UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia
Laurea Magistrale in Fisica

FINAL DISSERTATION

Convolutional Neural Network data analysis development for the Large Sized Telescope of CTA and broadband study of the blazar 1ES 1959+650

Author:
Pietro GRESPAN

Supervisor:
Prof. Mosè MARIOTTI

Co-Supervisors:
Dr. Rubén LÓPEZ-COTO
Dr. Elisa PRANDINI

Date of defense
December 1, 2020

Alla mia amata famiglia.

Declaration

I herewith declare that this thesis titled, "Convolutional Neural Network data analysis development for the Large Sized Telescope of CTA and broadband study of the blazar 1ES 1959+650", which is submitted in fulfillment of the requirements for the Master Degree in Physics, represents my own work except where due acknowledgement have been made. I further declared that it has not been previously included in a thesis, dissertation, or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Padova, 1 December 2020.

Abstract

For centuries, the human knowledge about the Universe was derived observing the sky only in the tiny optical window of the electromagnetic spectrum. It was only when astronomers started putting together the information coming from different electromagnetic bands, that they could unveil the sky in all its magnificence. The recent detection of gravitational waves and cosmic neutrinos gave a further substantial contribution towards a comprehensive view of the sky, and marked the birth of new, exciting discipline: the multi-messenger astrophysics.

The very-high-energy (VHE > 50 GeV) part of the electromagnetic spectrum was explored only in the last three decades. VHE γ-ray radiation is produced by the most violent processes in the Universe, for example in the region surrounding supermassive black holes, or in the vicinities of neutron stars. Whenever one of these photons reaches the Earth, it interacts with the air molecules in the high atmosphere initiating a cascade of ultra-relativistic charged particles that produce a faint and blue flash of Cherenkov light that reaches the ground. VHE γ-ray astronomy was boosted in the early 2000s by the advent of the Imaging Atmospheric Cherenkov Telecopes (IACTs) - such as MAGIC, VERITAS and H.E.S.S. - which detect the Cherenkov light and convert it into a shower image than can be analyzed to extract the properties of the primary γ ray. The Large Sized Telescope (LST) is one of the next-generation IACTs that will be part of the Cherenkov Telescope Array (CTA) observatory, designed to reach the lowest energy threshold ever achieved by a Cherenkov telescope.

This thesis is the outcome of my work experience in the Group of Multimessenger Astrophysics in Padova (INFN) and is articulated in two, distinct parts reflecting two research activities that I carried in parallel:

- The **application of Convolutional Neural Networks (CNNs) to the full event reconstruction for the LST**. Dominating background for IACTs is constituted by images produced by cosmic hadrons, with typical noise-to-signal ratios of several orders of magnitude. The standard machine learning technique used in the LST analysis for separating γ rays from hadronic background and to reconstruct their energy and arrival direction is the Random Forest method. It is based on a set of parameters extracted from the images, whereas recent Deep Learning techniques like CNNs are able to autonomously learn how to extract information from raw images, deciding by themselves which features or patterns of the dataset are meaningful for the task addressed. The aim of this work is to investigate whether this latter approach is able to outperform the standard parameter-based analysis.

- The **analysis of VHE γ-ray emission from an active galaxy detected by the MAGIC telescopes, the blazar 1ES 1959+650, in a multiwavelength context**. Blazars emit two extremely energetic, collimated jets closely to the observer's line of sight, and are characterized by a rapid variability across the entire electromagnetic spectrum. The study of bright blazars such as 1ES 1959+650 is very important to probe the mechanisms at work within the jet.

After an introduction on cosmic ray and γ-ray physics in **Chapter 1**, in **Chapter 2** I give an overview on the imaging atmospheric technique, the IACTs (MAGIC, CTA) and a detailed explanation of IACT data analysis.

For the first part of the thesis, I had the chance to be member of the Machine Learning group of CTA. In **Chapter 3**, after a quick review of Deep Learning and CNNs and their application to IACT data analysis, I illustrate our setup. This includes the implementation of a handcrafted VGG-style network to tackle the problems of **signal/background separation**, **energy and direction reconstruction**, and the data pre-processing and handling routine I contributed to implement, comprehending image interpolation, indexing and selection.

In **Chapter 4** I report the training procedures, the test strategies and the results for each task, where I show that the CNN we implemented perform better than the standard Random Forests in all three tasks. Moreover, the outcomes of our analysis are very similar to the results independently obtained by two other research groups (the γ-ray group at LAPP and the DL group of UCM, developing the *GammaLearn* and *CTLearn* libraries respectively) that carried out a similar work.

For the second part of the thesis, I joined the research group of the MAGIC collaboration that is performing a multiwavelength observation campaign on the blazar 1ES 1959+650. **Chapter 5** is concerned with a summary on the physics of active galactic nuclei and blazars in particular, giving a historical overview of our source.

In **Chapter 6** I firstly illustrate the analysis of the observations of 1ES 1959+650 performed by the MAGIC telescopes in 2017, describing the careful data selection applied and the procedures used to extract the signal. Then I present the study of the broadband emission of the source using data taken with different instruments between 2016 and 2020 including my analysis, where we noted an enhanced variability in VHE γ rays and in X-rays. Investigating the simultaneous emission in these two bands, we found two distinct correlation trends ascribable to a changed condition in the acceleration environment. Eventually, I constructed the spectral energy distribution of the source observed on a particular night and modeled its emission.

The conclusive **Chapter 7** summarizes all the results obtained in this work and lists prospects related with the topics treated in this thesis.

Riassunto

Per secoli, l'uomo ha studiato l'Universo osservandone soltanto la luce visibile. Solo quando gli astronomi cominciarono a unire le informazioni provenienti da diverse bande dello spettro elettromagnetico, il cosmo ha cominciato a rivelarsi in tutto il suo splendore. Le recenti osservazioni di onde gravitazionali e di neutrini cosmici hanno poi dato un'ulteriore spinta in avanti verso una più completa visuale del cielo, segnando la nascita di una nuova ed entusiasmante disciplina: l'astrofisica multi-messaggero.

La parte più energetica dello spettro elettromagnetico, sopra i 50 GeV, è stata investigata solo negli ultimi tre decenni. Questa radiazione, a cui ci si riferisce come raggi gamma, viene prodotta durante i fenomeni più violenti dell'Universo, come quelli che accadono nelle regioni circostanti un buco nero supermassiccio, oppure nelle vicinanze di stelle di neutroni. Quando uno di questi fotoni super energetici raggiunge la Terra, nell'interazione con le molecole dell'alta atmosfera produce uno sciame di particelle cariche, le quali, muovendosi più velocemente di quanto si muova la luce nell'aria, producono un lampo di fioca luce blu, chiamata luce Cherenkov, che raggiunge il terreno. I telescopi Cherenkov (IACT, Imaging Atmospheric Cherenkov Telescopes) raccolgono questa luce e la convertono in immagini dalle quali è possibile ricostruire le caratteristiche del raggio gamma originario. L'astronomia gamma ha visto uno sviluppo senza precedenti grazie all'avvento di telescopi come MAGIC, VERITAS e H.E.S.S. nei primi anni 2000. Il Large Sized Telescope (LST) è uno degli IACT di nuova generazione, ed andrà a far parte di un grande osservatorio chiamato Cherenkov Telescope Array (CTA).

Questa tesi è il risultato della mia esperienza nel gruppo di Astrofisica Multimessaggero di Padova (INFN) svolta durante il 2020, ed è articolata in due parti separate che descrivono due attività di ricerca che ho portato avanti in parallelo durante questo periodo:

- l'**applicazione di reti neurali convoluzionali (CNN) all'analisi dati di LST**. Il flusso entrante di raggi gamma è sovrastato da un flusso di adroni cosmici che danno luogo a immagini simili. Questo comporta un rapporto tra rumore e segnale dell'ordine di qualche migliaio. Nell'analisi dati di LST, la tecnica standard per la **separazione** di questi due tipi di eventi è il metodo delle Random Forest, una tecnologia di apprendimento automatico che viene utilizzata anche per **ricostruire l'energia** e la **direzione d'arrivo** degli eventi di tipo gamma. Questa tecnica è basata sull'analisi di parametri estratti dalle immagini. Algoritmi di recente sviuppo nell'ambito del Deep Learning, come le CNN, sono invece in grado di apprendere in maniera autonoma come estrarre utili informazioni direttamente dalle immagini, decidendo quali sono le caratteristiche più rilevanti per risolvere il problema che viene posto loro davanti. L'obiettivo di questo lavoro

è determinare se questo approccio può dare prestazioni migliori rispetto al metodo attualmente implementato.

- l'**analisi in un contesto multibanda dell'emissione gamma prodotta da una galassia attiva (il blazar 1ES 1959+650) e rivelata dai telescopi MAGIC**. I blazar sono nuclei galattici attivi che emettono un getto estremamente energetico e collimato nella direzione di vista dell'osservatore. La loro emissione è caratterizzata da una marcata variabilità in tutto lo spettro elettromagnetico. Lo studio di blazar molto luminosi come 1ES 1959+650 è di grande importanza per determinare i meccanismi fisici che accadono all'interno del getto.

Dopo un'introduzione sui raggi cosmici e sulla fisica dei raggi gamma riportata nel **Capitolo 1**, nel **Capitolo 2** viene data una panoramica sulla tecnica dei telescopi Cherenkov, su MAGIC e su CTA, e viene fornita una spiegazione dettagliata dell'analisi dati di questo tipo di strumenti.

Per la prima parte della tesi, ho avuto l'opportunità di diventare membro del gruppo di Machine Learning di CTA. Nel **Capitolo 3**, dopo una rapida escursione sul Deep Learning, le CNN e il loro attuale utilizzo nell'ambito dei telescopi Cherenkov, illustro i metodi che ho realizzato per quest'analisi. Questi comprendono l'implementazione di una rete VGG per affrontare i tre compiti (**separazione gamma/adroni**, la **ricostruzione dell'energia** e la **ricostruzione della direzione**), e il sistema di gestione dei dati che ho contribuito a sviluppare, come l'interpolazione delle immagini, la loro indicizzazione e selezione.

Nel **Capitolo 4** illustro le procedure di allenamento delle reti, le strategie per testarle e le loro prestazioni, le quali si dimostrano migliori rispetto alle Random Forest in ognuno dei tre compiti assegnati. Inoltre, dimostro che i risultati sono molto compatibili con quelli ottenuti indipendentemente da due altri gruppi di ricerca - *GammaLearn* (LAPP) e *CTLearn* (UCM) - che hanno sviluppato metodologie di analisi simili a quelle presentate un questo lavoro.

Per la seconda parte della tesi, mi sono unito al gruppo di ricerca della collaborazione MAGIC che si sta occupando di una campagna di osservazione multibanda (a diverse lunghezze d'onda) del blazar 1ES 1959+650. Nel **Capitolo 5** riassumo la fisica dei nuclei galattici attivi e dei blazar in particolare, dando inoltre una panoramica storica sulle osservazioni della sorgente.

Nel **Capitolo 6** descrivo dettagliatamente l'analisi delle osservazioni di 1ES 1959+650 svolte con i telescopi MAGIC nel 2017, dalla selezione dei dati alle procedure usate per estrarre il segnale. Presento poi lo studio dell'emissione della sorgente usando dati raccolti con diversi strumenti nel periodo tra il 2016 e il 2020, inclusa la mia analisi. Con questo studio abbiamo notato una spiccata variabilità nelle bande gamma e dei raggi X. Investigando la correlazione tra i due flussi, abbiamo trovato due diversi andamenti legati a due periodi diversi, ascrivibili a una mutazione delle condizioni fisiche nell'ambiente di accelerazione del getto. Infine, ho costruito la distribuzione spettrale di energia della sorgente osservata in una particolare notte, modellizzandone l'emissione.

In conclusione, nel **Capitolo 7** presento un resoconto dei risultati in questa tesi e ne discuto i possibili sviluppi futuri.

Acknowledgements

This work would not have been possible without the help of many people who accompanied me through this journey in the field of VHE γ-ray astrophysics.

First of all, I want to heartfully thank my two co-supervisors: Dr Rubén López-Coto for the great help and support given during the period of this thesis, constantly encouraging me to do better, and for being patient with me so many times; Dr Elisa Prandini for the several fruitful discussions, for introducing me to the fascinating world of AGNs and for the great enthusiasm for Astrophysics she irradiates. I also acknowledge them for giving me the chance to work in an international scientific community.

Special thanks to my supervisor Prof Mosè Mariotti for the great opportunity of attending the First Padova Excellence School of Physics of the Universe in Asiago, and for being always present and willing to help me with his advice.

Thanks to Nicola Marinello for all the time spent helping me moving the first steps in the field of Deep Learning, to Ettore Mariotti for his support whenever I needed some advice on CNNs, to Vanessa De Bortoli and Alessia Spolon for the many conversations and for supporting me during the most stressful moments

I am also grateful to all the Group of Multimessenger Astrophysics of Padova (INFN) for the human support and for providing me the all the resources necessary to carry out this work.

I would like to acknowledge the many scientists and analyzers who gave a significant contribution to this thesis: Tjark Miener, Dr Daniel Nieto Castano, Mikael Jacquemont and Dr Thomas Vuillaume for gathering and sharing the results of their experiments with CNNs; i give also credits to Thomas Vuillaume for the production of the Random Forests; Mr Shunsuke Sakurai for the MAGIC analysis of the years 2018 and 2019 used to build the final light curve of 1ES 1959+650, and for the cross-check of my analysis; Mrs Mari Takahashi who kindly provided the *Fermi*/LAT light curve and SED points used in the last part of the thesis; Prof Bidzina Kapanadze for the analysis of *all* the *Swift*/XRT and UVOT data here presented; Dr Vandad Fallah Ramazani for providing the Tuorla observations; Dr Andrea Tramacere for the precious help in modeling and understanding the mechanisms at work in blazars. Last of this list, but not least, big thanks go to Dr Daniela Dorner for the big support and the several discussions, and to all members of the analysis group of the MAGIC collaboration devoted to 1ES 1959+650.

Pietro GRESPAN
Università di Padova
December 1, 2020

Contents

Chapter 1

An introduction to γ-ray astronomy

Astrophysics particle began in 1912, when Austrian physicist V. F. Hess, while studying atmospheric ionization with a series of balloon flights, found that ionization rates rise with rising altitudes. Excluding the Sun and the Earth, he concluded that some other sources in the outer space must originate this radiation - later called Cosmic Rays (CRs) by R. Millikan, underlining its extraterrestrial nature. The discovery won him the Nobel prize in Physics in 1936 [Hess, 2018]).

CRs are the most energetic particles known, covering an energy range spanning from 10^8 eV up to more than 10^{20} eV; the spectrum shows a flux decreasing by 21 orders of magnitude, following approximately a power law of index -3 (see Fig. 1.1). Two spectral breaks appear in the spectrum, a knee and an ankle: the spectral index between these two points is $\Gamma = 3.1$, while it has a value of $\Gamma = 2.7$ outside. CRs are produced and accelerated by many sources and then propagated and diffused across all the Universe, within and without the Milky Way: their study can thus potentially unveil information about Galaxy formation, population, and all the objects and processes involved in their creation (e.g. Active Galactic Nuclei (AGNs), stellar explosions...). However, the vast majority of the CRs we observe consists of charged particles - in particular protons (87%) and Helium nuclei (12%) - that are deflected by the randomly oriented magnetic fields they find in their travel to the Earth. This results in almost randomly distributed arrival directions, and it makes it not possible to locate their origin in space.

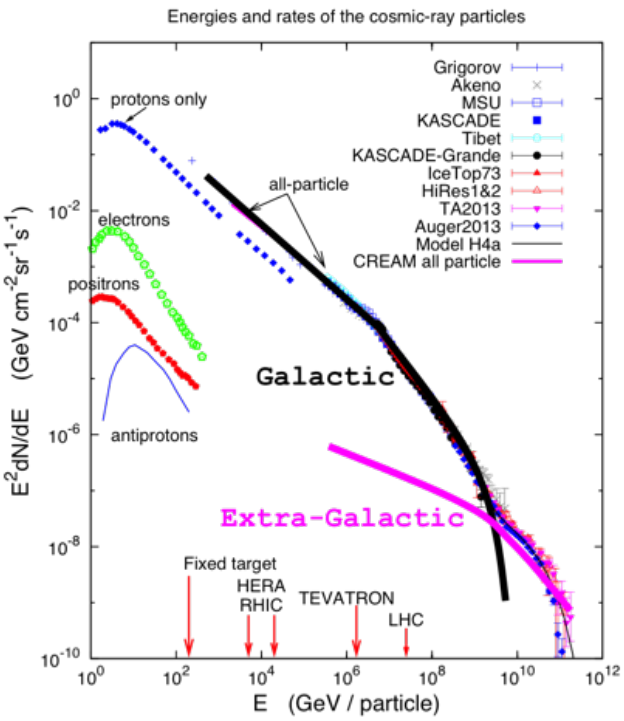


Figure 1.1: CR spectra taken by various experiments. The knee can be spotted around 10^7 eV, the ankle at the intersection between the galactic (black) and extra-galactic (pink) branches. Image from masterclass.icecube.wisc.edu.

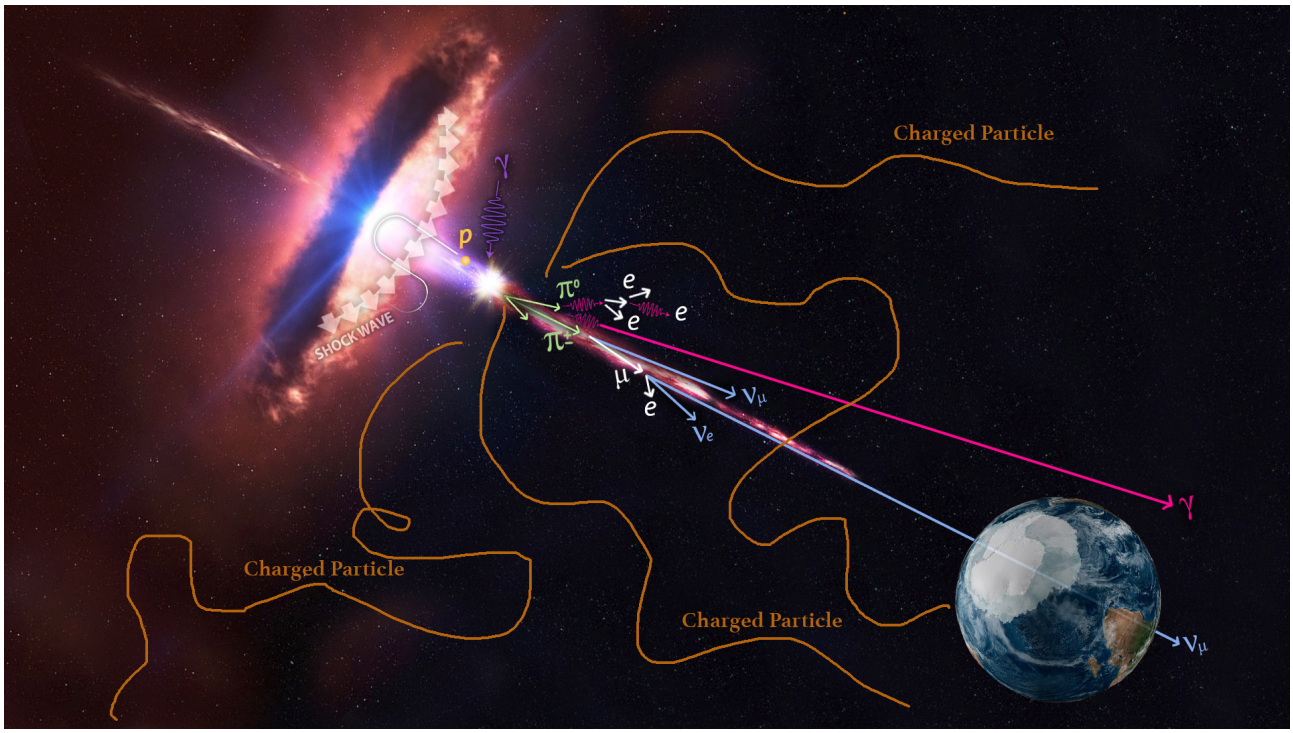


Figure 1.2: Representation of the cosmic paths of different CRs towards the Earth. Note that neutral messengers, like photons and neutrinos, follow straight lines.

In order to locate and study the sources producing CRs and the acceleration mechanisms (needed to explain such high energies), **neutral CRs** such as **photons**, neutrons and **neutrinos** are the ideal messenger, since they are not deflected by interstellar magnetic fields (). Of the three, photons are the "traditional" ideal messengers having been the only ones successfully observed for a long time: in fact, neutrons are relatively short-lived ($\sim$ 882 s in the rest frame), so that only the very energetic ones can make it to the Earth; neutrinos on the other hand are very difficult to detect since they undergo only Weak Interaction (WI). However, in very recent years an enormous interest has grown around them in the astronomers community, since the detection of the first ultra-high energy neutrino associated with an extra-galactic source in 2017 (the blazar TXS 0506+056) [Aartsen et al., 2018].

Photons, neutrinos and CRs, together with gravitational waves [Abbott et al., 2016] are nowadays the key actors of the era of **Multi-Messenger Astronomy**, the science gathering the efforts of the whole Astrophysics community to "probe the rich physics of transient phenomena in the sky" in the next decades [Branchesi, 2016].

1.1 γ rays

The most energetic form of electromagnetic radiation is produced in the most violent non-thermal processes of the Universe: the electromagnetic. Despite representing only a tiny fraction of all the CRs arriving to the Earth, they are, together with neutrinos, the only particles we can observe that directly point to their generators, as stated before, and are object of study of the so called γ-ray

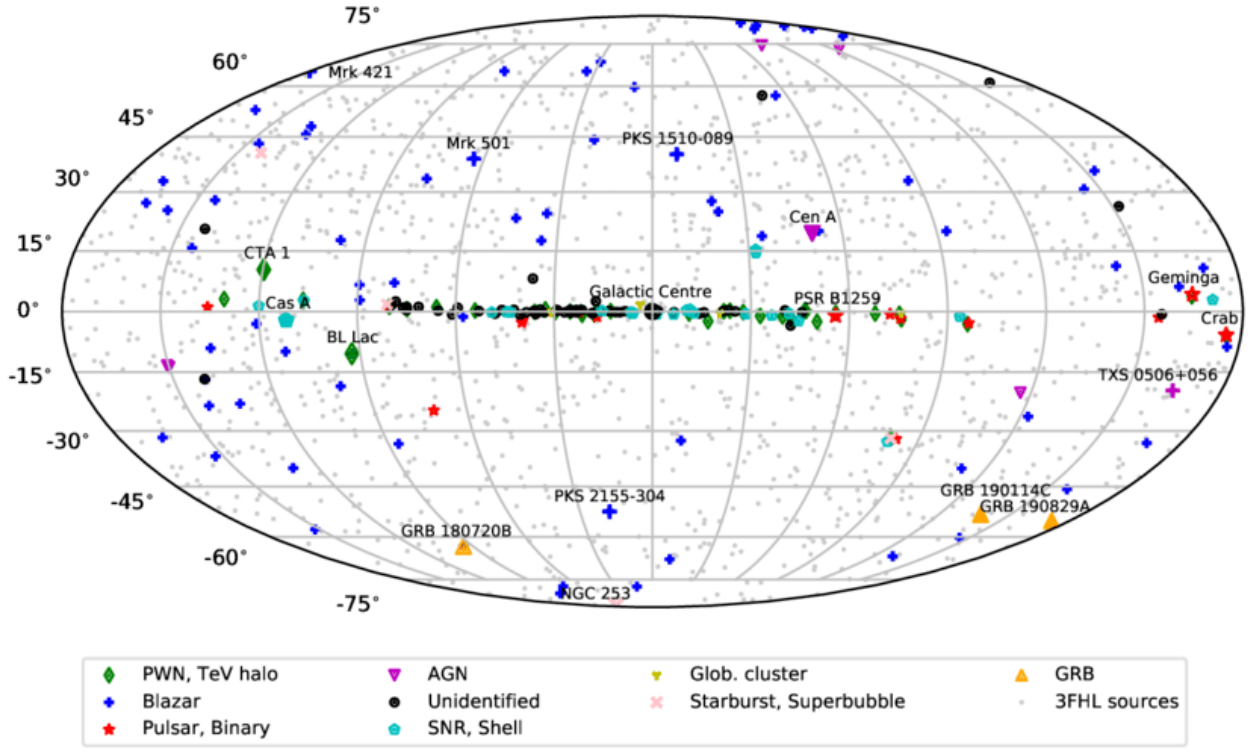


Figure 1.3: A map of the VHE γ-ray sky in mid-2019. Credits: J. Hinton.

astrophysics. γ rays populate the most energetic part of the EM spectrum, spanning from 500 keV up to hundreds of TeV. We can classify them accordingly to the five energy domains in which the CR spectrum is divided [Lopez-Coto, 2015]:

Name	Abbreviation	Energy range
low-energy	LE	500 keV - 30 MeV
high-energy	HE	30 MeV - 50 GeV
very high-energy	VHE	50 GeV - 100 TeV
ultra high-energy	UHE	100 TeV - 100 PeV
extremely high-energy	EHE	> 100 PeV

The work of this thesis is focused on VHE γ-astronomy.

1.2 VHE sources, production and absorption mechanisms

By now, around 160 VHE sources have been documented[1]. In Figure 1.3 a view of the VHE sky is shown.

[1]Source: http://tevcat2.uchicago.edu/

Galactic sources

- **Supernova remnants (SNRs)**: extended objects, leftovers of the explosion of stars in a supernova, where particles get accelerated at the shock between the ejected material and the interstellar medium. The origin of most of the emission is for sure hadronic, while whether there is a contribution of leptonic origin or not is still under debate.
- **Pulsars**: fast rotating, magnetized Neutron Stars (NSs) emitting EM waves beamed in narrow cones from the area near the magnetic poles. We observe the emission when the beam crosses the line of sight, resulting in a pulsating signal - which is responsible of their name.
- **Pulsar Wind Nebulae (PWNe)**: also referred as plerions, are a kind of nebula found inside SNRs directly fed by the Neutron Star at the center. The rotational energy of the NS is converted into a relativistic $e^{\pm}$ wind, which powers up the VHE γ-ray emission via IC (see below) scattering of ambient photons.
- **Binary systems**: they are composed by a massive star and a compact object (a NS or a Black Hole] that accretes or interacts with the companion's wind. The extreme conditions reached by the heating gas causes the emission of γ rays.

Extragalactic sources

- **Active Galactic Nuclei (AGNs)**: super massive black holes (SMBHs) accreting matter at the center of galaxies. Particles get accelerated in two ultra-relativistic collimated plasma jets, emitted perpendicular to the accretion disk. **Blazars** are AGNs with the jets aligned with the line of sight of the observer, i.e. pointing to the Earth. A more detailed review on blazars will be given in Chapter 5.
- **Starburst galaxies**: galaxies with uncommonly high star formation rate. The consequent higher supernova density results makes them ideal sites for acceleration of charged CRs that emit VHE γ rays.
- **Gamma Ray Bursts (GRBs)**: highly luminous outbursts of extragalactic origin. The name is due to the extremely large amount of energy that these transients emit in a short time. They are classified in two main types:
  - short GRBs, during less than 2 s;
  - long GRBs, lasting up to hundreds of seconds.

  The most likely production mechanisms are thought to be either catastrophic core collapse of massive stars for the long GRBs, or the merger of two compact objects originating short GRBs. On 14$^{\text{th}}$ January 2019 the MAGIC telescopes detected the long-duration GRB 190114C [V. Acciari et al., 2019], the first GRB ever detected by an IACT.

1.2.1 Production mechanisms

A VHE γ ray in the Universe is always produced by a highly energetic parent particle. The processes involved are schematically described in the following:

- **Synchrotron radation**: emitted by charged relativistic particle undergoing the effect of Lorentz's force in the presence of strong magnetic fields. The power emitted is:

$$P = \frac{q^4}{6\pi\epsilon_0 m^4 c^5} B^2 (E^2 - m^2 c^4) \tag{1.1}$$

  being B the magnetic field, m, q and E the rest mass, electric charge and total energy of the particle. This process is 10^{13} times more efficient for electrons than in the proton case, due to the m^{-4} dependence. Usually, these photons do not reach VHE, but can be up-scattered by VHE electrons through the Inverse Compton mechanism (see below).

- **Bremsstrahlung**: when charged particles in a gas interact electromagnetically with other particles, they produce a radiation with amplitude proportional to centripetal acceleration. This mechanism is very efficient only in the electron case too.

- **Inverse Compton (IC) scattering**: relativistic electrons can transfer part of their energy to photons, thus producing a γ emission up to Very High-Energy. Considering the process in the electron frame of the electron at rest, the energy of the scattered photon is:

$$h\nu' = \frac{m_e c^2}{1 - \cos\theta + m_e c^2/(h\nu)} \tag{1.2}$$

  being θ the angle of the scattered photon. The cross section is given by the Klein-Nishina formula, that at low or very-high energies reads:

$$h\nu \ll m_e c^2 : \qquad \sigma_T = \frac{8}{3}\pi r_e^2 \qquad \text{Thomson regime}$$

$$h\nu \gg m_e c^2 : \qquad \sigma_{KN} = \pi r_e^2 \frac{m_e c^2}{h\nu}\left(\ln\frac{2h\nu}{m_e c^2} + 0.5\right) \qquad \text{Klein-Nishina regime}$$

  If the electrons follow a power law distribution of index α and interact with a γ-ray field, it can be shown that the up-scattered photons follow a distribution proportional to $E^{-(\alpha+1)/2}$ and peaking at $E_{\rm peak} = \gamma_e^n h\nu$, where γ_e is the electron bulk Lorentz factor and $n = 1, 2$ depending whether we are in the Klein-Nishina or Thomson regime respectively.

- **Neutral pion decay**: a small fraction of CRs are relativistic protons. Interacting with ambient gas, they prouduce π^+, π^- and π^0 with roughly the same branching ratio; neutral pions mostly decay through EM channel in two photons of equal energy $E_\gamma \simeq 67.5$ MeV in the rest frame, and, in the lab frame:

$$E_\gamma = \frac{1}{2} m_{\pi^0} c^2 \frac{1 + \beta\cos\theta}{\sqrt{1-\beta^2}} \tag{1.3}$$

being β the velocity of the pion, and θ the angle between its direction and the photon. These photons can thus reach Very High-Energy, depending on β.

1.2.2 Absorption mechanisms

Despite they are not deflected by magnetic fields during their journey through intergalactic space, the most energetic γ rays undergo two main absorption mechanisms: *photon-matter* pair production, which is negligible, and *photon-photon* pair production, much more effective, causing a horizon for γ rays depending on their energy (the higher it is, the closer the horizon) [De Angelis and Mallamaci, 2018].

Concerning photon-photon interaction, γ rays come across two photon fields: the Cosmic Microwave Background (CMB] - which is the remnant EM radiation of the Big Bang, with an average energy of $\sim$ 0.6 meV - and the so-called Extragalactic Background Light (EBL) [Dwek and Krennrich, 2013] emitted during star formation, peaking at $\sim$ 8 meV and at $\sim$ 1 eV. The cross section is maximized for [Breit and Wheeler, 1934]:

$$\epsilon(E) \simeq \left(\frac{900 \text{ GeV}}{E}\right) \text{ eV} \tag{1.4}$$

so, for VHE γ rays, the interaction with EBL is the dominant absorption mechanism.

1.3 γ rays in the atmoshpere: the Cherenkov effect

When a charged CR, or a photon, reaches the Earth, its interaction with the air molecules gives birth to a cascade of particles each carrying a fraction of the original energy, whose number rapidly increases until the multiplication process stops and the shower is completely absorbed.

This phenomenon is also known as Extended Atmospheric Shower (EAS). The originated charged particles move at a speed greater than the speed of light in the atmosphere, producing a faint and blue flash of light that lasts a few nanoseconds, known as Cherenkov radiation [Watson, 2011]. The Cherenkov effect occurs whenever a charged particle moves faster than the phase velocity of light through a dielectric material (e.g. air, or water); the light generated by the interaction with the surrounding atoms cannot propagate forward from the particle, thus producing a shock front - much like the sonic boom produced by a supersonic aeroplane.

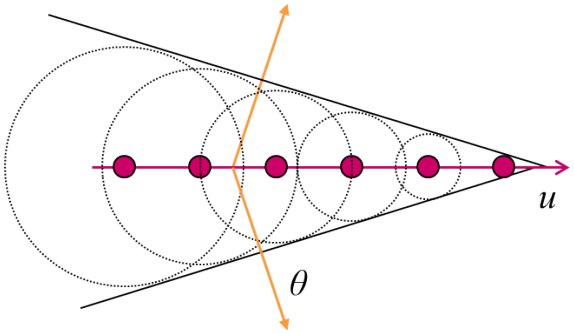


Figure 1.4: Scheme of the Cherenkov emission by a relativistic electron.

With reference to the Fig. 1.4 a blue light cone follows the electron, with an emission angle:

$$\cos\theta = \frac{1}{n(\nu,h)\beta} \tag{1.5}$$

being $n(\nu,h)$ the refractive index of the dielectric as function of frequency and atmospheric altitude and $\beta = v_p/c$ the ratio between particle velocity and speed of light in vacuum.

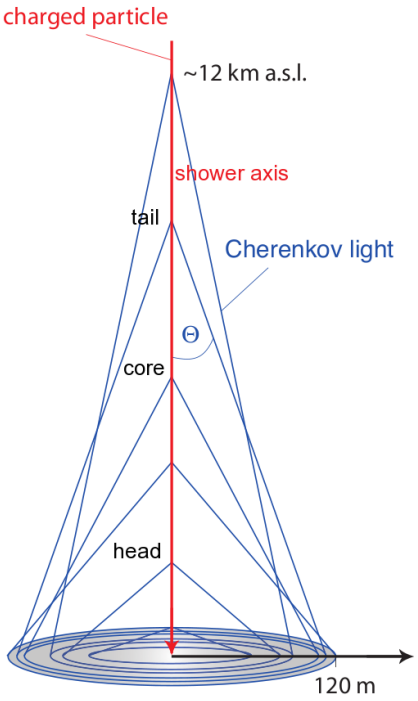


Figure 1.5: Illustration of the variation of the Cherenkov emission angle along the path. The superimposition of light cones produces the $\sim$ 120 m light pool at ground level. Credit: C. Schultz.

Cherenkov photons are emitted at an altitude of approximatively $h = 10 - 15$ km. A typical value of the emission angle is about $\theta \approx 0.7°$ in the electromagneti case, varying with height h due to variation of refractive index n along the atmosphere and to particles' energy loss along the path, thus producing at ground level a light pool of radius $r = h\tan\theta \approx 120$ m.

There exist two kinds of EASs:

- **hadronic showers**: originated mostly by protons that mainly decay into pions, and in other secondaries as kaons or nucleons.
- **electromagnetic (EM) showers**: a sequence of concatenated γ rays and e^+ / e^- pairs. Photons convert into electron/positron pairs that in turn emit EM radiation through bremsstrahlung.

Cascades originated by γ rays have a different development with respect to the hadron-initiated ones (see Fig. 1.6), and result on different shape: the formers are *pure electromagnetic* and much narrower, whereas the latter are less concentrated, due to the larger transverse momentum of the secondary hadrons produced, thus presenting a more irregular shape. Plus, γ-cascades develop faster ($\sim$ 3 ns against $\sim$ 10 ns of the hadronic case) - these facts are useful to identify the primary particle, as we will see later on in Chapters 2 and 3. Part of the initial hadron's energy can also be lost in EM showers, mimicking γ-initiated ones.

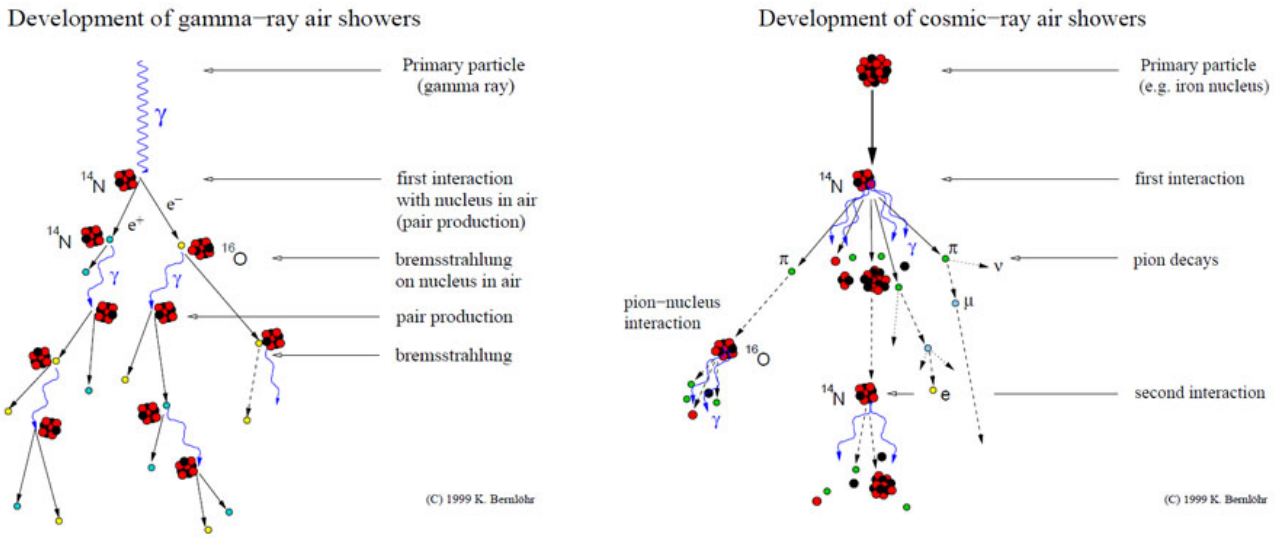


Figure 1.6: Two sketches of EASs development. Credit: K. Bernlöhr.

Chapter 2

Imaging Atmospheric Cherenkov Telecopes (IACTs)

There are two categories of experiments for the hunting of γ rays: *space-borne* satellites and *ground-based experiments*.

Experiments on board of satellites, such as ASI's *AGILE* or NASA's *Fermi*, catch photons before they enter the atmosphere and interact with the air molecules. Amongst their main strengths there are long observations times - not being scheduled by alternating day and night - and high background rejection power. However, they are not very effective at collecting VHE γ rays: due to the small photon fluxes at energies > 100 GeV, their detection area ($\sim 1\ m^2$) is not big enough to allow detections with sufficient statistics and timescales less than 1 year.

Here ground-based experiments come into play: they trace photons through the Extended Atmospheric Showers (EASs) produced in the atmosphere, and this allows to increase the collection areas of order of magnitudes with respect to detectors on satellites. There are two main techniques:

- detect the shower particles that reach the ground. Such long travelling EASs are originated by photons with an energy of above 100 GeV, so this technique is well suited for VHE and UHU γ rays. Water cherenkov arrays like the HAWC observatory[1] reach collection areas large enough to detect the fluxes of photons up to hundreds of TeV. They are located at high altitude (> 3000 m a.s.l.);
- detect the Cherenkov light emitted by the showers using the atmosphere itself as a calorimeter [Watson, 2011]. Imaging Atmospheric Cherenkov Telecopes (IACTs) such as VERITAS[2], H.E.S.S.[3] and MAGIC (see 2.2), are designed precisely with this aim, exploiting the features of the shower to retrieve information on its parent γ ray. Their targe are VHE γ rays: typical energy range of operation for an IACT is between 100 GeV (50 GeV for MAGIC) and some tens of TeV.

[1] High Altitude Water Cherenkov observatory.
[2] Very Energetic Radiation Imaging Telescope Array System
[3] High Energy Stereoscopic System

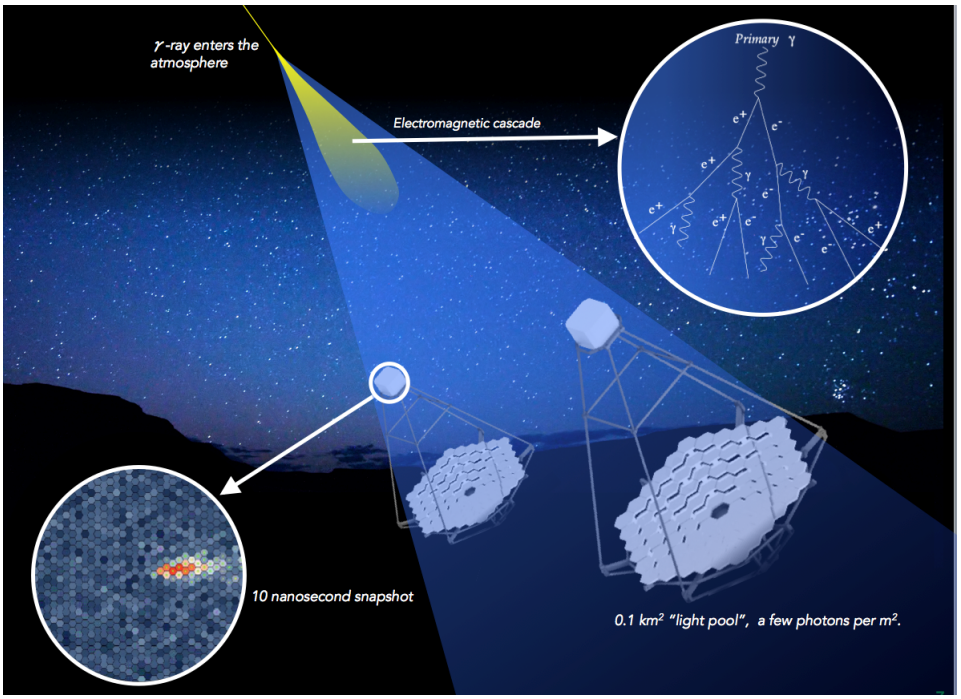


Figure 2.1: Rendering of two IACTs enlightened by an electromagnetic shower. Picture courtesy of CTA observatory.

2.1 The Imaging Atmospheric Cherenkov technique

Imaging Atmospheric Cherenkov Telecopes (IACTs) are facilities located at high altitude (about 2000-3000 m a.s.l.) in various observatories around the globe. IACT experiments usually consist on arrays of 2-5 telescopes in order to benefit of the enhanced performance of stereo observations [Caraveo, 2020]. Like shown in Figure 2.1, they are equipped with large reflectors and a camera installed on the focal plane, composed of a number of photodetectors - usually photomultipliers (PMTs). The effective area for a single IACT is of the order 10^5 m^2, while arrays can reach $\sim$ 1km^2. If an IACT is inside the Cherenkov light-pool of the γ ray, the photons are collected and focused onto the camera in such a way that their arrival directions are transformed in points of a (pixelized) shower image (see Fig. 2.1). Each photodetector converts the collected light into an electrical signal, which is preamplified and sent to the readout board and the triggering system. The triggering strategy, as well as the analysis, is based on the features of each image, i.e. geometrical shape, position, intensity, arrival times, which contain information on the primary particle. Due to the short duration of showers, IACTs must be equipped with properly responsive electronics and acquisition systems.

Nowadays, currently operating IACTs reach an angular resolution of below 0.1° for a pointlike source, and a sensitivity of the order of 1% of the Crab Nebula flux[4] in the energy range between 100 GeV and 1 TeV, worsening at energies below 100 GeV due to poor signal/background separation, and at energies above few tens of TeV due to low statistics [Acharya et al., 2013].

[4]The Crab Nebula is the "standard candle" of VHE astronomy, thanks to its strong and steady flux.

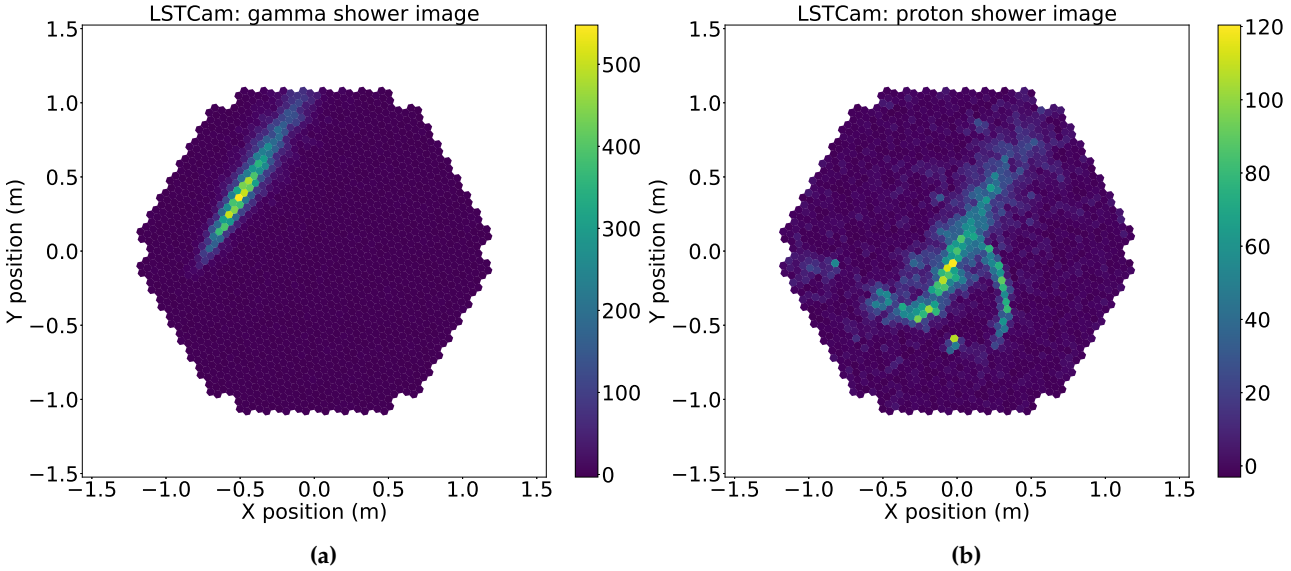


Figure 2.2: Shower images of gamma (left) and proton (right) initiated EASs on the camera of the LST.

2.1.1 The hadronic background

As we know from 1.3, there are two types of EASs, *(a)* pure EM showers, mostly induced by gamma rays and *(b)* hadronic showers. The flux of incoming γ rays is dwarfed by the flux of charged CRs entering the atmosphere that can mimic the signal which we are interested in (for example neutral pions produced in the atmosphere - see 1.2.1), resulting in a signal-to-noise ratio of about 1/1000 for VHE γ-ray sources. Fortunately, the different development of the cascades result in images of different shapes (see Fig. 2.2), that can be discriminated by means of Machine Learning techniques as described in 2.4 and 3.

2.1.2 Operation modes

An IACT performs observations in two different pointing modes:

- the **ON mode**, in which the telescope points directly to the target, so that the source is tracked at the center of the camera. In order to estimate the background, dedicated OFF observations need to be taken in sky regions where no gamma sources are expected, in the same conditions of Zenith Angle (ZA) and Night Sky Background (NSB: sources of ambient light during the night, such as the Moon);
- the **Wobble mode** [Fomin et al., 1994], in which the target is observed with an offset with respect to the center of the camera (for MAGIC, it is 0.4°). The advantage of this pointing mode is that there is no need to run dedicated OFF observations, since the background can be estimated using one or more regions in the camera - called Wobble positions - located on the circle of the same offset radius around the center of the camera, like illustrated in Fig. 2.6(a). This saves observation time, and assures a measurement of the background with the same exact

Figure 2.3: A picture of the MAGIC Telescopes. Credit: D. Lòpez.

telescope and sky conditions. The main drawbacks are the loss of detection efficiency due to the shifted position, and the problem of inhomogeneities in the camera that lead to systematic uncertainties in the background estimation. To overcome this latter issue, the source position in the camera is rotated around the center by an angle of 180° or 90° every 20 minutes (in MAGIC).

2.2 The MAGIC telescopes

The Florian Goebel Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) is a system of two IACTs built and mantained by an international collaboration involving over twenty institutions in Germany, Spain, Italy, Switzerland, Croatia, Finland, Poland, India, Bulgaria and Armenia. It is located in the Canary island of La Palma, on the Roque de Los Muchachos mountain at 2225 m a.s.l. The location is such that for most of the time clouds are below the site level, not disturbing observations. MAGIC I started operation in 2003, followed by his companion MAGIC II in 2009. The telescopes are equipped with 17 m diameter parabolic reflectors and cameras with a field of view (FoV) of 3.5 degree diameter and 1039 pixels made by photomultipliers (PMTs). The stereoscopic observations allow to reach a high angular resolution and sensitivity, and the large mirror area brings the energy threshold down to about 50 GeV. The design is highly technologic: the structure is made of carbon fiber tubes and the camera is held by two carbon-fiber archs, while electronics and readout system are located outside the camera, in a separate control house. In this way, the whole structure is kept solid yet lightweight, allowing telescope repointing of 180 degree in under 20 seconds.

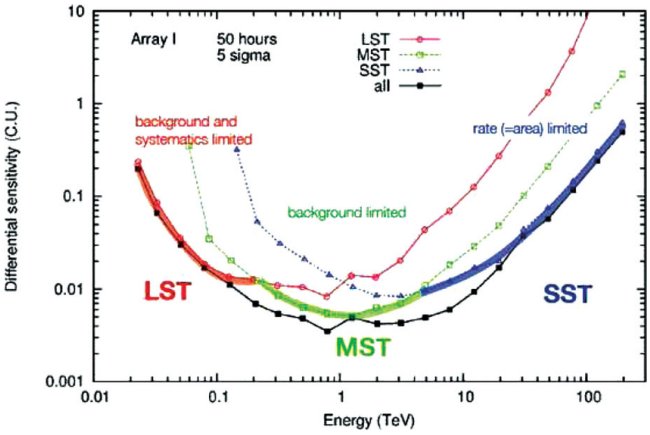


Figure 2.4: CTA differential sensitivity requirement in 50h of observation. The three types of telescopes will contribute in three overlapping energy ranges: LST (red) in the low-energy, MST (green) in the medium range, SST (blue) in the highest band. Taken from Acharya et al., 2013.

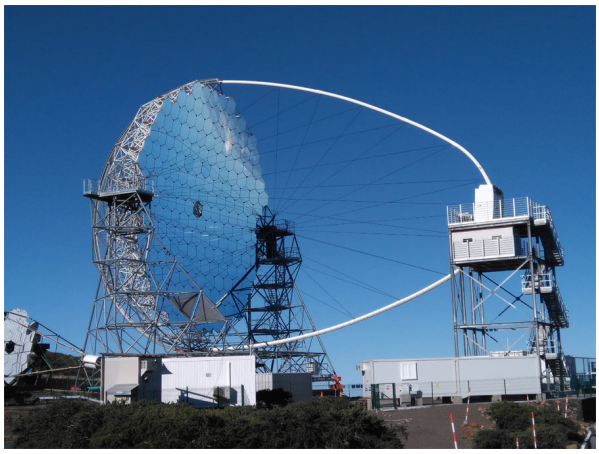


Figure 2.5: A picture of the Large Sized Telescope prototype (LST1) built in La Palma, Canary Islands. It has been inaugurated in 2018, detecting the Crab Nebula for the first time in November 2019 and the pulsated signal of the Crab Pulsar in June 2020. Credit: Ivan Jimenez

2.3 Future is now: CTA and its Large Sized Telescope (LST)

Cherenkov Telescope Array (CTA) is the observatory that will host the next generation of IACTs [Acharya et al., 2013]. It will be built in two sites, one in the Northern hemisphere, at La Palma (where MAGIC is also located), the other one in the Southern hemisphere, in the Atacama desert, in order to achieve full sky coverage. The northern array, consisting of 19 IACTs, will be optimized for the extra-galactic sources, while the southern, focused on galactic studies, will host 99 telescopes. The project has four main and ambitious aims:

- improve the sensitivity of a factor 10 in the 100 GeV - 10 TeV energy range (Fig. 2.4);
- extend the accessible energy range from few tens of GeV ($\sim$ 20) to 300 TeV;
- improve the angular resolution of about 5 times with respect to the current facilities;
- improve the temporal resolution: thanks to CTA's large collection area, time-variable and flaring VHE activities will be resolved on a sub-minute scale.

For pursuing this scope, in each site three types of IACTs will be deployed:

- **Large Sized Telescopes (LSTs)**, optimised to reach the lowest achievable energy thresholds. There will be 4 LSTs in both arrays working in stereo mode, in order to increase the collection area and enhance the signal/background separation power. Stiff carbon-fiber tubes hold a 23 m diameter reflector with 27.8 m focal length, making the structure of the telescope very solid and light-weight; such a design, similar to MAGIC, is needed in order to rapidly repoint the telescopes to follow up fast transients like GRBs. The mirror area for a single LST is of about 400 m^2. The camera is composed of 1855 PMTs, for a total Field of View of 4.5 degree diameter.

The first one, LST1, has been inaugurated in 2018 at La Palma (see Fig. 2.5). On November 23, 2019, the telescope was pointed towards the Crab Nebula for the first successful observing run, achieving a significance higher than 10σ with an exposure of 269 min[5] and on June 22, 2020 detected the VHE emission from the Crab pulsar with 5.2 σ. significance[6].

- **Medium Sized Telescopes (MSTs)**, designed for the aforementioned sensitivity improvement in the middle energy range, will be equipped with 12 m reflectors with 16 m focal length and 7-8 degree diameter Field of View, similarly to the concept of IACTs HESS and VERITAS;
- **Small Sized Telescopes (SSTs)**, smaller IACTs whose main purpose is to extend the energy range up to 300 TeV. Due to the low flux of the highest-energy gamma rays, a large number of them (70 in the Southern emisphere) will be needed in order to achieve the largest collection area possible.

The enhanced performances will boost the research in a number of topics, such as CRs acceleration and propagation, jetted-AGNs, Dark Matter annihilation searches, and many others.

2.4 IACT data analysis

In the context of IACTs, the event reconstruction consists in successfully fulfilling three tasks: **separate signal from background**, **reconstruct its energy** and its **arrival direction**. The current methods of analysis are based on parametrization of the images and in retrieving its properties by means of algorithms that operate in the space of these extracted parameters. The analysis of MAGIC data is performed through the `MARS` software package [Zanin, 2013], which is written in C and based on the `ROOT` framework. The analysis tools for LST1 are instead under development, and are written in Python; the high level analysis repository is `cta-lstchain`, based on `ctapipe`[7], which is the low-level data processing pipeline software for CTA. The essential scheme of the standard analysis pipeline is presented here below:

2.4.1 Low-Level data processing

1. **Calibration**: once an IACT is triggered, the analog signal of each PMT is sampled by a Analog to Digital Converter (ADC) and stored by the data acquisition system; from these raw data, signal is integrated and the total number of photons (*charge*) and their averaged arrival times (*peak times*) in the PMT are estimated. The calibration is performed using a calibration run;
2. **Image cleaning**: images are affected by the Night Sky Background (NSB) - namely all the sources of diffuse light of the night, like stars in the field of view of the IACT, or the moon - by dark noise of the PMTs and electronic noise of the readout system. Thus, most pixels contain noise, so images need to undergo a cleaning procedure. The standard one sets the values

[5] https://www.cta-observatory.org/lst1-detects-first-gamma-ray-signal/
[6] https://www.cta-observatory.org/lst1-detects-vhe-emission-from-crab-pulsar/
[7] DOI: 10.5281/zenodo.3372211

Image parameters

A typical γ-like event has an elliptical shape. During image parameter reconstruction, an ellipse is fitted to the clusters of pixels that survived the cleaning, in order to extract the so-called **Hillas parameters**, i.e. momenta of the fit up to second order containing meaningful information about the event. There are then other important parameters, containing other information on time properties and image quality.
Hillas parameters (some are shown in Fig. 2.6(b)):

- **Size/Intensity**: number of photo-electrons contained in the image. It is approximately proportional to the γ ray's energy (for a fixed zenith angle);
- **Width**: minor semi-axis of the ellipse, related to the lateral development of the shower;
- **Length**: major semi-axis of the ellipse. It measures the longitudinal development of the shower;
- **Center of Gravity (CoG)** of the shower image;
- **Conc(N)**: fraction of the total charge concentrated in the N brightest pixels, usually higher in the case of γ rays.

There are then two parameters related to the source expected position in the camera, also shown in Fig. 2.6(b):

- **Dist**: angular distance from the source expected position and the CoG. Linked to the impact parameter;
- **Alpha**: angle between the ellipse major axis and the line connecting the center of the ellipse with the source expected position;

Time parameters, useful for background rejection since the EM showers have a different temporal development with respect to hadronic showers:

- **Time RMS**: dispersion of the arrival times of the pixels that survive the cleaning;
- **Time Gradient**: angular coefficient of the linear interpolation to the projection of the arrival times on the ellipse major axis. It carries information on the direction of the shower development.

Image quality parameters:

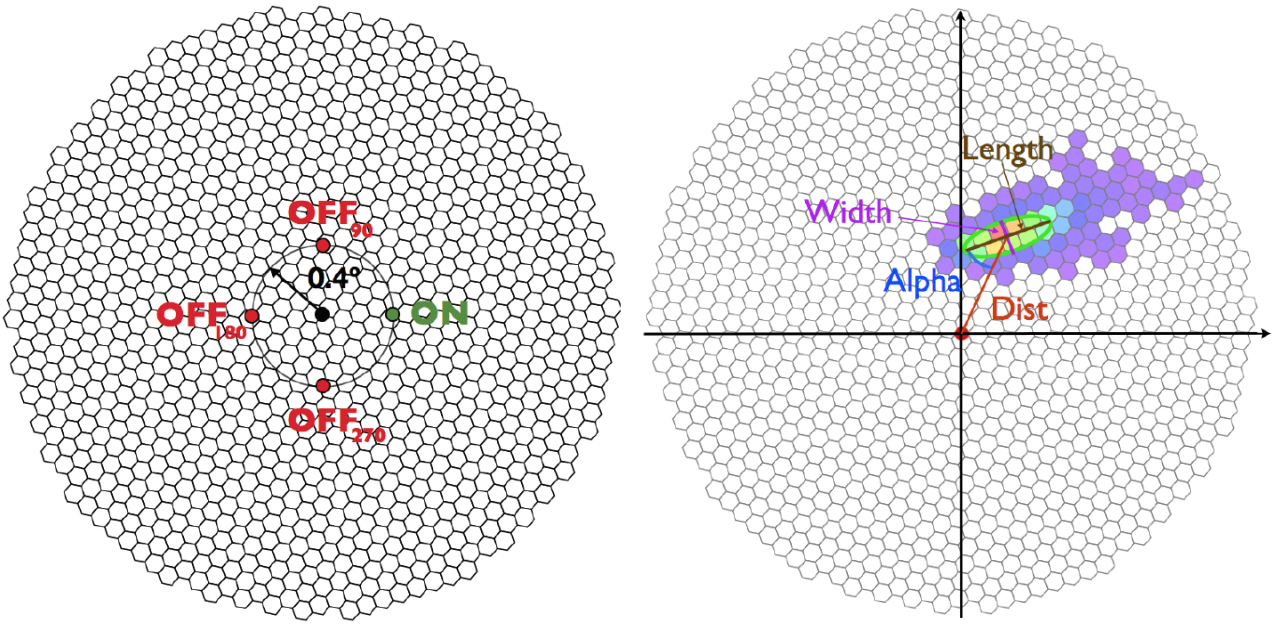
- **LeakageN**: fraction of the image size/intensity contained in the N outermost rings of pixels;
- **Number of islands**: an island is an isolated group of pixels that survives image cleaning.

Table 2.1: Image parameters fit to the shower images.

of two signal thresholds, i.e. number of *phe*: one for selecting the "core pixels", and a lower one to select the boundary pixels, in order to select only extended patterns (called *islands*) and get rid of isolated above-threshold pixels.

3. **Image parameters calculation**: then, image is parametrized and the information contained in the cleaned image is "translated" into the so-called Hillas parameters [Hillas, 1985] and other parameters, listed in Table 2.1.

4. **Stereo parameters calculation**: in the case of an array of two or more telescopes like MAGIC, there are more images of the same event, and the respective single-telescope parameters are merged;



(a) (b)

Figure 2.6: (a): the wobble pointing mode as seen in the MAGIC camera plane for one wobble. In this example 3 OFF regions are used. **(b)**: some of the Hillas parameters a camera of MAGIC. Images adapted from Lopez-Coto, 2015.

2.4.2 Intermediate data processing

In the case of MAGIC, the standard routine applicated by the analyzers usually starts at this point, while the previous steps are performed automatically by the On-Site analysis.

5. **Training of Random Forests (RFs)** and production of **Look Up Tables (LUTs)** for event reconstruction. The Random Forest method is the standard Machine Learning technique adopted in MAGIC [Albert et al., 2008] and LST1 analysis packages.
   For the γ/hadron separation task, a RF needs to be trained on *(a)* a sample of Monte-Carlo simulations of γ-ray shower events and *(b)* a sample of *real* hadronic events, acquired observing dark patches of the sky, where no gamma-signal is expected[8]. Then, a number of so-called *decision trees* is grown (ususally 100), leading to the estimation of a parameter called `hadronness` for each event, taking values between 0 and 1 and thus related to as the probability that the event is of hadronic origin.
   For direction and energy estimation, two RFs are trained only on MC gamma-events in order to learn how to retrieve the respective quantities.
   While RFs are used for energy reconstruction in single-telescope observations, the LUTs method is more efficient for stereo observations: the sample of (MC simulated) gammas is binned according to some meaningful parameters. For every bin, the average energy and its dispersion is

[8] In the case of LST1, MC simulated hadronic events are still used, instead of real ones.

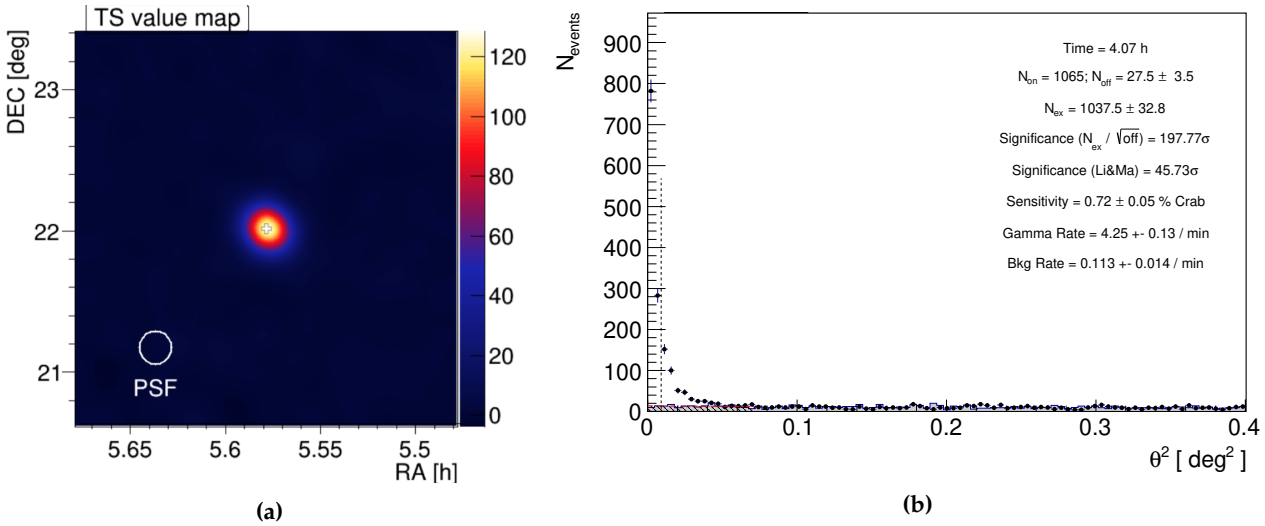


Figure 2.7: (a): Example of Sky Map of a pointlike source (the Crab Nebula). The circle represents the Point Spread Function (PSF) of the telescope to be compared to the source extension. TS stands for Test Statistics. (b) Example of θ^2 plot obtained with data of the Crab Nebula: the dashed vertical line defines the *signal region*.

calculated. Further in the analysis, the real events are binned in the same way, and their energy consequently evaluated.

6. **Event reconstruction**: application of the trained RFs and LUTs to the data to be analyzed.

2.4.3 High-Level data processing

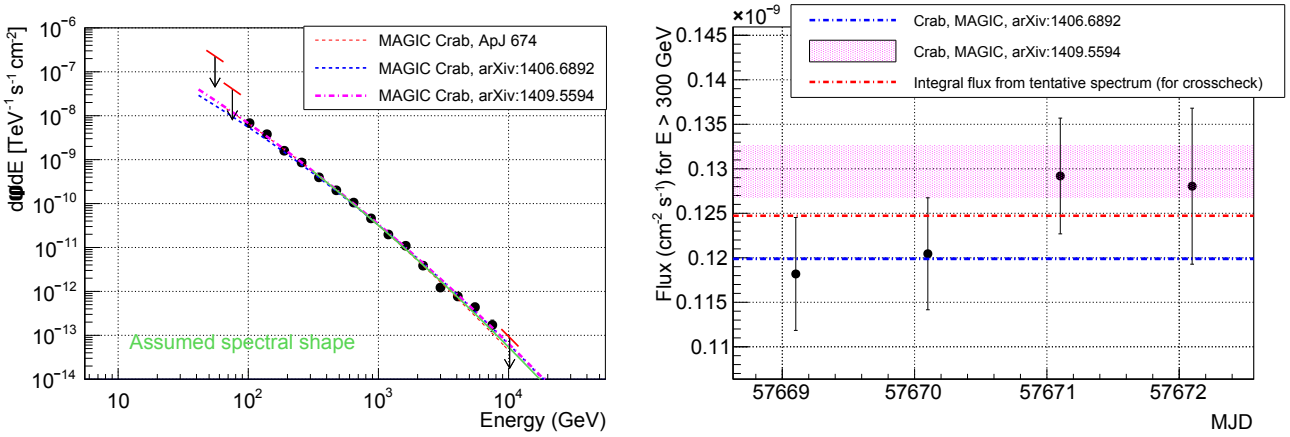
7. **Skymaps**: it is a 2-dimensional histogram of the arrival directions of the gamma-like events is produced, in sky coordinates. The background due to the hadronic component is subtracted, leaving a useful visual representation of the source, from which can be easily inferred whether it is pointlike or extended, or if there are other sources in the FoV. An example is shown in Fig. 2.7(a).

8. θ^2 **plot**: a histogram of the squared angular distance (θ^2) between the estimated shower arrival direction and the expected source position (see Fig. 2.7(b)). Cuts on some parameters (such as energy, `hadronness`, `size` or ZA) are applied to the data sample; in MAGIC analysis, three standard cut levels in `hadronness`, `size` and θ^2 define the three energy bands summarized in Table 2.2: Low-Energy, Full-Range, High-Henergy. Gamma rays coming from the source will have small θ^2 values, thus peaking at 0, the hadronic, residual background events will be almost uniformly distributed over the histogram.

   A signal region is defined by the *theta*^{2} cut, containing a number of events N_{ON}, to which is subtracted an estimation of the number of hadrons in the signal region thus obtaining the signal, i.e. the excess events N_{EX}. The background is estimated filling θ^2 plots with respect to

Level	`hadronness`	`size` (phe)	θ^2	$E_{\mathbf{reco}}$ (GeV)
LE	< 0.28	> 60	< 0.02	-
FR	< 0.16	> 300	< 0.009	-
HE	< 0.1	> 400	< 0.007	> 1000

Table 2.2: Standard cut levels used in MAGIC analysis.



(a) **(b)**

Figure 2.8: (a) Example of differential spectrum of the Crab Nebula; **(b)** example of light curve of the Crab Nebula, showing the integral flux as a function of time. The time bins are one-day long. MJD is a time calendar unit, standing for Modified Julian Date; its day zero is 17$^{\text{th}}$ Nov, 1858 at 00:00.

the OFFs position in the camera.The number of excess events $N_{\mathbf{EX}}$ is given by:

$$N_{\mathbf{EX}} = N_{\mathbf{ON}} - \alpha N_{\mathrm{OFF}} \tag{2.1}$$

where $\alpha = 1/N_{\mathbf{WobbleOFF}}$, i.e. the reciprocal of the number of OFF positions.

The significance of this signal is computed through a Likelihood Test Ratio [Neyman and Pearson, 1933] following a formula given by Li and Ma, 1983:

$$\sigma_{\mathrm{Li\&Ma}} = \sqrt{2N_{\mathrm{ON}} \ln\left[\frac{1+\alpha}{\alpha}\frac{N_{\mathrm{ON}}}{N_{\mathrm{ON}}-N_{\mathrm{OFF}}}\right] + 2N_{\mathrm{OFF}} \ln\left[(1+\alpha)\frac{N_{\mathrm{OFF}}}{N_{\mathrm{ON}}-N_{\mathrm{OFF}}}\right]} \tag{2.2}$$

In VHE gamma astronomy, 5σ is the minimum significance required to consider a source detected.

9. **Spectrum:** differential flux of VHE γ rays, i.e. number of photons per unit of area, time and energy:

$$\frac{d\Phi}{dE} = \frac{dN_\gamma(E)}{dA(E)\, dt_{\mathrm{eff}}\, dE} \tag{2.3}$$

being $dN_\gamma(E)$ the number of excess events in an energy bin dE (obtained with proper cuts adapted in order to reach a certain efficiency, much like illustrated in the previous paragraph on θ^2), $A(E)$ the collection area of the instrument and $t_{\mathbf{eff}}$ the *effective time* in which the source

has been observed. All these quantities need to be carefully computed.

- *collection area:* it is defined as the geometrical area around the telescope in which a γ ray triggers the instrument. In order to compute it, a sample of N_{sim} MC γ rays is needed. The expression in a given energy range is given by:

$$A_{\text{coll}}(E) = A_{\textbf{sim}} \frac{N_{\text{sel}}(E)}{N_{\text{sim}}(E)} \tag{2.4}$$

  $A_{\textbf{sim}}$ is the simulated area, $N_{\text{sim}}(E)$ the simulated photons in the energy bin and $N_{\text{sel}}(E)$ the simulated photons that survived the same aforementioned cuts applied to the real data. This expression holds only if the MC gammas are produced with the same VHE spectrum as the source (which is often unknown, but can be estimated through proper fit of common spectral shapes).
  Collection area has typical values of $\sim 10^4 - 10^5$: it depends, other than energy, also on ZA: for smaller values of ZA, we have smaller Cherenkov light pools and higher photon densities, and, conversely, for higher ZA values, larger light pools but lower photon densities. In this latter case, the light has also to travel a longer path in the air, undergoing a higher atmospheric absorption. For this reason, low-energy events are lost for the highest ZA, resulting in a wider collection area but a higher energy threshold.

- *effective time:* this is *not* the elapsed time from the beginning to the end of observation, because everytime an event triggers the telescope, a dead time d is needed by the readout system for the acquisition ($d = 26\mu s$ for MAGIC). Thus, the effective observation time is given by:

$$t_{\text{eff}} = \frac{t_{\text{elapsed}}}{1 + \lambda d} \tag{2.5}$$

  where λ is the event rate.

10. **Light curve:** it is the evolution of the integral flux of photons, selected in a given energy range[9]. over time. Basically, it is the integral of the differential flux over the energy interval:

$$\Phi = \int_{E_1}^{E_2} \frac{d\Phi}{dE} dE \tag{2.6}$$

    One of its main purposes is to study the variability of sources' emissions in time.

11. **Unfolding:** due to the uncertainties introduced by instrumental limitations and statistical analysis, the reconstructed fluxes needs to be carefully corrected through an unfolding procedure. It is a technique to transform the distribution of the estimated energy $R(E_R)$ back into the distribution of the true energy $T(E_T)$, very important in order to get reliable estimations of spectra,

[9] For MAGIC, usually $E > 300$ GeV (150 GeV for distant sources).

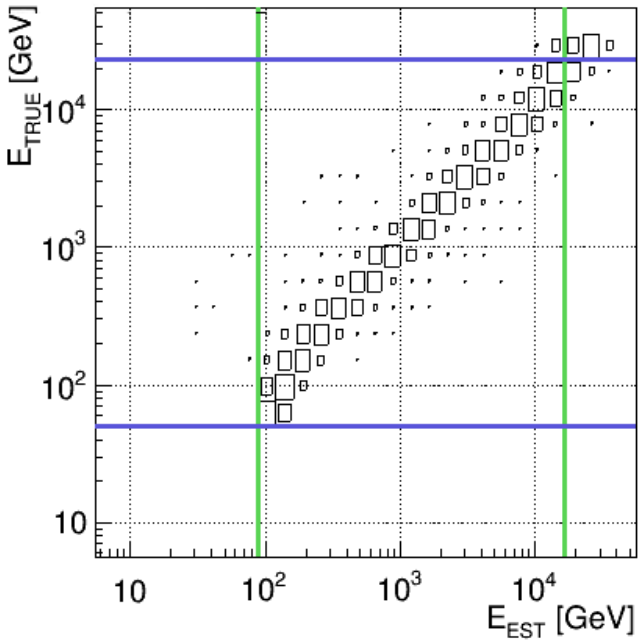


Figure 2.9: Example of migration matrix. The area of the rectangles represents the number of MC events contained in that bin.

fluxes and SEDs for real observations. Binning the two distributions in E_{RECO} and E_{TRUE} respectively, the relation between the two quantities can be written:

$$R_i = \sum_j M_{ij} T_j + B_i \tag{2.7}$$

where B_i is a parameter related to background, and the tensor M_{ij} is the so called *migration matrix* estimated on the simulated MC γ rays, of which we have an example in Fig. 2.9. The unfolding operation aims at inverting the migration matrix; several algorithms have been developed and are now used in MAGIC analysis, like described in Bertero, 1989, Tikhonov and Arsenin, 1977 and Schmelling, 1994. Another method is used in MAGIC, called *forward unfolding*, and it is very robust. Technically, it is not an unfolding procedure, but a fit: a spectral shape for the observed source is assumed, and the best parameters for that function are calculated by means of the least squares method. It returns the best spectral parameters but *no* spectral points.

Chapter 3

Deep Learning and Convolutional Neural Networks

Although the CTA concept contemplates two arrays of tens of telescopes working in stereo mode, at the moment only the LST1 prototype is at operation in La Palma, so analysis tools dedicated to single-telescope event reconstruction are under development. The Random Forest method, although being robust and reliable, works on the space of the Hillas parameters, and does not have access to the additional information potentially contained in the images and washed out during parametrization, like pixel-wise information. This motivates the testing and improvement of modern, state-of-the-art Deep Learning analysis techniques, in particular Convolutional Neural Networks (CNNs) that can exploit those hidden features. DL is explored in frontier research topics, such as speech recognition, facial recognition, medical imaging, and there is a growing interest in the particle physics community, which makes extensive use of these techniques for data analysis [Belayneh et al., 2020; Bourilkov, 2019; Guest et al., 2018]. Here follows a quick review of Deep Learning and a survey of the tools I employed in my analysis.

3.1 Deep learning

Deep Learning (DL) is a branch of Machine Learning, and it refers to a class of algorithms that mimics the mechanisms of the human brain involved in decision making and estimations of quantities. These algorithms, also referred to as Artificial Neural Networks (ANNs), are able to autonomously learn how to extract information from raw data, deciding by themselves which features or patterns of the dataset are meaningful for the task addressed. The term *deep* is referred to the use of multiple layers stacked one on the other, called hidden layers, each representing a different level of abstraction.

In this work, I used the approach of *supervised learning*, that is to train the algorithm providing the *labels* along with the input data, namely the real values we want it to estimate. At its most basic level, it is a minimization task: given a **batch** of the input data x, a functional returns the predictions $\boldsymbol{s}$, or *scores*, of the values we are interested in:

$$s = F\big[W, x\big]$$

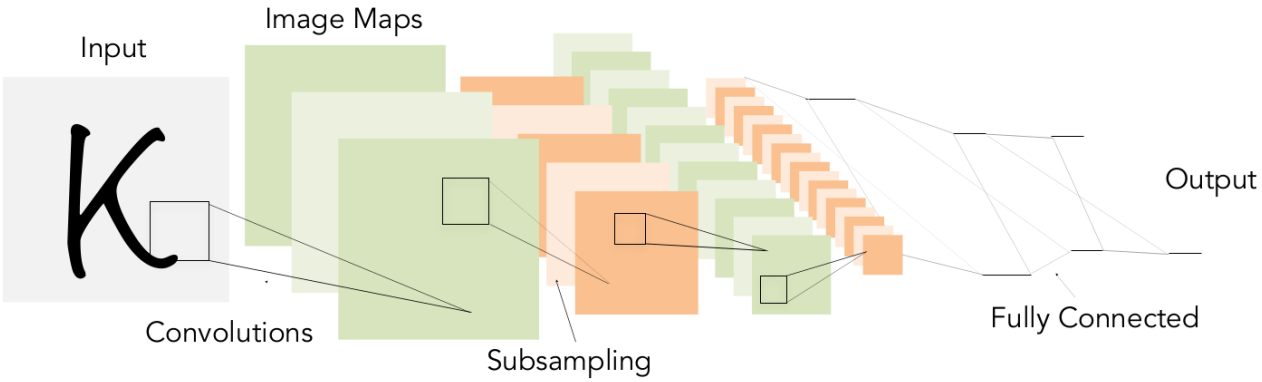


Figure 3.1: A scheme of how a Convolutional Neural Network works. Many convolution and subsampling operations extract relevant features from the input image and its image maps, leading to the final prediction (output).

where W represents the **weights**, i.e. the parameters contained in the hidden layers. Who decides how good the algorithm is performing is a **loss function** $\mathrm{L}(\boldsymbol{s}, \hat{\boldsymbol{s}})$ measuring how far the predicted score values $\boldsymbol{s}$ are from the true values $\hat{\boldsymbol{s}}$: the closer they are, the smaller is $\mathrm{L}(\boldsymbol{s}, \hat{\boldsymbol{s}})$. Hence, we want to minimize it. To do that, after each iteration, the weights W are properly updated through a back-propagation algorithm.

What we try to do with Deep Learning may resemble "fitting the data", but it is actually something different. The key word is learning: while "to fit" is all about finding the values of the parameters that best reproduce the input data - indeed, fit the data - "to learn" means instead to find the parameters that give the machine the ability of making correct predictions on previously unseen data.

3.2 A quick dive into CNNs

Convolutional Neural Networks [LeCun et al., 1998] are a particular tool of DL that show their full power in image recognition. They have grown very popular in the very recent years, after the network called AlexNet won the ImageNet Large Scale Visual Recognition Challenge in 2012 [Alom et al., 2018; Krizhevsky et al., 2012]. CNNs take images as input, or generally speaking rank-3 tensors of shape $h \times w \times d$, where $h \times w$ is the number of pixels of the image, and d is the depth (usually 3, corresponding to the 3 RGB channels). What makes them so peculiar among other techniques is that they have access to the spatial information of the input tensor. The building block of a CNN is the **convolutional block**, composed of:

1. a **convolutional layer**: basically the convolution between the image and a number of small, square filters ($l \times l \times d$ tensors), which contain the weights to be trained;
2. a **pooling layer**, that selects the most important features by subsampling the image, also reducing its size (for computational economy's sake);

3. an **activation function**: a non-linearity applied before the next convolutional block. Typical functions are tanh, the sigmoid function and the ReLU $\equiv \max\{x, 0\}$[1].

The filters in the conv-layer can be looked as feature detectors with a receptive field walking all across the image, performing a sort of dot product at each step and returning as output a resized representation of the picture (the *feature map*). This is what is called *local connectivity*: the neurons connecting the input layer to the feature map are looking only on the respective receptive field. Moreover, the fact that the weights of a given filter do not depend on the position on the image, but are instead shared by each neuron, assure that the whole operation is *translation-invariant*.

Many convolutional blocks (let's call them layers from now on) are stacked on each other, usually adding at the end Fully Connected layers (the so-called decision layers) that output the predictions: this is the basic structure of CNNs. Depth is one of their core features, and in the recent years the deepest and most sophisticated architectures have been developed, reaching outstanding results in image recognition, like VGGs [Simonyan and Zisserman, 2015], that reach up to 19 layers, Deep Residual Networks (ResNets [He et al., 2015]) and Densely Connected Convolutional Networks (DenseNets [Huang et al., 2018]) that can feature hundreds of layers.
VGG is a popular network using very small receptive fields in the convolutional layers (3×3), with a convolution stride of 1, and three Fully Connected (FC) layers at the end; between each layer, the ReLU activation function is used.

Putting it down naively, a CNN is thought to behave like this: the first layer is likely looking for the basic features like corners and edges, the second arranges these features in more complex shapes, and so on until the last layers that represent all the information into more abstract categories (something like "car-ness", "human-ness" - in our case, it will be `gammaness`, `energy`, `direction`...). If there was no activation function (non-linearity) between each layer, the whole network would be equivalent to a single big linear system, thus losing the ability to build such a hierarchical representation.

3.2.1 The training process

The dataset is split in three, namely a *training set*, a *validation set* and a *test set*. The algorithm sees the whole training set a number of times, called *epochs*, during which the weigths are adjusted. After each epoch, the network is tested on the validation set to monitor the performances on unseen data, and the training images are shuffled for the next iteration. Once all the process completes, the model is benchmarked on the test set.

The optimization of a CNN is based on the gradient of the loss function $\nabla \mathrm{L}(s, \hat{s})$ with respect to the weights W. The baseline method is the Stochastic Gradient Descent (SGD), that updates the weights towards the direction opposite to the gradient evaluated on the current batch of data:

$$W_{n+1} = W_n - \lambda \nabla \mathrm{L}(W_n)$$

[1] ReLU = **Re**ctified **L**inear **U**nit

Figure 3.2: 1-D representation of the impact of different values of LR on the minimization of a loss function $J(\theta)$. Image courtesy of J. Jordan.

where the step-size is controlled by the λ parameter, called **learning rate**. This is one of the most decisive parameters, and has to be carefully tuned in order to optimize the learning process, as shown in fig. 3.2.

3.2.2 Optimization: the Adam algorithm and Cyclical Learning Rate policy

Many refined optimization strategies have been developed, among which I used the popular *Adam* (ADAptive Moment estimation) algorithm, which is an extension to SGD. While the latter makes use of a single fixed LR value during the whole training time, *Adam* starts from an initial value and updates multiple learning rates based on the recent training history, or more specifically it "computes individual adaptive learning rates for different parameters (weights) from estimates of first and second moments of the gradients" [Kingma and Ba, 2017].

Coupled with *Adam*, I also adopted the *CLR* (Cyclical Learning Rate) policy [Smith, 2017], which periodically varies the initial learning rate of each epoch between two user-defined boundary values. Cycle's length is usually set between 3 and 8 epochs. The advantages of using *CLR* are that *(a)* it solves the annoying problem of finding the perfect learning rate, *(b)* it speeds up the training by making use of higher LR values [Smith and Topin, 2018] and *(c)* it makes the model explore different minima of $L(s, \hat{s})$, in fact obtaining a number of different network configurations with just one training [Huang et al., 2017] (fig. 3.3). *CLR* has also been shown to have a generalization effect (see next paragraph).

Another technique often used in the neural network field is *Batch Normalization* [Ioffe and Szegedy, 2015]: usually inserted before the non-linearity after a convolutional or FC layer, is used to reduce the dependence on weight initialization and make each dimension at some layer unit gaussian. Considering a batch of data, the following normalization is applied to the outputs at some layer:

$$\hat{x} = \frac{x - E[x]}{\sqrt{\text{Var}[x]}} \tag{3.1}$$

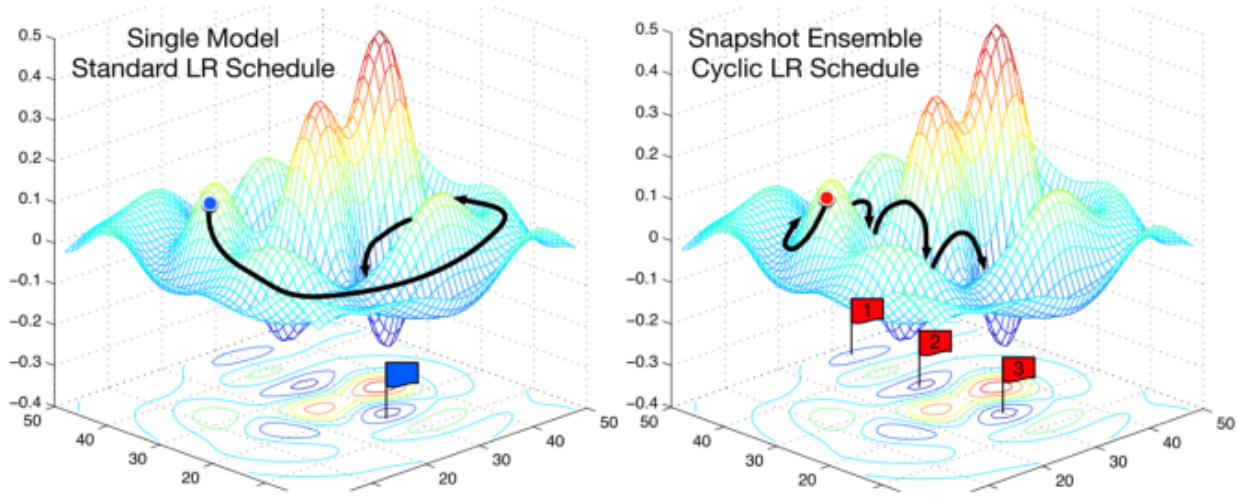


Figure 3.3: Visual comparison between standard optimization process and cyclical learning rate schedule. CLR allows the model to explore different minima of the loss function.

where $E[x]$ and $\sqrt{\text{Var}[x]}$ are the mean and the standard deviation of the layer units, and then the network is allowed to rescale and shift the range:

$$y = \alpha\hat{x} + \beta \tag{3.2}$$

where α and β are two free parameters to be learned. In this way, the network can also recover the identity mapping. I will use this technique at the beginning of our model, in order to let it arbitrarily decide if standardize the input images, depending on the task addressed.

3.2.3 The problem of over-training and regularization techniques

If not properly adjusted, CNNs tend to "memorize" the training set, specializing in reproducing its labels but not improving the actual ability of making correct prediction on unknown images - many times, they even lose it. This scenario is called *over-fitting*.

One idea to prevent this complex co-adaptation of the weights to the training data is *dropout* [Srivastava et al., 2014], that randomly switches off (or drops out) a given percentage of nodes (neurons) at each training example or at each mini-batch, preventing them to update. Also the previously discussed CLR, involving high learning rates, perturbs the training with some kind of "noise", helping the optimizer escaping and explore other local minima.

Over-training can be avoided also by means of other regularization techniques that aim at constraining the network parameters, like *L1* and *L2* norms, also called weight decay, that introduce penalties for too large weight values adding regularization terms to the the loss function, namely $L1 = -\eta \sum_i |W_i|$ and $L2 = -\eta \sum_i W_i^2$, $\eta \in [0, 1]$.

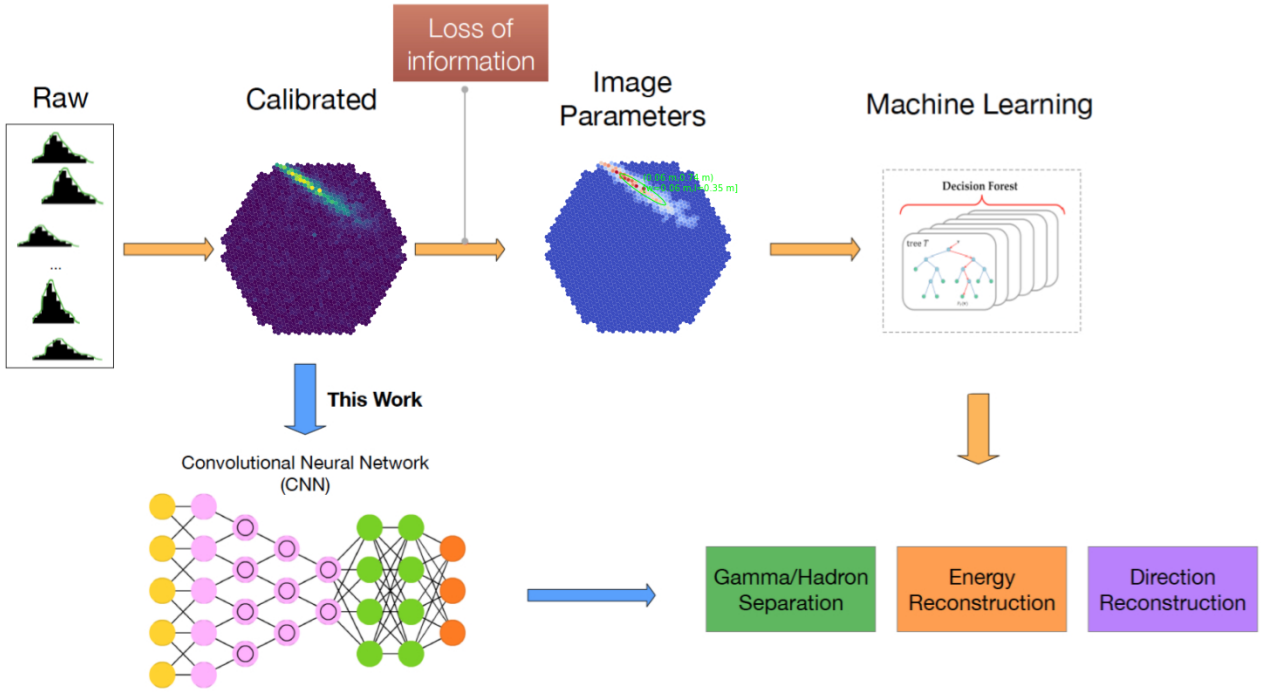


Figure 3.4: Diagram comparing the standard analysis with our approach. The aim of work is to perform IACT event reconstruction using CNNs, exploiting pixel-wise information and hidden features washed out by image parametrization.

3.3 LST event reconstruction with CNNs

Recalling chapter 2, the flux of incoming γ rays is dwarfed by the flux of charged CRs with a signal-to-noise ratio of 1/1000. Thus, in order to do VHE astronomy with IACTs, one has to be able to pick the γ rays coming from a certain direction among the overwhelming sea of hadrons, and evaluate their energy in order to study the spectral properties of the source of interest. Summarizing, event reconstruction for IACTs is decomposable in three parts:

1. **signal/background separation**, a *binary classification* task. Each input event is given a score called `gammaness`, i.e. the probability to be γ-induced;
2. **energy reconstruction** of the γ-initiated events, a *linear regression* task with one output;
3. **direction reconstruction**, a *linear regression* task with two outputs, i.e. the coordinates of the source of the arrival directions of the γ rays.

As already mentioned, the novel approach using CNNs aims to accomplish these tasks starting directly from the raw images, skipping the parametrization step, as illustrated in Fig. 3.4.

3.3.1 Our setup

A number of different network models from the repository `keras.applications` has been tested, among which the VGG16 and VGG19 showed the best and most stable results in all the three jobs. I therefore implemented a slightly tweaked VGG, composed of 13 convolutional layers and no FC layers, for a total of $\sim 1.4 \cdot 10^7$ parameters. The last activation function is different depending on the task: a *sigmoid* for the classification network, namely $S(x) = (1 + e^{-x})^{-1}$, and a *linear function* for regression. In Appendix A the final model architecture can be found. Usually, it's a best practice for Neural Networks to normalize and center the input images, however this is not the case for our data: our pixels are not bound between the canonical 255 values of RGB images, but contain instead precious information such as *charge* that is roughly proportional to the shower energy. So, in order to let the network decide whether or not to perform a standardization of the input, a Batch Normalization layer has been added right after the input and before the convolutional blocks. In this way, I aim to give our models more flexibility.

Computationally speaking, the training process (convolution, backpropagation...) is quite simple, but very expensive. Luckily, Graphics Processing Units (GPUs) massive parallelize the calculations and reduce the training time. This work was carried out using the INFN's cloud computing service Cloud Veneto - Area PD, joined with a Quadro RTX 6000 GPU. The analysis tools were developed using the Python API Keras 2.25 [Chollet, 2017], running on top of the Tensorflow 1.12 backend [Abadi et al., 2016]. Details and results of the analysis are presented in Chapter 4.

3.3.2 Related works: recent applications of DL to IACT analysis

CNNs and other DL methods are being widely explored to tacfkle the different problems of IACT data analysis. Nieto et al., 2017 investigate the ability of state-of-the-art CNN architectures in classifying simulated events as detected by an array of CTA. Parsons and Ohm, 2020 address the same task for stereoscopic H.E.S.S. observations (both MC and real) using Recurrent Neural Networks. To quantify the effects of the adding more image information with respect to the standard parameter-based analysis, the cleaned images are concatenated with the parameters in the input of the model, obtaining a 30-40% improvement in background rejection.

Shilon et al., 2019 apply CNNs and Convolutional Recurrent Neural Networks (CRNNs) to signal/background separation and direction reconstruction for the four H.E.S.S. telescopes, obtaining a significant improvement in background rejection and similar performances in angular regression with respect to traditional techniques. They also demonstrate two different methods to feed the data to the networks: by resampling images to square grids, and using custom convolution kernels adapted to the lattice of the camera images. A full event reconstruction for the MAGIC telescopes with CNNs has been performed by Mariotti, 2019, improving energy resolution for energies above $\sim$ 1 TeV and angular resolution up to 20%. A similar work for LST has been carried out by Marinello, 2019, who successfully explores the performances of different CNN architectures, however the dataset used was unrealistic under two main aspects: the first is that it was composed of events that triggered *at least* 2 out of the 4 LSTs in the array, whereas the telescope in operation is

only one; the second is an issue related to the simulated gain of the electronics, which led to a saturation of all the signal containing more than 100 phe. In this thesis is presented for the first time such analysis performed on data produced for single-telescope studies, with events triggering at least one telescope; the groups of INFN - University of Padova, Laboratoire d'Annecy De Physique Des Particules (LAPP) and Universidad Complutense de Madrid (UCM) are currently working for improving the performances of CNNs in the full event reconstruction for LST.

3.4 Handling the dataset

When training a neural network, knowing and understanding the available dataset is of utmost importance, as much as learning strategies and architecture choice, if not more. To achieve its best performance a CNN needs to have access to the largest possible amount of image samples during training.

The work of this thesis is performed on simulated shower images, produced in April 2019, for the first time focusing on single-telescope LST studies. They are not fully representative of the real observed data, as the technical conditions of the LST1 had still to be fully determined at that time. The data production was performed by Y. Ohtani, using the `CORSIKA` program [Heck et al., 1998] with the `IACT/ATMO` extension for the atmospheric simulation, while the whole mirrors-electronics-readout response was simulated with `sim-telarray` package. The shower images belong to three classes of particles:

- **protons:** $\sim 9 \cdot 10^5$ events generated in a view cone of 10° around the telescope pointing position, with energies spanning from 10 GeV up to 100 TeV;
- **diffuse γ rays:** $\sim 2 \cdot 10^6$ events generated in a view cone of 10° (see Fig. 3.6), with energies spanning from 5 GeV up to 50 TeV. They are used during training time;
- **pointlike γ rays:** $\sim 2 \cdot 10^6$ events generated with an offset of 0.4° with respect to the telescope pointing posistion, with energies spanning from 5 GeV up to 50 TeV. They are used during testing time.

The point in having two different event types for training and test is that the LST (and IACTs in general) points to the observed sources, thus the VHE γ rays are supposed to come from a precise direction in the sky[2], and it is under *this* condition that we want to test the reconstruction power of our analysis. Nevertheless, training the network directly on pointlike γ rays would introduce a bias in the learning process towards events that point to a specific direction, virtually losing the ability of recognizing and reconstructing γ events coming from a different source in the FoV, while we want the models to infer gammaness based on shape and development properties of the shower image.

Each image is composed of two channels: one is the *charge* image, in which every pixel carries the number of photoelectrons collected by the PMT, the other one is the *time* image, in which each pixels contains the time information extracted during calibration (see 2.4.1).

[2]We are considering only pointlike sources.

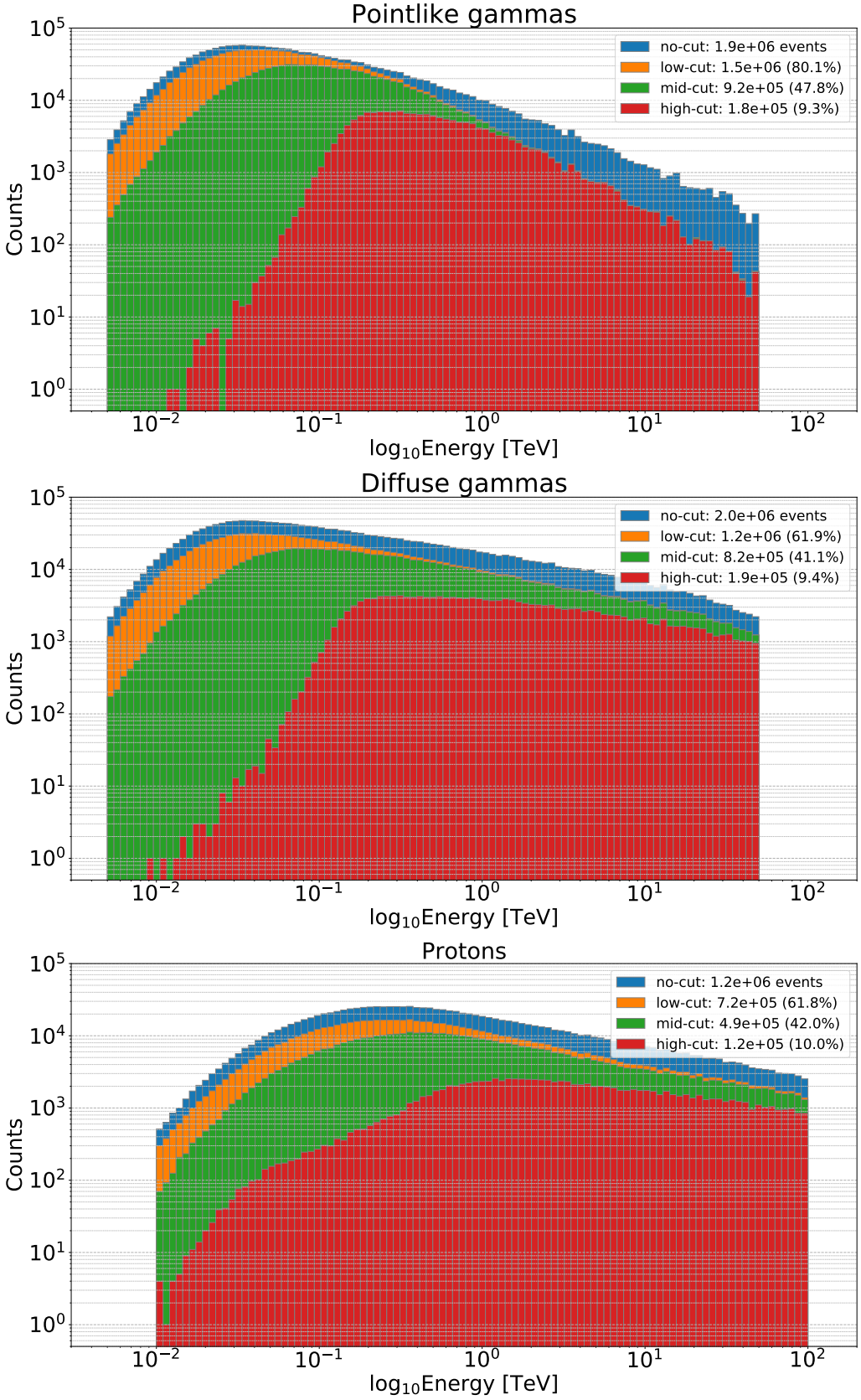


Figure 3.5: Energy distributions of the three data types: higher cut-levels translate into higher energy thresholds. Note that, although having different spectra, *protons* and *diffuse* γ rays are affected by the quality cuts almost in the same way (percentage-wise).

Some remarks

Since the production of the higher-energy events is computationally expensive, to come at terms with resources economy the simulations are performed in such a way that the same amount of energy is simulated within each energy decade, rather than the same number of events. Even though the events are not uniformly distributed over energy (which in principle would be preferable, as for the best training conditions dataset should be homogeneous, ideally speaking), this will not constitute an issue for our analysis: image quality has a much higher impact than these differences in our case, and I do not expect the best performances to be in the lower-energy bands (few GeV - few hundreds GeV), despite being the most represented (fig. 3.5).

Another point worth mentioning is that the dataset is composed of events that triggered *at least* one LST. Having 4 of them in the array, it means there can be up to 4 different images for each event that will nevertheless be considered as independent events - again, due to computational cost optimization.

3.4.1 Pre-processing of the data

The state-of-the-art Python API for CNNs, such as Keras, Tensorflow or PyTorch, rank 3 tensors as input, but the LST camera is hexagonal, composed of hexagonal pixels, so we need to rearrange the image to a rectangular lattice.

In Shilon et al., 2019 two different approaches have been proposed for image pre-processing: re-sampling methods, such as rebinning and interpolation, and modified hexagonal convolution kernels aimed at preserving the camera grid properties. Each method has its strenghts comparing with the others, such as low computational cost (hexagonal kernels), intensity conservation (rebinning) and shape conservation (interpolation). Finding which method represents the best trade-off would be a good subject/matter of investigation in sight of the definitive LST (or even CTA) analysis pipeline. We used the interpolation method, fitting each channel of the images with 2D-cubic splines, and then resampling the function onto a rectangular grid. The sampling grid step has been set to half the distance between the centers of two adiacent camera pixels (see Figure 3.7). In the end, the camera area is wrapped with the least possible zero-padding, resulting in a 3-rank tensor of 88×96 pixels and 2 channels[3]; an example is shown in Fig. 3.8. The images are then stored in HD5F files together with all the events information (MC energy, direction, class, `leakage2`...).

3.4.2 Data selection

The dataset comprises many low-quality images: several with low `intensity`, especially at the lower energies, and many others with the shower image not contained in the camera - indeed selecting images with a `leakage2` < 0.2, we lose roughly the 30%! Hence, it does make sense to perform some quality cut - but how much? In this work, I investigate the impact of four different cut levels on

[3] A further step in optimization would be to rotate the hexagonal camera of an angle that maximizes the area ratio $A_{\text{camera}} / A_{\text{rectangle}}$.

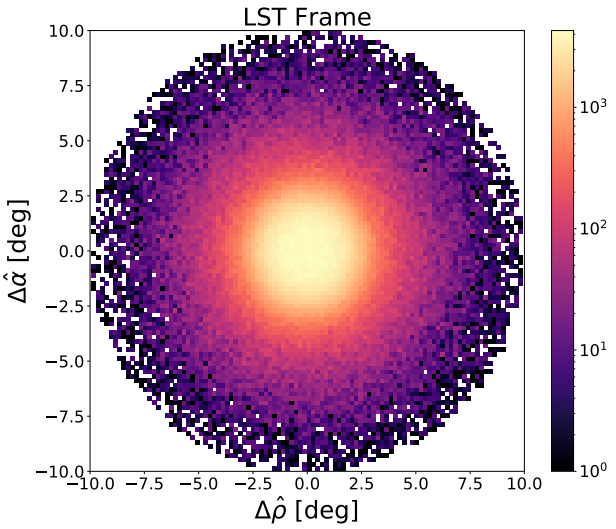


Figure 3.6: 2D histogram of the diffuse γ rays' arrival directions, in the telescope's alt-azimuthal frame of reference (α: alt, ρ: az).

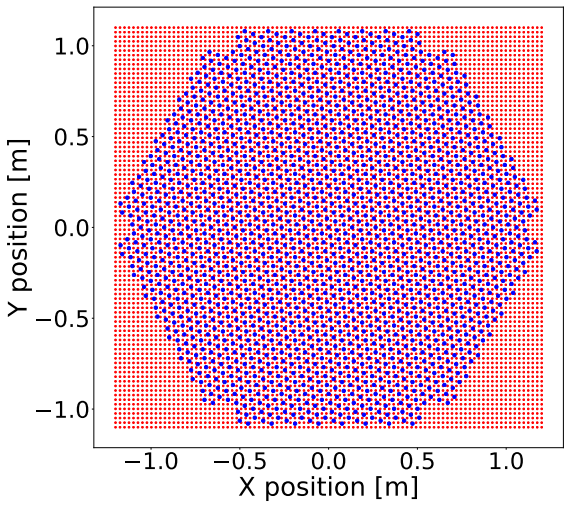


Figure 3.7: Superposition of the camera pixels position (blue points) and the resampling grid (red points).

the network performances, cutting on `intensity` and `leakage2`: `no-cuts`, `low-cuts`, `mid-cuts` and `high-cuts`:

Cut level	Intensity threshold	Max Leakage2	Energy threshold (`diffuse`)
`no-cuts`	`None`	`None`	13 GeV
`low-cuts`	50	0.2	14 GeV
`mid-cuts`	200	0.2	26 GeV
`high-cuts`	1000	0.2	169 GeV

Table 3.1: The four cut levels adopted in the analysis.

The effect of the different cuts on the dataset can be observed on the event energy distributions, shown in Fig. 3.5: a higher `intensity` threshold translates into a higher energy threshold.

3.4.3 Data management: the generator

During the training, images are usually loaded in the Random Access Memory (RAM); however, the whole interpolated dataset consists of several gigabytes, and it would not fit in RAM, so a custom Python data generator has been adapted from [Afshine and Shervine, 2018]. It performs an indexing of all the images, along with all the useful information (particle type, MC energy and direction, intensity, leakage2...) and stores the result in a *pickled Pandas dataframe* (a structured file) for future use. This allows Keras to load a given number of training/testing batches at a time and continuously hand it to the GPU. The generator handles all the operations that concern data management: selection (e.g. applying the cuts on intensity and leakage2), shuffle the images at each epoch, give the user access to data information and so on.

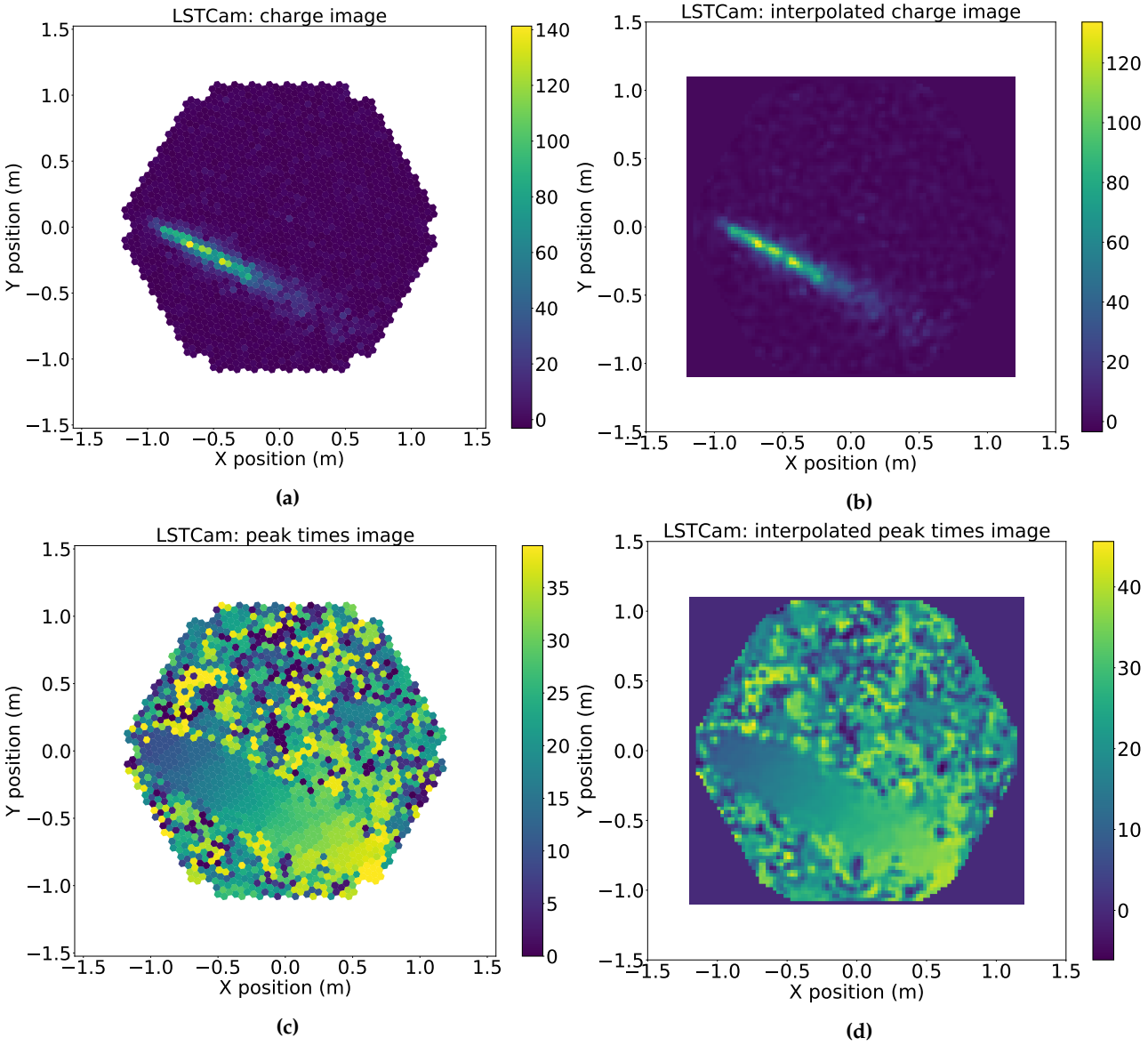


Figure 3.8: Camera event images of a γ-ray shower (left) and their interpolated versions (right). (a) is the charge image of the camera, in which each pixel contains the photoelectrons collected by the respective PMT (c) is the time image of the camera, containing the time information of each pixel; (b) and (d) are the obtained by fitting the camera images with 2D-cubic splines and resampling on a 88×96 rectangular lattice.

Chapter 4

Analysis and results

In this chapter I present the results of the full event reconstruction for the LST using Convolutional Neural Networks, that consists, again, on **separating** γ rays from hadron events and reconstructing their **energy** and **arrival direction**. For each of the three tasks, I used a **VGG-style network of 13 layers (VGG13)** described in Appendix A. The networks have been trained and tested on different selections of data, according the cut levels defined in 3.4.2; the performances have been then confronted with the ones obtained using the standard Random Forest method. Finally, in order to assess the quality and reliability of the work, at the end of the chapter I show the comparison between my results and the results by *CTLearn* and *GammaLearn*, two other research teams that are carrying out a similar analysis.

4.1 Signal/background separation

In a classification task each class needs to be equally represented, so a set of $\sim 9 \cdot 10^5$ protons and an equal selection of diffuse γ rays[1] was prepared, for a total of $\sim 1.8 \cdot 10^6$ events with a training/validation split of roughly 90/10. One concern was that some unbalancing may have been introduced by applying each cut; luckily, looking again at Fig. 3.5 we can see that the two sets remain even in terms of number of events, although being different in terms of energy distribution.

Each training lasted for 50 epochs, during which the networks were fed with batches of 128 events and regularized with a dropout rate of 0.5 and L2 weight decay of 10^{-5}. The optimizer was *adam* with a *learning rate* scheduled with the CLR policy introduced in 3.2.1, ranging from $5 \cdot 10^{-5}$ up to 0.005 with step size of 3. The loss function to minimize chosen for this task is the **binary cross-entropy (BCE)**, commonly used in binary classification. Given a *batch* of N events, the loss on that batch is computed like:

$$\mathrm{L_{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\hat{s}_i \ln s_i + (1-\hat{s}_i)\ln(1-s_i)\right] \tag{4.1}$$

where s_i is the estimated `gammaness` for the i-th event, and $\hat{s}_i$ is the true label value, namely 1 if it is a γ ray, 0 if it is a proton.

[1] Proton set is less numerous because protons trigger the system much less often than γs.

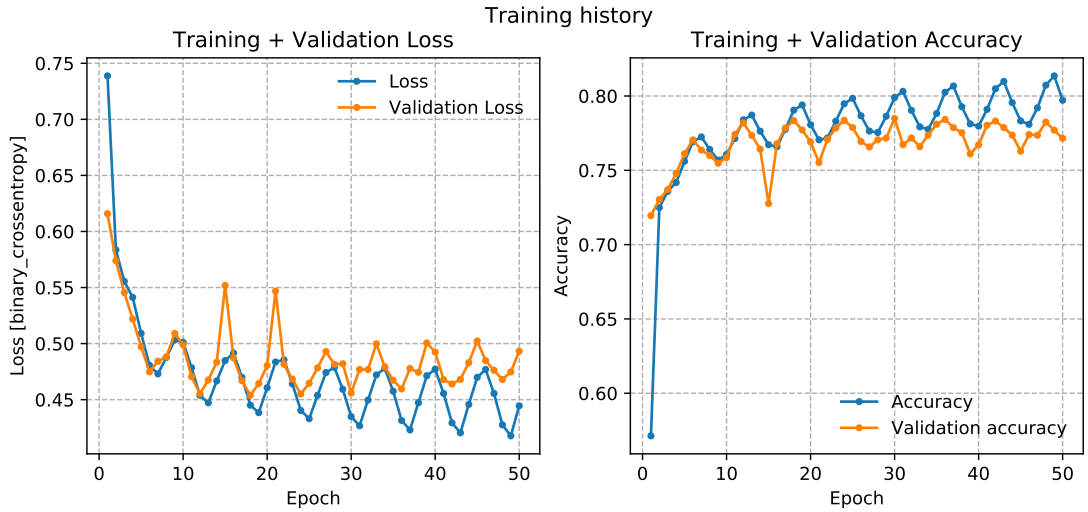


Figure 4.1: Training history of the `low-cut` VGG13. The wavy behaviour is an effect of the CLR policy.

cut level	train loss	valid loss	train accuracy	valid accuracy
`no-cuts`	0.498	0.532	0.762	0.743
`low-cuts`	0.427	0.477	0.799	0.785
`mid-cuts`	0.375	0.391	0.848	0.838
`high-cuts`	0.175	0.264	0.957	0.925

Table 4.1: Losses and accuracy reached by the models on maximum validation.

Given a `gammaness`-threshold, events above that value will be classified as γ rays, or as protons otherwise. The metric adopted to monitor training is the classic *binary accuracy*, defined as the ratio between the total correct predictions and the total number of predictions with a threshold of 0.5:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

where T/F stand for True/False and P/N for Positive/Negative.

In Table 4.1 are reported the values of loss and accuracy reached by the maximum-validation-accuracy models for the different cut levels, and in Fig. 4.1 is depicted the training history of the *low-cut* model. Once the trainings were completed, the models with the higher validation accuracy were tested on a total set of $2.6 \cdot 10^5$ protons (uncut value) and an equivalent set of *pointlike* γ rays. The test sets needed to be balanced for each cut level, since the cuts in this case have different effects on the two populations (Fig. 3.5). To evaluate the performances during test time, rather than binary accuracy I use the **Receiver Operating Characteristic** (**ROC**) curve, that is a graph of the true positive rate ($TPR = TP/(TP + FN)$) against the false positive rate ($FPR = FP/(FP + TN)$), calculated at different `gammaness` thresholds.

The area under the ROC curve (AUC) takes values between 0 and 1, and quantifies the diagnostic ability of the models: the closer to 1 it is, the better is the algorithm at distinguishing between the two

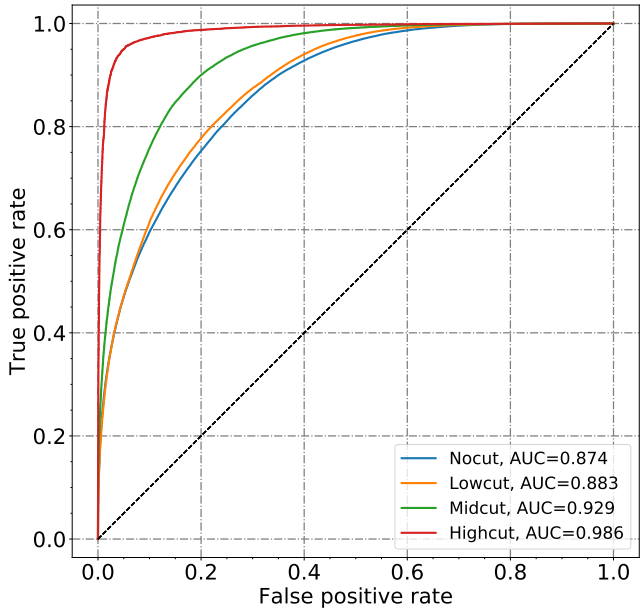


Figure 4.2: ROC curves for the different cut-levels. The higher the cut, the higher the discrimination power.

classes[2]. If AUC = 0.5 instead, the model is unable to make any prediction, behaving like a random decision maker. In Fig. 4.2 are plotted the ROC curves for all the cut-levels. The worst classifier is the `no-cut` (AUC = 0.868), whereas the best is the `high-cut` model, with AUC = 0.986. This hierarchy was expected, since higher cut levels increase the energy threshold (see Tab. 3.1) and select higher quality images. The drawback is that many events are also lost, losing the possibility to exploit their information.

Moreover, the discrimination power of a single network is energy-dependent. Dividing the test events into four energy bins, each model gives the following AUCs:

	5 - 50 GeV	50 - 500 GeV	500 GeV - 5 TeV	5 - 50 TeV
`no-cuts`	0.750	0.930	0.980	0.994
`low-cuts`	0.758	0.933	0.991	0.996
`mid-cuts`	0.811	0.940	0.991	0.994
`high-cuts`	-	0.980	0.994	0.999

The models are almost perfect at discriminating high-energy events ($5 < E < 50$ TeV) reaching AUC scores above 0.99, while they lose accuracy in the lowest energy range below 50 GeV (the `high-cut` model isn't evaluated in this bin because it is under its threshold). In Fig. 4.3(a) the ROC curves of the `mid-cut` VGG calculated in the four different energy domains are shown.

An important caveat: I use the estimated energy to evaluate the ROCs because when we apply this analysis to the data, this is what we will have. To calculate it, I apply the models for energy regression obtained in the next section 4.2 both to protons and to γ rays, even though they were

[2] AUC = 0 represents the inverted situation, in which the model is good at separating the two classes, but it is swapping them.

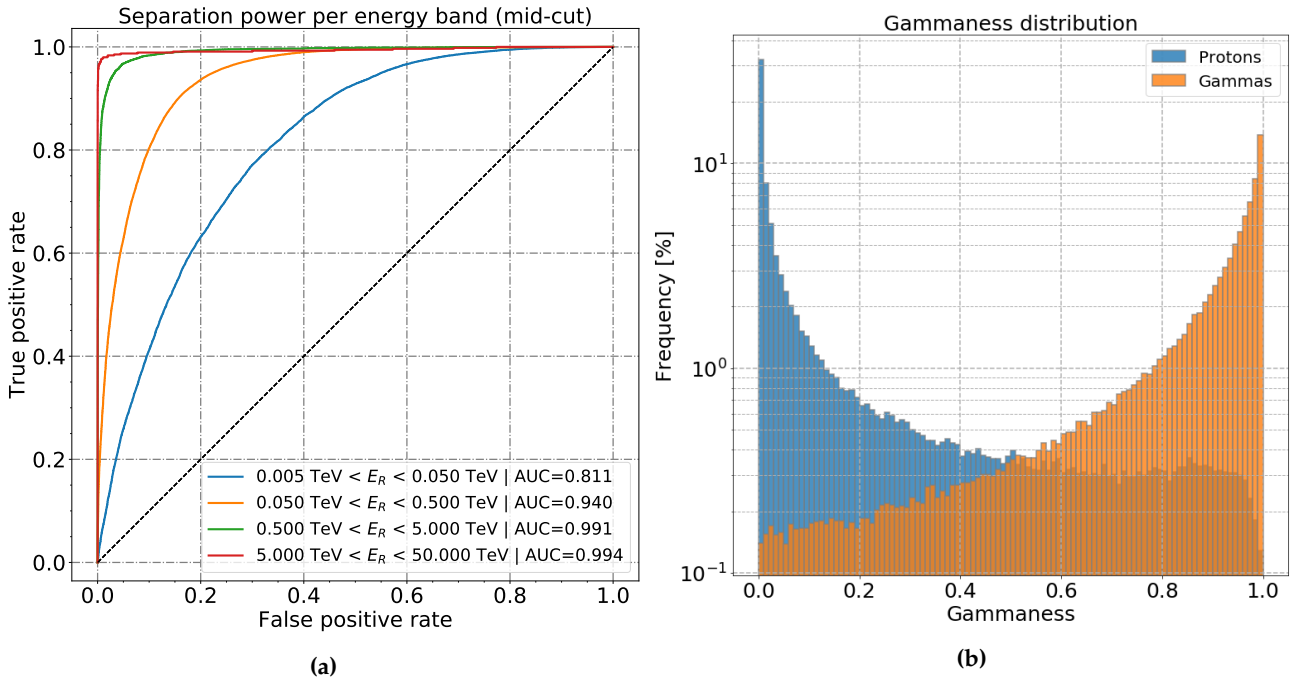


Figure 4.3: Performances of *mid-cut* model: (a) ROC curves in low, mid and high energy bands and (b) `gammaness` distribution of the pointlike γs and protons test sets.

trained only on γ-like events. Indeed, some protons lose a large part of their energy in EM showers and are likely to be classified and reconstructed as γ rays, thus affecting the curves.

Comparing the AUCs reached by my networks with the Random Forest, we find that the `low-cut`, `mid-cut` and `high-cut` models overtake the respective RFs by 15.6%, 9.5% and 3.4% respectively. The performance comparison is shown in Fig. 4.10 at the end of this chapter, where we also see that even the `no-cut` VGG13 outperforms the `mid-cut` RF.

4.2 Energy reconstruction

For this task, the whole set of $\sim 2 \cdot 10^6$ `diffuse` γ rays has been exploited, with a training/validation split of 80/20, together with the whole set of $\sim 1.9 \cdot 10^6$ `pointlike` γ rays for testing the models. Each training lasted 50 epochs, with a batch size 128, adam optimizer and CLR settings as in the previous section, except the `mid-` and `high-cut` models whose max LR was lowered down to 0.001. The networks have been regularized with a dropout rate of 0.3 and no L2 regularization.

The output of the models is the logarithm of the event's energy $\epsilon = \log_{10} E$, and the chosen loss function is the **mean absolute error (MAE)**:

$$L_\epsilon = \frac{1}{N} \sum_{i=1}^{N} |\epsilon_i - \hat{\epsilon}_i| \tag{4.3}$$

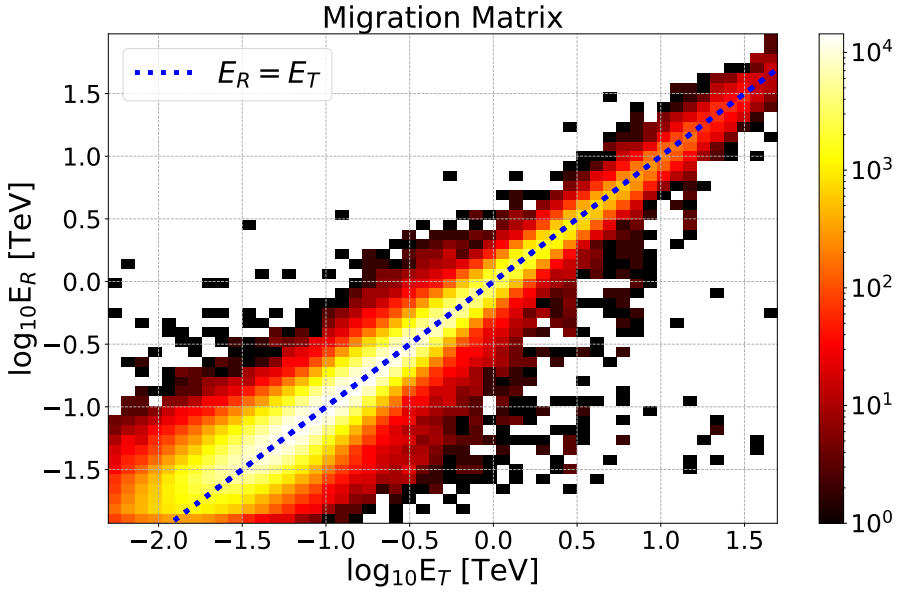


Figure 4.4: Energy migration matrix of `mid-cut` VGG. The blue dotted line represents the ideal reconstruction.

	`no-cut`	`low-cut`	`mid-cut`	`high-cut`
train	0.142	0.118	0.095	0.049
valid	0.150	0.125	0.104	0.067
test	0.142	0.129	0.107	0.070

Table 4.2: Mean Absolute Error reached by the models on training, validation and test sets.

In this way the algorithm tries to minimize L_ϵ keeping the same percentage error over all the energy range; indeed, propagating the model's prediction uncertainty $\Delta\epsilon$ - provided that it is sufficiently small - one obtains:

$$\frac{\Delta E}{E} = \frac{1}{E}\frac{dE}{d\epsilon}\Delta\epsilon \simeq \ln 10 \Delta\epsilon \tag{4.4}$$

This would be true if this error was uniformly distributed, but, as expected, it comes out to be energy dependent. In Tab. 4.2 are listed the training and validation losses reached by the models on the minimum validation epoch, and in Fig. 4.4 is shown the *migration matrix* of the `mid-cut` model (i.e. the 2D-histogram of the reconstructed energy against the MC energy already introduced in 2.4.3).

To evaluate the performances, I split the energy spectrum in 5 bins per decade, and for each bin the *relative energy error* is histogrammed, namely:

$$\frac{\Delta E}{E} = \frac{E_R - E_T}{E_T} \tag{4.5}$$

where T stands for "true" and R for "reconstructed". The histograms for the `mid-cut` model are shown in 4.5. The distribution is then bias-corrected, i.e. the median is subtracted, and we define the **energy resolution** as the 68^{th} percentile of the histogram $|E_R - E_T|_{\text{corr}}/E_T$.

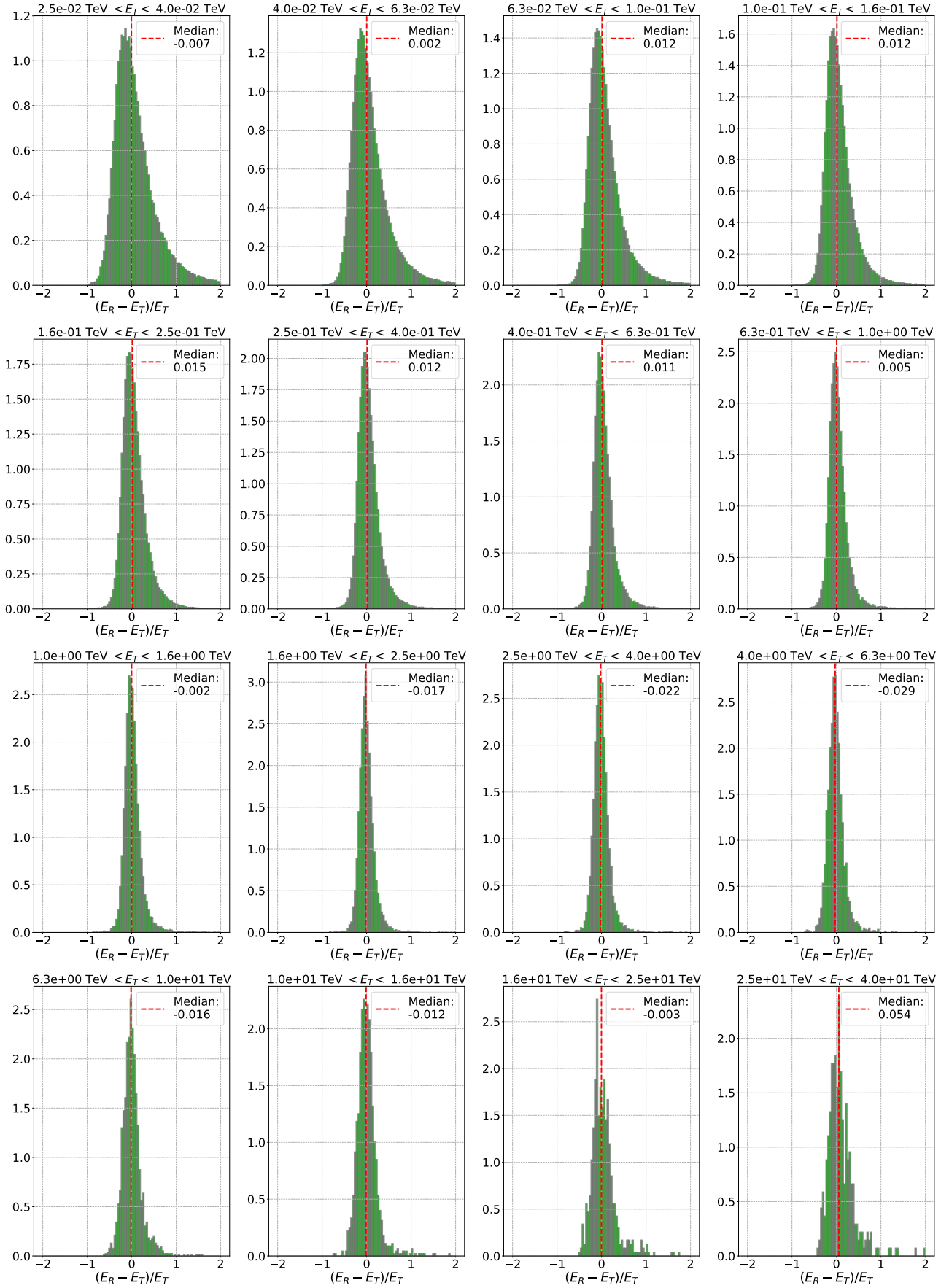


Figure 4.5: Histograms of the relative energy error for each energy bin, obtained with the `mid-cut` model. The median of each distribution represents the relative energy bias.

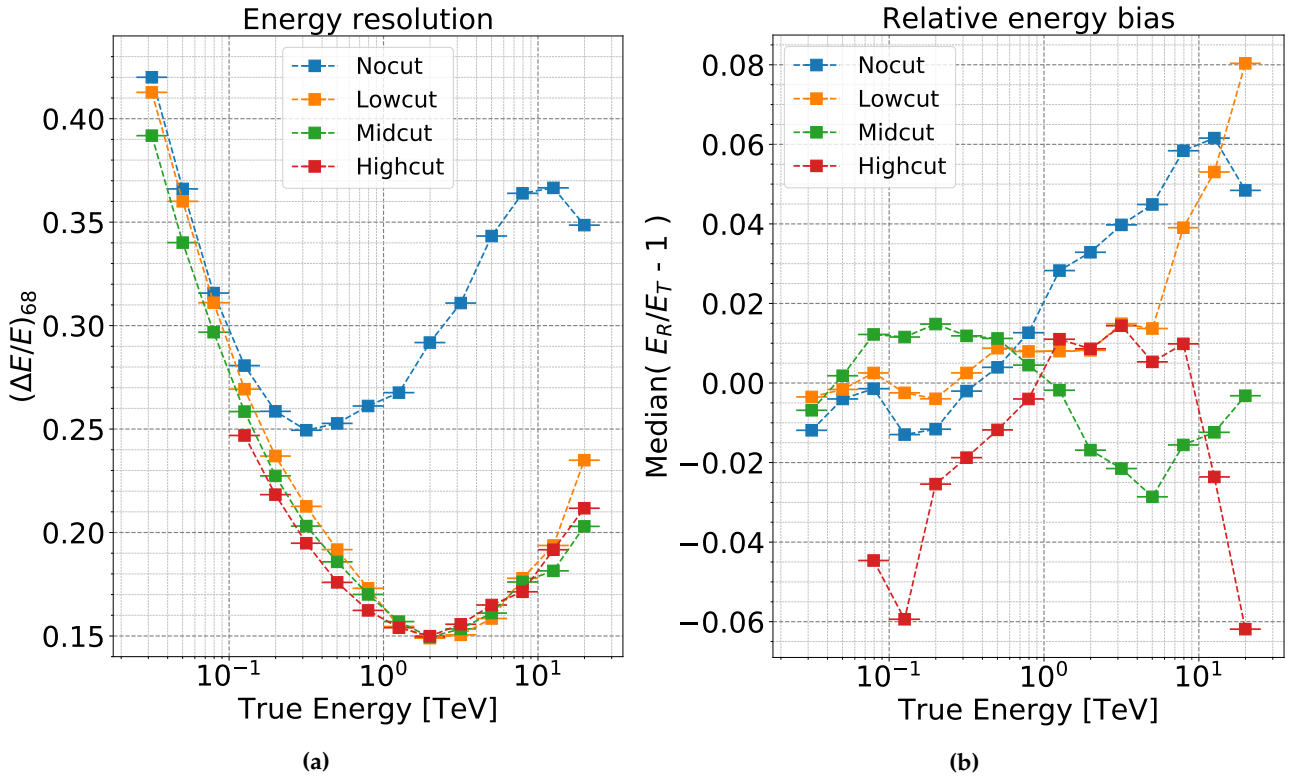
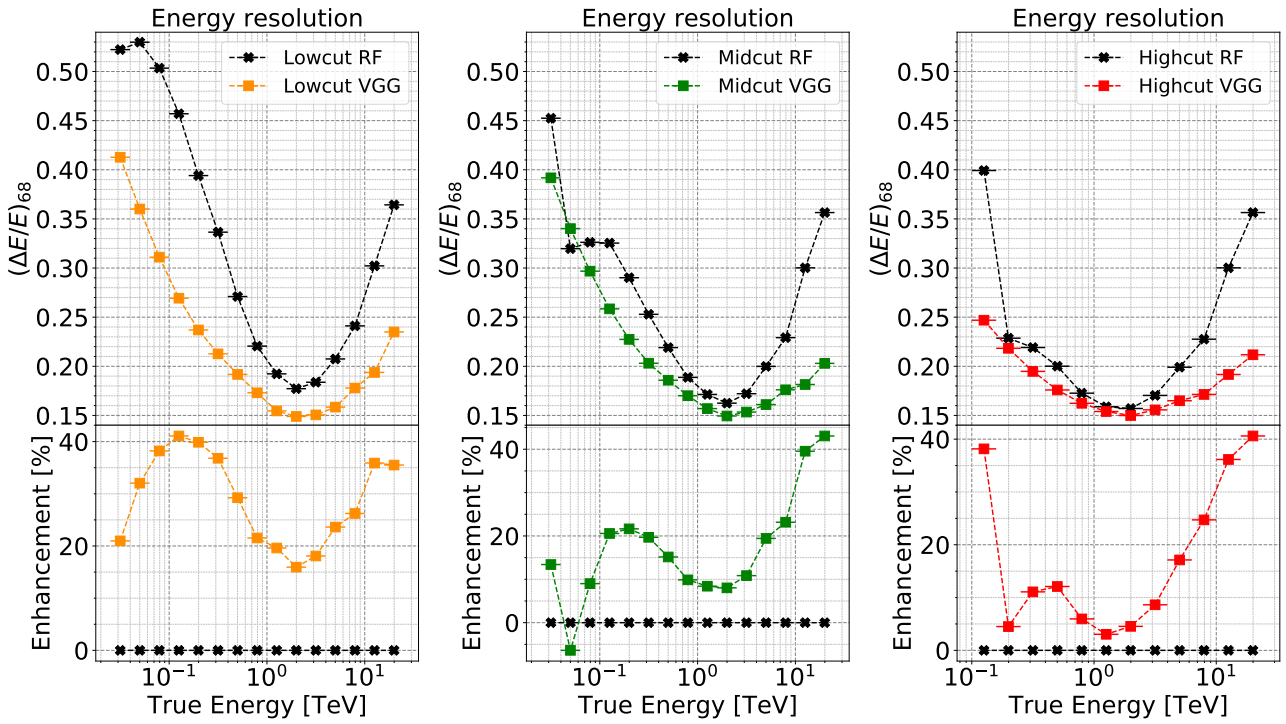


Figure 4.6: Energy resolutions (left) and relative energy biases (right) of the VGG13 for the different cut levels. All the models show the same trend and similar performances (except the `no-cut` network). The relative biases stay under 8%.

The **relative energy bias**, defined as the median of Eq. 4.5, vary less than 8% as depicted in Fig. 4.6(b). In Fig. 4.6(a) the energy resolutions for the different cut levels are shown; the first three points of the `high-cut` model have been excluded due to poor statistics (induced by the higher energy threshold - see Fig. 3.5) Also, I neglected the energies above 20 TeV, at which the simulated gamma events are scarce and that are not in the energy range that is the target of the LST. The `low`, `mid` and `high-cut` show the same trend and similar performances across all the energy ranges, especially in the 500 GeV-10 TeV where a resolution of around $15 - 18\%$ is reached; the `no-cut` reconstruction ability on the other hand starts degrading quite soon. This is probably due to the high number of leaked shower images: being a large part of them outside the camera, the number of actually contained photoelectrons is not representative of the events' energies. In general, at lower energies (< 500 GeV) resolution is poorer since low-energy γ rays produce less luminous cascades, resulting in less light collected in the camera.

But why does energy resolution worsen also at the *highest* energies (> 10 TeV)? The reason is that, since showers are brighter, there are more events with a large impact parameter able to trigger the telescope, producing shower images with few photoelectrons, thus lowering the quality of the dataset in this range.



(a) (b) (c)

Figure 4.7: Comparison between RF and VGG's energy resolutions (lower is better). The two methods show similar trends, although the CNN exhibit an overall better performance, especially at the higher energies.

4.2.1 Comparison with RF performance

To get a measure of how good my results are, I compared the energy resolutions reached by the VGG-networks with the ones obtained with the Random Forests produced via the `cta-lstchain` software package. Opposite to CNNs, that exploit pixel-wise information, RFs cannot be trained on the whole `no-cut` dataset: image parametrization fails for too low-quality images, hence they are discarded and it is not possible to make a fair comparison in that case.

The comparison is shown in Fig.4.7: the VGG models perform better in all three cases. They all largely outperform the RFs above 10 TeV, while they give similar performances in the few TeV regime. The overall largest improvement - defined as $(\epsilon_{RF} - \epsilon_{VGG})/\epsilon_{RF}$, being ϵ the resolution - is reached by the `low-cut` level, spanning from 15% up to 40% around 100 GeV; the enhancement is still appreciable in the other two cases, although it is less dramatic especially the `high-cut`. All my networks exhibit roughly the same trend as the RFs, as a prove that the reconstruction was done properly.

	no-cut	low-cut	mid-cut	high-cut
train	0.246	0.218	0.159	0.066
valid	0.255	0.223	0.180	0.100
test	0.210	0.196	0.142	0.056

Table 4.3: Values of Mean Absolute Error for direction reconstruction reached by the minimum-validation models on training, validation and test sets.

4.3 Direction reconstruction

To accomplish this task, the network needs to learn two coordinates describing the arrival direction of the primary γ ray generating the cascade. The true coordinates provided along the MC data are *altitude* $\hat{\alpha}$ and *azimuth* $\hat{\rho}$: these need to be converted into the LST coordinate-frame, taking the alt-azimuthal difference between the source position and the pointing direction of the telescope. Each event will thus be labeled with a tuple $z = (\Delta\hat{\alpha}, \Delta\hat{\rho})$, and consequently the last layer of the networks will have a double output. The dataset used for training and test is the same as the energy regression, as well as the training strategies (number of epochs, CLR, adam...).

The choice for the loss function is the **mean absolute error**. Given a batch of N events, it reads:

$$L_z = \frac{1}{N}\sum_{i=1}^{N} \|z_i - \hat{z}_i\| \tag{4.6}$$

where $\hat{z}$ represents the true coordinates and z the reconstructed counterpart. The overall values of L_z of the models on the training, validation and test sets are reported on table 4.3.

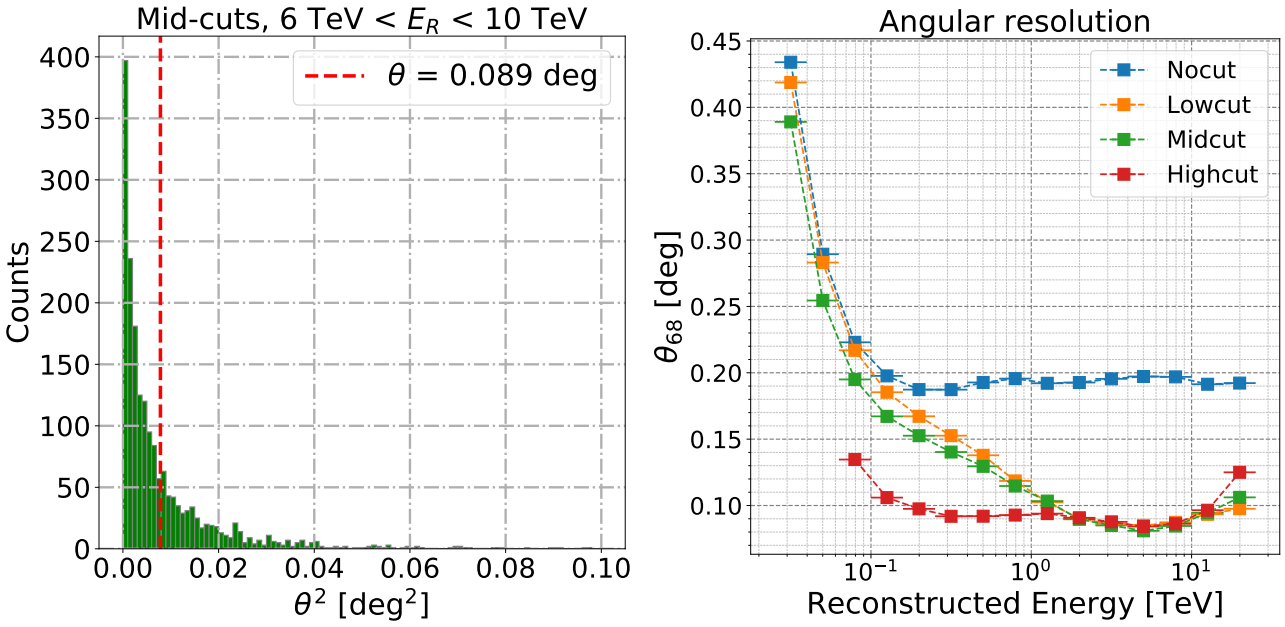
In order to evaluate the performances on the test sets, I produce a θ^2 plot (already defined in section 2.4.3) for each energy bin, namely the histogram of the quantity:

$$\theta_i^2 \equiv \|z_i - \hat{z}_i\|^2 \tag{4.7}$$

.

As an example, in Fig. 4.8(a) is shown the θ^2 histogram in the 6-10 TeV range obtained with the `mid-cut` model. In order to reproduce results as much as possible close to real life, note that the events are binned accordingly to the respective *reconstructed* energy, and *not* to the MC energy - in the same way I did for separation in section 4.1. Real data analysis is indeed affected by distorted energy reconstruction, like illustrated in the migration matrix (Fig. 4.4).

From each θ^2 histogram is then extracted the **angular resolution** for that energy bin, defined as the square root of the 68% containment radius shown in Fig. 4.8(b). The performances are sensitive to the different quality cuts up to $E_R \sim 1$ TeV. With increasing energies, all the models except the `no-cuts` gain pretty much the same resolution, reaching a value below a tenth degree in the range between 2 - 10 TeV. The `high-cuts` network's resolution goes down to this value already at 200 - 300 GeV, because cutting at 1000 phe only the very bright images are selected.



(a) (b)

Figure 4.8: (a) θ^2 histogram of test events in the 6-10 TeV range, obtained with the `mid-cut` model. The red-dotted line marks the 68^{th} percentile, used to calculate angular resolution. **(b)** Angular resolutions for *all* the cut-levels (lower is better). Models gain precision as energy increases, until $E \sim 10$ TeV.

4.3.1 Comparison with RF perfomance

As I did for the energy reconstruction, in Fig. 4.9 I compare my performances with the Random Forest's. The improvement brought by the VGG in this case is definitely evident, boosting the perfomances up to more than 60% with respect to the RFs. Again, the major improvement is given by the `low-cut` model, with a $35-55\%$ enhancement in the 1-20 TeV range (where the best resolution is achieved), but also the `mid-cut` and `high-cut` show a significant $20-50\%$ improvement in the same range.

4.4 Cross-check with LAPP and UCM groups

During the period of this work, similar analysis has been independently developed for LST by two other research groups: the γ-ray group at *Laboratoire d'Annecy De Physique Des Particules (LAPP)* that develops and uses the *GammaLearn*, the DL group at *Universidad Complutense de Madrid (UCM)* that develops and uses *CTLearn*. The *GammaLearn* team designed a multi-purpose single network based on the ResNet architecture that addresses all the three tasks at once, while *CTLearn* built three single-task ResNet-style networks. They implemented advanced techniques such as mechanisms of self-attention (squeeze and excitation networks [Hu et al., 2019]). Both the groups do not use the cubic resampling described in section 3.4.1 like I did: *CTLearn* processed the images via linear interpolation, *GammaLearn* used another method called *IndexedConv* (similar to hexagonal convolution used

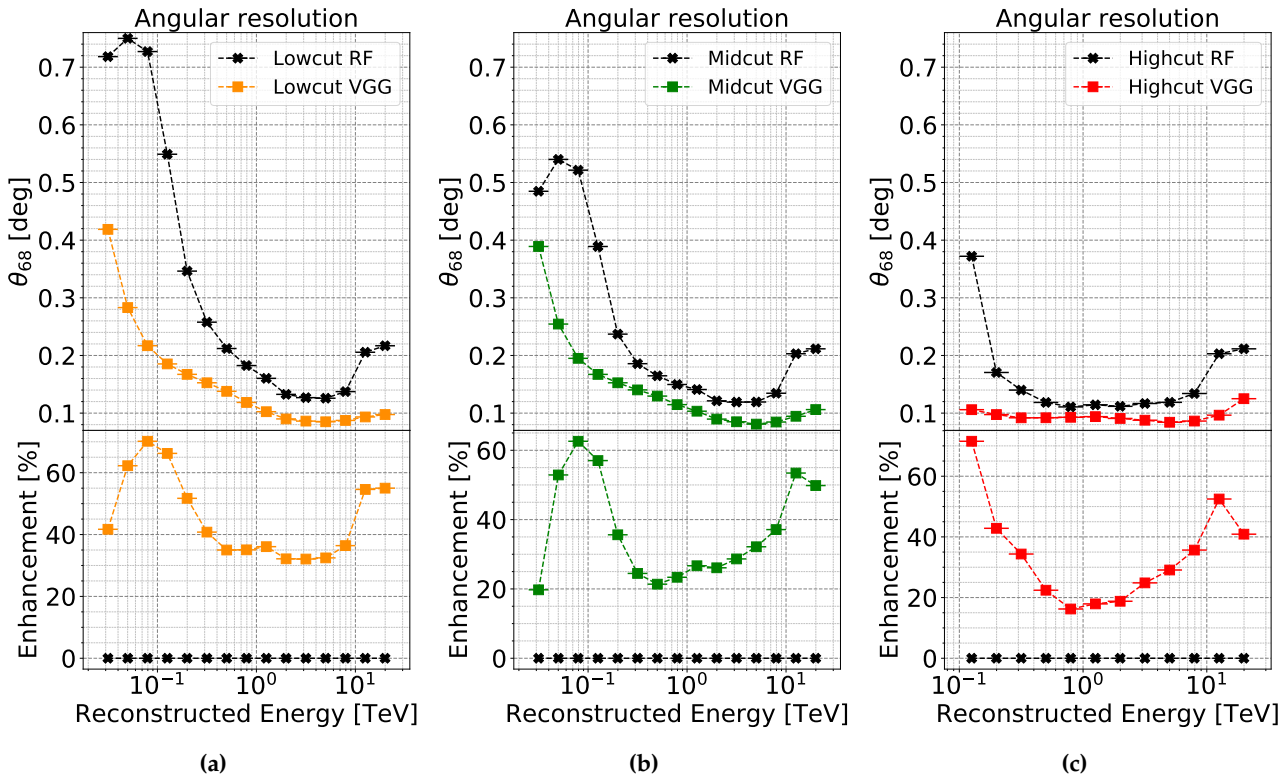


Figure 4.9: Comparison between RF and VGG's angular resolutions (lower is better). The CNN evidently outperforms the RF algorithm in this task, improving the resolution up to more than 60% in some ranges.

in Shilon et al., 2019): knowing the nearest neighbor of each pixel, it allows to perform convolution and pooling operations on non-Euclidean grid of data [Jacquemont. et al., 2019].

We compared the outcomes of our efforts as a cross-check of the robustness of our work. In Fig. 4.10 I show the results of signal/background separation for all the cut levels. Despite some small difference, the results are very similar, and this is even more evident when comparing to the AUCs obtained with the Random Forests: all of our `no-cut` models perform even better than the `mid-cut` RF. In Fig. 4.11 and Fig. 4.12 I show the energy and angular resolutions reached by each of the three groups for all the cut levels. Again, the results are consistent with each other, showing almost identical behaviours. The error bands displayed in the graphs are the 16^{th} and 84^{th} percentiles around the median of their results, as they repeated each experiment 10 times with 10 different seeds for the weight initialization of the networks.

The performance derived by the CNN analysis with respect to that of the RF, although better in all the tested tasks, it is less robust against variations in real data due to a small effects present in the data that are not properly simulated, as the noise present in each Field of View the telescope is looking at. I have performed all these tests on simulated data available at the time of writing this thesis, but the LST is currently under the commissioning phase, taking data and fine-tuning its performance. Whenever this period is finished, the CNN analysis will be tested using real data.

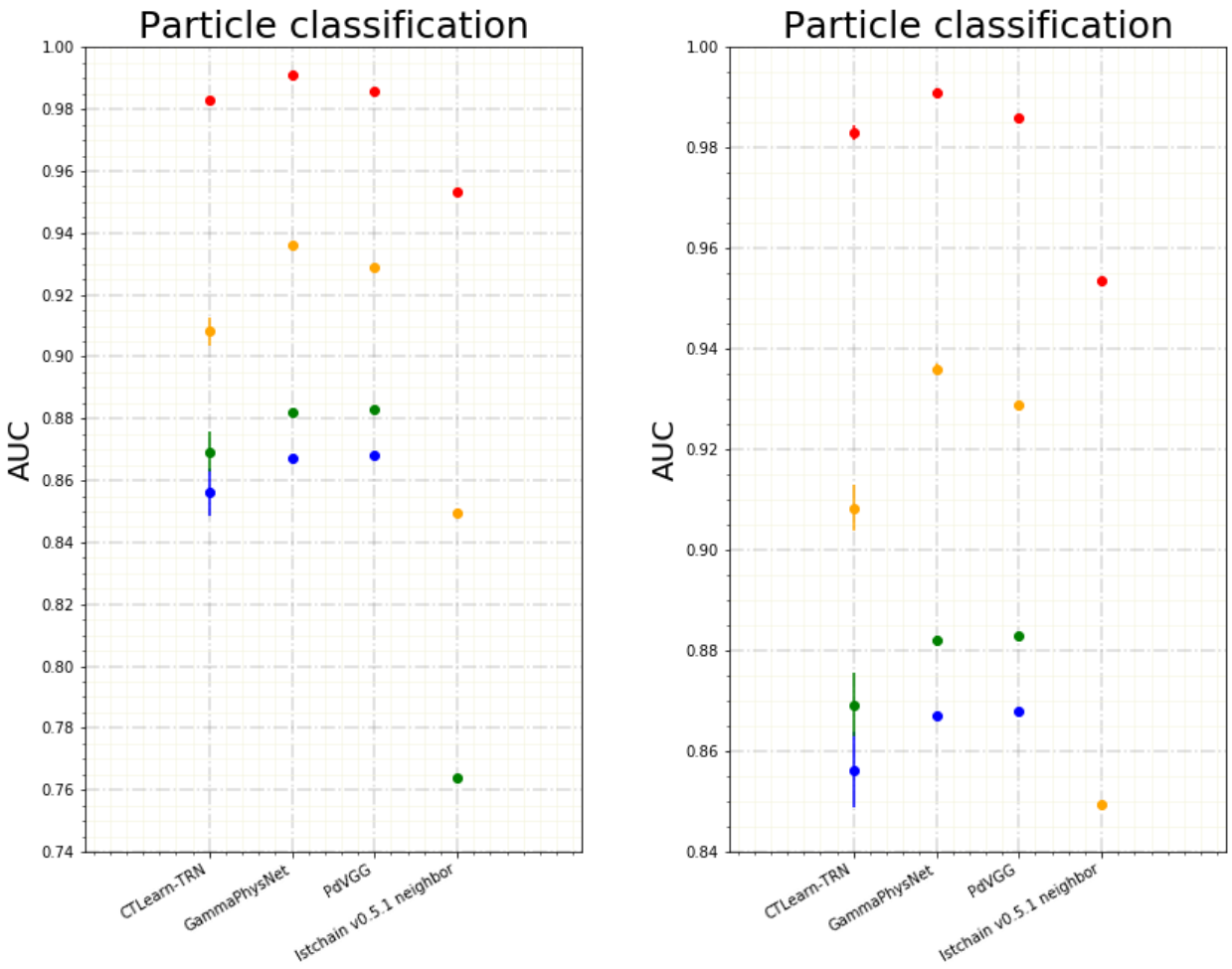


Figure 4.10: Comparison of the signal/background separation results obtained with the different approaches of INFN-UniPD (us), CTLearn and GammaLearn, and further compared to the Random Forests produced via the `cta-lstchain` software package. On the right I show the same picture as the one on the left, but zoomed. The AUC scores obtained by the CNNs are very similar, especially when compared to the results of RFs - even the `no-cut` CNN models outperform the `mid-cut` RF. Blue points: `no-cut` models; green: `low-cut`; yellow: `mid-cut`; red: `high-cut`.

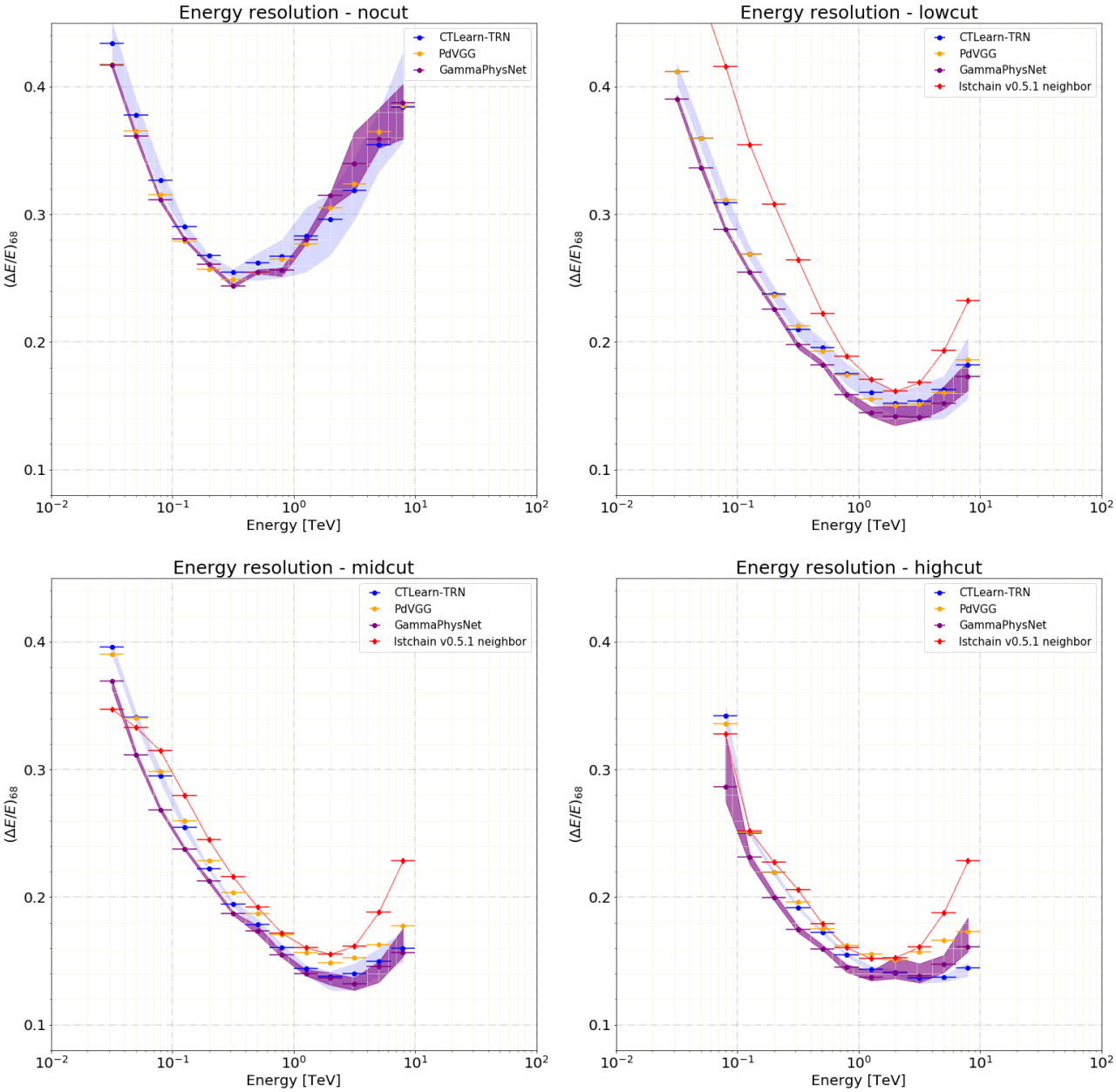


Figure 4.11: Comparison of the energy resolutions obtained with the different approaches of INFN-UniPD (us), CTLearn and GammaLearn, and further compared to the Random Forests produced with the `cta-lstchain` software package.

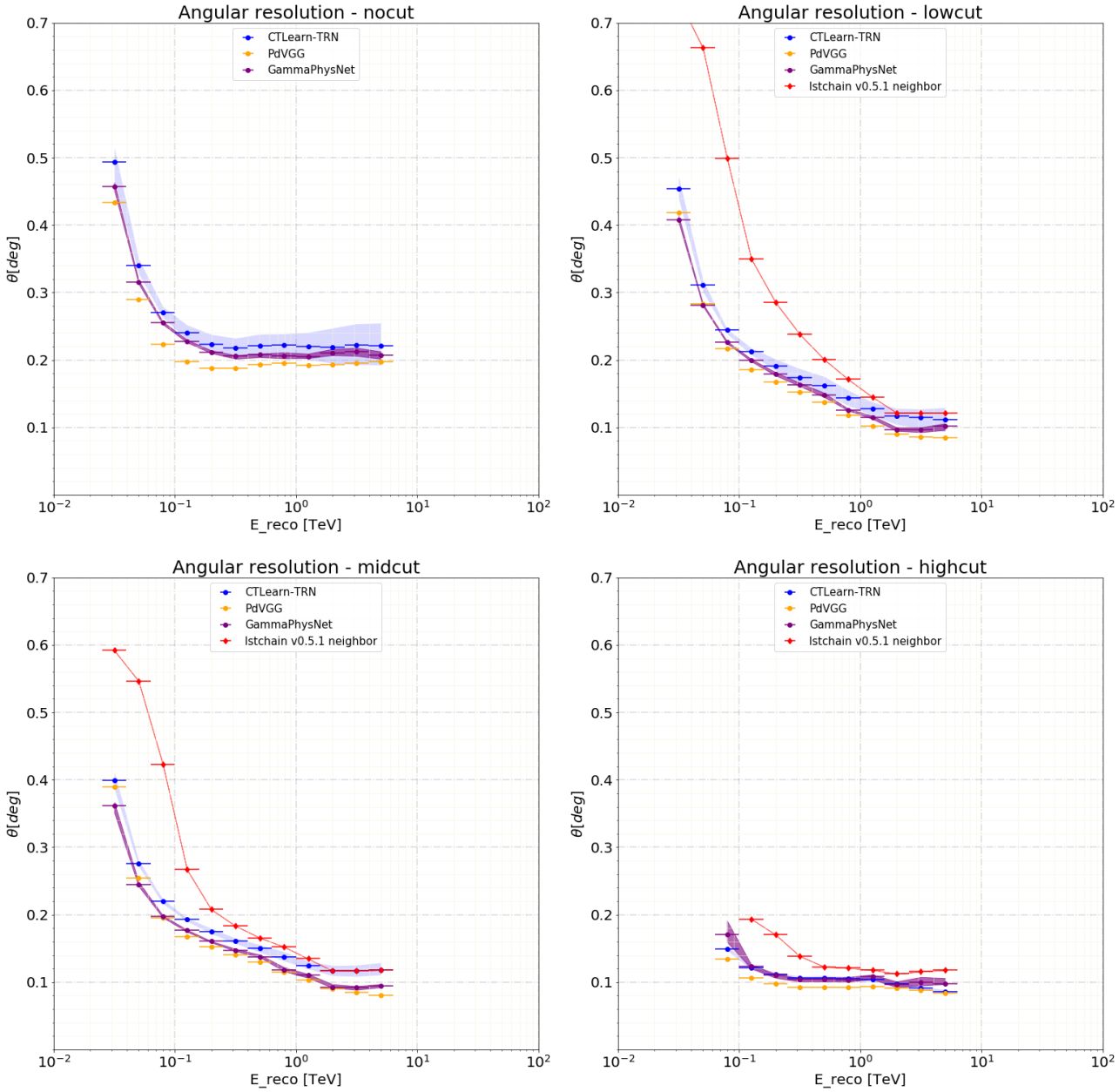


Figure 4.12: Comparison of the angular resolutions obtained with the different approaches of INFN-UniPD (us), CTLearn and GammaLearn, and further compared to the Random Forests produced with the `cta-lstchain` software package.

Chapter 5

Blazars

Active Galactic Nuclei (AGNs) are amongst the most intriguing objects in the Universe: they are cosmic accelerators producing extremely energetic particles, arousing enormous interest in astrophysicists. The most general definition of AGN refers to all the galaxies exhibiting a core luminosity that exceeds the luminosity due to stellar emission, comprehending a large family of objects; this is due to the presence of a central super massive black hole accreting matter around itself. One main subdivision is made on the basis of the presence of jetted emission [Padovani, 2016]: **non-jetted AGNs** are the vast majority currently observed, representing roughly the 90% of the population, and were classified mostly as radio-quiet AGNs; **jetted AGNs** represent the remaining 10%, and are usually radio-loud[1] [Kellermann et al., 1989]. From now on in this thesis we will call AGN *only the jetted active nuclei*.

5.1 The unified scheme of AGNs

The model describing the physical structure of an AGN is depicted in Fig. 5.1. It exhibits axial symmetry, due to angular momentum. Depending on the angle at which it is observed, the observational properties of an AGN change dramatically: in the past AGN viewed ad different angles were classified as different objects, until the following *unified scheme* was proposed [Urry and Padovani, 1995]. The main components are:

- a **super massive black hole** at the center, accreting the sourrouding material onto itself. The mass of super massive black holes span from about 10^6 up to 10^9 solar masses, and together with the angular momentum determines the geometry of the spacetime around it. In 2019 the Event Horizon Telescope performed the first imaging of such an object, the black hole at the center of the radiogalaxy M87 [Akiyama et al., 2019].
- a geometrically-thin **accretion disk** of external matter falling into the black hole. Friction caused by turbulent motion heats up the material as it approaches the innermost stable orbit. Being the disk optically thick, the result is a superposition of black-body emissions, that extend from optical light up to soft x-rays in the proximity of the black hole. The accretion efficiency of an AGN is linked to the emission power of the jets.

[1] i.e. exhibiting a radio flux at 5 GHz more than 10 times higher than the optical flux in the B band.

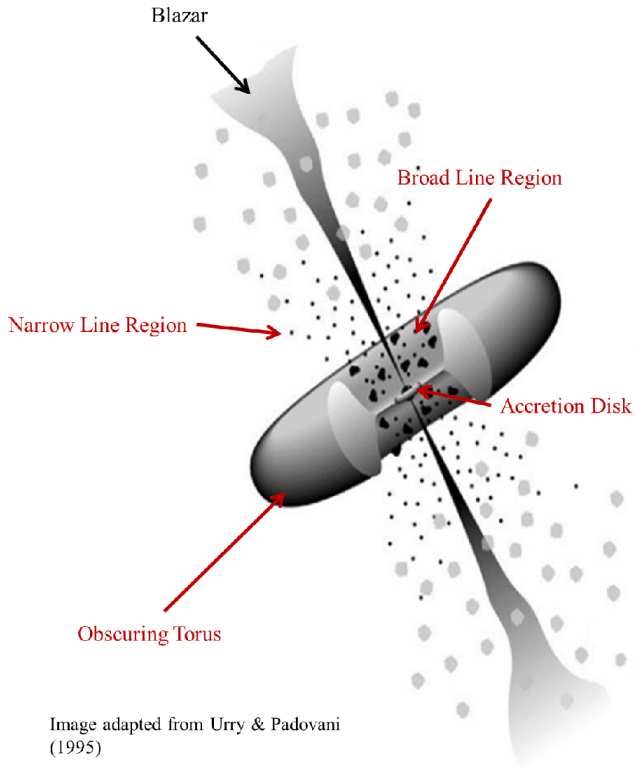


Figure 5.1: Scheme of a jetted AGN. Image adapted from Carroll, 2015 and Urry and Padovani, 1995.

- two **jets of ultra-relativistic plasma**, streaming outwards from the vicinitiy of the black hole in the direction perpendicular to the accretion disk. They are characterized by a *non-thermal emission* that covers all the electromagnetic spectrum, from radio up to VHE γ rays, commonly attributed to synchrotron radiaton and other mechanisms such as Inverse Compton (IC) (see 1.2.1). Jets are highly collimated by strong magnetic fields, travelling up to distances of the order of Mpc where they form two extended lobes emitting at radio frequencies.
- a **torus** made of dust and gases that covers and obscures the central region; a **corona** around the disk that reprocesses its emission; a **Broad Line Region (BLR)** where hot gas emits Doppler-widened optical spectral lines and a **Narrow Line Region (NLR)**, characterized by thin emission lines. All of these four features are almost absent in the case of BL Lac objects, a particular class of AGN of our interest that will be discussed in the next section.

Jetted AGNs are usually classified on the basis on the viewing angle with respect to the jets: when it is large ($> 20°$) they are called **radiogalaxies**, and we can observe the radio lobes originating from the jets; when the angle is relatively small ($< 20°$, see Fig. 5.1), they are called **blazars** [Prandini, 2017].

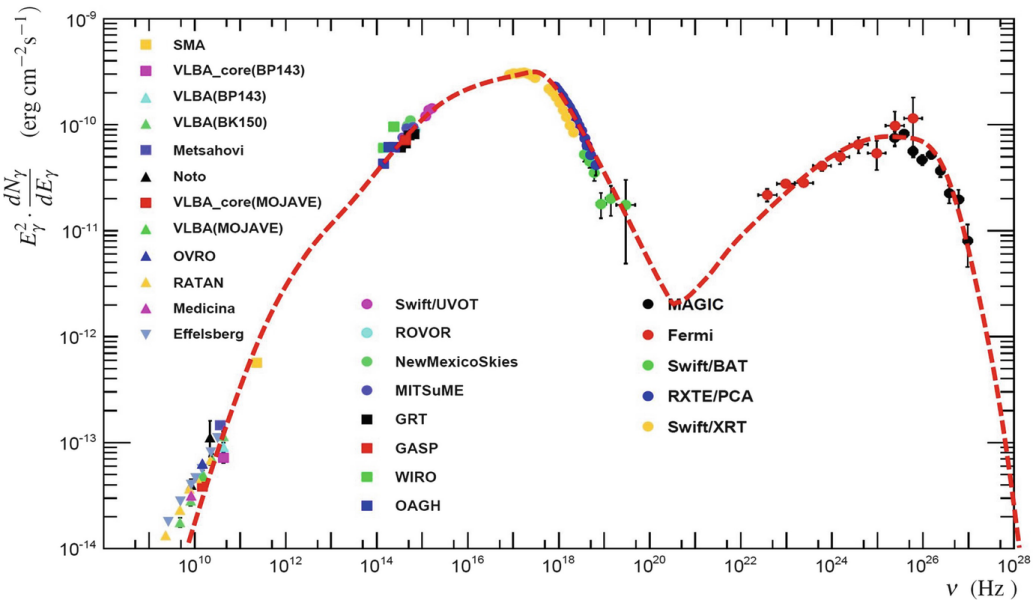


Figure 5.2: Spectral energy distribution of the blazar Markarian 421 - a BL-Lac object located in constellation Ursa Major - measured by various experiments. It shows the typical double-peaked non-thermal continuum.

5.2 The blazar family

Blazars are AGNs with the jets pointing towards the direction of the observer's line of sight. They are objects of utmost interest for astrophysicists: the jets are practically not obscured by the surrounding material, if any, so we can probe the properties of the emitting region.

The Spectral Energy Distribution (SED, see Appendix C.1) of blazars is dominated by the non-thermal continuum emitted by the jets, covering almost all the electromagnetic spectrum. It features **two broad peaks**: the first extends from the radio to the hard X-rays and exhibits a strong polarization, and it is commonly ascribed to synchrotron radiation by accelerated electrons; the second instead reaches the γ-ray domain and is still subject of debate. However, it is often explained assuming Inverse Compton (IC) up-scattering of photons by high-energy electrons (see 5.2.2). As an example, in Fig. 5.2 is shown the SED of the blazar Markarian 421 measured by different experiments. Another key characteristic of blazars emission is their **rapid variability** at all wavelengths, with typical timescales ranging from years in the radio down to minutes in the VHE and X-ray domains: they are known for exhibiting **flaring states**, that are sudden and unpredictable rises of the flux at some wavelength, that can exceed up to tens of times its average flux.

Blazars are divided in two classes:

- **Flat Spectrum Radio Quasars (FSRQs)**, showing strong optical emission lines coming from the Broad Line Region and Narrow Line Region, and the thermal spectrum of the accretion disk.
- **BL Lacertae** (or **BL-Lac**) **objects**, showing *almost no optical emission lines*. They exhibit the

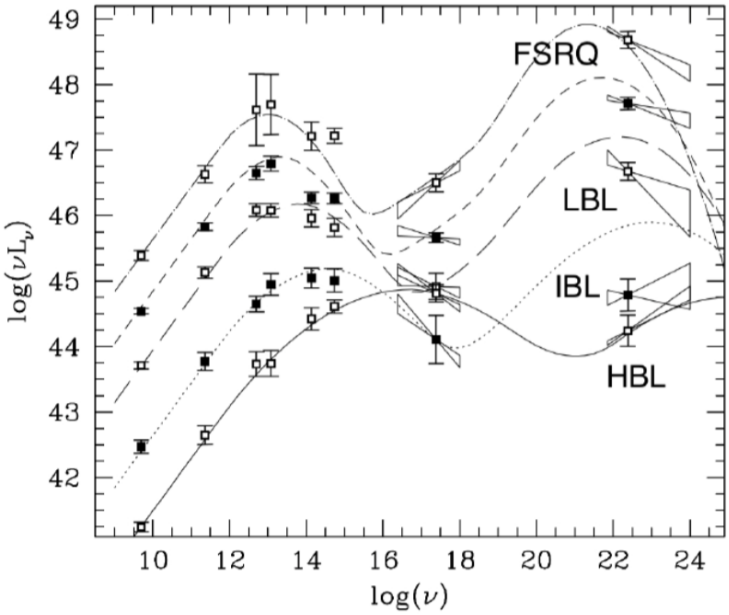


Figure 5.3: The Blazar sequence. The fainter is the source, the higher is the energy of the peaks (and viceversa). Adapted from Fossati et al., 1998.

strongest and most rapid variability among blazars; for this reason, they are among the favorite subjects of observation of multi-wavelength (MWL) campaigns. BL Lacs are subdivided in **LBL**, **IBL** and **HBL** (**L**ow-, **I**ntermediate- and **H**igh-peaked **BL**-Lac) on the basis of where the low-energy peak is located: $< 10^{14}$ eV, $10^{14} - 10^{15}$ eV, $> 10^{15}$ eV respectively [Padovani and Giommi, 1995].

Due to strong relativistic beaming effect, the observed power is strongly enhanced with respect to the emitted power, and also the energy of the photons and the time variability are amplified in the observer's frame of reference. See Appendix B for a further explanation of this phenomenon.

5.2.1 The blazar sequence

The overall luminosity of blazars seems to be anti-correlated with the position of the peaks of the SED: FSRQs are brighter, and their peaks lean towards lower energies with respect to BL-Lacs, that are conversely much less powerful and peaking at higher energies. As illustrated in Fig. 5.3, it seems that also the BL-Lac subcategories (LBL, IBL, and HBL) occupy regions of the plot accordingly to this trend. Such a phenomenon is called **blazar sequence** [Fossati et al., 1998; Ghisellini et al., 2017]. Moreover, there is another trend that can be spotted in 5.3: the brighter is the spectrum (e.g. FSQRs), the higher is the luminosity of the high-energy bump with respect to the low-energy.

These fenomena may be correlated to accretion efficiency: FSRQs are more efficient and, according to the AGN unification scheme depicted in 5.1, the enviroment around the central region is crowded by the BLR and by photons emitted by the accretion disk. When the jet crosses this region, these external photons can serve as seeds for IC scattering by the electrons of the jet, thus accounting

for the increased high-energy luminosity; in this way the electrons suffer a strong radiative cooling and cannot reach very-high energies, thus the SED peaks shift towards lower frequencies. This does not happen in a BL-Lac, where the ambient is much cleaner due to less accretion efficiency, and so the jet can travel unaltered: this explains both the lower luminosity and the higher energies reached by the peaks [Ghisellini, 2016].

From these observations, and in particular the second peak position, it appears clear that HBLs are the most detected blazars in the VHE γ rays.

5.2.2 Jet emission in BL Lac objects

The current paradigm of jet emission involves the presence of shock waves in the vicinity of the SMBH, accelerating blobs of plasma to ultra-relativistic velocities in regions filled with strong magnetic fields. The electron population of the ionized gas interacts with such fields and emit synchrotron radiation, which is the origin of the first peak of the SED of blazars.

The origin of the second peak is instead still largely discussed, and a plethora of emission models have been proposed, subdivided in *leptonic models* and *hadronic models*. Hadronic models include different mechanisms such as synchrotron emission by extremely-energetic protons [Aharonian, 2000] and γ-ray emission from decay of neutral pions originated from proton/proton or proton/γ interaction [Mannheim, 1993]; these models also foresee neutrino emissions, that have become of enormous interest since the detection of a neutrino event in association with the blazar TXS 0506+056 [Aartsen et al., 2018]. On the other hand, according to leptonic models, the VHE radiation is produced via IC scattering (see 1.2.1). In the case of BL-Lacs, in which there are no diffuse photons in the surroundings, the most accredited leptonic model is the Synchrotron Self-Compton (SSC), where the electrons present within the jets up-scatter the same synchrotron photons they emit [Ghisellini, 2013; Konigl, 1981; Tavecchio et al., 1998]. This bond foresees that the fluxes associated with the two bumps are correlated.

Synchrotron Self Compton

The simplest version of this model is the *one-zone SSC*: it consists on a population of ultra-relativistic electrons confined in a spherical emitting region of radius R, immersed in a uniform magnetic field B. The energy spectrum of the electrons is often assumed to be a *broken power law*:

$$N(\gamma) = \begin{cases} N_0 \left(\frac{\gamma}{\gamma_b}\right)^{-p_1} & \text{if} \quad \gamma \leq \gamma_b \\ N_0 \left(\frac{\gamma}{\gamma_b}\right)^{-p_2} & \text{if} \quad \gamma > \gamma_b \end{cases} \tag{5.1}$$

where γ is the Lorentz factor of the electrons and N_0 is the electron density parameter. p_1 and p_2 are the spectral indexes respectively before and after the spectral break, that occurs at $\gamma = \gamma_b$. Alternatively, it has been shown that the spectrum of particles accelerated in environments where shock waves, magnetic fields and cooling processes take place, is well approximated by a *log-parabola with*

a power law branch [Massaro, E. et al., 2006 and references therein], namely:

$$N(\gamma) = \begin{cases} N_0 \left(\frac{\gamma}{\gamma_0}\right)^{-s} & \text{if} \quad \gamma \leq \gamma_b \\ N_0 \left(\frac{\gamma}{\gamma_0}\right)^{-(s+r\log\gamma/\gamma_b)} & \text{if} \quad \gamma > \gamma_b \end{cases} \tag{5.2}$$

where γ_0 is called *turn-over energy*, s is the spectral index and r the spectral curvature.

Relativistic beaming effects (see Appendix C) are encoded in the Doppler factor:

$$\delta = \frac{1}{\Gamma_{\text{bulk}}(1 - \beta\cos\theta)} \tag{5.3}$$

where Γ_{bulk} is the bulk Lorentz factor of the electrons, and θ the angle between the jet and the observer's line of sight.

Summarizing, the model is completely determined by **7 parameters**: the **region radius** R, the **magnetic field** B, the **Doppler factor** δ, the **electron density** parameter N_0, the **spectral break** γ_b and the **two spectral indexes** p_1 and p_2 (or the **turn-over energy** γ_0, the **index** s and the **spectral curvature** r).

Additional parameters are the source's *redshift* z, a *low-energy cutoff* and a *high-energy cutoff* selecting a window of the electron spectrum.

5.3 The blazar 1ES 1959+650

The blazar 1ES 1959+650 is a BL-Lac object whose synchrotron emission peaks at X-rays, so it belongs to the HBL subclass. The high-energy bump of the SED peaks in the VHE domain, as shown in Fig. 5.4. It is part of an elliptic galaxy situated nearby, at a redshift $z = 0.048$, hosting a supermassive black hole of mass $M \approx 3.6 \cdot 10^8$ solar masses [Falomo et al., 2003; Perlman et al., 1996].

It has been discovered in radio frequencies in 1991 by Gregory and Condon, 1991, while a hint of emission at TeV energies was detected for the first time in 1999 by the Utah Seven Telescope Array experiment, reaching a significance of 3.9σ [Nishiyama, 1999]. The first observation by the MAGIC-I telescope was carried out in 2004 by Tonello and the Magic collaboration, 2006, for a total observing time of $\sim$ 6 h with a 8.2σ significance. This source is one of the brightest TeV emitters, exhibiting typical low-state fluxes above 300 GeV between $10 - 35\%$ of the Crab level [V. A. Acciari et al., 2020a and references therein].

1ES 1959+650 is a favorite target of intensive observation campaigns at different wavelengths, mostly for two reasons: on one side, thanks to its high luminosity, it is relatively easy to get a full broadband sampling of its spectrum, being very luminous. On the other hand, the source stands out for its strong variability, and in particular for past episodes of **orphan flares**, that are strong outburts with no simultaneous activity increase in the X-ray flux. The first episode happened in 2002,

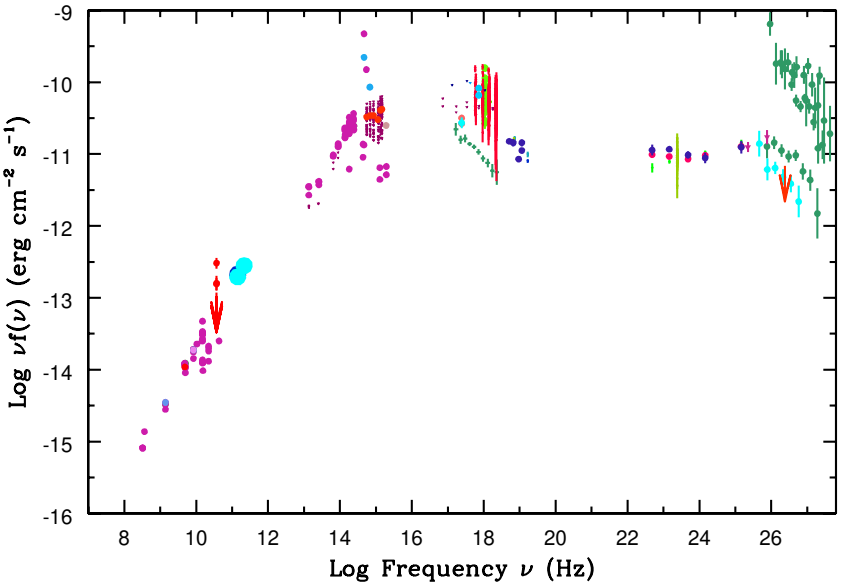


Figure 5.4: Broadband SED of the BL Lac object 1ES 1959+650 showing all the published data collected during years of observations. From the figure it can be inferred that the average low-energy peak lays in the X-ray band, while the high-energy is located in the VHE (for a reference, 1 keV $\simeq 2.42 \cdot 10^{17}$ Hz, 100 GeV $\simeq 2.42 \cdot 10^{25}$ Hz). Data taken from: tools.ssdc.asi.it.

when the Whipple telescope registered an increased activity in the TeV γ rays; the subsequent multi-wavelength campaign conducted by Krawczynski et al., 2004 found that the X- and γ-ray fluxes were generally correlated, except for an abrupt, strong γ-ray emission growth reaching almost 5 times the Crab Nebula flux, not accompanied by an X-ray counterpart. The light curve built in that study is shown in Fig. 5.5. After a quiescent state, another orphan flare was reported in 2012 by Aliu et al., 2014.

Other periods of intense VHE γ-ray activity have followed, such as the one observed in 2016, during which the γ-ray flux reached values up to 3 times the Crab Nebula level, but exhibited no orphan flares [V. A. Acciari et al., 2020a].

Such phenomena suggest that the dynamics of the jet emissions cannot be simply explained with the SSC model, that foresees a correlation between the fluxes of the two SED bumps, but instead more complex scenarios have to be taken into account, either leptonic or hadronic such as the hadronic synchrotron model proposed by Böttcher, 2005 for the orphan flares of 2002 and 2012. For all these reasons, the source is also a candidate source for the observation of high-energy neutrinos.

Intense dedicated studies on the spectral and timing charateristics of 1ES 1959+650, allowed by its brightness, are of enormous interest as they could unveil important information on the structure and mechanisms at work in the jet of BL Lac objects.

In the next chapter I will present the analysis of the observations of 1ES 1959+650 performed by the MAGIC telescopes during 2017, together with a study of the broadband emission of the source using data taken with different instruments between 2016 and 2020.

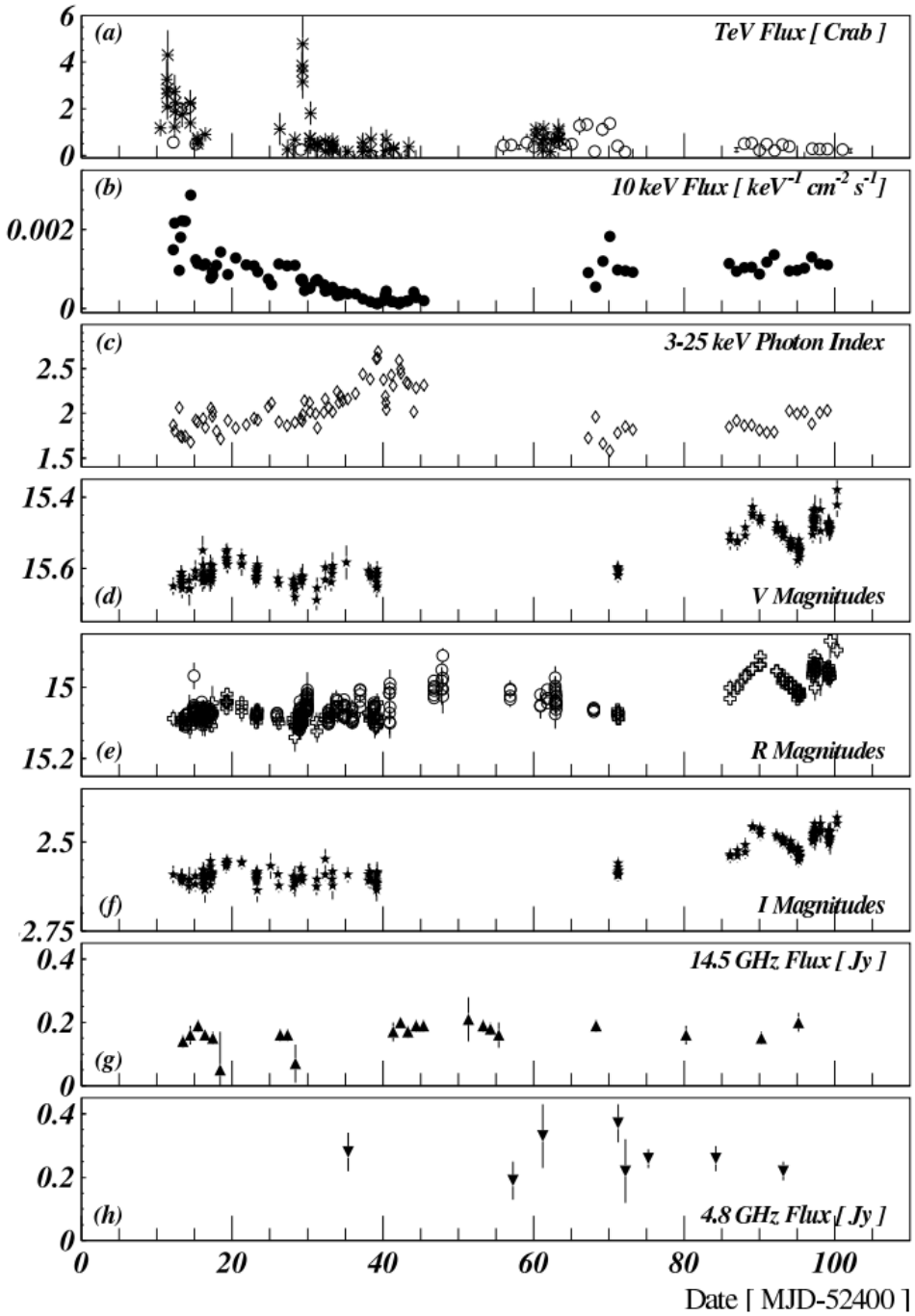


Figure 5.5: Multi-wavelength light curve of 1ES1959+650, May 18 - August 14 2002. It displays fluxes on the y-axis versus time on the x-axis. The upper panel shows the γ-ray fluxes, the second panel the X-rays: a simultaneous high activity between the two can be spotted right at the beginning of the figure, while the second γ-ray outburst (just before MJD[2] 524030) is the *orphan flare*, reaching 6 Crab Units. Image taken from Krawczynski et al., 2004.

Chapter 6

Broadband study of the blazar 1ES 1959+650

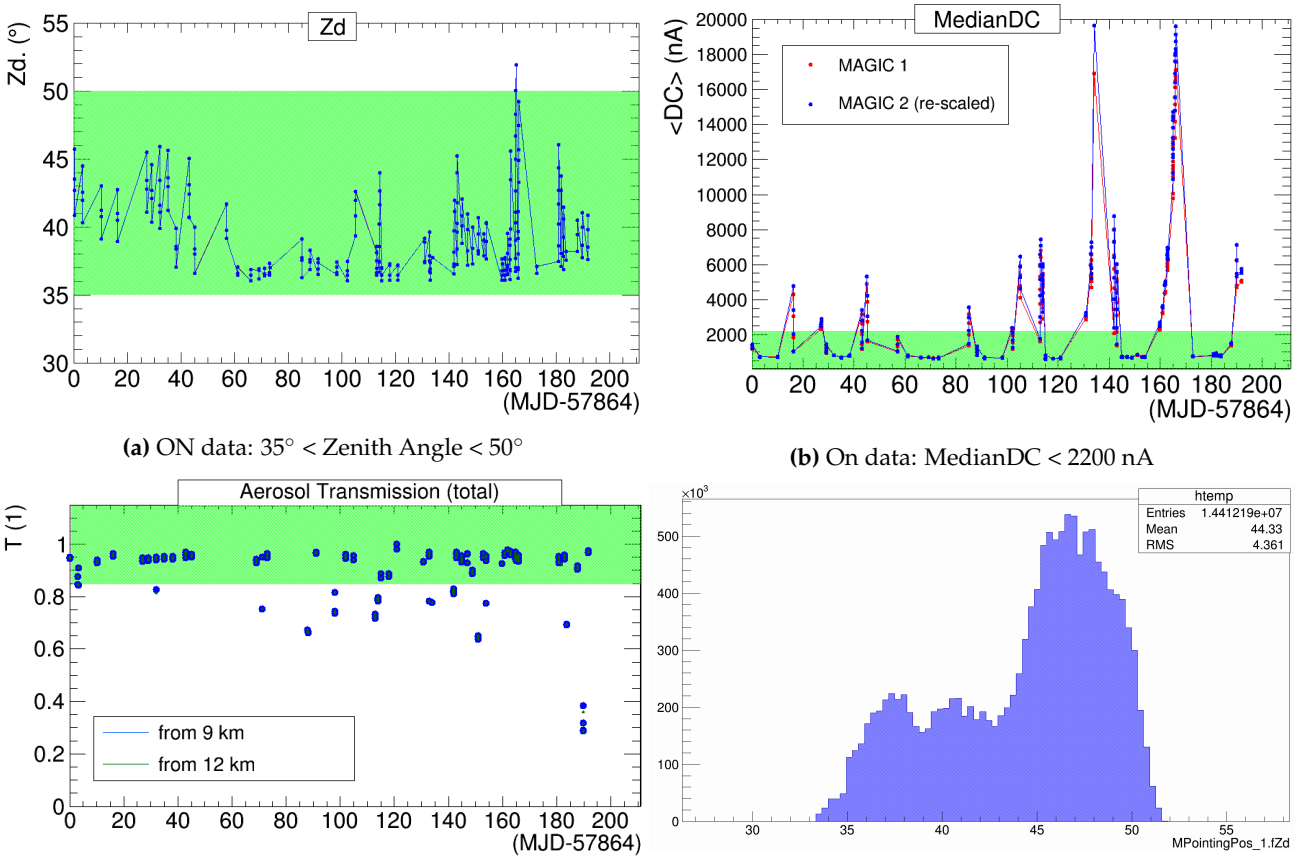
As part of the work for this thesis, I joined the group of the MAGIC collaboration that is conducting a multi-year campaign on 1ES1959+650. I performed the analysis of the data collected by the MAGIC telescopes during 2017 and started investigating the emission of the source in a multi-wavelength context, focusing on the period between 2016 and 2019.

In this chapter I will describe in detail the analysis procedure of the data sample and the main results obtained, then I will discuss the multi-wavelength behaviour of the source in the last four years of observations with different instruments. Finally, I will assemble the broadband SED observed on 13^{th} September 2017, when 1ES 1959+650 was in a high-activity state, and will model the emission with the SSC model.

6.1 Analysis of MAGIC observations taken in 2017

The analysis of MAGIC data of the source 1ES 1959+650 follows the procedure described in 2.4, using the `MARS` software package. The low level standard analysis is automated and directly performed on site, and the data - i.e. the calculated stereo parameters - are stored in `ROOT` files. As a first step of the analysis, I downloaded the `ROOT` files from the MAGIC Datacenter (PIC: Port d'Informació Cientifica). Here follows the list of the observation nights (each consisting in several 20-minute long observation *runs*):

2017/04/21	2017/06/03	2017/07/18	2017/08/20	2017/09/19	2017/10/11
2017/04/24	2017/06/05	2017/07/21	2017/08/30	2017/09/21	2017/10/19
2017/05/01	2017/06/17	2017/07/28	2017/09/01	2017/09/22	2017/10/20
2017/05/07	2017/06/21	2017/08/01	2017/09/02	2017/09/28	2017/10/21
2017/05/18	2017/06/26	2017/08/04	2017/09/10	2017/09/29	2017/10/22
2017/05/20	2017/06/29	2017/08/12	2017/09/11	2017/09/30	2017/10/26
2017/05/23	2017/07/01	2017/08/13	2017/09/13	2017/10/01	2017/10/28
2017/05/26	2017/07/03	2017/08/14	2017/09/15	2017/10/03	2017/10/30
2017/05/29	2017/07/15	2017/08/17	2017/09/17	2017/10/04	



(a) ON data: 35° < Zenith Angle < 50°

(b) On data: MedianDC < 2200 nA

(c) ON data: T(9km) > 0.85

(d) OFF data (Period 1): Zenith Angle coverage.

Figure 6.1: ON data selection: (a) Zenith angles of the observations; (b) Mean DC current of PMTs during observations. Only dark and low-moon data (DC < 2200 nA) were selected for analysis; (c) total aerosol transmission. Acceptance bands for each parameter are highlighted in green. OFF data selection: (d) ZA coverage of OFF data selected for Period 1.

Then, before processing the data with the intermediate level analysis, I performed some quality cuts. The quality selection is done according to the sky conditions, namely the Moon level and the weather conditions. The Moon level is related to the average DC current of PMTs in the camera, so in order to select observations taken under *dark/low-Moon* conditions, I excluded from the analysis data with a mean DC above 2200 nA. Indeed, the presence of the Moon increases the Night Sky Background (NSB, see 2.1.1), and consequently raises the current[1] of the camera.

The weather conditions (humidity, dust, clouds, etc.) affect the atmospheric transmittance, thus I selected data taken with an *aerosol transmission* at 9km below 85%[2].

The observations have always been taken with a ZA spanning from 35° to 50°, with few exceptions as shown in Fig. 6.1(a). The energy threshold of the MAGIC telescopes depends on ZA (in such

[1] For DC higher than the value we defined, one should apply appropriate image cleaning procedures depending on the DC level

[2] Aerosol transmission is evaluated with a LIDAR for different heights above the ground level of the telescope.

range, the threshold lays between 120-160 GeV [Aleksić et al., 2016]), so the few observations with ZA $> 50°$ have been discarded.

The total observation time after data selection amounts to 28.32 h. All the cuts applied are summarized and illustrated in Fig. 6.1.

6.1.1 MC and OFFs selection, training of the Random Forests

The training and production of the Random Forests and Look Up Tables is performed with an executable called `coach`. As input, it needs MC simulated gamma-induced shower images and a sample of OFF data containing no signal, chosen on the basis of two criteria:

- **Period:** the conditions of the telescopes vary overtime, due to instrumentation degradation (optical properties of the reflectors, point spread function..) and to hardware upgrades. Simulated data must resemble the real conditions to a high precision degree, so different periods need dedicated MC productions and OFF samples. Our observations cover two MC periods: **Period 1** from 2016/04/29 to 2017/08/02, and **Period 2** from 2017/08/02 until 2017/11/02, so I need two sets of RFs.
- **ZA:** the effective area and the energy threshold of an IACT depend on the ZA (see 2.4), so the MC and OFF data must cover adequately the same ZA range as the ON data, namely 35°-50°. See in Fig. 6.1(d) the ZA coverage of the OFF samples for Period 1.

Each MC sample is divided in two parts, one for training, the other to be used later in the analysis (6.1.3). The OFF data must undergo the same cuts applied to the ON data defined in the previous section.

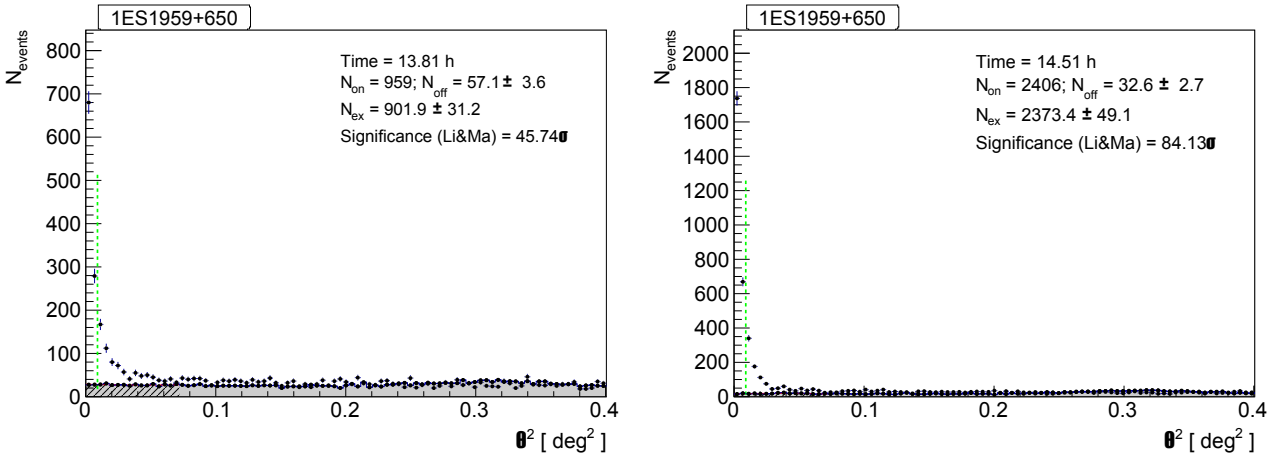
The RFs and LUTs are then used to reconstruct the data through a program called `melibea`.

6.1.2 Search of the signal

The reconstructed data are analyzed with the `odie` executable, that produces the θ^2-parameter histograms and computes the signal significance given by Eq. 2.2. The results for the three energy bands (High-Energy, Full-Range and Low-Energy), obtained applying the three sets of cuts defined in Table 2.2, are the following:

	Observation Time	Significance (Li&Ma)		
		LE	FR	HE
Period 1	13.81 h	56.25σ	45.74σ	24.41σ
Period 2	14.51 h	116.09σ	84.13σ	50.37σ

In Fig. 6.2 the two histograms obtained applying FR cuts are depicted. The high significance of the signal obtained in in both periods confirms that 1ES 1959+650 is very bright in the VHE range, and in all three energy bands investigated.



(a) Period 1 **(b)** Period 2

Figure 6.2: θ^2 plots of the two observation periods of the source 1ES 1959+650 in 2017. The histograms were filled applying Full Range cuts (see Tab. 2.2); the green dashed line defines the signal region. The high significance of the signal obtained in both the periods confirms that the source is very vright in the VHE range.

6.1.3 Spectra and light curve

As next step of the analysis I produce the differential energy spectrum and the light curve, useful to study the source emission as a function of energy and time (see 2.4.3).

In order to do that we use a program called `flute`. To build a differential energy spectrum and light curve, the data are binned in energy, and energy dependent cuts are applied (see Fig. 6.3). The cuts (on `hadronness` and θ^2 parameters) are based on the efficiency of MC data - in other words, we define the fractions of MC events that we require to survive the individual cuts, for *each* energy bin. For this analysis, I adopted a `hadronness` efficiency of 0.9 and a θ^2 efficiency of 0.75. I also applied a cut on `intensity/size` parameter, requiring a minimum of 50 phe for each event.

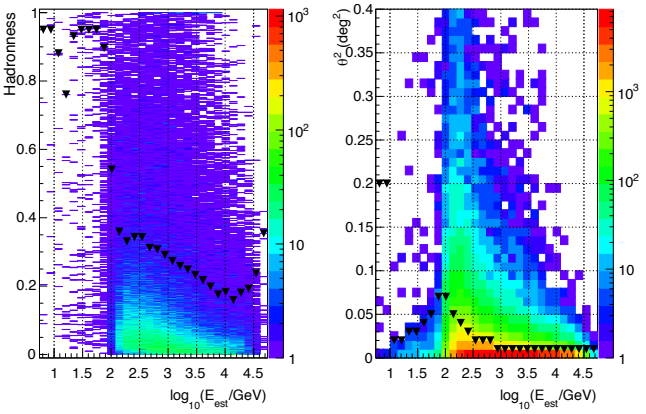


Figure 6.3: 2-D histograms of `hadronness` (left) and θ^2 (right) against the reconstructed energy. The black marks show the cut applied in every energy bin.

Now, to calculate the differential flux (Eq. 2.3), the *energy-dependent collection area* $A_{\text{eff}}(E)$ is evaluated on MC data. To get the best estimation, the MC sample should follow the same spectral shape as the source, so we:

- run `flute` a first time assuming a tentative spectrum
- fit different functions to the so obtained spectral points (listed in Tab. 6.1)

Name		Mathematical expression	
Power law	PWL	$\frac{d\Phi}{dE} = K\left(\frac{E}{E_0}\right)^{-A}$	(6.1)
Log-parabolic PWL	LP	$\frac{d\Phi}{dE} = K\left(\frac{E}{E_0}\right)^{-A-B\log_{10}(E/E_0)}$	(6.2)
PWL with exp. cut-off	EPWL	$\frac{d\Phi}{dE} = K\left(\frac{E}{E_0}\right)^{-A}\exp\left(\frac{-E}{E_{\rm CUT}}\right)$	(6.3)
PWL with super exp. cut-off	SEPWL	$\frac{d\Phi}{dE} = K\left(\frac{E}{E_0}\right)^{-A}\exp\left[-\left(\frac{E}{E_{\rm CUT}}\right)^{B}\right]$	(6.4)

Table 6.1: Analytical expressions of the typical spectral functions used in spectral analysis.

	K ($\times 10^{-10}$) [TeV^{-1} cm^{-2} s^{-1}]	E_0 [GeV]	A	$E_{\rm CUT}$ [TeV]	B	χ^2/NDF
Period 1 (EPWL)	2.2 ± 0.1	321.46	2.15 ± 0.08	2.5 ± 0.7	-	8.3/11
Period 2 (SEPWL)	5.7 ± 0.6	306.99	1.90 ± 0.16	5.1 ± 2.3	1.3 ± 1.4	14.4/12
low state (EPWL)	2.3 ± 0.2	286.50	2.2 ± 0.1	1.9 ± 0.7	-	9.3/9
highest state (LP)	7.3 ± 0.2	321.46	2.05 ± 0.05	-	0.31 ± 0.08	21.4/12
2017/09/13 (LP)	3.9 ± 0.2	344.45	1.81 ± 0.08	-	0.32 ± 0.10	12.4/13

Table 6.2: Best fit parameters to the differential photon spectra. Note that the parameters share the same letters for simplicity, but are intrinsically different belonging to different functions.

- select the one with the least $\chi^2/$*degrees of freedom*)
- re-run the executable with the updated, more appropriate spectral shape.

For each period I determined the best average spectral shape among the functions defined in Appendix 6.1: a **power law with exponential cut-off** (EPWL, see Eq. 6.3) and a **power law with super exponential cut-off** (SEPWL, Eq. 6.4) for Period 1 and Period 2 respectively. The fit results are listed in Table 6.2.

In Fig. 6.4 the night-wise light curve is illustrated, that shows the evolution of the integral flux of photons $>$ 300 GeV over time in 2017. The weighted averaged flux is $(4.45 \pm 0.08) \cdot 10^{-11}$ photons/cm^2/s. 1ES 1959+650 shows a the erratic trend typical of HBLs: after a general low-state during the first part of the year, exhibiting fluxes around 1/4 of the Crab level, around the half of August the activity increased approaching the Crab level during some nights of September, and exceeding it during the second half of October. In order to visualize the spectral variation of the source, in Fig. 6.5 I show two SEDs corresponding to the flux extrema: the red curve belongs to 2019/10/19 - the night of highest activity, with an integral flux of $(1.46 \pm 0.05) \cdot 10^{-10}$ cm^{-2} s^{-1} - whereas the blue one is the *low-state* SED, obtained averaging on the nights exhibiting an integral flux below $0.4 \cdot 10^{-11}$ cm^{-2} s^{-1}. The best fit to the spectra are an EPWL and a **log-parabolic power**

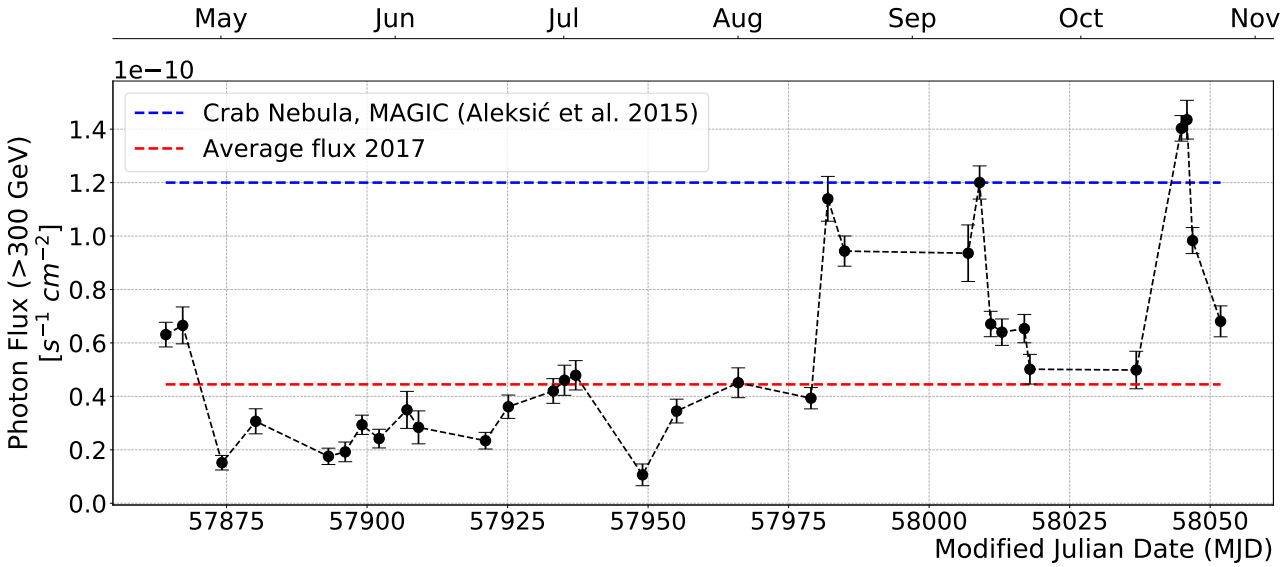


Figure 6.4: VHE light curve of 1ES 1959+650 during 2017. The activity increased in August (MJD 57975), with fluxes reaching and exceeding the Crab level in October (MJD 58025).

law (LP, Eq. 6.2) respectively, whose parameters are also reported in Tab. 6.2.

VHE Spectral Energy Distribution during 2017-09-13

For the purpose of the multi-wavelength study, I focused on 13^{th} September, a particular night in which 1ES 1959+650 was simultaneously observed in different bands by other instruments (see 6.2.4). During that night the blazar was in a high state, exhibiting an integral flux above 300 GeV of $(1.20 \pm 0.06) \cdot 10^{-10}$ cm^{-2} s^{-1}.

Since VHE γ rays are absorbed by the EBL, as described in 1.2.2, the differential flux is underestimated, especially at the highest energies. Giving the redshift of the source - $z = 0.048$ - as input of `flute`, the points are corrected using the absorption model described in Domínguez et al., 2011. In Fig. 6.6 both the observed and the deabsorbed SED are shown. The deabsorbed spectrum has been fitted with a LP (Eq. 6.2) whose parameters are reported in Tab. 6.2.

6.1.4 Unfolding

Once we have the spectral points as function of the reconstructed energy, an unfolding procedure needs to be applied in order to retrieve the spectrum in the true energy domain, as described at the end of the section 2.4. This task is fulfilled using a `ROOT` macro called `combUnfold.C`. Two energy ranges in both the E_{RECO} and E_{TRUE} are determined by the user on the basis of the data, and the migration matrix is then inverted within those ranges by means of an unfolding algorithm. I unfolded the spectrum of 13^{th} September with 4 different methods, whose results are presented in Fig. 6.7. As we can see from the figure, the spectral points obtained are in very good agreement within the errorbars.

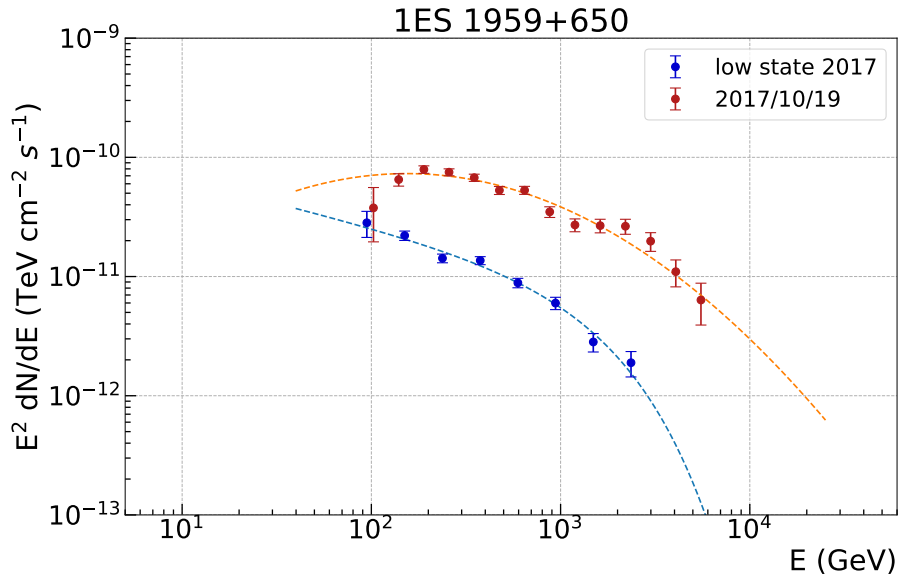


Figure 6.5: Observed SED of 1ES 1959+650 during 19th October and low-state average SED. The energy binning is different in the two cases: the former SED is obtained with the default 30 bins from 5 GeV to 50 TeV, whereas the latter with just 20 bins, due to poor statistics.

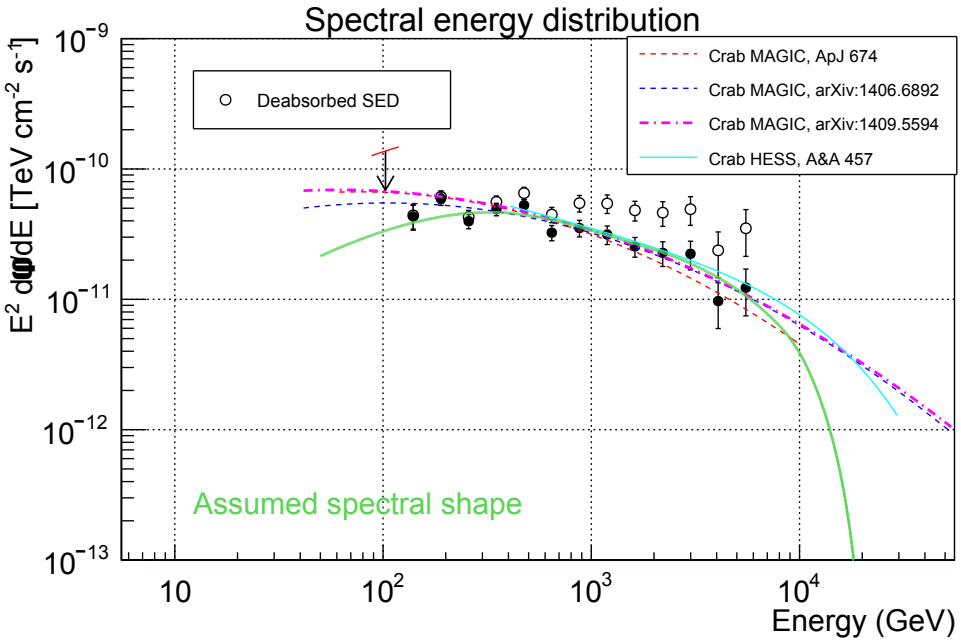


Figure 6.6: Spectral energy distribution of 1ES 1959+650 in 13th September 2017.

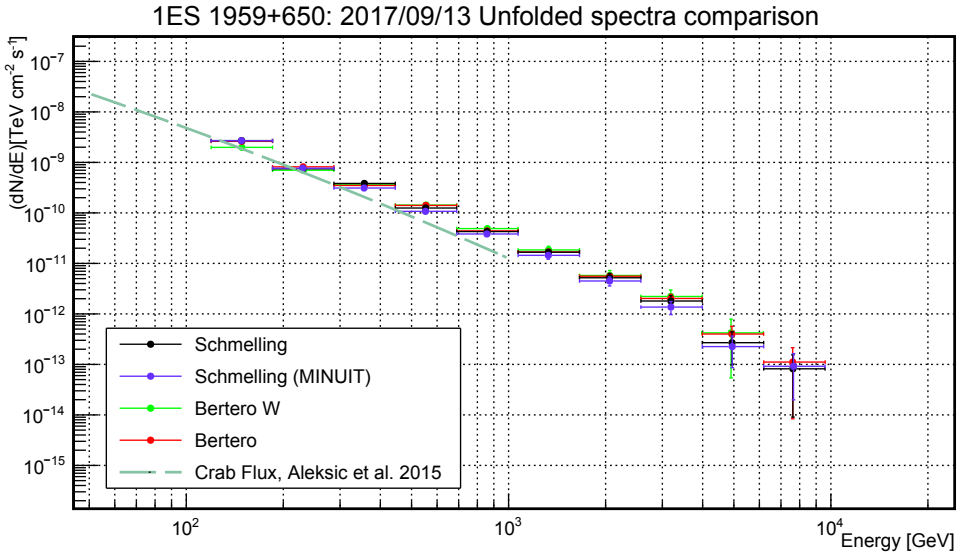


Figure 6.7: Differential flux of 2017/09/13 unfolded with four different methods; the obtained spectral points are in good agreement.

6.1.5 Crab test

Before proceeding with the data analysis, in order to make sure that the Random Forests and Look Up Tables work properly I performed a test analysis on observations of the Crab Nebula - the standard candle of VHE gamma astronomy. Concerning Period 1, I analyzed the observations of 8^{th}, 9^{th}, 10^{th}, 11^{th} and 23^{rd} of October 2016, to which were applied the same Zenith, Mean DC and quality cuts defined for 1ES 1959+650. I collected a total of 3.74 hours with 34.74σ significance in the FR energy band. The final results agree with the values published in literature (see also Fig. 6.8):

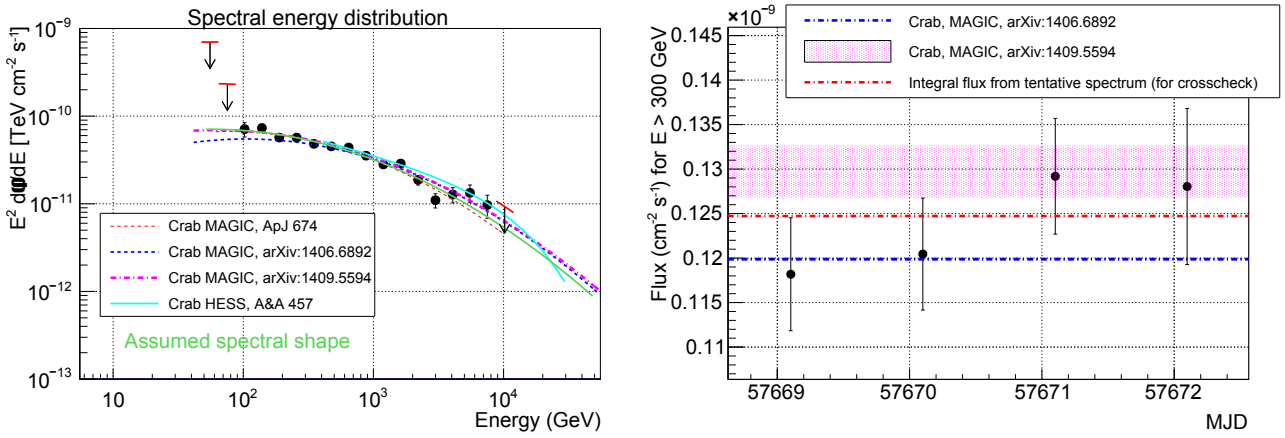
	Avg. integral flux $[10^{-10}$ cm^{-2} s$^{-1}]$	Compatibility
2017	$1.23 \pm 0.04_{stat}$	0.30
Aleksić et al., 2015	$1.20 \pm 0.08_{stat}$	

Unfortunately, no Crab Nebula data was taken in dark conditions during Period 2, so this test could not be done in that case.

6.1.6 Cross-check analysis

Reproducibility is a cornerstone of the scientific method, so to validate my analysis I cross-checked my light curve with the one produced independently by another analyzer of the analysis group. In order to make a fair comparison, we both applied the same quality cuts described in 6.1 and in 6.1.3[3], and assumed the same spectral shape (a power law of index A = -2, see Eq. 6.1).

[3]except for the `size` cut, set to 60 phe in this case.



(a) **(b)**

Figure 6.8: (a) SED of the Crab Nebula (black markers) compared with spectral measurements published by the H.E.S.S. and MAGIC collaborations; **(b)** light curve of the Crab Nebula. The integral flux (red dashed line) is in very good agreement with the values published in literature.

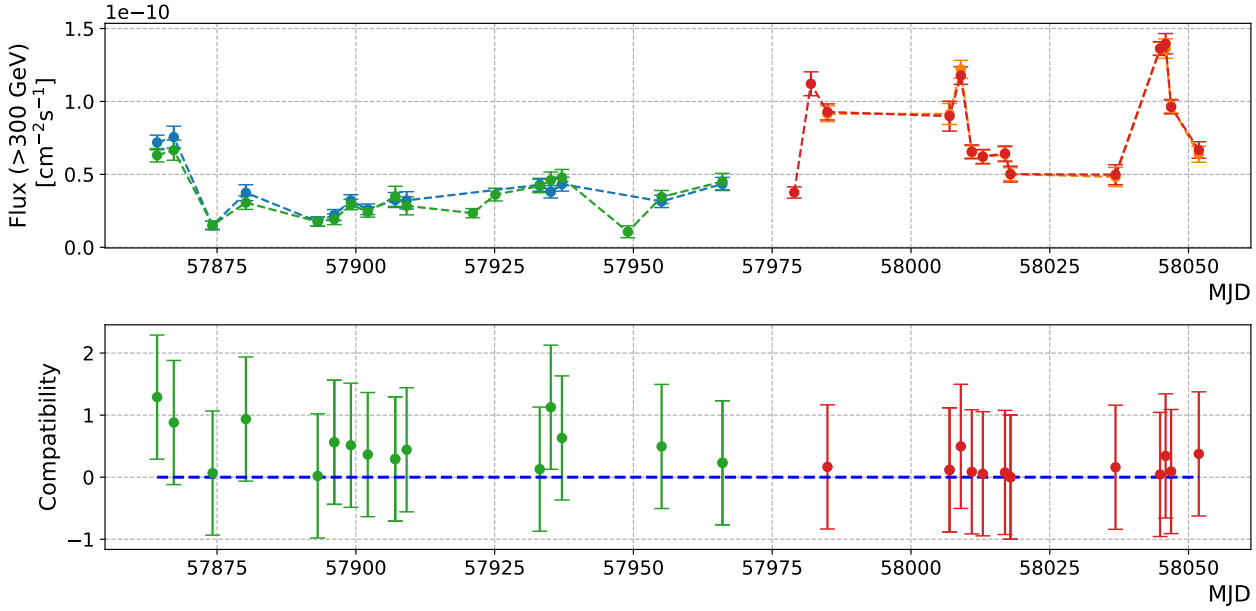


Figure 6.9: Comparison with the cross-check analysis. The two analysis periods are separated: blue and orange points are by the cross-checker, green and red are from this analysis

The daily fluxes are plotted together in the upper panel of Fig. 6.9, while the lower panel shows the compatibility defined as:

$$\eta = \frac{|F_1 - F_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{6.5}$$

being F_1 and F_2 the fluxes, σ_1 and σ_2 the respective errors. The light curves are overall very consistent with each other, specially in Period 2: this is a reassuring sign that the analysis was performed properly, and in a sense overcomes the lack of Crab Nebula data discussed above. There is however a mismatch in the number of nights considered: I analyzed more than the cross-checker. Since we used two different programs for data selection, this issue is probably due to a bug in the software, and it is currently under investigation. Anyway, this does not affect at all the good agreement of the results.

6.2 The multi-wavelength view

In order to characterize the emission of a source it is important to monitor its behaviour at different energy bands, taking simultaneous observations with different instruments. This motivates long-term **multi-wavelength campaigns** that allow to have deeper insight into the emission mechanisms at work in the jets and test/probe different models.

I joined my results together with the analysis of the MAGIC observations taken between 2018 and 2019, to build a long-term VHE γ-ray light curve of 1ES 1959+650. Collecting also the data already published in V. A. Acciari et al., 2020a, the period was further extended back to 2016, when the source exhibited an exceptionally high activity at TeV energies. The whole data cover the period from 2016/04/29 to 2019/12/04.

Then, I gathered the data taken by several facilities, covering the electromagnetic spectrum from the radio band up to the high-energy domain, to build the **long-term light curve** shown in Fig. 6.10.

I was able to collect this whole amount of data only thanks to the precious work of members of the analysis group of MAGIC and external collaborators in the multiwavelength field; all these analysis will be part of a publication.

6.2.1 Observations by other facilities

***Fermi* LAT**

The **Large Area Telescope (LAT)** is a detector on board of the *Fermi* satellite [Atwood et al., 2009]. Focused on HE γ rays, with energy ranging from 100 MeV to 300 GeV, its principle of detection is based on pair production. It is on operation from its launch in 2008, conducting all-sky surveys. The signal of 1ES 1959+650 was selected between 2016/11/24 and 2019/12/06 and in the **energy range 0.3 - 300 GeV**, in a region of interest of 20° around the position of the source. The data were analyzed with two time binnings, one with **bins of 3 days**, the other of **7 days**. We also collected

the 3-day-binned data between 2016/04/28 and 2016/11/24, already published in V. A. Acciari et al., 2020a.

Swift XRT

The **X-Ray Telescope (XRT)**, on board of the *Swift* satellite, is on operation since 2004 [Burrows et al., 2005]. It detects photons in the **energy window 0.3 - 10 keV**. We collected the XRT observations of 1ES 1959+650 carried out between 2016/04/30 and 2016/12/06, and obtained the **daily fluxes** by spectral analysis, applying the Log-Parabolic model (eq. 6.2) and taking galactic absorption into account.

Swift UVOT

The **Ultra Violet/Optical Telescope (UVOT)** on board of *Swift* observes the sky in 6 different photometric bands: the **optical V, B, U** and the **ultraviolet W1, M2 and W2**. Observations of the source were carried out in the same period as XRT, and were corrected on the galactic absorption.

Tuorla blazar monitoring program

Data in the **R optical band** were collected through the **Tuorla Observatory blazar monitoring program**, that builds optical light curves using many telescopes. The observations, taken from V. A. Acciari et al., 2020b, cover the period 2016/05/01 - 2019/12/09; the fluxes are corrected for galactic extinction and the host galaxy flux is subtracted.

OVRO

The **Owen Valley Radio Observatory (OVRO)** is located near Big Pine, California (USA) [Richards et al., 2011]. The OVRO 40 m radio telescope is used for blazar monitoring, observing at **15 GHz**. I collected the light curve publicly available produced from 2016/05/01 to 2020/01/23.

6.2.2 Long-Term Light Curve

The light curve in Fig. 6.10 represents the integral emission of the source at different wavebands. In the top panel is shown the VHE emission measured with MAGIC. We can clearly distinguish the four observation cycles corresponding to 2016, 2017, 2018 and 2019. The "holes" between a cycle and another are due to the duty cycle of MAGIC: during the period November-April the source crosses our sky during daytime. Same reasoning applies to Tuorla (fifth panel), that uses ground-based telescopes. The OVRO telescope, observing at radio frequencies, isn't affected by this problem, and neither are LAT, XRT and UVOT being space-borne experiment.

VHE: 2016 was characterized by the highest VHE activity, with an average flux of $(8.78 \pm 0.08) \cdot 10^{-11}$ photons/cm^2/s, then the emission slowly decreased, except for a bump in the second half of 2017 (see also 6.1.3) and a smaller one towards the end of 2018. In 2019 we observe the quietest state, with an average flux of $(1.60 \pm 0.03) \cdot 10^{-11}$ photons/cm^{2}s. The maximum to minimum ratio across the whole period is $max/min = 47.52$.

HE: the high state of 2016 seems to be confirmed also by the fluxes registered by the LAT, shown in panel 2. The HE activity follows the same descending trend of the VHE counterpart; the max to min ratio, calculated for the 7-day binned data, is $max/min = 14.29$. The VHE and HE fluxes are related to the high-energy peak of the SED.

X-rays: on the third panel the energy fluxes measured with XRT are displayed; they are associated with the emission of the first peak of the SED. The behaviour here is quite different, with an exceptionally high outburst in the second half of 2017, reaching more than two times the average flux of the other periods, generally below 10^{-1} erg/cm^{2}s. The ratio is $max/min = 13.06$.

UV/Optical: the UV and optical fluxes (UVOT, panel 4; TUORLA, panel 5) show an erratic trend; the higher max-to-min variation is exhibited by the V band, with a ratio 4.13, while for the other bands it lays around 2.3-2.7.

Radio: finally, the last panel shows the emission observed by OVRO. Even though a little enhanced radio activity seems to be spotted during 2019, this band has the lowest max-to-min ratio of 1.76.

6.2.3 X-rays and γ rays correlation

Detailed intra-band studies looking for possible correlations between the fluxes are of great importance for blazars: indeed, the manifestation of correlated variability between bands suggest that the emitting particles have a common origin. Unfortunately, these investigations are limited by lack of simultaneity: "zooming" the light curve, one immediately realizes that observations are often taken with non-neglectable time gaps.

The first relation to be investigated is between X-ray and VHE γ-ray emissions, that for a HBL are related to the SED peaks. In order to do that, we need to re-define simultaneity. I proceeded like this: to each X-ray flux, I assigned the closest-in-time VHE flux, with an acceptance gap of half

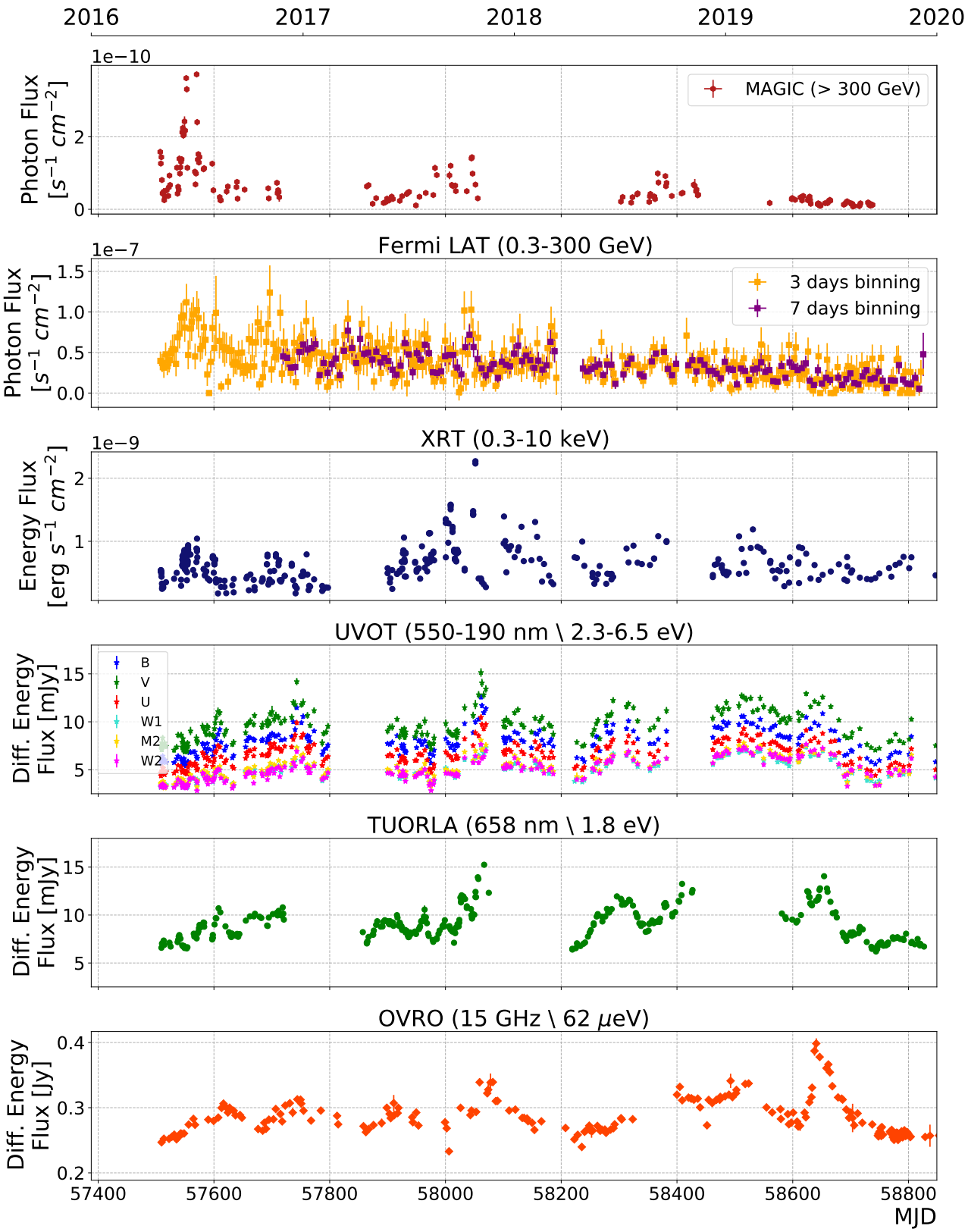


Figure 6.10: Long-term multi-wavelength light curve of 1ES 1959+650 between 2016 and 2020.

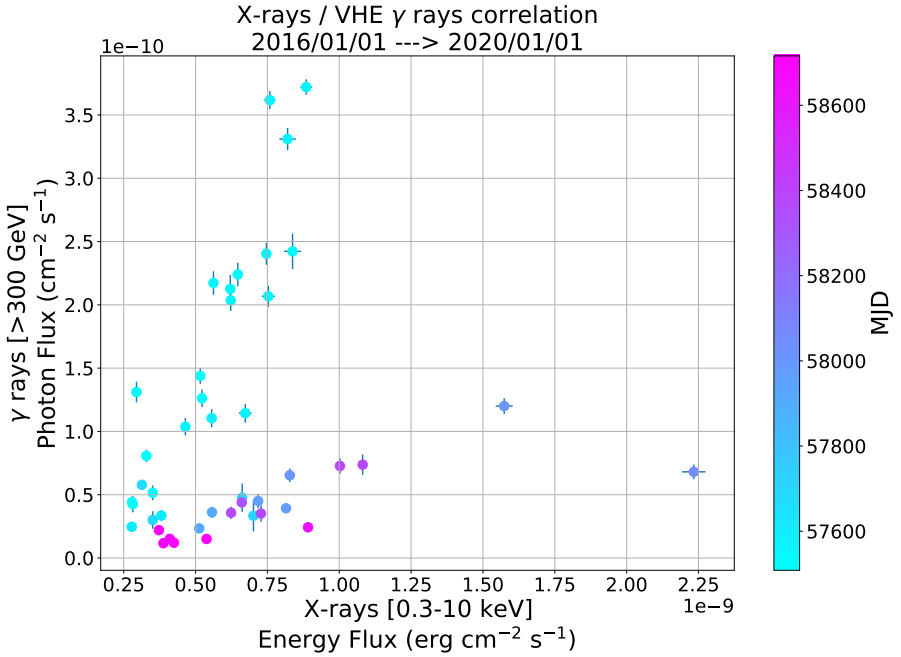


Figure 6.11: VHE fluxes against X-ray fluxes for simultaneous observations conducted between 2016 and the end of 2019. Color represents the time axis. Two distinct trends are clearly visible: the cyan-blue points belong to 2016, the others to the 2017-2019 period.

a day. In this way I got 43 simultaneous observations (over 4 years) with an average time gap of 3.4 hours. In Fig. 6.11 the γ-ray fluxes are plotted against the respective X-ray counterparts. The colour of each point represents time (see the colorbar). The result is intriguing: two distinct trends are observed, related to different periods. The steepest trend belongs to 2016, year of the strong γ-ray flares, whereas the other one is populated by the data of 2017 - when the blazar showed strong X-ray outbursts - 2018 and 2019. It is interesting to note that although the ratio between the two fluxes has changed, still they are correlated, but in a different way. To quantify the correlation, I computed the **weighted Pearson correlation coefficient** (R_W, see Appendix C.2) for two periods separately: the data from 2016 return a coefficient $R_W = 0.77$, confirming the value 0.76 measured by V. A. Acciari et al., 2020a[4]. The data from 2017 to 2020, displayed separately in Fig. 6.12, return $R_W = 0.83$. Technically, one should compare the same quantity (i.e. energy flux, or photon flux), and this will be addressed in the near future, although I don't expect dramatic changes to happen, as confirmed by the fact that for the 2016 data I get the same result as the paper.

The SSC model (see 5.2.2), that succeeds in modeling many of the observed blazar emissions, foresees a correlation between the two fluxes. If we assume to be in the context of this model, such a change in the trend may be explained with a changed condition in the acceleration environment, that could be either the electron density N_0, or the magnetic field B. These two parameters indeed separately regulate the *compton dominance*, i.e. the ratio of the IC peak's emission (VHE) to the synchrotron peak's emission (X-ray).

[4]The small difference is ascribable to the different time windows used (2.5 days in the paper, versus our 1-day-wide bins).

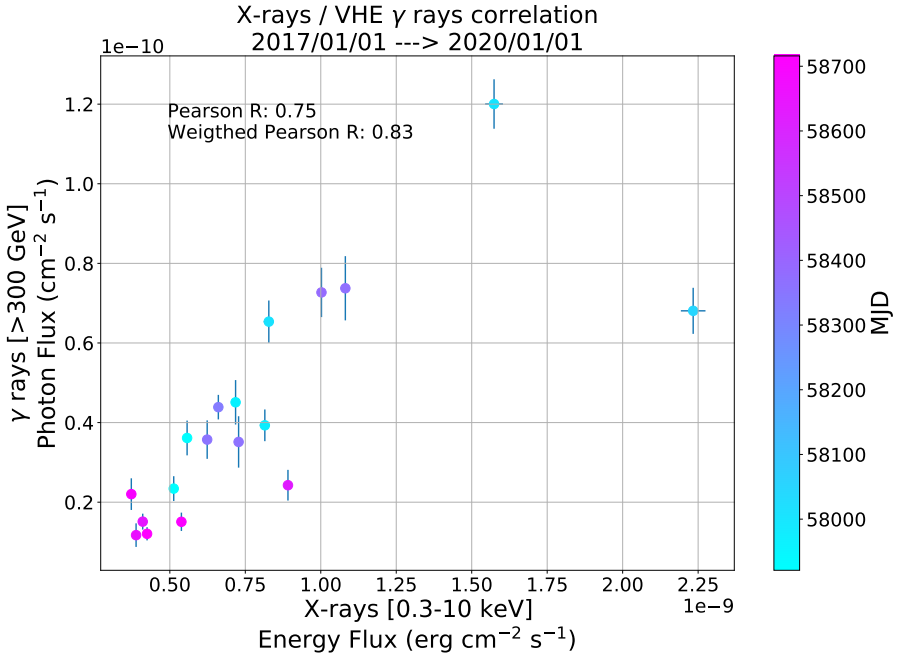


Figure 6.12: VHE fluxes against X-ray fluxes for simultaneous observations conducted between 2017 and the end of 2019. Color represents the time axis. The values of correlation factors clearly indicate a correlation between the two fluxes.

No correlation was found between X-rays and UV/optical bands, nor with radio; anyway, it is not to be excluded that a delayed correlation may exist between these bands, due to reprocessing mechanisms. That could be investigated with tools such as the Discrete Correlation Function (DCF), left for future work.

6.2.4 Fitting the SSC to the SED of 2017/09/13

As a final step, I modeled the SED of 1ES 1959+650 measured on 13th September 2017, when the source was showing one of the highest VHE activity of 2017 (it is the highest point in Fig. 6.12). Unfortunately, the nights of 18th-19th October - which had the highest fluxes of the year - could not be used due to lack of simultaneous observations in the X-rays.

In Fig. 6.13 the SED that I assembled is presented. The groups of spectral points, from left to right, are:

- **five spectral points in the optical/UV**, taken by UVOT in the same time window of the XRT data (see below). These are obtained from the differential fluxes also plotted in Fig. 6.10, from which I removed the contribution of the host galaxy as follows: 1.1 mJy, 0.4 mJy and 0.1 mJy were subtracted for V, B and U filters respectively (according to Tagliaferri et al., 2008 and S. Patel et al., 2017). At UV frequencies (W1 and W2 filters) this contribution is negligible.
- **26 X-ray spectral points**; XRT observed the source on 2017/09/12 at 23:03 CET for a total time 680s.

- **three *Fermi*/LAT points**. To collect enough statistics, the signal was integrated across 4 days in which the source was in a high state: from 2017/09/09, 12:00 to 2017/09/13, 12:00.
- **ten MAGIC points** obtained from the unfolded spectrum calculated in 6.1.4 with the Bertero algorithm; MAGIC started observing the source on 2017/09/12 at 22:59 CET for a total time of 74 min.

Comparing our spectral points (black markers) with the past emissions (grey points), we find an extraordinary, unprecedented X-ray activity; the X-spectrum is harder with respect to the past, i.e. the peak position is clearly higher than the typical value of $\sim 10^{17.5}$ Hz. The γ-ray activity is instead consistent with archival values.

For the modeling, I used the `JetSeT` framework[5] (**Jet**s **SE**D modeler and fitting **T**ool), an open source C/Python framework to simulate radiative and accelerative processes in relativistic jets, and to fit numerical models to multi-wavelength SEDs [Tramacere et al., 2009; Tramacere et al., 2011]. I assumed the one-zone SSC model described in 5.2.2, that consists in a spherical emission zone of radius R moving with a bulk Lorentz factor $\Gamma_{\mathbf{bulk}}$ and forming an angle θ with the observer's line of sight; the chosen electron spectral distribution is the log-parabola with power-law branch defined in Eq. 5.2.

The fitting procedure is based on three steps:

- calculation of phenomenological parameters, such as the peak frequencies and emissions, and the spectral indexes in the different bands;
- numerical model constraining: the physical parameters of the model (the radius R, the Doppler factor δ...) are constrained by the previously extracted parameters;
- numerical model optimization: the best parameters are determined through a minimization method. Amongst the minimizers implemented in the framework, I used two: the *least-squares bound minimizer* of SciPy, and `MINUIT`.

In the first step, the synchrotron peak and the IC peak are interpolated with a log-parabolic power law (Eq. 6.2). The resulting frequencies and fluxes are:

	$\log(\nu_P)$ [Hz]	$\log(\nu_P F_{\nu_P})$ [erg cm^{-2} s^{-1}]
Synchro	18.3 ± 0.4	-9.23 ± 0.03
IC	25.4 ± 0.2	-10.09 ± 0.04

The physical parameters were not severely constrained due to the lack of data in some of the energy bands, for instance in the hard X-rays; thus I needed to fix some of them before the minimization process, looking for typical values in literature.

Both the minimizers used reach a reduced chi-square $\chi^2/NDF \approx 20/9$; the obtained curves are shown in Fig. 6.13. The SSC model has also been configured to take EBL absorption into account,

[5] https://github.com/andreatramacere/jetset

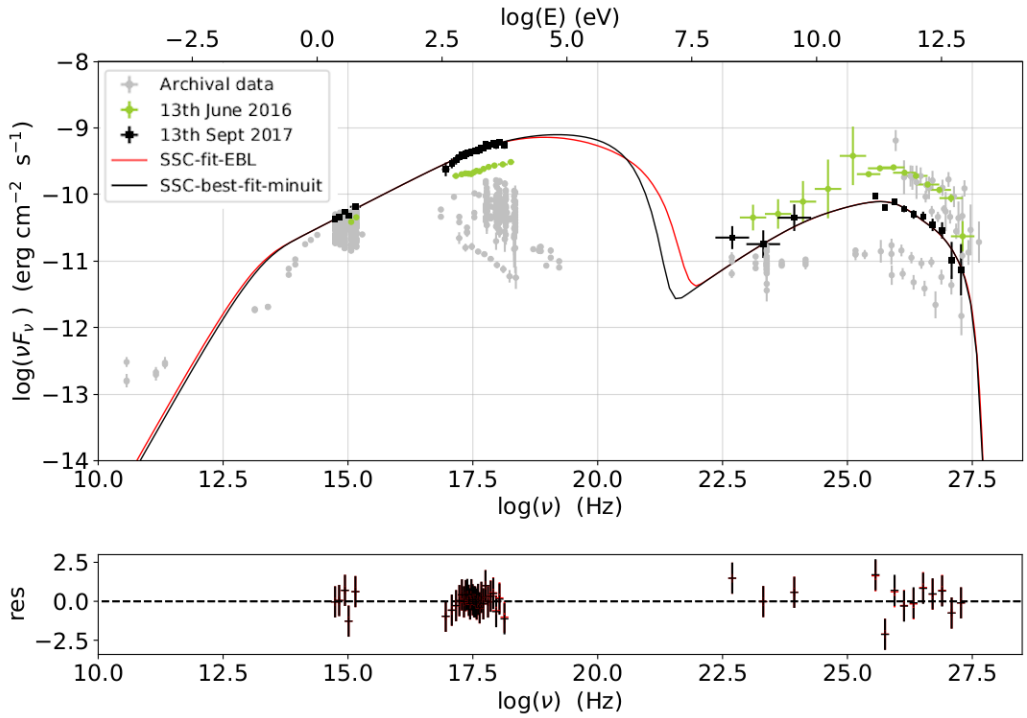


Figure 6.13: Broadband SED of 1ES 1959+650. Black markers are from 13^{th} September 2017, grey points are from past observations. Black and red lines are the SSC models obtained in this work, minimized with `MINUIT` and *SciPy least square bound* minimizers respectively. The SED observed on 13^{th} June 2016 (light green points) is shown for comparison.

using the EBL model proposed by Domínguez et al., 2011, and its the effect can be seen in Fig. 6.14, where the synchrotron and the IC components of the emission are separately shown.

The results of the fits are reported in Table 6.3, where they are compared to the parameters obtained by V. A. Acciari et al., 2020a for the flares of 2016/06/13 and 2016/06/14. These two flares were explainable with the SSC model - except for 2016/06/13, that required a relatively too high Doppler factor (> 30) to explain the spectral shape of the IC bump. The common parameters I obtained are of the same orders of magnitude.

The modeling I presented gives reasonable results that are consistent with typical literature values, suggesting that the SSC model could be the mechanism underlying the emission of 1ES 1959+650 during the night of 13^{th} September 2017. However, there is a clear degeneracy in the space of parameters that does not allow to derive strong conclusions on the emission mechanisms at work. Further steps to probe the jet physics include the collection of more data covering a wider energy range, and the study of the SED evolution over time with the highest time resolution possible. The advent of CTA, combined with strictly simultaneous MWL observations, will certainly boost the research capabilities in this field.

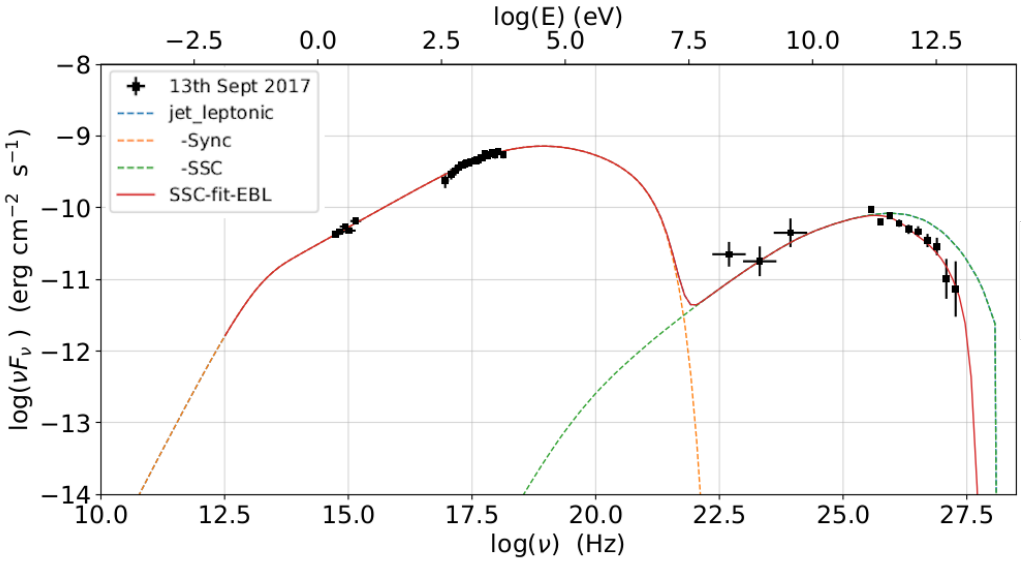


Figure 6.14: Broadband SED of 1ES 1959+650 in 2017/09/13 fitted with the SSC model, minimizing the *least squares function*. The synchrotron and IC contributions to the emission are shown separately. The modeling took EBL absorption into account, as can be deduced from the deviation between the IC component and the VHE spectral points.

Parameters	2017/09/13	2016/06/13	2016/06/14
$\gamma_{\rm min}$	$(9.4-9.9)\cdot 10^2$	$7\cdot 10^2$	$(3-7)\cdot 10^2$
$\gamma_{\rm max}$	$(5.0-8.8)\cdot 10^6$	$(1-7)\cdot 10^6$	$(1-7)\cdot 10^6$
γ_0	$(1.08-1.33)\cdot 10^2$	-	-
s	2.23	-	-
r	$0.38-0.48$	-	-
γ_b	-	$4-10\cdot 10^5$	$(1-5)\cdot 10^5$
n_1	-	$2.2-2.3$	$2.2-2.3$
n_2	-	$3.2-3.3$	$3.2-3.3$
R (cm)	$(1.00-1.05)\cdot 10^{16}$	$(0.7-1)\cdot 10^{15}$	$(0.8-1)\cdot 10^{15}$
B (G)	$0.14-0.17$	$0.10-0.25$	$0.2-0.4$
δ	$27-28$	$40-60$	$30-50$

Table 6.3: Parameters of one-zone SSC modeled on our SED, and on the two flares of 2016 reported by V. A. Acciari et al., 2020a. The parameters generally share similar values, and the same order of magnitude. Note that for the 2016 flares a different electron spectral shape was used, i.e. a broken power law (see Eq. 5.1).

Chapter 7

Conclusions and outlook

This thesis has approached the topic of VHE γ-ray astronomy by not only performing a full analysis and interpretation of the data taken on 1ES 1959+650 using MAGIC, but also developing a novel analysis technique applied to a new generation of telescopes.

In the first part of this thesis, I addressed the problem of event reconstruction for the Large Sized Telescope of CTA using the technology of CNNs: all the VGG-style networks I implemented perform better than the standard Random Forest technique. I also assessed the impact of data quality on the analysis, defining four cut levels (`no-`, `low-` `mid-` and `high-cuts`) involving four different energy thresholds. With no surprise, the `no-cut` model is always the least performing in all the tasks, since it is trained and tested on many not well-contained and faint images.

On the **background rejection** task, we show that the networks are very good at discriminating high-energy events above 500 GeV - where images have a higher quality - reaching AUCs > 0.99, while they lose accuracy in the lowest energy range below 50 GeV. The CNNs outperform the RFs at all the cut levels, leading up to an AUC improvement of 15.5%, 9.6% and 3.4% in the `low-cut`, `mid-cut` and `high-cut` case respectively. The CNNs are so powerful in this task with respect to the standard analysis that even the `no-cut` VGG turns out to be a better classifier than the `mid-cut` RF.

Concerning **energy reconstruction**, the `low-`, `mid-` and `high-cut` models show similar trends, achieving the best resolution around 2 TeV, with a value around 15%. We get very good results in this task comparing to the RFs, especially in the lowest (below 150-200 GeV) and in the highest energy ranges (above 10 TeV), where the networks lead to improvements of more than 40%. In the middle-range between 200 GeV and 2 TeV, the CNNs perfrom better than the RFs by 10% up to 25%.

The models give their best performance on the arrival **direction reconstruction**, where the enhancement with respect to the standard analysis is very remarkable: the RFs are overtaken by up to *more than* 60% for energies under 150 GeV, by 20 - 40% between 150 GeV - 10 TeV and by more than 40% above 10 TeV. With `low-cuts`, that focus on the lowest energies, we get an angular resolution spanning from 0.4° down to 0.18° when approaching 150 GeV; above 150 GeV and below 2 TeV, using `mid-cuts` we get values between 0.11° - 0.16°. A very interesting result is achieved in the 2-10 TeV range, where an angular resolution of 0.08° - 0.09° is reached by the models - the range is even extended down to 300 GeV if we consider only high-quality images.

Finally, we showed that the outcome of this analysis is consistent with the results independently obtained by two other research groups that have developed a similar analysis using different network architectures implemented in different libraries: the DL group at the Universidad Complutense de Madrid (UCM) that develops and uses the library *CTLearn* and the γ-ray group at the Laboratoire d'Annecy De Physique Des Particules (LAPP) that develops and uses the library *GammaLearn*. All this confirms that CNNs are an effective and reliable tool when applied to IACT event reconstruction, providing solid and replicable results, and, most importantly, performing remarkably better than the standard parameter-based analysis method currently used. There is surely room for improvement and for testing new, powerful state-of-the-art DL techniques, even though it seems that analysis perfomances at this point are dictated more by data quality than complexity of the algorithms. Next steps of this work will explore custom models concatenating in the input the images with the extracted parameters, in order to fully exploit all the information contained. Moreover, the LST is currently under the commissioning phase, taking data and fine-tuning its performance, thus a next stage of this work will be the application of the CNN analysis to real data.

The second part of this thesis was devoted to the detailed analysis of the VHE γ-ray emission of the blazar 1ES 1959+650 collected by the MAGIC telescopes during 2017. I first tested the analysis chain on a sample of Crab Nebula data, obtaining results compatible with those published by the MAGIC collaboration. After that, I applied the analysis to the source, obtaining the differential energy flux and the light curve above 300 GeV. These results were validated by a cross-check analysis independently performed by another analyzer of the MAGIC collaboration.

Then I conducted a broadband study of the emission of 1ES 1959+650 in the last four years at different wavelengths. Studying the long-term multiwavelength light curve, I could observe the high variability typical of this kind of objects, in particular in the VHE γ-ray and X-ray bands. 2016 was characterized by a very high activity at VHE γ rays, then the emission slowly decreased, except for a smaller bump in the second half of 2017. Conversely, the source exhibited a moderate X-ray activity in 2016, showing an outburst during 2017 and returning back again to a quiescent state in the next years. Related to these two phases, I found two distinct correlation trends between the γ-ray and X-ray fluxes: in both periods, an increased emission in one band was accompanied by an increase in the other, but with a different ratio. Such a change in the trend may be explained with a changed condition in the acceleration environment, and in particular, in the context of the Synchrotron Self Compton model, it can be ascribed to a different intensity of the magnetic field or a different electron density within the jet.

Finally, I assembled the broadband SED observed on 13th September 2017, modeling its emission with the SSC model: the physical parameters I found are consistent with the typical literature values foreseen for a BL Lac object like 1ES 1959+650.

These results are a contribution to a wider, promising research which will probe the emission mechanisms at work in this source. Amongst the next steps to be done there are the extension of the monitoring period back to 2015 and a detailed intra-band study to find other potential delayed

correlations between the fluxes. A topic of particular interest for the future is the study and characterization of the time evolution of the SED of this blazar, which will give deeper insights on the jet physics of AGNs.

Appendix A

VGG13 architecture

The architecture I implemented is a VGG13, adapted from Simonyan and Zisserman, 2015. VGG-style networks use very small receptive fields in the convolutional layers (3×3), with a convolution stride of 1. My VGG13 is composed of a first Batch Normalization layer used for input standardization, 13 stacked convolutional layers and pooling layers. The activation function after each convolutional block is the ReLU. The activation function of the output is either a *sigmoid*, used for classification (background rejection), or a *linear* activation function used for the regression tasks. In case of direction reconstruction, we need a double linear output.

Layer type	**Layer**
Batch Normalization	-
Conv2D 1	3×3, 64, stride 1
Conv2D 2	3×3, 64, stride 1
Max Pooling	2×2, stride 2
Conv2D 3	3×3, 128, stride 1
Conv2D 4	3×3, 128, stride 1
Max Pooling	2×2, stride 2
Conv2D 5	3×3, 256, stride 1
Conv2D 6	3×3, 256, stride 1
Conv2D 7	3×3, 256, stride 1
Max Pooling	2×2, stride 2
Conv2D 8	3×3, 512, stride 1
Conv2D 9	3×3, 512, stride 1
Conv2D 10	3×3, 512, stride 1
Max Pooling	2×2, stride 2
Conv2D 11	3×3, 512, stride 1
Conv2D 12	3×3, 512, stride 1
Conv2D 13	3×3, 512, stride 1
Max Pooling	2×2, stride 2
Flatten layer	-
Fully Connected	1-d (2-d) sigmoid (linear)

Table A.1: Architecture of the VGG13.

Appendix B

Beaming effect

In the case of blazars, where the jets cross the line of sight of the observer with a small angle θ, the radiation is collimated in a cone of angle $\alpha = 1/\Gamma_{\text{bulk}}$ in the observer's frame of reference, where Γ_{bulk} is the bulk Lorentz factor of the emitting particles. It happens that the the energy of the photons is boosted, the time variability amplified and the observed intensity is enhanced: this is called *beaming effect* and it is due to the ultra-relativistic nature of the jet. First of all,

In fact, to calculate the time interval between two events happening in the blob, one has to take in consideration the time dilation predicted by SR, but also that the photons - that deliver the information to the observer - travel with a finite speed c: the second event occurs when the blob has travelled a certain distance from the first event, so the respective photons have to cover different space intervals before reaching the Earth. The result is that the observed time interval is:

$$\Delta t_{\text{obs}} = \Gamma_{\text{bulk}}(1 - \beta\cos\theta)\Delta t_{\text{em}} \equiv \frac{\Delta t_{\text{em}}}{\delta} \tag{B.1}$$

where Γ_{bulk} is the Lorentz factor of the blob, and $\delta = 1/\Gamma_{\text{bulk}}(1 - \beta\cos\theta)$ is defined as the *Doppler factor*. This affects also energy, being the frequency the inverse of time:

$$\nu_{\text{obs}} = \delta\nu_{\text{em}} \tag{B.2}$$

For a blazar δ can be of the order of tens, and this in part explains the high energetic emissions we observe from such objects.

The beaming effect causes an even more pronounced enhancement on the luminosity: it can be shown that $L_{\text{obs}} = \delta^p L_{\text{em}}$, with $p > 1$.

Appendix C

Mathematical definitions

C.1 Spectral Energy Distribution (SED)

The Spectral Energy Distribution (SED) is the representation of the energy emitted by a source as a function of the energy, useful to compare at a glance the energy contribution of each wavelength (or frequency).

It is obtained multiplying the differential photon flux by E^2:

$$\frac{dN}{dE}(E) \longrightarrow E^2 \frac{dN}{dE}(E) \tag{C.1}$$

The so defined quantity has the units of an energy density: it is a spectrum in which each photon is weighted with its squared energy.

Alternative equivalent expressions of this quantity are given by:

$$\nu F_\nu \qquad \text{or} \qquad \lambda F_\lambda \tag{C.2}$$

where F_ν (F_λ) is the *differential energy flux*, namely:

$$F_\nu = \frac{dE}{dt\, dA\, d\nu} \qquad \text{or} \qquad F_\lambda = \frac{dE}{dt\, dA\, d\lambda} \tag{C.3}$$

It can be easily shown that $\nu F_\nu = \lambda F_\lambda$.

C.2 Weighted Pearson correlation coefficient

Given two quantities x and y with errors σ_x and σ_y, we calculate the *weighted mean* and the *weighted covariance* as follows:

$$mean(x, \sigma_x) = \frac{\sum_i 1/\sigma_{xi}^2 x_i}{\sum_i 1/\sigma_{xi}^2} \tag{C.4}$$

and similar expression for y, and:

$$cov(x, y, w) = \frac{\sum_i w_i(x_i - mean(x, \sigma_x))(y_i - mean(y, \sigma_y))}{\sum_i w_i} \tag{C.5}$$

being $w_i = 1/(\sigma_{xi}\sigma_{y_i})$.
The *weighted Pearson correlation coefficient* is then defined as:

$$R_W(x, y, w) = \frac{cov(x, y, w)}{\sqrt{cov(x, x, w)cov(y, y, w)}} \tag{C.6}$$

Bibliography

Aartsen, M. et al. (2018). Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science*, *361*(6398), 147–151. https://doi.org/10.1126/science.aat2890

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems.

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., Affeldt, C., Agathos, M., Agatsuma, K., Aggarwal, N., Aguiar, O. D., Aiello, L., Ain, A., Ajith, P., … Zweizig, J. (2016). Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, *116*, 061102. https://doi.org/10.1103/PhysRevLett.116.061102

Acciari, V. A., Ansoldi, S., Antonelli, L. A., Arbet Engels, A., Baack, D., Babić, A., Banerjee, B., Barres de Almeida, U., Barrio, J. A., & et al. (2020a). Broadband characterisation of the very intense tev flares of the blazar 1es 1959+650 in 2016. *Astronomy & Astrophysics*, *638*, A14. https://doi.org/10.1051/0004-6361/201935450

Acciari, V. A., Ansoldi, S., Antonelli, L. A., Arbet Engels, A., Baack, D., Babić, A., Banerjee, B., Barres de Almeida, U., Barrio, J. A., & et al. (2020b). Testing two-component models on very high-energy gamma-ray-emitting bl lac objects. *Astronomy Astrophysics*, *640*, A132. https://doi.org/10.1051/0004-6361/202037811

Acciari, V. et al. (2019). Teraelectronvolt emission from the γ-ray burst GRB 190114C. *Nature*, *575*(7783), 455–458. https://doi.org/10.1038/s41586-019-1750-x

Acharya, B. et al. (2013). Introducing the CTA concept. *Astropart. Phys.*, *43*, 3–18. https://doi.org/10.1016/j.astropartphys.2013.01.007

Afshine, A., & Shervine, A. (2018). A detailed example of how to use data generators with keras. https://stanford.edu/~shervine/blog/keras-how-to-generate-data-on-the-fly

Aharonian, F. (2000). Tev gamma rays from bl lac objects due to synchrotron radiation of extremely high energy protons. *New Astronomy*, *5*(7), 377–395. https://doi.org/10.1016/s1384-1076(00)00039-7

Akiyama, K., Bouman, K., & Woody, D. (2019). First m87 event horizon telescope results. i. the shadow of the supermassive black hole. *Astrophysical Journal Letters*, *875*.

Albert, J., Aliu, E., Anderhub, H., Antoranz, P., Armada, A., Asensio, M., Baixeras, C., Barrio, J., Bartko, H., Bastieri, D., & et al. (2008). Implementation of the random forest method for the imaging atmospheric cherenkov telescope magic. *Nuclear Instruments and Methods in Physics*

Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, *588*(3), 424–432. https://doi.org/10.1016/j.nima.2007.11.068

Albert, J. et al. (2006). Observation of very high energy gamma-ray emission from the active galactic nucleus 1es 1959+650 using the magic telescope. *THE ASTROPHYSICAL JOURNAL*, *639*, 761–765. https://doi.org/10.1086/499421

Aleksić, J., Ansoldi, S., Antonelli, L. A., Antoranz, P., Babic, A., Bangale, P., Barceló, M., Barrio, J. A., Becerra González, J., Bednarek, W., Bernardini, E., Biasuzzi, B., Biland, A., Bitossi, M., Blanch, O., Bonnefoy, S., Bonnoli, G., Borracci, F., Bretz, T., … Zanin, R. (2016). The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula. *Astroparticle Physics*, *72*, 76–94. https://doi.org/10.1016/j.astropartphys.2015.02.005

Aleksić, J., Ansoldi, S., Antonelli, L., Antoranz, P., Babic, A., Bangale, P., Barrio, J., Becerra González, J., Bednarek, W., Bernardini, E., & et al. (2015). Measurement of the crab nebula spectrum over three decades in energy with the magic telescopes. *Journal of High Energy Astrophysics*, *5-6*, 30–38. https://doi.org/10.1016/j.jheap.2015.01.002

Aliu, E., Archambault, S., Arlen, T., Aune, T., Barnacka, A., Beilicke, M., Benbow, W., Berger, K., Bird, R., Bouvier, A., Buckley, J., Bugaev, V., Cerruti, M., Chen, X., Ciupik, L., Collins-Hughes, E., Connolly, M., Cui, W., Dumm, J., & Fumagalli, M. (2014). Investigating broadband variability of the tev blazar 1es 1959+650. *The Astrophysical Journal*, *797*. https://doi.org/10.1088/0004-637X/797/2/89

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Esesn, B. C. V., Awwal, A. A. S., & Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches.

Arrabito, L., Bernlöhr, K., Bregeon, J., Carrère, M., Khattabi, A., Langlois, P., Parello, D., & Revy, G. (2020). Optimizing cherenkov photons generation and propagation in corsika for cta monte-carlo simulations.

Atwood, W. B., Abdo, A. A., Ackermann, M., Althouse, W., Anderson, B., Axelsson, M., Baldini, L., Ballet, J., Band, D. L., Barbiellini, G., Bartelt, J., Bastieri, D., Baughman, B. M., Bechtol, K., Bédérède, D., Bellardi, F., Bellazzini, R., Berenji, B., Bignami, G. F., … Ziegler, M. (2009). The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission., *697*(2), 1071–1102. https://doi.org/10.1088/0004-637X/697/2/1071

Belayneh, D., Carminati, F., Farbin, A., Hooberman, B., Khattak, G., Liu, M., Liu, J., Olivito, D., Pacela, V. B., Pierini, M., & et al. (2020). Calorimetry with deep learning: Particle simulation and reconstruction for collider physics. *The European Physical Journal C*, *80*(7). https://doi.org/10.1140/epjc/s10052-020-8251-9

Bertero, M. (1989). Linear inverse and iii-posed problems. In P. W. Hawkes (Ed.). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2539(08)60946-4

Böttcher, M. (2005). A Hadronic Synchrotron Mirror Model for the "Orphan" TeV Flare in 1ES 1959+650., *621*(1), 176–180. https://doi.org/10.1086/427430

Bourilkov, D. (2019). Machine and deep learning applications in particle physics. *International Journal of Modern Physics A*, *34*(35), 1930019. https://doi.org/10.1142/s0217751x19300199

Branchesi, M. (2016). Multi-messenger astronomy: Gravitational waves, neutrinos, photons, and cosmic rays. *Journal of Physics: Conference Series*, *718*, 022004. https://doi.org/10.1088/1742-6596/718/2/022004

Breit, G., & Wheeler, J. A. (1934). Collision of two light quanta. *Phys. Rev.*, *46*, 1087–1091. https://doi.org/10.1103/PhysRev.46.1087

Burrows, D. N. et al. (2005). The Swift X-ray Telescope. *Space Sci. Rev.*, *120*, 165. https://doi.org/10.1007/s11214-005-5097-2

Caraveo, P. (2020). The golden age of high-energy gamma-ray astronomy: The cherenkov telescope array in the multimessenger era. *La Rivista del Nuovo Cimento*, *43*. https://doi.org/10.1007/s40766-020-00006-3

Carroll, C. (2015). A feasibility study of photometric reverberation mapping using meter-class telescopes.

Chollet, F. (2017). *Deep learning with python*. Manning.

De Angelis, A., & Mallamaci, M. (2018). Gamma-ray astrophysics. *The European Physical Journal Plus*, *133*(8). https://doi.org/10.1140/epjp/i2018-12181-0

Domínguez, A., Primack, J. R., Rosario, D. J., Prada, F., Gilmore, R. C., Faber, S. M., Koo, D. C., Somerville, R. S., Pérez-Torres, M. A., Pérez-González, P., Huang, J.-S., Davis, M., Guhathakurta, P., Barmby, P., Conselice, C. J., Lozano, M., Newman, J. A., & Cooper, M. C. (2011). Extragalactic background light inferred from AEGIS galaxy-SED-type fractions. *Monthly Notices of the Royal Astronomical Society*, *410*(4), 2556–2578. https://doi.org/10.1111/j.1365-2966.2010.17631.x

Dwek, E., & Krennrich, F. (2013). The extragalactic background light and the gamma-ray opacity of the universe [Seeing the High-Energy Universe with the Cherenkov Telescope Array - The Science Explored with the CTA]. *Astroparticle Physics*, *43*, 112–133. https://doi.org/https://doi.org/10.1016/j.astropartphys.2012.09.003

Falomo, R., Kotilainen, J. K., Carangelo, N., & Treves, A. (2003). Black hole masses and the fundamental plane of BL lacertae objects. *The Astrophysical Journal*, *595*(2), 624–630. https://doi.org/10.1086/377432

Foffano, L. (2019). *The extreme blazar phenomenon in a multi-messenger context* (Doctoral dissertation). Università di Padova.

Fomin, V., Stepanian, A., Lamb, R., Lewis, D., Punch, M., & Weekes, T. (1994). New methods of atmospheric Cherenkov imaging for gamma-ray astronomy. I. The false source method. *Astroparticle Physics*, *2*(2), 137–150. https://doi.org/10.1016/0927-6505(94)90036-1

Fossati, G., Maraschi, L., Celotti, A., Comastri, A., & Ghisellini, G. (1998). A unifying view of the spectral energy distributions of blazars. *Monthly Notices of the Royal Astronomical Society*, *299*(2), 433–448. https://doi.org/10.1046/j.1365-8711.1998.01828.x

Fruck, Christian, & Gaug, Markus. (2015). Atmospheric monitoring in magic and data corrections. *EPJ Web of Conferences*, *89*, 02003. https://doi.org/10.1051/epjconf/20158902003

Ghisellini, G., Righi, C., Costamante, L., & Tavecchio, F. (2017). The fermi blazar sequence. *Monthly Notices of the Royal Astronomical Society*, *469*(1), 255–266. https://doi.org/10.1093/mnras/stx806

Ghisellini, G. (2013). Synchrotron self-compton. *Radiative processes in high energy astrophysics* (pp. 89–93). Springer International Publishing. https://doi.org/10.1007/978-3-319-00612-3_6

Ghisellini, G. (2016). The blazar sequence 2.0.

Gregory, P. C., & Condon, J. J. (1991). The 87GB Catalog of Radio Sources Covering 0 degrees < delta < +75 degrees at 4.85 GHz., *75*, 1011. https://doi.org/10.1086/191559

Guest, D., Cranmer, K., & Whiteson, D. (2018). Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, *68*(1), 161–181. https://doi.org/10.1146/annurev-nucl-101917-021019

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition.

Heck, D., Knapp, J., Capdevielle, J., Schatz, G., & Thouw, T. (1998). *CORSIKA: a Monte Carlo code to simulate extensive air showers.*

Hess, V. (2018). On the observations of the penetrating radiation during seven balloon flights.

Hillas, A. (1985). Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei. *19th International Cosmic Ray Conference (ICRC19), Volume 3*, *3*, 445.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-excitation networks.

Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J. E., & Weinberger, K. Q. (2017). Snapshot ensembles: Train 1, get m for free.

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2018). Densely connected convolutional networks.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift.

Jacquemont., M., Antiga., L., Vuillaume., T., Silvestri., G., Benoit., A., Lambert., P., & Maurin., G. Indexed operations for non-rectangular lattices applied to convolutional neural networks. In: *Proceedings of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications - volume 5: Visapp,* INSTICC. SciTePress, 2019, 362–371. ISBN: 978-989-758-354-4. https://doi.org/10.5220/0007364303620371.

Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green, R. (1989). VLA Observations of Objects in the Palomar Bright Quasar Survey., *98*, 1195. https://doi.org/10.1086/115207

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization.

Konigl, A. (1981). Relativistic jets as X-ray and gamma-ray sources., *243*, 700–709. https://doi.org/10.1086/158638

Krawczynski, H., Hughes, S. B., Horan, D., Aharonian, F., Aller, M. F., Aller, H., Boltwood, P., Buckley, J., Coppi, P., Fossati, G., & et al. (2004). Multiwavelength observations of strong flares from the tev blazar 1es 1959+650. *The Astrophysical Journal*, *601*(1), 151–164. https://doi.org/10.1086/380393

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in*

neural information processing systems 25 (pp. 1097–1105). Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Li, T., & Ma, Y. (1983). Analysis methods for results in gamma-ray astronomy., *272*, 317–324. https://doi.org/10.1086/161295

Lopez-Coto, R. (2015). *Very-high-energy gamma-ray observations of pulsar wind nebulae and cataclysmic variable stars with MAGIC and development of trigger systems for IACTs* (Doctoral dissertation). Universitat Autonoma de Barcelona.

Mannheim, K. (1993). The proton blazar., *269*, 67–76.

Marinello, N. (2019). *Convolutional neural network single-telescope reconstruction for the large size telescope of cta* (Master's thesis). Università di Padova.

Mariotti, E. (2019). *Deep learning on magic: A performance evaluation for very high energy gamma-ray astrophysics* (Master's thesis). Università di Padova.

Massaro, E., Tramacere, A., Perri, M., Giommi, P., & Tosti, G. (2006). Log-parabolic spectra and particle acceleration in blazars - iii. ssc emission in the tev band from mkn 501. *A&A*, *448*(3), 861–871. https://doi.org/10.1051/0004-6361:20053644

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*, 289–337. http://www.jstor.org/stable/91247

Nieto, D., Brill, A., Kim, B., & Humensky, T. B. (2017). Exploring deep learning as an event classification method for the cherenkov telescope array.

Nishiyama, T. (1999). Detection of a new TeV gamma-ray source of BL Lac object 1ES 1959+650. *26th International Cosmic Ray Conference (ICRC26), Volume 3*, *3*, 370.

Padovani, P. et al. (2017). Active galactic nuclei: what's in a name? *Astron. Astrophys. Rev.*, *25*(1), 2. https://doi.org/10.1007/s00159-017-0102-9

Padovani, P. (2016). The faint radio sky: radio astronomy becomes mainstream. *Astron. Astrophys. Rev.*, *24*(1), 13. https://doi.org/10.1007/s00159-016-0098-6

Padovani, P., & Giommi, P. (1995). The Connection between X-Ray– and Radio-selected BL Lacertae Objects., *444*, 567. https://doi.org/10.1086/175631

Parsons, R. D., & Ohm, S. (2020). Background rejection in atmospheric cherenkov telescopes using recurrent convolutional neural networks. *The European Physical Journal C*, *80*(5). https://doi.org/10.1140/epjc/s10052-020-7953-3

Patel, S. R., Shukla, A., Chitnis, V. R., Dorner, D., Mannheim, K., Acharya, B. S., & Nagare, B. J. (2018). Broadband study of blazar 1es 1959+650 during flaring state in 2016. *Astronomy & Astrophysics*, *611*, A44. https://doi.org/10.1051/0004-6361/201731987

Patel, S., Shukla, A., Chitnis, V., Dorner, D., Mannheim, K., Acharya, B., & Nagare, B. (2017). Broadband study of blazar 1es 1959+650 during flaring state in 2016. *Astronomy Astrophysics*, *611*. https://doi.org/10.1051/0004-6361/201731987

Perlman, E., Stocke, J., Schachter, J., Elvis, M., Ellingson, E., Urry, C., Potter, M., Impey, C., & Kolchinsky, P. (1996). The einstein slew survey sample of bl lacertae objects. *Astrophysical Journal, Supplement Series*, *104*(2), 251–285. https://doi.org/10.1086/192300

Poole, T. S., Breeveld, A. A., Page, M. J., Land sman, W., Holland, S. T., Roming, P., Kuin, N. P. M., Brown, P. J., Gronwall, C., Hunsberger, S., Koch, S., Mason, K. O., Schady, P., vanden Berk, D., Blustin, A. J., Boyd, P., Broos, P., Carter, M., Chester, M. M., … Still, M. (2008). Photometric calibration of the Swift ultraviolet/optical telescope., *383*(2), 627–645. https://doi.org/10.1111/j.1365-2966.2007.12563.x

Prandini, E. (2017). Highlights from tev extragalactic sources.

Richards, J. L., Max-Moerbeck, W., Pavlidou, V., King, O. G., Pearson, T. J., Readhead, A. C. S., Reeves, R., Shepherd, M. C., Stevenson, M. A., Weintraub, L. C., Fuhrmann, L., Angelakis, E., Zensus, J. A., Healey, S. E., Romani, R. W., Shaw, M. S., Grainge, K., Birkinshaw, M., Lancaster, K., … Bustos, R. (2011). BLAZARS IN THE FERMI ERA: THE OVRO 40 m TELESCOPE MONITORING PROGRAM. *The Astrophysical Journal Supplement Series*, *194*(2), 29. https://doi.org/10.1088/0067-0049/194/2/29

Schmelling, M. (1994). The method of reduced cross-entropy a general approach to unfold probability distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *340*(2), 400–412. https://doi.org/https://doi.org/10.1016/0168-9002(94)90119-8

Shilon, I., Kraus, M., Büchele, M., Egberts, K., Fischer, T., Holch, T., Lohse, T., Schwanke, U., Steppa, C., & Funk, S. (2019). Application of deep learning methods to analysis of imaging atmospheric cherenkov telescopes data. *Astroparticle Physics*, *105*, 44–53. https://doi.org/10.1016/j.astropartphys.2018.10.003

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Smith, L. N. (2017). Cyclical learning rates for training neural networks.

Smith, L. N., & Topin, N. (2018). Super-convergence: Very fast training of neural networks using large learning rates.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Tagliaferri, G., Foschini, L., Ghisellini, G., Maraschi, L., and, G., Albert, J., Aliu, E., Anderhub, H., Antoranz, P., Baixeras, C., Barrio, J. A., Bartko, H., Bastieri, D., Becker, J., Bednarek, W., Bedyugin, A., Berger, K., Bigongiari, C., Biland, A., & Zapatero, J. (2008). Simultaneous multiwavelength observations of the blazar 1es 1959+650 at a low tev flux. *The Astrophysical Journal*, *679*, 1029. https://doi.org/10.1086/586731

Tavecchio, F., Maraschi, L., & Ghisellini, G. (1998). Constraints on the Physical Parameters of TeV Blazars., *509*(2), 608–619. https://doi.org/10.1086/306526

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. W.H. Winston.

Tonello, N., & the Magic collaboration. (2006). Observation of the AGN 1es1959+650 with the MAGIC telescope. *Journal of Physics: Conference Series*, *39*, 457–459. https://doi.org/10.1088/1742-6596/39/1/120

Tramacere, A., Giommi, P., Perri, M., Verrecchia, F., & Tosti, G. (2009). Swift observations of the very intense flaring activity of Mrk 421 during 2006. I. Phenomenological picture of electron acceleration and predictions for MeV/GeV emission., *501*(3), 879–898. https://doi.org/10.1051/0004-6361/200810865

Tramacere, A., Massaro, E., & Taylor, A. M. (2011). Stochastic Acceleration and the Evolution of Spectral Distributions in Synchro-Self-Compton Sources: A Self-consistent Modeling of Blazars' Flares., *739*(2), Article 66, 66. https://doi.org/10.1088/0004-637X/739/2/66

Urry, C. M., & Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei., *107*, 803. https://doi.org/10.1086/133630

Wagner, R., Backes, M., Satalecka, K., Bonnoli, G., Doert, M., Steinke, B., Strah, N., Terzić, T., Tescaro, D., & Uellenbeck, M. (2011). Monitoring of bright, nearby active galactic nuclei with the magic telescopes. *Proceedings of the 32nd International Cosmic Ray Conference, ICRC 2011*, *8*. https://doi.org/10.7529/ICRC2011/V08/1030

Watson, A. A. (2011). The discovery of cherenkov radiation and its use in the detection of extensive air showers. *Nuclear Physics B - Proceedings Supplements*, *212-213*, 13–19. https://doi.org/10.1016/j.nuclphysbps.2011.03.003

Zanin, R. (2013). MARS, the MAGIC analysis and reconstruction software. *33rd International Cosmic Ray Conference*, 0773.