UNIVERSITY OF MODENA AND REGGIO EMILIA

International Doctorate in Information And Communication Technologies

Cycle XXXIII

Analysis methods and data management for real-time gamma-ray astrophysics.

Candidate: Nicoló PARMIGGIANI

Advisor: Prof. Domenico BENEVENTANO

Co-Advisor: PhD Andrea BULGARELLI

Coordinator of the Doctorate Program: Prof. Sonia BERGAMASCHI

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dottorato di ricerca in Information And Communication Technologies

---

Ciclo XXXIII

Metodi per l'analisi e la gestione dei dati dell'astrofisica gamma in tempo reale.

Candidato: Nicoló PARMIGGIANI

Tutor: Prof. Domenico BENEVENTANO

Co-Tutor: PhD Andrea BULGARELLI

Coordinatore del Corso di Dottorato: Prof. Sonia BERGAMASCHI

Keywords:

Gamma-ray astronomy
Gamma-ray bursts
Real-time analysis
Deep Learning
Convolutional Neural Network

The context of this PhD is the data analysis and management for gamma-ray astronomy, which involves the observation of gamma-rays, the most energetic form of electromagnetic radiation. From the gamma-ray observations performed by on ground or spacecraft based observatories, it is possible to study catastrophic events involving compact objects, such as white dwarves, neutron stars, and black holes. These events are called gamma-ray transients. The today astrophysics is multi-wavelength, i.e. astrophysical observations are performed using the electromagnetic radiation detected by different observatories at different wavelengths (gamma-ray, x-ray, ultraviolet, infrared, radio etc.). Since 2017 with the first gravitational wave event GW170817 and neutrino event IceCube-170922A having a gamma-ray counterpart, the astrophysical community is experiencing a new era called "multi-messenger astronomy". In this new era, the astronomical sources are observed by observatories that receive different signals: gravitational waves, electromagnetic radiation, and neutrinos.

In the multi-messenger era, astrophysical observatories share science alerts through different communication networks. A science alert is an immediate communication from one observatory to other observatories that an interesting astrophysical event is occurring in the sky. The coordination of different observatories done by sharing science alerts is mandatory to understand the physical phenomena observed. For this reason, the speed is crucial, and automated data analysis pipelines are developed to detect gamma-ray transients and generate science alerts during the astrophysical observations or immediately after.

This automated analysis can follow three main patterns: (i) executing a blind search to detect transient event using only the observatory data without any suggestion about the source position, (ii) analysing known sources using only the observatory data to search transient, and (iii) react to detection performed by other observatories to search for a counterpart into the incoming or archive data (a.k.a external science alerts). The first two patterns aim to detect transient events and share them with the scientific community, while the third pattern aims to manage external science alerts. In this thesis, both scenarios (the generation of internal science alerts, and the management of external science alerts) are described and implemented developing the analysis software.

Observatories have to manage the follow-up of these external science alerts by developing dedicated software. During this PhD, the research activity focused on the AGILE space mission, operational since 2007, and on the Cherenkov Telescope Array Observatory (CTA), currently in the construction phase. The follow-up of external science alerts received from Gamma-Ray Bursts (GRB) and Gravitational Waves (GW) detectors is one of the

AGILE Team's current major activity. The software developed during this PhD uses the Gamma-ray Coordinates Network (GCN) to receive GRB and GW external science alerts from other observatories. Future generations of gamma-ray observatories like the CTA or the ASTRI Mini-Array can take advantage of the technologies developed for AGILE. This research aims to develop analysis and management software for gamma-ray data to achieve the observatories scientific goal.

The first chapter introduces the basic concept used in this thesis, and the observatories where the work for this PhD was completed.

The second chapter of this thesis describes the software and the web platform developed during this PhD used by AGILE researchers to prepare the Second AGILE Catalog of Gamma-ray sources. The results of the catalogue analyses are stored in a database, and the web platform visualises these results querying the database. The software developed for the catalogue analyses follows the pattern of known sources analysis described before. The work completed in this context is an improvement of the software developed during my Master Thesis; in particular, the database is based on the database developed during my Master Thesis to manage detections of gamma-ray sources and light curve generated by the AGILE real-time scientific analysis pipeline.

The third chapter presents the RTApipe framework designed to facilitate the development of real-time scientific analysis pipelines that follow the three analysis patterns described before. The framework provides a common pipeline architecture and automatisms that can be used by observatories to develop their pipelines. This framework was used to develop the pipelines for the AGILE space mission, develop a prototype of the scientific pipeline of the Science Alert Generation system of the CTA Observatory, and develop a prototype of the pipeline for the scientific analysis the ASTRI Horn telescope. In addition, the third chapter shows the scientific results obtained with the software developed for AGILE during this PhD research activities. The AGILE real-time analysis pipelines are mandatory to perform a fast follow-up of science alerts (internal and external) and send scientific alerts through the GCN network or Astronomer's Telegram when the pipelines detect a transient event. The most important scientific results obtained with the support of the automated pipelines were published in referred journals.

The fourth chapter describes a new method to detect GRBs in the AGILE/GRID data using the Convolutional Neural Network. With this Deep Learning technology, it is possible to improve the AGILE/GRID instrument's detection capabilities. This method was also integrated as a science tool in the AGILE pipelines.

The last chapter presents the conclusions and the future works that will be carried on after this PhD.

Sommario

Il contesto delle attività svolte per il PhD sono l'analisi e la gestione dei dati per l'astronomia dei raggi gamma, la quale coinvolge l'osservazione dei raggi gamma, la forma più energetica di radiazione elettromagnetica. Dalle osservazioni dei raggi gamma effettuate con osservatori terrestri o a bordo di satelliti, è possibile studiare eventi catastrofici generati da oggetti compatti come nane bianche, stelle di neutroni e buchi neri. Questi eventi sono chiamati transienti di raggi gamma. L'astrofisica di oggi è multi-wavelength, le osservazioni utilizzano dati provenienti da diversi osservatori che raccolgono informazioni a differenti lunghezze d'onda (raggi gamma, raggi x, ultravioletti, infrarossi, onde radio, etc.). Dal 2017 con il primo evento gravitazionale GW170817 e con il primo neutrino IceCube-170922A per i quali sono state trovate controparti elettromagnetiche, la comunità astrofisica è entrata nella nuova era multi-messenger, dove le sorgenti astronomiche sono osservate raccogliendo differenti segnali: onde gravitazionali, radiazioni elettromagnetiche e neutrini.

In questa multi-messenger era, gli osservatori astrofisici condividono le loro allerte scientifiche tramite reti di comunicazione. L'allerta scientifica è una comunicazione immediata da un osservatorio ad altri osservatori per segnalare che un evento astrofisico interessante sta avvenendo nel cielo. Per comprendere questi fenomeni, essi devono essere osservati durante la loro evoluzione. Per questa ragione, la velocità è cruciale e vengono sviluppate pipeline automatiche di analisi dei dati per identificare questi transienti e generare allerte scientifiche. Queste analisi automatiche possono essere eseguite con tre modalità principali: (i) ricercando in modo casuale all'interno dei dati raccolti dall'osservatorio un evento transiente senza avere indicazioni sulla possibile posizione (ii) analizzando sorgenti conosciute utilizzando i dati dell'osservatorio per ricercare fenomeni transienti e (iii) reagendo alle allerte scientifiche ricevute da altri osservatori per ricercare una controparte transiente nei propri dati in arrivo o di archivio. In questa tesi vengono trattati e utilizzati entrambe gli scenari (generazione di allerte interne e gestione delle

allerte esterne) per sviluppare il software di analisi.

Gli osservatori devono gestire queste allerte scientifiche sviluppando software dedicato. Durante questo PhD l'attività di ricerca è stata focalizzata principalmente sulla missione spaziale AGILE, attualmente in operazione, e sull'osservatorio Cherenkov Telescope Array (CTA), in fase di costruzione. Il follow-up di allerte scientifiche esterne ricevute dagli strumenti che identificano Gamma-Ray Bursts (GRB) e Gravitational Waves (GW) è una delle maggiori attività del Team di AGILE. I software sviluppati durante questo PhD utilizzano il Gamma-ray Coordinates Network (GCN) per ricevere allerte scientifiche da altri osservatori. Le generazioni future degli osservatori di raggi gamma come CTA o ASTRI Mini-Array possono trarre vantaggio dalle tecnologie sviluppate per AGILE. Questa ricerca ha l'obiettivo di sviluppare software per l'analisi e la gestione dei dati gamma.

Nel primo capitolo vengono introdotti i concetti fondamentali utilizzati in questa tesi e descritti gli osservatori per i quali è stato portato avanti il lavoro del PhD.

Il secondo capitolo della tesi descrive il software e la piattaforma web sviluppati durante questo PhD e utilizzati dai ricercatori AGILE per preparare il secondo catalogo di sorgenti gamma di AGILE. I risultati delle analisi eseguite per questo catalogo sono memorizzati in un database dedicato e la piattaforma web consente la visualizzazione di questi risultati interrogando il database. Il software di analisi per questo catalogo segue la modalità di analisi di sorgenti conosciute descritta in precedenza. Il lavoro svolto in questo ambito è una evoluzione di quanto sviluppato durante la mia Tesi di Laurea Magistrale; in particolare il database dedicato si basa su quello sviluppato durante la mia Tesi di Laurea Magistrale per gestire le detection di sorgenti gamma e le curve di luce generate dalla pipeline di analisi real-time della missione spaziale AGILE.

Nel terzo capitolo viene presentato il framework RTApipe progettato per facilitare lo sviluppo di pipeline per l'analisi scientifica in tempo reale e per la gestione di allerte scientifiche esterne. Il framework offre un'architettura comune e automatismi che possono essere utilizzati dagli osservatori per sviluppare le loro pipeline. Questo framework è stato usato per sviluppare le pipeline della missione spaziale AGILE, per sviluppare un prototipo per la Science Alert Generation (SAG) dell'osservatorio CTA e per sviluppare il prototipo della pipeline di analisi scientifica del telescopio ASTRI Horn. Inoltre, il terzo capitolo riporta anche i risultati scientifici ottenuti con le pipeline di analisi in tempo reale sviluppate per AGILE durante le attività di ricerca di questo PhD. Queste pipeline sono necessarie per poter eseguire un rapido follow-up degli eventi transienti (interni ed esterni) e per inviare allerte scientifiche tramite il GCN network o l'Astronomer's Telegram quando il software identifica dei transenti gamma. I risultati scientifici più importanti ottenuti utilizzando questi software come supporto sono stati pubblicati su riviste scientifiche referate.

Il quarto capitolo descrive un nuovo metodo per identificare i GRB nei dati dello strumento AGILE/GRID utilizzando le Convolutional Neural Network. Con questa tecnologia Deep Learning è possibile migliorare la capacità di detection di AGILE. Questo metodo è anche stato inserito come tool scientifico all'interno della pipeline AGILE.

L'ultimo capitolo riporta le conclusioni e gli sviluppi futuri che verranno portati avanti dopo il periodo del PhD.

Acknowledgments

A special thanks to Dr. Andrea Bulgarelli and Prof. Domenico Beneventano for the support received during these years of the PhD and for the inestimable advices.

Thanks to Prof. Marco Tavani, Principal Investigator of the AGILE space mission, for allowing me to work for this project and thanks to all the AGILE team with whom I have collaborated during the PhD.

Furthermore, I thank the researchers of INAF / OAS Bologna with whom I have had the pleasure of collaborating over these three years and with whom I hope to continue my career.

I want to thank my parents and my girlfriend Alice for supporting and encouraging me during these years in the path of professional but also personal growth.

The AGILE space mission is funded by the Italian Space Agency (ASI) with scientific and programmatic participation by the Italian Institute of Astrophysics (INAF) and the Italian Institute of Nuclear Physics (INFN). Investigation supported by the ASI grant I/089/06/2, I/028/12/5 and I/028/12/6. Thank to the ASI management for unfailing support during AGILE operations. I acknowledge the effort of ASI and industry personnel in operating the ASI ground station in Malindi (Kenya), and the data processing done at the ASI/SSDC in Rome: the success of AGILE scientific operations depends on the effectiveness of the data flow from Kenya to SSDC and INAF/OAS Bologna and the data analysis and software management.

Ringraziamenti

Un particolare ringraziamento al Dr. Andrea Bulgarelli e al Prof. Domenico Beneventano per il sostegno ricevuto durante questi anni di dottorato e per i preziosi consigli.

Ringrazio il Prof. Marco Tavani, Principal Investigator della missione spaziale AGILE, per avermi dato la possibilità di lavorare per questo progetto e tutto il team di AGILE con il quale ho collaborato durante questi anni.

Inoltre ringrazio i ricercatori dell'INAF/OAS Bologna con i quali ho avuto il piacere di collaborare in questi anni e con i quali spero di poter continuare il percorso lavorativo.

Voglio ringraziare i miei genitori e la mia fidanzata Alice per avermi accompagnato durante questi anni nel percorso di crescita personale e professionale.

Contents

List of Figures

List of Tables

Table 1: List of abbreviations.

Acronym	Description
ACADA	Array Control and Data Acquisition
ACS	AntiCoincidence System
AGILE	Light Imager for Gamma-Ray Astrophysics
AGNs	Active Galactic Nuclei
AMON	Astrophysical Multi-messenger Observatory Network
ASI	Italian Space Agency
ASP	Automated Sience Processing
ATel	Astronomer's Telegram
AUC	Area Under the Curve
BAT	Burst Alert Telescope
CMB	Cosmic Microwave Background
CNN	Convolutional Neural Network
CR	Cosmic Ray
CTA	Cherenkov Telescope Array Observatory
CV	CNN output value
DAQ	Data AQuisition
DBMS	DataBase Management System
DL	Deep Learning
DL3	Photon event list
DM	Data Model
DNN	Deep Neural Network
E-R	Entity Relationship
ETL	Extract Transform Load
EVT	AGILE event file
FOV	Field Of View
FPR	False Positive Rate
FTP	File Transfer Protocol
GBM	Gamma-Ray Burst Monitor
GCN	Gamma-ray Coordinates Network
GeV	Giga elettronvolt
GPU	Graphical Processing Unit
GRB	Gamma-Ray Burst
GRID	Gamma-Ray Imaging Detector
GTI	Good Time Intervals
GUI	Graphical User Interface
GW	Gravitational Waves
IAC	Instituto de Astrofisica de Canarias
IACT	Imaging Atmospheric Cherenkov Telescope

Acronym	Description
INAF	Italian National Institute for Astrophysics
IPN	Third Interplanetary Network
IRFs	Instrument Response Functions
ISOC	Fermi Instrument Science Operation Center
ISRF	Interstellar Radiation Field
LAT	Large Area Telescope
LIGO	Laser Interferometer Gravitational-Wave Observatory
LOG	AGILE auxiliary file
LST	Large-Sized Telescope
LVC	LIGO / VIRGO Collaboration
MA	Mini Array
MAGIC	Major Atmospheric Gamma-ray Imaging Cherenkov Telescopes
MCAL	Mini-Calorimeter
MeV	Mega elettronvolt
MJD	Modified Julian Date
ML	Machine Learning
MLE	Maximum Likelihood Estimator
MM	Multi-Messenger
MOC	Mission Operation Center
MOLA	MAGIC Online Analysis
MST	Medium-Sized Telescope
MW	Multi-Wavelength
PSF	Point Spread Function
RM	Ratemeters
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SA	Super AGILE
SAG	Science Alert Generator
SOC	Science Operation Center
SSC	Satellite Control Center
SSDC	ASI Space Science Data Center
SST	Small-Sized Telescope
ST	Silicon Tracker
STE	Sub-Threshold Event
TDRSS	Tracking and Data Relay Satellite System
TGF	Terrestrial Gamma-ray Flash
TPR	True Positive Rate
UV	Ultra Violet
VHE	Very High Energy

Chapter 1

Introduction

The context of this PhD is the data management and analysis of astronomical data for gamma-ray astronomy. This astronomy branch studies the universe's physics and the nature of celestial objects using the gamma-ray electromagnetic radiation. Several technologies to design and develop software used to manage and analyse astronomical data were studied during this PhD.

The gamma-rays are the most energetic form of electromagnetic radiation. From the gamma-ray observations, it is possible to study catastrophic events involving compact objects, such as white dwarves, neutron stars, and black holes. These events are called gamma-ray transients.

In astronomy, observatories and experiments are organisations established to builds and operates one or more facilities (e.g. on ground-based telescopes, space telescopes etc.). An observatory allows external entities to propose observations and access to acquired data. On the other hand, an experiment manages one or more facilities such as the observatory but differs from the observatory for the data dissemination policy because it usually reserves the data only for the internal collaboration.

During this PhD, software for the gamma-ray data management and analysis, mainly for gamma-ray transients, was developed. In Sec. 2 it is described the software developed in collaboration with the researchers of the AGILE space mission (Sec. 1.2) to manage the gamma-ray data produced by the instruments onboard the spacecraft. These data are analysed with different methods. The results are stored into a database, and then visualised using a web Graphical User Interface (GUI).

The work for this PhD was carried on in collaboration with researchers of the Italian Institute of Astrophysics[1] (INAF/OAS Bologna). The collaboration started with the AGILE space mission and then continued with ground-based gamma-ray observatories: CTA (Sec. 1.3) and ASTRI Mini Array (Sec. 1.4). In Sec. 2 it is described the software developed in collaboration with the researchers of the AGILE space mission (Sec. 1.1) to manage the gamma-ray data produced by the instruments onboard the spacecraft. These data are analysed with different methods. The results are stored in a database, and then visualised using a web GUI.

The analysis and the management of the astrophysical data are becoming a challenge because the next generation of gamma-ray observatories, like the Cherenkov Telescope Array (Sec. 1.3) and the ASTRI Mini Array (Sec. 1.4), will generate a large amount of data. These data can be considered Big Data for their characteristics (Sec. 1.5). For this reason, the next generation of observatories must exploit new information technologies to satisfy the Big

[1] https://www.oas.inaf.it/en/

Data requirements and to manage this Big Data production.

As described in [ZZ15] the Big Data era is influencing the astrophysical research field and leads to collaborations between astronomers, computer scientists, and data scientists. The data collected by the observatories is increasing, and researchers need new techniques to manage and analyse this amount of data.

The observatories share information about transients phenomena with the scientific community to study the same physical event with different instruments collecting data at different wavelengths (e.g. infrared, ultraviolet, X-ray, gamma-ray and more). In recent years, the astrophysical landscape changed, and now the so-called ”multi-messenger era” (Sec. 1.6) is leading part of the astronomical observatories activities. In this context, the observatories share their information with the astrophysical community to study the same transient event with different ”messenger” signals: electromagnetic radiation, gravitational waves, and neutrinos. The observatories develop their own real-time scientific analysis pipeline to identify possible transient phenomena (e.g. Gamma-ray Bursts) and react to external alerts from other observatories (Sec. 1.7) as soon as possible. The Sec. 3 describes the RTApipe framework designed and developed during this PhD to help researchers realise real-time analysis pipelines that satisfy the scientific goals of the gamma-ray observatories.

One of the key technologies that can improve the analysis capability of astrophysical data is Machine Learning (ML) (Sec. 1.8). The Deep Learning (DL) techniques, a branch of ML, are growing up in recent years due to several concomitant factors like the increased computing power of the Graphical Processing Units (GPU) and the huge amount of data available to train these DL models. During this PhD, a Convolutional Neural Network (CNN) was developed to improve the gamma-ray detection capability of the AGILE space mission. This CNN is described in Sec. (4). The results obtained with this CNN are compared with results obtained applying the standard analysis method used by the AGILE Team showing the improvement in this real-time analysis context.

1.1 Gamma-Ray Astronomy

As described in [Fun15], the gamma-rays are the highest-energy form of electromagnetic radiation. The study of gamma-rays is an established discipline in modern astrophysics that involves compact objects like neutron stars and black holes or cosmological scale object such as galaxies. Gamma-ray astronomy also studies cosmology and particle physics. In fact, the cosmic rays (charged particles) are deflected by the electromagnetic fields present in the universe before being observed from Earth, while the gamma-rays travel in straight lines and can be used to study the origin of cosmic rays.

In recent years the Multi-Messenger astronomy described in Sec. 1.6 studies the same physical phenomena collecting signals from different messengers. The gamma-rays have an essential role in detecting a counterpart of these phenomena with electromagnetic radiation.

The gamma-rays can be produced by several mechanisms:

- Proton-proton collision: this mechanism produces gamma-ray when a proton or a cosmic ray strikes another proton or atomic nucleus. This collision produces several things from which pions are unstable particles that decay into a pair of gamma-rays.

- Matter-antimatter annihilation: when a particle and its anti-particle (e.g. electron and positron) collide, this annihilation process produces pions that decay in gamma-rays.

- Radioactive decay: when an element changes to another element, it leaves the nucleus in an excited state. When the excited atom decays into the ground state, it emits gamma-ray.

- Acceleration of charged particles: in this mechanism, a magnetic field accelerate a charged particle that emits radiation. This mechanism associated with electrons produces gamma-rays. If the electron is accelerated in the electrostatic field around a nucleus, the phenomenon is called Bremsstrahlung. When a static magnetic field produces the acceleration, the resulting radiation is called Synchrotron radiation. The process is called Compton scattering when the acceleration occurs in the electromagnetic field of a photon that deviates the radiation.

These mechanisms can occur in Galactic (e.g. pulsars, supernova remnants, and gamma-ray binaries) or extra-Galactic sources such as the Gamma-Ray Bursts described in Sec. 1.6.

The Earth's atmosphere blocks gamma-rays, and for this reason, the first studies about gamma-rays were carried out using space-based missions. Few gamma-ray space observatories are currently in operations. The AGILE satellite, described in Sec. 1.2, is one of them. The researchers also developed instruments to detect gamma-rays from ground-based telescopes, exploiting the fact that when the gamma-rays interact with the atmosphere, the primary gamma-ray can not be seen, but it creates secondary products that can be detected from the ground and used to infer the properties of the original gamma-ray (e.g. direction, energy and more). This secondary product of a gamma-ray that interact with the atmosphere produces a Cherenkov light, a shower of light produced when charged particles (products of gamma-ray interaction with the atmosphere) pass through a medium at speeds greater than the speed of light in that medium. The hard part of the gamma-ray study from the ground is distinguishing the shower generated by cosmic rays (the greatest part) from those generated by gamma-rays. The use of pixelated cameras was essential to visualize the showers' different shapes and classify them to discriminate between proton-based and gamma-ray-based showers. The researchers obtained another improvement by connecting several Cherenkov telescopes to study the same shower from different angles and also to suppress the background. These technologies are the fundamentals for the new gamma-ray observatories such as the Cherenkov Telescope Array (1.3) and the ASTRI Mini Array (1.4).

1.2 AGILE Space Mission

AGILE (Astrorivelatore Gamma ad Immagini LEggero - Light Imager for Gamma-Ray Astrophysics) is a scientific mission of the Italian Space Agency (ASI) that was launched on 23rd Apr 2007 [Tav+08] [Tav+09]. The AGILE payload detector consists of the Silicon Tracker (ST) [Bar+01; Pre+03; Bul+10; Cat+11] the SuperAGILE X-ray detector [Fer+07], the CsI(Tl) Mini-Calorimeter (MCAL) [Lab+09], and an AntiCoincidence System (ACS) [Per+06]. The combination of ST, MCAL, and ACS forms the Gamma-Ray Imaging Detector (GRID). The AGILE/GRID is used for observations in the 30 MeV-50 GeV energy range. The Precise Positioning System and the two Star Sensors provide accurate timing, positional, and attitude information. The ST is the core of the AGILE/GRID, and it relies on the process of photon conversion into electron-positron pairs. It consists of 12 trays, the first 10 of which include a tungsten converter followed by a pair of silicon microstrip detectors with strips orthogonal to each other, the last two consisting only of silicon detectors. The gamma-rays are converted in the tungsten (silicon) layers, and a readout electronics acquire and process the data.

The data produced by these instruments onboard AGILE is downlinked in the ASI ground station (Malindi, Kenya) almost at each orbit (about every 90 minutes) when the connection is possible. This connection between AGILE and the ground station is called "contact". Usually, it contains more than one orbit data because the data are not downloaded at each orbit for technical reasons but are stored onboard the spacecraft until the next contact. From here on for contact it is intended the data package received to the ground station. The data is immediately transferred to Italy via ASINet with a dedicated bi-directional IntelSat link to Telespazio, Fucino (AQ, Italy). The data is then sent to the ASI Space Science Data Center[2] (SSDC), which performs the reduction, archiving, and distribution of this data. The data preprocessing is performed by the AGILE Preprocessing System [Tri+08]. This data stream is shown in Fig. 1.1, which taken from [Bul19].

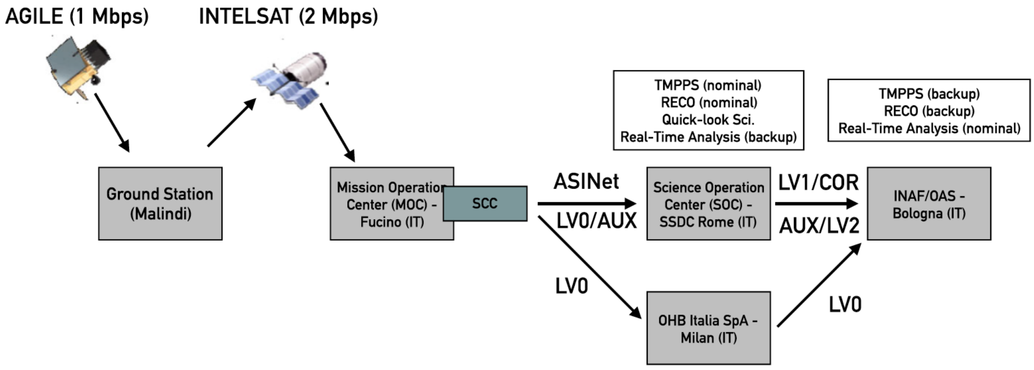

Figure 1.1 AGILE Telemetry data flow. Every $\sim$ 96 minutes the satellite passes over the Malindi (Kenya) Ground Station. At each pass, the satellite telemetry is downlinked and forwarded in near real-time to the Mission Operations Centre (MOC) via INTELSAT. Telemetry is then transferred by the MOC Satellite Control Centre (SSC) to the AGILE Science Operation Center (SOC) at the ASI Space Science Data Centre (SSDC) and then to the INAF/OAS Bologna site. A backup chain from OHB Italia SpA in Milan is also present. This figure is taken from [Bul19].

The reconstruction step performed for the AGILE/GRID data uses an implementation of the Kalman Filter technique. This reconstruction is performed on AGILE/GRID raw data acquired during the observation and downloaded from the satellite to the on-ground station. The background event filter FM3.119 is then applied to assign to each event a classification flag. The AGILE auxiliary file (LOG) is then created with all the spacecraft information relevant to the computation of the effective exposure and Good Time Intervals (GTI).

Finally, the event direction in Galactic coordinates is reconstructed and written in the AGILE event files (EVT), excluding events flagged as charged background particles. This procedure, described in [Bul+19], produces the

[2] https://www.ssdc.asi.it

Level-2 (LV2) archive of LOG and EVT files. The AGILE/GRID data are publicly available from the ASI/SSDC website[3].

During this PhD, the software for the gamma-ray data analysis of the AGILE space mission was developed in collaboration with the AGILE Team. The Sec. 2 describes the software developed to perform part of the analysis for the Second AGILE catalogue of gamma-ray sources, store the results in a database, and visualise the results through a web GUI. AGILE is used as a case study in Sec. 3 to show how the RTApipe framework was used to develop the real-time analysis pipelines for AGILE. The gamma-ray sky maps produced with AGILE/GRID data are also the input for the Convolutional Neural Network described in Sec. 4.

1.3 Cherenkov Telescope Array

The Cherenkov Telescope Array[4] (CTA) is an observatory for the ground-based gamma-ray astronomy, for the first time in Very-High-Energy (VHE) field this observatory will be driven by proposals and will have open access for data. This observatory aims to build and operate more than 100 Imaging Atmospheric Cherenkov Telescope (IACT) [Kre09].

The IACT telescopes will be of three different class: Large-Sized Telescope (LST), Medium-Sized Telescope (MST) and Small-Sized Telescope (SST). The different size of the telescope changes their capability to detect gamma-rays at different energies. The telescopes will be installed in two different sites on in the northern hemispheres (La Palma, Spain) and the other in southern hemispheres (Paranal, Chile). The project office of CTA decided the sites after years of study, and extensive investigations of the environmental conditions, simulations of the science performance and evaluations of construction and operation costs.

CTA provides a unique gamma-ray observation sensitivity to study crucial open questions in astrophysics, cosmology, and fundamental physics as described in [Che+20]. It will also be able to perform unprecedented observations of the transient event in the very high-energy sky. The sensitivity performances and the analysis tools for this short-term analysis are described in [Fio+19].

CTA will have several science use cases that are described in the white book [Che+19]. The three main scientific themes are:

- Understanding the Origin and Role of Relativistic Cosmic Particles.

[3]https://www.ssdc.asi.it/mmia/index.php?mission=agilemmia
[4]https://www.cta-observatory.org

- Probing Extreme Environments.
- Exploring Frontiers in Physics.

Some of the main targets that CTA will observe to search for discoveries are pulsars, black holes, supernova remnants, dark matter, binary systems and more.

The CTA receives contributions from the CTA Consortium, including 1500 members from more than 200 institutes in 31 countries. The work for this PhD is inserted in this international research context where the INAF is one of the main contributors.

During this PhD, the research activities on CTA were carried on with the INAF/OAS Bologna Team devoted to developing the real-time analysis pipeline of CTA, described in Sec. 3.7, that will be part of the Science Alert Generation system. This system aims to detect candidate science alerts during the CTA observations as soon as the telescopes acquire the data.

1.4 ASTRI Mini Array

The ASTRI Mini-Array (MA) [Par+16] is an INAF observatory aiming to construct and operate an experiment to study gamma-ray sources emitting at very high energy in the TeV spectral band and intensity interferometry observations. The ASTRI MA consists of an array of nine innovative IACT telescopes that are an evolution of the two-mirror ASTRI Horn telescope successfully tested since 2014 at Serra La Nave Astronomical Station of the INAF Observatory of Catania, Italy. Each telescope will be equipped with the new ASTRICAM Silicon photomultiplier Cherenkov Camera [Cat+18].

The ASTRI Mini-Array's nine telescopes will be distributed at one hundred meters of distance from each other at the Teide Astronomical Observatory, operated by the Instituto de Astrofisica de Canarias (IAC), on Mount Teide (2400 m a.s.l.) in Tenerife (Canary Islands, Spain). INAF will operate the ASTRI MA on the basis of a host agreement with IAC.

The main scientific goal of the ASTRI MA is to perform high-energy ($E > 1TeV$) observations of galactic and extragalactic sources with a sensitivity better than that reachable by the other IACT telescopes currently in operation (HESS, MAGIC, VERITAS). Furthermore, the ASTRI MA will also perform Intensity Interferometry of a selected sample of bright sources. The Intensity Interferometry observation using an IACT telescope (MAGIC) is described in [Acc+20].

The ASTRI MA must be operated remotely, and no human presence is foreseen on-site during observations. A data centre will be installed on-site to have computing power close to the telescopes. The data acquired from the telescopes are sent to the on-site data centre for a fast quick look analysis to check the data quality and have feedback during the observations. The data centre installed on-site must be optimised to reduce costs without compromising security and safety operations and the integrity of the collected data.

The ASTRI MA will benefit from the high-speed networking connection already present at Teide to deliver all data to the Italian ASTRI MA data centre, limiting the number of storage devices on-site. When the data arrives at the main data centre at INAF/OAR Roma, it is analysed with automated software to provide fast results.

During this PhD, a collaboration with the INAF/OAS Bologna Team devoted to developing the ASTRI MA software was carried on. The collaboration aimed to develop the automated analysis pipeline prototype to analyse the ASTRI Horn telescope [Gir+19] (a pathfinder for the MA) simulated gamma-ray data. This work is described in Sec. 3.8.

1.5 Big Data and Astronomy

One of the most used definitions of Big Data was given by Doug Laney [Lan01] in 2001 analysing three properties of Big Data called "The 3 Vs": Volume, Velocity and Variety. Volume means the amount of data that must be managed, stored and analysed. The technological breakthroughs improve the capability to handle high data volume; thus, there is no constant threshold data quantity that defines Big Data.

Data Velocity is the fast rate at which the data are received from the data sources. Some contexts require real-time or near-real-time analysis and results.

Variety refers to a different type of data that can be received (e.g. text, images, videos etc.). Usually, data are not structured and can be generated from several sources. Traditional integration architectures, such as Extract/Transform/Load (ETL), designed to work with the standard relational database, have low performances managing not structured Big Data. For this reason, new frameworks are developed to integrate, manage and analyse Big Data (Hadoop, Spark, Flink and more).

In recent years, two Vs were added to the definition of Big Data: Veracity and Value. Big Data Value is a key factor for companies. Some use cases can be predictive maintenance, fraud detection, machine learning, business decision making and more. It is possible to extract valuable information from Big Data only if reliable data are available. Veracity is a property of

Big Data which ensures that the data and data sources are reliable and meet all the security requirements. Big Data have a lot of Value for companies, but it is essential to achieve a high level of Veracity to make crucial decisions based on the analysis of these data.

The new generation of terrestrial observatories (e.g. CTA and ASTRI Mini Array) will consist of tens or hundreds of telescopes. The analysis and the management of the astrophysical data generated by this new generation of observatories will be a challenge. For this reason, the observatories are developing technologies to analyse and manage their Big Data in collaboration with computer scientist and data scientist, as described in [ZZ15].

These data can be considered Big Data for their properties. In fact, regarding the definition of Big Data with the 5Vs, the Volume (Petabytes/year) and Velocity (Terabytes/second) with which these observatories will generate data are very high. In addition, these observatories will analyse the data in real-time to produce scientific results as soon as possible.

Each observatory has its data format, and usually, it is required to analyse data acquired by several observatories managing their different data formats. The data produced by the observatories during the whole analysis process consist of a Variety of different types:

- Raw data generated directly by gamma-ray detectors.
- Plots: histogram (1D, 2D, and 3D), time series, spectra, and more.
- Aggregated data: light curve, counts maps and science alerts.

The Veracity of the data produced during the observation is verified with data quality analyses performed automatically during the data acquisition. These analyses aim to verify that the quality requirements for the data used to produce scientific results are satisfied.

In the context of a real-time analysis pipeline, the data have a great Value because the results of the analyses can generate insight or used to take decision and actions. The software can automatically react to improve the observation or perform the follow-up of science alerts. In addition, when a transient event is detected inside the data, the software automatically sent to the astrophysical community a science alert. This fast reaction is possible only with software that manages the data in real-time and after an on-line data quality check to verify the Veracity of the data.

The RTApipe framework described in Sec. 3 can be configured to work with the Big Data acquired by the next generation observatories (e.g. CTA).

The CNN described in Sec. 4 and developed during this PhD, can be used after the training phase to analyse a huge amount of gamma-ray data, exploiting the computing capabilities of the modern GPUs.

1.6 Multi-wavelength and Multi-Messenger Astronomy

The today astrophysics is Multi-Wavelength (MW), i.e. astrophysical observations are performed using the data of different observatories that collect electromagnetic radiation at different wavelengths (gamma-ray, x-ray, ultraviolet, infrared, etc.).

In recent years, the astrophysical landscape changed due to two main transient events. The first event was the Gamma-Ray Burst (GRB) detected on 17-Aug-2017 by the Fermi Gamma-ray Space Telescope [Aje+18] after the LIGO-Virgo Collaboration detection of the Gravitational Wave (GW) GW170817, produced during a binary neutron star coalescence [Abb+17a]. This GW event was observed with multi-messenger instruments during the following days [Abb+17b]. The GWs are an oscillating strain in space-time and travel through space at the speed of light as described in [Abb+09].

GRBs [GM12] are the most extreme explosive events in the Universe. The duration can be a few milliseconds to several minutes. GRBs have luminosity hundreds of times brighter than a typical supernova and about a million trillion times as bright as the Sun. The energy released during the prompt phase is $\sim 10^{51} ergs$. GRBs can be classified in long-duration GRBs and short-duration GRBs. The threshold between the two classes of GRBs is about 2 sec.

The second event was the neutrino detected by IceCube (IceCube-170922A) which led to an extensive campaign of observations, carried out by different wavelength electromagnetic observatories [Aje+18]. During the campaign, the Fermi-LAT instrument detected an enhanced gamma-ray emission from the blazar TXS 0506+056 at the time the neutrino arrived. These events started the so-called "Multi-Messenger era" (MM) [Mes+19].

In the MW and MM astronomy, the observatories study these GRBs with the results obtained from several instruments to understand their origin and behaviour. AGILE can detect GRBs with all its instruments GRID [Del+11], MCAL [Urs+19c] and SA [Del+08].

The coordination of different observatories in the MW and MM context is possible thanks to the sharing of science alerts through the Gamma-Ray Coordinates Network[5] (GCN). The GCN can be used for automated software and fast reaction to external science alerts. The astrophysical community also uses other networks (the Astronomer's Telegram[6], AMON[7] and more) but

[5] https://gcn.gsfc.nasa.gov
[6] http://www.astronomerstelegram.org
[7] https://www.amon.psu.edu

for different purposes. More details on the GCN and AMON networks can be found in Sec. 3.1.

A science alert is a communication from/to the astrophysical community that a transient phenomenon occurs in the sky. In this context, the observatories share their information with the community to study the same transient event with different "messenger" signals: electromagnetic radiation, gravitational waves, and neutrinos.

During this PhD, software to manage and analyse these science alerts was designed and developed. The Sec. 3 describes the RTApipe framework developed and used to create the real-time analysis software for the gamma-ray data acquired by observatories like AGILE or CTA.

1.7 Real-time Scientific Analysis Pipelines

In the MW and MM context, observatories share information about transients phenomena with the scientific community through the GCN network. The coordination of different observatories done by sharing science alerts is mandatory to understand and study these physical phenomena. For these reasons, each observatory has developed a real-time scientific analysis pipeline to identify possible transient phenomena detected by instruments (e.g. GRBs), send alerts to the astrophysical community, and speed up external alerts' reaction time. A review of the real-time analysis pipeline developed for some observatories is covered in Sec. 3.1.

These real-time pipelines can work following three main patterns. The first pattern executes a blind search to detect transient events using only the observatory data without any suggestion about the source position. The second pattern analyse known sources using only the data of the observatory to search for transient events. The third reacts to detection performed by other observatories to search for a counterpart into the incoming or archive data (a.k.a external science alerts). In this thesis, both scenarios (the generation of internal science alerts, and the management of external science alerts) are described and used to develop analysis software.

During this PhD, the RTApipe framework (described in Sec. 3) was designed and developed to help researchers implementing flexible real-time scientific analysis pipelines to fulfil gamma-ray observatories' requirements (Sec. 3.3). This framework can be used to develop pipelines that follow the analysis patterns described before. With these pipelines, it is possible to obtain fast results and interact with this astrophysical community.

The RTApipe framework has two main use cases. The first is that the pipeline reacts to an external science alerts, received through the GCN network, starting scientific analysis on the data available or waiting for data coverage. The second is that the pipeline executes periodically scientific analysis performing a blind search of transient events inside the observatory data stream or performing analysis at a fixed position in the sky to generate science alerts. If the pipeline detects a science alert with a high statistical significance in both scenarios, it sends this alert to the scientific community through the GCN network.

The framework is developed to minimise the reaction time and manage the different analysis patterns in parallel. This rapid reaction is crucial in this context, where other observatories need to be informed to start their observations on the same transient event.

This framework can run analyses using different software called Science Tools developed by researchers. One of these Science Tools that can be implemented into the real-time analysis pipeline is the CNN, described in Sec. 4, developed during this PhD to detect GRBs inside the AGILE gamma-ray sky maps.

During this PhD, the RTApipe framework was used to develop several real-time analysis pipelines for the AGILE space mission described in Sec. 3.6. These pipelines implement both the analysis patterns to perform a blind search for transient events and react to external science alerts. This automated software is mandatory to enable the AGILE Team for a fast reaction to external and internal science alerts. Without these pipelines, it is impossible for the team to rapidly follow-up the transient events and share the science alerts with the scientific community because all the analyses should be performed manually.

The AGILE researchers published several science alerts through the GCN network and the ATel web portal using these pipelines' results. The most important scientific results obtained with the support of the pipelines are published on refereed journals. The Sec. 3.6.4 describes the results obtained with these pipelines and published by the AGILE Team.

1.8 Machine Learning and Astronomy

Machine Learning (ML) is a computer science field that aims to give computers the ability to learn from experience performing specific actions without being explicitly programmed. The classic programming approach requires writing rules and instructions (if-then) provided to the computer to perform a series of tasks. With the ML approach, programmers write an algorithm capable of defining a mathematical model based on the input data. This procedure is based on statistical inference [GBC16], and there is no human intervention during the learning process. With this approach, computers can learn how to solve a complex problem learning from a big dataset of examples.

ML has been used in astronomy for many years. In recent years a new type of ML is gaining momentum: Deep Learning (DL). The DL is a subset of the ML that uses algorithms capable of creating large mathematical models; these models' training requires thousands or millions of input data. The DL is based on the Deep Neural Networks (DNN) inspired by the human brain and neurons. DL algorithms are able to perform a complex task such as image classification, speech recognition, self-driving car and more.

In astronomy, the DL techniques are commonly used in image analysis. For this specific task, the most used DL architecture is the CNN. The Sec. 4.5.1 covers several use cases of DL techniques in the astronomy research field. In [Bar19], many examples of ML and DL techniques that can be used for classification, clustering, and regression problems in astronomy are described.

The CNN developed during this PhD is used to classify gamma-ray maps produced with AGILE/GRID data. This CNN is described in detail in sec 4 with all the steps performed to simulate the huge datasets prepared to train the model and the evaluation of the trained model with real data.

Chapter 2

Gamma-ray Data Analysis and Management

This chapter describes the software developed during this PhD in support of the AGILE Team's analyses to prepare the Second AGILE/GRID catalogue of gamma-ray sources [Bul+19]. This catalogue contains all the gamma-ray sources detected with the AGILE/GRID instrument during the first two years of AGILE gamma-ray observations. The software created for this catalogue is divided into different components developed during three phases following the AGILE Team researchers' needs. The first part of the software includes scripts used to automatise the analyses and collect the results (described in Sec. 2.1). The second part is a MySQL[1] database (Sec. 2.2) used to store all the results of the analyses and the scripts needed to import the results into the database. The last part is a Graphical User Interface (GUI) that researchers can use to visualise the results with plots an tables (Sec. 2.3).

The AGILE/GRID archive data used to produce this catalogue can be analysed in the future using new technologies like the CNN developed to detect GRBs as described in Sec. 4. More details about this future work are described in the Conclusion and Future Work section (Sec. 5).

2.1 Analysis for the Second AGILE catalogue of gamma-ray sources

The AGILE Team published, on the Astronomy and Astrophysics journal, the second catalogue of high-energy gamma-ray sources [Bul+19] detected during the first 2.3 years of AGILE operations (2007-2009) in the so-called 'pointing mode'. This catalogue, called 2AGL, contains sources detected in the energy range 100 MeV - 10 GeV. The sources are detected performing several analyses on the raw data acquired by the GRID instrument onboard the AGILE satellite (Sec. 1.2). The data used for the analysis are generated during the gamma-ray event reconstruction step performed with the AGILE/GRID implementation of the Kalman Filter technique described in Sec. 1.2

During this PhD, several scripts to manage the execution of these analyses performed with the BUILD25 of the AGILE Science Tools [Bul+19] were developed. The version control of these scripts is obtained with Git[2], and the software is uploaded on GitHub[3]. These analyses are performed on the AGILE/GRID archive data using the position of gamma-ray sources identified

[1]https://www.mysql.com
[2]https://git-scm.com
[3]https://github.com

during the work of the 2AGL catalogue. The scripts' goal is to parallelise the analyses' execution and divide the whole process into different phases. The procedure starts with the gamma-ray data observed outside the Galactic plane and then analyses the Galactic plane's data.

After the computational analysis, the results obtained are imported into a database (AGILE-GRID-DB) to store them and allow the researchers to manage the results easily. In addition, a web GUI was developed to visualise these results with plots and tables. The Sec. 2.2 describes the AGILE-GRID-DB used to store these results, while the Sec. 2.3 describes the web GUI. This web GUI is password protected and can be accessed only by the AGILE Team, but future open access is under evaluation.

2.2 Scientific Results Database

The AGILE-GRID-DB is developed to store all the results obtained from the analysis of the AGILE/GRID data performed during the realisation of the catalogue as described in Sec. 2.1. The AGILE-GRID-DB is used as an archive but also as a data source for a web GUI developed to show the results to the AGILE Team. This web GUI is described in detail in Sect 2.3. A first version of the AGILE-GRID-DB was developed during my Master Thesis, in collaboration with AGILE researchers. During the Master Thesis, the software to import the AGILE real-time analysis results, called SPOT6 [Bul+14], and a web GUI to visualise these results with plots and tables were developed. The SPOT6 pipeline executes a blind search of gamma-ray transients inside the AGILE/GRID data with a timescale of days. The AGILE/GRID catalogue described in this chapter is used into the SPOT6 pipeline to associate the transient phenomena with a known source.

subsectionDatabase Schema

The AGILE-GRID-DB is implemented using MySQL because the requirements for this database fit with the relational database architecture. The analysis results are imported automatically into the AGILE-GRID-DB using software developed during this PhD with the Ruby[4] programming language. The first step was to create the Entity-Relationship (E-R) model, starting from the information obtained during the interaction with the researchers that developed the analysis software. With the E-R schema, it is possible to represent a conceptual view of the data through a list of entities, their attributes and the relationships that exist between them. This model is very close to reality and can be useful to understand the data schema. The E-R scheme is divided into various parts, which represent sets of information

[4]https://www.ruby-lang.org/en/

with different purposes. Some information is static and inserted just one time, others dynamic and updated over time.

The Fig. 2.1 shows an high-level schema of the AGILE-GRID-DB. The blue part of the schema consists of static entities; the main ones are:

- Sky region: identifies the sky region where the sources are detected. The regions are established a priori and rarely updated.

- Template: a series of predefined parameters used by the analysis software.

- Filter: the name of the filter used to discriminate the background from gamma-ray photons inside the data acquired by the AGILE/GRID instrument.

- Archive: represents the archive in which the AGILE/GRID data are stored. This field is required to reproduce the results.

Two entities representing the analysis process are indicated in orange in the schema shown in Fig. 2.1. The Analysis entity refers to the various types of analysis performed on the data received from AGILE. They differ in the analysis integration interval, which can be 1, 2, 4, or 7 days or for the Science Tools used to perform the analyses.

For each analysis, the software performs several runs searching for new detections of gamma-ray sources inside the AGILE/GRID instrument's data. The Analysis entity is in relation (1, N) with the Run entity representing these runs. One analysis can have multiple runs. Each run can identify more than one detection. For this reason, the Run entity is in a relationship (1, N) with the Detection entity.

Each run is identified by a name and a unique identification number. It also has other attributes, such as status and start-stop processing times.

The analysis software processes the data collected by the AGILE/GRID instrument searching for gamma-ray emissions. The emissions of a gamma-ray flux from a specific position in space are represented by the Detection entity (green part of the Fig. 2.1). All the information defined inside a detection is necessary to study these astrophysical phenomena.

The main attributes of the Detection entity are:

- Coordinates (l,b,r): l,b define the centre of the source error localisation region in Galactic coordinates where the gamma-ray emission was detected, while r defines the size of this region.

- Time: the start-stop times of the detection.

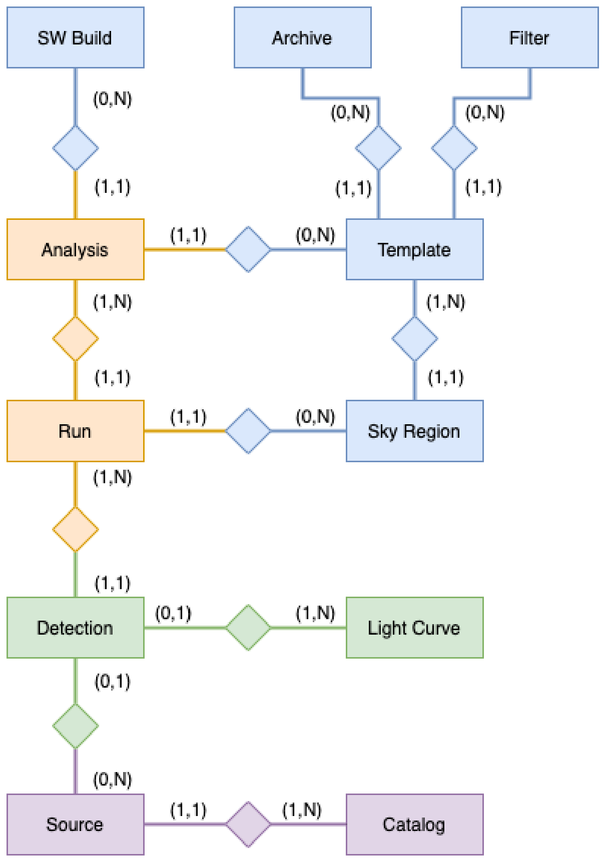

Figure 2.1 High-level schema of the database created to store the AGILE/GRID analysis results. The blue and violet entities represent a static part of the model that is imported before the analysis. The orange entities represent the dynamic part of the model that stores all the analyses execution information. The green entities are used to store the results of the analyses.

- Attributes: additional information about the detection.
- $\sqrt{TS}$: the statistical significance of the detection.

The general attributes of a detection contain all the most important parameters to characterise the detection and allow the researchers to query the AGILE-GRID-DB for several purposes. Researchers need to visualise the data using a time interval or a sky region defined in Galactic coordinates (l,b). The detections can be aggregated by the sky position, creating light curves. The light curves represent the temporal evolution of the gamma-ray flux in a specific sky region (usually associated with a gamma-ray source).

They are inserted into the AGILE-GRID-DB within the Light Curve table. The light curves can be visualised with plots through a web GUI (Fig. 2.2).

The violet entities of the diagram in Fig. (2.1) shows a static part of the AGILE-GRID-DB containing the catalogues of sources detected from different observatories. This part consists of two entities: Source and Catalog. The Catalog entity is in a relationship (1, N) with the Source entity because each catalogue contains several sources (thousands). This part of the AGILE-GRID-DB is needed to associate a detection with one or more known sources when possible. If there are no known sources compatible with the detection, it is associated with a new source, and no relationship is established with the Source entity. The Source and Catalog entities contain static information inserted before starting the data import process.

2.3 Web Graphical User Interface

During this PhD work, a web GUI was developed to visualise the gamma-ray analysis results stored in the AGILE-GRID-DB described in the previous section. This section describes the web GUI and all its functionalities used by the AGILE researchers to prepare the AGILE/GRID Second Catalogue.

The web GUI is implemented within a LAMP environment (Linux, Apache, MySQL, PHP). The whole system is installed inside a server machine located inside the INAF/OAS Bologna data centre. The AGILE Team can access this website remotely using the login credentials provided only to team members. The plots are implemented with the Plotly.js[5] framework and the used programming languages are Javascript, HTML, PHP, and CSS. The GUI layout is implemented with the Bootstrap[6] framework enabling a responsive behaviour compatible with smartphones and tablets.

2.3.1 Query sources from database

The website's home page shows a table containing all the gamma-ray sources of the 2AGL catalogue. The user can download the catalogue information in .txt or .reg formats. The text format includes the following columns: source name, coordinates in galactic projection, error region ellipse parameters, spectral index with error, exposure, association with other catalogues, and statistical significance of the source. The .reg format contains the information required to plot the sources using visualisation tools such as DS9

[5]https://plotly.com/javascript/
[6]https://getbootstrap.com

[7]: ellipse parameters of the error localisation region and the labels of the sources. The table has a quick search field where the user can specify the source's name or coordinates to explore the catalogue. The sources table contains the following information:

- Open Detail: this button is used to open the web page related to the source's details.
- Id: this value represents the id that the source has into the AGILE-GRID-DB. It can be used to run queries directly on the AGILE-GRID-DB and for customised analyses.
- AGILENAME: this is the catalogue name given to the sources following the format 2AGL JHHMM + DDMM.
- Tec Name: This is the source's technical name generated during the data analysis and can be used for querying and searching.
- Confirmed: this attribute is used to confirm a source after several manual analyses performed to verify its presence.
- Comment: any comments entered by researchers.
- $\sqrt{TS}$: this value represents the statistical significance of the analysis that confirmed the source.
- Typefun: this value represents the function type used to calculate the source's energy spectrum [Bul+19].
- Analysis Names: this name refers to the analysis performed to identify the source and can be used to query the results.
- Source Position: this field defines the source's galactic coordinates and the parameters to define the ellipse that represents the error region of the source's position.

2.3.2 Source details visualisation

The detail page related to a source shows several groups of aggregated information about the analyses performed on the AGILE/GRID data related to that source. The core results are the light curves and spectral analyses. The web page is divided into several information blocks. The first block shows

[7]https://sites.google.com/cfa.harvard.edu/saoimageds9

the source's basic information such as the name, gamma-ray flux, galactic coordinates, spectral index, and other attributes.

The second block shows more technical information about the analyses, which parameters are used and more detailed results. These results are focused for researchers that want to repeat the analysis.

The Fig. 2.2 shows a light curve of the selected source calculated with a predefined time binning. The user can choose to visualise the light curve between different integration intervals (1, 2, 4, and 7 days). The X-axis of the diagram represents time in the Modified Julian Date (MJD) format, while the Y-axis represents the gamma-ray flux in ph cm^{-2} s^{-1}. Each detection is visualised with a coloured dot, a vertical bar representing the error on the detection's flux, and a horizontal bar indicating the detection's time windows. The diagram's title shows the analysis's name to generate the light curve and the essential parameters to reproduce the result. The diagram also shows two dashed green horizontal lines representing the average flux of the source $\pm$ the flux error. These lines are useful to visualise when the source flux is greater or lower than this average level. There are sources with variable gamma-ray emission fluxes and sources with stable fluxes. The plot is implemented using the Plotly framework described in Sec. 2.3. This framework offers many features, some of which are very useful:

- Hovering the mouse on a detection shows a label with information.
- The plot can be zoomed and panned.
- The plot can be downloaded in PNG format.

The user can download the detection data with the green buttons located below the plot. The blue button "Upload File and Plot LC" can be used to upload a file containing a light curve and show it within the plot with the light curve of the source. The user can compare two light curves with this functionality. The light curve file to upload must contain tstart, tstop, sqrrts, flux, and fluxerr parameters separated by a space. All the detections aggregated within the light curve are presented with a table after the light curve plot. This table shows the following information:

- $\sqrt{TS}$: the statistical significance of the maximum likelihood analysis of the detection.
- L-peak, B-peak: the Galactic coordinates of the detection.
- Exp: the exposure of the AGILE-GRID instrument during the time window of the detection expressed in cm^2 s .

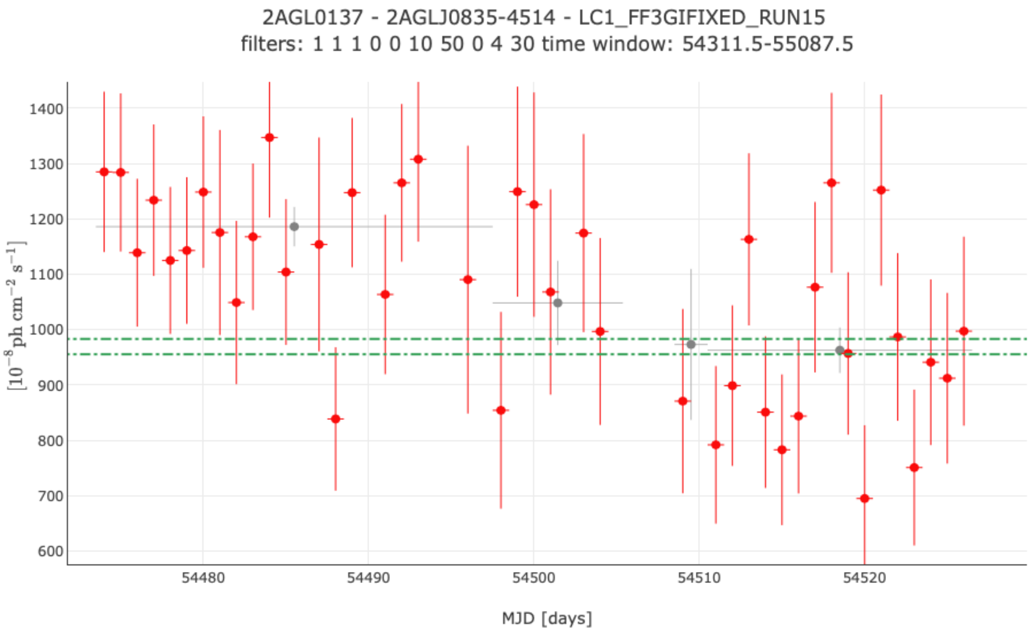

Figure 2.2 Light curve of a gamma-ray source from the 2AGL catalog representing the flux behaviour in function of the time. Each point indicates a detection of the source with the gamma-ray flux and its error. The two horizontal green dashed lines represent the mean flux value of the source ± the flux error.

- Time: the integration time window for the detection analysis expressed in various formats.
- Flux, flux-err: detection flux with its error, both expressed in ph cm^{-2} s^{-1}.
- Flux-ul: the upper limit on the flux expressed in ph cm^{-2} $s^{-1}$and calculated when the $\sqrt{TS}$ is less than three and the detection is not confirmed.
- Counts ± error Counts error: the gamma-ray photon counts of the detection with its error.
- Spectral-index ± error: the spectral index with its error.
- A button to open the gamma-ray sky map centred in the source position.

The Fig. 2.3 shows the result of the spectral analysis. The axes of the diagram are represented with a logarithmic scale. The energy bins are expressed in MeV and shown on the X-axis, while the Y-axis shows the flux in ph cm^{-2} s^{-1}. The user can upload a file containing the information of another spectrum to plot and compare it with the spectrum of the 2AGL source. The format of the file to upload the spectrum must contain $\sqrt{TS}$, flux [ph cm^{-2} s^{-1}], flux-error, flux-erg [erg cm^{-2} s^{-1}] flux-erg-error, energy min, energy max, energy bin logarithmic centre, exposure, and flux upper limit.

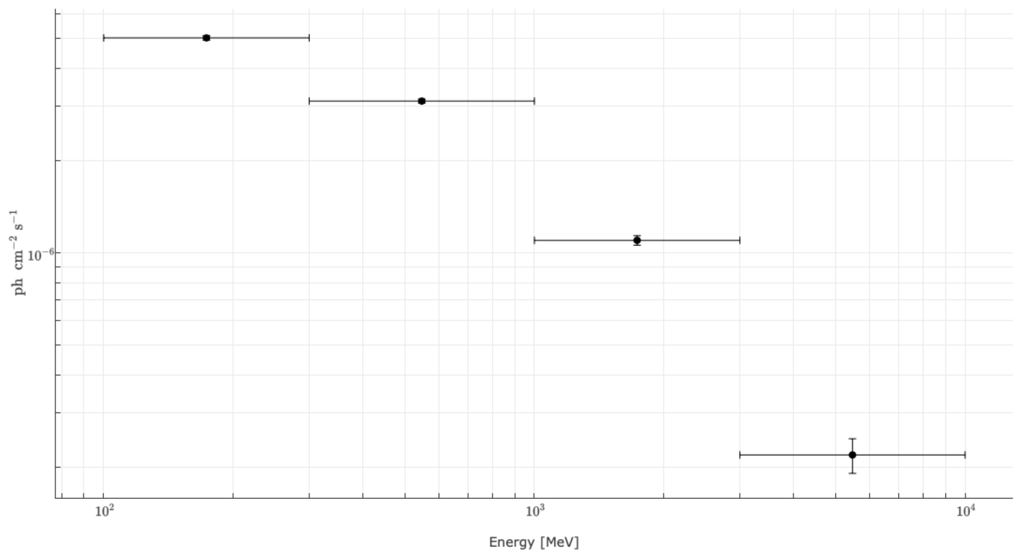

Figure 2.3 Gamma-ray spectrum of the 2AGL source with four energy bins. The X-axis represents the flux expressed in ph cm^{-2} s^{-1}while the Y-axis represents the energy bins: 100-300 Mev, 300-1000 MeV, 1000-3000 MeV, and 3000-10000 MeV.

The data related to the spectrum are shown with a table in which each row represents one energy bin of the spectrum. The table shows the following attributes:

- $\sqrt{TS}$: the statistical significance of the energy bin analysis.
- Flux, flux-err: the flux with its error expressed in ph cm^{-2} s^{-1}(the vertical error bar for each detection in the plot).
- Erg, erg-err: the flux with its error expressed in erg cm^{-2} s^{-1}.
- E-min, E-max: the energy range expressed in MeV (horizontal error bar for each detection in the plot).

- E-c, E-geom: the geometric and logarithmic centres concerning the energy interval, both expressed in MeV.
- Exp: the exposure of the AGILE/GRID instrument expressed in cm^2 s .
- Flux-ul: the upper limit on the flux when the statistical significance is less than three sigma.

2.3.3 Web page to export catalogue data

The information about the 2AGL sources imported into the AGILE-GRID-DB during the AGILE Team's analyses can be downloaded in several formats from the web GUI. With this functionality, the user can work with the information of the 2AGL catalogue to perform further manual analyses. This website was also used as a support tool during the 2AGL paper creation. It is possible to download several tables in LaTex format, that can be inserted directly into the paper. With this platform, the AGILE researchers are able to use an iterative workflow, which consists of performing analyses on the data, viewing the results obtained and exporting the latex tables to insert them directly into the catalogue paper.

2.3.4 Diagrams for statistics analysis

This web GUI can generate a group of statistical plots on the sources' parameters of this catalogue. Some of these plots are also reported in the catalogue paper. The researchers use these plots to perform checks and inspections during the analyses.

The plot in the top-left panel of the Fig. 2.4 shows the correlation between the statistical significance of the sources $\sqrt{TS}$ and the error radius with which the sources are positioned. The greater the significance with which a source is identified, the smaller the error range with respect to the positioning should be. If there are sources with the opposite behaviour, the user can visualise this problem inside the plot.

The plot in the top-right panel of the Fig. 2.4 shows the correlation between the sources' integral flux and the flux calculated by summing the flux of each spectral analysis energy bin. The two results should tend to be equal. The dashed line represents the quadrant's bisector, and the sources should be plotted as close as possible to this line.

The histogram shown in the bottom-left panel of the Fig. 2.4 shows the distribution of the spectral index of all sources that have a Power Law as their spectral function.

The histogram shown in the bottom-right panel of the Fig. 2.4 shows the distribution of the error radius of the sources' positions, calculated with a 95% confidence level.

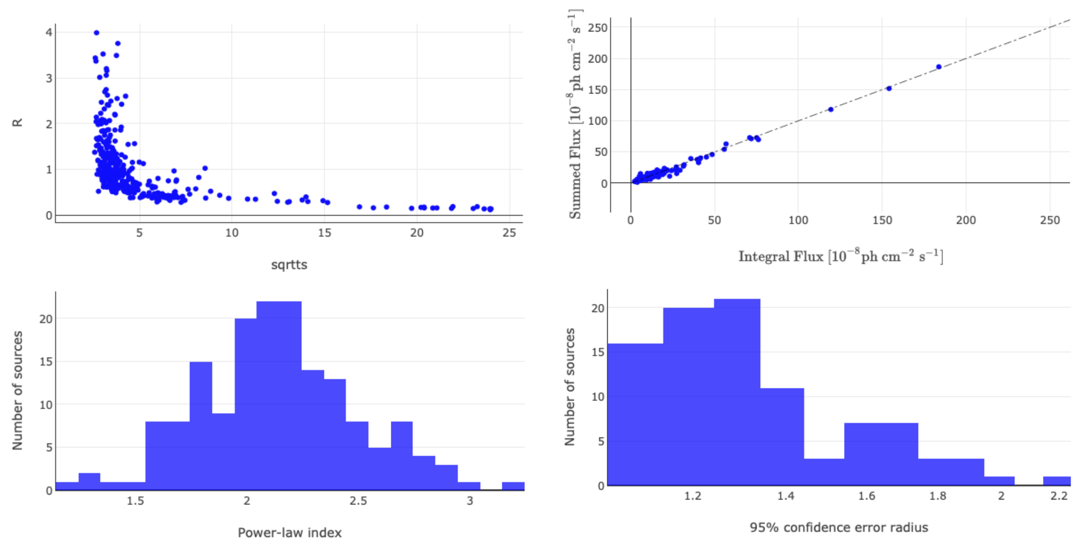

Figure 2.4 Statistical information calculated on the attributes of the 2AGL sources. The top-left panel shows the correlation between the statistical significance of the sources $\sqrt{TS}$ and the error radius with which the sources are positioned. The top-right panel shows the correlation between the sources' integral flux and the flux calculated by summing each spectral analysis energy bin's flux. The bottom-left panel shows the distribution of all sources' spectral indexes that have a Power Law as their spectral function. The bottom-right panel shows the distribution of the error radius of the sources' positions, calculated with a 95% confidence level

2.3.5 Check associations with other catalogues

This last part of the web GUI can be used to verify the associations of the 2AGL sources with the sources present in other catalogues imported into the AGILE-GRID-DB. The user can select the catalogue that wants to compare with 2AGL and then press the button to load the associations' table. The table shows all the 2AGL sources aggregated by type (BLAZAR, SNR, etc.) with the relative associations to the external catalogue selected by the user.

Chapter 3

Framework for Real-Time Scientific Analysis Pipelines

As already introduced in Sec. 1.7 the observatories share information about transients phenomena with the scientific community to study the same physical event with different instruments. This information sharing grow up in the last years due to the so-called Multi-Wavelength (MW) and Multi-Messenger (MM) era, described in Sec. 1.6. The observatories develop their own real-time scientific analysis pipelines to identify possible transient phenomena (e.g. GRBs) and react to external alerts sent from other observatories.

During this PhD, a framework called RTApipe is designed and developed in collaboration with researchers of the AGILE space mission. This framework aims to enable the developers to implement flexible real-time scientific analysis pipelines to fulfil the scientific goal of gamma-ray observatories. The solutions adopted in this framework derive from the know-how acquired during the development of several real-time analysis pipelines for the AGILE space mission, which are described in the work [Bul19].

The following sections describe the RTApipe framework. The AGILE space mission is used as a case study to show how the framework can satisfy the gamma-ray observatories requirements. Several real-time scientific analysis pipelines for AGILE were implemented using this framework, and this work is described in detail in the following sections of this chapter. The RTApipe framework is also described in a paper we submitted to the Astronomical Data Analysis Software and Systems 2020 Conference [Par+a].

The gamma-ray observatories have a Scientific Team composed of Astronomers that shall visualise and check the analysis results produced by the real-time analysis pipelines as soon as they are available. The Astronomer on-duty is the person that is in charge of the follow-up of science alerts. If the pipeline detects a transient event, the Astronomer on-duty is informed and shall manage this alert. The AGILE space mission has a Scientific Team dedicated to the follow-up of transients detected by the observatory or science alerts received by other observatories. The team is organised with daily shifts (24 hours on 24) in which the Astronomer on-duty has the responsibility for the follow-up of the real-time analyses results. The Astronomer on-duty must guarantee a rapid response when internal and external alerts are received to validate the automatic analyses' results and send the scientific community a communication concerning the event.

3.1 Context and Related Works

In the today MW and MM era, the real-time analysis pipeline is mandatory for the observatories that aim to receive and send science alerts through networks like the Gamma-Ray Coordinates Network (GCN) and the Astrophysical Multi-messenger Observatory Network (AMON).

The GCN [Bar+00] is a service developed for the first time in 1993 to distribute the sky position coordinates of GRBs to the observatories community in real-time (a few seconds) while the transient event is still ongoing. The observatories that receive this science alert can follow-up the events with their instruments performing MW simultaneous observations. In recent years this network was updated to share other science alert types (GWs and neutrinos) to perform MM observations of the same physical transient events. Two types of communications can be managed through the GCN network: (i) Circulars in a human-readable format and (ii) Notices to be interpreted by automated software. The GCN system is fully automated, and there are no humans involved. This architecture enables low-latency management of the science alerts with a delay between 1 and 60 seconds. The full list of the science alerts sent to the GCN is available on the web portal[1]. The observatories can subscribe to the notification channel, defining the observatories' list for which the science alerts must be forwarded. In [Bar+00] are reported all the available filters and configurations.

The GCN service is integrated with the RTApipe framework described in this chapter. It is used by the real-time analysis developed during this PhD to send and receive science alerts related to transient events.

The AMON [Kei+17] is a consortium of many observatories that acquire data with different messenger signals. This consortium aims to search correlations between the sub-threshold detections in one messenger and other sub-threshold signals in other messengers. A sub-threshold signal is a transient event with insufficient statistical significance to become a reliable science alert. If more observatories detect a sub-threshold signal simultaneously, the AMON real-time analysis pipeline can correlate these events and find reliable above-threshold detection. When a reliable detection is found, it is rapidly distributed to all the other observatories and the community, to facilitate rapid follow-up. AMON is connected with the GCN network to receive and send science alerts. The AMON system implements the world's first science alerts database with a terabyte-scale in the MM context. The database is designed to store events received in real-time from the AMON member observatories as well as their archival data. The AMON analysis system, described in [Kei+17], is an example of a real-time analysis pipeline that analyses the data acquired by several observatories to increase the transients event's detection capability.

The Swift mission [Geh+04] is an MW observatory for GRB astronomy.

[1] https://gcn.gsfc.nasa.gov/gcn3_archive.html

It is an autonomous rapid-slewing satellite to study the transient astronomy. It observes more than 100 GRB/year and can perform a detailed X-ray and UV/optical afterglow follow-up in timescales ranging from 1 minute to several days after the GRB.

Swift has four main scientific goals:

- Determine the origin of GRBs.
- Classify GRBs and search for new types.
- Study the interaction of the ultrarelativistic outflows of GRBs with their surrounding medium.
- Use GRBs to study the early universe.

This spacecraft consists of three instruments: a wide-field gamma-ray (15–150 keV) detector that detects GRBs, and triggers autonomous spacecraft slews; a narrow-field X-ray telescope to perform spectroscopy in the 0.2–10 keV band; and a narrow-field UV/optical telescope that will operate in the 170– 600 nm band.

The rapid data processing system allows other observatories to follow-up all types of high-energy transients. Swift data processing results are rapidly distributed to the astronomical community, and all interested observatories can perform follow-up measurements.

The Burst Alert Telescope (BAT) [Bar+05] is one of three instruments onboard the Swift spacecraft. The BAT first detects the GRB and localises the burst direction. It takes an onboard decision if the GRB is worth a new pointing to perform a follow-up observation, and if it decides to follow-up the GRB it sends that position to the spacecraft attitude control system. The spacecraft slews to point the two narrow Field Of View (FOV) instruments at the GRB location within 20–70 sec. The pointing change allows the spacecraft to perform a follow-up X-ray and optical observation.

The results generated in real-time onboard the Swift spacecraft (GRB position, light curve, spectral, and image information) are transmitted to the ground via the multiple access system within Tracking and Data Relay Satellite System [2] (TDRSS). Once received from the on-ground processing software these alerts are sent to the GCN network. The GCN network distributes them to other observatories or person that can perform follow-up observations. The total time elapsed since the BAT detects the GRBs and the GRBs positions are available to the scientific community is $\sim 20 sec$. Other information about the GRBs is produced with a time delay.

[2]https://www.nasa.gov/directorates/heo/scan/services/networks/tdrs_main

There are two main stages in the data processing pipeline:

- the quick-look processing happens within a few hours after the data are transmitted on the ground.

- the final processing results are available a week later.

Swift perform 7–10 telemetry contacts per day, and the processing pipelines perform the quick-look analysis in three hours. Usually, the data products are available after a range of ∼4 to 24 hours since the GRB detection onboard the spacecraft. The results can be visualised and downloaded by the scientific community using web portals and File Transfer Protocol (FTP) connections from several data centre all around the world. The results are available while ground-based follow-up activities are in progress or in the planning stages. These data are in standard FITS[3] format and can be analysed with standard Science Tools (e.g., ftools[4])

The Fermi Gamma-Ray Space Telescope [McE+12] was launched on 2008 June 11 into a 565 km orbit with an inclination of 25.6 deg. The payload comprises two science instruments, the Large Area Telescope (LAT) and the Gamma-Ray Burst Monitor (GBM).

The Fermi/GBM [Mee+09] has two main scientific goals. The first is to perform a joint analysis of GRBs observed by both the GBM and the LAT instruments. The secondary objectives are to provide a real-time GRBs location to allow the spacecraft to repoint on the gamma-ray transient to perform LAT observations and share the burst locations as soon as possible with the community of others observatories through the GCN network.

The real-time analysis pipeline that analyses the data acquired by the Fermi/GBM instrument is described in [Gol+20]. This pipeline, called RoboBA, is operational since early 2016. It is composed of a set of algorithms developed to run autonomously replacing the Astronomers intervention during the processing of GRBs. The interest in the detection and localisations of GRBs is increasing. For this reason, the results of these analyses are desirable as soon as possible. The human-in-the-loop processing requires 1-2 hours to have the final localisations, so the fully automated RoboBA pipeline was developed to provide localisations for GRBs with a reduced latency within 10 minutes after the transient event and accuracy comparable to which obtained with manual analysis. The pipelines send a GCN Circular and a machine-readable Notice to the GCN network when a GRB is detected. The Fermi

[3] https://fits.gsfc.nasa.gov/fits_documentation.html
[4] https://heasarc.gsfc.nasa.gov/ftools/

gamma-ray observatory sends to the GCN network several types of communications [5] that can be received from other observatories to follow-up the GRBs detected by Fermi/GBM.

The Fermi/LAT instrument [Ack+12] is the primary instrument onboard the Fermi spacecraft. It is a wide field-of-view imaging telescope, designed for the gamma-ray detection in the energy range from 20 MeV to more than 300 GeV. The LAT instrument consists of three components: a tracker/converter, a calorimeter, and an anticoincidence detector. It has a system onboard the spacecraft for the triggering and data acquisition to select and record the most gamma-ray candidate events transmitted to the ground. When received on the ground, these events are processed by a pipeline that verifies the data integrity at each step and produces and make available all the data products related to calibration and performance monitoring of the LAT instrument. This processing pipeline of the LAT data utilises $\sim$ 300 Terabytes of storage each year.

In [Fla+09; Zim+12] it is described the Data Processing Pipeline that automatically processes the Fermi/LAT data down-linked from the spacecraft. This pipeline automates the production of data quality monitoring quantities, reconstruction and routine analysis of all data received from the spacecraft. It delivers the science products to the collaboration and the Fermi Science Support Center[6]. The reconstruction of raw data and the data processing at each data level requires huge computing resources. This pipeline receives new data every 3 hours and executes about 2000 jobs to process each data arrival, completing the processing before the next data arrives. The real-time pipeline workflow is a time-critical operation and automatically triggers the analysis for each new data arrival. This system exploits the parallel processing executing a graph of analyses described in an XML file that is interpreted by the pipeline to transform this graph of required processing to a list of tasks.

The RTApipe framework implements this parallel processing management using the Slurm open-source workload manager described in Sec. 3.4.6 instead of developing dedicated software with messaging communication and thread pools as done for the Fermi Data Processing Pipeline pipeline. In addition, Slurm can be scaled to a cluster of machines with a few changes in the configurations.

The software has a three-tier architecture and is written almost entirely in Java[7]. The layers are the back-end components, the middle-ware and the

[5]https://gcn.gsfc.nasa.gov/fermi.html
[6]https://fermi.gsfc.nasa.gov/ssc/
[7]https://www.oracle.com/java/

front-end user interfaces. The three-tier architecture [8] separates the application into three layers, and it is largely used for client-server software. This architecture organises the software into three logical and physical computing tiers:

- the data tier, where the data associated with the application is stored and managed.
- the application tier, where data is processed.
- the presentation tier, where the user can visualise the data.

The back-end components (data tier) is implemented with an Oracle[9] relational database that stores all processing states. The Oracle environment is used to develop a scheduler that runs periodic jobs (e.g. for system monitoring). The user can monitor the state of the system with several plots. Stored procedures (Java or PL/SQL) perform within the database query-intensive tasks (e.g. evaluate dependencies between jobs, execute or skip jobs etc.). A similar approach is used to implement the Science Logic component (Sec. 3.4.3) of the RTApipe framework; this component is implemented within a MySQL relational database.

The back-end has an asynchronous message system to send asynchronous messages between the batch-jobs to the Pipeline Server. When a new job is executed, a message is sent containing the worker node's hostname and information. When the job is finished, a message is sent back. These messages can contain additional commands to execute sub-tasks.

The Pipeline Server is the second layer of the system (application tier) and has two pools of Java threads, a work pool and an admin pool. The work pool manages the execution of the jobs allocating threads and submitting tasks to these threads. The admin thread pool searches for jobs waiting for the execution and submitting these jobs to the worker pool. The admin thread also monitors and stores the status of jobs in the database and queries the database to get the list of jobs that are ready to be executed.

A similar component is implemented within the RTApipe framework called Pipeline Manager and described in Sec. 3.4.4.

The last layer of this pipeline is the front-end Graphical User Interface (GUI) (presentation tier). This GUI is implemented with both web and command-line applications that can interact with bottom layers through APIs. The web GUI has a password protected worldwide access and is developed with a cluster of Apache Tomcat servers, and developed with the JSP

[8] https://www.ibm.com/cloud/learn/three-tier-architecture

[9] https://www.oracle.com//database/technologies/

technology[10]. It uses SQL queries to read data from the Oracle database and can interact with Pipeline Server.

The same approach is used during this PhD to develop the real-time analysis prototypes for the Science Alert Generation of CTA and for the science analysis pipeline of the ASTRI Horn telescope. The analyses' results are stored into a MySQL relational database, and the web GUI queries this database to read the required information and visualise them with plots and tables.

As described in [Ban+09] the Fermi/LAT instrument can detect GRBs with an onboard procedure. The software on board the spacecraft can detect GRBs, localise them, and report the information about their positions to the ground with the telemetry. These GRBs are communicated to the scientific community through the GCN network, allowing other space or ground-based telescopes the follow-up of these events.

Chapter 4 of the Book [Way+12] describes the ground-based real-time analysis pipeline that searches for short and long transient events in the Fermi/LAT data. The LAT instrument performs continuous monitoring of the gamma-ray sky with all-sky coverage every three hours. This survey enables the Fermi spacecraft the search for transient gamma-ray events from known and unknown sources. In addition, the data are archived increasing the information acquired on the same source or sky region.

The Fermi Team developed an Automated Science Processing (ASP) system, deployed at the Fermi Instrument Science Operations Center (ISOC), part of the automated data pipeline that processes the raw data. The ASP use the processed data as input to detect transient phenomena and characterise them. When the ASP System detect a transient event, it sends a Notice to the GCN network to enable other observatories to follow-up the flaring sources to detect a possible counterpart in their data.

The ASP software can detect signals produced by point sources like GRBs and Active Galactic Nuclei (AGNs).

The GRBs can be detected with more accuracy at energies below the LAT range. The GBM instrument onboard the Fermi spacecraft is designed to detect these GRBs at lower energies and alert the LAT to perform an analysis to define a better position of the GRBs detected by the GBM. This refinement task is one of the analyses performed by the ASP system.

The ASP software uses as an input to search for GRBs the Notices received through the GCN network by other observatories (INTEGRAL, Swift and more). The information received within the GCN Notices (GRB position, trigger time etc.) are stored in a database and then processed by the

[10]https://www.oracle.com/java/technologies/jspt.html

ASP system.

The AGILE real-time analysis pipeline developed during this PhD described in Sec. 3.6 implement a similar workflow, receiving science alerts from the GCN network, storing them into a MySQL database and processing them with several Science Tools.

The analysis performed by the ASP software extracts the photons in a radius of 15° centred in the position notified with the alert and in a 200 sec time windows centred on the trigger time of the alert event. These photons are analysed with the MLE the search for a transient event.

The ASP software also performs a blind search analysis to detect GRBs not detected by the GBM instrument. The software analyses the data stream from LAT to search for clusters of photons. For each cluster, the probability that these events are generated by a uniform distribution in time and space domain is calculated. It takes into account the probability of the time correlation of the photon and the spatial distribution. The candidate bursts are inserted into the database as if they are alerts received by the GCN network and then the ASP software analyses them.

The ASP software performs an analysis of known sources to monitor their gamma-ray flux. If a source's flux exceeds a certain threshold, a notification is sent to the Astronomer's Telegram. The pipeline that searches for new flaring sources uses data integration of six hours, one day, and one week. These time windows are defined to accommodate humans. For each integration time, the software produces an all-sky counts map that is analysed using continuous wavelet transform to smooth out statistical fluctuations and enhance the signal at the location of point sources. Then a MLE analysis is executed find transient events [Aje+19]. This pipeline is implemented with low-level analysis tools, Python scripts, scheduling systems and databases. A java-based pipeline manages the whole automated system.

This part of the ASP software that monitors known sources and searches gamma-ray flares with a longer time interval than a GRB event is similar to the AGILE SPOT6 pipeline's described in [Bul+14].

The Fermi Team is studying new ML approaches that can be incorporated in his pipeline. These ML models can be trained with human analysis, achieving better results than completely automated analyses. The human feedback involved during the ML training phase is important to reduce false positive.

The science alerts sent by the Fermi/LAT and the Fermi/GBM detectors to the GCN network are received and managed by the real-time analysis pipeline developed for AGILE and described in Sec. 3.6.1.

The Third Interplanetary Network (IPN) [Hur+10] is a group of nine spacecraft equipped with instruments able to detect GRBs. The AGILE and

Fermi space missions are part of this network. The IPN aims to coordinate the observation of the same GRB performed by several observatories to localise the GRB sky position with high precision. The IPN can be considered an all-sky, full-time monitor of transient gamma-ray activity because it eliminates the Earth and Sun occultation problem merging all the spacecraft observations. Being part of this network means that observatories shall share their detections as soon as possible to allow the IPN to localise the GRB position.

As described in Sec. 3.6.4 the real-time analysis pipelines developed for AGILE during this PhD produced several GRB science alerts that contributed for the IPN triangulation to identify a small error localisation region for the GRB.

The IceCube Neutrino Observatory [Aar+17] is a cubic-kilometre detector built into the ice at the South Pole to detect neutrinos. The construction phase ended in December 2010, and the commissioning phase was completed in 2011. This observatory's main scientific goal is the discovery of astrophysical neutrinos and the identification of their sources. IceCube collaborate with MM observatories: optical, X-ray, gamma-ray, radio, and gravitational wave. The detection of neutrinos has an important role in the MM context because it can trigger the transient event's follow-up with other observatories to search for a GRB or GW counterparts.

The IceCube real-time analysis system, described in [Aar+17], is mandatory to achieve the observatory scientific goals and share the results with the community. IceCube is built in a remote place; for this reason, an automated analysis system is required to process data without human intervention. This analysis systems run continuously, performing analyses on data acquired by the detector as soon as possible using a dedicated computing cluster located near the detector site. The analyses perform reconstruction and filtering of the events detected by IceCube. The events that pass the filtering process are sent to the online alert system with a mean delay of 20 seconds. The alert system prepares JSON-formatted[11] messages that are transmitted to the northern hemisphere datacenter. The median total messages latency is about 33 seconds, including the reconstruction, filtering and transmission processes. These event messages are stored in a database and then distributed using the ZMQ[12] framework to a list of real-time analysis pipelines that process the events to determine if a science alerts should be generated and shared with other observatories through the AMON and GCN networks.

The Laser Interferometer Gravitational-Wave Observatory (LIGO) [Abb+09]

[11] http://www.json.org/
[12] http://zeromq.org

is constituted of three specialised Michelson interferometers. Two detectors (H1 and H2) are located on the Hanford site in Washington and have an arms' length of respectively 4 and 2 km. The third detector (L1) is sited on Livingston Parish, Louisiana and has an arms' length of 4 km. The GWs are perturbations in the space-time curvature that propagate at the speed of light as waves and are generated by accelerated masses (e.g. during the of binary black holes or binary neutron stars). When a GW passes through the detectors' arms, their length change and this mutation impress a phase modulation on the laser light transmitted in each Michelson interferometer arm. This phase change is detected, analysing the data acquired by the interferometers. The arm's length change during a GW is only $\sim 10^{-8}m$. The detectors are built-in separated sites to reject instrumental and environmental perturbations into the data. Having more detectors enable the observatory to perform a triangulation between the different sites to localise the source of the GWs.

As described in [Mes+17] the need for low-latency analyses to detect GWs becomes critical in the MW and MM era because the GWs can trigger the follow-up of the transient events by other observatories. The electromagnetic observations executed as the follow-up of GWs may lead to the detection of the GRBs prompt and neutrinos within seconds. In a longer time scales (days to years) these observations are followed by X-ray, optical, and radio afterglows. The data acquired by the LIGO sites are processed within ~ 12 seconds from the data acquisition to obtain low-latency results. The analysis for compact binary coalescences requires huge computing resources. For this reason, it is parallelised on hundreds or thousands of compute nodes. The data must be broadcasted to all these computing nodes as soon as it is available. This fast analysis provides rapid results to the LIGO Astronomer on-duty and enables sharing these science alerts through the GCN network for MM follow-up.

Both the IceCube and LIGO observatories send their science alert to the GCN in the MM astronomy context.

The real-time analysis pipeline developed for AGILE during this PhD and described in Sec. 3.6.1 has a specific workflow for the follow-up of GW science alerts and neutrinos. The AGILE Team is informed via email when the pipeline receives a GW science alert. The Astronomer on-duty is in charge of the follow-up to send a Circular to the GCN network in response to the alert. In several occasions, the AGILE Team was the first team to react and share results to the scientific community in response to this kind of alert.

In addition to spacecraft gamma-ray detectors, there are many on-ground based telescopes for studying the gamma-rays (e.g. MAGIC, HESS, VERI-

TAS, and more). Two on-ground gamma-ray observatories are in the construction phase: the Cherenkov Telescope Array and the ASTRI Mini Array.

The Major Atmospheric Gamma-ray Imaging Cherenkov Telescopes [Tri+10] (MAGIC) observatory is composed of two 17 m diameter mirror Imaging Atmospheric Cherenkov Telescope (IACT) telescopes (MAGIC-I and MAGIC-II). The two telescopes can observe the high energy gamma-ray sky independently or in stereoscopic mode.

In [Tes+13] the Data AcQuisition (DAQ) and real-time analysis software of the MAGIC observatory is described. Each telescope has a DAQ system written in C/C++ language that can manage parallel processing. There are two dedicated server computers where these DAQs runs, one for each telescope. The DAQ system receives the data stream from the telescope electronic boards and performs a list of analyses to reconstruct the events and check the data quality. Checking the data quality in real-time is important to identify hardware or software problems. The reconstructed events are then analysed with the MAGIC Online Analysis (MOLA) [Tes+13] software to obtain an estimate of the gamma-ray flux from sources in the FOV of the telescope in real-time. Two MOLA software systems read the data stream from the DAQ systems of each telescope. Bot the DAQ and MOLA software run independently for each telescope and in two different server computers. The MOLA software has a multithread architecture. Two threads are devoted to reading the data stream from the two DAQs, while a third thread performs the stereoscopic analysis of the data acquired from both telescopes.

The MAGIC Team developed an automatic alert system [Ber+19] to receive science alerts from the GCN network to allow the observatory to follow-up these transient events. Initially, this system was designed to follow-up the GRBs in the MW context, but it was recently updated to receive alerts from neutrino and GW observatories in the MM context. The MAGIC automated alerts systems have three main tasks: (i) receive alerts from the GCN network; (ii) process the alerts using pre-configured analyses to check if the events can be seen with the MAGIC telescopes; (iii) if it is possible, start the automatic follow-up of the transient event. This procedure aims to point the telescopes, if possible, to the sky position of the science alerts received and start the data acquisition and analysis with the software described before.

The software system and real-time analyses described in this section show that space and ground-based observatories developed real-time analysis systems to receive and send their analysis results through the GCN network. During this PhD, the RTApipe framework was designed to help observatories developing real-time analysis pipelines. The RTApipe framework can be used in several contexts (e.g. space or ground-based observatories). The software and solutions developed for the observatories cited in this section are used as a starting point to define the use cases (Sec. 3.2) and the requirements (Sec. 3.3) for the RTApipe framework. At the end of the chapter, the development of real-time analysis pipelines in different contexts (AGILE space mission, CTA and ASTRI Mini Array ground-based observatories) is described. These pipelines confirm this framework's flexibility and the ability to meet the needs of general gamma-ray observatories.

3.2 Use Cases

This framework shall satisfy two main use cases:

- it shall react to an external science alert, received through one of the MW and MM network (e.g. the GCN network) from one of the astrophysical instruments that have detected the transient phenomena, and start scientific analysis on the data available or wait for data coverage.
- it shall execute periodical scientific analysis searching for transient events inside the data stream of the observatory to generate science alerts and sent them to the scientific community through the GCN network.

These two use cases cover the gamma-ray observatories' major activities, and the work for this PhD is principally involved in these activities. On the other hand, these use cases can represent the common gamma-ray scenario found in other observatories, as described in Sec. 3.1. For this reason, this framework will be used to develop the real-time analysis of other observatories like the CTA and projects like AFISS (part of the Horizon 2020 program[13]).

In this context, the RTApipe framework aims to reduce the reaction time to external science alerts and improve the capability to detect transients in real-time inside the observation's data in an automated way.

3.3 Requirements

The RTApipe framework is designed to fulfil the requirements of the gamma-ray observatories real-time scientific analysis software in the MW and MM context. The main requirements are:

1. The analysis software shall implement the use cases described in Sec. 3.2.

[13] https://ec.europa.eu/programmes/horizon2020/en

2. The framework shall support both real-time and offline data processing.
3. The framework shall be usable in different gamma-ray observatory contexts (e.g. space or ground-based, single telescope or array of telescopes).
4. The software shall analyse the data acquired by different arrays of telescopes that are performing parallel observations.
5. The new generation of gamma-ray observatories will produce Big Data. The analysis software shall be able to manage and analyse these data.
6. The software shall be able to execute a great number of analyses at the same time, exploiting parallel processing.
7. The computing power is a constraint; for this reason, the software shall manage queues and prioritise processes to optimise the available resources.
8. Different gamma-ray observatories can have different computing power needs; for this reason, the software developed with this framework shall be scalable in a cluster of servers.
9. The software shall allow the developers the configuration of different analysis workflows.
10. The software shall be able to perform different analyses for each type of science alerts or instruments that triggered it (e.g. GW, neutrino, Fermi/GBM etc.).
11. The analysis software shall be configurable with many different Science Tools and shall allow the developers to update the Science Tools configurations.
12. The analyses shall be performed with several time scales from seconds to hours using the same Science Tools.
13. The real-time analysis software is crucial for gamma-ray observatories, and a failure recovery shall be possible using backup deployments of the pipelines in different data centres.

The RTApipe framework is designed to satisfy these requirements. This framework, designed during the PhD, was used to develop the analysis software for three gamma-ray observatories, one space-based and two ground-based. The software developed with the RTApipe is described in detail in the following sections. The Sec. 3.9 uses the software developed with this framework to show that all the requirements are satisfied.

3.4 Framework architecture

In this section, the architecture of the RTApipe framework and all its components are described in detail. The components implemented within the RTApipe framework are Data Model, Science Logic, Pipeline Manager, and Task Manager. This framework requires external components: Data Sources, Science Tools, Analysis Results, and External Interfaces. The Data Model (DM) defines the framework's data structure in terms of entities and relationships between entities. The DM is very flexible and allows the RTApipe framework to implement real-time scientific analysis pipelines for different contexts (space missions, an array of telescopes, and more). The Science Logic provided within the framework implements the most frequent use cases of a gamma-ray observatory (Sec. 3.2) and satisfies the requirements defined in Sec. 3.3. The Pipeline Manager and the Task Manager are the components implemented to execute the analysis and manage priority between different tasks and queues. Fig. 3.1 shows a high-level overview of the framework and its component. In the next subsections, it is described in detail each component of this framework.

3.4.1 Data Model

The DM consists of all the entities present in this framework: Data Sources, Instruments, Observations, Analysis, Science Tools, and more. The Fig. 3.2 shows a high-level schema of the DM. This DM is developed to fit with different contexts varying from a space mission to an array of ground-based telescopes. The developers of the pipeline can configure the DM to fit the observatory's workflow. Part of this DM is static (showed with green in Fig. 3.2) and can be configured by the developers before starting the data processing. In contrast, the other part is dynamically updated during the operations (showed with blue in Fig. 3.2).

The Data Model is implemented with a MySQL relational database. The relational database fit with the requirements of the framework. MySQL is an open-source relational database whit a big community, and it is possible to assert that it will be maintained in the next years. In addition, this DBMS is well known by the AGILE Team because it is used in other activities like for the database described in Sec. 2.2. The complete DM consists of 43 tables with 57 relationships between them.

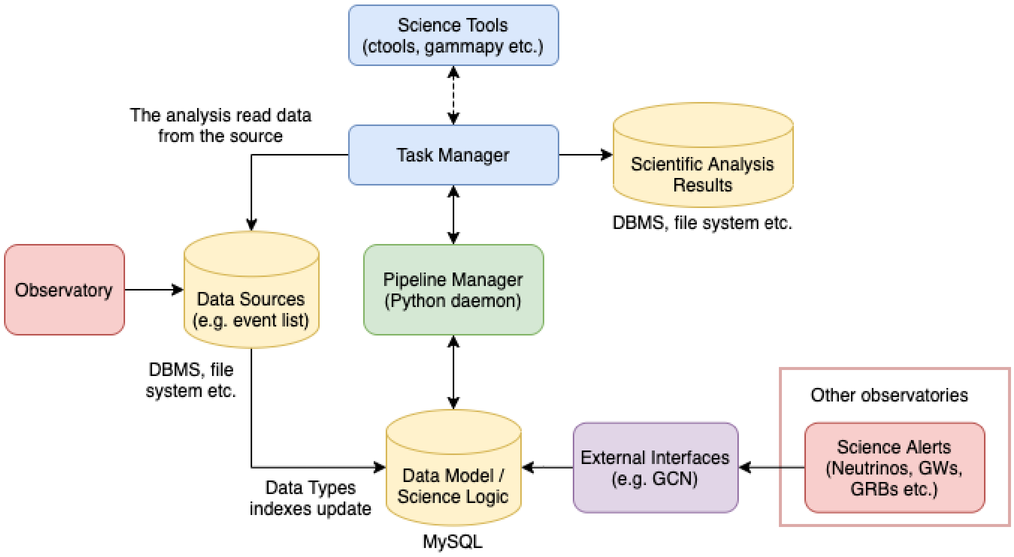

Figure 3.1 High-level schema of the RTApipe framework software architecture. The schema shows all the RTApipe framework components and external entities that are interfaced with the framework. It also represents the storage systems used by the software.

The DM static entities are:

- Instrument: it contains all the observatory instruments producing data or instruments that send science alerts to the GCN network.
- Data Type: all types of data managed by the pipeline and required for analyses execution.
- Data Source: it contains the sources that produce Data Types(e.g. gamma-ray detectors).
- Science Tools: the software packages used for the analyses (e.g. how to execute the software, the input requested etc.).
- Analysis Type: the configurations that define the usage of the Science Tools (e.g. creating sky maps with a time window of 100 sec).

The DM dynamic entities are:

- Observation: This entity is used to store observation's information (e.g. pointing coordinates, start time, end time etc.).

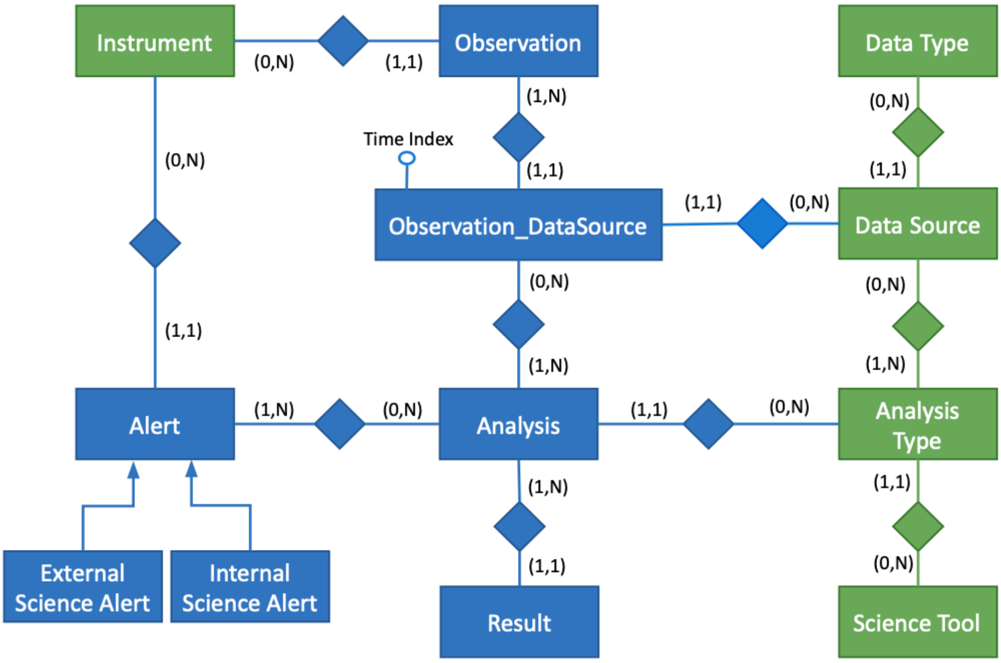

Figure 3.2 High-level schema of the Data Model. Static entities and relationships are green, while dynamic entities and relationships are blue.

- Observation_DataSource: This entity implements the N, N relationship between Observations and Data Sources. For each observation, many Data Sources can generate data. Each Data Source can generate data during many Observations.
- Analysis: this entity stores the analyses parameters.
- Results: this entity contains the references for the results stored in external storage systems (e.g. file system paths).
- Alert: this table contains all science alerts, received by the GCN network (external) or generated by an observatory instrument (internal).

The developers must configure all the DM's static information before starting the data acquisition and the real-time analysis with the pipeline. Despite the very complex astrophysical context, the DM is easy to understand. The pipeline's components update the DM's dynamic part in real-time during the operations. The dynamic part includes the indexes of the various Data Sources, the analyses performed by the pipeline, the status of these analyses, any science alerts received, and other auxiliary information. The Science Logic algorithms described in detail in the following section update most of this dynamic part of the DM.

Case Study

The AGILE case study can be used to understand how the developers have to configure this DM. AGILE has several instruments, and they can be configured into the DM's Instrument table. Each instrument can generate one or more Data Sources.

Now AGILE operates in spinning-mode so it can not point a specific target; for this reason, only one continuous observation is considered for each instrument.

In the case of a pointing instrument, the system would create a new observation in the Observation entity for each new pointing. When new data are received from an instrument (e.g. at each contact of the satellite with the Ground Station), the system updates the Observation-DataSource table of the DM with the data index for each Data Source in relationship with that specific observation. At this point, the pipeline knows that there are new data to analyse. Another scenario occurs when external or internal science alerts are received or generated. In this case, the information about these alerts is inserted into the Alert entity, and pipeline starts new analyses.

AGILE can generate internal science alert and receive external science alerts from External Interfaces like the GCN network. The External Interfaces are described more in detail in Sect. 3.4.8. All the analyses performed by the pipeline are in a relationship with a particular Science Tools. The analyses can be configured to use the same Science Tools in different ways and with different parameters. The AGILE pipeline has twenty Science Tools analysing the different data received from the AGILE instruments. The Sec. 3.4.3 describes how the analysis configuration is used by the pipeline to analyse data.

Finally, the analysis results can be stored in the Result entity of the DM. This storage can be done directly or by saving the information to retrieve results into another storage system.

3.4.2 Data Sources

A Data Source is a component that generates one or more Data Types. These data types are the input for the analysis of the scientific pipeline. The pipelines developed with the RTApipe framework can be interfaced with one or more Data Sources.

The framework does not specify where data should be stored but requires an updated index for each Data Source saved into the ObservationDataSource entity of the DM described in the Sec. 3.4.1. A good example of Data Source can be the event list generated from an instrument and stored into a storage system. In the case of an observatory, the event list is the list of physical events with their parameters. The physical events are the phenomena detected by the instrument (e.g. gamma-ray photon). The event list data are not stored into the pipeline database but in a separated storage system. The pipeline database requires only an updated index for the data sources. This index can be the time of the last event detected by the instrument.

The pipeline knows when new data arrives into the system because the Data Source updates the index into the pipeline database, and it also knows the time window of these data on which it can start the analysis. The update of this index can activate the MySQL triggers that are used to implement the Science Logic described in Sec. 3.4.3. The pipeline can manage different Data Types from different Data Sources.

The data can be stored in many storage systems: files, databases, cloud storage, and more. For this reason, Big Data sources are not a problem for the pipeline because it will only see the updated index and not the entire data content. Big Data will be saved in dedicated service, and the pipeline needs the update of the index value inside the database when new data arrives from the Data Sources.

Case Study

The real-time scientific analysis pipeline developed for AGILE can perform analysis for all the instruments onboard the satellite. This pipeline manages eight different Data Sources, each with its Data Type. Each Data Source has its preprocessing software; for this reason, not all data inside the same contact are received at the same moment into the scientific pipeline.

As described in Sec. 3.4.3, the RTApipe framework can set up one or more Data Types required for each analysis, and the pipeline will wait until all the necessary data are available. When a Data Source updates the index into the pipeline database, a MySQL trigger is activated and starts new analysis on the new data. For this reason, the analysis starts automatically when the Data Source update the index into the database pipeline, while the full data is stored in another storage system.

3.4.3 Science Logic

The Science Logic (SL) defines the set of rules used by the pipeline to know *When* and *How* execute the configured analysis during operations. All the rules to manage the most frequent use cases in the MW and MM context are implemented in the RTApipe framework. If required, the users can add or modify the existing rules according to their own needs.

The Science Logic rules are implemented with 12 MySQL triggers. In this way, the Science Logic is implemented in the same DM's service, and MySQL can execute Science Logic algorithms with high performances directly on the DM when triggers start. The triggers can be used to implement the desired analysis workflows.

This component interacts with most of the dynamic part of the Data Model. It manages pipeline behaviour during the operations without requiring human intervention. This set of rules, defined in advance, allows the pipeline to operate in real-time autonomously. These rules are developed to perform new analyses when one of the following two scenarios occurs:

- The Data Source updates the index of a specific Data Type configured for one or more analyses. This scenario is described in Fig. 3.3.
- A new science alert is received by the pipeline and inserted into the DM. This scenario is described in Fig. 3.4.

The first scenario occurs when the data source updates the index of one or more Data Types into the pipeline database communicating that new data arrived. The user can configure a minimum time window of data required to start the analyses (e.g. 100 sec, 1000 sec etc.). If no time window is configured, the pipeline performs the analyses at each index update. The pipeline trigger checks if there are analysis configured into the DM that must be executed for the Data Type arrived. If yes, the trigger prepares these analyses and sets the state to 'runnable'. If the new data does not cover all the time window requested, the pipeline waits for new data. Once in 'runnable' state, the analyses are ready to be executed from the Pipeline Manager component described in Sec. 3.4.4.

The second scenario occurs when the pipeline receives an internal science alert or an external science alert from other observatories through the External Interfaces defined in Sec. 3.4.8. This last scenario is implemented to search for a science alert counterpart inside the incoming or archive data. In this scenario, the pipeline trigger checks if there are analyses configured for the science alert received. If yes, it creates the analyses with a 'pending for data' state. The trigger checks if the data archive contains the required time

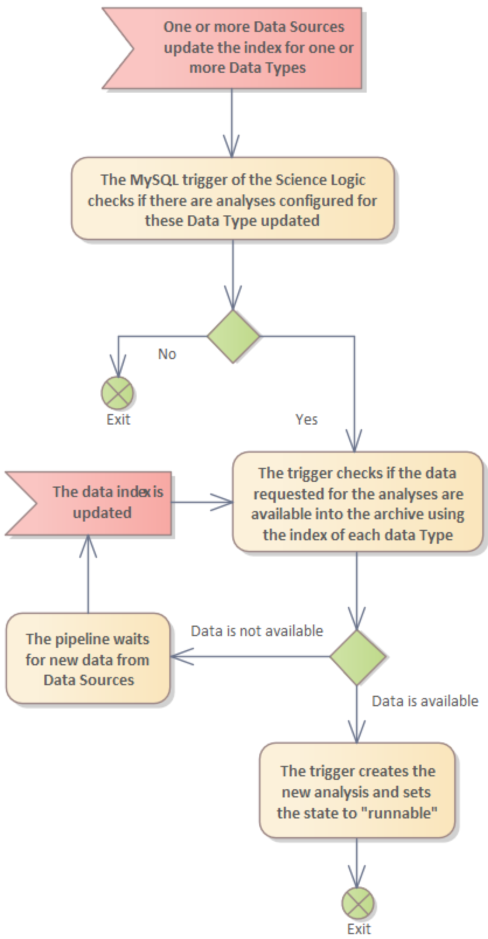

Figure 3.3 Activity diagram of the Science Logic workflow when one or more Data Sources update the index of one or more Data Types.

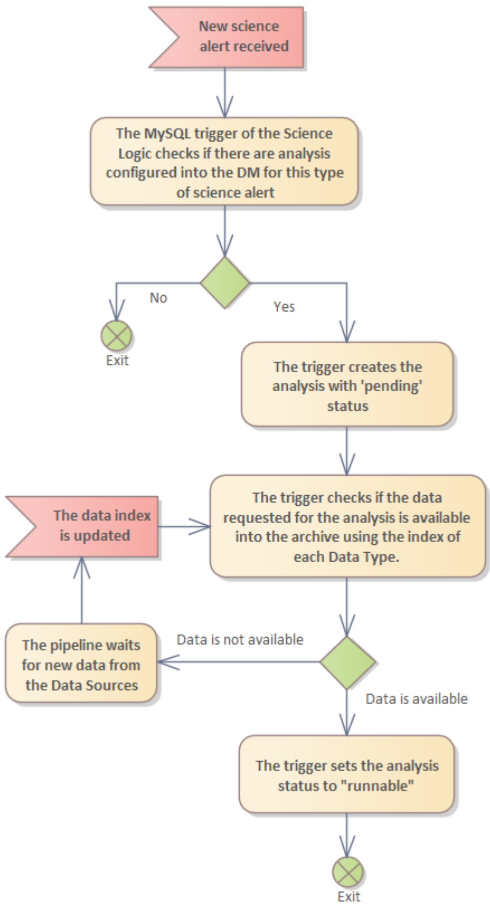

Figure 3.4 Activity diagram of the Science Logic workflow when a new science alert (internal or external) is received.

window for the analyses and if possible sets the analyses status to 'runnable'; otherwise, the pipeline waits until the required data arrives, and then starts the suspended analyses. In both the scenarios, the pipeline creates the analysis starting from the configuration defined by the user in the AnalysisType entity.

The main rules that can be configured by the user to perform the desired analyses are:

1. It is possible to configure a group of Data Types required to start a

specific analysis and the trigger check for the data availability or wait until the needed data is available.

2. The analysis can be configured with an integral time window. The time windows start at the begin of the observation and integrate all data to the last data received. This functionality can be useful to accumulate data and statistics since the start of the observation.

3. It is possible to configure a flexible time window without predefined size that starts with the end time of the last analysis time window.

4. The analyses can have a single fixed time window or a large time window divided in various bins (e.g. perform analysis from -1000 sec to $+1000$ sec with bins of 100 sec).

5. In case of a science alert it is possible to configure a list of analyses with a predefined time windows centred in the physical event's time communicated within the science alert (e.g. $\pm 100 sec$).

6. It is possible to configure a list of analyses that are executed only when a science alert is received from a particular observatory (e.g. from IceCube).

7. It is possible to configure a partial analysis that starts if the data coverage is not complete. Then when the entire data arrives, another analysis is performed with full coverage; if the partial analysis is still running, it is stopped to release resources for the full coverage analysis.

8. It is possible to configure the sky position coordinates used for the analyses. This position can be the position of a known gamma-ray source, the telescope FOV, or a list of coordinates defined with a grid of points (e.g. HEALPix [14]).

9. It is possible to define an analysis that is executed when a list of linked analyses ends without error. This can be useful to import results on a database or to merge results from different analyses.

10. It is possible to define the priority between different analysis types and the queues of the scheduler where the analyses must be submitted. High-priority analysis can suspend low-priority analysis.

11. The analyses and trigger can be disabled or enabled just changing a flag into the DM.

[14] https://lambda.gsfc.nasa.gov/toolbox/tb_pixelcoords.cfm

The pipeline developer has to configure these parameters and analysis types. After this initial configuration, all these triggers manages the analysis automatically during the operations.

Case Study

The pipelines developed for AGILE implement both the scenarios. The first scenario (Data Types indexes updated) is implemented configuring several analyses with the different Data Types acquired by the instruments onboard AGILE. The AGILE/MCAL pipeline (Sec. 3.6.2) waits that the index of a list of Data Types is updated and then executes a list of analyses. These analyses' time windows are flexible because the AGILE/MCAL telemetry data can have different time coverage. The AGILE ratemeters pipeline has the same behaviour.

The AGILE real-time analysis pipeline that reacts to science alerts uses almost all the rules described before. When the pipeline receives a new science alert by the GCN network, it executes a list of analyses configured for that type of science alert (e.g. GW science alert). Each type of science alert has a list of configured analyses executed when the science alert is received. These analyses are performed on the AGILE instrument's data (AGILE/MCAL, AGILE/GRID and ratemeters).

The time windows for the analyses is fixed and centred with the event time defined into the science alert. If there is no full data coverage for all the Data Types required for the analysis, a partial analysis is executed. When new data arrives and the requested time windows are covered, a new analysis is started terminating the partial one if it is still running. The partial analysis aims to process as soon as possible the data available because the transient event can be inside these data, so there is no reason to wait for the next contact. If the time window is not fully covered for just 10 seconds, but the transient event is still detectable, the partial analysis can detect it. All these automatism can be configured using the Science Logic rules.

Usually, the localisation error region received with the science alert is small, and the pipeline can execute the analyses in that sky region.

In the case of GW science alert, the analysis uses a grid of coordinates based on the HEALPix system to define the analysis sky position because this kind of science alerts have a large sky region where the gamma-ray source can be detected. Hence, the pipeline performs different analyses in different positions. The results are then merged by a process that starts if all the analyses end without error.

3.4.4 Pipeline Manager

The main task of this component is to execute the analyses generated by the Science Logic rules on the available computational resources. The Pipeline Manager can manage several pipelines (e.g. one for each different instrument onboard a satellite). Each pipeline has its database containing the Data Model, and the Pipeline Manager reads the connection information to this database from a configuration file. This component performs a simple task but must guarantee high performance to avoid delays. In the case of hundreds of analyses per second, this component could become a bottleneck for the system so that the implementation technologies may vary based on the pipeline requirements.

The Pipeline Manager is a software written in Python that periodically queries the database to check if the triggers have created new analyses in 'running' state and are ready to be executed. When the Pipeline Manager finds new analyses, it creates a new directory in the file system. Then it queries the configuration database to obtain all the information necessary to execute these analyses (time window, Science Tools, and more) and saves this information in a configuration XML file. The Fig. 3.5 shows an XML file created by the Pipeline Manager. This XML file is parsed to read all parameters by a Python wrapper that is the interface between the Science Tools and the pipeline during the analysis execution. One template for each Science Tools used by the pipeline is configured in the DM so that the Pipeline Manager shall produce it at run-time. The XML file's red section contains dynamical parameters with #@@# tags that are defined by the Pipeline Manager for each run. At the same time, the blue part can be customised by the developer of the pipeline and can contain fixed parameters for a specific Science Tool. After preparing all the configuration files necessary to interface the pipeline with the Science Tools, the Pipeline Manager submits the analysis to the Task Manager, described in Sec. 3.4.6.

Case Study

AGILE has different real-time analysis pipelines. There is one pipeline that reacts to external or internal science alerts. The other pipelines execute analysis when new data arrives from the satellite to search for transient events. With the RTApipe framework, it is straightforward to set up different pipelines and manage them. It is possible to stop a pipeline temporarily, just changing a flag inside the configuration file. During this PhD, new pipelines were added without the need to modify or halt running pipelines. The system is securely split into different components (one Pipeline Manager and several

```xml
<run id="#@runid@#">
  <parameter name="RegionOfInterest" ra="#@ra@#" dec="#@dec@#" skypositiontype="#@skypositiontype@#"
  <parameter name="TimeIntervals" tmin="#@tstart@#" tmax="#@tstop@#" timeunit="#@timerefunit@#" times
  <parameter name="ScienceToolReference" timestart="51544.5" timeunit="day" timesys="mjd" timeref="LO
  <parameter name="Energy" emin="#@emin@#" emax="#@emax@#" energyBinID="" />
  <parameter name="DirectoryList" run="#@diroutputrun@#" results="#@diroutputresults@#" runprefix="#@
  <parameter name="WorkInMemory" value="0" />
  <parameter name="MakeCtsMap" value="1" />
  <parameter name="CountsMap" enumbins="1" nxpix="" nypix="" binsz="0.02" coordsys="CEL" proj="CAR" e
  <parameter name="HypothesisGenerator3GH" value="0" />
    <parameter name="CtsMapToPng" value="1" />
  <parameter name="DeleteRun" value="0" />
  <parameter name="Binned" value="0" />
</run>
```

Figure 3.5 Example of XML configuration file created by the Pipeline Manager. The XML file is divided into two sections. The red section is dynamically generated at run-time while the blue section is static and configured before the operations.

pipelines DMs) to avoid this and other problems.

3.4.5 Task Manager

The Pipeline Manager executes the analyses prepared by the Science Logic. Since the number of these analyses can be potentially very high, the pipeline manager submits these processes to a Task Manager component which executes them in parallel optimising the computational resources. The Task Manager is a job scheduler that executes analyses following the priority between them defined by the user into the DM. To scale the system within a cluster, the Task Manager performs analyses on multiple computing nodes and coordinates them. The Task Manager can temporarily suspend low-priority running analyses to execute higher priority analyses (e.g. when an external alert is received).

The Task Manager is implemented using Slurm [15], an open-source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters. The Pipeline Manager submits all the analysis created by the Science Logic to Slurm, Slurm manages the resources and the execution of the analysis. Slurm can manage thousands of jobs and can be easily scaled from one machine to thousands of servers. It can execute jobs in parallel and manage different priorities between jobs or working queues. Another key feature of Slurm is the capability to log inside a

[15] https://slurm.schedmd.com

database all the information about every process. This logging system is useful to monitor the pipeline in real-time, viewing statistics about the execution of tasks. Scaling the computing power with Slurm does not require to change the Pipeline Manager software. Within the DM, it is possible to configure the priorities between jobs, which will then be respected by Slurm, and reserve resources for a specific type of analysis.

Case Study

As already said, AGILE has different pipelines executing scientific analysis as soon as the data or science alerts are received. For this reason, the Task Manager component provided by the RTApipe framework is necessary to manage all the analyses and run them in parallel. When a science alert is received, the Task Manager automatically suspends all the low-priority analyses from other pipelines and starts high-priority analyses triggered by the alert. Then when the high priority analyses ended, the low priority analyses are restarted. This automatic behaviour is critical to ensure a fast reaction to science alert optimising the computational resources.

3.4.6 Science Tools

The Science Tools are software developed to perform scientific analysis. The RTApipe framework can manage different Science Tools, and it is possible to analyse the same data with different parameters or configurations. Usually, the Science Tools are updated during the observatory lifetime by different development teams. Thus, it is important to have the flexibility to change the Science Tools inside the pipeline. The Science Tools are configured within the DM. The pipeline automatically performs the analyses configured with the required Science Tools. As described in Sec. 3.4.4, the DM contains templates of the configuration files for each Science Tool enabled to work with the pipeline. This template includes all the parameters needed by the Science Tools to perform the analysis. A part of these parameters is fixed and can be modified by the developer before the pipeline starts processing data. The second part of the settings is dynamically managed by the Pipeline Manager when it creates the configurations for the analysis.

Case Study

The AGILE Team developed several pipelines specialised on different instruments. Usually, there is a team of developers for each pipeline and each instrument. Specific Science Tools are developed for scientific analysis on a particular instrument data format. The RTApipe framework allows the management of different Science Tools and the smooth integration of new Science Tools that will be developed during the observatory lifetime. One of these Science Tools that can be implemented into this framework is the CNN, described in Sec. 4, developed during this PhD to detect GRBs inside the AGILE gamma-ray sky maps. The AGILE Team releases new Science Tools updates when needed, and deploys these changes into the pipelines. The RTApipe framework features used to build these pipelines support this update without extended downtime or significant configuration changes. The upgrade of one pipeline does not affect other pipelines that can run without stopping.

3.4.7 Analysis Results

The RTApipe framework does not require the direct management of the analysis results. The analyses performed by the Task Manager can store the results in a database or into simple ASCII files. This flexibility is essential to analyse several types and formats of data. Each Science Tool can save its results in different storage systems: database, file system, cloud services, and more.

Case Study

The pipelines developed for AGILE save the results in different storage systems. The pipelines DMs contain only the references to the storage systems, allowing the researchers to link the results with the DMs entities. Part of the results is stored in a MySQL relational database different from that used as the pipeline DM. The remaining results are stored directly into the file system. The results can be visualised with plots and tables shown by a web GUI that queries the database or the file system.

3.4.8 External Interfaces

Not all scientific pipeline components are integrated inside the RTApipe framework to provide higher flexibility for the developers. For this reason, the framework must be integrated with External Interfaces. One of the External Interfaces must inform the pipeline when an observation begins or ends. The Data Sources can be considered External Interfaces because they are not provided within the framework, but they need to update the data index inside the pipeline database. In the MW and MM context, it is crucial to have an interface with other observatories to receive external science alerts.

The RTApipe does not require a specific system to receive alerts, and the developers can decide which one to use and then configure it (e.g. connect the system with the GCN network).

Case Study

The AGILE Team decided to integrate the pipeline with the GCN network to receive all the external science alerts that can have a counterpart inside the AGILE instruments' data. The AGILE Team goal is to react rapidly to external science alerts searching for a transient counterpart event. If a significant detection is found by the AGILE/MCAL pipeline (Sec. 3.6.2), the pipeline send an automated Notice to the GCN network. All the pipelines developed for AGILE have an interface with the preprocessing system (described in Sect. 3.6). The pipelines receive updates for the data index of each Data Source from this External Interface.

3.5 Framework key features

The RTApipe is a framework designed to help the development of real-time science analysis pipelines for both ground-based and space gamma-ray observatories. This framework satisfies the MW and MM context's high demanding requirements described in Sec. 3.3 with a list of key features described in this section.

Parallel analysis

With this framework, it is possible to develop pipelines capable of performing several analyses in parallel. The only constraints are the available computational resources. The user must configure the various types of analyses to be performed in the configuration database. Then the pipeline will automatically execute all the configured analyses in parallel. It is possible to use the same Science Tool to perform analyses with different parameters or use different Science Tools to run the same analysis type. Parallel analyses can concern the same input data stream or multiple input data streams. Consider a telescopes array divided into sub-arrays, each with a different pointing in the sky. The framework can manage a list of analyses performed in parallel for each sub-array. The developers can use the RTApipe framework to develop a pipeline that can handle this workflow complexity.

Flexibility

The flexibility allows the pipelines to be developed with this framework to support many updates quickly and straightforwardly. Pipelines can receive several input data streams and perform analyses with many Science Tools, which can be configured in just a few steps. The core of the framework is isolated from all those components that can vary from observatory to observatory to achieve a high flexibility level. The framework does not ask developers to use a data format or a specific input or output storage system. The pipeline can work with any Science Tools and allows the user to configure new Science Tools when needed.

Scalability

Slurm is the job scheduler of this framework. Slurm can be configured in a single server or on thousands of servers by modifying a few lines in the Slurm configuration file. Increasing computing power does not need to change the DM or the Science Logic. If Slurm is configured in a cluster consisting of multiple nodes, it can also manage the nodes' failure, keeping the pipeline operational.

Task Priority

The RTApipe framework allows the configuration of execution priority between different types of analysis. If the computing resources are not enough to run all analyses, it is possible to set up priority between them to execute first the analyses with the most priority results. This feature is important to start high-priority jobs, suspending low priority jobs (e.g. when a Scientific Alert is received). It is also possible to configure job queues to manage the priorities within a queue (e.g. a queue for each Data Source).

Pipeline Monitoring

The Task Manager logs in a database all the information related to the execution of the analyses. Therefore, it is possible to perform queries on the database to obtain real-time monitoring or statistical analysis. It is possible to calculate the average time needed to perform a specific analysis or get the list of analyses ended with error to search for bugs. Since this information is updated in real-time, it is also possible to create a GUI that periodically queries the database and shows the analyses progress.

Containerization and easy setup

The RTApipe framework is deployed within a Singularity[16] container. All the services needed by the pipelines (MySQL, Apache, Slurm, Anaconda, Python, and more) are installed and configured inside the container. This system allows a quick and straightforward setup on different types of hardware and Operating Systems. The use of containers is also necessary to integrate the pipeline into continuous integration and delivery workflow.

Different layers are created to build the Singularity container, each with a specific purpose. This approach optimises the build process of the container. All the system configuration and packages are installed in the first layers, while the next layer install Services and Science Tools. With this architecture, it is possible to add software in the last layer without rebuild all the container but starting from previous layers and reduce building time. The container built with the last layer can be moved to one or more servers where the pipeline must run.

3.6 AGILE pipelines

During the PhD, the RTApipe framework was used to develop three AGILE real-time analysis pipelines. One of these pipelines react to internal or external science alerts (e.g. GRBs and GWs) and is described in Sec. 3.6.1. The other two pipelines perform a blind search of transient events (mainly GRBs) into the AGILE/MCAL instruments' data (Sec. 3.6.2) and the ratemeters of all detectors onboard AGILE (Sec. 3.6.3). The ratemeters are the raw information about the acquisition event rate over the time (e.g. the number of photons or particles detected per second) of all the instruments onboard AGILE and can be used to identify transient events. When a transient event is detected, it increases the acquisition rate of AGILE instruments.

These practical examples of pipelines developed using the RTApipe framework demonstrate this framework's flexibility in different contexts. In the following sections, the implementation of each of these pipelines and the use cases are described in detail.

The pipeline software packages are installed into a Singularity container following the procedure described in [Par+20]. An efficient backup system is implemented by deploying the Singularity containers on more than one server and geographically distributing the backups. These servers are configured with the procedure described in [PB20a]. The Fig. 3.6 shows the AGILE real-time analysis pipelines' deployment diagram. The pipelines are deployed

[16] https://sylabs.io/docs/

in two different data centre (ASI/SSDC Roma and INAF/OAS Bologna) to have a backup system always available for failures recovery.

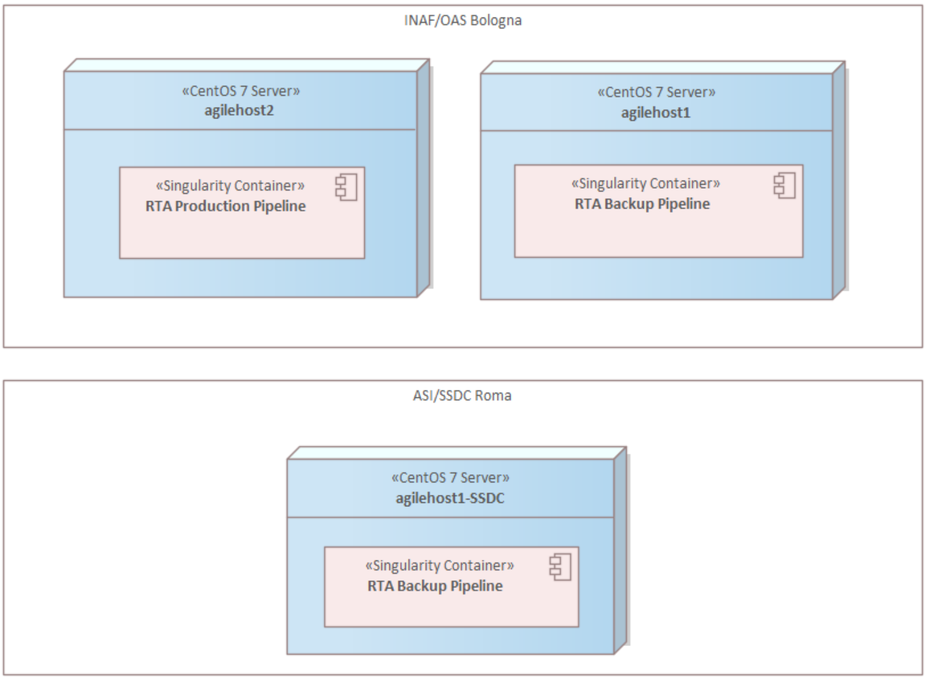

Figure 3.6 The deployment diagram of the AGILE real-time analysis Singularity containers to have a geographically distributed backup. Two containers are deployed at INAF/OAS Bologna while the third container is deployed at ASI/SSDC Roma.

The AGILE real-time analysis pipelines receive external science alerts from other observatories through the GCN Network (with Notice format) or internal science alert from the AGILE/MCAL pipeline (Sec. 3.6.2).

The pipelines are part of the AGILE space mission real time analysis context shown in Fig. 3.7 (taken from [Bul19]) that starts with the preprocessing and reconstruction pipelines executed at ASI/SSDC and ends with the detection of transient events that are shared through science alerts to the GCN network or email to the AGILE Team.

3.6.1 Science Alert Pipeline

This pipeline is designed to respond to internal or external alerts and analyse the data received from multiple instruments onboard the AGILE satellite. The internal alert can be received by other AGILE pipelines when a transient event is detected (Sec. 3.6.2, 3.6.3). The pipeline reacts to internal

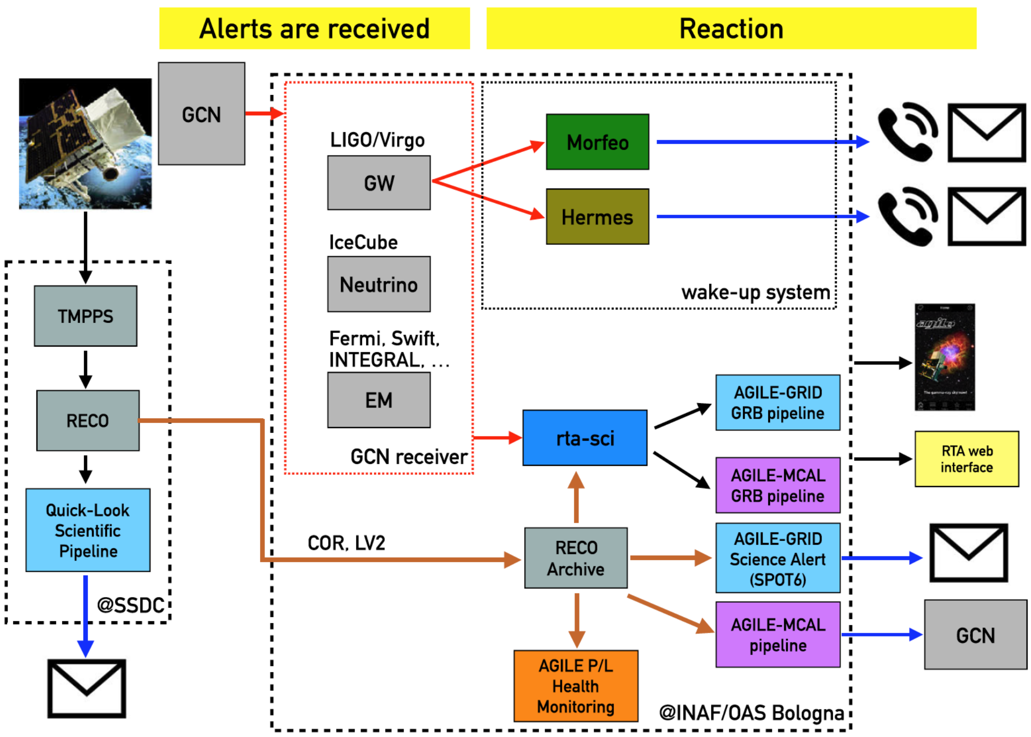

Figure 3.7 The second version of the AGILE Real-Time Analysis workflow at INAF/OAS Bologna, with the reconstructed data received from SSDC. Science alerts are received from the GCN network. Two independent wakeup systems call the AGILE Team if a new alert from LIGO/Virgo collaboration is received. If new alerts and or data are received, the real-time analysis pipeline performs the required scientific analysis. The GRID, MCAL and SA GRB alert pipeline react to external science alerts. The AGILE-GRID Science Alert System (SPOT6) and the AGILE/MCAL pipelines process new data as soon as it is received from SSDC, which performs the data's preprocessing and reconstruction. In the presence of a GRB identified by MCAL pipeline, an automated GCN notice is submitted to the GCN network. A web GUI and the AGILEScience App show the results of the automated processing. A backup of the AGILE RTA is also running in SSDC. This figure is taken from [Bul19].

science alerts to search for a transient counterpart in other instruments' data onboard AGILE (e.g. a science alert received by AGILE/MCAL triggers analysis on AGILE/GRID data). The external science alerts are received from other observatories through the GCN network with Notice format. The pipeline can receive external science alerts from several observatories, including FERMI, LIGO/Virgo Collaboration (LVC), and IceCube (described in Sec. 3.1).

It is possible to configure the pipeline to execute a different list of analyses for each science alerts types (e.g. GWs or GRBs) or instruments that sent the alert (e.g. IceCube or Fermi). Simultaneously the pipeline alerts the AGILE Team and the Astronomer on-duty to follow-up the science alerts with these analyses' results. Suppose the pipeline detects a transient event with a compatible trigger time and position with the science alert and with a statistical significance over a threshold. In that case, a science alert is sent automatically to the GCN network to inform other observatories.

When an external science alert is received, the AGILE real-time analysis pipeline performs more than 100 analyses with 20 different Science Tools that use the AGILE-GRID, AGILE-MCAL, and ratemeters data. Slurm manages the execution of these analyses in parallel to obtain the results as soon as possible. If other low-priority jobs are running, Slurm suspends them and executes the high-priority jobs related to the science alert. This pipeline's performances satisfy the requirement of AGILE to follow-up external science alerts and produce quick results for the researchers. In a few seconds or minutes (depending on the task) since the arrival of the alert, researchers can visualise the results using the web GUI.

Science Tools are developed with different programming languages, but it is possible to integrate them into this framework with small effort. Often the AGILE Team creates new Science Tools or updates the existing ones, and with this framework, it is effortless to add and configure new types of analysis.

The Fig. 3.8 shows the workflow of this pipeline. The internal/external science alerts are received and managed by the Pipeline Manager. The Pipeline Manager follows the Science Logic described in Sec. 3.4.3 and trigger automatic analyses to search for a transient counterpart within the AGILE data. If the requested data are available into the archives, the analyses are executed. Otherwise, the Pipeline Manager waits until new data is received. The AGILE/GRID analyses use the LOG and EVT input data. The AGILE/MCAL analyses use the COR 3908/3916. The ratemeters analyses use the COR 3913, and the Auxiliary analyses use the AUX data input.

All these data types are sent to the INAF/OAS Bologna from the Science Operation Center (SOC) at ASI/SSDC Rome after the reconstruction processes. This data flow is described in Sec. 1.2.

These analyses produce results stored in a MySQL database and in the server's file system where the analyses are executed. The Astronomer on-duty can visualise these results through the web GUI described in Sec. 3.6.1

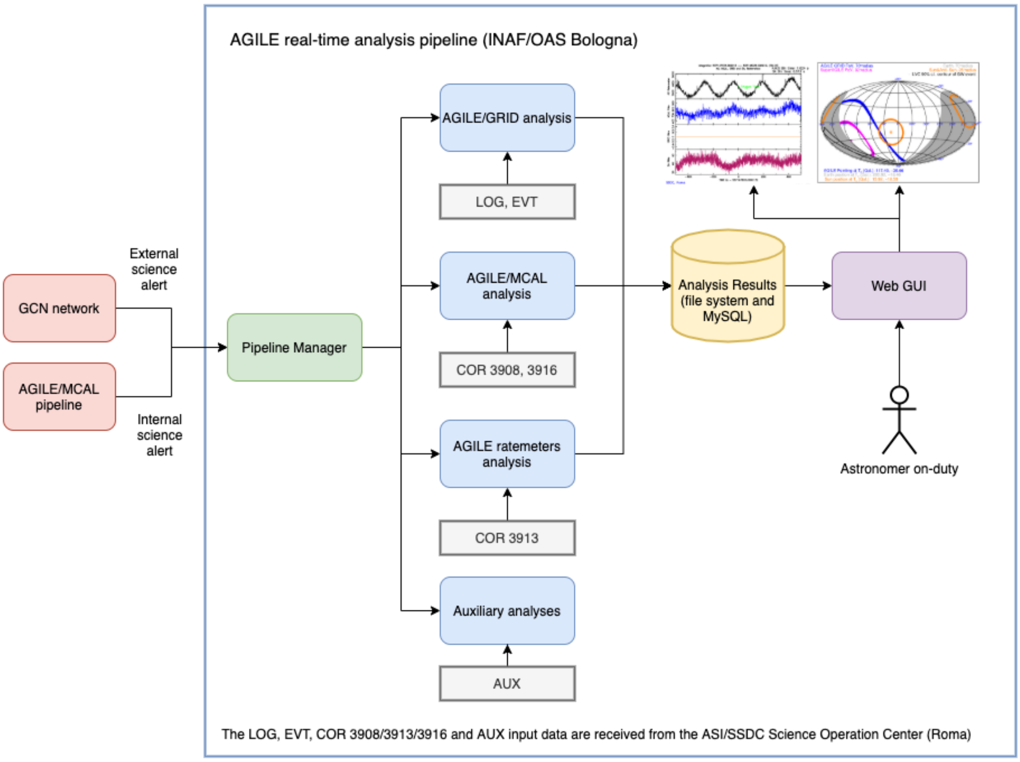

Figure 3.8 Workflow of the AGILE real-time analysis pipeline that reacts to science alerts. This pipeline executes the analysis when an internal or external alert is received processing AGILE/MCAL, AGILE/GRID and ratemeters data. The results are stored in a MySQL database and in the file system. The Astronomer on-duty can visualise these results using the web GUIs.

and follow-up the science alerts.

AGILE Control Room

This web GUI, developed during the PhD, is not part of the RTApipe framework but is essential to allow AGILE researchers to access information related to the satellite and data acquisition status and the real-time analysis pipelines results. To access the Control Room, the user must connect to a public website and log in with the username and password provided only to the AGILE Team members. The connection address does not change, but the system administrators can connect it to different servers in case of failures. This web GUI is based on a previous version of the web GUI developed by AGILE researchers and described in [Zol+19]. The technology used to implement this web portal are described in Sec. 2.3.

The Control Room's home page shows several plots and tables that can be used to check the data flow status. The Fig. 3.9 shows the status of the data flow arriving from the ASI/SSDC data centre to the INAF/OAS Bologna data centre. This figure shows the delay (hours) accumulated since the last data arrived. If there is a problem during the data transfer, the delay starts to grow up, indicating to the Astronomer on-duty that there is a system failure. This plot track two Data Types (LOG and COR 3916) that are used to check the data coverage of the AGILE/GRID and AGILE/MCAL instruments.

This web page also shows the status of different archives where the acquired data is stored. If a science alert is received, the Astronomer on-duty can use the Control Room to check if there is data coverage on the alert time or if it is required the new contact data. It is possible to visualise information about past and next contact, when they will arrive and the time windows of their data coverage. It is possible to view the name of the Astronomer on-duty that is in charge to perform the follow-up and manual analysis when a science alert is received from other instruments.

A section of the web page shows preliminary results from the AGILE/GRID and AGILE/MCAL pipelines. The gamma-ray sky map generated by the SPOT6 pipeline [Bul+14] analysing the last contact AGILE/GRID data.

A table is shown to visualise the AGILE/MCAL pipeline results for the last five contacts. This table contains a summary of any GRB and Terrestrial Gamma-ray Flashes (TGF) detected by these analyses. TGFs [Mar+14] are intense and brief emissions of gamma-rays associated with thunderstorm activity and lightening. They are largely detected from space mission designed for astrophysical purposes. These events are described as millisecond time scale bursts of gamma rays with the incoming direction compatible with the Earth and a spectral hardness typically much higher than that of cosmic gamma-ray bursts. AGILE can detect these physical phenomena with the MCAL instrument [Mar+10].

The last section of the Control Room shows the real-time analysis results related to the last science alert received. If the users need more information, they can open a detail page on that science alert.

The Control Room's home page contains information that the Astronomer on-duty can use to check the ongoing operations and preliminary results from all the AGILE pipelines.

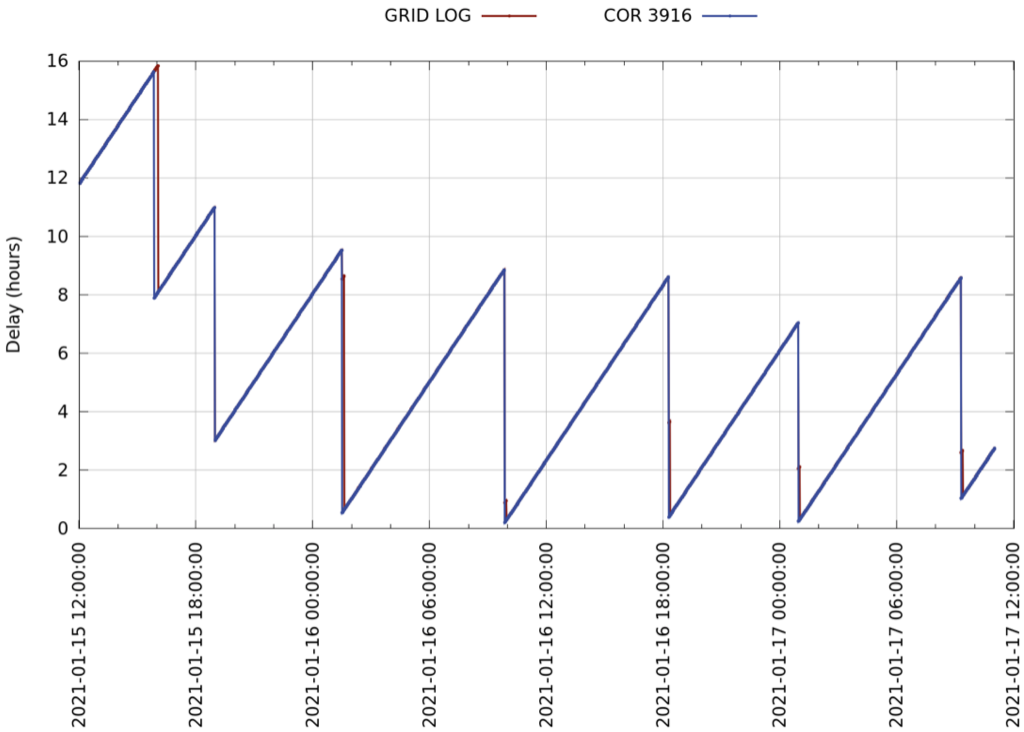

Figure 3.9 The status of the data flow arriving from the ASI/SSDC data centre to the INAF/OAS Bologna data centre. The plot shows the delay (hours) accumulated since the last data arrived. If there is a problem during the data transfer, the delay starts to grow up over the average value indicating to the Astronomer on-duty that there is a system failure. This plot track two Data Types (LOG and COR 3916) that are used to check the data coverage of the AGILE/GRID and AGILE/MCAL instruments.

RTA Analysis Results

The Control Room contains specific web pages to visualise a table showing the list of the science alerts received and processed by the real-time analysis pipeline. From this table, the user can enter into the detail page of a specific science alerts to visualise related analyses results performed by the real-time analysis pipeline on the AGILE data. It is possible to visualise the table with all the science alerts received or filter for a specific instrument that sends internal or external alerts (e.g. Fermi/LAT, Fermi/GBM, LIGO, IceCube, AGILE/MCAL).

The real-time analysis pipeline performs two main types of analyses when a new science alert is received:

- Prompt Analysis: this analysis is performed with data of all the instruments onboard AGILE using a time window centred in the transient event time. These results are important to check if a counterpart detection is found near the transient event time notified by the science alert and in the same sky position.
- Full Analysis: this analysis allows the Astronomer on-duty to visualise the results obtained in a larger time window (±1000 seconds). Considering that the satellite continually spins around its axis and the event time may occur when the AGILE instruments have no exposure in the event position it is possible that AGILE can observe the sky position after many seconds.

Prompt Analysis Results This web page shows all the results of the analyses performed with a time interval close to the event time noticed with the science alert. The first part of the web page shows sky maps useful to identify the sky region where the event occurred and the visibility that the different AGILE instruments have on that area. A table with the archive status of different Data Types is shown allowing the user to understand if there is the data coverage for the event time or whether the pipelines must wait for the next contact.

The second block shown in Fig. 3.10 contains the results of the data analysis of the AGILE/MCAL instrument with different time scales (±5 sec, ±10 sec, ±100 sec, ±1000 sec) to rapidly check if there is a burst detected with this instrument. The Astronomer on-duty can perform detailed manual analysis if a burst is detected into these plots.

The third block of this web page (Fig. 3.11) shows plots related to the results of the analyses performed on the AGILE/GRID instrument's data. The analyses are performed on different integration intervals (±2 sec, ±5 sec, ±10 sec, ±100 sec).

The last block of diagrams illustrated in Fig. 3.12 is related to the analyses performed on the different AGILE instruments' ratemeters. These diagrams allow the user to have a complete overview of all instruments. They can be useful to detect bursts in the instruments' ratemeters even when the high-level analyses can not detect any burst.

The pipeline prepares two templates of the texts for the GCN Circulars. The Astronomer on-duty, who is in charge of the alert's follow-up, can use these prepared Circulars, check them, and eventually modify them with more precise manual analyses and then send the Circulars to the GCN network.

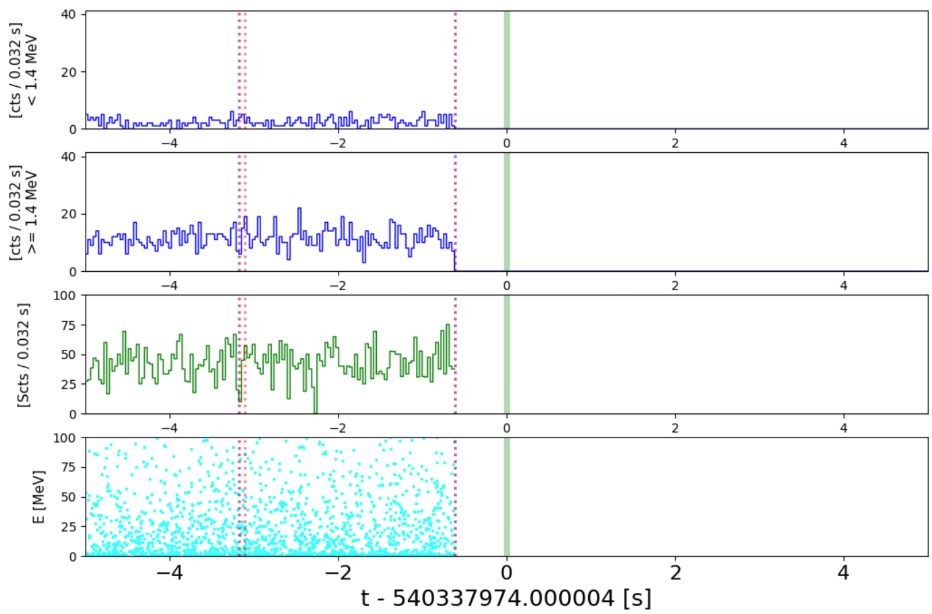

Figure 3.10 Plot containing the results of the AGILE/MCAL analyses centred in the time sent with the science alert. This plot shows the photon counts detected by AGILE/MCAL at different energy ranges in function of the time.

These templates are prepared with results of both MCAL and GRID instruments data.

Full Analysis Results This page shows a sequence of plots generated by a list of analyses performed with a longer time interval than the prompt analyses (± 1000). This time window is divided into segments of 200 sec, and the pipeline executes several analyses with the AGILE/GRID and AGILE/MCAL instruments' data for each segment. The GUI shows four plots for each time segment (AGILE/GRID counts map, AGILE/GRID exposure map, AGILE/MCAL light curve, and two visibility plots). The user can visualise the data acquired by AGILE before and after the science alert.

3.6.2 AGILE/MCAL Pipeline

The MCAL instrument has a dedicated real-time analysis pipeline developed during this PhD in collaboration with the AGILE Team using the RTApipe framework. The workflow of the pipeline is described in Fig. 3.13. This pipeline's input Data Types are received for each contact by the preprocessing and reconstruction pipeline executed in the ASI/SSDC Science Opera-

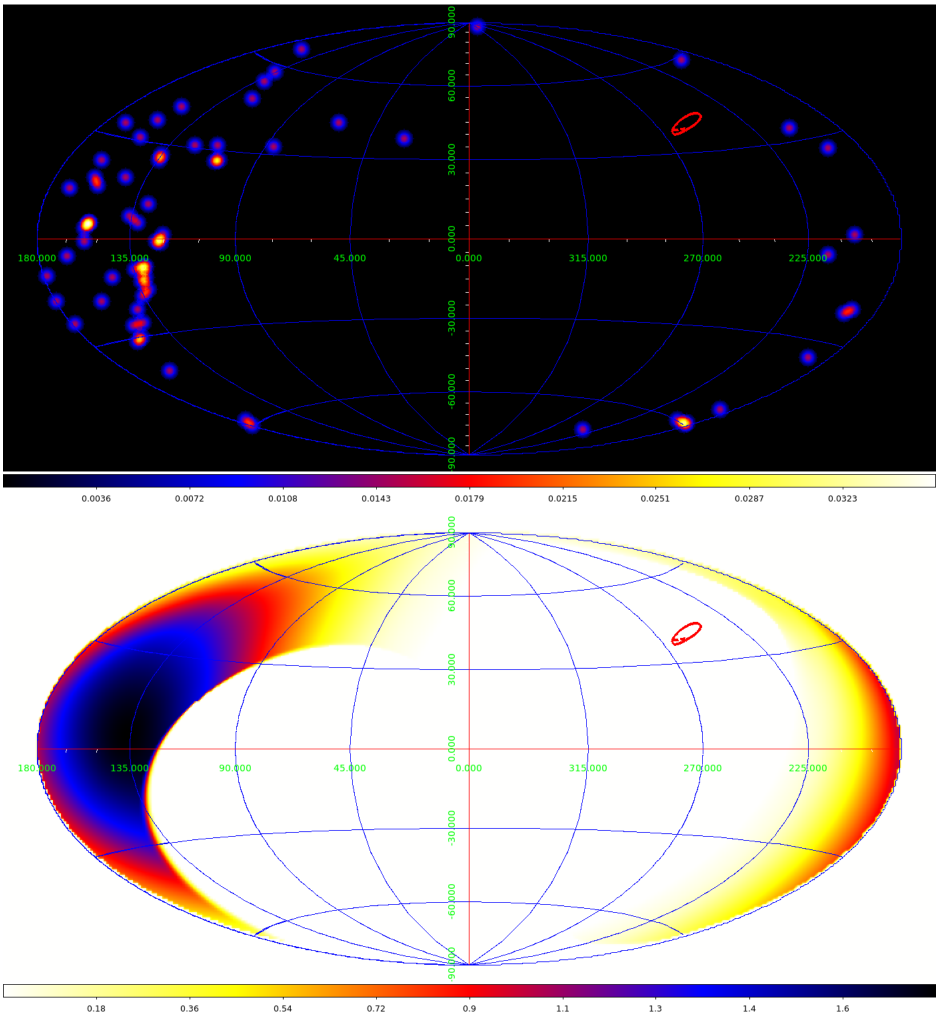

Figure 3.11 Results obtained with the AGILE/GRID real-time analysis pipeline in different time windows centred in the time sent with the science alert. The pipelines prepare different types of gamma-ray sky maps. The top plots show the counts maps of AGILE/GRID where each pixel represents the number of photons detected in that region. The bottom plots show the exposure maps of AGILE/GRID where each pixel represents the AGILE/GRID instrument's exposure in that region expressed in cm^2 s . The sky regions defined with red lines represent the error localisation region of the science alerts.

tion Center (Roma) (Fig. 1.1). The COR 3908/3916 and the LOG Data Types are the AGILE/MCAL pipeline's input data. When the contact is transferred from the SSDC to the INAF/OAS Bologna, the indexes of the

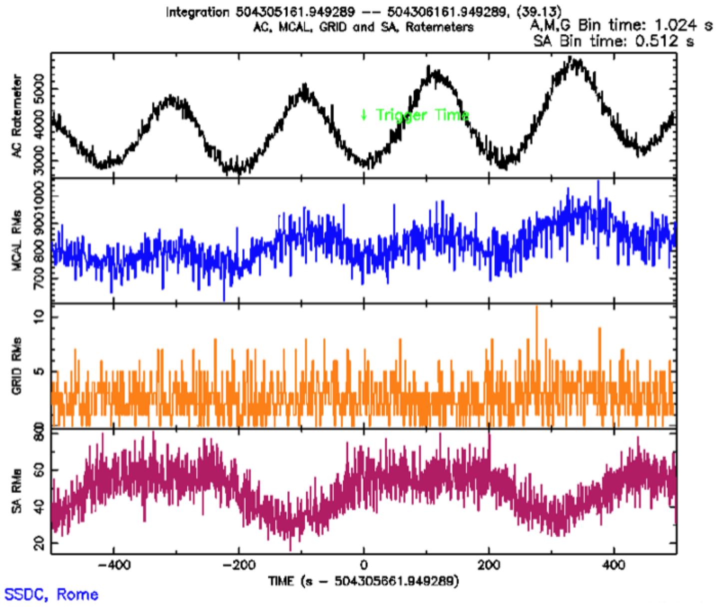

Figure 3.12 Results obtained with the AGILE instruments' ratemeters centred in the time sent with the science alert. Each panel represents the acquisition rate of photons or particles by the instrument onboard AGILE: MCAL, SuperAGILE, GRID, and ACS.

Data Types received are updated and the pipeline's triggers, described in Sec. 3.4.3 are executed. The triggers check if there are analyses configured for the Data Types and create the analyses. The Pipeline Manager checks if there are analysis in 'runnable' status and submits these analyses to the Task Manager. The pipeline automatically performs analyses searching for GRBs and TGFs at each contact. In addition to GRBs, events with a lower significance threshold are identified as GRBLike and Sub-Threshold Events (STE). The results of these analyses are then imported in a MySQL database and stored into the server's file system where the analyses are executed. The Astronomer on-duty can visualise the AGILE/MCAL pipeline results through the web GUI (Sec. 3.6.2) that queries the database and shows the images present in the file system.

AGILE/MCAL Results

The AGILE Team can visualise the AGILE/MCAL pipelines' results using the web GUI developed during this PhD. The home page of this GUI consists of a table showing the list of satellite's contacts. For each contact, thus table shows the time windows of the contact data coverage and preliminary results

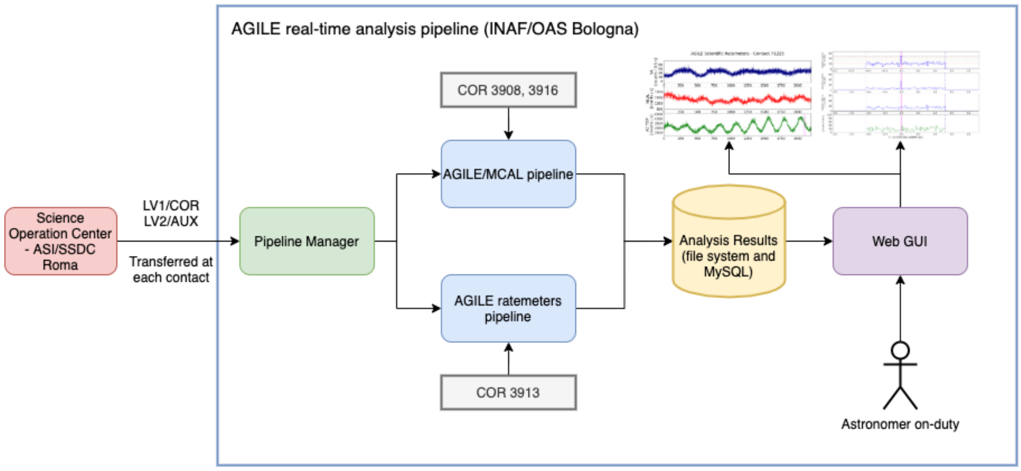

Figure 3.13 The workflow of the AGILE/MCAL and AGILE ratemeters pipelines. The input data are received by the ASI/SSDC Science Operation Center at each contact acquired by the ground station (Fig. 1.1). When the data are received, the two pipelines start a list of configured analyses that produce results stored in a MySQL database and in the file system. The Astronomer on-duty can visualise these results using the web GUIs.

of the analyses performed on the contact data (e.g. the number of GRBs or TGFs if detected).

It is possible to load a web page showing the table of all the GRBs detected since the pipeline's first run. For each GRB, this table shows several data (e.g. signal to noise ratio of the GRB). In addition to the information shown directly in the table, it is possible to use a button which opens a web page for a specific GRB showing the light curve (Fig. 3.14). This light curve represents the number of photon counts at different energy ranges.

The real-time analysis software that analyses AGILE/MCAL data is also designed to identify the TGFs. A web page shows the table with all TGFs detected by the RTA pipeline. This table contains information about each TGF (e.g. the duration, and the average energy) and has a button that user can press to open the detail page with the TGF light curve and position map. The results shown in this detail page (Fig. 3.15) are contained in three plots. The first plot shows the TGF light curve in terms of counts versus time aggregated in 50-microsecond bins. The second plot highlights the energy of each photon of the light curve. The last plot shows a red dot representing the satellite position on the earth map. This plot can be used to identify the TGF position and compare it to a weather map (TGF are generated during

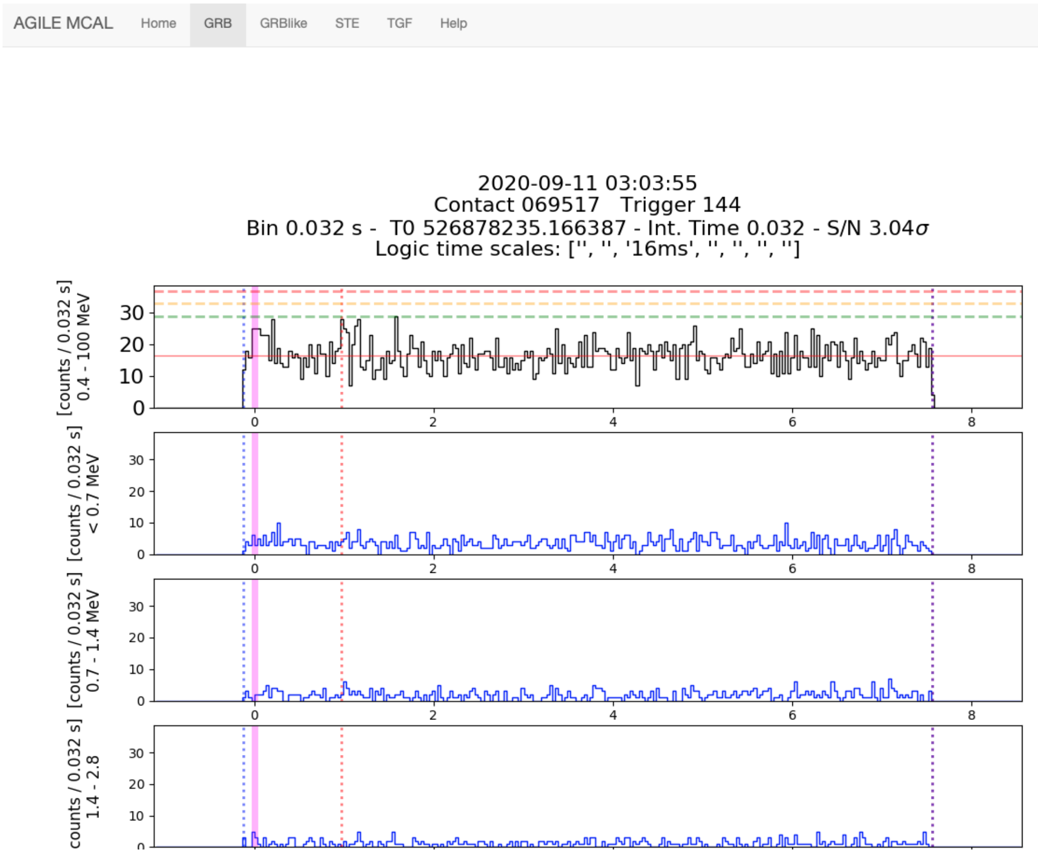

Figure 3.14 Light curves of the AGILE/MCAL instrument. It shows the photons counts at different energy ranges in function of the time.

tropical storms).

3.6.3 AGILE Ratemeters Pipeline

During this PhD, the real-time pipeline to perform analysis on the ratemeters (RM) acquired by all instruments on board the AGILE satellite was developed in collaboration with the AGILE Team. This pipeline analyses each instrument's RM data contact-by-contact and can detect fast transients, GRBs, and solar flares. This pipeline also includes a dedicated web GUI to visualise the results of the analyses.

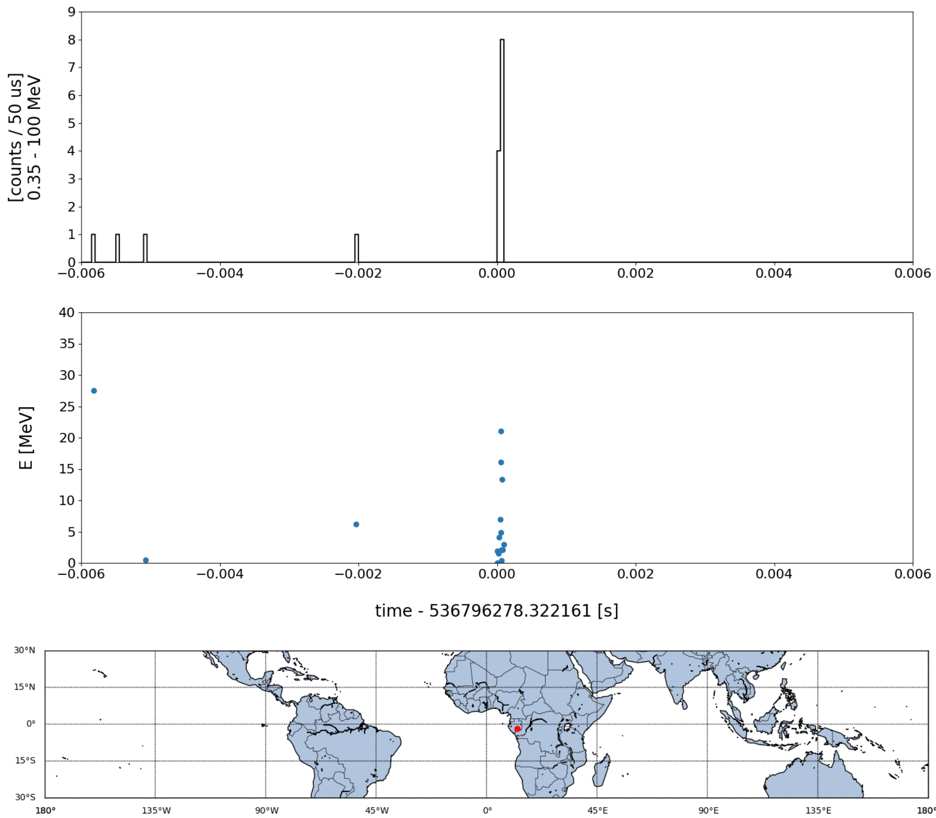

Figure 3.15 The first panel represents the light curve of a TGF with the photons counts in the Y-axis and the time in the X-axis. The second panel represents the energy of each gamma-ray detected for the TGF in the Y-axis. The last panel represents the position of the satellite at the moment of the TGF detection.

RM pipeline Results

The AGILE Team uses this web GUI to visualise the results of the RM pipeline. The home page of the GUI contains a table showing the list of contact. For each contact, the table reports information (e.g. the time windows of data coverage) and allows the user to open a detailed web page for a specific contact.

The RM detail page (Fig. 3.16) shows a list of plots in which there are multiple panels. Each panel represents the RM related to an instrument onboard the satellite. The panels show the original data aggregated with 0.5 sec bins. The web page also showst the panels with the same data but after the trend removal with the Fourier transform method. The RM plots show the data of the following instruments: Super AGILE, MCAL, and the

AGILE Anticoincidence system described in Sec. 1.2.

The Astronomer on-duty is in charge of monitoring the automated RM pipeline results periodically during day and night. If a burst is displayed within these diagrams and, in particular, when the burst is present in more than one panel, the Astronomer on-duty must perform more detailed manual analyses to investigate the phenomena. The peaks are highlighted within the diagrams with a magenta band.

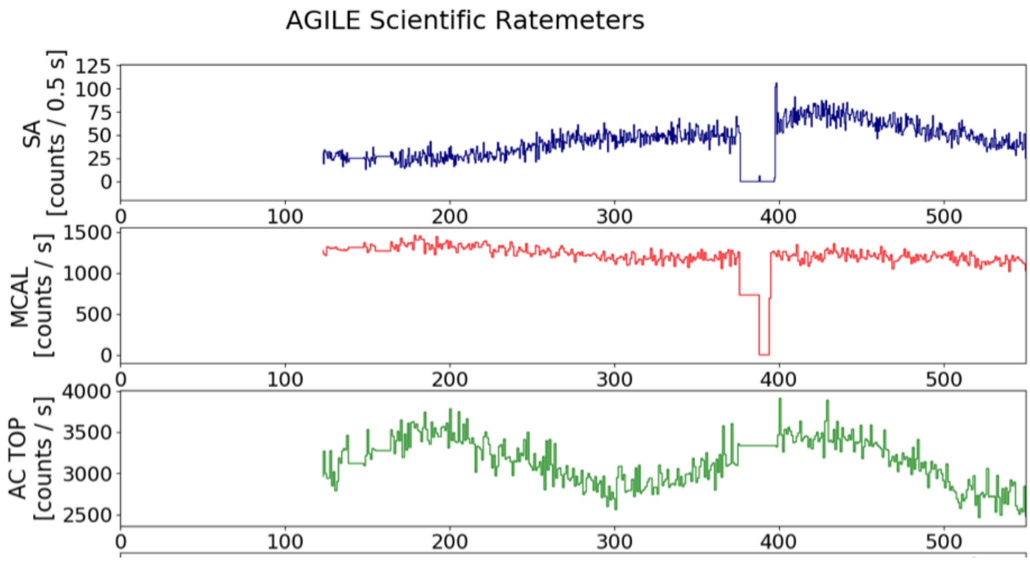

Figure 3.16 Results of the ratemeters pipeline. The contact data is divided into time windows of 550 seconds, and for each time window, the plots represents the acquisition rate of photons or particles by the instrument onboard AGILE: MCAL, SuperAGILE, and Anticoincidence.

3.6.4 Results obtained with the AGILE RTA Pipelines

The AGILE researchers use the pipelines described in Sec. 3.6 to monitor the scientific analysis performed as soon as the AGILE data are available to the software. With these web GUI, a fast follow-up of the external science alerts received through the GCN network from other observatories is possible. When the automated pipelines detect a transient during the analysis, a notification email is sent to the AGILE Team. Researchers can visualise the analysis results using the web GUI and then perform manual analysis for more details if necessary. The AGILE Team can follow up these external science alerts with a rapid workflow thanks to this automated software that react to these alerts and analyses data automatically preparing results to the Astronomer on-duty. Otherwise, the reaction time would be longer because the Astronomer on-duty should manually perform all analyses.

In addition, the AGILE automated pipelines search for transient events into the data of AGILE/MCAL (Sec. 3.6.2) and from ratemeters (Sec. 3.6.3) at each contact received from the satellite. When the AGILE/MCAL pipeline detects a transient event, it automatically sends a Notice to the GCN network to inform other observatories. This sharing of internal alerts with the scientific community is possible only with the real-time analysis pipelines developed during this PhD that automatically search for transient events. Both the pipelines send an email to the AGILE Team to inform researchers of the alert. They can perform manual analyses and send further communication to the GCN network or the ATel portal with more details. Without automated software, this workflow would not be possible.

The next sections show some communication examples sent by the AGILE Team and the AGILE pipelines to the scientific community.

GCN Circulars and Notices

During the PhD, the AGILE Team was involved in the follow-up of GWs detected by the LIGO-Virgo collaboration with a daily commitment of 24 hours out of 24. One of the main goals of the real-time analysis pipelines developed during this PhD is to receive external science alerts and react to them analysing the AGILE data automatically. The AGILE Team shares the results of this follow-up with Circulars sent to the GCN network. These Circulars are written manually after verifying the automated pipelines' results and containing information about the possible detection of a counterpart of the science alert received through the GCN network or about the gamma-ray flux upper limit. The pipelines prepare a draft that the Astronomer on-duty can use as a starting point to write the Circular.

The Circulars [Urs+18; Ver+18] were published to the GCN network by the AGILE Team after the detection of the GRB180914B with AGILE/MCAL and AGILE/GRID instruments respectively. AGILE was the first observatory to send a science alert to the GCN network about this GRB. Other observatories started the transient event's follow-up in an MW context (gamma-ray, x-ray, ultraviolet, optical, and infrared). Many Circulars were published by other observatories that found counterparts to this GRB. A shortlist of these Circulars for the follow-up of other observatories:

Fermi-LAT [BL18], Swift-XRT[17] [D'A+18], and Swift-UVOT[18] [PK18].

The GCN Circular [Urs+19b] was published by the AGILE Team on the GRB 190114C detection by the AGILE/MCAL instrument. This alert is important because the MAGIC observatory also detected the same GRB 190114C [Mir+19] and for the first time, a GRB was detected at very high energy greater than 300 GeV from an IACT telescope [MAG+19b].

The GRB190606A was reported by the AGILE Team with the GCN Circular [Urs+19a]. This GRB was also detected by Fermi-LAT [BLF19], and Fermi-GBM [RMF19]. The GRB observation from multiple observatories leads to a triangulation by the IPN to detect a small error region where the GRB could be located [Koz+19].

In addition to the Circulars, the AGILE automated pipelines send to the GCN network another kind of communication called Notice that is automatically generated and meant to be read and parsed by machines. This communication is sent as soon as the automated analysis detects a transient (e.g. GRBs) without the human intervention. Other observatories receive these Notices and can perform a follow-up starting from the detection made with AGILE instruments, in the same way as AGILE perform the follow-up of other observatories detections. Since May 2019, when this feature was developed inside the AGILE/MCAL pipelines, more than 40 Notices of GRBs have been sent to the GCN network[19].

The Astronomer's Telegrams

The Astronomer's Telegram[20] (ATel) is an internet-based service where observatories can publish their short communication about transients. This service allows the researchers to follow-up the results of other observatories and shares their results. This communication channel is not intended only for the GRBs similar to the GCN network. Indeed through this portal, it is possible to notify the community for a new source or a source that increases its flux during a period of hours or days. Telegrams as available instantly on the service's website and distributed to subscribers. These ATels are sent via email by the researchers (with authorized access) to this portal and are available in less than one second.

During the period of this PhD, the AGILE Team sent many ATels. Most of them are for transient events that were detected with the automated pipelines and then verified manually by researchers.

[17] https://swift.gsfc.nasa.gov/about_swift/xrt_desc.html
[18] https://swift.gsfc.nasa.gov/about_swift/uvot_desc.html
[19] https://gcn.gsfc.nasa.gov/agile_mcal.html
[20] http://www.astronomerstelegram.org

The two ATels [Urs+20b; Tav+20a] can be used as an example of bursts detected by AGILE instruments from a source called SGR 1935+2154. These bursts were detected with the real-time analysis RM pipeline and shared with the scientific community to allow other observatories to follow-up the transient event.

The ATel [Luc+19b] shows the results of a MM follow-up that the AGILE Team performed on an external science alerts received from the GCN network. This science alert was sent by IceCube and is related to neutrino detection.

Papers

The results of the follow-up of external science alerts are usually published in referred papers. These papers can be produced by a collaboration between different observatories that share their data where the AGILE Team reports the results obtained with manual analysis and the automated pipeline.

During this PhD, the AGILE Team published several papers on the results obtained during the follow-up of external science alerts performed with the support of the real-time analysis pipelines developed for this PhD activity. The main papers are: [Urs+20a], [Tav+20b], [Cas+20], [MAG+19a], [Ver+19b] and [Luc+19a].

3.7 CTA high-level reconstruction prototype

The prototype of the real-time analysis pipeline for the CTA Science Alert Generation (SAG) system was developed during this PhD. CTA, described in Sec. 1.3 is an observatory for the ground-based gamma-ray astronomy. This software is one of the sub-systems of the Array Control, and Data Acquisition System (ACADA) [Oya+19] of the CTA observatory.

The data produced by the CTA analysis software are divided into different categories:

- A (real-time): data produced and distributed with low latency for science alert generation and fast data quality evaluation.
- B (next-day): data produced on-site by with off-line software during the next day.
- C (final): data produced with full high-quality data processing pipelines in the off-site data centre.
- S (Simulation): data produced during the simulation.

For all these categories, the data are classified with different data processing level:

- R0-R1: data level of the hardware and data acquisition software. These data are not stored for long-term analysis.

- DL0-DL1-DL2: intermediate data product generated by the reconstruction pipeline. This pipeline aims to reconstruct the energy, direction and other parameters for each event.

- DL3: this data level represent the selected Air-Shower Events (event list) and the associated instrumental response characterizations. This is the first data level provided to the user for the analysis and also used by the automated scientific analysis pipeline.

- DL4-DL5-DL6: the DL4 data are intermediate data (e.g. counts and exposure cubes) produced by the automated scientific analysis starting from DL3 data level. The DL4 data are then used to generate the DL5 data level, which is the first level of scientific quick-look result containing information in physical units (e.g. light curves and spectra). The DL6 data level contains high-level legacy data products such as survey sky maps, diffuse gamma-ray background models or catalogues.

The SAG system is divided into different components: low-level reconstruction pipeline, online data quality, and high-level reconstruction pipeline. The high-level reconstruction pipeline is involved in the scientific analysis of the data acquired by the CTA observatory during the observations. This component is called SAG-SCI. The SAG-SCI pipeline receives DL3 input data from the low-level reconstruction pipeline and produces DL4 and DL5 results. This pipeline analyses data in real-time as soon as they are received and produces data of category A.

The CTA observatory, described in Sec. 1.3 is a gamma-ray observatory and will be involved in the MW and MM astronomy.

The full list of SAG requirements and functionalities is described in [Bul+13] and [Bul+15].

The SAG-SCI main requirements are:

- The SAG-SCI is a software system that analyses CTA data during the observation in quasi-real-time.

- This software will run on-site near the telescope's location. For this reason, the computational resources are limited.

- The SAG-SCI has a latency of five seconds to perform the analyses and to detect candidate science alerts since the event list is reconstructed from raw data.

- CTA can observe different target at the same time with eight sub-array. The software must perform in parallel the analyses for these observations.

- The analysis must search for transient phenomena on different timescales from 10 seconds to 30 minutes.

- The sensitivity of the analysis is required not to be worse than the one of the off-line analysis by more than a factor of 2.

The CTA will observe the gamma-ray sky with tens of telescopes. Each telescope will produce a huge amount of data (up to 5 GB/s) and thousands of events per second (up to 15 kHz). All this data must be analysed in real-time to generate science alerts and follow-up science alerts received from other observatories in the MW-MM context. Considering the "5 Vs" Big Data definition, the data produced by CTA telescopes can be classified as Big data. The Volume and Velocity are very high, and it is required a real-time analysis to detect gamma-ray transient events within seconds. The Veracity of the data is evaluated in real-time by a SAG component called Online Data Quality. The Value of the results obtained with the real-time analysis can be used to share the science alerts with the scientific community and to have feedback on the ongoing observations.

During this PhD, a prototype of the SAG-SCI system was developed in collaboration with the INAF/OAS Bologna team using the RTApipe framework. The use cases (Sec. 3.2) and the requirements (Sec. 3.3) of the RTApipe framework are in common with all gamma-ray observatories including CTA. For this reason, this prototype was developed to evaluate the RTApipe framework capabilities in a challenging context like the CTA/SAG. In addition, this prototype can be used as a starting point to develop the final SAG-SCI software system.

The configuration of the environment used to implement this pipeline with the RTApipe framework is described in the INAF Technical Report [PB20b].

Future works are planned to test the performances of the SAG-SCI prototype, developed with the RTApipe framework, simulating eight parallel observations as defined by requirements. During these performance tests, the SAG-SCI will be in charge of analysing the data stream received from the low-level reconstruction pipelines of more than 100 telescopes in real-time. The system will be deployed into several servers exploiting the scalability feature of the RTApipe framework. The Slurm scheduler can balance the workload on the cluster of servers and prioritise the processes. This test can be performed with simulated data to reach the maximum performances capability of the system. Failures can be simulated to evaluate the pipeline capability to recover failure events without losing data or results. The SAG-SCI has a requirement to complete the analyses in five seconds. This requirement can be verified during the test. The infrastructure to enable this kind of test is being developed in collaboration with other team from the CTA consortium.

The SAG-SCI prototype receives a simulated event list. The event list (DL3) is the input data for this software and contains all the gamma-ray events detected by the telescopes and reconstructed from raw data. The software executes a list of pre-configured analyses as soon as new DL3 data arrives into the system. The Science Tools used to perform these analyses are the ctools [21]. The prototype search for gamma-ray source detection using the Maximum Likelihood Estimator (described in [Mat+96]) and generates different results (sky maps, light curves and more).

The diagram in Fig. 3.17 shows a high-level functional schema of the SAG-SCI pipeline. The Pipeline Manager is the orchestrator of the pipeline. When new events are inserted into the DL3 database, the Pipeline Manager executes scientific analyses on new data. The analyses are created by the Science Logic when the DL3 Data Type index is updated into the pipeline database.

To run an analysis with a Science Tool, the Pipeline Manager needs a Wrapper which provides an interface between the Pipeline Manager and the Science Tool.

This wrapper read the configuration XML file, described in Sec. 3.4.4, to obtain the parameters for the analysis. These parameters can be customised on the specific Science Tools. Then the wrapper can execute the analysis with the Science Tool configured inside the pipeline database. The pipeline can be configured with several Science Tools developing one wrapper for each Science Tool. It is possible to add new Science Tools without modifying the pipeline architecture; the user needs to add a few rows into the pipeline database. This interface between the pipeline and the Science Tools provide great flexibility to update the pipeline with new analysis software when required.

The results of the analysis are stored in a MySQL database. The Astronomer on-duty can visualise these results through the web GUI developed as part of the prototype (Fig. 3.18). The GUI will be not part of the work package of the SAG. Thus, the GUI presented here is developed just to vi-

[21] http://cta.irap.omp.eu/ctools

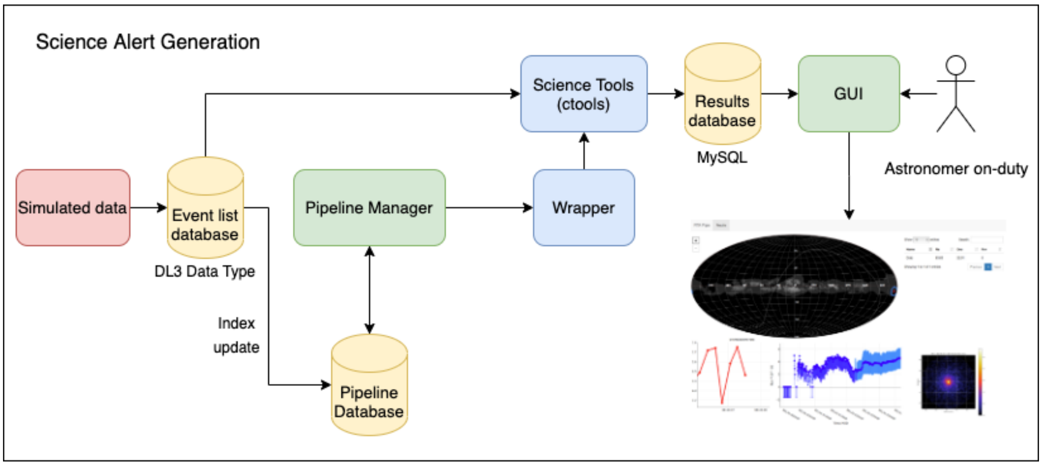

Figure 3.17 High-level functional schema of the SAG-SCI pipeline prototype for the CTA real-time scientific analysis. This pipeline executes the analysis when simulated data are sent to the pipeline. The pipeline executes analyses with different Science Tools. The results are stored in a MySQL database and in the file system. The Astronomer on-duty can visualise these results using the web GUIs.

sualise the results obtained with the pipeline and not propose a solution for the CTA production software. The GUI is developed starting from the GUI developed for the AGILE real-time analysis pipelines and with the base environment described in Sec. 2.3 plus the d3-celestial[22] library used to plot the interactive sky map in galactic coordinates.

The Astronomer on-duty during the observation can use this GUI to visualise the results of the real-time scientific analysis. The GUI shows a map of the sky in galactic coordinates where all the ongoing observations are displayed. Each observation has a blue circle representing the FOV of the sub-array that is pointing a target (red dot). The Fig. 3.18 shows four observations that are executed simultaneously with four different sub-array of telescopes. For each observation, a pipeline analyses the incoming data to search for detection of transient events that can be classified as science alerts. The GUI also shows a table with all targets, their names and sky positions. The bottom part of the GUI represents three different plots:

- The left plot shows the average event rate (Hz) of the observation.
- The central plot represents the light curve (gamma-ray flux in the func-

[22]https://github.com/ofrohn/d3-celestial

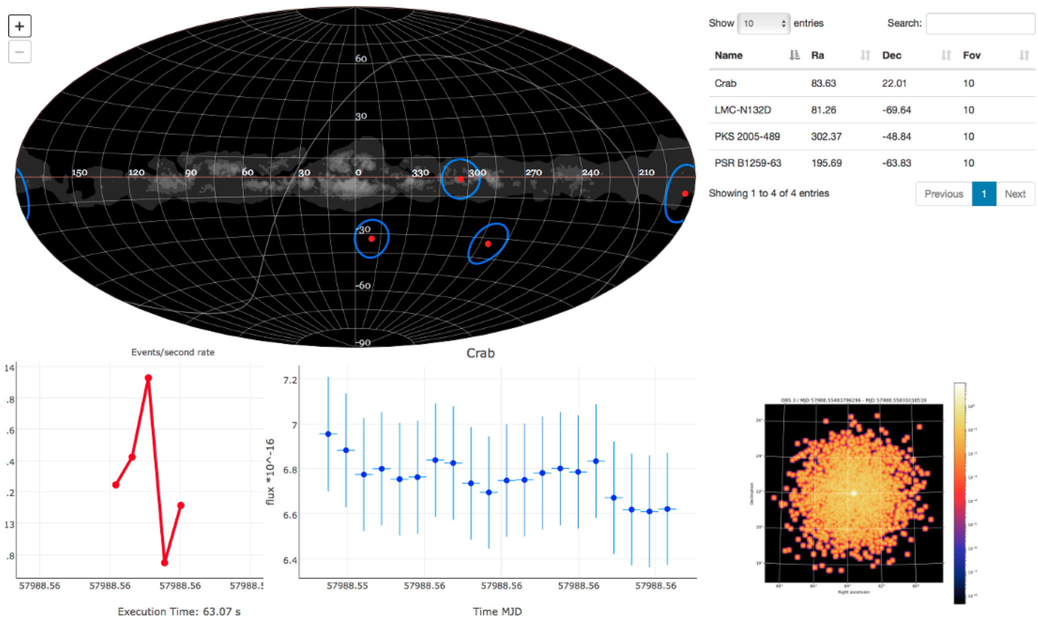

Figure 3.18 Graphical user interface of the SAG-SCI real-time analysis prototype. The GUI shows a map of the sky in Galactic coordinates with all the active observations (blue circles), a light curve and a counts map in the FOV of each observations, and a plot to show the acquisition rate of the telescopes (Hz).

tion of the time) of the target that the sub-array is pointing. This light curve is updated in real-time as soon as the last data received analysis is terminated.

- The right plot shows a counts map. Each pixel of the map represents the counts detected during the observation in that sky region.

3.8 ASTRI Horn analysis software prototype

The RTApipe framework is used to develop a prototype of the scientific analysis of the ASTRI Horn telescope (described in Sec. 1.4). This prototype and the know-how acquired during the development can be used as the starting point to develop the scientific analysis pipeline of the ASTRI Mini Array.

This software is developed to analyse ASTRI Horn simulated data to test the performances and functionalities the Science Tools used to analyse data. The prototype's architecture is very close to the architecture developed for the SAG prototype shown in Fig. 3.17.

The simulator that sends data into the scientific analysis pipeline was developed to test the analysis, and the results obtained simulating a real observation. The pipeline executes scientific analyses on the event list produced starting from the raw simulated data. This event list is analysed with common gamma-ray Science Tools such as ctools. These analysis results are saved into a database, and a GUI shows these results to the Astronomers on-duty (Fig. 3.19). The GUI is similar to the one shown in the Sec. 3.7 developed for the CTA observatory. In this case, there is only one observation because a single telescope composes the observatory. Both the GUIs are implemented with the same base environment described in Sec. 2.3 plus the d3-celestial[23] library used to plot the interactive sky map in galactic coordinates.

This prototype can be used to test the Science Tools but also the interfaces between different components. It is possible to implement scientific use cases and visualise the analysis results updated in real-time during the simulation through the GUI.

[23]https://github.com/ofrohn/d3-celestial

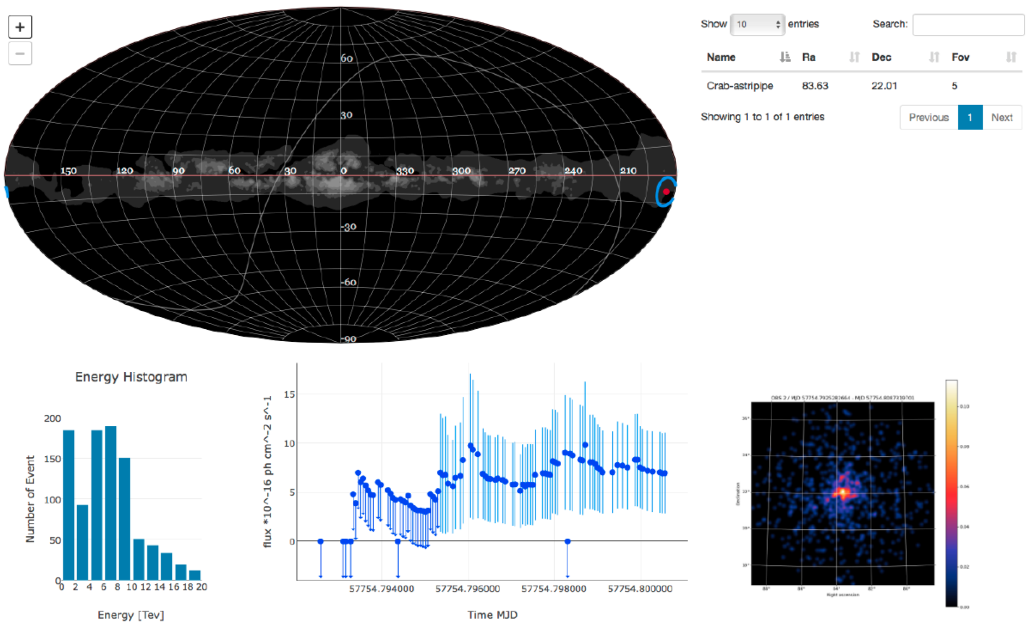

Figure 3.19 Graphical user interface of the ASTRI Horn scientific analysis prototype. The GUI shows a map of the sky in galactic coordinates with the pointing of the telescope, a light curve for the gamma-ray flux calculated in the pointed sky region, a counts map of the instrument FOV, and a histogram of the energy of the events acquired by the telescope.

3.9 Requirements verification

The RTApipe framework developed during this PhD is designed to fulfil the 13 requirements described in Sec. 3.3. The software architecture (Sec. 3.1) and the key features (Sec. 3.5) of the RTApipe framework aim to satisfy these requirements in the gamma-ray astronomy context for which this framework is designed.

The software developed during this PhD using the RTApipe framework for the AGILE space mission, for the CTA observatory, and for the ASTRI Horn telescope can be used to verify that all these requirements are satisfied.

The AGILE real-time analysis pipelines described in Sec. 3.6 implement the use cases referred to the requirement 1. The science analysis pipeline prototype developed for the ASTRI Horn telescope shows that the RTApipe framework can work in both on-line and off-line mode, and with spacecraft or ground-based telescopes. This behaviour satisfies the requirements 2 and 3. The real-time analysis prototype for the CTA observatory (Sec. 3.7) confirms that the RTApipe framework can analyse data acquired by several sub-array of telescopes observing different sky region at the same time (requirement 4) and work with the Bid Data produced by the CTA telescopes (requirement 5). The Task Manager component (Sec. 3.4.6) aims to perform parallel analyses and to manage priority between analyses (requirements 6 and 7). In addition, the Task Manager is implemented with the Slurm workload manager that is scalable to a cluster of servers changing a configuration file (requirement 8).

The AGILE real-time analysis pipelines described in Sec. 3.6 implement different analysis workflows to start analyses when a science alert is received or when the observatory instruments acquire new data. These real-time analysis pipelines can execute the analyses with several Science Tools and can be configured to perform a list of analysis for each science alert type (e.g. GWs or GRBs) or instruments that sent the alert to the GCN network (e.g. IceCube or Fermi). The pipelines are configured to perform analyses with different time scales (e.g. 100 or 1000 sec). These features implemented using the RTApipe frameworks confirms that the framework is compliant with requirements 9, 10, 11, and 12. Several of these features are also implemented in the prototype for the real-time analysis of the CTA observatory described in Sec. 3.7.

As described in Sec. 3.6 it is possible to deploy the Singularity container, where all the software packages are installed, on backup servers geographically distributed (requirement 13).

The real-time analysis software developed with the RTApipe framework and described in this section confirms that all the requirements are satisfied. Gamma-ray observatories can use the RTApipe framework to develop scientific analysis pipeline in the MM and MW astronomy context.

Chapter 4

Deep Learning Method for GRB Detection

As described in Sec. 3, during this PhD, the real-time analysis pipeline to react to external science alerts received from the GCN network was developed in collaboration with the AGILE Team. A science alert is a communication from/to the astrophysical community that a transient phenomenon occurs in the sky. This pipeline performs several analyses using different Science Tools. To analyse the data acquired by the AGILE/GRID instrument to search for transient events in a short term scale (seconds) the currently used method is the Li&Ma, described in Sec. 4.1.

This chapter describes a new GRB detection method developed during this PhD in collaboration with AGILE Team based on Deep Learning (DL) technology. This method has the goal to improve the detection capability of these AGILE real-time analysis pipelines. The results obtained on real AGILE data with the new method are compared with those obtained using the standard Li&Ma method on the same GRBs. These results are described in Sec. 4.7. The DL algorithms require a training phase where huge datasets of simulated data are used. To simulated these datasets the model of the AGILE observing pattern (Sec. 4.3) and the physical model of the GRBs that can be detected by AGILE (Sec. 4.4) are studied. The new method uses a Convolutional Neural Network (CNN) to detect GRBs into the AGILE sky maps, the architecture of this CNN is described in Sec. 4.5, while the procedure used to evaluate the CNN is described in Sec. 4.6. This CNN can be installed as Science Tool into the real-time analysis pipeline developed for AGILE (Sec. 3.6) to detect GRBs counterparts when external science alerts are received. More details about the work presented in this chapter can be found in this paper under review ([Par+b]).

4.1 Introduction and state of the art algorithms

The real-time analysis pipeline developed for AGILE can react to external science alerts and detect a possible GRB counterpart in AGILE/GRID short term (< 1000 sec) observations. A detection occurs when the pipeline finds a transient event with a statistical significance above a defined threshold (usually three sigma). The GRB position and the trigger time are known in advance because the pipeline reacts to external science alerts sent by other observatories through the GCN network. The analysis is performed using the aperture photometry method [LM83], evaluating the gamma-ray photon counts detected inside a time window containing the target (T_{on}) and a time window containing only background (T_{off}). The counts are selected from the AGILE/GRID photon list in a radius of 10° from the centre of the error localisation region reported by the external science alert to include the source emission considering the instrument Point Spread Function (PSF). The AGILE/GRID instrument has a PSF $< 10°$ at energy > 50 MeV [Sab+15]. The background is evaluated before the GRB trigger time (the instant when the physical event occurred) because the effective duration of the GRB in the AGILE/GRID energy-range is unknown.

In the current method used by the AGILE/GRID automated pipeline, the significance of the event is calculated with the Li&Ma formula [LM83] using the counts extracted with the On/Off procedure described before. This method, from now on called Li&Ma has two main limits.

The Li&Ma method does not use the shape of the PSF during the analysis, but just the event number inside a region defined large enough to include the PSF. Furthermore, the Li&Ma method requires the counts in both T_{on} and T_{off} time windows to be not too few, with a threshold of ten photons usually applied [LM83].

The AGILE Team accepted this threshold of ten events defined in [LM83] for the real-time analysis pipeline. The detections with less than ten counts are not considered.

This chapter describes a new detection method to overcome these limits and, in general, improve the AGILE/GRID automated pipeline's ability to follow-up GRB alerts. This new method can be implemented into the AGILE real-time analysis pipeline to detect GRBs inside the AGILE/GRID maps when an external science alert is received.

This new method uses a class of DL algorithms called Convolutional Neural Network (CNN) described in detail in Sec. 4.5. In this new method, the CNN is used to classify AGILE/GRID intensity maps and detect GRBs into the field. Intensity maps are counts maps divided by the exposure of the AGILE/GRID instrument and report the measurement in ph cm^{-2} $\mathrm{s}^{-1} sr^{-1}$ for each pixel, allowing the normalisation of the observation for the exposure.

The CNN requires a training phase with large simulated datasets of sky maps representing the average background level and the GRB flux distribution expected in the AGILE/GRID energy range (100 MeV - 10 GeV). The study of the observing pattern is described in Sec. 4.3. The Sec. 4.4 describes the GRB model used to simulate GRBs for the CNN training. After performing the CNN training, the *p-value* distribution from only background maps is computed in different observational conditions (Sec. 4.6). The CNN is then applied to real data using the GRBs of the Swift/BAT[1], Fermi/LAT

[1] Swift/BAT Gamma-Ray Bursts Online Catalog https://swift.gsfc.nasa.gov/archive/grb_table/

[Aje+19] and Fermi/GBM [2] catalogues, considering them as external alerts. This analysis is described in Sec. 4.7.

The main reasons why the CNN method improves the AGILE GRBs detection capabilities are:

1. The CNN can be trained on a specific instrument, learning from huge datasets of simulated data, while the Li&Ma is a general method, and it is not trained on a specific dataset. The PSF of the AGILE instrument is used during the CNN training phase to define the convolution process's kernels' size.

2. The CNN is trained with dataset simulated using the background level calculated during real AGILE/GRID observation. In addition, the flux of the simulated GRBs are simulated following the values extracted from the Fermi/LAT GRB catalogue [Aje+19]. All this information is learned by the CNN, while the Li&Ma is applied as-is.

3. The CNN does not require a minimum number of events to be used. On the contrary, the Li&Ma requires at least ten events in both the T_{on} and T_{off} time windows.

4. The results described in Sec. 4.7 show that the CNN can detect 21 GRBs from real AGILE data as counterparts to GRBs detected by other observatories (Fermi and Swift). The Li&Ma from the same list of GRBs can detect only two counterparts. This detection capability is an excellent result for the AGILE/GRID instruments considering the short number of GRBs that this instrument detected during the spacecraft's lifetime.

5. The results obtained show that the CNN improves the AGILE/GRID instrument's detection capability during a follow-up of external science alerts.

6. Starting from the method described here, future works are planned to classify gamma-ray sky maps containing more than one source and perform a regression analysis to determine the GRB position and the flux. These kinds of analyses can not be performed with the Li&Ma method.

This is the first attempt to use a CNN to classify the AGILE/GRID gamma-ray sky maps. With more computational resources, it is possible to improve the CNN method's detection capability training the CNN in more

[2]Fermi/GBM Gamma-Ray Bursts Catalog https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.htm

complex conditions. The results presented here proves the CNN's capability to detect GRBs and justify future research commitment. Future works are planned to use this CNN to perform a blind search into the AGILE/GRID data archive, which contains the data acquired during 13 years of observations.

The methods described in this section to train and evaluate a CNN for the gamma-ray sky maps classification can be used to create a CNN able to classify the sky maps produced by the next generation of observatories such as CTA (Sec. 1.3) and the ASTRI Mini Array (Sec. 1.4). These observatories will produce more complex gamma-ray sky maps collecting a larger number of events and background information. The CNN can be trained on these observatories' new observing condition, learning from their PSF and background levels.

4.2 Assumptions for the analysis

The work presented in this chapter is based on a list of assumptions that fix the parameters used for the analysis. Many of these parameters are inherited from the AGILE real-time analysis pipeline where the CNN method will be installed. In fact, the CNN method is compared with the standard method used in this pipeline. The parameters that are not defined into the AGILE real-time analysis pipeline are defined here to test the CNN model with common conditions for the AGILE/GRID observations. In this work, rare and complex conditions that will be treated in future researches are avoided when possible. The main assumption made for this work are:

1. The gamma-ray sky maps simulated to train and evaluate the CNN have a time window of 200 seconds. This value is selected after the analysis of the AGILE observing pattern described in Sec. 4.3.1.

2. The mean AGILE/GRID exposure in 200 sec sky maps is calculated, excluding exposure levels under a threshold defined to avoid limit conditions that are not the goal for this work.

3. The energy range considered is 0.1 GeV - 10 GeV. This energy range is the standard one used by the AGILE Team to perform analysis on AGILE/GRID data and is supported by the AGILE Science Tools' simulation software.

4. This work is focused on GRBs in the extra-Galactic region ($|b| > 10$, where b is the latitude) to avoid regions with several sources in the

field of view and to avoid the diffuse Galactic gamma-ray background (Sec. 4.3.2) .

5. The background level, used during the evaluation of a map with the CNN, is calculated by performing a Maximum Likelihood Estimator (MLE) analyses. This analysis's time window is defined in an automated way to have at least ten counts in the radius of interest. The AGILE Team defined a minimum ten counts threshold to perform an MLE analysis on AGILE/GRID data.

6. As described in [Bul+12], the AGILE Team's MLE analyses on the AGILE/GRID data have an analysis radius of 10° centred on the source position to limit the effect of systematic errors far from the position of the hypothesised source. In addition, with this radius, all the PSF of the AGILE/GRID instrument at these energies is included in the analysis.

7. The external science alerts considered for this work have a maximum error region of 1°. This scenario covers more than 90% of GRBs presented in the Second Fermi/LAT GRBs catalogue and 100% of the GRBs reported in the Swift/BAT GRBs catalogue. The science alerts with a larger error region are excluded because the CNN developed in this work has not the goal to find the source position performing a blind search.

8. The CNN uses intensity maps (counts maps divided by the exposure maps) as input to make the CNN exposure-independent;

4.3 Modelling the Observations

The AGILE satellite's observation pattern is studied to extract the parameters to simulate the datasets used to train the CNN.

The AGILE orbit (quasi-equatorial with an inclination angle of 2.5° and an average altitude of 500 km, 96 min period) is optimal for low-background gamma-ray observations. From 2007 July to 2009 October, AGILE observed the gamma-ray sky in 'pointing mode', characterised by quasi-fixed pointing with a slow drift ($\sim 1°$/day) of the instrument boresight direction following solar panel constraints. Due to a change in the satellite pointing system control, since 2009 November, the AGILE gamma-ray observations have been obtained with the instrument operating in 'spinning mode' (i.e., the satellite axis sweeps a 360° circle in the sky approximately every 7 min). The axis of this circle points toward the Sun, therefore the whole sky is exposed every six months.

This new mission configuration provides a unique capability to the AGILE satellite to discover transient events. The actual spinning configuration of the satellite, together with a large field of view and the sensitivity of $F = (1-2) \times 10^{-8}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ for 100 sec time integration, grants the coverage of 80% of the sky, with each sky position covered 200 times per day with 100 sec of integration time.

4.3.1 Parameters identification

The complex observing pattern of AGILE in 'spinning mode' is studied with the AGILE Team's collaboration to identify a range of observing conditions (average exposure level, the range of the background) required to perform the Monte Carlo simulations of training, validation, and test datasets. This work is focused on extra-galactic sky regions, and for this reason, a sky region centred at Galactic coordinates $(l, b) = (45, 30)$ is selected to study the exposure and background level. The AGILE/GRID exposure in the centre of this sky region is calculated during one spinning revolution. The Fig. 4.1 shows the typical AGILE/GRID exposure pattern for a sky region that is not covered by the Earth or the Sun: the exposure values are calculated with time windows of 1 sec and a radius of 10°. The exposure is highly variable with a well-defined peak, due to the AGILE rotation.

The study of the exposure pattern during the AGILE spacecraft spinning is used to determine the time window to simulate the sky maps. A time window of 200 seconds is selected because it contains the exposure of an entire AGILE spin, thus with this duration, the probability to have exposure after the trigger time of a transient event is maximised.

The exposure analysis is performed dividing one year of data (1st Jan 2018, 1st Jan 2019) into integration time windows of 200 sec to obtain the exposure distribution. Excluding intervals with zero exposure, the mean value obtained is $\sim 20 \times 10^3\ cm^2 s$ (Fig.4.2). All exposure levels lower than $20 \times 10^3\ cm^2 s$ are excluded considering the histogram in Fig. 4.2. Then a new distribution is calculated without the values lower than $20 \times 10^3\ cm^2 s$ obtaining with a mean value of $\sim 40 \times 10^3\ cm^2\ s$. This value is used to simulate all the datasets.

4.3.2 AGILE/GRID background estimation

During this work, two background components are taken into account. The diffuse gamma-ray background (g_{gal}) produced by the interaction of Cosmic

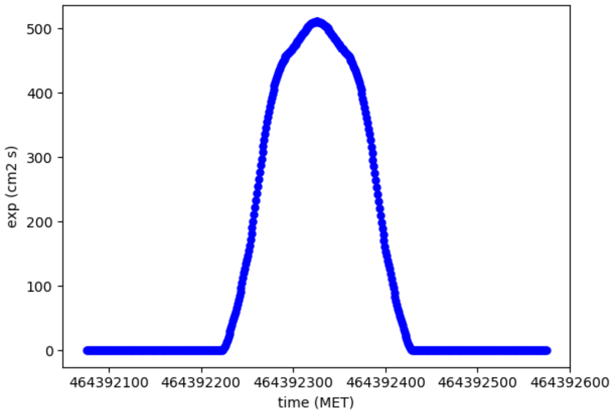

Figure 4.1 Typical pattern of the AGILE/GRID exposure for a fixed accessible sky region given in values of [cm^2 s] as a function of time during the AGILE spinning mode.

Rays (CR) with the Galactic interstellar medium, the Cosmic Microwave Background (CMB), and the Interstellar Radiation Field (ISRF). The second background component is the (quasi) isotropic background (g_{iso}) which includes both a contribution from the cosmic extra-galactic diffuse emission as well as a component of noise due to residual cosmic-ray induced backgrounds at the detector level. In the extra-galactic regions, the isotropic background dominates the AGILE/GRID data. For this reason, the g_{gal} value is considered equal to zero. More details about the AGILE/GRID background model can be found in [Bul+19].

One year of data (1st Jan 2018, 1st Jan 2019) is analysed using time windows of 6 hours to obtain the distribution of the g_{iso} background values in the extra-galactic position defined in the previous section. The time window length is defined to have a mean of ten counts in a radius of 10°. This value is required to perform the background level's statistical analysis using the MLE method. Fig. 4.3 shows the histogram of the g_{iso} distribution excluding time windows with less than ten counts in a radius of 10°. The mean of the distribution is 10.4×10^{-5}(ph cm^{-2} s^{-1}) and the standard deviation is 3.0×10^{-5}(ph cm^{-2} s^{-1}). This distribution is used to simulate the datasets to train and evaluate the CNN, more detail in Sec. 4.5.2.

4.4 GRB Model

Since the launch of the AGILE satellite (23rd Apr 2007) the AGILE/GRID instrument detected eleven GRBs: GRB080514B [Giu+08], GRB090401B

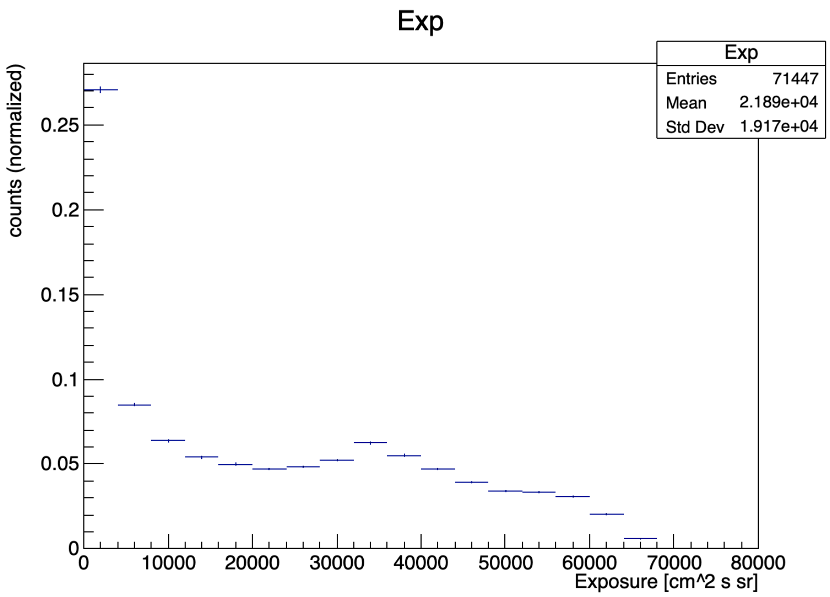

Figure 4.2 Histogram of exposure values [cm^2 s] for 200 sec integrations during one year of AGILE/GRID data.

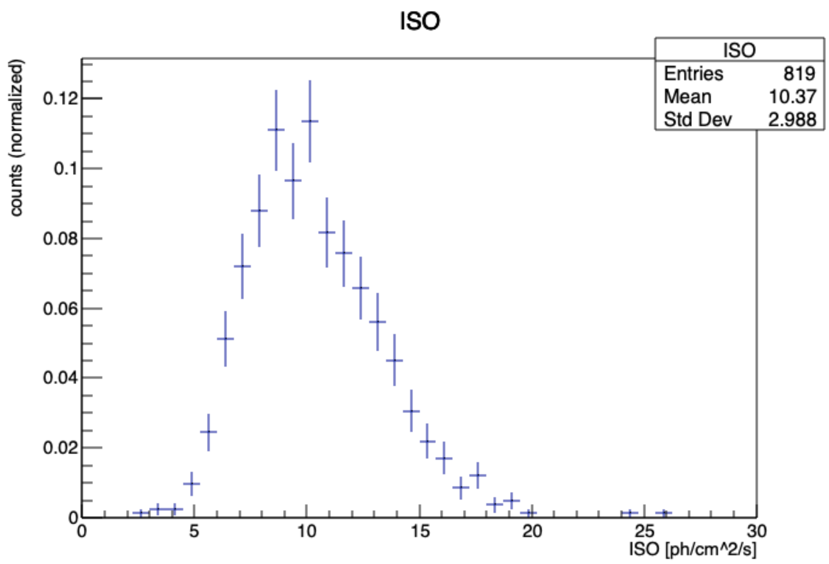

Figure 4.3 Histogram of g_{iso} values [10^{-5}(ph cm^{-2} s^{-1})] for 6 hours integrations during one year of AGILE/GRID data.

[Mor+09; Giu+09], GRB090510 [Giu+10], GRB100724B [Del+11], GRB130327B [Lon+13], GRB130427A [Ver+13], GRB131108A [Giu+13], GRB170115B [Ver+17], GRB180914B [Ver+18], GRB190501A [Luc+19c], and GRB1905030A

[Ver+19a].

These GRBs have some of the main properties of the larger GRBs population detected by the Fermi/LAT instrument and discussed in its First GRB catalogue [Ack+13] and confirmed more recently in its Second GRB catalogue [Aje+19]. In particular, the GRBs' main characteristics at energies greater than 100 MeV, as detected by LAT and used in this study, are the spectral model and its temporal decay. The first one shows a clear flattening of the spectrum to a value around -2 at late times, independent of other GRB properties, and a typical larger duration concerning lower energies, extending up to 1000 sec in the First catalogue and up to 10000 sec in the second one. The temporal power-law decay index is clustered around -1. The GRB detected by the LAT, in its First catalogue, were among the brightest detected by the Fermi/GBM, with a fluence generally greater than a few $10^{-6}\,\mathrm{erg\,cm^{-2}}$ (see discussion in the First LAT GRB catalogue). In the Second LAT GRB catalogue, the fluence limit decreased up to around $10^{-6}\,\mathrm{erg\,cm^{-2}}$ for long GRBs. From here on, the Fermi/LAT catalogue will refer to the Second Fermi/LAT GRBs catalogue.

4.4.1 Flux estimation from LAT catalog

Under the assumption that the simple power law model from [Aje+19] is a good enough estimate of the spectral shape of the Fermi/LAT detected events, the F_{ph}^{LAT} (ph cm^{-2} s^{-1}) within 0.1-100 GeV energy range observed by Fermi/LAT is scaled to a F_{ph}^{GRID} (ph cm^{-2} s^{-1}) value within the AGILE/GRID energy range (0.1-10 GeV). The photon flux in AGILE energy range was evaluated weighting each event by its spectral index. The photon fluence was then computed as $f_{ph} = F_{ph} \times T_{100}$[3]. The fraction of the integrated photon flux emitted by the source in 200 sec exposure time is required to perform the simulation of the datasets. For events with T_{100} greater than 200 sec, the photon flux emitted within the said exposure time is calculated, assuming the catalogued simple power-law evolution model to weight the loss of later emission. The average photon flux of events with T_{100} less than 200 sec is instead mediated over the exposure time to preserve the total photon fluence (Figure 4.4).

The simulation process generates random fluxes for the GRBs starting from a function fitted from real GRBs data contained in the Fermi/LAT catalogue . This function is $F > 6.6 \times 10^{-6}$ (ph cm^{-2} s^{-1}). This exponential function is obtained through Levenberg-Marquardt method: $y = a \cdot e^{-x/b}$

[3]The duration of the burst (T_{100}) is defined as the time between the first and last photon detection in the 0.1-10 GeV energy range to be associated with the GRB with probability $p > 0.9$ [Aje+19].

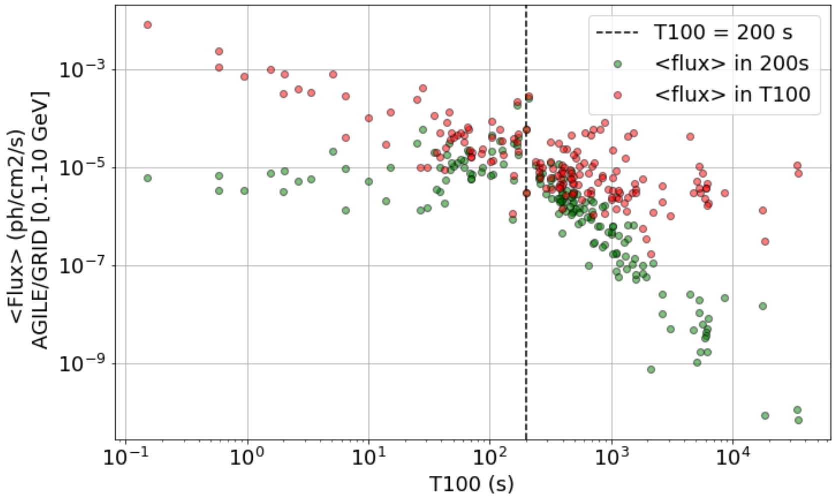

Figure 4.4 Average photon flux of the second Fermi/LAT GRB catalog population as a function of the LAT duration. Red data represent the cataloged flux values within the AGILE/GRID energy range (0.1-10 GeV). In green is the evaluation of the average flux of each event for 200 sec emission.

where $x = log_{10}(F_{ph})$, $a = 2.7 \times 10^{-5}$ and $b = 0.523$. In Figure 4.5 the distribution and fitting function are displayed in the top panel, with residuals shown in the bottom panel.

4.5 Convolutional Neural Network

This new method to detect GRBs into the AGILE/GRID gamma-ray sky maps is developed using a DL approach. The DL architecture used for this work is the CNN, a class of Deep Neural Network (DNN) specifically developed to analyse and classify images as described in [KSH12; GBC16]. The CNN is trained to detect GRBs into the intensity maps produced with data acquired by the AGILE/GRID instrument. This CNN has a multiple layer architecture where each layer can identify a specific feature inside the sky maps. The CNN is trained with a supervised learning technique. This technique requires the CNN training with a labelled dataset. These kinds of datasets contain the results of the classification for each element. Various astrophysical and gamma-rays data analysis contexts already use this CNN method and more in general DL models. In the next section, three works related to this topic are described to find a similar approach to this method.

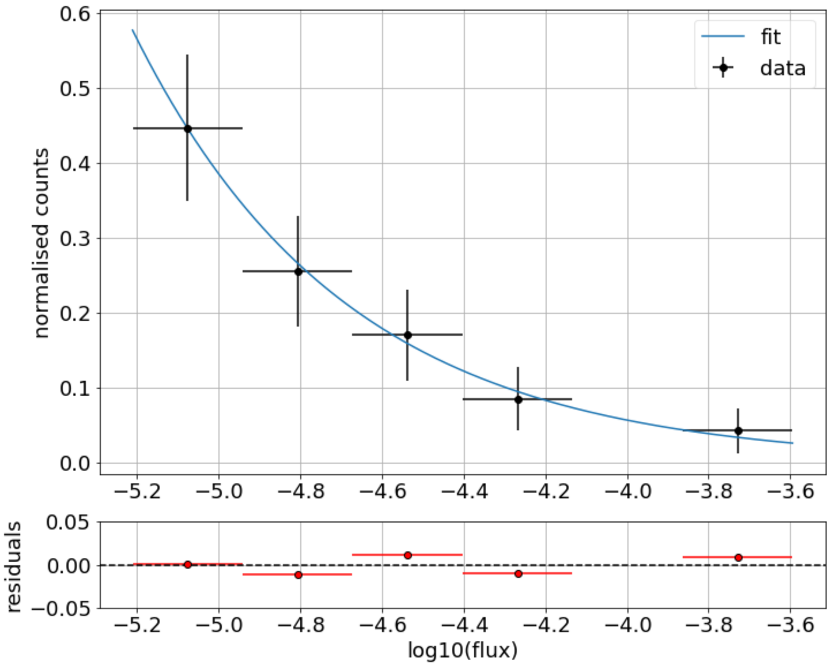

Figure 4.5 Gamma-ray bursts population of the second Fermi/LAT catalog, with photon flux above $6.6 \cdot 10^{-6}$ ph cm^{-2} s^{-1}. The distribution fit (continuous blue line) is achieved with an exponential law. Residuals are shown in the bottom panel.

4.5.1 Related Works

There are several examples of DL techniques used in the astrophysical context and more specifically, in gamma-ray astronomy. The work [Mar+20] uses the DNNs to detect the position of gamma-ray sources in sky maps. This work compares the results obtained with different CNN architectures. The CNNs are trained and tested on simulated datasets. These datasets are created with Fermitools[4] starting from the data presented in the Second Fermi gamma-ray sources catalogue. For the CNN developed during this PhD, a similar approach is used and described in section 4.5.2. The AGILE maps are simulated containing GRBs compatible with the GRBs described by the second Fermi/LAT GRB catalogue. This procedure is described in Sec. 4.4.1.

The source's properties taken by the Fermi GRB catalogue are convolved with the observational capability of AGILE to obtain simulations very close to the real data. The gamma-ray sky maps must be simulated because thou-

[4] https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/

sands of maps are needed to train the CNN, and so many images of real data are not always available. This PhD's work on DL techniques aims to develop a method that can be used to detect sources into sky maps and calculate the statistical significance of this detection. This method can provide results that can be communicated to the scientific community. The first phase of the work uses simulated datasets, but in the second phase, the CNN is validated using real AGILE/GRID data, confirming the obtained results. The CNN developed focuses on very short astrophysical events with a duration of seconds. These events are science alerts notified by other instruments through the GCN network. For this reason, the spatial research of the source is not important because the system knows in advance where the transient phenomena occurred.

The work [Car+18] uses CNNs to classify gamma-ray sky map of the galactic centre produced with the Fermi/LAT instrument. The results obtained with the CNNs are compared with standard techniques for the gamma-ray analysis to understand if it is possible to exploit these innovative technologies to improve gamma-ray data analysis. In all these related works, a big amount of data are simulated to test the CNN. This is a common practise when the real data are not enough to perform the training or are not balanced (the objects' classes are not present in the same quantity).

Furthermore, the gamma-ray observatories have their data format, and it is not possible to use real data taken from other observatories without a change in the data format. In other fields (e.g. optical observations), it is possible to use other instruments' images because the information's acquisition is the same. Usually, the astronomical projects have their own software to simulate data, for this work, the AGILE science tools are used, and this procedure is described in Sec. 4.5.2.

In [Car+18] the CNN is trained with simulated data and then tested on real data. This procedure was also used to verify the CNN developed during this PhD and is described in this chapter. Both related works use CNN to classify gamma-ray sky maps because this class of DNN was explicitly developed to work with images. The CNN developed for AGILE uses the same DNN architecture.

In the work [FKM20] the DL techniques are used to classify the gamma-ray sources of the last released Fermi/LAT catalogue ([Abd+20]). In this work, the authors did not use the CNN architecture but a Recurrent Neural Network (RNN), developed to work with time-series data because they don't use images as input but the sources' parameters. This example shows how the DL techniques can be used in other contexts from image classification. In

[FKM20], the researchers used the DL frameworks Keras[5] and Tensorflow[6], the same used during this PhD. In addition, the same evaluation methods cited in [FKM20] are used in this work. In fact, the CNN is evaluated using the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC) described in Sec. 4.5.4.

4.5.2 Datasets simulation

The dataset for the training, validation and test phases are generated with three Monte Carlo simulations. Each simulation created 40000 AGILE/GRID intensity sky maps. Intensity maps (obtained by dividing counts maps by exposure map) makes the CNN exposure-independent.

The first step is the counts maps simulation, and from these maps, the intensity maps are calculated. This simulation is performed using the BUILD25 of the AGILE Science Tools [Bul+19], which includes a sky simulator called AG_multisim . The background event filter called FM3.119, and the Instrument Response Functions (IRFs), called H0025, are used. The energy range used for these simulations is 100 MeV - 10 GeV. The simulator applies a Poisson-distributed noise to each pixel and produces each resulting counts map exactly as flight data. One of the inputs required for the simulation is the exposure map. The simulated maps have an integration time of 200 sec and size of 100×100 pixels with a bin size of 0.5°, i.e. $50° \times 50°$. The integration time is defined in Sec. 4.3.1. The datasets are labelled to train the CNN with a supervised learning procedure. The 200 sec exposure map (Fig. 4.6) is obtained from AGILE/GRID data, centring the map in the sky region defined before and searching for a time window with an exposure level similar to the mean level found in Sec. 4.3.1.

The Big Data (Terabytes) simulated during this phase and for the *p-value* evaluation (Sec. 4.6) require a dedicated software developed to execute these simulations in parallel and exploit all the available computing power. Two GPUs are used to speed-up the CNN training procedure. The parallel analyses are managed with Slurm, the same workload manager used for the RTApipe framework describe in Sec. 3.4.6. The data are managed in-memory by means of an iterative workflow to improve the simulations' speed. At each cycle of this workflow, only a bunch of data are analysed. This iterative workflow is the key feature to analyse this Big Data reducing the required time and space. The whole procedure is automated with several scrips that read configuration parameters from a configuration file. The results can be

[5] https://keras.io
[6] https://www.tensorflow.org

easily reproduced thanks to this automation.

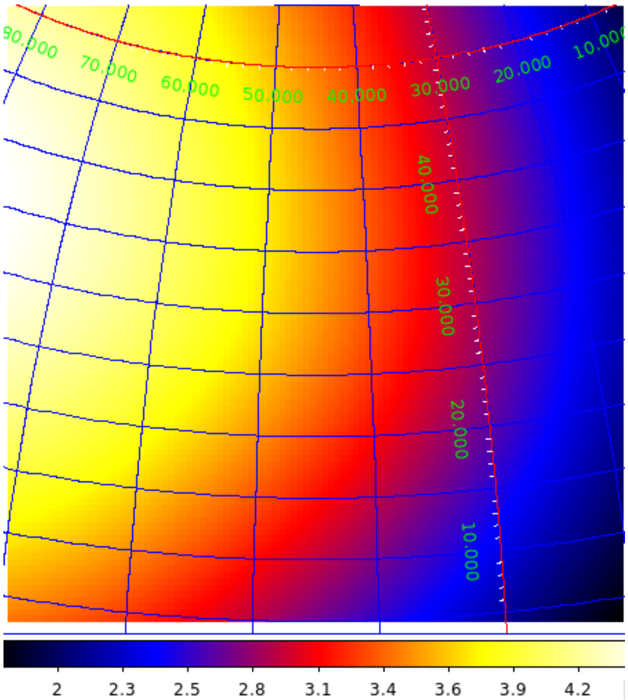

Figure 4.6 Exposure map used as input for the Monte Carlo simulations, expressed in $cm^2 ssr$. The map is represented in ARC projection and Galactic coordinates, with a bin size = 0.5°.

Only external science alerts with a maximum error localisation region of 1° are analysed in this work. This scenario covers more than 90% of GRBs presented in the Second Fermi/LAT GRBs catalogue and 100% of the GRBs reported in the Swift/BAT GRBs catalogue. For this reason, the GRB maps are simulated by placing a source in a random position with a radius lower than 1° from the maps' centre. The external alert error localisation region is assumed to be in the centre of the sky maps.

The background level for the simulations is obtained from the isotropic background distribution calculated in Sec. 4.3.2; no Galactic diffuse emission is considered.

The fluxes of the sources are randomly generated using the fit function described in Sec. 4.4.1. The minimum flux value for the maps with a GRB is defined to reach a 2σ significance over the background. The GRBs' fluxes and positions simulated with this method are thus compatible with real external science alerts. This approach simulates the datasets with the background levels and GRBs fluxes obtained from real data, improving the CNN's transfer learning from simulated datasets to real data.

The datasets contain half of the maps with a simulated GRB and the other half background only.

Fig. 4.7 shows the counts' distribution of simulated sky maps with a GRB (blue) and background only (red).

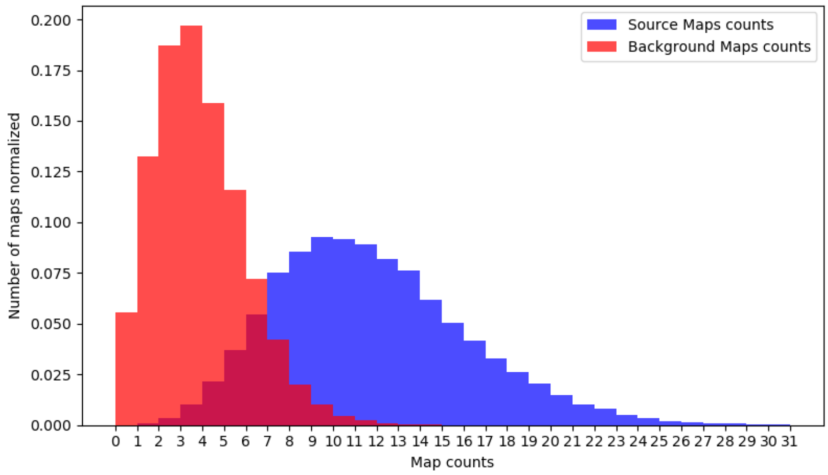

Figure 4.7 Histogram of the sum of photon counts inside the simulated counts maps. The red histogram for the background only maps, the blue histogram for the maps with a simulated GRB.

Image pre-processing

Fig. 4.8 shows an example of smoothed intensity maps simulated with the procedure described before. The datasets are processed performing a Gaussian smoothing of the intensity maps with a radius of 6°, assuming it as twice the value of the AGILE/GRID instrument PSF for this energy range.

Two 3D histograms are created by summing up all the intensity maps, pixel by pixel, to verify the counts' spatial distribution in the dataset maps. The X and Y axes refer to the map pixels reference system, and the Z-axis refers to the normalised value of the summed pixel's values from all maps in the dataset. Fig. 4.9 shows the histogram obtained from the intensity maps containing a simulated GRB. In this histogram, the peak in the centre of the represents the simulated GRBs. Fig. 4.10 shows the histogram for the background only maps.

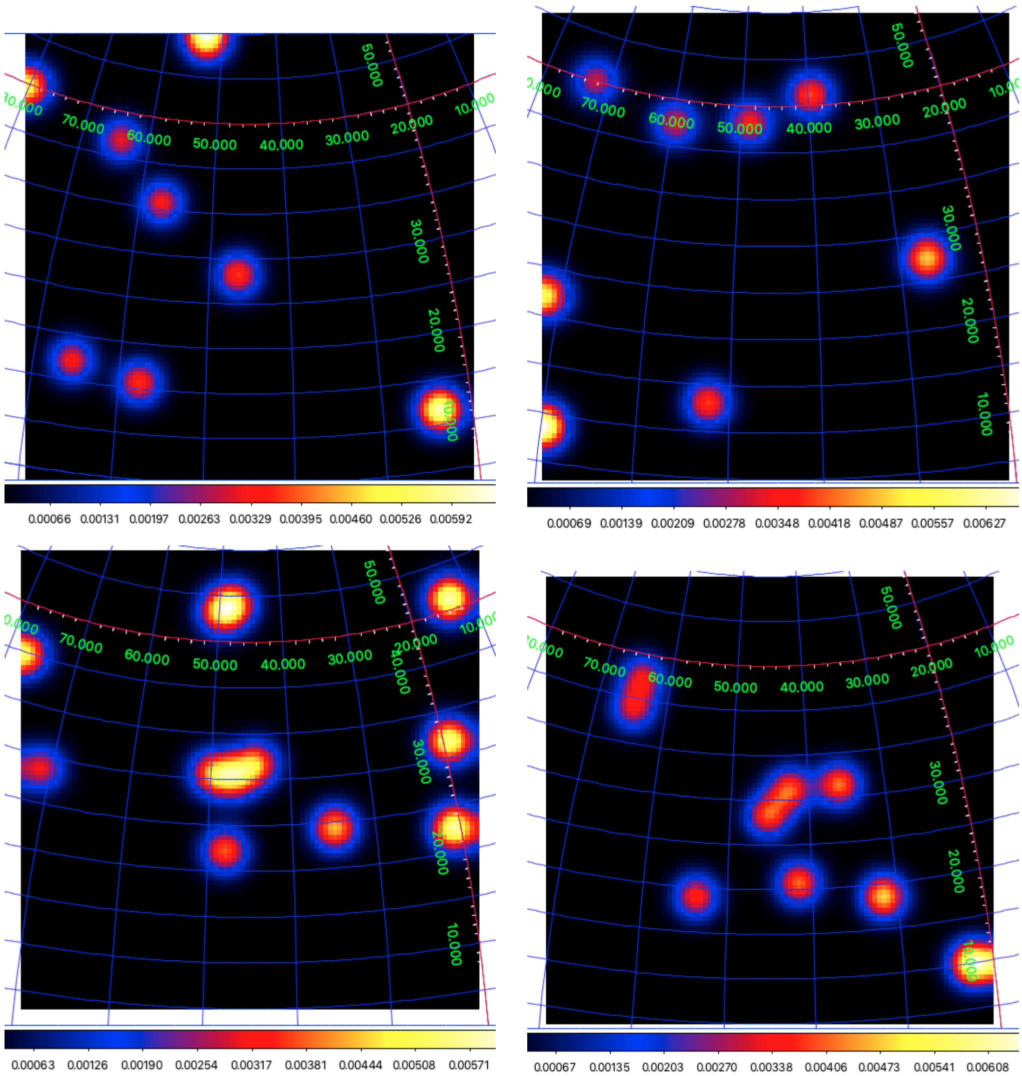

Figure 4.8 Smoothed intensity maps from the simulated dataset used to train the CNN. The top two images are background only. The bottom two images contain a simulated GRB. The maps are represented in ARC projection and Galactic coordinates, with a bin size = 0.5°.

4.5.3 CNN Architecture

More than seven hundred different parameterisations are tested on a separated validation dataset to find the best CNN architecture for the GBRs detection. In particular, each architecture changed for at least one of the following parameters:

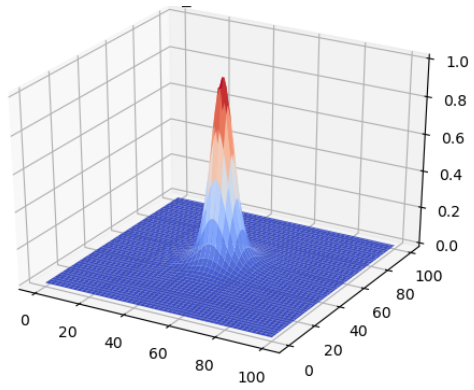

Figure 4.9 The 3D histogram obtained summing all the counts of the smoothed maps of the dataset with a GRB. X and Y axes represent the pixels of the maps while the Z axis represents the normalized summed counts.

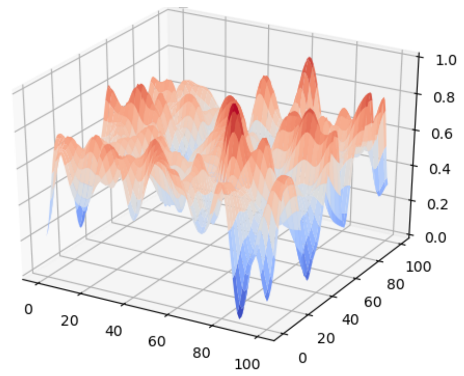

Figure 4.10 The 3D histogram obtained summing the smoothed counts maps of the background only dataset. X and Y axes represent the maps pixels while the Z axis represents the normalized summed counts.

- Epochs number
- Number of convolutional layers
- Number of filters in convolutional layers
- Number of dense layers
- Batch size
- Learning rate

After this evaluation, it is possible to note that increasing the neural network's complexity does not always lead to better results. This behaviour usually happens when working with real-world problems. The final CNN, displayed[7] in fig. 4.11, is chosen according to the best trade-off between training time and validation performance. This network is composed of ten layers, and it is implemented using two open-source frameworks, Keras running on top of Tensorflow.

The final architecture is defined as follows. The first layer receives an array of maps, each with a size of 100 × 100 pixels, then a Convolution2D layer with twenty filters is applied. The CNN uses filters with a kernel size

[7] http://alexlenail.me/NN-SVG/LeNet.html

of 12 × 12 pixels to identify features within the AGILE/GRID simulated intensity maps. The size of these filters is defined starting from the PSF of the AGILE/GRID instrument, described in [Bul+12]. A 12×12 pixels kernel size with 0.5°-side pixels is used to cover an area ($6° \times 6°$) approximately equal to twice the value of the AGILE/GRID PSF in this energy-range. The next layer consists of a MaxPooling2D operation with a kernel size of 2×2 pixels. This layer aims to reduce the size of the image to reduce the computations that must be performed in the subsequent layers.

Three additional Convolution2D layers and a Max-Pooling2D layer are applied to find new features and reduce the image size. At this point, a Dropout layer with the probability of 25% is applied as a regularisation technique to prevent overfitting. The model also implements a Dense layer that flats the 2D tensors in a single dimension array of 1000 elements and a Dropout layer with a probability of 50%. Finally, a two-neurons Dense layer is used with a Softmax activation function that provides the predicted probabilities of the two classes of intensity maps: background or GRB. The CNN's output value, here on called CV, is a number between 0 and 1. If the CNN has an output $CV = 0$, it means that the map is classified background only with 100% probability. Conversely, if the CNN has an output $CV = 1$, it means that the map is classified containing a GRB with 100% probability. Usually, the CVs are numbers between these two opposite situations, and the 0.5 value is the standard threshold between the two classifications.

All the convolutional layers use ReLu activation functions. This activation function improves computing efficiency during the training and has a good gradient propagation. The Keras initialiser used to set the initial random weights is the Variance Scaling Initialiser, while the biases are initialised to zero. All the analyses are performed using Python 3.6 on an NVIDIA Tesla K80 GPU.

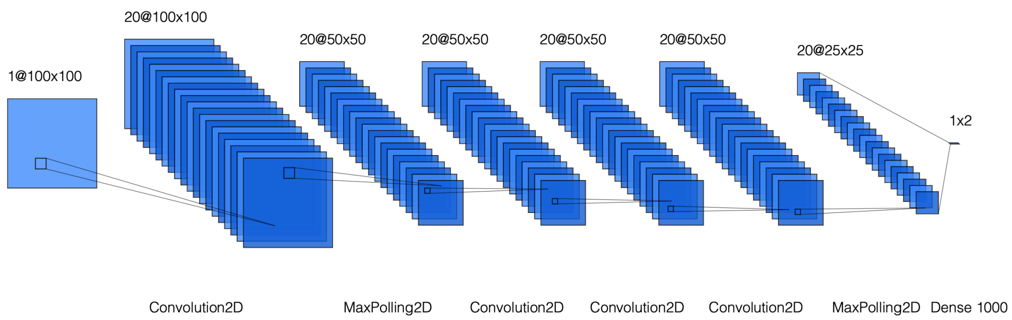

Figure 4.11 Schema of the CNN architecture created with a graphical tool.

4.5.4 CNN Training and Testing

With the previous analysis, the hyperparameters of the CNN are fixed. The final training is performed using a batch size of 200 maps, and the model achieves convergence after five epochs (Fig. 4.12). As loss function, the CNN implements the Sparse Categorical Crossentropy and uses the Adam optimiser with a learning rate of 0.001. The Accuracy and AUC are considered as performance metrics for the CNN. The accuracy is the percentage of input maps that the CNN classifies correctly with respect to all the maps tested using the model. Instead, the AUC is calculated with respect to the ROC curve, which is a graphical plot that shows the discrimination ability of a binary classifier, in this case, the CNN. The ROC curve's X-axis represents the False Positive Rate (FPR), while the Y-axis represents the True Positive Rate (TPR). Thus, the AUC of the ROC curve is a metric of the classifier that considers the overall possible thresholds probabilities to define the final output label.

Fig. 4.12 shows the Accuracy and the Loss for training and test datasets. As expected, the training accuracy increases with the number of epochs, while the opposite behaviour is observed when considering the training loss function. This means that the model gradually learns how to classify the maps correctly. The final model provides a test accuracy of 98.2%, which means that the model correctly classifies the 98.2% of unseen maps, and it performs accurately in both classes, GRB and background. The AUC calculated with the ROC Curve is equal to 0.997. This value is very close to 1 and indicates a high separability level reached by the CNN model.

The CNN reaches a value of accuracy similar to the optimal one after the first epoch of the training. The accuracy improves of small quantities in the following epochs. The CNN has this behaviour because the 40 000 gamma-ray sky maps that compose the training dataset are enough to train the CNN how to classify maps with and without a GRB. The CNN can learn from this dataset all the information required with few epochs. Many different CNN architectures with more layers are tested, but these additional layers are not enough to obtain better results. The additional layers increase the training time, and for this reason, the CNN architecture with best-trade off between training time and results is selected.

From a technical point of view, these good results obtained with a low number of epochs, indicate that the problem is relatively simple to solve for the CNN.

However, this work represents a new approach for AGILE/GRID GBRs detection and, as shown in Sec. 4.7, the use of a CNN in the context of an automated pipeline lead to significant improvements compared with the

Li&Ma approach currently in use by the AGILE Team.

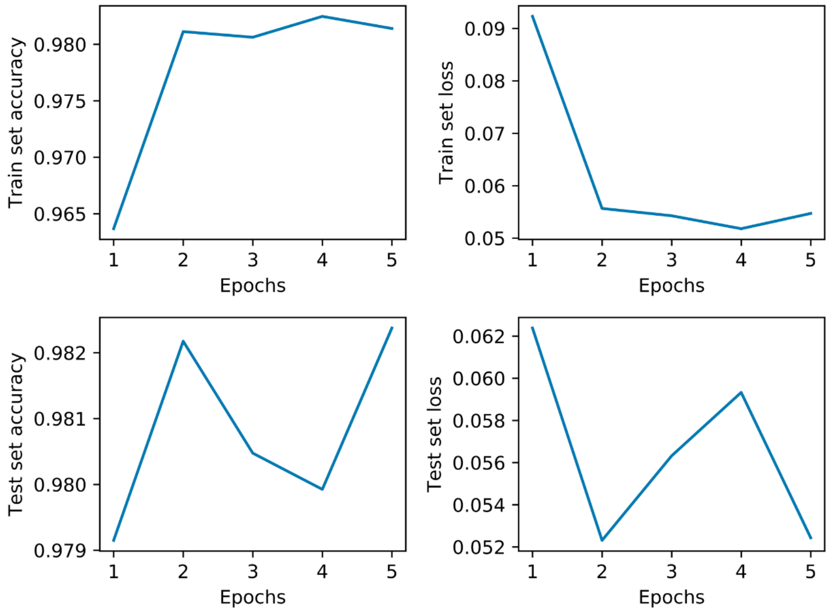

Figure 4.12 Accuracy and Loss values obtained during the five epochs training of the CNN, with both train and test datasets.

4.6 CNN *p-value* evaluation

This CNN network is developed to work as part of the AGILE/GRID real-time analysis pipeline for detecting GRBs into the sky maps starting from external scientific alerts received from other projects. If the CNN detects a GRB counterpart, the AGILE Team can communicate this detection to the community through the GCN network.

The CNN's output value (CV) cannot be used directly to determine the significance level of a GRB detection. The threshold of 0.5 on the CV means that some maps with background-only can be classified as GRB for statistical fluctuation of the background. Thus, a better threshold on the CVs must be calculated to achieve the requested statistical significance. An evaluation using background only maps to determine the *p-value* of the CNN is performed to find different levels of CV thresholds. The p-value is the probability of finding the observed, or better, results when the null hypothesis is true. In this case, where the CNN is searching GRB into the sky map, the null hypothesis is background only map without a GRB inside.

A similar approach used to calculate the *p-value* for a MLE method is described in detail in [Bul+12]. The *p-value* distributions are calculated using the CNN output values obtained with background only maps in different conditions.

The distribution Φ of the CV values resulting from the CNN analysis procedure on background only simulated maps with a defined level of background is calculated to evaluate the *p-value* . The probability that the result of an analysis in an empty field has $CV \geq h$ (that is the complement of the cumulative distribution function) is:

$$P(CV \geq h) = \int_{h}^{+\infty} \Phi(x)dx \tag{4.1}$$

which is also called the *p-value* $p = P(CV \geq h)$ and defines the probability of obtaining that value or larger when the null hypothesis is true.

4.6.1 *P-value* determination for different background conditions

The *p-value* distribution is strongly affected by the background level. Different *p-value* distributions are calculated to determine the statistical significance of a CNN detection in different background conditions, allowing this method to be applied on real AGILE/GRID observations.

Three different background levels from background distribution defined in Sec. 4.3.2 are selected and reported in Fig. 4.3: the mean level ($g_{iso} = 10.4 \times 10^{-5}$ph cm^{-2} s^{-1}) and two 1σ deviations adding or subtracting the standard deviation of 3.0×10^{-5}ph cm^{-2} s^{-1}, obtaining $g_{iso} = 7.4 \times 10^{-5}$ph cm^{-2} s^{-1} and $g_{iso} = 13.4 \times 10^{-5}$ph cm^{-2} s^{-1}.

One dataset of ten million background only maps is simulated for each background level. The maps are simulated using the parameters and the same observational model used to create the training dataset and described in the Sec. 4.5.2, and analysed with the trained CNN. Using the CNN classification results obtained from this procedure, a *p-value* distribution is calculated for each of the observational conditions defined. The *p-value* distribution of CV values is shown in Fig. 4.13. Fig. 4.14 shows a zoom of the *p-value* distribution when $CV >= 0.9$ to better understand the *p-value* behaviour at high significance.

During this PhD, the number of different observing conditions is limited for constraints on computing power and time, but more *p-value* analyses are planned for the future to improve the accuracy of this method. The exposure level is fixed because this method evaluates intensity maps, which are exposure independent. The exposure level for this simulation is fixed at $44 \times 10^3 [cm^2 s]$.

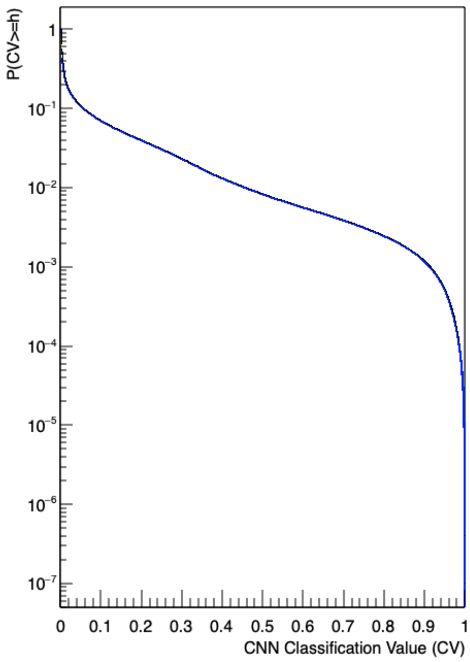

Figure 4.13 The *p-value* distribution of the CV values with a background level of $g_{iso} = 10.4 \times 10^{-5}$ph cm^{-2} s^{-1}.

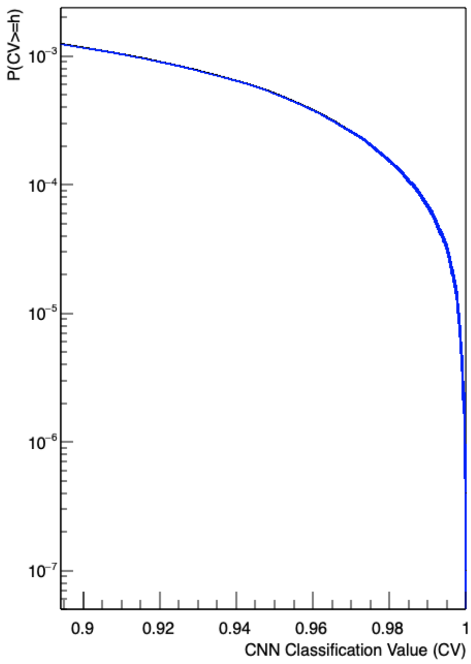

Figure 4.14 Zoom of the *p-value* distribution of the CV values shown in Fig. 4.13.

Table 4.1 reports the thresholds of the CNN classification values (CV) written as (1-CV) in relation to different σ levels for different background conditions that are defined before. The algorithm thresholds depend on the background level. This behaviour is expected, given that the detection of a GRB depends strongly on the background conditions. The results reach a maximum significance level of 5σ due to constraints on computing power and time. The fitting function between the three p-value distributions is calculated. With this fitting function, it is possible to estimate the CV threshold for different g_{iso} values.

4.7 AGILE/GRID GRB search and results

The trained CNN is tested with real GRBs and real AGILE/GRID data. The GRB catalogues of Swift/BAT, Fermi/LAT, and Fermi/GBM are used to perform these tests.

Table 4.1 Relation between σ and CNN Classification Value threshold for different observing conditions.

σ	*p-value*	$g_{iso} = 7.4$	$g_{iso} = 10.4$	$g_{iso} = 13.4$
3	1.35×10^{-3}	2.4×10^{-1}	1.1×10^{-1}	6.5×10^{-2}
3.5	2.32×10^{-4}	5.5×10^{-2}	2.7×10^{-2}	1.4×10^{-2}
4	3.17×10^{-5}	1.5×10^{-2}	4.9×10^{-3}	2.0×10^{-3}
4.5	3.40×10^{-6}	2.2×10^{-3}	8.9×10^{-4}	2.8×10^{-4}
5	2.86×10^{-7}	2.1×10^{-4}	3.5×10^{-5}	2.5×10^{-5}

This table shows the CNN values thresholds as (1-CV) for different statistical significance levels. These thresholds are calculated in different background and exposure conditions defined in Sec. 4.6. The g_{iso} are expressed in 10^{-5}ph cm^{-2} s^{-1}.

AGILE/GRID intensity maps are generated for each GRB contained in these catalogues using the GRB trigger time and the centre of the error localisation region defined inside the catalogues. The integration time and the map size are fixed and are defined in Sec. 4.5.2. This analysis is performed using the consolidated AGILE/GRID data archive. This archive covers a time window that starts on 1st Jan 2010 and ends on 30th Nov 2019. A list of 193 GRBs is obtained from these catalogues after applying four filters: (i) the AGILE/GRID map with 200 sec of integration starting from the GRB trigger time must have an exposure value greater than the minimum value of $20 \times 10^3 [cm^2 s]$, fixed in Sec. 4.3.1, (ii) the localisation error region radius of the GRB must be $<= 1°$, (iii) the GRB trigger time must be inside the AGILE/GRID consolidated archive time window, and (iiii) the GRBs must be extra-Galactic, so GRBs with $|b| < 10$ are excluded. From these 193 GRBs, the CNN detected 21 GRBs with $\sigma \geq 3$. The AGILE/GRID can detect not all GRBs detected by other observatories due to the different energy-range or the lower AGILE/GRID sensitivity.

Table 4.2 reports the detected GRBs' list, including the statistical significance of the CNN detection. The statistical significance is obtained with the procedure described in Sec. 4.6.1. The background level (g_{iso}), used to determine the right *p-value* distribution, is evaluated in a time window (T_{off}) preceding the GRB trigger time, starting from 6 hours and expanding it until at least ten counts are found in a radius of 10° from the GRB position. The background level (g_{iso}) is evaluated using an MLE. The number of ten counts is required to perform the MLE analysis. The significance is then calculated with g_{iso} and CV values using the fitting function for the p-value threshold defined in Sec. 4.6.1.

Table 4.2 also reports the detection with $\sigma \geq 3$ calculated using the Li&Ma method on the same list of 193 GRBs and with the same On-Off parameters used for the CNN analysis. The Li&Ma is a likelihood ratio method applied to aperture photometry. It is largely used in gamma-ray astronomy and by the AGILE Team as the standard analysis for GRB detection. The Li&Ma method applied to the same GRBs and with the same parameters can detect with a significance greater than 3σ only two GRBs. This comparison reveals that in this context of an automated pipeline with fixed parameters, the CNN detects more GRBs counterparts than the Li&Ma algorithm.

The N_{on} values are the number of photons inside a $10°$ area from the centre of the GRB's alert error localisation region. As already stated in Sec. 1, the Li&Ma method requires at least ten counts to be applied. Using this threshold for both Li&Ma and CNN results will lead to the same detection number, but the CNN has not this requirement because it uses different criteria to analyse the data, like the spatial distribution of the photons inside the intensity maps. For this reason, the CNN can detect GRBs even when the counts of photons are less than ten. The CNN is more flexible and more suitable in this context of short time detection with few photons.

This CNN method developed during this PhD in collaboration with the AGILE researchers improves the AGILE/GRID detection capabilities from the currently used Li&Ma method. This CNN can be implemented into the real-time analysis pipeline of AGILE. Future works are planned to use this CNN with the AGILE/GRID archive data searching for GRBs that have not been detected with the Li&Ma and for this reason, are still undetected.

Table 4.2 List of GRBs detected with CNN and Li&Ma from catalogs.

GRB	CNN T_{on}	CNN N_{on}	CNN sigma	Li&Ma T_{on}	Li&Ma N_{on}	Li&Ma sigma
100724B	200	9	5			
110530A	200	3	3			
120711A	200	2	3			
121202A	200	2	3.5			
130427A	200	5	5			
130518A	200	2	3.5			
130828A	200	5	5			
131108A	200	11	5	200	11	4.4
141012A	200	3	4			
141028A	200	4	3			
160325A	200	4	4			
160804A	200	2	3.5			
160912A	200	8	4			
170115B	200	4	4			
170127C	200	3	4.5			
170522A	200	5	5			
170710B	200	4	3.5			
180418A	200	2	3.5			
180720B	200	13	5	200	13	4.7
190324B	200	5	3			
190530A	200	6	4.5			

[Note]: This table shows the comparison between the results obtained with the CNN and the Li&Ma methods searching for GRBs in AGILE-GRID data starting from GRBs catalogs. The T_{on} is expressed in seconds. N_{on} indicates the number of counts in a radius of $10°$ from the center of the GRB error localization region.

Chapter 5

Conclusions and Future Works

Several research activities were carried on during the PhD in collaboration with researchers of national and international gamma-ray observatories (AGILE space mission, Cherenkov Telescope Array, and ASTRI Mini Array). This section describes the conclusions on these activities and future works.

As a general overview, the thesis work can be divided into two parts. The first part (Sec. 2) describes the data analysis and management framework of the AGILE space mission archive to produce the second gamma-ray catalogue of the sources detected with the AGILE/GRID instrument. The second part (Sec. 3 and 4) deals with the real-time analysis of gamma-ray data acquired by observers and for the management of external scientific alerts. A science alert is an urgent communication from one observatory to other observatories that an interesting astrophysical event is occurring in the sky.

The Sec. 2 describes the software developed during this PhD to support the AGILE Team's to perform the analyses for the Second AGILE/GRID catalogue of gamma-ray sources [Bul+19]. This catalogue contains all the gamma-ray sources detected with the AGILE/GRID instrument during the first two years of the AGILE space mission during the so-called 'pointing mode'. The software developed for the realisation of this catalogue is divided into different components. The first part of the software that was developed includes scripts used to automatise the workflow and aggregate the analysis results. The second part is a MySQL database used to store all the results obtained with the analyses and the scripts that import the results into the database. The last part is a Graphical User Interface (GUI) that allow the researchers to visualise the results with plots and tables.

In Sec. 3 it is described the RTApipe framework designed to help researchers develop real-time analysis pipelines for the gamma-ray observatories. This framework has several key features that satisfy the gamma-ray observatory requirements in the multi-wavelength and multi-messenger era. In this context, the observatories that acquire information with different electromagnetic wavelength or other messenger signals (electromagnetic radiations, neutrinos and gravitational waves) share their results about transient events sending science alerts through the Gamma-ray Coordinates Network (GCN). The sharing of these science alerts allows the observatories' automated software to react rapidly to these alerts searching for a counterpart in the observatory incoming or archive data. This workflow can be used to study the same astrophysical events with the results obtained from different observatories observing the same event through many messenger signals. The in-depth study of these events is possible due to this information sharing.

The Sec. 3 also describes the prototypes of real-time analysis pipeline developed during this PhD for the Science Alert Generation (SAG) of the Cherenkov Telescope Array (CTA) and for the science analysis pipeline of the ASTRI Horn telescope as a feasibility study for the ASTRI Mini Array. These two gamma-ray observatories are in the construction phase and the INAF/OAS Bologna, where this PhD was carried on, is leading several working groups within these projects.

In more details, three real-time analysis pipelines for the AGILE space mission were developed using the RTApipe framework in collaboration with the AGILE Team. This software is described in Sec. 3.6.1. These pipelines aim to react to external science alerts and analyse the instruments' data acquired onboard AGILE as soon as they are available searching for gamma-ray transients. The AGILE researchers use these pipelines' results during the daily work to check if new transient events are detected. The Sec. 3.6.1 also shows the most important scientific results that the AGILE researchers reached using this software as a support tool and then published on scientific journals or shared with the scientific community through networks such as the Astronomer Telegrams or the GCN network.

The AGILE real-time analysis pipeline performs several analyses using different Science Tools. The Li&Ma, described in Sec. 4.1, is the current standard method for this pipeline to analyse the data acquired by the AGILE/GRID instrument searching for transient events in a short term scale (seconds). The last activity of this PhD was to study a new Gamma-ray Bursts (GRBs) detection method based on Deep Learning technology. This method has the goal to improve the detection capability of this AGILE real-time analysis pipelines. The results obtained on real AGILE data with the new method were compared with those obtained using the standard Li&Ma method on the same GRBs; this comparison highlighted the improvement obtained with the new techniques based on Deep Learning.

The Sec. 4 describes the workflow implemented to train and evaluate a Convolutional Neural Network (CNN), a particular Deep Learning architecture, for the detection of GRBs within the AGILE/GRID gamma-ray sky maps in the energy range 100 MeV - 10 GeV. This study aims to improve the AGILE real-time analysis pipeline's detection capability that reacts to external science alerts. Three datasets of 40000 gamma-ray sky maps are simulated to train this CNN, half of the maps contain background-events only while the other half contains a GRB. The simulation of these datasets is performed using the AGILE Science Tools. The parameters used for these simulations are defined studying the AGILE satellite's observing pattern and the characteristics of the GRBs that can be detected by the AGILE/GRID instrument using the Second catalogue of GRBs detected by Fermi/LAT as the baseline. One of these three datasets (the validation dataset) is used to test different CNN architecture varying the CNN's hyperparameters (e.g. number of layers, number of neurons etc.). Once found the best CNN architecture that optimises the trade-off between the training time and performances, the final training was performed using the train and test datasets. The CNN obtained great results during the training phase with an accuracy on the test dataset of 98.2%. To calculate the statistical significance of the detections done with this CNN the p-value of the CNN is calculated in different background conditions simulating ten millions of sky maps for each condition. The CNN is then evaluated with real AGILE/GRID archive data using the GRBs presented within the Fermi/LAT, Fermi/GBM, and Swift/BAT catalogues as external science alerts. The results obtained shows that the CNN can detect 21 GRBs from the list of 192 GRBs selected from these catalogues and compatible with the AGILE/GRID instrument. These 192 GRBs are selected applying four filters to obtain GRBs visible for the AGILE instruments, with an error localisation radius $<= 1°$, a trigger time within the AGILE/GRID data archive time widows and an extra-Galactic position. On the opposite, the standard analysis method used by the AGILE Team (Li&Ma) detected just 2 GRBs from the same list of events. Not all the GRBs can be detected by AGILE/GRID because different observatories have different observing conditions and performances. The results confirm that the CNN improves the GRBs detection capability into the AGILE/GRID sky maps. Thus, this CNN can be implemented within the AGILE real-time analysis pipeline.

5.1 Future Works

The future works can be divided into three main directions.

The first future concerns the new analysis method based on Deep Learning; the CNN developed during this PhD and described in Sec. 4 will be used in the AGILE real-time analysis pipeline. It is planned to increase the detection capability of this CNN calculating additional p-value distribution in more background conditions. This was not possible during this PhD duo time and computing constraints.

The CNN can be updated to perform a blind search of gamma-ray sources within sky maps without any suggestion about the source's position. This new version of the CNN could be used as Science Tool of the real-time analysis pipeline for the CTA and ASTRI Mini Array observatories. These observatories will produce gamma-ray sky maps very close to the AGILE/GRID sky maps, and for this reason, the same method used to train the CNN for AGILE can be used.

In collaboration with AGILE researchers, this CNN will be used to search for GRBs into the AGILE/GRID data archive starting from the GRBs catalogues of AGILE/MCAL or performing a blind search. This analysis aims to detect new GRBs that have not been detected with standard methods used before.

Another future work concerns the analysis pipelines developed during this PhD. The real-time analysis pipelines developed for AGILE, described in Sec. 3.6 will be updated with new Science Tools (e.g. the CNN) to reduce the reaction time to generate science alerts. In the coming years, the AGILE real-time analysis pipelines will adapt and evolve together with the scientific context as they have been developed with the RTApipe framework.

The prototypes of real-time analysis pipelines developed during this PhD and described in Sec. 3.7, 3.8 will be used for the development of the production software for the CTA and ASTRI Mini Array observatories.

The real-time analysis for CTA part of the SAG system will be based on the RTApipe framework described in Sec. 3. The list of requirements used to develop this framework includes the SAG system's requirements. For this reason, the framework is compliant with these requirements. The prototype developed during this PhD will be completed in future activities to include more scientific use cases for the scientific analysis performed during the CTA observations. The prototype can be used to test both the scientific and computation performances of the Science Tools available for the gamma-ray analysis (e.g. ctools[1], Gammapy[2] and more), considering the SAG twenty seconds constraint to produce candidate science alerts since the input data arrives.

The software for the ASTRI Mini Array's scientific pipeline will be developed starting from the prototype developed during this PhD. In the next few years, this observatory will be built and will begin to acquire data. The raw data acquired will be transferred from the site where telescopes are installed to the Archive System sited at INAF/OAR Roma. Once the data are available into the Archive System, an automated pipeline will start the reconstruction from the raw data to a data format suitable for scientific analysis. Then the pipeline will execute scientific analyses. The science users will be able to visualise these results as soon as they are produced using a web GUI. All the work performed during this PhD can be used as a starting point to complete these software packages for the ASTRI Mini Array observatory.

Concerning the last future work, it is under evaluation by the AGILE Team the possibility to open the 2AGL catalogue web GUI described in Sec. 2.3 to the public. This will require a cleaning of the interfaces to reduce the quantity of information visualised in the GUI. If this work is approved by the

[1]http://cta.irap.omp.eu/ctools/
[2]https://gammapy.org

AGILE Team, the external users will be able to visualise the details of the 2AGL sources or to download the data for personal analysis. This work will be carried on in collaboration with the ASI/SSDC Team that is in charge of the Italian space missions like AGILE public data management.

Bibliography

[Aar+17] M. G. Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.3 (Mar. 2017), P03012. DOI: 10.1088/1748-0221/12/03/P03012. arXiv: 1612.05093 [astro-ph.IM].

[Abb+09] B P Abbott et al. "LIGO: the Laser Interferometer Gravitational-Wave Observatory". In: *Reports on Progress in Physics* 72.7 (2009), p. 076901. DOI: 10.1088/0034-4885/72/7/076901. URL: https://doi.org/10.1088/0034-4885/72/7/076901.

[Abb+17a] B. P. Abbott et al. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". In: Phys. Rev. Lett. 119.16, 161101 (Oct. 2017), p. 161101. DOI: 10.1103/PhysRevLett.119.161101. arXiv: 1710.05832 [gr-qc].

[Abb+17b] B. P. Abbott et al. "Multi-messenger Observations of a Binary Neutron Star Merger". In: ApJ 848.2, L12 (Oct. 2017), p. L12. DOI: 10.3847/2041-8213/aa91c9. arXiv: 1710.05833 [astro-ph.HE].

[Abd+20] S. Abdollahi et al. "Fermi Large Area Telescope Fourth Source Catalog". In: ApJS 247.1, 33 (Mar. 2020), p. 33. DOI: 10.3847/1538-4365/ab6bcb. arXiv: 1902.10045 [astro-ph.HE].

[Acc+20] V. A. Acciari, M. I. Bernardos, E. Colombo, J. L. Contreras, J. Cortina, A. De Angelis, C. Delgado, C. Díaz, D. Fink, M. Mariotti, S. Mangano, R. Mirzoyan, M. Polo, T. Schweizer, and M. Will. "Optical intensity interferometry observations using the MAGIC Imaging Atmospheric Cherenkov Telescopes". In: MNRAS 491.2 (Jan. 2020), pp. 1540–1547. DOI: 10.1093/mnras/stz3171. arXiv: 1911.06029 [astro-ph.IM].

[Ack+12] M. Ackermann et al. "THE FERMI LARGE AREA TELESCOPE ON ORBIT: EVENT CLASSIFICATION, INSTRUMENT RESPONSE FUNCTIONS, AND CALIBRATION".

In: *The Astrophysical Journal Supplement Series* 203.1 (2012), p. 4. DOI: 10.1088/0067-0049/203/1/4. URL: https://doi.org/10.1088/0067-0049/203/1/4.

[Ack+13] M. Ackermann et al. "The First Fermi-LAT Gamma-Ray Burst Catalog". In: ApJS 209.1, 11 (Nov. 2013), p. 11. DOI: 10.1088/0067-0049/209/1/11. arXiv: 1303.2908 [astro-ph.HE].

[Aje+18] M. Ajello et al. "Fermi-LAT Observations of LIGO/Virgo Event GW170817". In: ApJ 861.2, 85 (July 2018), p. 85. DOI: 10.3847/1538-4357/aac515.

[Aje+19] M. Ajello et al. "A Decade of Gamma-Ray Bursts Observed by Fermi-LAT: The Second GRB Catalog". In: ApJ 878.1, 52 (June 2019), p. 52. DOI: 10.3847/1538-4357/ab1d4e. arXiv: 1906.11403 [astro-ph.HE].

[Ban+09] D. L. Band et al. "Prospects for GRB Science with the Fermi Large Area Telescope". In: ApJ 701.2 (Aug. 2009), pp. 1673–1694. DOI: 10.1088/0004-637X/701/2/1673. arXiv: 0906.0991 [astro-ph.HE].

[Bar+00] S. D. Barthelmy, T. L. Cline, P. Butterworth, R. M. Kippen, M. S. Briggs, V. Connaughton, and G. N. Pendleton. "GRB Coordinates Network (GCN): A status report". In: *AIP Conference Proceedings* 526.1 (2000), pp. 731–735. DOI: 10.1063/1.1361631. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.1361631. URL: https://aip.scitation.org/doi/abs/10.1063/1.1361631.

[Bar+01] G. Barbiellini et al. "The AGILE scientific instrument". In: *Gamma 2001: Gamma-Ray Astrophysics*. Ed. by Steven Ritz, Neil Gehrels, and Chris R. Shrader. Vol. 587. American Institute of Physics Conference Series. Oct. 2001, pp. 774–778. DOI: 10.1063/1.1419498.

[Bar+05] Scott D. Barthelmy, Louis M. Barbier, Jay R. Cummings, Ed E. Fenimore, Neil Gehrels, Derek Hullinger, Hans A. Krimm, Craig B. Markwardt, David M. Palmer, Ann Parsons, Goro Sato, Masaya Suzuki, Tadayuki Takahashi, Makota Tashiro, and Jack Tueller. "The Burst Alert Telescope (BAT) on the SWIFT Midex Mission". In: Space Sci. Rev. 120.3-4 (Oct. 2005), pp. 143–164. DOI: 10.1007/s11214-005-5096-3. arXiv: astro-ph/0507410 [astro-ph].

[Bar19] Dalya Baron. "Machine Learning in Astronomy: a practical overview". In: *arXiv e-prints*, arXiv:1904.07248 (Apr. 2019), arXiv:1904.07248. arXiv: 1904.07248 [astro-ph.IM].

[Ber+19] A. Berti et al. "Following up Transient sources at Very High Energies with MAGIC". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. International Cosmic Ray Conference. July 2019, p. 633. arXiv: 1909.02798 [astro-ph.HE].

[BL18] E. Bissaldi and F. Longo. "GRB 180914B: Fermi-LAT detection." In: *GRB Coordinates Network* 23232 (Jan. 2018), p. 1.

[BLF19] E. Bissaldi, F. Longo, and Fermi-LAT Team. "GRB 190606A: Fermi-LAT detection". In: *GRB Coordinates Network* 24761 (June 2019), p. 1.

[Bul+10] A. Bulgarelli et al. "The AGILE silicon tracker: Pre-launch and in-flight configuration". In: *Nuclear Instruments and Methods in Physics Research A* 614.2 (Mar. 2010), pp. 213–226. DOI: 10.1016/j.nima.2009.12.051.

[Bul+12] A. Bulgarelli, A. W. Chen, M. Tavani, F. Gianotti, M. Trifoglio, and T. Contessi. "Evaluating the maximum likelihood method for detecting short-term variability of AGILE γ-ray sources". In: A&A 540, A79 (Apr. 2012), A79. DOI: 10.1051/0004-6361/201118023. arXiv: 1201.2602 [astro-ph.IM].

[Bul+13] A. Bulgarelli et al. "The Real-Time Analysis of Cherenkov Telescope Array Observatory". In: *International Cosmic Ray Conference*. Vol. 33. International Cosmic Ray Conference. Jan. 2013, p. 3099. arXiv: 1307.6489 [astro-ph.IM].

[Bul+14] A. Bulgarelli et al. "The AGILE Alert System for Gamma-Ray Transients". In: ApJ 781.1, 19 (Jan. 2014), p. 19. DOI: 10.1088/0004-637X/781/1/19. arXiv: 1401.3573 [astro-ph.IM].

[Bul+15] A. Bulgarelli et al. "The On-Site Analysis of the Cherenkov Telescope Array". In: *34th International Cosmic Ray Conference (ICRC2015)*. Vol. 34. International Cosmic Ray Conference. July 2015, p. 763. arXiv: 1509.01963 [astro-ph.IM].

[Bul+19] A. Bulgarelli et al. "Second AGILE catalogue of gamma-ray sources". In: A&A 627, A13 (July 2019), A13. DOI: 10.1051/0004-6361/201834143. arXiv: 1903.06957 [astro-ph.HE].

[Bul19] Andrea Bulgarelli. "The AGILE Gamma-Ray observatory: software and pipelines". In: *Experimental Astronomy* 48.2-3 (Dec. 2019), pp. 199–231. DOI: 10.1007/s10686-019-09644-w.

[Car+18] Sascha Caron, Germán A. Gómez-Vargas, Luc Hendriks, and Roberto Ruiz de Austri. "Analyzing γ rays of the Galactic Center with deep learning". In: J. Cosmology Astropart. Phys. 2018.5, 058 (May 2018), p. 058. DOI: 10.1088/1475-7516/2018/05/058. arXiv: 1708.06706 [astro-ph.HE].

[Cas+20] C. Casentini et al. "AGILE Observations of Two Repeating Fast Radio Bursts with Low Intrinsic Dispersion Measures". In: ApJ 890.2, L32 (Feb. 2020), p. L32. DOI: 10.3847/2041-8213/ab720a. arXiv: 1911.10189 [astro-ph.HE].

[Cat+11] P. W. Cattaneo et al. "First results about on-ground calibration of the silicon tracker for the AGILE satellite". In: *Nuclear Instruments and Methods in Physics Research A* 630.1 (Feb. 2011), pp. 251–257. DOI: 10.1016/j.nima.2010.06.078. arXiv: 1112.2600 [astro-ph.IM].

[Cat+18] Osvaldo Catalano et al. "The ASTRI camera for the Cherenkov Telescope Array". In: *Ground-based and Airborne Instrumentation for Astronomy VII*. Ed. by Christopher J. Evans, Luc Simard, and Hideki Takami. Vol. 10702. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2018, p. 1070237. DOI: 10.1117/12.2314984.

[Che+19] Cherenkov Telescope Array Consortium et al. *Science with the Cherenkov Telescope Array*. 2019. DOI: 10.1142/10986.

[Che+20] The Cherenkov Telescope Array Consortium et al. "Sensitivity of the Cherenkov Telescope Array for probing cosmology and fundamental physics with gamma-ray propagation". In: *arXiv e-prints*, arXiv:2010.01349 (Oct. 2020), arXiv:2010.01349. arXiv: 2010.01349 [astro-ph.HE].

[D'A+18] A. D'Ai, A. Melandri, S. J. Laporte, J. A. Kennea, B. Sbarufatti, S. L. Gibson, Z. Liu, J. P. Osborne, P. D'Avanzo, and P. A. Evans. "GRB 180914B: Swift-XRT afterglow detection." In: *GRB Coordinates Network* 23236 (Jan. 2018), p. 1.

[Del+08] E. Del Monte et al. "GRB 070724B: the first gamma ray burst localized by SuperAGILE and its Swift X-ray afterglow". In: A&A 478.1 (Jan. 2008), pp. L5–L9. DOI: 10.1051/0004-6361:20078816. arXiv: 0712.0500 [astro-ph].

[Del+11] E. Del Monte et al. "The AGILE observations of the hard and bright GRB 100724B". In: A&A 535, A120 (Nov. 2011), A120. DOI: 10.1051/0004-6361/201117053. arXiv: 1109.3018 [astro-ph.HE].

[Fer+07] M. Feroci et al. "SuperAGILE: The hard X-ray imager for the AGILE space mission". In: *Nuclear Instruments and Methods in Physics Research A* 581.3 (Nov. 2007), pp. 728–754. DOI: 10.1016/j.nima.2007.07.147. arXiv: 0708.0123 [astro-ph].

[Fio+19] V. Fioretti, D. Ribeiro, T. B. Humensky, A. Bulgarelli, G. Maier, A. Moralejo, and C. Nigro. "The Cherenkov Telescope Array sensitivity to the transient sky". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. International Cosmic Ray Conference. July 2019, p. 673. arXiv: 1907.08018 [astro-ph.IM].

[FKM20] Thorben Finke, Michael Krämer, and Silvia Manconi. "Classification of Fermi-LAT sources with deep learning using energy and time spectra". In: *arXiv e-prints*, arXiv:2012.05251 (Dec. 2020), arXiv:2012.05251. arXiv: 2012.05251 [astro-ph.HE].

[Fla+09] D. L. Flath, T. S. Johnson, M. Turri, and K. A. Heidenreich. "GLAST (FERMI) Data-Processing Pipeline". In: *Astronomical Data Analysis Software and Systems XVIII*. Ed. by D. A. Bohlender, D. Durand, and P. Dowler. Vol. 411. Astronomical Society of the Pacific Conference Series. Sept. 2009, p. 193.

[Fun15] Stefan Funk. "Ground- and Space-Based Gamma-Ray Astronomy". In: *Annual Review of Nuclear and Particle Science* 65 (Oct. 2015), pp. 245–277. DOI: 10.1146/annurev-nucl-102014-022036. arXiv: 1508.05190 [astro-ph.HE].

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[Geh+04] N. Gehrels et al. "The Swift Gamma-Ray Burst Mission". In: ApJ 611.2 (Aug. 2004), pp. 1005–1020. DOI: 10.1086/422091. arXiv: astro-ph/0405233 [astro-ph].

[Gir+19] Enrico Giro et al. "The ASTRI-Horn telescope validation toward the production of the ASTRI Mini-Array: a proposed pathfinder for the Cherenkov Telescope Array". In: *Optics for EUV, X-Ray, and Gamma-Ray Astronomy IX*. Ed. by Stephen

L. O'Dell and Giovanni Pareschi. Vol. 11119. International Society for Optics and Photonics. SPIE, 2019, pp. 446 –460. DOI: 10.1117/12.2530896. URL: https://doi.org/10.1117/12.2530896.

[Giu+08] A. Giuliani et al. "AGILE detection of delayed gamma-ray emission from GRB 080514B". In: A&A 491.2 (Nov. 2008), pp. L25–L28. DOI: 10.1051/0004-6361:200810737. arXiv: 0809.1230 [astro-ph].

[Giu+09] A. Giuliani et al. "GRB 090401B: AGILE refined analysis." In: *GRB Coordinates Network* 9075 (Jan. 2009), p. 1.

[Giu+10] A. Giuliani et al. "AGILE Detection of Delayed Gamma-ray Emission From the Short Gamma-Ray Burst GRB 090510". In: ApJ 708.2 (Jan. 2010), pp. L84–L88. DOI: 10.1088/2041-8205/708/2/L84. arXiv: 0908.1908 [astro-ph.HE].

[Giu+13] A. Giuliani et al. "GRB 131108A: AGILE/GRID observation." In: *GRB Coordinates Network* 15479 (Jan. 2013), p. 1.

[GM12] Neil Gehrels and Péter Mészáros. "Gamma-Ray Bursts". In: *Science* 337.6097 (Aug. 2012), p. 932. DOI: 10.1126/science.1216793. arXiv: 1208.6522 [astro-ph.HE].

[Gol+20] A. Goldstein et al. "Evaluation of Automated Fermi GBM Localizations of Gamma-Ray Bursts". In: ApJ 895.1, 40 (May 2020), p. 40. DOI: 10.3847/1538-4357/ab8bdb. arXiv: 1909.03006 [astro-ph.IM].

[Hur+10] K. Hurley et al. "The Third Interplanetary Network". In: *AIP Conference Proceedings* 1279.1 (2010), pp. 330–333. DOI: 10.1063/1.3509301. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.3509301. URL: https://aip.scitation.org/doi/abs/10.1063/1.3509301.

[Kei+17] A. Keivani, H. Ayala, J. DeLaunay, and AMON Core Team. "Astrophysical Multimessenger Observatory Network (AMON): Science, Infrastructure, and Status". In: *35th International Cosmic Ray Conference (ICRC2017)*. Vol. 301. International Cosmic Ray Conference. Jan. 2017, p. 629. arXiv: 1708.04724 [astro-ph.IM].

[Koz+19] A. Kozlova et al. "IPN Triangulation of GRB 190606A (short)". In: *GRB Coordinates Network* 24765 (June 2019), p. 1.

[Kre09] Frank Krennrich. "Gamma ray astronomy with atmospheric Cherenkov telescopes: the future". In: *New Journal of Physics* 11.11 (2009), p. 115008. DOI: 10.1088/1367-2630/11/11/115008. URL: https://doi.org/10.1088/1367-2630/11/11/115008.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012, pp. 1097–1105. URL: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[Lab+09] C. Labanti, M. Marisaldi, F. Fuschino, M. Galli, A. Argan, A. Bulgarelli, G. Di Cocco, F. Gianotti, M. Tavani, and M. Trifoglio. "Design and construction of the Mini-Calorimeter of the AGILE satellite". In: *Nuclear Instruments and Methods in Physics Research A* 598.2 (Jan. 2009), pp. 470–479. DOI: 10.1016/j.nima.2008.09.021. arXiv: 0810.1842 [astro-ph].

[Lan01] Doug Laney. "3-D Data Management: Controlling Data Volume, Velocity, and Variety". In: *META Group Res Note 6* 6 (Jan. 2001).

[LM83] T. P. Li and Y. Q. Ma. "Analysis methods for results in gamma-ray astronomy." In: ApJ 272 (Sept. 1983), pp. 317–324. DOI: 10.1086/161295.

[Lon+13] F. Longo et al. "GRB 130327B: gamma-ray detection by AGILE." In: *GRB Coordinates Network* 14344 (Jan. 2013), p. 1.

[Luc+19a] F. Lucarelli et al. "AGILE Detection of Gamma-Ray Sources Coincident with Cosmic Neutrino Events". In: ApJ 870.2, 136 (Jan. 2019), p. 136. DOI: 10.3847/1538-4357/aaf1c0. arXiv: 1811.07689 [astro-ph.HE].

[Luc+19b] F. Lucarelli et al. "AGILE follow-up of the high-energy neutrino candidate event IceCube-190730A". In: *The Astronomer's Telegram* 12970 (July 2019), p. 1.

[Luc+19c] F. Lucarelli et al. "GRB 190501A: AGILE/GRID detection." In: *GRB Coordinates Network* 24361 (Jan. 2019), p. 1.

[MAG+19a] MAGIC Collaboration et al. "Observation of inverse Compton emission from a long γ-ray burst". In: Nature 575.7783 (Nov. 2019), pp. 459–463. DOI: 10.1038/s41586-019-1754-6. arXiv: 2006.07251 [astro-ph.HE].

[MAG+19b] MAGIC Collaboration et al. "Teraelectronvolt emission from the γ-ray burst GRB 190114C". In: Nature 575.7783 (Nov. 2019), pp. 455–458. DOI: 10.1038/s41586-019-1750-x. arXiv: 2006.07249 [astro-ph.HE].

[Mar+10] M. Marisaldi et al. "Detection of terrestrial gamma ray flashes up to 40 MeV by the AGILE satellite". In: *Journal of Geophysical Research (Space Physics)* 115.53, A00E13 (Mar. 2010), A00E13. DOI: 10.1029/2009JA014502.

[Mar+14] M. Marisaldi et al. "Properties of terrestrial gamma ray flashes detected by AGILE MCAL below 30 MeV". In: *Journal of Geophysical Research (Space Physics)* 119.2 (Feb. 2014), pp. 1337–1355. DOI: 10.1002/2013JA019301.

[Mar+20] Drozdova Mariia, Broilovskiy Anton, Ustyuzhanin Andrey, and Malyshev Denys. "A study of neural networks point source extraction on simulated Fermi/LAT telescope images". In: *Astronomische Nachrichten* 341.8 (Oct. 2020), pp. 819–826. DOI: 10.1002/asna.202013788. arXiv: 2007.04295 [cs.CV].

[Mat+96] J. R. Mattox et al. "The Likelihood Analysis of EGRET Data". In: ApJ 461 (Apr. 1996), p. 396. DOI: 10.1086/177068.

[McE+12] Julie E. McEnery, Peter F. Michelson, William S. Paciesas, and Steven Ritz. "Fermi Gamma-Ray Space Telescope". In: *Optical Engineering* 51.1 (2012), pp. 1 –10. DOI: 10.1117/1.OE.51.1.011012. URL: https://doi.org/10.1117/1.OE.51.1.011012.

[Mee+09] Charles Meegan et al. "The Fermi Gamma-ray Burst Monitor". In: ApJ 702.1 (Sept. 2009), pp. 791–804. DOI: 10.1088/0004-637X/702/1/791. arXiv: 0908.0450 [astro-ph.IM].

[Mes+17] Cody Messick et al. "Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data". In: *Phys. Rev. D* 95 (4 2017), p. 042001. DOI: 10.1103/PhysRevD.95.042001. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.042001.

[Mes+19] Peter Meszaros, Derek B. Fox, Chad Hanna, and Kohta Murase. "Multi-messenger astrophysics". In: *Nature Reviews Physics* 1.10 (Oct. 2019), pp. 585–599. DOI: 10.1038/s42254-019-0101-z. arXiv: 1906.10212 [astro-ph.HE].

[Mir+19] R. Mirzoyan, K. Noda, E. Moretti, A. Berti, C. Nigro, J. Hoang, S. Micanovic, M. Takahashi, Y. Chai, A. Moralejo, and MAGIC Collaboration. "MAGIC detects the GRB 190114C in the TeV energy domain." In: *GRB Coordinates Network* 23701 (Jan. 2019), p. 1.

[Mor+09] E. Moretti et al. "AGILE gamma-ray detection of GRB 090401B." In: *GRB Coordinates Network* 9069 (Jan. 2009), p. 1.

[Oya+19] Igor Oya et al. "The Array Control and Data Acquisition System of the Cherenkov Telescope Array". In: *ICALEPCS 2019 - 17th Biennial International Conference on Accelerator and Large Experimental Physics Control Systems*. New York, NY, United States, Oct. 2019. URL: https://hal.inria.fr/hal-02421340.

[Par+a] N. Parmiggiani, A. Bulgarelli, D. Beneventano, L. Baroncelli, A. Addis, and M. Tavani. "RTApipe, a framework to develop astronomical pipelines for the real-time analysis of scientific data". submitted to the Astronomical Data Analysis Software and Systems 2020 Conference.

[Par+b] N. Parmiggiani, A. Bulgarelli, A. Giuliani, V. Fioretti, A. Di Piano, F. Longo, F. Verrecchia, M. Tavani, D. Beneventano, and A. Macaluso. "A Deep Learning Method for AGILE-GRID GRB Detection". Under review on the Astrophysical Journal.

[Par+16] Giovanni Pareschi, Giacomo Bonnoli, Stefano Vercellone, ASTRI Collaboration, and CTA Consortium. "The mini-array of ASTRI SST-2M telescopes, precursors for the Cherenkov Telescope Array". In: *Journal of Physics Conference Series*. Vol. 718. Journal of Physics Conference Series. May 2016, p. 052028. DOI: 10.1088/1742-6596/718/5/052028.

[Par+20] N. Parmiggiani, Bulgarelli A., Tampieri. S., L. Baroncelli, and A. Addis. *AGILE Real-Time Analysis Setup*. Tech. rep. 49. Italian National Institute for Astrophysics, 2020.

[PB20a] N. Parmiggiani and A. Bulgarelli. *Centos 7 Configuration for AGILE Pipelines*. Tech. rep. 50. Italian National Institute for Astrophysics, 2020.

[PB20b] N. Parmiggiani and A. Bulgarelli. *CTA Real-Time Analysis Pipeline Configuration*. Tech. rep. 46. Italian National Institute for Astrophysics, 2020.

[Per+06] F. Perotti, M. Fiorini, S. Incorvaia, E. Mattaini, and E. Sant'Ambrogio. "The AGILE anticoincidence detector". In: *Nuclear Instruments and Methods in Physics Research A* 556.1 (Jan. 2006), pp. 228–236. DOI: 10.1016/j.nima.2005.10.016.

[PK18] N. Paul and M. Kuin. "GRB180914B: Swift UVOT detection of an optical afterglow." In: *GRB Coordinates Network* 23241 (Jan. 2018), p. 1.

[Pre+03] M. Prest, G. Barbiellini, G. Bordignon, G. Fedel, F. Liello, F. Longo, C. Pontoni, and E. Vallazza. "The AGILE silicon tracker: an innovative /γ-ray instrument for space". In: *Nuclear Instruments and Methods in Physics Research A* 501.1 (Mar. 2003), pp. 280–287. DOI: 10.1016/S0168-9002(02)02047-8.

[RMF19] O. J. Roberts, C. Meegan, and Fermi GBM Team. "GRB 190606A: Fermi GBM detection". In: *GRB Coordinates Network* 24764 (June 2019), p. 1.

[Sab+15] S. Sabatini et al. "On the Angular Resolution of the AGILE Gamma-Ray Imaging Detector". In: ApJ 809.1, 60 (Aug. 2015), p. 60. DOI: 10.1088/0004-637X/809/1/60. arXiv: 1507.01475 [astro-ph.HE].

[Tav+08] M. Tavani et al. "The AGILE space mission". In: *Nuclear Instruments and Methods in Physics Research A* 588.1-2 (Apr. 2008), pp. 52–62. DOI: 10.1016/j.nima.2008.01.023.

[Tav+09] M. Tavani et al. "The AGILE Mission". In: A&A 502.3 (Aug. 2009), pp. 995–1013. DOI: 10.1051/0004-6361/200810527. arXiv: 0807.4254 [astro-ph].

[Tav+20a] M. Tavani et al. "AGILE detection of a hard X-ray burst in temporal coincidence with a radio burst from SGR 1935+2154". In: *The Astronomer's Telegram* 13686 (Apr. 2020), p. 1.

[Tav+20b] M. Tavani et al. "An X-Ray Burst from a Magnetar Enlightening the Mechanism of Fast Radio Bursts". In: *arXiv e-prints*, arXiv:2005.12164 (May 2020), arXiv:2005.12164. arXiv: 2005.12164 [astro-ph.HE].

[Tes+13] Diego Tescaro, Alicia Lopez-Oramas, Abelardo Moralejo, Daniel Mazin, and Daniela Hadasch. "The MAGIC Telescopes DAQ Software and the On-the-Fly Online Analysis Client". In: *International Cosmic Ray Conference*. Vol. 33. International Cosmic Ray Conference. Jan. 2013, p. 2803. arXiv: `1310.1565 [astro-ph.IM]`.

[Tri+08] M. Trifoglio, A. Bulgarelli, F. Gianotti, F. Lazzarotto, G. Di Cocco, F. Fuschino, and M. Tavani. "Architecture and performances of the AGILE Telemetry Preprocessing System (TMPPS)". In: *Space Telescopes and Instrumentation 2008: Ultraviolet to Gamma Ray*. Ed. by Martin J. L. Turner and Kathryn A. Flanagan. Vol. 7011. International Society for Optics and Photonics. SPIE, 2008, pp. 1033 –1040. DOI: `10.1117/12.789338`. URL: `https://doi.org/10.1117/12.789338`.

[Tri+10] D. Borla Tridon, T. Schweizer, F. Goebel, R. Mirzoyan, and M. Teshima. "The MAGIC-II gamma-ray stereoscopic telescope system". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 623.1 (2010). 1st International Conference on Technology and Instrumentation in Particle Physics, pp. 437 –439. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2010.03.028`. URL: `http://www.sciencedirect.com/science/article/pii/S0168900210005929`.

[Urs+18] A. Ursi et al. "GRB 180914B: AGILE/MCAL detection of a burst." In: *GRB Coordinates Network* 23226 (Jan. 2018), p. 1.

[Urs+19a] A. Ursi et al. "Fermi trigger No 581478912 / GRB 190606A: AGILE/MCAL observations". In: *GRB Coordinates Network* 24759 (June 2019), p. 1.

[Urs+19b] A. Ursi et al. "GRB 190114C: AGILE/MCAL detection." In: *GRB Coordinates Network* 23712 (Jan. 2019), p. 1.

[Urs+19c] Alessandro Ursi, Marco Tavani, Francesco Verrecchia, Martino Marisaldi, Andrea Argan, Alessio Trois, and Patrizio Tempesta. "A New AGILE MCAL Configuration to Detect Gamma-Ray Bursts and Sub-threshold Events in the Multimessenger Era". In: ApJ 871.1, 27 (Jan. 2019), p. 27. DOI: `10.3847/1538-4357/aaf28f`.

[Urs+20a] A. Ursi et al. "AGILE and Konus-Wind Observations of GRB 190114C: The Remarkable Prompt and Early Afterglow Phases". In: ApJ 904.2, 133 (Dec. 2020), p. 133. DOI: 10.3847/1538-4357/abc2d4.

[Urs+20b] A. Ursi et al. "AGILE observations of the SGR 1935+2154 "burst forest"". In: *The Astronomer's Telegram* 13682 (Apr. 2020), p. 1.

[Ver+13] F. Verrecchia et al. "GRB 130427A: high energy gamma-ray detection by AGILE." In: *GRB Coordinates Network* 14515 (Jan. 2013), p. 1.

[Ver+17] F. Verrecchia, A. Ursi, F. Lucarelli, F. Longo, and M. Tavani. "GRB 170115B: AGILE detection." In: *GRB Coordinates Network* 20474 (Jan. 2017), p. 1.

[Ver+18] F. Verrecchia et al. "GRB 180914B: AGILE/GRID detection." In: *GRB Coordinates Network* 23231 (Jan. 2018), p. 1.

[Ver+19a] F. Verrecchia et al. "GRB 190530A: AGILE/GRID analysis." In: *GRB Coordinates Network* 24683 (Jan. 2019), p. 1.

[Ver+19b] Francesco Verrecchia, Marco Tavani, Andrea Bulgarelli, Martina Cardillo, Claudio Casentini, Immacolata Donnarumma, Francesco Longo, Fabrizio Lucarello, Nicoló Parmiggiani, Giovanni Piano, Maura Pilia, Carlotta Pittori, Alessandro Ursi, and Agile Team. "AGILE search for gamma-ray counterparts of gravitational wave events". In: *Rendiconti Lincei. Scienze Fisiche e Naturali* (Feb. 2019). DOI: 10.1007/s12210-019-00854-0.

[Way+12] Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, and Ashok N. Srivastava. *Advances in Machine Learning and Data Mining for Astronomy.* 2012.

[Zim+12] S. Zimmer, L. Arrabito, T. Glanzman, T. Johnson, C. Lavalley, and A. Tsaregorodtsev. "Extending the Fermi-LAT Data Processing Pipeline to the Grid". In: *Journal of Physics Conference Series.* Vol. 396. Journal of Physics Conference Series. Dec. 2012, p. 032121. DOI: 10.1088/1742-6596/396/3/032121. arXiv: 1212.4115 [astro-ph.IM].

[Zol+19] A. Zoli, A. Bulgarelli, M. Tavani, V. Fioretti, M. Marisaldi, N. Parmiggiani, F. Fuschino, F. Gianotti, and M. Trifoglio. "The AGILE Pipeline for Gravitational Waves Events Follow-up". In: *Astronomical Data Analysis Software and Systems XXVI*. Ed. by Marco Molinaro, Keith Shortridge, and Fabio Pasian. Vol. 521. Astronomical Society of the Pacific Conference Series. Oct. 2019, p. 139.

[ZZ15] Yanxia Zhang and Yongheng Zhao. "Astronomy in the Big Data Era". In: *Data Science Journal* 14 (May 2015), p. 11. DOI: `10.5334/dsj-2015-011`.