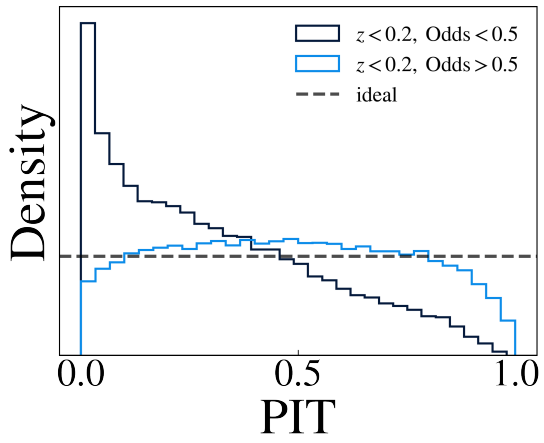

Figure 5.6: Contribution of low-Odds PDFs to the deviation from uniformity in the PIT distribution.

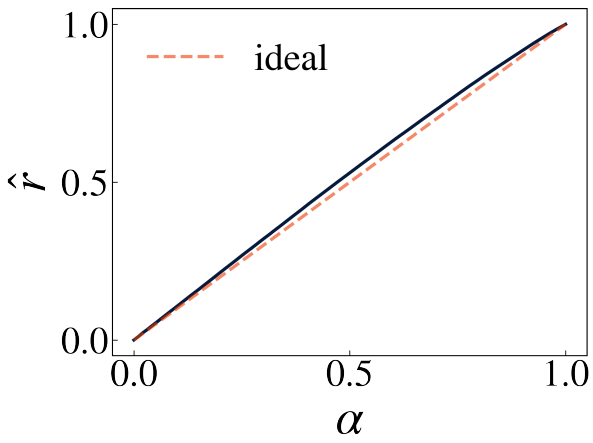

Figure 5.7: Coverage test comparing nominal credibility levels with empirical coverage for the MDN model, its AE-decoded PDFs, and LS-DR10-derived PDFs. The dashed line indicates perfect calibration.

The coverage test in Figure 5.7, which probes global calibration, shows that the empirical coverage $\hat{r}$ closely follows the diagonal relation with the nominal credibility levels. This indicates that, on average, the predicted PDFs are well calibrated (Hermans et al., 2022), with only a slight tendency toward underconfidence. This behavior is consistent with the interpretation of the PIT and Odds diagnostics.

The consistency between the PIT and coverage diagnostics indicates that the MDN PDFs are well calibrated globally, while still exhibiting localized deviations associated with specific redshift regimes. The residual skewness in the PIT distribution can be attributed primarily to low-redshift objects and to PDFs with low Odds values, suggesting that both systematic biases and increased uncertainty contribute to the observed behavior.

The MDN model provides accurate photometric redshifts with well-calibrated and informative PDFs, although it exhibits systematic effects at both low and high redshift extremes. At high redshift, these trends are likely driven by the limited representation of objects in the training set, which hampers the model's ability to generalize in this regime. At low redshift, the observed systematics may be associated with subtle biases in the learned mapping between photometric features and redshift. Understanding the origin of these effects is the subject of ongoing investigation–in Section 6 we show that this problems may be related to the complexity of the model and the way it is interpreting the colors.

6 Applications and Extensions

The photo-z estimates developed in this thesis are now validated and ready for scientific use, exhibiting robust performance across most of the test sample, with the exception of a localized miscalibration affecting a small subsample, as discussed in Chapter 5. As briefly outlined in Chapter 1, photo-z estimates enable a wide range of astrophysical and cosmological applications. The photo-z presented here (and their subsequent extensions) have been employed in two distinct contexts: *dark sirens* and *galaxy cluster science*.

Dark sirens constitute one of the primary research directions of our research team in Laboratory of Artificial Intelligence for Physics (LAB-IA), providing an independent probe of cosmology through the combination of gravitational wave (GW) observations and galaxy catalogs. Galaxy cluster applications are connected to my participation in the *Chilean Cluster Galaxy Evolution Survey* (CHANCES) collaboration, where accurate photometric redshifts play a key role in cluster membership assignment. In this chapter, we explore their impact in these contexts.

6.1 | Dark Sirens

The year 2017 was marked by the first GW detection with an electromagnetic counterpart. The event GW170817, a binary neutron star merger detected by the LIGO and Virgo interferometers, and its associated kilonova provided for the the first direct measurement of luminosity distance, without the need for any distance ladder, enabling an independent way of measuring the expansion rate of the universe (Abbott et al., 2017a).

This is possible because the GW signal encodes the luminosity distance d_L directly in its waveform (Abbott et al., 2020; Blanchet, 2002). In a flat ΛCDM model, the relationship between the luminosity distance and the redshift is given by

$$d_L(z) = \frac{(1+z)c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}, \tag{6.1}$$

where Ω_m and Ω_Λ are the matter and dark energy density parameters, respectively. GW sources from compact binary mergers are self-calibrating distance indicators, often referred to as *standard sirens* (Schutz, 1986; Holz and Hughes, 2005) – the gravitational-wave analogue of standard candles (Branch and Miller, 1993) – in which distances are measured

from GW signals rather than electromagnetic observations. Standard sirens therefore offer a pathway to constrain H_0 and other cosmological parameters independently of the cosmic distance ladder (Chen et al., 2018).

With the electromagnetic counterpart of GW170817, it was possible to obtain the redshift of the host galaxy, NGC 4993, spectroscopically (Hjorth et al., 2017), yielding a measurement of $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1} Mpc^{-1} (Abbott et al., 2017b). A GW event associated with an observed electromagnetic counterpart and an identified host galaxy is referred to as a *bright siren*.

In contrast to the bright siren case, which has so far been confirmed only once, the vast majority of GW events reported by the LIGO-Virgo-KAGRA (LVK) collaboration originate from binary black hole (BBH) mergers – systems that are not expected to produce an electromagnetic counterpart in most scenarios (Abbott et al., 2023), although some models in the literature predict possible optical counterparts under specific conditions (Rodríguez-Ramírez et al., 2023). The standard sirens with no eletromagnect counterpart are called *dark sirens*.

The *statistical dark standard sirens* method (Gair et al., 2023) addresses the absence of electromagnetic counterparts by statistically marginalising over all galaxies within the GW localization volume as potential hosts. Rather than identifying a unique host, each galaxy in the localization region contributes its redshift to a combined posterior on H_0, weighted by its probability of hosting the merger (e.g., by stellar mass or star-formation rate). While less precise than the bright siren approach on an event-by-event basis, the dark siren constraint tightens as the number of detected events grows (Palmese et al., 2023), making it a promising statistical probe of cosmic expansion in the era of large GW catalogues.

In Alfradique et al. (2024), we used the photometric redshifts developed for the DELVE second data release (DELVE DR2) and their associated PDFs to constrain the value of H_0 considering 10 different GW events from the first and third LVK observing runs. This study employed the galaxy catalogue dark siren method, in which the marginalisation over H_0 accounts for the actual distribution of galaxies in an observed catalogue, rather than assuming a modelled redshift distribution.

In this framework, we model the posterior distribution of H_0 adopting a flat ΛCDM cosmology with $\Omega_m = 0.3$ (and therefore $\Omega_\Lambda = 0.7$) as our fiducial model, following the Bayesian formalism described in detail in Chen et al. (2018) and adapted in Soares-Santos et al. (2019); Palmese et al. (2023) (see Section 2.4.3 for a brief overview of Bayesian statistics). The H_0 posterior given the GW data $\mathcal{D}_{\rm GW}$ and the electromagnetic galaxy catalogue data $\mathcal{D}_{\rm EM}$ is written via Bayes' theorem as

$$p(H_0 \mid \mathcal{D}_{\rm GW}, \mathcal{D}_{\rm EM}) \propto p(\mathcal{D}_{\rm GW}, \mathcal{D}_{\rm EM} \mid H_0)\, p(H_0), \tag{6.2}$$

where $p(H_0)$ is a uniform prior on H_0 over the interval $[20, 140]$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and $p(\mathcal{D}_{\rm GW}, \mathcal{D}_{\rm EM} \mid H_0)$ is the joint GW-electromagnetic (EM) likelihood. Here, $\mathcal{D}_{\rm GW}$ encodes the GW observables – in particular the luminosity distance posterior $p(d_L)$ and the sky localisation (Ω) probability map $p(\Omega)$ provided by the LVK interferometers – while $\mathcal{D}_{\rm EM}$ represents the eletromagnectic information from the galaxy catalogue, including the photometry, photo-z PDF, and its associated bias Δz for each potential host galaxy.

This framework relies on three main assumptions: that the GW source is hosted by one of the galaxies in the catalogue; and that the catalogue EM and GW information are independent. The full H_0 posterior, marginalising over all galaxies in the catalogue and over their photometric redshift PDFs, is written as

;

$$p\left(H_0 \mid \mathcal{D}_{\rm GW}, \mathcal{D}_{\rm EM}\right) \propto \frac{p(H_0)}{\beta(H_0)} \sum_{i=1}^{N_{\rm gal}} \frac{1}{\mathcal{Z}_i} \int dz \int d\Delta z \times A(H_0, z)\, p\left(\mathcal{D}_{\rm GW} \mid d_L\left(z, H_0\right), \hat{\Omega}_i\right)\, p\left(\Delta z\right)\, \hat{p}\left(z - \Delta z\right), \tag{6.3}$$

where $\beta(H_0)$ is the selection function accounting for observational biases in the GW-EM detection process; the sum runs over all $N_{\rm gal}$ galaxies in the catalogue within the localization volume of the GW event; and $\mathcal{Z}_i$ is a normalization constant for the i-th galaxy, ensuring that the redshift likelihood is properly normalized. $p\left(\mathcal{D}_{\rm GW} \mid d_L\left(z, H_0\right), \hat{\Omega}_i\right)$ is the likelihood of the GW data given a source located at luminosity distance $d_L(z, H_0)$ and sky position $\hat{\Omega}_i$.

The function $A(H_0, z)$ encodes the astrophysical weight assigned to each galaxy, which may depend on its luminosity or other properties (e.g., stellar mass or star formation rate), reflecting the expectation that BBH mergers preferentially occur in certain types of galaxies. In Alfradique et al. (2024), this is taken as $A(H_0, z) = r^2(z, H_0)/H(z, H_0)$, where $r(z, H_0)$ is the comoving distance to redshift z and $H(z, H_0)$ is the Hubble parameter at that redshift.

Therefore, weighting galaxies by $A(H_0, z)$, in this case, is equivalent to assuming that BBH mergers trace the galaxy distribution uniformly in comoving volume – that is, each galaxy is weighted proportionally to the available cosmic volume at its redshift. Our recent analyses of dark sirens using LVK detections incorporate more realistic assumptions for $A(H_0, z)$. For example, in Alfradique et al. (2026), the weighting accounts for the fact that BBH mergers are more likely to occur in more luminous galaxies.

Finally, I would like to call the attention for the term $p(\Delta z)\hat{p}(z_i - \Delta z)$. These terms comes directly from the work presented in this thesis. $p(\Delta z)$ is the empirical distribution of the residual bias $\Delta z = z_{phot} - z_{spec}$ obtained from the evaluation of the redshifts over our training sample, and $\hat{p}(z - \Delta z)$ is the PDF of the i-th galaxy estimated from our MDN at the bias-corrected redshift $z - \Delta z$.

Alfradique et al. (2024) argues that the use of the full probabilistic information improves the quality of the constrain on the H_0 posteriors, in contrast to the standard approach of assuming a Gaussian PDF centered on the z_{phot} values with a given deviation $\sigma_{\text{photo-z}}$ from the estimated photo-z errors. We inferred the posterior distribution for 2 events of the LVK third observing run (O3): GW190924_021846 and GW200202_154313. Figure 6.1 shows the H_0 posterior distribution for these two events, as well as the combined posterior between the dark sirens and the bright siren (addapted from Nicolaou et al. (2020), including pecular velocity corrections).

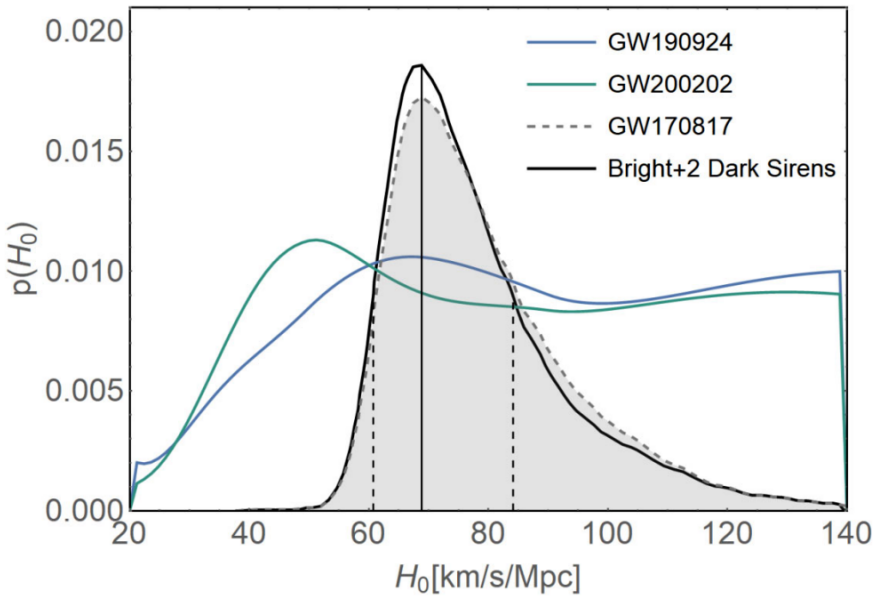

Figure 6.1: Posterior distributions of the Hubble constant for the dark siren events GW190924_021846 and GW200202_154313, obtained using full photometric redshift PDFs. The individual posteriors are shown in green and blue, the bright siren measurement adapted from Nicolaou et al. (2020) is shown as the shaded region, and the combined posterior distribution is shown in black. Image from Alfradique et al. (2024)

The final result of this analysis yields two constraints on H_0. The first is obtained from the combined posterior shown in Figure 6.1, corresponding to $H_0 = 68.84^{+15.51}_{-7.74}$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The second combines the posterior from these two events with those from 8 dark sirens in previous observing runs presented in Palmese et al. (2023), yielding $H_0 = 76.00^{+17.64}_{-13.45}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

At the time of publication, this represented one of the most precise measurements of H_0 obtained using dark sirens. This result highlights the critical role of well-calibrated redshift catalogs, particularly the accurate characterization of photometric redshift PDFs, underscoring their impact on cosmological inference.

6.1.1 | Advances in Dark Sirens and Extension to LS-DR10

During the fourth observing run (O4) of the LVK Collaboration, we performed a similar analysis including five GW events from O4a and three updated sky maps from O3. In this case, we did not employ the photometric redshifts described in this thesis; instead, we used an improved version derived from LS-DR10 photometry, motivated by its significantly larger overlap with the GW localization regions compared to DELVE-DR2.

We also adopted an LMU-MDN neural network with reduced complexity, reducing both the number of neurons and the number of Gaussian components to six. The results are presented in Bom et al. (2024), yielding a constraint on the Hubble constant of $H_0 = 68.0^{+4.4}_{-3.8}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from the combined analysis of all dark sirens and one bright siren.

Finally, our most recent contribution is presented in Alfradique et al. (2026), which introduces a refined analysis incorporating more realistic, data-driven assumptions on the galaxy distribution and observational selection effects, including magnitude weighting in the r band. This work also presents a new photometric redshift estimation for the northern hemisphere based on Legacy Survey Data Release 9 photometry. The combined analysis yields a constraint of $H_0 = 69.9^{+4.1}_{-4.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

6.2 | Galaxy Cluster Membership

The CHANCES (Haines et al., 2023; Sifón et al., 2025; Méndez-Hernández et al., 2026) is a *4-metre Multi-Object Spectroscopic Telescope* (4MOST) (de Jong et al., 2019) Community Survey designed to investigate the connection between galaxy evolution and hierarchical structure formation through deep and wide multi-object spectroscopy of galaxy clusters and their environments. Over its five-year duration, the survey will target approximately 5×10^5 cluster galaxies, probing regions out to $\sim 5\,R_{200}$ and enabling detailed studies of environmental effects such as galaxy pre-processing prior to cluster infall.

CHANCES is divided into three subsurveys:

- *The CHANCES low-z subsurvey* is designed to target 50 nearby galaxy clusters at $z \leqslant 0.07$, including wide regions within the Shapley and Horologium–Reticulum superclusters. It provides a detailed and statistically complete view of the local cosmic web, sampling clusters over a wide mass range and enabling studies of galaxy properties and environmental effects down to low stellar masses.

- *The CHANCES evolution subsurvey* is designed to target 50 of the most massive galaxy clusters uniformly distributed over $0.07 < z < 0.45$. This subsurvey aims to trace the evolution of cluster galaxies across cosmic time, leveraging a homogeneous sample primarily selected from Sunyaev–Zel'dovich surveys, and enabling the study of how galaxy populations and cluster properties evolve with redshift.

- *The CHANCES circumgalactic medium (CGM) subsurvey* is designed to investigate the circumgalactic medium in and around galaxy clusters by observing galaxies in the foreground of background quasars. It aims to characterize the origin of absorption features (e.g., Mg II) and their connection to galaxy properties and environment, providing a complementary view of how gas in dense environments influences galaxy evolution.

To each sub-survey is assigned a target list to be observed by the 4MOST instrument. Before the beginning of 4MOST observations, each subsurvey adopted its own strategy to construct this target list (Sifón et al., 2025). In particular, the low-z subsurvey strategy relied on a photo-z based approach.

The initial plan was to use publicly available photometric redshifts from the LS-DR10 and S-PLUS footprints to assign cluster membership. The S-PLUS survey provides one of the most accurate photo-z catalogues in the literature (Lima, 2025), largely due to its use of 12 photometric bands, and in fact served as the primary inspiration for the work developed in this thesis. However, despite its high precision, S-PLUS covers only $\sim 9000\,\mathrm{deg}^2$ of the sky. Therefore, LS-DR10 offers a valuable alternative for complementing photo-z information in regions outside the S-PLUS footprint.

For each cluster, they considered the solid angle corresponding to $5 \times R_{200}$, where R_{200} is the radius within which the mean density of the cluster is 200 times the critical density of the Universe, commonly used as an approximation to the virial radius of the system. All galaxies within this area, brighter than the magnitude limit $r < 20.4$ and with available photometric redshifts, were then tested against cluster membership criteria based on their consistency with the cluster redshift. More explicitly, the galaxy is considered a potential member of the cluster if

$$|z_{\mathrm{phot}} - z_c| < N \times \sigma_{NMAD}(m_r) \tag{6.4}$$

where z_{phot} is the photometric redshift of the galaxy and z_c is the cluster central redshift. The right-hand side of the criterion accounts for the scatter σ_{NMAD} (see Section 4.3) as a function of the r-band magnitude m_r.

This scatter is modeled through an empirical relation obtained by fitting a third-order polynomial to the dependence between σ_{NMAD} and m_r, using the training sample of a given photo-z estimator. This approach provides an estimate of the expected dispersion σ_{NMAD} for any galaxy, including those without spectroscopic redshifts, thereby allowing

the tolerance to be modulated according to galaxy brightness while accounting for systematic effects of the photo-z estimation. The constant N is a tunable tolerance factor, adjusted according to the source of the photo-z data in order to optimize completeness, while balancing the intrinsic performance differences between distinct photo-z estimators.

Figure 6.2 presents the color–magnitude diagram (CMD) for the target selection performed in the Abell 85 (A85) cluster (Bravo-Alfaro et al., 2009), comparing the different photo-z estimators considered – in the end, an extension of the photo-z developed in this thesis was adopted. This cluster was chosen as the primary benchmark because it lies within the overlapping footprint of both S-PLUS and LS-DR10, allowing for a direct and fair comparison between methods.

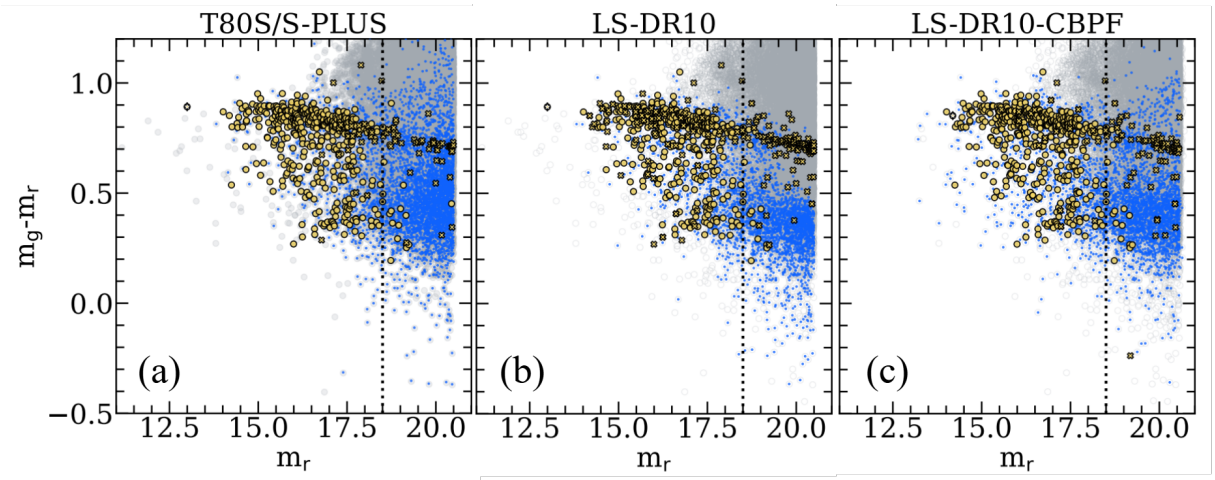

Figure 6.2: CMDs for the A85 cluster, illustrating the selection of cluster member candidates using different photometric redshift estimators. Blue points indicate galaxies selected as members according to Equation 6.4, while gray points correspond to all background galaxies within $5 \times R_{200}$*. Yellow points represent spectroscopically confirmed cluster members: circles denote those selected by the respective method, while crosses indicate confirmed members that were not selected. Figure adapted from Méndez-Hernández et al. (2026)*

Analysing these results in the CMD space is particularly informative because galaxy clusters exhibit well-known features, most notably the *red sequence*: a tight, nearly linear relation populated by passively evolving, early-type galaxies that have ceased star formation. This sequence is a hallmark of dense environments and provides a robust diagnostic of cluster membership and galaxy evolution, as it traces the buildup of quiescent galaxies over cosmic time (Graves et al., 2009).

A clear difference emerges when comparing the panels. In contrast to the S-PLUS results (Figure 6.2a), the LS-DR10 results (Figure 6.2b) show a noticeable deficit of objects at the faint end of the red sequence (FRS). This behavior is not expected, as the red sequence is known to extend to low luminosities, being populated by quiescent dwarf elliptical galaxies, which are abundant in cluster environments (Graves et al., 2009).

The observed deficit (hereafter called *FRS problem*) likely reflects limitations in the photo-z performance at faint magnitudes and low redshift in LS-DR10, possibly associated

with the underrepresentation of this specific population (i.e., faint, red galaxies at low z) in the training samples of the LS-DR10 photo-z estimators. This behavior does not appear in the S-PLUS results, most likely because as the survey accounts with additional narrow filters, degeneracies inside the the dataset may be mitigated.

The use of the redshifts developed as an extension of this thesis takes place as an attempt to address, or at least better understand, the FRS problem. At the time I joined the collaboration, photometric redshifts developed for Bom et al. (2024) were already available, derived from LS-DR10 photometry. When these redshifts – obtained using an LMU-MDN model with overall performance comparable to publicly available ones – were directly applied to the target selection algorithms, they reproduced the same behavior observed in Figure 6.2b for public photo-z estimates. In fact, the results were slightly worse, as the "hole" in the red sequence became more pronounced.

After several months of interdisciplinary collaboration and extensive testing of different approaches – including alternative models and architectures for photo-z estimation, data augmentation, and loss weighting strategies – I identified that the issue was primarily related to how the model interprets color information. In the low-z FRS regime, the models appeared to place greater emphasis on the magnitude-redshift relation than on the color-redshift relation.

To address this, I designed a model that processes colors and magnitudes separately, and subsequently concatenates their latent representations for joint learning. As shown in Figure 6.3, the proposed solution relies on a relatively simple architecture based exclusively on MLPs, but explicitly structured to extract information independently from colors and magnitudes. These representations are then concatenated and further processed through an additional dense layer, followed by a mixture density output layer with six Gaussian components.

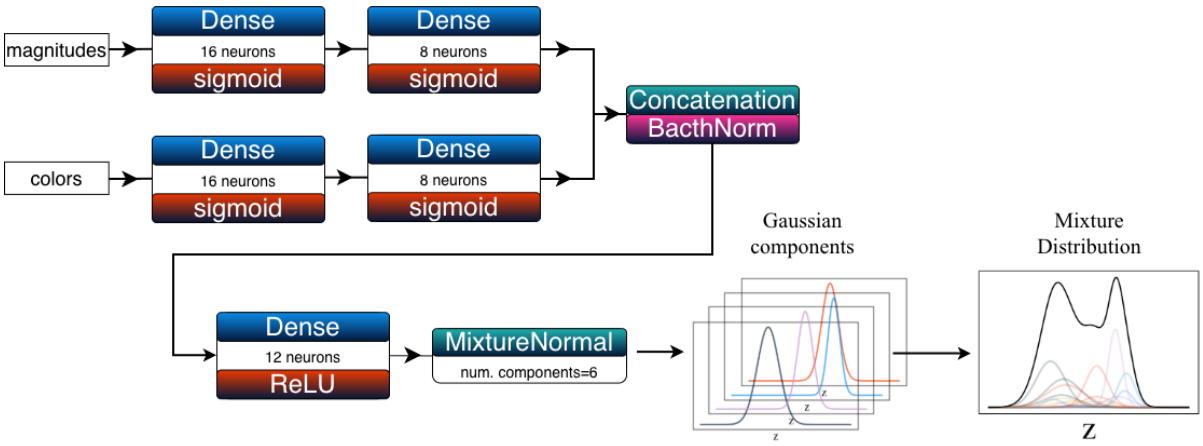

Figure 6.3: Architecture of the proposed model. Colors and magnitudes are processed independently through dedicated MLP branches to extract latent representations. These are then concatenated and passed through an additional dense layer, followed by a mixture density output layer with six Gaussian components.

The redshifts coming from this approach are refered as LS-DR10-CBPF in Méndez-Hernández et al. (2026) and the CMD from the target selection is show in Figure 6.2c, where we can see that the FRS region is indeed more populated, presenting a behavior closer to the S-PLUS candidate selection. In Table 6.1 we show the comparison between the quality metrics on the test sample for the public LS-DR10, and the LS-DR10-CBPF, proving that the new redshifts developed by us outperforms the official results presented in LS-DR10 in the faint-low-z region.

Table 6.1: Overall and faint low-z performance metrics for LS-DR10-CBPF and LS-DR10 photo-z estimates. Table from Méndez-Hernández et al. (2026)

Metric	**Selection**	**CBPF-z**	**LS-DR10**
Mean Bias	$m_r < 20.5$	0.00032	0.05869
Median Bias		−0.0021	0.0013
σ_{NMAD}		0.01531	0.01534
Outlier Fraction (%)		1.4	9.0
Mean error		0.035	0.105
Mean Bias	$18.5 < m_r < 20.5$;	0.028	0.146
Median Bias	$z_{spec} < 0.3$	0.0078	0.0136
σ_{NMAD}		0.028	0.032
Outlier Fraction (%)		5.6	19.6
Mean error		0.062	0.189
Mean Bias	$18.5 < m_r < 20.5$;	0.07	0.19
Median Bias	$z_{spec} < 0.07$	0.021	0.031
σ_{NMAD}		0.031	0.042
Outlier Fraction (%)		16.4	21.0
Mean error		0.106	0.234

These redshifts were exhaustively tested against the public ones for the target selection, and after rigorous validation were, became, together with the S-PLUS redshifts, the ones used for the target selection for the fainter objects ($m_r > 18.5$). More specifically, the final target list was divided into three sub-surveys:

- **Low-z bright (S1501):** bright ($m_r < 18.5$) photometric members selected using S-PLUS photometric redshifts, public photometric redshifts from LS-DR9 and/or LS-DR10, and/or photometric redshifts from LS-DR10-CBPF, and objects with $z_{spec} < 0.1$.

- **Low-z faint (S1505):** faint ($18.5 \leqslant m_r < 20.4$) photometric members selected using S-PLUS photometric redshifts when available and/or CBPF photometric redshifts, and objects with $z_{spec} < 0.1$

- **Low-z faint supplementary (S1506):** equivalent to S1505, with the addition of red sequence galaxies not captured by the photometric redshift selection. Although the CBPF photometric redshifts significantly improved the target selection at the

faint red end of the colour-magnitude diagram relative to LS-DR10, the very faint end ($m_r \gtrsim 19.5$) of the red sequence still presents lower target densities, motivating a complementary colour-based selection. Red sequence galaxies were identified within $\pm 1\sigma$ of a linear fit on the colour-magnitude diagram – details in Section 3.3 of Méndez-Hernández et al. (2026).

Additionally, for the Fornax cluster (Maddox et al., 2019), where the photometric redshift selection was not appropriate due to its proximity, our LS-DR10-CBPF photometric redshifts were adopted as the baseline for background cleaning, excluding objects with $z_{\rm phot} > 0.14$. The target selection for the Fornax cluster is datailed in Section 3.4 of Méndez-Hernández et al. (2026).

The photometric redshift catalogue presented here, developed as an extension of this thesis, represents a key contribution to CHANCES, and was recognised with builder status within the collaboration. A dedicated publication describing the photometric redshift pipeline, further methodological developments, and the full catalogue with associated PDFs is currently in preparation.

Right after the submission of the target list, even before any observations were conducted by 4MOST, these redshifts already proved to be valuable for LSS studies based solely on the target sample. In Baier-Soto et al. (2025), the authors explored the connection between the large-scale filamentary structure and the morphology of galaxy clusters in the Shapley and Horologium–Reticulum superclusters, based on the target list developed for the low-z subsurvey. They found that 82% of X-ray clusters are associated with optically detected filaments, and that the elongation of clusters is strongly aligned with the orientation of nearby filaments. This result provides observational evidence that matter accretion through filaments not only drives galaxy infall but also shapes the gravitational potential of clusters.

Similarly, in the recent work of Dulcien et al. (2026), the authors investigated the role of galaxy mergers in the pre-processing of galaxies prior to cluster infall, using a sample of $\sim 4.4 \times 10^4$ galaxies from the CHANCES survey around 33 low-redshift clusters ($z < 0.07$). Their analysis revealed that merging galaxies are preferentially located closer to cosmic filaments compared to the general galaxy population, particularly beyond the cluster virial radius, supporting a scenario in which filaments provide favorable conditions for galaxy mergers.

These results highlight that, even in the absence of spectroscopic confirmation, the target selection based on the photometric redshifts developed as an extension of this thesis already enables meaningful investigations of the cosmic web and its role in galaxy evolution.

7 Conclusion

In this thesis, we introduced a combination between RNNs and MDNs to estimate photometric redshifts in large scale ($> 17,000$ deg^2 of the sky). Our approach arises as an innovative way to generate reliable estimates of photometric redshifts even using only 3-band coverage.

We presented that the quality of our point estimates z_{phot} for DELVE overcomes usual ML techniques, and is comparable to the publicly available redshifts of LS-DR10. In Section 5.1, we explored the influence of the z distribution in the training set on the final results. We show that even using less data than the other training sets, the best results came from the A-trained model (DLTRAIN-A), showing that the homogeneity of the training set is critical to avoid significant biases in the high-z region, in agreement with the results from Zou et al. (2019).

Our model provides photo-z PDFs, which allow error estimations and statistical assessments for any DELVE galaxy in MCAT with at least three bands (including g and r). The main reason for using MDNs is the capability to generate mathematically well-defined PDFs for photo-z. This enables possible analytical procedures and facilitates uncertainty estimation. In Section 5.2 we validate the reliability of the PDFs through several diagnostics, showing that our method provides well-calibrated PDFs, except for a small subsample of galaxies with low odds, whose PDFs tend to be overly broad and poorly constrained. These objects are not representative, and are easily identifiable and can be excluded from any analysis by applying a simple odds threshold.

The final product of this work is one of the most extensive catalogs of photometric redshifts with complete probability density estimations, containing 347M new measurements of z_{phot} using DELVE photometry.

We show in Chapter 6 the impact of these photometric redshift methods, as well as extensions of the work presented in this thesis. Perspectives include formalizing the discussion conducted in the same chapter, providing detailed explanations regarding the development of the photo-z using LS-DR10 photometry.

Although the results presented here are consistent, a deeper investigation into the causes of the observed biases in the photo-z estimates is still needed. Furthermore, it remains unclear how the LMU – a network primarily designed to interpret continuous time series – processes magnitudes and colors, which carry no temporal correlation a priori. Our first tests suggest that the network somehow interprets the sequence of magnitudes as tracing the shape of the spectrum, however, no conclusive result has been found yet.

Another potential improvement for the photo-z estimates relies on including mor-

phological features as input. As shown in Lima (2025) and other studies such as Tagliaferri et al. (2003),morphological parameters also impact network performance. In this work we chose to proceed using only magnitudes and colors; however, we are already working on implementing morphology to evaluate its impact on DELVE-DR2 photo-z. In particular, Santana-Silva (in prep.) are currently developing a DL-based morphological classification for DELVE images, which we aim to combine with our photo-z models.

This work is published in Teixeira et al. (2024), the photo-z catalog can be accessed in `https://datalab.noirlab.edu/data/delve`, and the code used to build, train, and evaluate the model, as well to carry most of the analysis described in this thesis is publicly available in `https://github.com/gsmteixeira/`.

Part II

Transient Characterization

8 SN 2022acko and the Properties of Its Red Supergiant Progenitor

8.1 | Introduction

"The nitrogen in our DNA, the calcium in our teeth, the iron in our blood, the carbon in our apple pies were made in the interiors of collapsing stars. We are made of starstuff."
— Carl Sagan

This quotation from Sagan (1980) highlights that most of the elements that constitute life were synthesized billions of years ago in the interiors of stars. As the first generations of stars began to form – around 100 Myr after the Big Bang (Loeb and Furlanetto, 2013) – nucleosynthesis processes progressively enriched the interstellar medium with metals. Among the main contributors to this chemical enrichment are core-collapse supernovae (CCSNe), which play a central role in dispersing these elements into their surroundings. Beyond their importance for cosmic chemical evolution, these events are also of great interest for a variety of physical processes, and unresolved questions we will explore in this thesis.

CCSNe represent the terminal explosions of massive stars with initial masses greater than $8\,M_\odot$ (Woosley and Weaver, 1986). Within this category, Type II supernovae (SNe II) are characterized by the presence of prominent hydrogen lines in their spectra, indicating hydrogen-rich progenitor stars (Filippenko, 1997). The vast majority of these transient events are believed to result from the explosions of red supergiants (RSG) (Burrows et al., 1995). Throughout their lifetimes, RSGs undergo successive stages of nuclear burning, synthesizing progressively heavier elements until reaching the iron peak. Once nuclear fusion ceases, outward pressure–sustained by electron degeneracy–eventually fails due to electron capture and photodissociation, triggering core collapse and a SN (Wheeler, 2007). These explosions play a crucial role in enriching the interstellar medium with heavy elements and regulating star formation in galaxies (Sahijpal, 2014; Elmegreen, 1998). However, the detailed physical processes within these stars, especially in the final centuries to years before core collapse and their terminal explosions, remain not fully understood (Davies and Beasor, 2020; Fryer et al., 2012; Heger et al., 2003; Woosley et al., 2002).

Most SNe II arise from RSGs with extended hydrogen envelopes, producing a

characteristic plateau phase in their light curves that lasts approximately $60 - 90$ days, further classifying them as Type II-P supernovae (SNe II-P) (Arcavi, 2017; Barbon et al., 1979; Falk and Arnett, 1973). This plateau phase occurs because, after core collapse, a shock wave propagates through the outer layers of the red supergiant, heating the hydrogen envelope to temperatures of several tens of thousands of Kelvin, fully ionizing the material and making it highly luminous. As the envelope expands and cools, the temperature eventually drops below the hydrogen recombination threshold. At this point, protons and electrons begin to combine into neutral hydrogen. This recombination process releases thermal energy, which slows the decline in luminosity, acting like a photosphere with quasi-constant temperature. Once the entire hydrogen envelope has recombined, the plateau ends. The light curve then drops sharply as it enters the radioactive decay phase, powered by the decay of $^{56}\mathrm{Co} \rightarrow {}^{56}\mathrm{Ni}$.

In contrast, those SNe II lacking a well-defined plateau and instead exhibiting a linearly declining brightness are categorized as Type II-L SNe (Barbon et al., 1979; Carroll and Ostlie, 2017). Although this classification has been traditionally used–and investigated from a spectral perspective as well (Kou et al., 2020)–observations suggest that SNe II may span a continuum of properties rather than forming two distinct subtypes (Galbany et al., 2016; Valenti et al., 2016).

The observed population of SNe II with identified RSG progenitors exhibits a zero-age main sequence mass (M_{ZAMS}) distribution spanning approximately 7–19 $\mathrm{M}_{\odot}$(Smartt, 2009; Valenti et al., 2016; Auchettl et al., 2019; Davies and Beasor, 2020). In contrast, the broader population of RSGs includes stars with initial masses up to $25 - 30\,\mathrm{M}_{\odot}$ (Humphreys and Davidson, 1979; Davies et al., 2018; Katsuda et al., 2018). This discrepancy between the initial masses of confirmed SNe II progenitors and the full RSG population is known as the *red supergiant problem*. Notably, this discrepancy is also evident not only in direct observations of progenitor stars but also in other independent mass indicators, such as nebular spectroscopy of SNe II (Fang et al., 2025; Rodríguez, 2022), which will be explored later.

Several explanations have been proposed to account for the RSG problem, one possible explanation is that some massive RSGs may undergo core collapse and collapse directly into a black hole without producing an observable SN, resulting in so-called "failed SNe" (Sukhbold et al., 2016). Another explanation is that a small fraction of RSGs may exhibit lower luminosities in their final stages due to extreme variability (Soraisam et al., 2018; Levesque and Massey, 2020; Jencson et al., 2022; Beasor et al., 2025).

Alternatively, the RSG problem may arises from observational biases, particularly the limited number of SNe with confidently identified progenitor stars (Davies and Beasor, 2020). In such cases, progenitor mass estimates derived from pre-explosion images may be systematically biased by factors such as variability or the absence of infrared observations, which can obscure or underestimate the intrinsic luminosity. As a result, in-

dependent methods for estimating progenitor masses (such as nebular spectroscopy, light curve modeling, or environment analysis) are critical for analyzing whether our detections yield accurate and consistent mass constraints. Resolving the RSG problem requires a larger sample of SNe observations with well-characterized progenitor systems. Such efforts are essential not only for clarifying the fate of massive stars but also for advancing our broader understanding of the SNe II-P population.

The ideal scenario for connecting SN explosions to their respective progenitor stars and estimating their mass involves having archival data that directly captures the progenitor in pre-explosion imaging (e.g., Dyk, 2017). Such data enable mass estimates through template fitting by comparing the observed luminosity with stellar evolution models across a range of initial stellar masses. The best-fitting template corresponds to a set of stars whose predicted luminosities are consistent with the observed value, thereby constraining the progenitor's initial mass. In the case of SNe II-P, which are known to originate from RSGs, the progenitor mass can also be estimated by fitting RSG SED models to the observed SED, particularly when multi-band photometry is available. While these methods offer direct constraints on progenitor mass, they are often not feasible–particularly for more distant SNe ($\gtrsim 28$ Mpc), as obtaining photometry of resolved massive stars becomes increasingly difficult beyond this range (Smartt, 2015).

In addition to using direct detections, several indirect methods can be used to infer progenitor properties from the observed SNe II-P. For example, on the absence of optically thick circumstellar material (CSM) around the progenitor, the early-time emission is primarily shaped by the physical structure of the progenitor star, particularly its radius and envelope mass. When the shock wave generated by the core collapse reaches the stellar surface, in the so-called shock breakout phase, it releases a brief burst of high-energy radiation. This is immediately followed by a cooling phase, during which the hot, expanded envelope radiates thermal energy as it cools and becomes progressively more transparent (Morag et al., 2023; Waxman and Katz, 2017; Sapir and Waxman, 2017). The duration, shape, and brightness of this early light curve are directly tied to the progenitor's properties.

For SNe II-P, the plateau phase of the light curve is powered by hydrogen recombination in the extended hydrogen-rich envelope (Kasen and Woosley, 2009; Popov, 1993). When recombination ceases and the ejecta become optically thin, the explosion transitions into the nebular phase. During this stage, the spectrum is dominated by emission lines from elements synthesized during the progenitor's lifetime and the explosion itself (Jerkstrand, 2017; Dessart et al., 2021). In this phase, prominent lines such as [O I] at 5577, 6300, and 6364 Å and [Ca II] at 7292 and 7324 Å correlate with the total ejecta mass of these elements (Jerkstrand, A. et al., 2012). By comparing observed nebular-phase spectra with theoretical non-local thermodynamic equilibrium radiative transfer (NLTERT) models informed by nucleosynthesis yields, one can estimate the progenitor's initial mass

by matching the synthetic spectra to the observed features (e.g., Jerkstrand et al., 2014; Tinyanont et al., 2021; Kilpatrick et al., 2023b).

Beyond the initial mass, other SN observables can also be linked to progenitor properties. For instance, studies by Childress et al. (2015) and Valenti et al. (2016) suggest that the duration and luminosity of the plateau phase, as well as the amount of synthesized ^{56}Ni correlate with the progenitor's mass. These correlations are often interpreted through empirical mappings derived from SNe II with direct progenitor detections, such as those presented in Eldridge et al. (2019); Rodríguez (2022). These photometric indicators thus provide additional constraints on the nature of the progenitor system, complementing our spectroscopic analyses.

In this part of the thesis – particularly in this chapter – we contribute to the effort of mapping SNe II to their RSG progenitor systems by analyzing a specific object: SN 2022acko. This SN was discovered in the galaxy NGC 1300 by the Distance Less Than 40 Mpc (DLT40; Tartaglia et al., 2018) survey on 2022 December 6. SN 2022acko was classified as a SNe II-P within the first 24 hr of detection, based on spectral data obtained with the Yunnan Faint Object Spectrograph and Camera (YFOSC) (Wang et al., 2019), as reported by Li et al. (2022).

This study focuses on constraining the progenitor's initial mass using both direct estimates from pre-SN imaging and indirect inferences based on the SN's photometric and spectroscopic evolution. Throughout our analysis, we adopt a redshift of $z = 0.00526$ (Springob et al., 2005) and a distance of 19.0 ± 2.9 Mpc (Anand et al., 2020) for NGC 1300. We will also present new optical and ultraviolet photometry of SN 2022acko spanning 1–350 days after detection, complemented by deep pre-explosion imaging previously reported by Van Dyk et al. (2023), as well as additional late-time spectroscopy covering approximately $200 - 600$ days post-explosion.

In Section 8.2, we describe the photometric and spectroscopic observations of SN 2022acko presented in this work. The characterisation of the progenitor system is then addressed through three complementary approaches: pre-explosion imaging in Section 8.3, shock-cooling modelling of the early light curve in Section 8.4, and nebular spectroscopy analysis in Section 8.5. In Section 8.6, we discuss the implications of our estimates for the progenitor properties and evaluate the role of mass loss during the progenitor's lifetime. Finally, Section 8.7 summarises our main findings.

8.2 | Observations

The published version of this work (Teixeira et al., 2026) was led by me. Although I conducted most of the analysis and was responsible for the main findings, this would

not have been possible without the efforts of those who carried out the observations. In particular, the main contributions to the data acquisition and processing were made by the co-authors Charles D. Kilpatrick[1] and André S. Santos[2].

In this section, we present the observations and data acquisition associated with SN 2022acko. This part of the work relies primarily on the efforts of the co-authors mentioned above, who led the data acquisition and initial processing, complemented by previously published datasets used in our analysis.

8.2.1 | Pre-explosion data

Our analysis includes pre-explosion imaging from Hubble Space Telescope (HST) and Spitzer Space Telescope (Spitzer). We analyzed all available observations obtained from these instruments at the site of SN 2022acko in NGC 1300. The HST imaging spans from 6 January 2001 to 4 January 2020, corresponding to approximately 22 to 3 yr prior to the explosion, and was acquired using Wide Field Planetary Camera 2 (WFPC2), Advanced Camera for Surveys (ACS), and Wide Field Camera 3 (WFC3), as summarized in Table 8.1. We downloaded all calibrated HST image frames from the Mikulski Archive for Space Telescopes[3] using our automated pipeline `hst123` (Kilpatrick, 2021). Each image was optimally aligned using `TweakReg` and individual epochs and filters were separately drizzled using `astrodrizzle` – both are tools from the `DrizzlePac` package (Hoffmann et al., 2021). We then performed photometry across the calibrated and aligned frames using `dolphot` (Dolphin, 2016) with the F814W image as a reference frame. We saved all photometry of point-like sources for analysis, which is further described in Section 8.3.

We similarly processed the Spitzer imaging from the Infrared Array Camera (IRAC) using our custom pipeline described in (Kilpatrick et al., 2023b; Kilpatrick and Rubin, 2023). There were four separate epochs of IRAC imaging obtained from 2009–2011, which we combined into two stacks for each filter using `mopex` (Makovoz and Khan, 2005). We performed photometry of point sources in each frame using the instrumental PSF from Spitzer.

We detect a single point-like counterpart close to the reported site of SN 2022acko in HST/F814W and F160W imaging, and we report the brightness of this source and upper limits in all remaining bands in Table 8.1. We assess whether this candidate is likely associated with SN 2022acko in Section 8.3.

[1] https://orcid.org/0000-0002-5740-7747
[2] https://orcid.org/0000-0002-1420-3584
[3] https://mast.stsci.edu/

MJD	Instrument	Filter	Brightness (AB mag)
51915.55779	WFPC2	F606W	>25.99
53269.61735	ACS/WFC	F555W	>26.86
53269.62340	ACS/WFC	F435W	>27.43
53274.49244	ACS/WFC	F814W	26.24± 0.07
53274.49845	ACS/WFC	F658N	>24.90
58053.50298	WFC3/IR	F160W	24.42±0.03
58852.38389	WFC3/UVIS	F336W	>26.33
58852.44509	WFC3/UVIS	F275W	>25.74
–	*Spitzer*/IRAC	Ch 1	>22.00
–	*Spitzer*/IRAC	Ch 2	>22.05

Table 8.1: Photometry of the pre-explosion counterpart to SN 2022acko from HST and Spitzer imaging.

8.2.2 | High-resolution imaging

We obtained a high-resolution image of SN 2022acko with the Gemini South Adaptive Optics Imager (GSAOI) (McGregor et al., 2004) in H-band on 11 January 2023, roughly 38 days post-explosion. All imaging was obtained in conjunction with the Gemini Multi-conjugate Adaptive Optics System (GeMS) (Rigaut et al., 2014) in laser guide star mode, with 43×30 s images and resulting in an average full-width at half-maximum (FWHM) of the SN 2022acko PSF of 100 mas. We processed all GSAOI imaging using Gemini `pyraf` methods (Labrie et al., 2010), including pixel-level calibration using dark and flat-field frames, sky subtraction from sky frames obtained 3′ from SN 2022acko and obtained in conjunction with the science frames, distortion correction with `disco_stu`[4], and image stacking to a common, undistorted image frame using `SWarp` (Bertin, 2010). SN 2022acko and several dozen unresolved astrometric calibrators in NGC 1300 are clearly detected in the final image stack, which we discuss further in Section 8.3.

8.2.3 | Optical/UV observations

We acquired optical photometry data from Bostroem et al. (2023), covering the first 150 days after explosion, obtained using several instruments. This dataset includes *UBgVri*-band observations from the Las Cumbres Observatory robotic telescope network (Brown et al., 2013), *co*-band data from the Asteroid Terrestrial-impact Last Alert System (ATLAS) (Tonry et al., 2018), and *BgVri* and *Open*-filter observations from the 0.4 m Panchromatic Robotic Optical Monitoring and Polarimetry Telescopes (PROMPT) telescopes as part of the DLT40 program (Tartaglia et al., 2018).

[4] `https://www.gemini.edu/sciops/data/software/disco_stu.pdf`

Additional ultraviolet and optical observations were obtained with the Ultraviolet/Optical Telescope (UVOT) aboard the Neil Gehrels Swift Observatory, in the *UVW2*, *UVM2*, *UVW1*, U_S, B_S, and V_S bands (Roming et al., 2005).

Furthermore, we supplemented the light curve dataset with *griz*-band imaging obtained using the 0.8m T80-South robotic telescope located at the Cerro Tololo Inter-American Observatory (CTIO), Chile, as part of the S-PLUS Transient Extension Program (STEP) (Santos et al., 2024). STEP emerges as a time-domain branch of the on-going S-PLUS (Mendes de Oliveira et al., 2019). Early observations were acquired starting 13 days after the discovery date on December 19, 25 and 30 of 2022, while late-time imaging was obtained between 300 days after explosions. We used a modified version of `DoPHOT`(Schechter et al., 1993) to perform forced photometry (i.e., measuring the flux at a given position regardless of whether a source is formally detected) at the supernova location, then measuring AB magnitudes after image subtraction using PAN-STARRS DR2 images as templates.

An additional set of late-time photometric data was also obtained 200 days after explosion using the same filter configuration with the Las Cumbres Observatory network, and in *G* and *Open* filters through the DLT40 survey. These datasets provide valuable constraints on the fading light curve, enhancing our temporal coverage during the decline phase. The novel light-curve information not included in Bostroem et al. (2023) were publicly released in Teixeira et al. (2026). All photometric observations of SN 2022acko used in the published work are shown in Figure 8.1. For this thesis, the most relevant portions of these data are the first ~ 15 d, used in Section 8.4, and the very late phases observed by STEP, which are used for spectral scaling in Section 8.5.

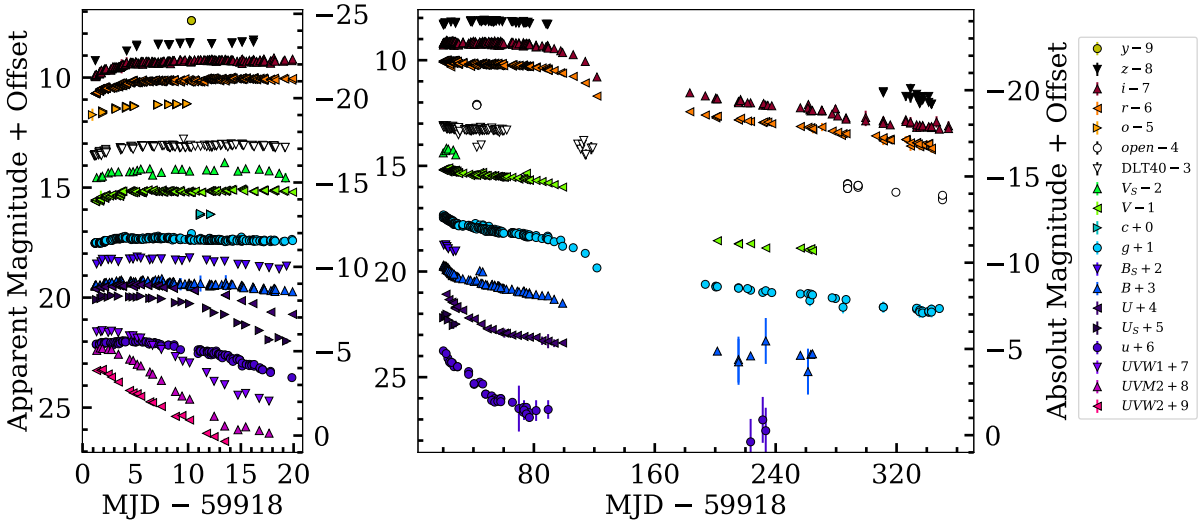

Figure 8.1: Apparent and absolute magnitudes of SN 2022acko over the first 350 days of observation. The left panel highlights the early-time evolution (first 20 days), while the right panel displays the later phases. Apparent magnitudes are shown on the left y-axis, and the corresponding absolute magnitudes are overlaid on the right y-axis. The y-axes labels and the legend include offset values for a better visual representation of the data.

8.2.4 | Spectroscopy

We observed SN 2022acko with the Keck-I and Keck-II 10 m telescopes at Mauna Kea Observatory, Hawaii. We used the DEIMOS spectrograph (Faber et al., 2003) on Keck-II on 2023 Sep 5 (275 days post-explosion) and the Low Resolution Imaging Spectrometer (LRIS) (Oke et al., 1995) on Keck-I 2024 Jan 6 and 2024 Aug 9 (396 and 612 days post-explosion, respectively). DEIMOS was used in a low-resolution 600ZD grating mode with a central wavelength approximately at 7500 Å and flux calibrated with the standard star BD 28+4211 [5]. Similarly, we used the B400/3400 grism and R400/8500 grating in conjunction with the d560 dichroic on LRIS, covering approximately 3000–10500 Å. We flux calibrated all epochs using spectra of BD 28+4211 obtained on the same night as the SN 2022acko observations. All data were processed with `pypeit` v1.16 (Prochaska et al., 2020), using dome flats and arc lamp exposures obtained on the same night and in the same instrumental setup as each science frame. We performed optimal coaddition of the one-dimensional spectra and telluric correction using `pypeit` atmospheric grids for Maunakea. The final calibrated spectra are analyzed in Section 8.5.

8.3 | Progenitor Mass from Pre-Explosion Imaging

In this section, we present the astrometric identification of a pre-explosion counterpart to SN 2022acko by aligning high-resolution Gemini/GSAOI imaging obtained shortly after its discovery with archival Hubble Space Telescope pre-explosion observations. Here, We describe the alignment procedures, present evidence for the progenitor identification, and characterize the physical properties of the identified counterpart.

This section constitutes a fundamental component of the analysis presented in this work. Most of this pre-explosion analysis was primarily conducted by our collaborator, Charles D. Kilpatrick (co-author in Teixeira et al. (2026)).

[5] https://www.eso.org/sci/observing/tools/standards/spectra/bd28d4211.html

8.3.1 | A pre-explosion counterpart to SN 2022acko

We aligned our GSAOI H-band image of SN 2022acko (Section 8.2.1) to pre-explosion *HST* imaging in F814W in order to determine whether a counterpart was present in pre-explosion imaging at the site of the supernova. We identified 27 sources present in both the GSAOI image and the *HST* imaging that appear point-like in both frames (Figure 8.2).

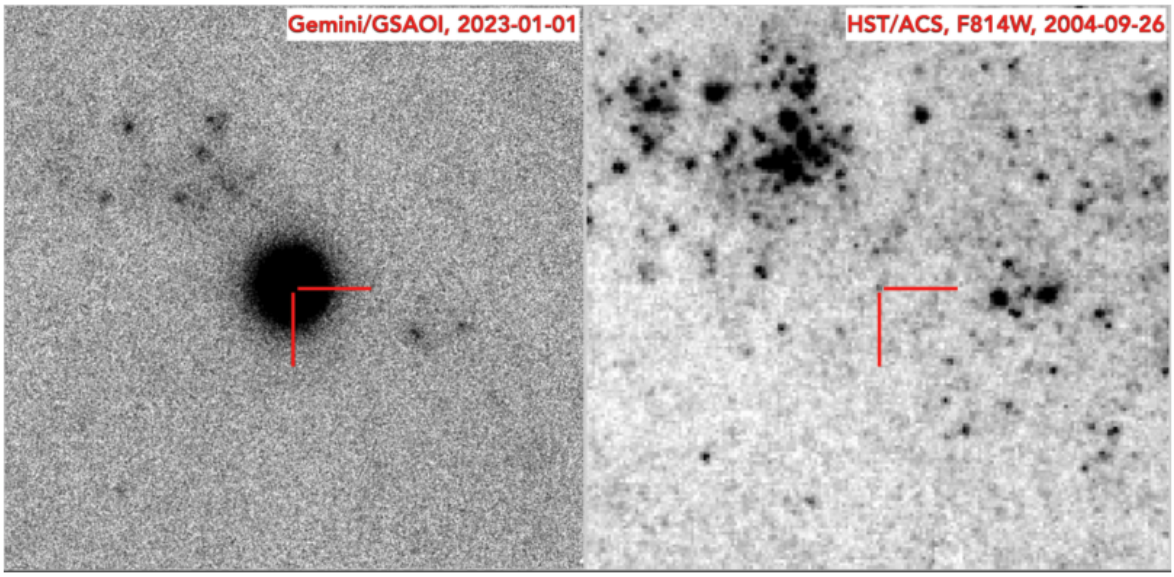

Figure 8.2: The site of SN 2022acko from 1 January 2023 as observed in high-resolution GSAOI H-band imaging (left) compared with the same site in pre-explosion ACS F814W imaging from 26 September 2004 (right). We identify a single, unblended, point-like source in the pre-explosion emission (shown at the position of the red crosshairs) that is astrometrically consistent with the supernova.

Following the methods in Kilpatrick et al. (2023b), we use these sources to align the two images frames in the following way. We first split the sample of common alignment sources in half and calculate an alignment solution that aligns the Gemini/GSAOI→*HST* image frame. We then estimate the systematic uncertainty from our solution by comparing the predicted coordinates of the remaining half to the sample using this solution. Repeating this procedure 1000 times, we take the average dispersion in right ascension and declination between the calculated GSAOI source coordinates and known *HST* source coordinates as the systematic uncertainty in our solution, which is approximately δR.A. $= 0.02''$ and δDec $= 0.03''$. We adopt the best-fitting alignment solution as the one calculated from all 27 sources.

Next, we determine where SN 2022acko is located in pre-explosion imaging based on the alignment solution estimated above. Due to the extremely high signal-to-noise detection of SN 2022acko in our GSAOI imaging, the uncertainties in our centroid from that detection are negligible compared with our alignment uncertainties (e.g., $< 0.001''$ in

the GSAOI frame). Thus, we find that the resulting position of SN 2022acko is coincident with a single, unblended point-source detected at $\approx 14\sigma$ significance in the F814W image to within $\approx 0.01''$(nominally 0.4σ of the astrometric uncertainty), well within the systematic uncertainties calculated above.

There are no other point-like sources detected at the $>3\sigma$ level within $\approx 0.2''$ (8σ of the astrometric uncertainty) of SN 2022acko. This pre-explosion counterpart identification is consistent with the independent alignment performed using *JWST*/NIRCam imaging reported in Van Dyk et al. (2023).

We estimate the probability of chance coincidence between SN 2022acko and the pre-explosion counterpart by considering that we detect 254 point-like sources at $>3\sigma$ significance in the F814W frame within $10''$ of SN 2022acko. This means that there is a $\approx 1.3\%$ chance that SN 2022acko would have intersected with a random point source to within 3σ astrometric significance by chance, which we take as a conservative estimate of the probability of chance coincidence. While nominally a low significance, this residual low likelihood of a chance detection reinforces the need for future *HST* follow up to determine whether the pre-explosion counterpart has in fact disappeared, confirming its association with SN 2022acko.

Finally, with the detection of a single, high-significance counterpart to SN 2022acko in our F814W image, we align this frame to the remaining *HST* and *Spitzer* imaging. The alignment uncertainties are generally small (≈a few mas) in *HST* imaging, but they approach $0.01''$ in both *Spitzer*/IRAC Channels 1 and 2. We obtain a detection of the counterpart in SN 2022acko as previously reported in Van Dyk et al. (2023), but no detections in any other *HST* or the *Spitzer* bands. As this source is aligned with the position of the counterpart constrained from the discovery coordinate and reported in Table 8.1, we use those data in the analysis described below.

8.3.2 | Characterization of the Pre-explosion Counterpart

With the detection of the SN 2022acko pre-explosion counterpart in F814W and F160W imaging as well as limits from F336W to IRAC Channel 2, we characterize the potential luminosity and spectral type of the putative counterpart using methods described in Kilpatrick et al. (2023b) and Kilpatrick et al. (2023a). We compare these data using a Markov Chain Monte Carlo (MCMC) fitter using both a simple blackbody model (i.e., fit with $T_{\rm eff}$ and $\log(L/\,{\rm L}_{\odot})$) and a grid of MARCS (Gustafsson et al., 2008) stellar SEDs combined with an opacity simulating a RSG stellar wind described in the papers above. The model is described using the intrinsic RSG luminosity (L), the effective temperature

of the RSG photosphere ($T_{\rm eff}$), a line-of-sight opacity through the RSG wind in V-band (τ_V), and a temperature describing cool blackbody emission from stellar light reprocessed in any dust present in the wind ($T_{\rm dust}$). We incorporate line-of-sight extinction from the Milky Way, but due to the lack of any apparent host extinction in the light curves and spectra of SN 2022acko (Bostroem et al., 2023), we do not include additional extinction from NGC 1300. The results of our best-fitting SED are presented in Figure 8.3. For comparison, we also show the results of the best-fitting model assuming pure blackbody emission between F814W and F160W.

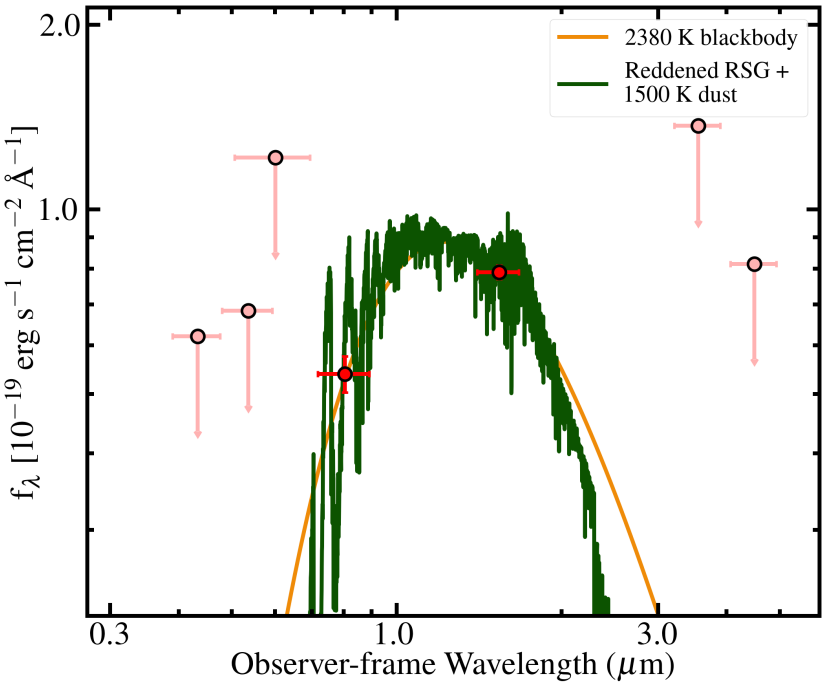

Figure 8.3: Photometry of the SN 2022acko pre-explosion counterpart in HST F814W and F160W (red circles) as described in Section 8.2.1. Upper limits from other HST and Spitzer bands are shown in pink. We fit these data using two models described in Section 8.3.2; a pure blackbody with a temperature of 2380 K and luminosity of $\log(L/\,{\rm L}_\odot) = 4.27$ *and a more realistic MARCS RSG with added circumstellar reddening from a shell of* 1500 *K dust and a total luminosity of* $\log(L/\,{\rm L}_\odot) = 4.3$.

The best-fitting blackbody results in a photosphere with effective temperature 2380 K and $\log(L/\,{\rm L}_\odot) = 4.27 \pm 0.18$, which is significantly cooler than realistic RSG SEDs from the MARCS models. We note that as these data are fit to two photometric points, there could be significant systematic uncertainties on these quantities due to overfitting. Assuming that the underlying source is dominated by a single RSG photosphere and given the lack of significant line-of-sight extinction, we infer that there is some additional source of reddening in the local environment around the SN 2022acko progenitor system that our RSG model may capture. Moreover, the implied luminosity is consistent with a 7.5 ± 0.5 $M_\odot$ RSG following models in Choi et al. (2016), which is extremely low for a bare RSG photosphere. This suggests that either the SN 2022acko progenitor system is anomalously low mass for a SN II-P progenitor star or we are missing a significant

fraction of the emission from this source at wavelengths beyond F814W and F160W. We note that this conclusion is consistent with findings in Van Dyk et al. (2023), who compare to Binary Population and Spectral Synthesis (BPASS) (Eldridge et al., 2017) models with luminosities as low as $\log(L/\,\mathrm{L}_\odot) = 4.3$ and initial mass 7.7 $\mathrm{M}_\odot$.

We instead turn to our MARCS RSG models to determine the most luminous the SN 2022acko progenitor system could be once we incorporate the *Spitzer*/IRAC Channels 1 and 2 limits. Combining systematic uncertainty in the model and uncertainty in the distance, the maximum luminosity model that is consistent with our photometry at 1σ is $\log(L/\,\mathrm{L}_\odot) \approx 4.5$.

8.4 | Progenitor Mass from Early Light Curve

The early light curves of SNe are dominated by Shock Cooling (SC) emission, which provides constraints on the progenitor structure and the energy deposited in the outer envelope during shock breakout (Sapir and Waxman, 2017). In this section, we analyze and compare the early light curve of SN 2022acko with the SC models developed in Morag et al. (2023) (hereafter MSW23).

MSW23 presents a unified framework that combines previous numerical and semi-analytical descriptions of shock-cooling evolution (Sapir and Waxman, 2017; Rabinak and Waxman, 2011; Katz et al., 2012). In this framework, the authors provide a simple analytic description for the time evolution of the bolometric luminosity, L, and the color temperature, $T_{\rm col}$, of SC emission produced by the explosion of RSGs. They also introduce modifications to account for line blanketing in the UV at early phases of the explosion by incorporating a suppression factor in the spectral energy distribution.

The model assumes a spherically symmetric RSG with total mass M and radius R, composed of a uniform core surrounded by a polytropic envelope of mass $M_{\rm env}$. In this framework, the bolometric luminosity is given by

$$L(\tilde{t}) = L_{\rm br}\left\{\tilde{t}^{-4/3} + 0.9\exp\left[-\left(\frac{2.0t}{t_{\rm tr}}\right)^{0.5}\right]\tilde{t}^{-0.17}\right\}, \tag{8.1}$$

and the color temperature by

$$T_{\rm col}(\tilde{t}) = T_{\rm col,br}\min(0.97\tilde{t}^{-1/3}, \tilde{t}^{-0.45}). \tag{8.2}$$

Here, $\tilde{t} = (t - t_0)/t_{\rm br}$, where t_0 is the time of shock breakout (i.e., when the shock reaches the stellar surface). The transparency timescale is defined as $t_{\rm tr} = (19.5\mathrm{d})(\kappa M_{\rm env}/v_{s*})^{1/2}$ corresponding to the epoch at which photons begin to escape from deeper layers of the

envelope. This model adopts $\kappa = 0.34\,\mathrm{cm^2\,g^{-1}}$, as being the opacity due to electron scattering in fully ionized H and He. The "br" subscripts denote breakout quantities, which are directly determined from the model parameters and take the forms

$$t_{\rm br} = (0.86\,\mathrm{h}) R^{1.26} v_{\rm s*}^{-1.13} (f_\rho M \kappa)^{-0.13}, \tag{8.3}$$

$$L_{\rm br} = (3.69 \times 10^{42}\,\mathrm{erg\,s^{-1}}) R^{0.78} v_{\rm s*}^{2.11} (f_\rho M)^{0.11} \kappa^{-0.89}, \tag{8.4}$$

and

$$T_{\rm col,br} = (8.19\,\mathrm{eV}) R^{-0.32} v_{\rm s*}^{0.58} (f_\rho M)^{0.03} \kappa^{-0.22}, \tag{8.5}$$

where $v_{\rm s*}$ is the characteristic shock velocity at the stellar surface (related to the explosion energy E by $v_{\rm s*} \sim (E/M_{\rm env})^{1/2}$), and f_ρ is a numerical factor associated with the polytropic structure of the progenitor. In this formalism, the product $f_\rho M$ always appears together and is therefore treated as a single parameter.

In this formulation, the RSG mass M can be interchangeably referred to as the ejecta mass, as the model assumes that the entire stellar mass (core and envelope) is ejected, leaving no compact remnant. Together, these equations define the SC emission of a SN originating from a RSG, parameterized by the set θ given by

$$\theta = (v_{\rm s*}, M_{\rm env}, f_\rho M, R, t_0) \tag{8.6}$$

The model was implemented using the `lightcurve_fitting` package (Hosseinzadeh et al., 2023a), and the parameters were inferred using MCMC methods. At each observed epoch t and filter f of our light curve (Figure 8.1), the algorithm computes L and $T_{\rm col}$ from Equations 8.1 and 8.2. From these quantities, it derives the blackbody radius $R_{\rm bb}$ by assuming a blackbody with effective temperature equal to $T_{\rm col}$.

The corresponding spectral energy distribution is then constructed as a Planck spectrum, L_λ, which is convolved with the filter transmission function f to obtain model fluxes (or equivalently, luminosities or magnitudes) that can be directly compared with the observations. The likelihood is evaluated, and the MCMC routine explores the parameter space accordingly. This methodology has also been applied to other SNe, including SN 2021yja, SN 2023ixf, and SN 2020jfo (Hosseinzadeh et al., 2022, 2023b; Kilpatrick et al., 2023b).

Preliminary runs showed that the model converges toward values of $f_\rho M$ close to zero for SN 2022acko, while $M_{\rm env}$ samples are concentrated around $\sim 3\ \mathrm{M_\odot}$. This behavior is inconsistent with the model assumptions, since, from the radiation-hydrodynamic simulations in Sapir and Waxman (2017) (Figure 5 in their work), the parameter f_ρ is expected to be of order unity.

To enforce a physically meaningful constraint, i.e., that the ejecta mass must be

at least comparable to the envelope mass, we modified the log-likelihood function to discard samples for which $f_\rho M < 0.1 M_{\rm env}$. This condition ensures that only models with physically plausible values of f_ρ are retained in the posterior distribution.

We ran the MCMC using 32 walkers, with 5000 burn-in steps followed by an additional 5000 sampling steps. The best-fit values and their corresponding prior distributions are presented in Table 8.2. In this run, we included an intrinsic scatter parameter, σ, which inflates the observational uncertainties by a factor of $\sqrt{1+\sigma}$.

The phases used in the fit span from MJD 59918.9 (first detection) to $t_{\rm max}$, where $t_{\rm max}$ is defined as the upper validity limit of the model. This threshold corresponds to the epoch at which the photospheric temperature drops below 0.7 eV, at which point the assumption of a fully ionized gas – and thus constant opacity – breaks down. These values can be derived from the model parameters (see Eq. 24 of Sapir and Waxman (2017) and Eq. 19 of Morag et al. (2023)).

We iteratively performed the fits while decreasing the maximum MJD included in the SN 2022acko light curve until the inferred $t_{\rm max}$ exceeded the latest epoch used in the fit. This ensures that the model remains valid over the full temporal range of the fitted data. We find $t_{\rm max} = 59931.54 \pm 0.34$, meaning that we used the first $\sim$ 13d of the light curve presented in 8.1 for this fit.

Figure 8.4 shows the results of our analytical fits compared to the observed data, while Figure 8.5 presents the posterior distributions of the model parameters. The fitted curves appear to be in good agreement with the data, although a slight mismatch is observed in the UV bands. This behavior is expected and may arise from UV line blanketing in SNe II. This discrepancy is mitigated in the MSW23 framework compared to earlier versions of SC models, as demonstrated by Hosseinzadeh et al. (2023b) for SN 2023ixf and by Teixeira et al. (2026) for SN 2022acko (the published version of this work).

The posterior distributions of the parameters are generally close to Gaussian, showing no significant pathological behavior. The exceptions are $M_{\rm env}$ and f_ρ, which exhibit slight skewness induced by the manual modification of the likelihood to exclude unphysical samples.

We acknowledge that this is not an optimal approach, as such constraints should be

Table 8.2: Parameter priors and best-fit values obtained with the shock-cooling model of (Morag et al., 2023) for the SN 2022acko.

		Prior			Best fit values	
Parameter	Variable	Shape	Min.	Max	MSW23	Units
description	v_{s*}	Uniform	0	10.0	$1.12^{+0.06}_{-0.06}$	10^3 km s^{-1}
description	$M_{\rm env}$	Uniform	0	10.0	$2.0^{+0.4}_{-0.3}$	$M_\odot$
description	$f_\rho M$	Uniform	$0.1 \times M_{\rm env}$	100	$0.21^{+0.04}_{-0.03}$	$M_\odot$
description	R	Uniform	0	14374	$580.0^{+20.0}_{-20.0}$	$R_\odot$
description	t_0	Uniform	59910.94	59920.44	$59917.44^{+0.06}_{-0.06}$	MJD
description	σ	LogUniform	0	10^2	$6.8^{+0.1}_{-0.1}$	Dimensionless

intrinsically incorporated into the modeling and sampling framework. In fact, an updated version of the `lightcurve_fitting` package is currently under development to address this issue (based on direct communication with collaborators).

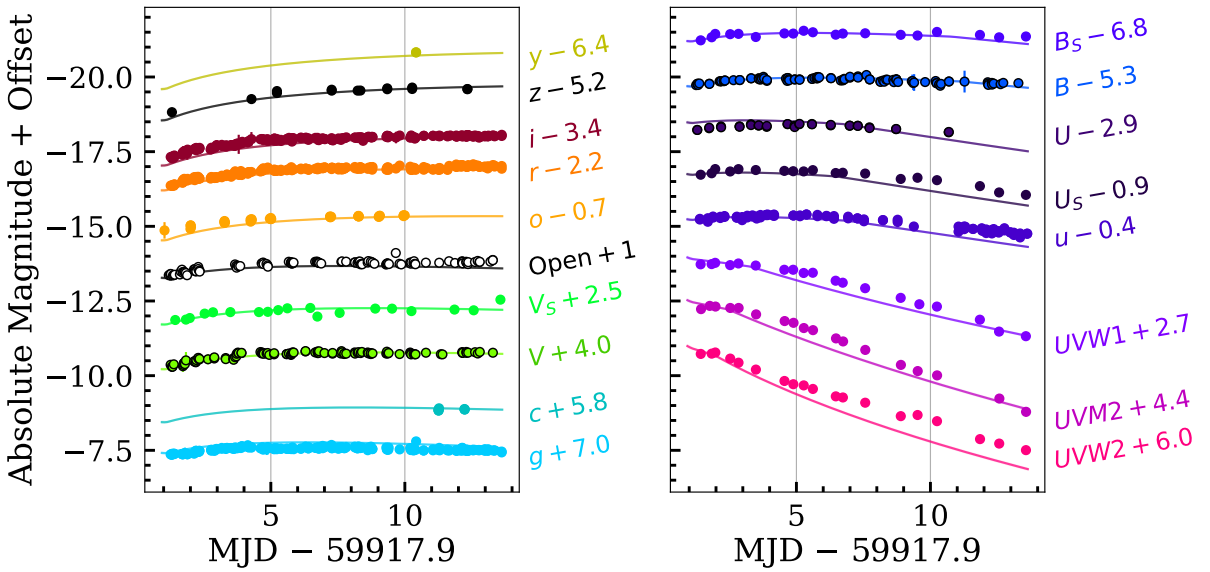

Figure 8.4: Upper panel: Best-fit models from MSW23 (solid lines) and SW17 (dashed lines) compared to the observed early-time light curve. Lower panels: Residuals (biases) between the models and the observed data, shown as a function of MJD. Points represent the mean residuals within MJD bins. All models are plotted only within their respective validity time ranges.

The best-fit parameters for the MSW23 model, as listed in Table 8.2, are $v_{s*} = 1.12^{+0.06}_{-0.06} \times 10^3\,\mathrm{km\,s^{-1}}$, $M_{\mathrm{env}} = 2.0^{+0.4}_{-0.3}\ \mathrm{M}_\odot$, $f_\rho M = 0.21^{+0.04}_{-0.03}\ \mathrm{M}_\odot$, $R = 580^{+20}_{-20}\ \mathrm{R}_\odot$, $t_0 = 59917.44^{+0.06}_{-0.06}\,\mathrm{MJD}$, and $\sigma = 6.8^{+0.1}_{-0.1}$. As discussed previously, $f_\rho M$ consistently trends toward the lower bound of $\sim 0.1\,M_{\mathrm{env}}$, which would imply an unusually low ejecta mass. Moreover, the explosion epoch inferred from the SC fit is inconsistent with the last nondetection of SN 2022acko (MJD 59918.17, from the ATLAS survey Bostroem et al., 2023), corresponding to ~ 17 hours prior to that observation.

The radius $R = 580^{+20}_{-20}\ \mathrm{R}_\odot$ inferred from the MSW23 model is consistent with the expected range of 100–1000 $\mathrm{R}_\odot$ (Levesque and Massey, 2020) for RSGs. To estimate the progenitor mass, we compared this radius derived from our SC modeling with the terminal radii of stars at their final evolutionary stages, as provided by the Modules for Experiments in Stellar Astrophysics (MESA) Isochrones & Stellar Tracks (Choi et al., 2016).

Assuming a metallicity of [Fe/H] $= -0.22$, consistent with measurements in the disk region of NGC 1300 (Rosado-Belza, D. et al., 2020), Figure 8.6 presents the terminal radius and luminosity of stars at the end of their evolution for different values of M_{ZAMS}, together with the progenitor constraints derived in Section 8.3. This comparison suggests a progenitor mass of $M_{\mathrm{ZAMS}} \sim 9$–$10\ \mathrm{M}_\odot$, consistent with typical values for RSGs (Davies and Beasor, 2018).

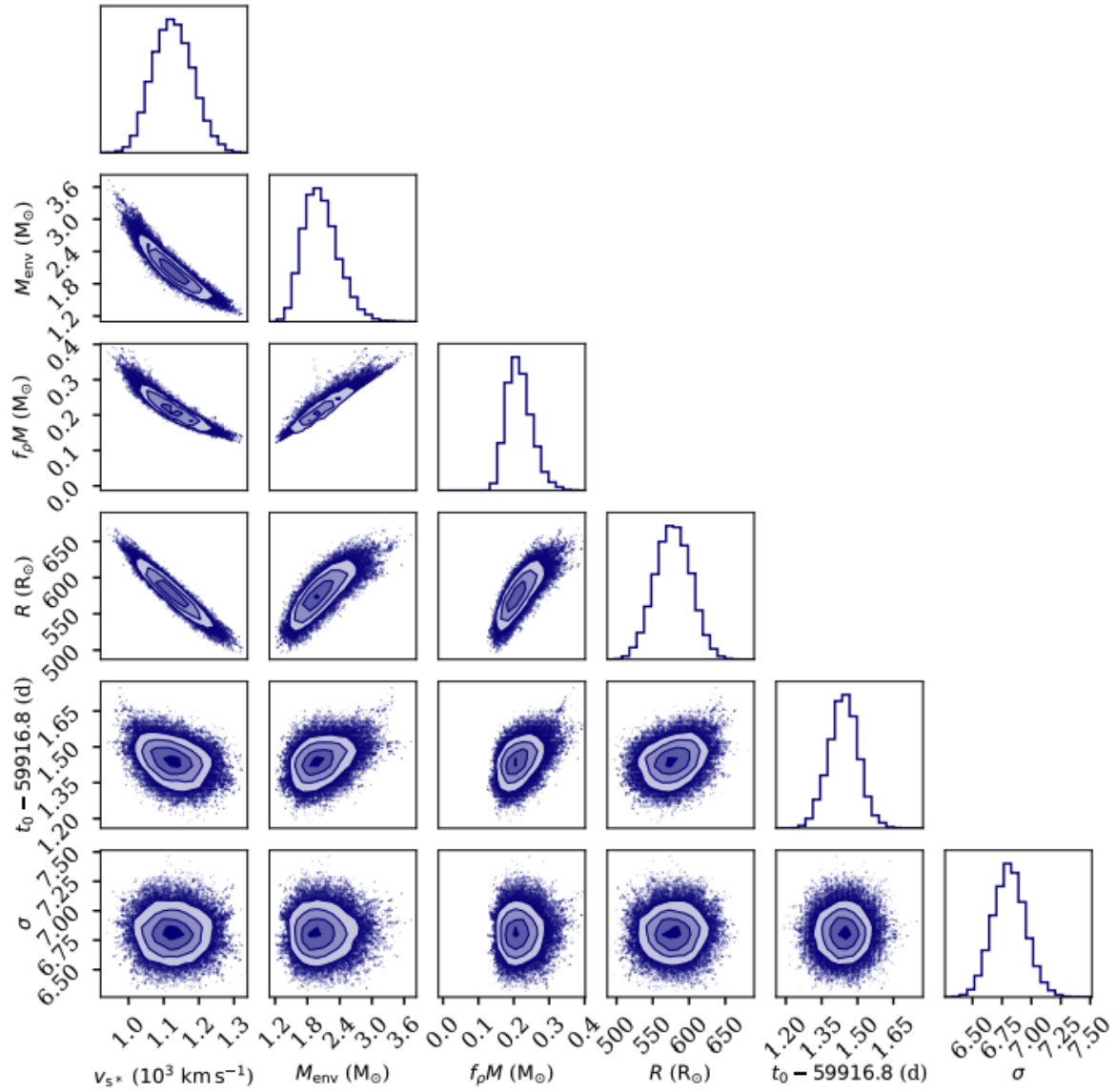

Figure 8.5: Corner plots for the best-fit parameters of SN 2022acko obtained from model MSW23 (left) and model SW17 (right).

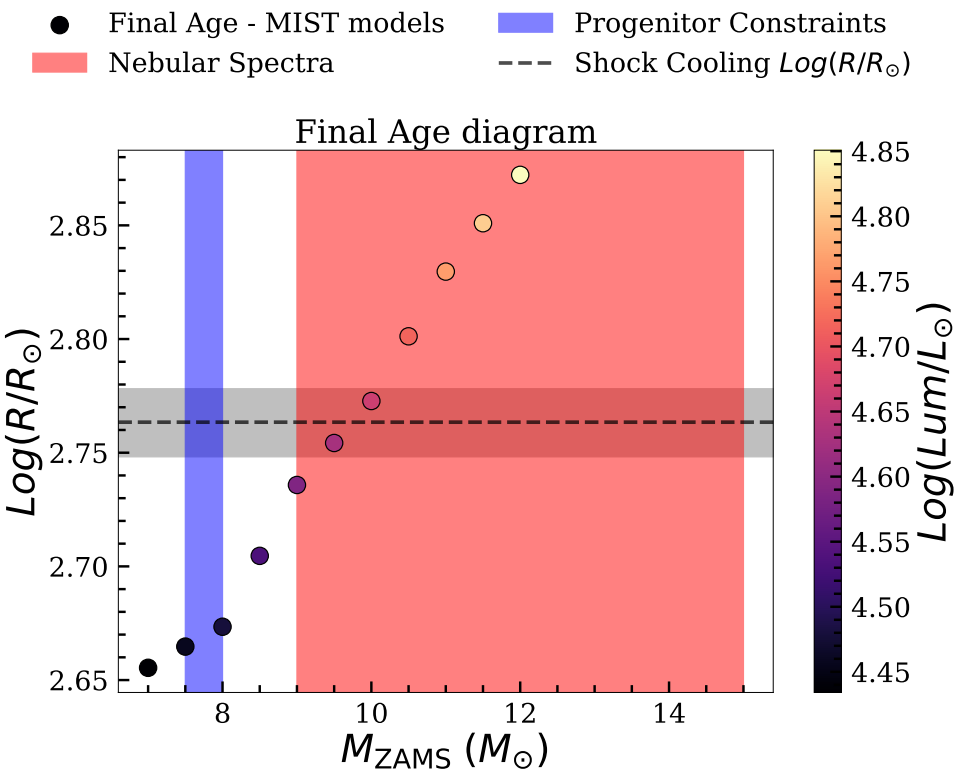

Figure 8.6: The terminal radius and luminosity from stars evolution tracks in function of the stars initial masses

8.5 | Progenitor Mass from Nebular Spectra

The nebular phase of SNe II begins when the ejecta become optically thin (Ferrari et al., 2024; Kilpatrick et al., 2023b; Tinyanont et al., 2021), typically several tens of days after the plateau phase in the case of SNe II-P. During this phase, the spectra are dominated by emission lines from the synthesized elements in the inner ejecta. Among these, the strength of the oxygen lines ([O I] at 6300 and 6364 Å) is particularly sensitive to the oxygen mass in the progenitor core, making it a key diagnostic for estimating the initial mass of the progenitor star (Jerkstrand, 2017).

Modeling these nebular spectra requires accounting for several physical processes, including the nebular composition, velocity fields, non-thermal excitation, and Non-Local Thermodynamic Equilibrium (NLTE) conditions (Jerkstrand, 2017). In this section, we compare synthetic nebular models for SNe II with the observed spectra of SN 2022acko (see Section 8.2.4) to estimate the initial mass of its progenitor.

We used two different sets of nebular models in this analysis. Both sets are based on NLTE radiative transfer and nucleosynthetic yields for CCSNe. The first set encompasses progenitor stars with initial masses of 9 $M_\odot$ (Jerkstrand et al., 2018) and 12-29 $M_\odot$ (Jerkstrand et al., 2014), assuming nebular spectra of different ages. The second set consists of

a grid of models for progenitor stars with initial masses between 9-29 $M_\odot$, corresponding to one-year-old SNe II (Dessart et al., 2021).

All considered models were scaled by the ratio between the total nickel mass derived for SN 2022acko ($M_{\rm Ni} = 0.014 \pm 0.004 M_\odot$) (Teixeira et al., 2026) and the $M_{0,\rm Ni}$ value assumed for generating each of the synthetic spectra. Furthermore, we scaled the modeled spectra to match the distance of SN 2022acko, $d_L = 19.0 \pm 2.9$ Mpc (Bostroem et al., 2023; Anand et al., 2020). Finally, we accounted for the difference between the observation dates of the SN 2022acko spectra and the ages of the SNe used to generate the synthetic spectra, δt, by scaling the fluxes by $\exp(-\lambda_{\rm Co}\delta t)$. At the end, the modeled fluxes were rescaled by the factor s_0, given by

$$s_0 = \frac{M_{\rm Ni}}{M_{\rm Ni,0}} \times \left(\frac{d_{L,0}}{d_L}\right)^2 \times e^{-\lambda_{\rm Co}\delta t}, \tag{8.7}$$

in which $d_{L,0}$ is the luminosity distance assumed for generating the synthetic spectra. The observed spectra were calibrated for the expected flux of the SN at each observation date as implied by the light curves. We used the `pysynphot` package (STScI Development Team, 2013) to generate synthetic photometry from the spectra. Specifically, we computed fluxes in the r and i bands for the 275-day spectrum, in the i band for the 396-day spectrum, and in the z band for the 612-day spectrum. The choice of bands was guided by the availability of photometric measurements closest in time to each spectrum. The resulting fluxes were then scaled so that the synthetic photometry matched the observed values.

The photometry dates, $t_{\rm phot}$, do not exactly match the spectroscopic dates, $t_{\rm spec}$. Therefore, after the rescaling, we further adjusted the flux by multiplying it by the factor $\exp(-\delta t_{cal}\lambda_{\rm Co})$, where $\delta t_{cal} = t_{\rm spec} - t_{\rm phot}$. We emphasize that the photometry used for rescaling was chosen to be as close as possible to $t_{\rm spec}$ for each spectrum. The values of δt_{cal} were approximately -2, +42, and +260 days, corresponding to the spectra dates in ascending order.

Figure 8.7 shows the flux-calibrated observed nebular spectra from Keck, alongside the models of Jerkstrand et al. (2018, 2014) and Dessart et al. (2021). Notably, for the two youngest spectra (panels *a* and *b*), there is a disagreement between the progenitor star mass suggested by the two models that best fit the observations. For the earliest spectra (275 d), the comparison with Jerkstrand et al. (2018, 2014) models suggests that the progenitor star of SN 2022acko is more likely a 12 $M_\odot$ star. On the other hand, there is no significant agreement between the spectra with any of Dessart et al. (2021) models in the oxygen doublet region, except for small intersection with the 9 $M_\odot$ model. However, if we consider that this disagreement may arise from calibration uncertainties and instead focus on comparing the spectral shape in this region, the 10 $M_\odot$ model provides a better match to the observed spectra.

When analyzing the 396 d old spectra, we have also a disagreement, but now the

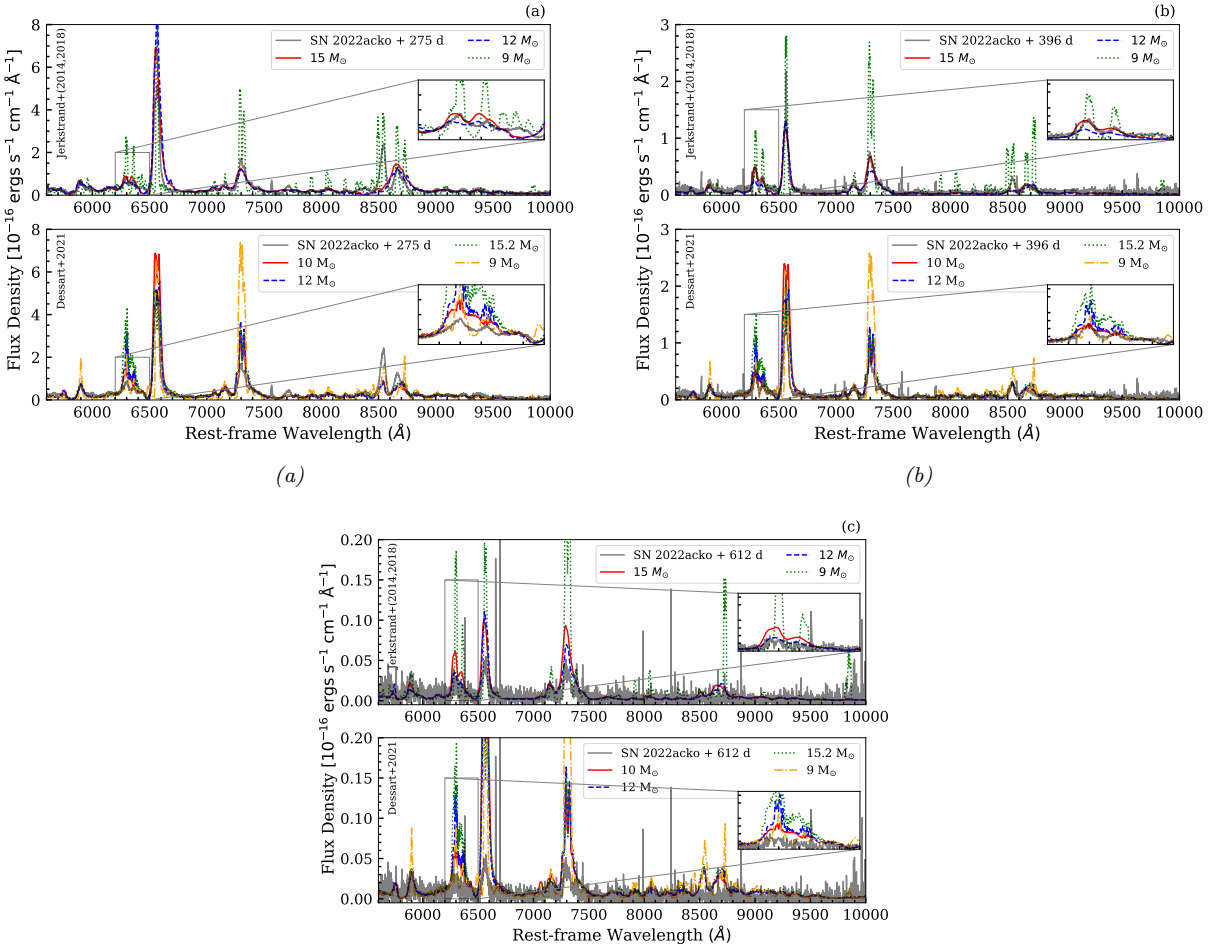

Figure 8.7: SN 2022acko Keck spectra (grey lines) from September 2023 (a), January 2024 (b), and August 2024 (c), at 275, 396, and 612 days from the rest-frame explosion date, respectively. We compare all spectra to nebular models from Jerkstrand et al. (2018) and Jerkstrand et al. (2014) (upper panels), and Dessart et al. (2021) (lower panels). Within each panel, we show an inset zoomed in on the [O I] λλ6300,6364 lines, which are the primary features used to match models of SN II nebular spectra.

best fit for the Jerkstrand et al. (2018, 2014) models corresponds to a 15 $M_{\odot}$progenitor, while the the spectra show better agreement with a 9 $M_{\odot}$ progenitor star when compared to the Dessart et al. (2021) models.

It is worth noting that both models suggest a progenitor star mass of $> 9; M_{\odot}$, trending to higher masses. These results are in disagreement with the estimates of the progenitor star mass derived from the direct observations (Section 8.3).

The analysis of the oldest spectrum presents two main caveats. First, we did not have SN 2022acko photometry available near the epoch of the spectrum ($\delta t_{cal} = 296$d), so the flux calibration may not accurately reflect the true values. Despite this limitation, it is encouraging to note that pronounced emission lines, including those from [O I] λ6303, Hα, and [Ca II] $\lambda\lambda$7922, 7324, are still visible.

The second caveat is that this spectrum corresponds to one of the latest phases ever observed for a SN II-P a regime where the validity of existing nebular models has not been thoroughly tested. Although the spectrum is significantly lower in signal-to-noise than the earlier phases, we detect the expected spectral features that are associated with nebular SN II-P in particular the [O I] feature. However, the nebular models we use only model the evolution of SNe II-P up to 500 days post-explosion in the case of the Jerkstrand et al. (2018, 2014) models, and around 350 days post-explosion for the Dessart et al. (2021) models.

Although this results may be strongly affected by the aforementioned caveats, for the late-time spectrum, the best agreement – in terms of the [OI] doublet – is found with the 12 $M_{\odot}$ progenitor model from Jerkstrand et al. (2018, 2014).

The trustworthiness of the comparison between the nebular spectra and theoretical models is compromised by significant uncertainties in flux calibration and by systematic errors introduced when rescaling the models to account for differences in nickel mass and luminosity distance. These limitations likely contribute substantially to the discrepancies discussed above, once the scaling factor alone introduces a relative uncertainty of approximately 45% .

8.6 | Discussion

The progenitor mass estimates from pre-explosion imaging, shock cooling modeling, and nebular spectroscopy are summarized in Table 8.3. These estimates rely on three different natures of data, and lead to a very broad range of reliable constrains for the progenitor mass.

Using early-phase photometry, we applied the shock cooling emission models from Morag et al. (2023) to constrain the progenitor star's radius to $\sim 580\,R_{\odot}$. By comparing

Table 8.3: Summary of progenitor mass estimates for SN 2022acko.

Analysis	**Mass Estimate ($M_\odot$)**	**Section**
Pre-explosion imaging	$7.5 - 8.0$	8.3
Shock cooling modeling	$9 - 10$	8.4
Nebular spectroscopy	$9 - 15$	8.5

this estimate with MESA evolutionary tracks, we inferred a progenitor mass of $9 - 10\,M_\odot$ for SN 2022acko. These findings are consistent with typical values of progenitor mass for SNe II, and are also consistent with the lower estimates from the nebular spectroscopy modeling.

Although SC modeling results are compatible with a higher mass progenitor, these models involve several approximations, and their systematics are not yet fully understood. In the specific case of SN 2022acko, the SC analysis suggests that the first detection occurred approximately 1.5 days after the estimated date of explosion. Since SC-based estimates of radius – and by extension, mass – depend critically on early-time data, we performed a rough test to estimate the impact of a delayed first detection. For this test, we used SN 2021yja (Hosseinzadeh et al., 2022), a well-observed event with low CSM density and a discovery only ∼5.4 hours post-explosion. Using the MSW23 model, we first fit the SC emission using the initial 9 days of data, as done in Hosseinzadeh et al. (2022). We then repeated the fit, excluding the first 1.5 days to simulate the data gap similar to SN 2022acko. The resulting posteriors (Figure 8.8) indicate that omitting early-time data can lead to a significantly larger (about $\sim 100\,R_\odot$) inferred progenitor radius. While this exercise is approximate, it suggests that the SC-derived radii–and thus masses–for SN 2022acko could be overestimated if early emission is missing.

We also compared the nebular-phase spectra of SN 2022acko with models from Jerkstrand et al. (2018, 2014) and Dessart et al. (2021), and found that none of the three independent spectra yielded consistent progenitor mass estimates across the different models. The variation in outcomes from these visual comparisons results in a poorly constrained progenitor mass range spanning approximately $9 - 15\,M_\odot$, trending towards $\sim 12\,M_\odot$. This highlights a tension between two distinct mass regimes: a higher-mass scenario ($9 - 15\,M_\odot$, higher relative to the very low mass estimate from progenitor observations) supported by shock-cooling modeling and nebular spectroscopy, and a lower-mass scenario (~ 7.5–$8.0\,M_\odot$) derived from direct detection of the progenitor through RSG SED fitting.

A possible explanation for the disagreement in progenitor star mass estimates is significant envelope loss during the lifetime of SN 2022acko progenitor star. In such a scenario, the progenitor could have had a higher initial mass, reflected in the nebular-phase spectra due to the strong correlation between progenitor mass and synthesized oxygen. While the observational features closer to the explosion would appear consistent

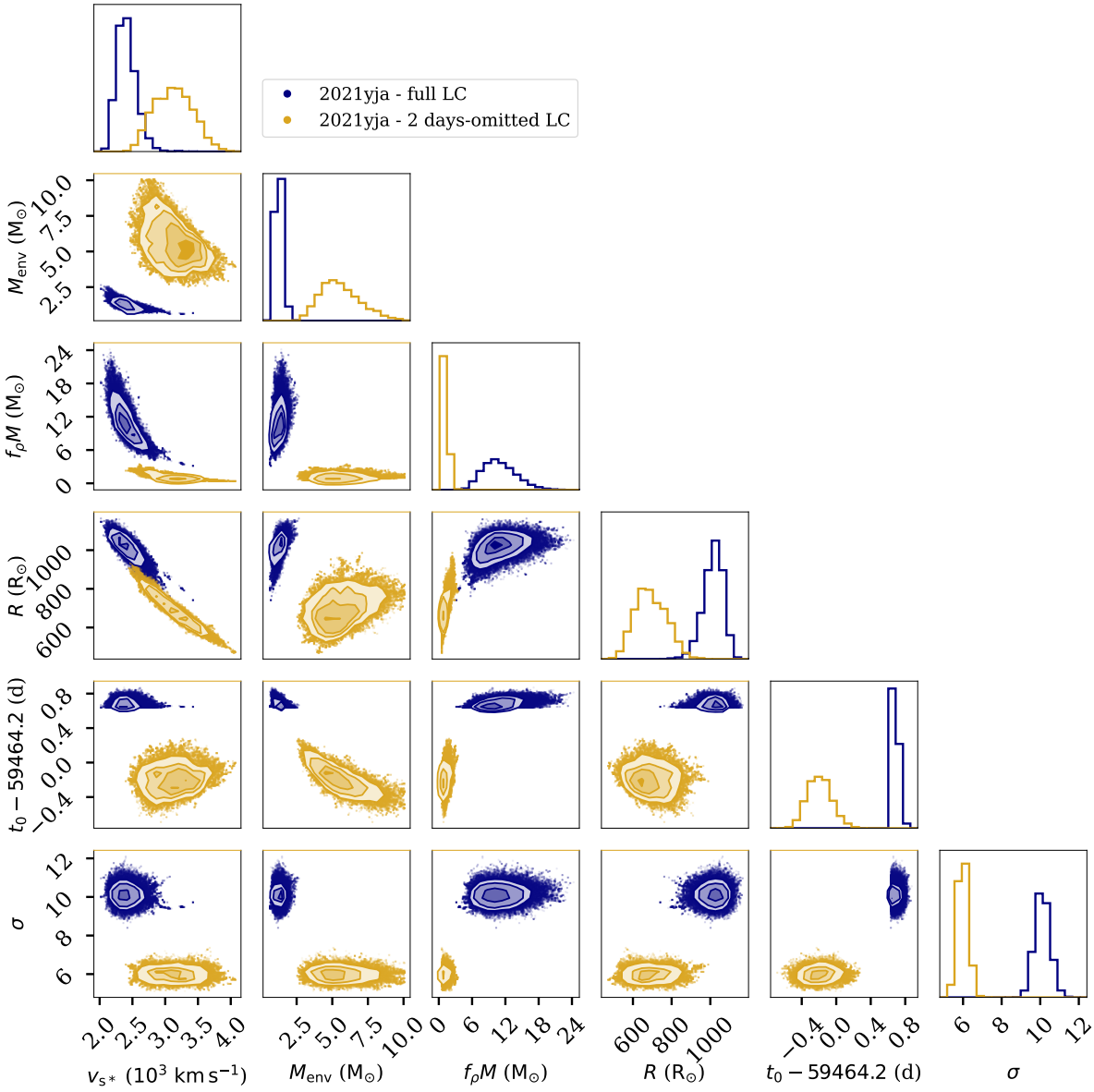

Figure 8.8: Posterior estimation for the Morag et al. (2023) model over the early light curve of SN 2021yja. The blue distributions represent the results using the full multi-band light curve, compatible with the original estimates from Hosseinzadeh et al. (2022). The golden distributions correspond to the estimation where the first day of observation was omitted from the light curve.

with a lower-mass star, after substantial envelope stripping.

This scenario would be further supported if SC modeling tends to overestimate the progenitor radius, since such a close-to-explosion measurement would be consistent with a low progenitor mass. However, to adopt this assumption with confidence, additional studies are needed to better characterize and quantify the systematic uncertainties and biases inherent in SC models.

If this significant mass loss were driven primarily by stellar winds or episodic outbursts (see review in Smith, 2014), we would expect clear evidence of CSM surrounding the progenitor, such as a secondary luminosity peak due to SC UV re-emission in the early light curve (Morag et al., 2023; Waxman and Katz, 2017). Even if CSM were present, however distant from the star, we would anticipate a signature on timescales of years assuming the mass loss occurred in the final 1000 yr prior to core collapse, i.e., during carbon burning or later (Limongi et al., 2024). The shock would heat this material, resulting in an increase in luminosity from thermal radiation. However, apart from fluctuations comparable to the plotted uncertainties, Figure 8.1 shows a nearly steady decay in luminosity for the later phases, which does not indicates the presence of CSM at extended radii from the progenitor star. At this stage, mass transfer to a binary star companion emerges as an alternative explanation (Ragosta et al., 2025), which we explore below.

Binary systems can reach mass transfer rates of $\sim 0.01\,\mathrm{M_\odot\,yr^{-1}}$ (Blagorodnova, Nadejda et al., 2021), potentially explaining the mass loss with absence of strong CSM effects in our observations. While no direct evidence of a binary companion at the site of SN 2022acko is available (see Section 8.3), we explored this hypothesis using BPASS (Byrne et al., 2022) models.

To constrain the progenitor system, we examined BPASS evolutionary models at the metallicity of NGC 1300, [Fe/H] $= -0.22$ (Rosado-Belza, D. et al., 2020). We compared the final predicted magnitudes of these models in the F814W and F160W filters with the progenitor photometry from Table 8.1, using a 3σ tolerance. Among the models meeting the photometric constraints, none with a terminal mass $\gtrsim 7\,\mathrm{M_\odot}$ were found to simultaneously match our progenitor constrains (Table 8.3). This suggests that no reliable high-mass binary evolution model in the BPASS grid adequately reproduces the observed properties of SN 2022acko.

While binary mass transfer remains a plausible mechanism, further evidence–such as late-time imaging or detailed modeling–will be necessary to confirm its role in shaping SN 2022acko's progenitor evolution and explosion properties.

Nevertheless, the main tension arises between the progenitor properties inferred from pre-explosion imaging and those derived from the SN data. It is important to note that, in addition to the substantial systematic uncertainties associated with the flux calibration of the nebular spectra, our progenitor mass estimate from pre-explosion imaging relies on only two photometric data points. Therefore, this apparent discrepancy

may not be as significant as it seems, particularly if the RSG SED fit underestimates the progenitor's luminosity or temperature.

8.7 | Conclusions

In this work we have presented an extended optical and UV light curve for SN 2022acko, covering approximately 350 days. The final ∼200 days of observations represent a novel compilation, including original data from the STEP follow-up program. Our dataset also includes pre-explosion imaging from *HST* at the SN 2022acko site, as well as three epochs of Keck spectroscopy obtained at 275, 396, and 612 days after detection. We used this dataset to derive independent constraints on the progenitor star's properties, with a particular focus on constraining the star's initial mass.

Using *Spitzer* and *HST* pre-explosion imaging of the SN 2022acko site –specifically the F814W and F160W bands from *HST*, along with photometric upper limits from other filters– we fit a RSG SED and found better agreement with models corresponding to a progenitor mass of approximately $7.5\,M_\odot$.

Applying SC emission models from Morag et al. (2023) to the SN early-phase photometry, we constrained a progenitor radius of $\sim 580\,R_\odot$. Assuming a typical terminal effective temperature for RSGs of ∼3500 K, this corresponds to a progenitor luminosity of $\log(L/L_\odot) \approx 4.66$. We compared these estimates with MESA evolutionary tracks, and found this radius consistent with progenitors star with mass within $9 - 10\,M_\odot$.

We compared the nebular-phase spectra of SN 2022acko with models from Jerkstrand et al. (2018, 2014) and Dessart et al. (2021), finding an inconsistency with the low-mass progenitor scenario suggested by the pre-explosion analysis. Instead, the spectral features show better agreement with models corresponding to progenitor masses in the range of $9 - 15\,M_\odot$.

To explain the discrepancy among the progenitor mass estimates, we considered the possibility that SN 2022acko originated from a binary system and had undergone significant mass loss through interaction with a companion. However, upon analyzing binary evolution models from the BPASS database, we found no scenario that simultaneously satisfies both the mass constraints and the pre-explosion *HST* observations at the SN 2022acko site.

The systematic uncertainties – approximately 45% in the flux calibration of the nebular spectra – limit the reliability of the higher progenitor mass estimates, and and help explain why the mass range inferred from this method is so broad. Nevertheless, the lower mass estimates are also subject to notable limitations. In addition to relying on upper limits, the pre-explosion SED fitting is based on only two photometric detections (F814W

and F160W), which might not be sufficient to robustly constrain the full spectral energy distribution. Similarly, the shock-cooling models are based on simplifying assumptions that may not fully reflect the physical conditions of SN 2022acko. Indeed, previous studies employing these models have encountered difficulties in accurately reproducing early-time light curves, especially in the UV regime (Meza-Retamal et al., 2024; Shrestha et al., 2024b; Andrews et al., 2024; Shrestha et al., 2024a). As a result, a tension arises among the progenitor mass estimates. However, the majority of the available evidence (post-explosion data) favors higher progenitor masses, consistent with typical values for SNe II-P.

This work presents a comprehensive analysis encompassing data from pre-explosion imaging, early-time photometry, and late-time nebular spectroscopy. The tension between different progenitor mass estimates appears to stem primarily from the limited pre-explosion data, which may have biased the lower-mass estimates, while the uncertainties from SN post-explosion analyses are likely dominated by intrinsic flux calibration errors and modeling limitations. A more detailed and physically realistic modeling, as well as a deeper investigation on the sources of systematic uncertainties may be required to better interpret the reported constraints.

9 PILCA: Physics-Informed ML Light Curve Analyzer

In Chapter 8 we discuss the importance of characterizing SNe II in terms of the properties of their progenitor stars. Particularly, in Section 8.4 we analyze the early light curve of SN 2022acko to infer its progenitor radius (and consequently its mass) by fitting the analytical models of Morag et al. (2023). We show in the same section that the fit was successful – except for a manual adjustment to the likelihood to preserve the physical meaning of the results, the regression was performed without further complications. This is expected given that, as seen in Figure 8.4, our light curve is very well sampled.

In the LSST era, an unprecedented number of transients will be discovered (some have already been discovered), many identified shortly after explosion. However, unlike SN 2022acko, their characterization is challenging due to sparsely and irregularly sampled light curves and the limited availability of spectroscopic follow-up for distant sources. Therefore, robust analysis tools capable of extracting physical information from incomplete data are essential.

In this chapter we introduce Physics-Informed Machine Learning Light Curve Analyzer (PILCA), a work in progress consisting of a framework that aims to extract physical information of transients by fitting complex analytical models to sparse light curves, with a particular focus on its application to LSST alerts. As it is in a very early stage of development, here we focus on the idea behind the project and present preliminary results.

9.1 | Methodology

The general problem here is straightforward: it consists of fitting an analytical model, such as the one described in Section 8.4, to an observed dataset. Let $\Phi(t, \boldsymbol{\theta})$ denote a physical model describing the light curve of a transient, parameterized by $\boldsymbol{\theta}$. The modeled photometry in filter f at time t is given by

$$\hat{y}_f(t) = \Phi_f(t, \boldsymbol{\theta}). \tag{9.1}$$

Given a dataset $\mathcal{D} = \{t_{\text{obs}}, y_f, f\}$ – note that the measured y_f does not carry a hat – the goal is to infer the parameters $\boldsymbol{\theta}$ that best match the modeled light curves with $\mathcal{D}$. In a Bayesian framework, this corresponds to sampling from the posterior $p(\boldsymbol{\theta} \mid \mathcal{D})$. Traditional

approaches rely on MCMC methods, which often become inefficient or fail when data are sparse or poorly constrained (Gelman et al., 2014). PILCA reformulates this problem by modeling the parameter inference process with a neural network.

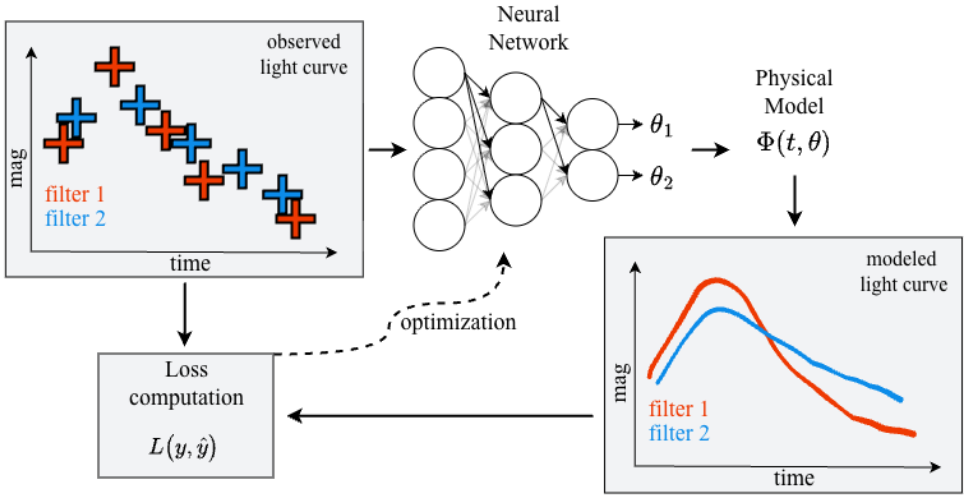

Figure 9.1: Schematic overview of the PILCA framework.

As schematized in Figure 9.1, the network takes the observed light curve as input and outputs an estimate $\hat{\boldsymbol{\theta}}$. The corresponding model prediction

$$\hat{y} = \Phi(t_{\text{obs}}, \hat{\boldsymbol{\theta}}) \tag{9.2}$$

is compared to the observed data through a loss function $L(\hat{y}, y)$. Since the physical model Φ is differentiable, the gradient with respect to the network parameters w explicitly depends on the physical information carried by Φ. Recalling the discussion in Section 2.3.2, the gradient can then be computed as

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \Phi} \frac{\partial \Phi}{\partial \hat{\boldsymbol{\theta}}} \frac{\partial \hat{\boldsymbol{\theta}}}{\partial w}. \tag{9.3}$$

This structure explicitly embeds physical knowledge into the optimization process (Section 2), guiding gradient descent toward physically meaningful parameter estimates.

9.2 | Network

The neural network maps multi-band, irregularly sampled light curves directly onto the physical parameters of the explosion model. As shown in Figure 9.2, photometric observations in different filters are first processed independently through filter-specific sub-networks. This design primarily accommodates the non-uniform and irregular time sampling across filters, allowing each sub-network to model filter-specific cadences and

missing observations before the representations are combined. To account for parameter uncertainties and correlations, the network was chosen to be a BNN (see Section 2.4.3), with selected weights modeled as Gaussian random variables, $w \sim \mathcal{N}(\mu, \sigma)$.

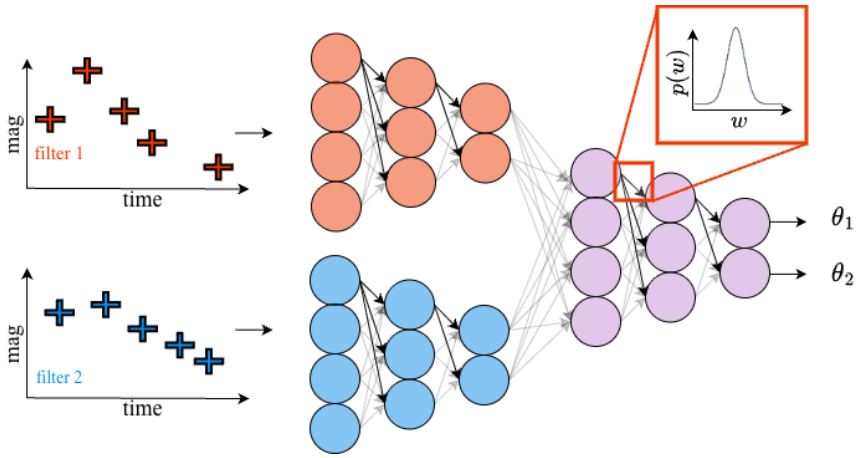

Figure 9.2: Network architecture of PILCA.

9.2.1 | Training

In this initial version of the framework, we choose the physical model to be the shock cooling models presented in Section 8.4. In that sense, our parameters are defined as

$$\hat{\boldsymbol{\theta}} = (R, v_{s*}, M_{\rm env}, f_\rho M, t_0)\,. \tag{9.4}$$

During the training stage, we generate the modeled light curve by passing the observed light curve through the model (with fixed weight distributions) $N = 100$ times, then taking the average of these N samples

$$\langle\hat{\boldsymbol{\theta}}\rangle = (\langle R\rangle, \langle v_{s*}\rangle, \langle M_{\rm env}\rangle, \langle f_\rho M\rangle, \langle t_0\rangle) \tag{9.5}$$

and computing the physical model for the averaged parameters. In that sense, the modeled light curve $\hat{y}$, to be compared with the observed one, is defined as

$$\hat{y} = \Phi(t_{\rm obs}, \langle\hat{\boldsymbol{\theta}}\rangle). \tag{9.6}$$

Here we omit the dependence on the observed filter f for simplicity, though the comparisons are time- and filter-specific.

Regarding the discussion in Section 2.4.3, the appropriate loss function should be the ELBO loss defined in Equation 2.34. However, as this first attempt was conceived as a step zero to verify whether the framework makes sense, we did not enforce strict rigor in

using the most appropriate loss function; instead, we defined the loss as the MSE between the modeled light curve $\hat{y}$ and the observed data y. This implies that the optimization process is not taking into consideration the full probabilistic nature of the parameter distribution, and is only optimizing the network to produce good mean estimates.

Furthermore, this modeling can be thought of as an unsupervised learning process: at no point during training do we have access to the true values of the physical parameters $\boldsymbol{\theta}$. The network learns entirely from the agreement between the modeled and observed light curves, without any labeled supervision on the parameters themselves. In that sense, PILCA is designed to perform the regression transient by transient, fitting an independent model for each object rather than learning a global mapping across a population.

9.3 | Preliminary Experiments

To assess the reliability of the network, we generated synthetic multi-band light curves using fixed physical parameters corresponding to those inferred for SN 2022acko within the shock-cooling model:

$$R = 580\,R_\odot, \quad v_{s*} = 1.12 \times 10^3\,\mathrm{km\,s^{-1}}, \quad M_{\mathrm{env}} = 2.0\,M_\odot,$$
$$f_\rho M = 0.21\,M_\odot, \quad t_0 = 3\ \mathrm{d}.$$

A random noise of $\mathcal{N}(0, 0.1\,\mathrm{mag})$ was added to the synthetic photometry. We restricted the analysis to the *ugrizy* filter set and adopted a cadence of ~ 1.4 observations per band per day, which remains optimistic relative to the expected LSST alert cadence. Training was performed until convergence of the loss function.

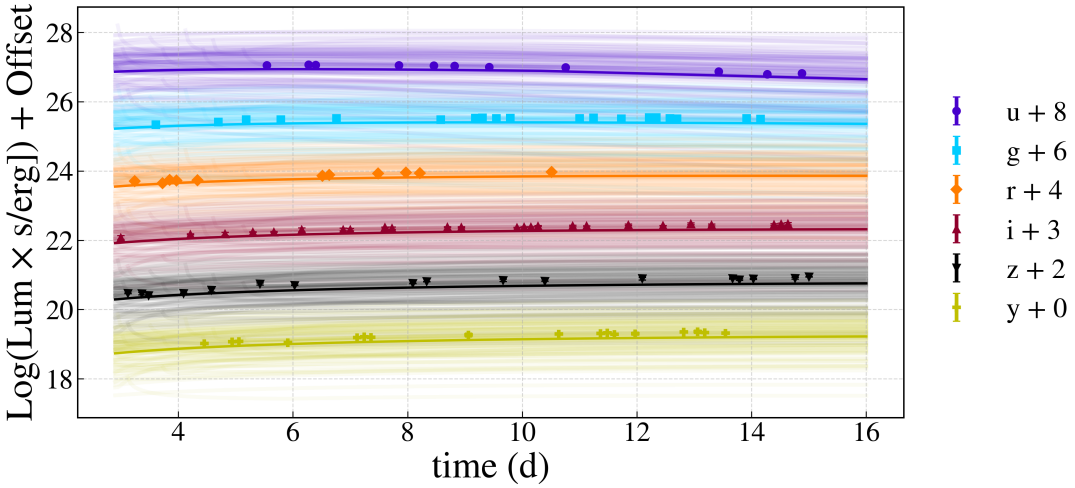

Figure 9.3: Synthetic multi-band light curves in the ugrizy filter set generated with the physical parameters of SN 2022acko (points), compared to the model light curves reconstructed from 1000 samples of $\hat{\boldsymbol{\theta}}$ (semi-transparent lines). An offset of 2 was added to each filter for a better visualization.

After training, we sampled $\hat{\boldsymbol{\theta}}$ 1000 times and converted the outputs from model units to physical units. Figure 9.3 shows the resulting model light curves compared to the simulated data, and Figure 9.4 presents the corresponding parameter distributions. The recovered parameters are

$$R \approx 495.96^{+254.36}_{-256.54}\, R_\odot, \quad v_{s*} \approx 1.57^{+1.16}_{-1.01} \times 10^3\,\mathrm{km\,s^{-1}}, \quad M_{\mathrm{env}} \approx 1.81^{+1.07}_{-1.09}\, M_\odot,$$
$$f_\rho M \approx 0.03^{+0.02}_{-0.03}\, M_\odot, \quad t_0 \approx 1.69^{+0.86}_{-0.93}\ \mathrm{d}.$$

PILCA recovers four out of five parameters with distributions that encompass the true values. The progenitor radius and envelope mass show moderate deviations (-14% and -9%, respectively), while v_{s*} and t_0 present larger mean offsets ($+40\%$ and -44%), though in all four cases the true values fall within the recovered uncertainties. The density parameter $f_\rho M$, however, is significantly underestimated (-86%), with the true value $0.21\,M_\odot$ lying well outside the recovered distribution $0.03^{+0.02}_{-0.03}\,M_\odot$. As discussed in Section 8.4, $f_\rho M$ presents intrinsic estimation difficulties even with MCMC on well-sampled light curves, and its degeneracy with M_{env} in shock-cooling models makes it the most challenging parameter to constrain.

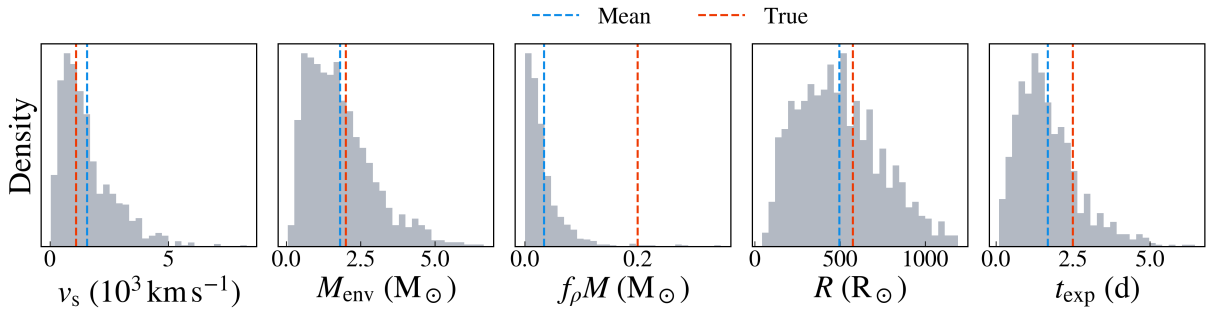

Figure 9.4: Density estimations of the shock-cooling physical parameters recovered by PILCA from 1000 samples of $\hat{\boldsymbol{\theta}}$. Vertical dashed lines indicate the true (red) input values and the predicted distribution mean (blue).

Beyond the mean deviations, the recovered distributions are broad and – th the exception of R and t_0 – non-symmetric, with no clear physical motivation for such asymmetries. These features are, however, fully consistent with the current formulation: since the MSE loss constrains only mean predictions, correlations and covariances between parameters are not captured, and the model is effectively learning the optimal mean solution. Properly recovering physically meaningful posterior distributions will require incorporating the full ELBO loss, enabling the BNN to model the posterior covariance structure. Even then, fully reproducing all correlations observed in Figure 8.5 is not guaranteed given the intrinsic complexity of the problem. Addressing these limitations remains a key direction for future work.

9.4 | Conclusions

Although still in its early stages, PILCA has the potential to become a robust ML framework for transient characterization. A natural question arises as to why MDNs were not adopted here, as in the photo-z case. The answer is that multi-dimensional MDNs operate at a substantially higher level of complexity, as they must model joint distributions over several correlated physical parameters simultaneously. As this represents one of the first implementations of the PILCA framework, we opted for the simplest viable approach – a BNN trained with an MSE loss – to establish an initial density estimator and verify that the framework produces physically meaningful results before introducing additional complexity.

Future developments will focus on incorporating the full ELBO loss to enable proper posterior estimation, and on exploring MDNs or simulation-based inference (SBI; Cranmer et al., 2020) approaches using large synthetic datasets. These approaches will allow us to generalize physical-parameter estimation to more complex and realistic supernova environments under an LSST-like cadence. If successful, the same methodology can naturally be extended to other transient classes and physical models, provided they remain differentiable.

Despite the limitations discussed above, the PILCA framework is fully automated, requiring no tuning of initial conditions as in the case of traditional MCMC algorithms, and thus serving as a valuable tool for automated Bayesian inference. In particular, it can be used to provide informed initial conditions for traditional sampling methods, as it already produces physically motivated parameter distributions – although these cannot yet be interpreted as true posterior distributions.

One of PILCA's next key steps is to be systematically compared with other gradient-based density estimators, such as Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Betancourt, 2018) and more advanced non-vanilla sampling techniques. Our main interest is to compare PILCA and HMC, as the two methods are conceptually related yet fundamentally different. HMC explores the full posterior distribution by defining a likelihood-based energy function that governs a pseudo-particle motion in parameter space. By analogy, PILCA instead learns to move this particle toward the mode that maximizes the likelihood, making it an efficient but non-posterior estimator.

In addition, we plan to integrate this framework into the Fink broker, enabling users to perform transient model fitting online in real time. The code used for both the framework and the analyses presented in this thesis is publicly available at `https://github.com/gsmteixeira` [1].

[1] This is still a work in progress, so I discourage high expectations regarding code design.

Final Remarks

In this thesis we developed a set of contributions encompassing AI-based methodologies, their astrophysical applications, and dedicated astrophysical analyses. Among them, we released one of the most extensive photometric redshift catalogs, which later proved to be among the most accurate, outperforming the official results of LS-DR10 in the low-z regime. These redshift estimates had a significant impact on dark siren methods for estimating the Hubble constant, a particularly relevant application in the context of the current Hubble tension. They also enabled members of the CHANCES collaboration to recover large-scale structure properties directly from a photo-z-based catalog. In addition, we introduced an innovative approach to photometric redshift estimation using non-traditional neural networks, specifically through the LMU framework, as well as a novel strategy for processing magnitudes and colors in photo-z inference (as implemented in the CHANCES model).

Beyond analyses directly driven by AI, this work includes a comprehensive study of a SN II-P, the SN 2022acko, in the context of the S-PLUS transient extension program. The analysis contributed with estimates of the progenitor mass associated with this SN, representing a relevant step toward shedding light on unresolved red supergiant problem. This study also motivated further work in time-domain astronomy, particularly in light curve analysis for the Rubin era, leading to the development of a novel framework for characterizing transients in sparse alert light curves, PILCA. This framework is still in its early stages and has not yet been systematically compared with standard methods in the literature. If successful, it has strong potential to support automated decision-making in the large-scale transient data streams from Rubin. Nevertheless, it is already established as a automatic density estimator for shock-cooling models parameters.

Future work will focus on further developing PILCA and expanding its applications in transient astronomy, such as implementing analytical models for other types of transients. I also intend to publicly release all photometric redshift products from this thesis, along with a more detailed assessment of systematic effects and possible mitigation strategies. In addition, I plan to investigate the impact of separating color and magnitude information into independent networks on model performance, which has not yet been explored in the literature.

Overall, this thesis illustrates several examples of the impact of AI in astrophysics, with contributions spanning *From Large-Scale Structure to Time-Domain Astronomy*. It also serves as a reminder that we still have much to explore regarding the potential of AI techniques in this field.

Bibliography

Abadi, M. et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", , 2015. Software available from tensorflow.org, URL https://www.tensorflow.org/.

Abbott, B. P. et al. (LIGO Scientific Collaboration and Virgo Collaboration), "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral", *Phys. Rev. Lett.*, vol. 119, p. 161101, 2017a. URL http://dx.doi.org/10.1103/PhysRevLett.119.161101.

Abbott, B. P. et al., "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals", *Classical and Quantum Gravity*, vol. 37, no. 5, p. 055002, 2020. URL http://dx.doi.org/10.1088/1361-6382/ab685e.

Abbott, B. P. et al., "A gravitational-wave standard siren measurement of the Hubble constant", *Nature*, vol. 551, no. 7678, pp. 85–88, 2017b, ISSN 1476-4687. URL http://dx.doi.org/10.1038/nature24471.

Abbott, R. et al., "Constraints on the Cosmic Expansion History from GWTC–3", *The Astrophysical Journal*, vol. 949, no. 2, p. 76, 2023. URL http://dx.doi.org/10.3847/1538-4357/ac74bb.

Abbott, T. M. C. et al., "The Dark Energy Survey: Data Release 1", ApJS, vol. 239, no. 2, 18, 2018. 1801.03181, URL http://dx.doi.org/10.3847/1538-4365/aae9f0.

Adame, A. et al., "DESI 2024 III: baryon acoustic oscillations from galaxies and quasars", *Journal of Cosmology and Astroparticle Physics*, vol. 2025, no. 04, p. 012, 2025a. URL http://dx.doi.org/10.1088/1475-7516/2025/04/012.

Adame, A. et al., "DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations", *Journal of Cosmology and Astroparticle Physics*, vol. 2025, no. 02, p. 021, 2025b. URL http://dx.doi.org/10.1088/1475-7516/2025/02/021.

Alfradique, V. et al., "Improved constraint on the Hubble constant from dark sirens with LIGO/Virgo/KAGRA O4a", , 2026. 2603.20195, URL https://arxiv.org/abs/2603.20195.

Alfradique, V. et al., "A dark siren measurement of the Hubble constant using gravitational wave events from the first three LIGO/Virgo observing runs and DELVE", *Monthly Notices of the Royal Astronomical Society*, vol. 528, no. 2, pp. 3249–3259, 2024, ISSN 0035-8711. URL http://dx.doi.org/10.1093/mnras/stae086.

Almosallam, I. A. et al., "A sparse Gaussian process framework for photometric redshift estimation", *Monthly Notices of the Royal Astronomical Society*, vol. 455, no. 3, pp. 2387–

2401, 2015, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/455/3/2387/9380211/stv2425.pdf, URL http://dx.doi.org/10.1093/mnras/stv2425.

Anand, G. S. et al., "Distances to PHANGS galaxies: New tip of the red giant branch measurements and adopted distances", *Monthly Notices of the Royal Astronomical Society*, vol. 501, no. 3, pp. 3621–3639, 2020, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/501/3/3621/35699965/staa3668.pdf, URL http://dx.doi.org/10.1093/mnras/staa3668.

Anbajagane, D. et al., "The DECADE cosmic shear project I: A new weak lensing shape catalog of 107 million galaxies", *The Open Journal of Astrophysics*, vol. 8, 2025. URL http://dx.doi.org/10.33232/001c.146158.

Andrews, J. E. et al., "SN 2022jox: An Extraordinarily Ordinary Type II SN with Flash Spectroscopy", ApJ, vol. 965, no. 1, 85, 2024. 2310.16092, URL http://dx.doi.org/10.3847/1538-4357/ad2a49.

Arcavi, I., "Hydrogen-Rich Core-Collapse Supernovae", in: *Handbook of Supernovae*, edited by A. W. Alsabti and P. Murdin, p. 239 (Springer International Publishing, 2017). URL http://dx.doi.org/10.1007/978-3-319-21846-5_39.

Arjovsky, M. et al., "Unitary Evolution Recurrent Neural Networks", in: *Proceedings of The 33rd International Conference on Machine Learning*, edited by M. F. Balcan and K. Q. Weinberger, *Proceedings of Machine Learning Research*, vol. 48, pp. 1120–1128 (PMLR, New York, New York, USA, 2016). URL https://proceedings.mlr.press/v48/arjovsky16.html.

Arnouts, S. et al., "LePHARE: Photometric Analysis for Redshift Estimate", Astrophysics Source Code Library, record ascl:1108.009, 2011. 1108.009.

Auchettl, K. et al., "Measurement of the Core-collapse Progenitor Mass Distribution of the Small Magellanic Cloud", *The Astrophysical Journal*, vol. 871, no. 1, p. 64, 2019. URL http://dx.doi.org/10.3847/1538-4357/aaf395.

Baier-Soto, R. et al., "The role of supercluster filaments in shaping galaxy clusters", *A&A*, vol. 704, p. A228, 2025. URL http://dx.doi.org/10.1051/0004-6361/202556957.

Barbon, R. et al., "Photometric properties of type II supernovae.", A&A, vol. 72, pp. 287–292, 1979.

Bassett, B. et al., "Baryon acoustic oscillations", in: *Dark Energy: Observational and Theoretical Approaches*, edited by P. Ruiz-Lapuente, p. 246 (Cambridge University Press, 2010).

Beasor, E. R. et al., "The Red Supergiant Progenitor Luminosity Problem", *The Astrophysical Journal*, vol. 979, no. 2, p. 117, 2025. URL http://dx.doi.org/10.3847/1538-4357/ad8f3f.

Beck, R. et al., "Photometric redshifts for the SDSS Data Release 12", MNRAS, vol. 460, no. 2, pp. 1371–1381, 2016. 1603.09708, URL http://dx.doi.org/10.1093/mnras/stw1009.

Beckwith, S. V. et al., "The Hubble ultra deep field", *The Astronomical Journal*, vol. 132, no. 5, p. 1729, 2006.

Bengio, Y. et al., "Greedy Layer-Wise Training of Deep Networks", in: *Advances in Neural Information Processing Systems*, edited by B. Schölkopf, J. Platt and T. Hoffman, vol. 19 (MIT Press, 2006). URL https://proceedings.neurips.cc/paper_files/paper/2006/file/5da713a690c067105aeb2fae32403405-Paper.pdf.

Benítez, N., "Bayesian Photometric Redshift Estimation", ApJ, vol. 536, no. 2, pp. 571–583, 2000. astro-ph/9811189, URL http://dx.doi.org/10.1086/308947.

Bertin, E., "SWarp: Resampling and Co-adding FITS Images Together", Astrophysics Source Code Library, record ascl:1010.068, 2010.

Bertin, E., "Automated Morphometry with SExtractor and PSFEx", in: *Astronomical Data Analysis Software and Systems XX*, edited by I. N. Evans, A. Accomazzi, D. J. Mink and A. H. Rots, *Astronomical Society of the Pacific Conference Series*, vol. 442, p. 435 (2011).

Bertin, E. et al., "SExtractor: Software for source extraction.", A&AS, vol. 117, pp. 393–404, 1996. URL http://dx.doi.org/10.1051/aas:1996164.

Betancourt, M., "A Conceptual Introduction to Hamiltonian Monte Carlo", , 2018. 1701.02434, URL https://arxiv.org/abs/1701.02434.

Bishop, C., "Mixture density networks", *Aston University*, p. 26, 1994. URL https://research.aston.ac.uk/en/publications/mixture-density-net.

Blagorodnova, Nadejda et al., "The luminous red nova AT 2018bwo in NGC 45 and its binary yellow supergiant progenitor", *A&A*, vol. 653, p. A134, 2021. URL http://dx.doi.org/10.1051/0004-6361/202140525.

Blanchet, L., "Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries", *Living Reviews in Relativity*, vol. 5, no. 1, p. 3, 2002, ISSN 1433-8351. URL http://dx.doi.org/10.12942/lrr-2002-3.

Blei, D. M. et al., "Variational Inference: A Review for Statisticians", *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017. https://doi.org/10.1080/01621459.2017.1285773, URL http://dx.doi.org/10.1080/01621459.2017.1285773.

Bolton, A. S. et al., "SPECTRAL CLASSIFICATION AND REDSHIFT MEASUREMENT FOR THE SDSS-III BARYON OSCILLATION SPECTROSCOPIC SURVEY", *The Astronomical Journal*, vol. 144, no. 5, p. 144, 2012. URL http://dx.doi.org/10.1088/0004-6256/144/5/144.

Bom, C. R. et al., "An extended catalogue of galaxy morphology using deep learning in southern photometric local universe survey data release 3", *Monthly Notices of the Royal Astronomical Society*, vol. 528, no. 3, pp. 4188–4208, 2023, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/528/3/4188/56629312/stad3956.pdf, URL http://dx.doi.org/10.1093/mnras/stad3956.

Bom, C. R. et al., "A dark standard siren measurement of the Hubble constant following LIGO/Virgo/KAGRA O4a and previous runs", *Monthly Notices of the Royal Astronomical Society*, vol. 535, no. 1, pp. 961–975, 2024, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/535/1/961/60414703/stae2390.pdf, URL http://dx.doi.org/10.1093/mnras/stae2390.

Bond, J. R. et al., "How filaments of galaxies are woven into the cosmic web", Nature, vol. 380, no. 6575, pp. 603–606, 1996. astro-ph/9512141, URL http://dx.doi.org/10.1038/380603a0.

Bostroem, K. A. et al., "SN 2022acko: The First Early Far-ultraviolet Spectra of a Type IIP Supernova", *The Astrophysical Journal Letters*, vol. 953, no. 2, p. L18, 2023, ISSN 2041-8213. URL http://dx.doi.org/10.3847/2041-8213/ace31c.

Brammer, G. B. et al., "EAZY: A Fast, Public Photometric Redshift Code", *The Astrophysical Journal*, vol. 686, no. 2, pp. 1503–1513, 2008. URL http://dx.doi.org/10.1086/591786.

Branch, D. et al., "Type IA Supernovae as Standard Candles", ApJ, vol. 405, p. L5, 1993. URL http://dx.doi.org/10.1086/186752.

Bravo-Alfaro, H. et al., "Galaxy evolution in Abell 85 - I. Cluster substructure and environmental effects on the blue galaxy population", *A&A*, vol. 495, no. 2, pp. 379–387, 2009. URL http://dx.doi.org/10.1051/0004-6361:200810731.

Breiman, L., "Random Forests.", *Machine Learning*, vol. 45, pp. 5–32, 2001. URL http://dx.doi.org/10.1023/A:1010933404324.

Breiman, L. et al., *Classification and Regression Trees* (Chapman and Hall/CRC, 1984), 1st ed. URL http://dx.doi.org/10.1201/9781315139470.

Breiman, L. et al., "Submodel Selection and Evaluation in Regression: The X-Random Case", *International Statistical Review*, vol. 60, no. 3, pp. 291–319, 1992. URL http://dx.doi.org/10.2307/1403680.

Brown, T. M. et al., "Las Cumbres Observatory Global Telescope Network", *Publications of the Astronomical Society of the Pacific*, vol. 125, no. 931, p. 1031, 2013. URL http://dx.doi.org/10.1086/673168.

Burke, D. L. et al., "Forward Global Photometric Calibration of the Dark Energy Survey", *The Astronomical Journal*, vol. 155, no. 1, p. 41, 2017. URL http://dx.doi.org/10.3847/1538-3881/aa9f22.

Burrows, A. et al., "On the Nature of Core-Collapse Supernova Explosions", ApJ, vol. 450, p. 830, 1995. astro-ph/9506061, URL http://dx.doi.org/10.1086/176188.

Byrne, C. M. et al., "The dependence of theoretical synthetic spectra on α-enhancement in young, binary stellar populations", *Monthly Notices of the Royal Astronomical Society*, vol. 512, no. 4, pp. 5329–5338, 2022, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/512/4/5329/43377969/stac807.pdf, URL http://dx.doi.org/10.1093/mnras/stac807.

Carliles, S. et al., "Photometric Redshift Estimation on SDSS Data Using Random Forests", in: *Astronomical Data Analysis Software and Systems XVII*, edited by R. W. Argyle, P. S. Bunclark and J. R. Lewis, *Astronomical Society of the Pacific Conference Series*, vol. 394, p. 521 (2008). 0711.2477, URL http://dx.doi.org/10.48550/arXiv.0711.2477.

Carliles, S. et al., "RANDOM FORESTS FOR PHOTOMETRIC REDSHIFTS", *The Astrophysical Journal*, vol. 712, no. 1, p. 511, 2010. URL http://dx.doi.org/10.1088/0004-637X/712/1/511.

Carroll, B. W. et al., *An introduction to modern astrophysics, Second Edition* (Cambridge University Press, 2017).

Chan, K. C. et al. (DES Collaboration), "Dark Energy Survey Year 3 results: Measurement of the baryon acoustic oscillations with three-dimensional clustering", *Phys. Rev. D*, vol. 106, p. 123502, 2022. URL http://dx.doi.org/10.1103/PhysRevD.106.123502.

Chang, C. et al., "The effective number density of galaxies for weak lensing measurements in the LSST project", *Monthly Notices of the Royal Astronomical Society*, vol. 434, no. 3, pp. 2121–2135, 2013, ISSN 0035-8711. https://academic.oup.com/mnras/

article-pdf/434/3/2121/18496619/stt1156.pdf, URL http://dx.doi.org/10.1093/mnras/stt1156.

Chen, H.-Y. et al., "A two per cent Hubble constant measurement from standard sirens within five years", Nature, vol. 562, no. 7728, pp. 545–547, 2018. 1712.06531, URL http://dx.doi.org/10.1038/s41586-018-0606-0.

Childress, M. J. et al., "Measuring nickel masses in Type Ia supernovae using cobalt emission in nebular phase spectra", *Monthly Notices of the Royal Astronomical Society*, vol. 454, no. 4, pp. 3816–3842, 2015, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/454/4/3816/18508278/stv2173.pdf, URL http://dx.doi.org/10.1093/mnras/stv2173.

Choi, J. et al., "Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models", ApJ, vol. 823, no. 2, 102, 2016. 1604.08592, URL http://dx.doi.org/10.3847/0004-637X/823/2/102.

Coe, D. et al., "Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology", *The Astronomical Journal*, vol. 132, no. 2, pp. 926–959, 2006. URL http://dx.doi.org/10.1086/505530.

Colless, M. et al., "The 2dF Galaxy Redshift Survey: spectra and redshifts", MNRAS, vol. 328, no. 4, pp. 1039–1063, 2001. astro-ph/0106498, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04902.x.

Cover, T. M. et al., *Entropy, Relative Entropy, and Mutual Information*, chap. 2, pp. 13–55 (John Wiley & Sons, Ltd, Hoboken, NJ, USA, 2005), ISBN 9780471748823. URL http://dx.doi.org/10.1002/047174882X.ch2.

Cranmer, K. et al., "The frontier of simulation-based inference", *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30055–30062, 2020. https://www.pnas.org/doi/pdf/10.1073/pnas.1912789117, URL http://dx.doi.org/10.1073/pnas.1912789117.

Crenshaw, J. F. et al., "Learning Spectral Templates for Photometric Redshift Estimation from Broadband Photometry", *The Astronomical Journal*, vol. 160, no. 4, p. 191, 2020. URL http://dx.doi.org/10.3847/1538-3881/abb0e2.

Dahlen, T. et al., "A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation", ApJ, vol. 775, no. 2, 93, 2013. 1308.5353, URL http://dx.doi.org/10.1088/0004-637X/775/2/93.

Dalmasso, N. et al., "Conditional density estimation tools in python and R with applications to photometric redshifts and likelihood-free cosmological inference", *Astronomy and Computing*, vol. 30, p. 100362, 2020, ISSN 2213-1337. URL http://dx.doi.org/10.1016/j.ascom.2019.100362.

Davies, B. et al., "The initial masses of the red supergiant progenitors to Type II supernovae", MNRAS, vol. 474, no. 2, pp. 2116–2128, 2018. 1709.06116, URL http://dx.doi.org/10.1093/mnras/stx2734.

Davies, B. et al., "The 'red supergiant problem': the upper luminosity boundary of Type II supernova progenitors", *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 1, pp. 468–476, 2020, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/493/1/468/32513078/staa174.pdf, URL http://dx.doi.org/10.1093/mnras/staa174.

Davies, B. et al., "The luminosities of cool supergiants in the Magellanic Clouds, and the Humphreys–Davidson limit revisited", *Monthly Notices of the Royal Astronomical Society*, vol. 478, no. 3, pp. 3138–3148, 2018, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/478/3/3138/25331202/sty1302.pdf, URL http://dx.doi.org/10.1093/mnras/sty1302.

Davis, M. et al., "A survey of galaxy redshifts. II. The large scale space distribution.", ApJ, vol. 253, pp. 423–445, 1982. URL http://dx.doi.org/10.1086/159646.

Dawson, K. S. et al., "THE BARYON OSCILLATION SPECTROSCOPIC SURVEY OF SDSS-III", *The Astronomical Journal*, vol. 145, no. 1, p. 10, 2012. URL http://dx.doi.org/10.1088/0004-6256/145/1/10.

Desai, S. et al., "The Blanco Cosmology Survey: Data Acquisition, Processing, Calibration, Quality Diagnostics, and Data Release", ApJ, vol. 757, no. 1, 83, 2012. 1204.1210, URL http://dx.doi.org/10.1088/0004-637X/757/1/83.

DESI Collaboration et al., "The Early Data Release of the Dark Energy Spectroscopic Instrument", *arXiv e-prints*, arXiv:2306.06308, 2023. 2306.06308, URL http://dx.doi.org/10.48550/arXiv.2306.06308.

Desprez, G. et al., "Euclid preparation-X. The Euclid photometric-redshift challenge", *Astronomy & Astrophysics*, vol. 644, p. A31, 2020.

Dessart, L. et al., "The explosion of 9-29 $M_\odot$ stars as Type II supernovae: Results from radiative-transfer modeling at one year after explosion", A&A, vol. 652, A64, 2021. 2105.13029, URL http://dx.doi.org/10.1051/0004-6361/202140839.

Dey, B. et al., "Calibrated Predictive Distributions for Photometric Redshifts", *Machine Learning for Astrophysics, proceedings of the Thirty-ninth International Conference on Machine Learning (ICML 2022)*, 2022. URL https://par.nsf.gov/biblio/10457174.

Dolphin, A., "DOLPHOT: Stellar photometry", Astrophysics Source Code Library, record ascl:1608.013, 2016.

Drlica-Wagner, A. et al., "Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology", ApJS, vol. 235, no. 2, 33, 2018. 1708.01531, URL http://dx.doi.org/10.3847/1538-4365/aab4f5.

Drlica-Wagner, A. et al., "The DECam Local Volume Exploration Survey: Overview and First Data Release", *The Astrophysical Journal Supplement Series*, vol. 256, no. 1, p. 2, 2021. URL http://dx.doi.org/10.3847/1538-4365/ac079d.

Drlica-Wagner, A. et al., "The DECam Local Volume Exploration Survey Data Release 2", ApJS, vol. 261, no. 2, 38, 2022. 2203.16565, URL http://dx.doi.org/10.3847/1538-4365/ac78eb.

Duane, S. et al., "Hybrid Monte Carlo", *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987, ISSN 0370-2693. URL http://dx.doi.org/https://doi.org/10.1016/0370-2693(87)91197-X.

Dulcien, C. et al., "Caught in the web: Galaxy mergers along cosmic filaments", *A&A*, vol. 708, p. L5, 2026. URL http://dx.doi.org/10.1051/0004-6361/202558037.

Duncan, K. J., "All-purpose, all-sky photometric redshifts for the Legacy Imaging Surveys Data Release 8", *Monthly Notices of the Royal Astronomical Society*, vol. 512, no. 3, p. 3662–3683, 2022, ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stac608.

Durrer, R., "The cosmic microwave background: the history of its experimental investigation and its significance for cosmology", *Classical and Quantum Gravity*, vol. 32, no. 12, p. 124007, 2015. URL http://dx.doi.org/10.1088/0264-9381/32/12/124007.

Dyk, S. D. V., "The direct identification of core-collapse supernova progenitors", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, no. 20160277, 2017. URL http://dx.doi.org/10.1098/rsta.2016.0277.

D'Isanto, A. et al., "Photometric redshift estimation via deep learning - Generalized and pre-classification-less, image based, fully probabilistic redshifts", *A&A*, vol. 609, p. A111, 2018. URL http://dx.doi.org/10.1051/0004-6361/201731326.

Efron, B. et al., *An Introduction to the Bootstrap* (Chapman and Hall/CRC, 1994), 1 ed. URL http://dx.doi.org/10.1201/9780429246593.

Einstein, A., "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie", *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp. 142–152, 1917.

Eisenstein, D. J. et al., "SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems", AJ, vol. 142, no. 3, 72, 2011. 1101.1529, URL http://dx.doi.org/10.1088/0004-6256/142/3/72.

Eldridge, J. et al., "Supernova lightCURVE POPulation Synthesis II: Validation against supernovae with an observed progenitor", *Publications of the Astronomical Society of Australia*, vol. 36, p. e041, 2019.

Eldridge, J. J. et al., "Binary Population and Spectral Synthesis Version 2.1: Construction, Observational Verification, and New Results", PASA, vol. 34, e058, 2017. 1710.02154, URL http://dx.doi.org/10.1017/pasa.2017.51.

Elman, J. L., "Finding structure in time", *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990, ISSN 0364-0213. URL http://dx.doi.org/https://doi.org/10.1016/0364-0213(90)90002-E.

Elmegreen, B. G., "Observations and Theory of Dynamical Triggers for Star Formation", in: *Origins*, edited by C. E. Woodward, J. M. Shull and J. Thronson, Harley A., *Astronomical Society of the Pacific Conference Series*, vol. 148, p. 150 (1998). astro-ph/9712352, URL http://dx.doi.org/10.48550/arXiv.astro-ph/9712352.

Faber, S. M. et al., "The DEIMOS spectrograph for the Keck II Telescope: integration and testing", in: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*, edited by M. Iye and A. F. M. Moorwood, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 4841, pp. 1657–1669 (2003). URL http://dx.doi.org/10.1117/12.460346.

Falk, S. W. et al., "A Theoretical Model for Type II Supernovae", *The Astrophysical Journal*, vol. 180, p. L65, 1973, ISSN 1538-4357. URL http://dx.doi.org/10.1086/181154.

Fang, Q. et al., "Red Supergiant problem viewed from the nebular phase spectroscopy of type II supernovae", *arXiv e-prints*, arXiv:2504.14502, 2025. 2504.14502.

Ferrari, L. et al., "Progenitor mass and ejecta asymmetry of supernova 2023ixf from nebular spectroscopy", *A&A*, vol. 687, p. L20, 2024. URL http://dx.doi.org/10.1051/0004-6361/202450440.

Filippenko, A. V., "Optical Spectra of Supernovae", ARA&A, vol. 35, pp. 309–355, 1997. URL http://dx.doi.org/10.1146/annurev.astro.35.1.309.

Friedmann, A., "Über die Krümmung des Raumes", *Zeitschrift für Physik*, vol. 10, no. 1, pp. 377–386, 1922, ISSN 0044-3328. URL http://dx.doi.org/10.1007/BF01332580.

Fryer, C. L. et al., "Compact Remnant Mass Function: Dependence on the Explosion Mechanism and Metallicity", ApJ, vol. 749, no. 1, 91, 2012. 1110.1726, URL http://dx.doi.org/10.1088/0004-637X/749/1/91.

Gair, J. R. et al., "The Hitchhiker's Guide to the Galaxy Catalog Approach for Dark Siren Gravitational-wave Cosmology", *The Astronomical Journal*, vol. 166, no. 1, p. 22, 2023. URL http://dx.doi.org/10.3847/1538-3881/acca78.

Galbany, L. et al., "UBVRIz Light Curves of 51 Type II Supernovae", AJ, vol. 151, no. 2, 33, 2016. 1511.08402, URL http://dx.doi.org/10.3847/0004-6256/151/2/33.

Gelman, A. et al., *Bayesian Data Analysis* (CRC Press, 2014).

Gers, F. A. et al., "Learning to Forget: Continual Prediction with LSTM", *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000, ISSN 0899-7667. https://direct.mit.edu/neco/article-pdf/12/10/2451/814643/089976600300015015.pdf, URL http://dx.doi.org/10.1162/089976600300015015.

Glorot, X. et al., "Deep Sparse Rectifier Neural Networks", in: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, edited by G. Gordon, D. Dunson and M. Dudík, *Proceedings of Machine Learning Research*, vol. 15, pp. 315–323 (PMLR, Fort Lauderdale, FL, USA, 2011). URL https://proceedings.mlr.press/v15/glorot11a.html.

"Goan, E. et al., *Bayesian Neural Networks: An Introduction and Survey*, pp. 45–87 (Springer International Publishing, Cham, 2020), ISBN 978-3-030-42553-1. URL http://dx.doi.org/10.1007/978-3-030-42553-1_3.

Goodfellow, I. et al., *Deep Learning* (MIT Press, 2016). http://www.deeplearningbook.org.

Graves, G. J. et al., "Dissecting the Red Sequence. I. Star-Formation Histories of Quiescent Galaxies: The Color-Magnitude versus the Color-σ Relation", ApJ, vol. 693, no. 1, pp. 486–506, 2009. 0810.4334, URL http://dx.doi.org/10.1088/0004-637X/693/1/486.

Gschwend, J. et al., "DES science portal: Computing photometric redshifts", *Astronomy and Computing*, vol. 25, pp. 58–80, 2018. 1708.05643, URL http://dx.doi.org/10.1016/j.ascom.2018.08.008.

Gustafsson, B. et al., "A grid of MARCS model atmospheres for late-type stars. I. Methods and general properties", A&A, vol. 486, no. 3, pp. 951–970, 2008. 0805.0554, URL http://dx.doi.org/10.1051/0004-6361:200809724.

Haines, C. et al., "CHANCES: A CHileAN Cluster galaxy Evolution Survey", *The Messenger*, vol. 190, pp. 31–33, 2023. URL http://dx.doi.org/10.18727/0722-6691/5308.

Heger, A. et al., "How Massive Single Stars End Their Life", *The Astrophysical Journal*, vol. 591, no. 1, p. 288, 2003. URL http://dx.doi.org/10.1086/375341.

Henghes, B. et al., "Benchmarking and scalability of machine-learning methods for photometric redshift estimation", *Monthly Notices of the Royal Astronomical Society*, vol. 505, no. 4, pp. 4847–4856, 2021, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/505/4/4847/38835633/stab1513.pdf, URL http://dx.doi.org/10.1093/mnras/stab1513.

Henghes, B. et al., "Deep learning methods for obtaining photometric redshift estimations from images", MNRAS, vol. 512, no. 2, pp. 1696–1709, 2022. 2109.02503, URL http://dx.doi.org/10.1093/mnras/stac480.

Hermans, J. et al., "A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful", , 2022. 2110.06581.

Hinton, G. E. et al., "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006. URL http://dx.doi.org/10.1162/neco.2006.18.7.1527.

Hjorth, J. et al., "The Distance to NGC 4993: The Host Galaxy of the Gravitational-wave Event GW170817", *The Astrophysical Journal Letters*, vol. 848, no. 2, p. L31, 2017. URL http://dx.doi.org/10.3847/2041-8213/aa9110.

Hochreiter, S. et al., "Long Short-Term Memory", *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, ISSN 0899-7667. https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf, URL http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Hoffmann, S. L. et al., "New Drizzlepac Handbook Version 2.0 Released In Hdox: Updated Documentation For HST Image Analysis", in: *American Astronomical Society Meeting Abstracts*, *American Astronomical Society Meeting Abstracts*, vol. 53, p. 216.02 (2021).

Holz, D. E. et al., "Using Gravitational-Wave Standard Sirens", *The Astrophysical Journal*, vol. 629, no. 1, p. 15, 2005. URL http://dx.doi.org/10.1086/431341.

Hosseinzadeh, G. et al., "Light Curve Fitting", , 2023a. URL http://dx.doi.org/10.5281/zenodo.8049154.

Hosseinzadeh, G. et al., "Weak Mass Loss from the Red Supergiant Progenitor of the Type II SN 2021yja", *The Astrophysical Journal*, vol. 935, no. 1, p. 31, 2022. URL http://dx.doi.org/10.3847/1538-4357/ac75f0.

Hosseinzadeh, G. et al., "Shock Cooling and Possible Precursor Emission in the Early Light Curve of the Type II SN 2023ixf", *The Astrophysical Journal Letters*, vol. 953, no. 1, p. L16, 2023b. URL http://dx.doi.org/10.3847/2041-8213/ace4c4.

Hu, W. et al., "Weak Lensing: Prospects for Measuring Cosmological Parameters", *The Astrophysical Journal*, vol. 514, no. 2, p. L65, 1999. URL http://dx.doi.org/10.1086/311947.

Hubble, E., "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae", *Proceedings of the National Academy of Science*, vol. 15, no. 3, pp. 168–173, 1929. URL http://dx.doi.org/10.1073/pnas.15.3.168.

Huchra, J. et al., "A survey of galaxy redshifts. IV - The data", ApJS, vol. 52, pp. 89–119, 1983. URL http://dx.doi.org/10.1086/190860.

Humphreys, R. M. et al., "Studies of luminous stars in nearby galaxies. III. Comments on the evolution of the most massive stars in the Milky Way and the Large Magellanic Cloud.", ApJ, vol. 232, pp. 409–420, 1979. URL http://dx.doi.org/10.1086/157301.

Ilbert, O. et al., "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey", *A&A*, vol. 457, no. 3, pp. 841–856, 2006. URL http://dx.doi.org/10.1051/0004-6361:20065138.

Jencson, J. E. et al., "An Exceptional Dimming Event for a Massive, Cool Supergiant in M51", ApJ, vol. 930, no. 1, 81, 2022. 2110.11376, URL http://dx.doi.org/10.3847/1538-4357/ac626c.

Jeon, M. et al., "A Conservative Approach for Unbiased Learning on Unknown Biases", in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16752–16760 (2022).

Jerkstrand, A., "Spectra of Supernovae in the Nebular Phase", in: *Handbook of Supernovae*, edited by A. W. Alsabti and P. Murdin, p. 795 (Springer International Publishing, 2017). URL http://dx.doi.org/10.1007/978-3-319-21846-5_29.

Jerkstrand, A. et al., "The nebular spectra of SN 2012aw and constraints on stellar nucleosynthesis from oxygen emission lines", MNRAS, vol. 439, no. 4, pp. 3694–3703, 2014. 1311.2031, URL http://dx.doi.org/10.1093/mnras/stu221.

Jerkstrand, A. et al., "Emission line models for the lowest mass core-collapse supernovae - I. Case study of a 9 $M_{\odot}$ one-dimensional neutrino-driven explosion", MNRAS, vol. 475, no. 1, pp. 277–305, 2018. 1710.04508, URL http://dx.doi.org/10.1093/mnras/stx2877.

Jerkstrand, A. et al., "The progenitor mass of the Type IIP supernova SN 2004et from late-time spectral modeling", *A&A*, vol. 546, p. A28, 2012. URL http://dx.doi.org/10.1051/0004-6361/201219528.

de Jong, R. et al., "4MOST: Project overview and information for the First Call for Proposals", *The Messenger*, vol. 175, pp. 3–11, 2019. URL http://dx.doi.org/10.18727/0722-6691/5117.

Jordan, M. I. et al., "An Introduction to Variational Methods for Graphical Models", *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999, ISSN 1573-0565. URL http://dx.doi.org/10.1023/A:1007665907178.

Kang, M., *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Harvard University Press, 2011), ISBN 9780674049352. URL http://www.jstor.org/stable/j.ctv1m46g4k.

Kasen, D. et al., "TYPE II SUPERNOVAE: MODEL LIGHT CURVES AND STANDARD CANDLE RELATIONSHIPS", *The Astrophysical Journal*, vol. 703, no. 2, p. 2205, 2009. URL http://dx.doi.org/10.1088/0004-637X/703/2/2205.

Katsuda, S. et al., "Progenitor Mass Distribution of Core-collapse Supernova Remnants in Our Galaxy and Magellanic Clouds Based on Elemental Abundances", ApJ, vol. 863, no. 2, 127, 2018. 1807.03426, URL http://dx.doi.org/10.3847/1538-4357/aad2d8.

Katz, B. et al., "NON-RELATIVISTIC RADIATION MEDIATED SHOCK BREAKOUTS. II. BOLOMETRIC PROPERTIES OF SUPERNOVA SHOCK BREAKOUT", *The Astrophysical Journal*, vol. 747, no. 2, p. 147, 2012. URL http://dx.doi.org/10.1088/0004-637X/747/2/147.

Kilpatrick, C. D., "charliekilpatrick/hst123: hst123", , 2021. URL http://dx.doi.org/10.5281/zenodo.5573941.

Kilpatrick, C. D. et al., "charliekilpatrick/forwardmodel: forwardmodel", , 2023. URL http://dx.doi.org/10.5281/zenodo.8060639.

Kilpatrick, C. D. et al., "SN 2023ixf in Messier 101: A Variable Red Supergiant as the Progenitor Candidate to a Type II Supernova", *The Astrophysical Journal Letters*, vol. 952, no. 1, p. L23, 2023a, ISSN 2041-8213. URL http://dx.doi.org/10.3847/2041-8213/ace4ca.

Kilpatrick, C. D. et al., "Type II-P supernova progenitor star initial masses and SN 2020jfo: direct detection, light-curve properties, nebular spectroscopy, and local environment", *Monthly Notices of the Royal Astronomical Society*, vol. 524, no. 2, pp. 2161–2185, 2023b.

Kim, B. et al., "Learning Not to Learn: Training Deep Neural Networks with Biased Data", , 2019. 1812.10352.

Kingma, D. P. et al., "Adam: A Method for Stochastic Optimization", , 2017. 1412.6980, URL https://arxiv.org/abs/1412.6980.

Kou, S. et al., "A New Method to Classify Type IIP/IIL Supernovae Based on Their Spectra", *The Astrophysical Journal*, vol. 890, no. 2, p. 177, 2020. URL http://dx.doi.org/10.3847/1538-4357/ab6601.

Kragh, H., "Hubble Law or Hubble-Lemaître Law? The IAU Resolution", , 2018. 1809.02557, URL https://arxiv.org/abs/1809.02557.

Kurzweil, R. (Ed.), *The Age of Intelligent Machines* (MIT Press, 1990).

Labrie, K. et al., "The Gemini Recipe System: a dynamic workflow for automated data reduction", in: *Observatory Operations: Strategies, Processes, and Systems III*, edited by D. R. Silva, A. B. Peck and B. T. Soifer, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7737, p. 77371U (2010). URL http://dx.doi.org/10.1117/12.857370.

Le, Q. V. et al., "A Simple Way to Initialize Recurrent Networks of Rectified Linear Units", , 2015. 1504.00941, URL https://arxiv.org/abs/1504.00941.

Leavitt, H. S. et al., "Periods of 25 Variable Stars in the Small Magellanic Cloud.", *Harvard College Observatory Circular*, vol. 173, pp. 1–3, 1912.

LeCun, Y. et al., "Backpropagation Applied to Handwritten Zip Code Recognition", *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989. URL http://dx.doi.org/10.1162/neco.1989.1.4.541.

Leijnen, S. et al., "The Neural Network Zoo", *Proceedings*, vol. 47, no. 1, 2020, ISSN 2504-3900. URL http://dx.doi.org/10.3390/proceedings2020047009.

Lemaître, G., "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques", *Annales de la Société Scientifique de Bruxelles*, vol. 47, pp. 49–59, 1927.

Levesque, E. M. et al., "Betelgeuse Just Is Not That Cool: Effective Temperature Alone Cannot Explain the Recent Dimming of Betelgeuse", ApJ, vol. 891, no. 2, L37, 2020. 2002.10463, URL http://dx.doi.org/10.3847/2041-8213/ab7935.

Li, C. et al., "Photometric redshift estimation of galaxies in the DESI Legacy Imaging Surveys", *Monthly Notices of the Royal Astronomical Society*, vol. 518, no. 1, pp. 513–525, 2022. URL http://dx.doi.org/10.1093/mnras/stac3037.

Li, L. et al., "LiONS Transient Classification Report for 2022-12-07", *Transient Name Server Classification Report*, vol. 2022-3549, p. 1, 2022.

Li, S. et al., "Deep Independently Recurrent Neural Network (IndRNN)", *arXiv e-prints*, arXiv:1910.06251, 2019. 1910.06251, URL http://dx.doi.org/10.48550/arXiv.1910.06251.

Lima, E., "ErikVini/SpecZCompilation: Southern Hemisphere Spectrocopic Redshift Compilation (2024.01.24) - Zenodo", , 2024. URL http://dx.doi.org/10.5281/zenodo.11641315.

Lima, E. et al., "Photometric redshifts for the S-PLUS Survey: Is machine learning up to the task?", *Astronomy and Computing*, vol. 38, p. 100510, 2022. URL http://dx.doi.org/10.1016/j.ascom.2021.100510.

Lima, E. V. R. d., *Reconstruction of the Large-Scale Structure of the Universe using photometric redshifts and machine learning techniques*, Phd thesis, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, Brazil, 2025. [cited 2026-02-23], URL http://dx.doi.org/10.11606/T.14.2025.tde-23042025-152242.

Lima, M. et al., "Estimating the redshift distribution of photometric galaxy samples", *Monthly Notices of the Royal Astronomical Society*, vol. 390, no. 1, pp. 118–130, 2008, ISSN 0035-8711. URL http://dx.doi.org/10.1111/j.1365-2966.2008.13510.x.

Limongi, M. et al., "Evolution and Final Fate of Solar Metallicity Stars in the Mass Range 7–15 M $_{\odot}$. I. The Transition from Asymptotic Giant Branch to Super-AGB Stars, Electron Capture, and Core-collapse Supernova Progenitors", ApJS, vol. 270, no. 2, 29, 2024. 2312.00107, URL http://dx.doi.org/10.3847/1538-4365/ad12c1.

Loeb, A. et al., *The First Galaxies in the Universe* (Princeton University Press, 2013).

Maddox, N. et al., "A spectroscopic census of the Fornax cluster and beyond: preparing for next generation surveys", MNRAS, vol. 490, no. 2, pp. 1666–1677, 2019. 1909.04379, URL http://dx.doi.org/10.1093/mnras/stz2530.

Makovoz, D. et al., "Mosaicking with MOPEX", in: *Astronomical Data Analysis Software and Systems XIV*, edited by P. Shopbell, M. Britton and R. Ebert, *Astronomical Society of the Pacific Conference Series*, vol. 347, p. 81 (2005).

Mandelbaum, R. et al., "Precision photometric redshift calibration for galaxy–galaxy weak lensing", *Monthly Notices of the Royal Astronomical Society*, vol. 386, no. 2, pp. 781–806, 2008.

McCarthy, J. et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955", *AI Magazine*, vol. 27, no. 4, p. 12, 2006. URL http://dx.doi.org/10.1609/aimag.v27i4.1904.

McCulloch, W. S. et al., "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943, ISSN 1522-9602. URL http://dx.doi.org/10.1007/BF02478259.

McGregor, P. et al., "Gemini South Adaptive Optics Imager (GSAOI)", in: *Ground-based Instrumentation for Astronomy*, edited by A. F. M. Moorwood and M. Iye, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 5492, pp. 1033–1044 (2004). URL http://dx.doi.org/10.1117/12.550288.

McLachlan, G. J. et al., *Mixture Models: Inference and Applications to Clustering* (Marcel Dekker, New York, United States, 1988).

Mendes de Oliveira, C. et al., "The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters", MNRAS, vol. 489, no. 1, pp. 241–267, 2019. 1907.01567, URL http://dx.doi.org/10.1093/mnras/stz1985.

Meza-Retamal, N. et al., "Circumstellar Interaction Signatures in the Low-luminosity Type II SN 2021gmj", ApJ, vol. 971, no. 2, 141, 2024. 2401.04027, URL http://dx.doi.org/10.3847/1538-4357/ad4d55.

Mitchell, T. M., *Machine learning*, 9 (McGraw-hill New York, 1997).

Molino, A. et al., "Assessing the photometric redshift precision of the S-PLUS survey: the Stripe-82 as a test-case", *Monthly Notices of the Royal Astronomical Society*, vol. 499, no. 3, pp. 3884–3908, 2020, ISSN 0035-8711. URL http://dx.doi.org/10.1093/mnras/staa1586.

Monsalves, N. et al., "Application of Convolutional Neural Networks to time domain astrophysics. 2D image analysis of OGLE light curves", *A&A*, vol. 691, p. A106, 2024. URL http://dx.doi.org/10.1051/0004-6361/202449995.

Morag, J. et al., "Shock cooling emission from explosions of red supergiants – I. A numerically calibrated analytic model", *Monthly Notices of the Royal Astronomical Society*, vol. 522, no. 2, pp. 2764–2776, 2023, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/522/2/2764/56878154/stad899.pdf, URL http://dx.doi.org/10.1093/mnras/stad899.

Morganson, E. et al., "The Dark Energy Survey Image Processing Pipeline", *Publications of the Astronomical Society of the Pacific*, vol. 130, no. 989, p. 074501, 2018. URL http://dx.doi.org/10.1088/1538-3873/aab4ef.

Moskowitz, I. et al., "Improving Photometric Redshift Estimates with Training Sample Augmentation", *The Astrophysical Journal Letters*, vol. 967, no. 1, p. L6, 2024. URL http://dx.doi.org/10.3847/2041-8213/ad4039.

Mucesh, S. et al., "A machine learning approach to galaxy properties: joint redshift–stellar mass probability distributions with Random Forest", *Monthly Notices of the Royal Astronomical Society*, vol. 502, no. 2, pp. 2770–2786, 2021. URL http://dx.doi.org/10.1093/mnras/stab164.

Méndez-Hernández, H. et al., "Targeting cluster galaxies for the 4MOST CHANCES Low-z sub-survey with photometric redshifts", *A&A*, vol. 706, p. A34, 2026. URL http://dx.doi.org/10.1051/0004-6361/202556796.

Nakazono, L. et al., "The Quasar Catalogue for S-PLUS DR4 (QuCatS) and the estimation of photometric redshifts", *Monthly Notices of the Royal Astronomical Society*, vol. 531, no. 1, pp. 327–339, 2024, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/531/1/327/57604639/stae971.pdf, URL http://dx.doi.org/10.1093/mnras/stae971.

Neal, R. M., *Bayesian Learning for Neural Networks*, *Lecture Notes in Statistics*, vol. 118 (Springer, New York, NY, 1996), ISBN 978-0-387-94724-2. URL http://dx.doi.org/10.1007/978-1-4612-0745-0.

Newman, J. A. et al., "Photometric Redshifts for Next-Generation Surveys", *Annual Review of Astronomy and Astrophysics*, vol. 60, no. Volume 60, 2022, pp. 363–414, 2022, ISSN 1545-4282. URL http://dx.doi.org/https://doi.org/10.1146/annurev-astro-032122-014611.

Nicolaou, C. et al., "The impact of peculiar velocities on the estimation of the Hubble constant from gravitational wave standard sirens", *Monthly Notices of the Royal Astronomical Society*, vol. 495, no. 1, pp. 90–97, 2020, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/495/1/90/33216608/staa1120.pdf, URL http://dx.doi.org/10.1093/mnras/staa1120.

Ochsenbein, F. et al., "The VizieR database of astronomical catalogues", A&AS, vol. 143, pp. 23–32, 2000. astro-ph/0002122, URL http://dx.doi.org/10.1051/aas:2000169.

Oke, J. B. et al., "THE KECK LOW-RESOLUTION IMAGING SPECTROMETER", *Publications of the Astronomical Society of the Pacific*, vol. 107, no. 710, p. 375, 1995. URL http://dx.doi.org/10.1086/133562.

Palmese, A. et al., "A Standard Siren Measurement of the Hubble Constant Using Gravitational-wave Events from the First Three LIGO/Virgo Observing Runs and the DESI Legacy Survey", *The Astrophysical Journal*, vol. 943, no. 1, p. 56, 2023. URL http://dx.doi.org/10.3847/1538-4357/aca6e3.

Pasquet, J. et al., "Photometric redshifts from SDSS images using a convolutional neural network", *A&A*, vol. 621, p. A26, 2019. URL http://dx.doi.org/10.1051/0004-6361/201833617.

Pedregosa, F. et al., "Scikit-learn: Machine Learning in Python", *J. Mach. Learn. Res.*, vol. 12, no. null, p. 2825–2830, 2011, ISSN 1532-4435.

Perlmutter, S. et al., "Measurements of Ω and Λ from 42 High-Redshift Supernovae", ApJ, vol. 517, no. 2, pp. 565–586, 1999. astro-ph/9812133, URL http://dx.doi.org/10.1086/307221.

Planck Collaboration et al., "Planck 2018 results - VI. Cosmological parameters", *A&A*, vol. 641, p. A6, 2020. URL http://dx.doi.org/10.1051/0004-6361/201833910.

Polsterer, K. L. et al., "Uncertain Photometric Redshifts", , 2016. 1608.08016.

Popov, D. V., "An Analytical Model for the Plateau Stage of Type II Supernovae", ApJ, vol. 414, p. 712, 1993. URL http://dx.doi.org/10.1086/173117.

Prochaska, J. et al., "PypeIt: The Python Spectroscopic Data Reduction Pipeline", *The Journal of Open Source Software*, vol. 5, no. 56, 2308, 2020. 2005.06505, URL http://dx.doi.org/10.21105/joss.02308.

Rabinak, I. et al., "THE EARLY UV/OPTICAL EMISSION FROM CORE-COLLAPSE SUPERNOVAE", *The Astrophysical Journal*, vol. 728, no. 1, p. 63, 2011. URL http://dx.doi.org/10.1088/0004-637X/728/1/63.

Ragosta, F. et al., "On the binary nature of the progenitor of SN2015ap: insights from its light curve and spectral evolution", *Monthly Notices of the Royal Astronomical Society*, vol. 541, no. 2, pp. 1048–1063, 2025, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/541/2/1048/63627506/staf1054.pdf, URL http://dx.doi.org/10.1093/mnras/staf1054.

Ranzato, M. a. et al., "Efficient Learning of Sparse Representations with an Energy-Based Model", in: *Advances in Neural Information Processing Systems*, edited by B. Schölkopf, J. Platt and T. Hoffman, vol. 19 (MIT Press, 2006). URL https://proceedings.neurips.cc/paper_files/paper/2006/file/87f4d79e36d68c3031ccf6c55e9bbd39-Paper.pdf.

Ren, Y. et al., "Photo-z Estimation with Normalizing Flow", *The Astrophysical Journal*, vol. 997, no. 1, p. 65, 2026. URL http://dx.doi.org/10.3847/1538-4357/ae201d.

Reynolds, D. A. et al., "Gaussian mixture models.", *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.

Riess, A. G. et al., "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant", AJ, vol. 116, no. 3, pp. 1009–1038, 1998. astro-ph/9805201, URL http://dx.doi.org/10.1086/300499.

Rigaut, F. et al., "Gemini multiconjugate adaptive optics system review - I. Design, trade-offs and integration", MNRAS, vol. 437, no. 3, pp. 2361–2375, 2014. 1310.6199, URL http://dx.doi.org/10.1093/mnras/stt2054.

Rindler, W., *Relativity: Special, General, And Cosmological Second Edition* (Oxford University Press, 2006), ISBN 9780198567318. URL http://dx.doi.org/10.1093/oso/9780198567318.001.0001.

Rodríguez, Ó., "Luminosity distribution of Type II supernova progenitors", MNRAS, vol. 515, no. 1, pp. 897–913, 2022. 2206.12974, URL http://dx.doi.org/10.1093/mnras/stac1831.

Rodríguez-Ramírez, J. C. et al., "Optical emission model for Binary Black Hole merger remnants travelling through discs of Active Galactic Nuclei", *Monthly Notices of the Royal Astronomical Society*, vol. 527, no. 3, pp. 6076–6089, 2023, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/527/3/6076/54130328/stad3575.pdf, URL http://dx.doi.org/10.1093/mnras/stad3575.

Roming, P. W. A. et al., "The Swift Ultra-Violet/Optical Telescope", Space Sci. Rev., vol. 120, no. 3-4, pp. 95–142, 2005. astro-ph/0507413, URL http://dx.doi.org/10.1007/s11214-005-5095-4.

Rosado-Belza, D. et al., "The kinematics of young and old stellar populations in nuclear rings of MUSE TIMER galaxies", *A&A*, vol. 644, p. A116, 2020. URL http://dx.doi.org/10.1051/0004-6361/202039530.

Rosenblatt, F., "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958. URL http://dx.doi.org/10.1037/h0042519.

Rumelhart, D. E. et al., "Learning representations by back-propagating errors", *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, ISSN 1476-4687. URL http://dx.doi.org/10.1038/323533a0.

Russell, S. et al., *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2010), 3 ed.

Sadeh, I. et al., "ANNz2: photometric redshift and probability distribution function estimation using machine learning", *Publications of the Astronomical Society of the Pacific*, vol. 128, no. 968, p. 104502, 2016.

Sagan, C., *Cosmos* (Random House, New York, 1980).

Sahijpal, S., "Contributions of type II and Ib/c supernovae to Galactic chemical evolution", *Research in Astronomy and Astrophysics*, vol. 14, no. 6, 693-704, 2014. 1401.1202, URL http://dx.doi.org/10.1088/1674-4527/14/6/008.

Salvato, M. et al., "The many flavours of photometric redshifts", *Nature Astronomy*, vol. 3, pp. 212–222, 2019. 1805.12574, URL http://dx.doi.org/10.1038/s41550-018-0478-0.

Samuel, A., "Some studies in machine learning using the game of checkers (Reprinted from Journal of Research and Development, vol 3, 1959)", *Ibm Journal of Research and Development*, vol. 44, pp. 207–226, 2000.

Santana-Silva, L., "From Pixels to Classes: Deep Learning Classification of 30 Million Galaxies in the DECam Local Volume Exploration Survey", *in prep.*, in prep. In prep.

Santos, A. et al., "The S-PLUS Transient Extension Program: imaging pipeline, transient identification, and survey optimization for multimessenger astronomy", *Monthly Notices of the Royal Astronomical Society*, vol. 529, no. 1, pp. 59–73, 2024.

Sapir, N. et al., "UV/Optical Emission from the Expanding Envelopes of Type II Supernovae", *The Astrophysical Journal*, vol. 838, no. 2, p. 130, 2017. URL http://dx.doi.org/10.3847/1538-4357/aa64df.

Schechter, P. L. et al., "DOPHOT, A CCD PHOTOMETRY PROGRAM: DESCRIPTION AND TESTS", *Publications of the Astronomical Society of the Pacific*, vol. 105, no. 693, p. 1342, 1993. URL http://dx.doi.org/10.1086/133316.

Schlegel, D. J. et al., "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds", *The Astrophysical Journal*, vol. 500, no. 2, p. 525, 1998. URL http://dx.doi.org/10.1086/305772.

Schmidt, S. et al., "Evaluation of probabilistic photometric redshift estimation approaches for The Rubin Observatory Legacy Survey of Space and Time (LSST)", *Monthly Notices of the Royal Astronomical Society*, vol. 499, no. 2, pp. 1587–1606, 2020.

Schuldt, S. et al., "Photometric redshift estimation with a convolutional neural network: NetZ", *A&A*, vol. 651, p. A55, 2021. URL http://dx.doi.org/10.1051/0004-6361/202039945.

Schutz, B. F., "Determining the Hubble constant from gravitational wave observations", *Nature*, vol. 323, no. 6086, pp. 310–311, 1986, ISSN 1476-4687. URL http://dx.doi.org/10.1038/323310a0.

Shannon, C. E., "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1948.tb01338.x, URL http://dx.doi.org/https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shlens, J., "A Tutorial on Principal Component Analysis", , 2014. 1404.1100, URL https://arxiv.org/abs/1404.1100.

Shrestha, M. et al., "Evidence of Weak Circumstellar Medium Interaction in the Type II SN 2023axu", ApJ, vol. 961, no. 2, 247, 2024a. 2310.00162, URL http://dx.doi.org/10.3847/1538-4357/ad11e1.

Shrestha, M. et al., "Extended Shock Breakout and Early Circumstellar Interaction in SN 2024ggi", ApJ, vol. 972, no. 1, L15, 2024b. 2405.18490, URL http://dx.doi.org/10.3847/2041-8213/ad6907.

Sifón, C. et al., "CHANCES, the Chilean Cluster Galaxy Evolution Survey: Selection and initial characterisation of clusters and superclusters", A&A, vol. 697, A92, 2025. 2411.13655, URL http://dx.doi.org/10.1051/0004-6361/202452710.

Slipher, V. M., "The radial velocity of the Andromeda Nebula", *Lowell Observatory Bulletin*, vol. 2, no. 8, pp. 56–57, 1913.

Slipher, V. M., "Spectrographic Observations of Nebulae", *Popular Astronomy*, vol. 23, pp. 21–24, 1915.

Slipher, V. M., "Nebulae", *Proceedings of the American Philosophical Society*, vol. 56, pp. 403–409, 1917.

Smartt, S. J., "Progenitors of Core-Collapse Supernovae", ARA&A, vol. 47, no. 1, pp. 63–106, 2009. 0908.0700, URL http://dx.doi.org/10.1146/annurev-astro-082708-101737.

Smartt, S. J., "Observational Constraints on the Progenitors of Core-Collapse Supernovae: The Case for Missing High-Mass Stars", *Publications of the Astronomical Society of Australia*, vol. 32, p. e016, 2015. URL http://dx.doi.org/10.1017/pasa.2015.17.

Smith, N., "Mass Loss: Its Effect on the Evolution and Fate of High-Mass Stars", ARA&A, vol. 52, pp. 487–528, 2014. 1402.1237, URL http://dx.doi.org/10.1146/annurev-astro-081913-040025.

Soares-Santos, M. et al., "First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary-Black-hole Merger GW170814", ApJ, vol. 876, no. 1, L7, 2019. 1901.01540, URL http://dx.doi.org/10.3847/2041-8213/ab14f1.

Soraisam, M. D. et al., "Variability of Red Supergiants in M31 from the Palomar Transient Factory", ApJ, vol. 859, no. 1, 73, 2018. 1803.09934, URL http://dx.doi.org/10.3847/1538-4357/aabc59.

Soumagnac, M. T. et al., "Star/galaxy separation at faint magnitudes: application to a simulated Dark Energy Survey", MNRAS, vol. 450, no. 1, pp. 666–680, 2015. 1306.5236, URL http://dx.doi.org/10.1093/mnras/stu1410.

Springob, C. M. et al., "A Digital Archive of H I 21 Centimeter Line Spectra of Optically Targeted Galaxies", ApJS, vol. 160, no. 1, pp. 149–162, 2005. astro-ph/0505025, URL http://dx.doi.org/10.1086/431550.

Srivastava, N. et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014a. URL http://jmlr.org/papers/v15/srivastava14a.html.

Srivastava, N. et al., "Dropout: a simple way to prevent neural networks from overfitting", *J. Mach. Learn. Res.*, vol. 15, no. 1, p. 1929–1958, 2014b, ISSN 1532-4435.

Stone, M., "Cross-Validatory Choice and Assessment of Statistical Predictions", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974. https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1974.tb00994.x, URL http://dx.doi.org/https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

STScI Development Team, "pysynphot: Synthetic photometry software package", Astrophysics Source Code Library, record ascl:1303.023, 2013.

Sukhbold, T. et al., "CORE-COLLAPSE SUPERNOVAE FROM 9 TO 120 SOLAR MASSES BASED ON NEUTRINO-POWERED EXPLOSIONS", *The Astrophysical Journal*, vol. 821, no. 1, p. 38, 2016. URL http://dx.doi.org/10.3847/0004-637X/821/1/38.

Sutton, R. S. et al., *Reinforcement Learning: An Introduction* (A Bradford Book, Cambridge, MA, 2018), ISBN 978-0-262-03924-6, 552 pp.

Tagliaferri, R. et al., "Neural Networks for Photometric Redshifts Evaluation", in: *Lecture Notes in Computer Science*, vol. 2859, pp. 226–234 (Springer, Berlin, Heidelberg, 2003). URL http://dx.doi.org/10.1007/978-3-540-45216-4_26.

Tartaglia, L. et al., "The Early Detection and Follow-up of the Highly Obscured Type II Supernova 2016ija/DLT16am", ApJ, vol. 853, no. 1, 62, 2018. 1711.03940, URL http://dx.doi.org/10.3847/1538-4357/aaa014.

Teixeira, G. et al., "Photometric redshifts probability density estimation from recurrent neural networks in the DECam local volume exploration survey data release 2", *Astronomy and Computing*, vol. 49, 100886, 2024. 2408.15243, URL http://dx.doi.org/10.1016/j.ascom.2024.100886.

Teixeira, G. et al., "SN 2022acko and the Properties of Its Red Supergiant Progenitor: Direct Detection, Light Curves, and Nebular Spectroscopy", *The Astrophysical Journal*, vol. 998, no. 1, p. 123, 2026. URL http://dx.doi.org/10.3847/1538-4357/ae27a0.

Tinyanont, S. et al., "Progenitor and close-in circumstellar medium of type II supernova 2020fqv from high-cadence photometry and ultra-rapid UV spectroscopy", *Monthly Notices of the Royal Astronomical Society*, vol. 512, no. 2, pp. 2777–2797, 2021, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/512/2/2777/43158830/stab2887.pdf, URL http://dx.doi.org/10.1093/mnras/stab2887.

Tonry, J. L. et al., "The ATLAS All-Sky Stellar Reference Catalog", *The Astrophysical Journal*, vol. 867, no. 2, p. 105, 2018. URL http://dx.doi.org/10.3847/1538-4357/aae386.

Troxel, M. et al., "The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology", *Physics Reports*, vol. 558, pp. 1–59, 2015, ISSN 0370-1573. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology, URL http://dx.doi.org/https://doi.org/10.1016/j.physrep.2014.11.001.

Turing, A. M., "I.—COMPUTING MACHINERY AND INTELLIGENCE", *Mind*, vol. LIX, no. 236, pp. 433–460, 1950, ISSN 0026-4423. https://academic.oup.com/mind/article-pdf/LIX/236/433/61209000/mind_lix_236_433.pdf, URL http://dx.doi.org/10.1093/mind/LIX.236.433.

Valenti, S. et al., "The diversity of Type II supernova versus the similarity in their progenitors", MNRAS, vol. 459, no. 4, pp. 3939–3962, 2016. 1603.08953, URL http://dx.doi.org/10.1093/mnras/stw870.

Van Dyk, S. D. et al., "Identifying the SN 2022acko progenitor with JWST", *Monthly Notices of the Royal Astronomical Society*, vol. 524, no. 2, pp. 2186–2194, 2023, ISSN 0035-8711. https://academic.oup.com/mnras/article-pdf/524/2/2186/50895458/stad2001.pdf, URL http://dx.doi.org/10.1093/mnras/stad2001.

Van Dyk, S. D. et al., "Identifying the SN 2022acko progenitor with JWST", *Monthly Notices of the Royal Astronomical Society*, vol. 524, no. 2, p. 2186–2194, 2023, ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stad2001.

Virtanen, P. et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Methods*, vol. 17, pp. 261–272, 2020. URL http://dx.doi.org/10.1038/s41592-019-0686-2.

Voelker, A. R. et al., "Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks", in: *Advances in Neural Information Processing Systems*, pp. 15544–15553 (2019).

Wang, C.-J. et al., "Lijiang 2.4-meter Telescope and its instruments", *Research in Astronomy and Astrophysics*, vol. 19, no. 10, p. 149, 2019. URL http://dx.doi.org/10.1088/1674-4527/19/10/149.

Waxman, E. et al., *Shock Breakout Theory*, p. 967–1015 (Springer International Publishing, 2017), ISBN 9783319218465. URL http://dx.doi.org/10.1007/978-3-319-21846-5_33.

Wheeler, J. C., *Index*, p. 328–339 (Cambridge University Press, 2007).

Woosley, S. E. et al., "The evolution and explosion of massive stars", *Reviews of Modern Physics*, vol. 74, no. 4, pp. 1015–1071, 2002. URL http://dx.doi.org/10.1103/RevModPhys.74.1015.

Woosley, S. E. et al., "The physics of supernova explosions.", ARA&A, vol. 24, pp. 205–253, 1986. URL http://dx.doi.org/10.1146/annurev.aa.24.090186.001225.

Yin, H. et al., "A Rapid Review of Clustering Algorithms", , 2024. 2401.07389, URL https://arxiv.org/abs/2401.07389.

York, D. G. et al., "The Sloan Digital Sky Survey: Technical Summary", *The Astronomical Journal*, vol. 120, no. 3, p. 1579, 2000. URL http://dx.doi.org/10.1086/301513.

Zhao, D. Y. et al., "Diagnostics for conditional density models and Bayesian inference algorithms", in: *Conference on Uncertainty in Artificial Intelligence* (2021). URL https://api.semanticscholar.org/CorpusID:235436347.

Zhou, R. et al., "The clustering of DESI-like luminous red galaxies using photometric redshifts", MNRAS, vol. 501, no. 3, pp. 3309–3331, 2021. 2001.06018, URL http://dx.doi.org/10.1093/mnras/staa3764.

Zhou, R. et al., "DESI luminous red galaxy samples for cross-correlations", J. Cosmology Astropart. Phys., vol. 2023, no. 11, 097, 2023. 2309.06443, URL http://dx.doi.org/10.1088/1475-7516/2023/11/097.

Zhou, X. et al., "Photometric Redshift Estimates using Bayesian Neural Networks in the CSST Survey", *Research in Astronomy and Astrophysics*, vol. 22, no. 11, p. 115017, 2022. URL http://dx.doi.org/10.1088/1674-4527/ac9578.

Zhuang, F. et al., "A Comprehensive Survey on Transfer Learning", , 2020. 1911.02685, URL https://arxiv.org/abs/1911.02685.

Zou, H. et al., "Photometric Redshifts and Stellar Masses for Galaxies from the DESI Legacy Imaging Surveys", *The Astrophysical Journal Supplement Series*, vol. 242, no. 1, p. 8, 2019. URL http://dx.doi.org/10.3847/1538-4365/ab1847.

Zuntz, J. et al., "The LSST-DESC 3x2pt Tomography Optimization Challenge", *The Open Journal of Astrophysics*, vol. 4, no. 1, 2021, ISSN 2565-6120. URL http://dx.doi.org/10.21105/astro.2108.13418.