UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences

Search for cosmic sources of high energy neutrinos with the AMANDA-II detector

Thèse présentée par
Mathieu LABARE
en vue de l'obtention du grade de
Docteur en Sciences

Promoteur
Pr. Daniel Bertrand

Membres du jury
Pr. Peter Tinyakov
Pr. Kael Hanson
Pr. Gaston Wilquet
Dr. Bruny Baret
Dr. Chad Finley

Décembre 2009

A Matthys

Résumé

AMANDA-II est un télescope à neutrinos composé d'un réseau tri-dimensionnel de senseurs optiques déployé dans la glace du Pôle Sud. Son principe de détection repose sur la mise en évidence de particules secondaires chargées émises lors de l'interaction d'un neutrino de haute énergie ($>$ 100 GeV) avec la matière environnant le détecteur, sur base de la détection de rayonnement Čerenkov.

Ce travail est basé sur les données enregistrées par AMANDA-II entre 2000 et 2006, afin de rechercher des sources cosmiques de neutrinos. Le signal recherché est affecté d'un bruit de fond important de muons et de neutrinos issus de l'interaction du rayonnement cosmique primaire dans l'atmosphère. En se limitant à l'observation de l'hémisphère nord, le bruit de fond des muons atmosphériques, absorbés par la Terre, est éliminé. Par contre, les neutrinos atmosphériques forment un bruit de fond irréductible constituant la majorité des 6100 événements sélectionnés pour cette analyse. Il est cependant possible d'identifier une source ponctuelle de neutrinos cosmiques en recherchant un excès local se détachant du bruit de fond isotrope de neutrinos atmosphériques, couplé à une sélection basée sur l'énergie, dont le spectre est différent pour les deux catégories de neutrinos.

Une approche statistique originale est développée dans le but d'optimiser le pouvoir de détection de sources ponctuelles, tout en contrôlant le taux de fausses découvertes, donc le niveau de confiance d'une observation. Cette méthode repose uniquement sur la connaissance de l'hypothèse de bruit de fond, sans aucune hypothèse sur le modèle de production de neutrinos par les sources recherchées. De plus, elle intègre naturellement la notion de facteur d'essai rencontrée dans le cadre de test d'hypothèses multiples.

La procédure a été appliquée sur l'échantillon final d'évènements récoltés par AMANDA-II. Comme aucun signal n'a été statistiquement détecté, une limite supérieure moyenne pour un flux de neutrinos muoniques a été évaluée à E^2 dΦ/dE = 4.5 10^{-11} TeV cm^{-2}s^{-1} avec un niveau de confiance de 99$\%$.

Abstract

AMANDA-II is a neutrino telescope which comprises a three dimensional array of optical sensors deployed in the South Pole glacier. Its principle rests on the detection of the Cherenkov radiation emitted by charged secondary particles produced by the interaction of a high energy neutrino ($>$ 100 GeV) with the matter surrounding the detector.

This work is based on data recorded by the AMANDA-II detector between 2000 and 2006 in order to search for cosmic sources of neutrinos. A potential signal must be extracted from the overwhelming background of muons and neutrinos originating from the interaction of primary cosmic rays within the atmosphere. The observation is limited to the northern hemisphere in order to be free of the atmospheric muon background, which is stopped by the Earth. However, atmospheric neutrinos constitute an irreducible background composing the main part of the 6100 events selected for this analysis. It is nevertheless possible to identify a point source of cosmic neutrinos by looking for a local excess breaking away from the isotropic background of atmospheric neutrinos; This search is coupled with a selection based on the energy, whose spectrum is different from that of the atmospheric neutrino background.

An original statistical approach has been developed in order to optimize the detection of point sources, whilst controlling the false discovery rate – hence the confidence level – of an observation. This method is based solely on the knowledge of the background hypothesis, without any assumption on the production model of neutrinos in sought sources. Moreover, the method naturally accounts for the trial factor inherent in multiple testing.

The procedure was applied on the final sample of events collected by AMANDA-II. Since no signal was statistically found, an upper limit on the neutrino flux was evaluated at $E^2\ \mathrm{d}\Phi/\mathrm{d}E = 4.5\ 10^{-11}$ TeV cm^{-2}s^{-1} with a confidence level of 99$\%$.

Acknowledgments

Albeit it is impossible to draw up an exhaustive list of every contributor, I want nevertheless to have a thought to the persons who helped me driving this work to its end.

First, I would like to thank my supervisor, Pr.Daniel Bertrand, who gave me the opportunity to carry out this thesis on a subject (I'm sure he did not know at the moment he made me the proposition) that covers two of my favorite topics. I'm very grateful to him for having taken the responsability of this thesis and keeping believing in me all along these four years.

I'm very indebted to Dr.Bruny Baret, *le plus belge des parisiens*, who took me under his brand-new Doctor's wings to steer me for more than two years and never let me down after his departure for other neutrinos skies.

I would also like to express all my gratitude to Pr.Kael Hanson, who suffered through the first draft of this book (in a weird language called *Frenglish*), and gave me countless corrections and suggestions to make it intelligible to any normal English reader. Thinking about that, you should thank him too!

The AMANDA/IceCube collaboration is an incredible experiment, located in an incredible place and run by wonderful people; So I thank my officemates Alfio and Sabrina, as well as the whole IceCube team in Brussels (Catherine, Daan, Olbren, and all the newcomers), in Belgium and abroad.

I acknowledge all the members of the IIHE, from the secretaries to the computer support team and other colleagues (amazingly, some people work on experiments that do not search for neutrinos!) for the great atmosphere in the unique lab in Belgium where 3 languages, and sometimes more, are daily used.

Du plus profond de mon coeur, je remercie Virginie, ma famille et mes amis. Ces quatres années auront été fantastiques, mais 2009 aura sans contestation aucune été la plus extraordinaire: Naissances, mariages, voyages, multiples défenses de thèse (quelle idée d'avoir des amis physiciens!), aucun grand moment ne m'aura été épargné! Merci à tous pour votre aide, votre soutien, votre patience et votre affection, dans les bons moments comme dans les (petites) périodes de doute.

For those I might have forgotten, and I surely have, I sincerely apology for this and they can complain by sending an e-mail, or recognize themselves in the category "friends".

Contents

Introduction

Astronomy is known as an ancestral science which has been long-limited to the visible part of the electromagnetic spectrum. From greatest antiquity, and long before the invention of the first optical instruments, astronomers observed and wondered about stars.

For centuries visible light has been the only gate we had to reach the Cosmos. At the end of the 19th century, the observable spectrum broadened with the emergence of radio, infrared and X rays invisible to the naked eye. Later the deployment of observational instruments in space have allowed astronomers to overcome issues caused by the opacity of the earth's atmosphere at several wavelength bands. Thus, at the dawn of this third millennium, *light*, both visible and invisible, allowed the discovery of cosmic phenomenons taking place over more than 18 orders of magnitude in wavelength along the electromagnetic spectrum.

In the meantime, the discovery and the study of cosmics rays have given us a complementary channel to probe the sky thanks to the observation of a wide variety of ionizing particles apparently coming from all regions of the Universe. Cosmic rays were the first source of particles for high energy physics investigations, long before the construction of man-made accelerators. Numerous searches were performed with detectors deployed at ground level, in high altitude balloons, or sent into space, in order to measure and understand the characteristics of cosmic rays and answer questions about their origin and composition. These investigations quickly led to the discovery of new particles.

Among all these particles rushing from the Cosmos, the neutrino combines two interesting advantages. The neutrino interacts only weakly with matter. As its cross section is very small, it can travel over large distances without being absorbed by the interstellar medium. Moreover, the neutrino has no electrical charge and can make its way from its original place without being deviated by magnetic fields. Therefore a correlation can be established with possible sources of neutrinos, in the case we can detect and reconstruct the neutrino direction. Since the eighties a new astronomy era is then being developed, with the construction of experiments specially dedicated to the detection of neutrinos.

Since neutrinos can only be detected through the particles emitted by their interaction with matter, very large target volumes are necessary to detect them in quantity. Various methods and materials have been used to detect neutrinos. Some experiments (GALLEX, HOMESTAKE) looked for excesses in the chemical abundances of germanium or argon caused by the interaction of neutrinos in the interaction volume. Other detectors (Reines-Cowan[Cow56]) use scintillators. A third category concerns Cherenkov radiation detection. The interaction medium can be a large transparent volume of pure water (Super-Kamiokande), heavy water (SNO), or natural environment like seas (ANTARES [Ant09],[HA03],[Ac07]), lakes (BAIKAL [Ba99],[Aa06b]), or in the instance of this thesis, polar ice with the IceCube Neutrino Observatory.

In Cherenkov detectors, the neutrino is identified by the Cherenkov photons emitted by charged leptons resulting from the interaction of the original neutrino with matter within or in the surroundings of the detector. This optical radiation is detected by photomultipliers deployed in or around the interaction medium. It is then important to understand the properties of this medium and the response of the detector in order to reconstruct accurately the original particle trajectory.

An extraterrestrial neutrino can either be produced in the sun or directly originates from cosmic objects. We are only interested in the last category of events. However, this weak signal is overwhelmed by the much stronger flux of events – mainly muons and neutrinos – produced by the interaction of cosmic rays with the atmosphere. It is therefore necessary to develop techniques to isolate interesting neutrino events from the large atmospheric background.

A series of cuts and selections, described in Chapter 2, are applied to the large number of data recorded by the detector. These selections finally lead to a final sample of reconstructed events composed mostly of atmospheric neutrinos constituting an irreducible background.

The main topic of this work consists in the development of a statistical method allowing the identification of any tiny but statistically significant discrepancy to the atmospheric background. The identification is only performed on the basis of the background hypothesis without any assumption on the possible signal. Moreover the procedure guarantees the confidence level of any possible discovery.

Therefore, the third chapter of this thesis starts with a general introduction to statistical methods. It also describes the original False Discovery Rate (FDR) controlling procedure. Then we thoroughly describe the different approaches and development to adapt this mathematical tool to the search for astrophysical neutrino point sources. Chapter 4 covers fixed grid approaches while Chapter 5 examines the successive enhancements leading to the final procedure. The construction of a reliable description of the background behavior is an important

matter which deserves a complete chapter (Chapter 6). Checks and expected performances of the procedure with the *Data Challenge* simulations are developed in Chapter 7. The method is eventually applied to the complete set of data collected by the AMANDA-II detector between 2000 and 2006 and results are presented in Chapter 8.

Part I

High Energy Neutrino Astrophysics : A new window on the Universe

We live in a changing Universe, and few things are changing faster than our conception of it.

Timothy Ferris

1

From photon to neutrino : Neutrino Astronomy and High Energy Point Sources

1.1 Photons

Man is naturally restricted to observe the very thin part of the spectrum of electromagnetic radiation in which he continuously lives. For centuries, the observation of cosmic phenomena was confined to the visible part of the electromagnetic spectrum. The end of the nineteen century witnessed the development of Maxwell's theory of electromagnetism. This new field kicked off the start of major technological advances progressively opening new windows on the observation and the understanding of objects crowding the Universe. Thus, the very first observation of non-visible electromagnetic waves from an astronomical object dates from the 1930s: the detection of the Milky Way radio signal by Karl G. Jansky. However the Earth's atmosphere limits mainly ground observations to the optical and a large part of the radio spectrum (Fig.1.1). The use of other bands needed beforehand the deployment of new observatories above the atmosphere.

It was necessary to wait until the end of World War II and its important repercussions in terms of electronic developments to really initiate radio astronomy. Balloons were sent in the highest part of the atmosphere to provide first observations of the sky in the X-ray and γ-ray domains. In the sixties, when

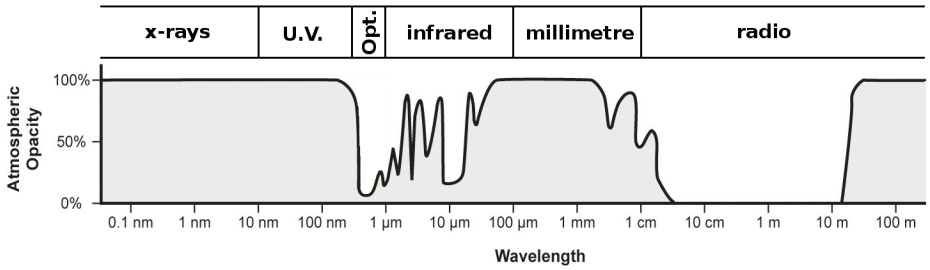


Figure 1.1: *Earth's atmosphere opacity for wavelength from X-rays to radio wave bands*

technology allowed us to send rockets and satellites into space, the sky survey in the ultra-violet (UV), X-rays and γ-rays became possible.

The radio emission can result from synchrotron radiation and then provides a bench of new information about the hottest plasmas in the Universe. Most of radio radiation seems to come from surroundings of galaxies and more precisely from active nuclei of galaxies (AGN) or neutron stars expelling energy in the form of relativistic plasma jets. This permitted the discovery of new powerful phenomena like blazars, quasars, pulsars, etc., discussed in section 1.3.2.

Those objects are not only radio emitters. On the other side of the electromagnetic spectra, the study of X-ray bands reveals galactic and extra-galactic sources to be potential strong, often variable X-ray emitters. Still more energetic, γ-ray emissions were also detected.

Finally, an important discovery in radio and millimeter astronomy was the Microwave Background Radiation observed by Penzias and Wilson in 1965 [WPH65] [PW65]. This radiation is the remnant of the equilibrium radiation spectrum from the early Universe and then an significant testimony in favor of the Big Bang theory. Its interesting features are a perfect black-body spectrum with radiation temperature of $2.7K$ and its remarkable isotropy.

1.2 Cosmic Rays

Electromagnetic radiation is not the sole open door we possess to probe the Cosmos. The very first hint about the existence of cosmic rays goes back to 1912, when the Austrian physicist Victor Hess performed a series of high-altitude balloon flights in order to quantify the variation of the ionizing radiation with altitude. Surprisingly, an increase of the signal with the altitude was observed [Hes12]. The conclusion of this phenomena was that an unexpected radiation penetrates the atmosphere. This kicked off the hunting of new particles and the

set-up of hundreds of experiments [NA37] dedicated to study this new radiation.

1.2.1 Primary Cosmic Rays

Cosmic radiation was first identified as a flux of high energy particles whose composition has been established [Mue91] to be mainly protons (86%), various nuclei (such as α-particles (11%) and heavy elements (1%)) and electrons (2%). The ratio of lighter elements is comparable to the primordial abundance of these elements in the Universe. However, lithium, beryllium and boron are found to be overabundant among cosmic rays compared to their small contribution in the solar system or the interstellar medium. These secondary elements are produced from spallation of primary heavier nuclei (like carbon and oxygen) during their way through space. One remarkable feature of cosmic rays is that their energies spectra span a very wide range (10^9 up to 10^{20} eV) and thirty orders of magnitude in flux (Fig.1.2).

The lower energy region ($<$ GeV) of the cosmic ray component is due to solar activity. There are also referred as Solar Energetic Particles (SEP) and consists mainly of protons, electrons and heavy ions accelerated through diffusive shock acceleration (or first-order Fermi acceleration mechanism, see further) or shock-drift mechanism [GP03].

Above this limit, the differential flux follows successive power laws $d\Phi/dE \propto E^{-\gamma}$. A first break is observed at an energy about 10^{15} eV, which is know as the "knee" of the cosmic ray spectrum. At that point, the spectral index steepens from $\gamma \approx 2.7$ to $\gamma \approx 3.1$. The most favored theory assumes this to be the maximum energy reachable by galactic accelerators [EW01]. Therefore, cosmic rays below the break are supposed to come from galactic objects. This theory is consistent with the assumption that the acceleration mechanism of the nuclei is Fermi acceleration [LS88]. At an energy of 10^{18} eV, refered as the "ankle", the spectrum flattens again. Preferred scenarios account for an extra-galactic origin of cosmic rays [WW90].

However the transition out of galactic contribution to highest energies happens progressively. Different options can explain this behavior. The Fermi acceleration mechanism allows heavier elements (up to iron) to reach higher energies ($E_{max} = Z \times R^*$, where R^* characterizes the maximum rigidity and Z the atomic number). Nevertheless this hypothesis has difficulty to explain the smoothness of the evolution to the ankle. A hypothesis taking this feature into account is a re-acceleration process (in multiple sources shocks area) like supernovae cluster or by galactic wind [ZV06]. Another solution is a change in the properties of diffusion of the interstellar medium [Ptu81].

The highest cosmic rays ever observed reached an energy over 10^{20} eV [Kam08].

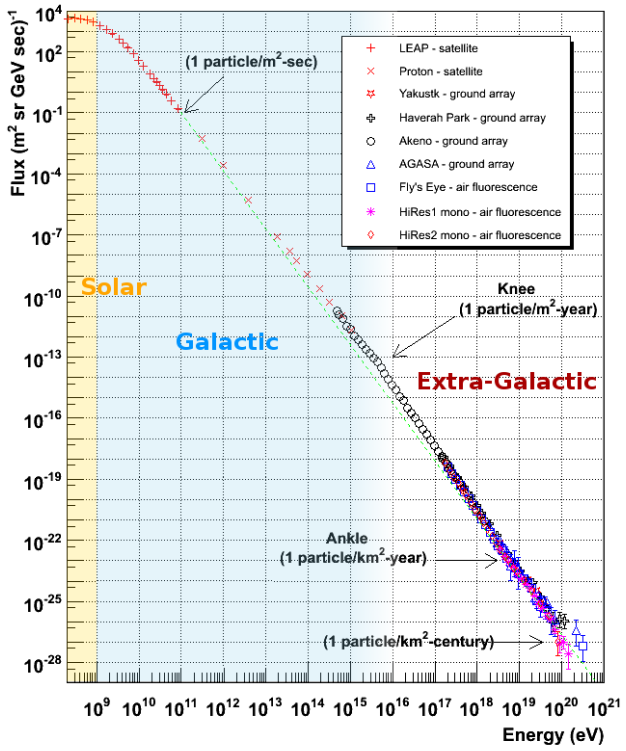


Figure 1.2: *Cosmic ray spectrum (modified from [Han])*

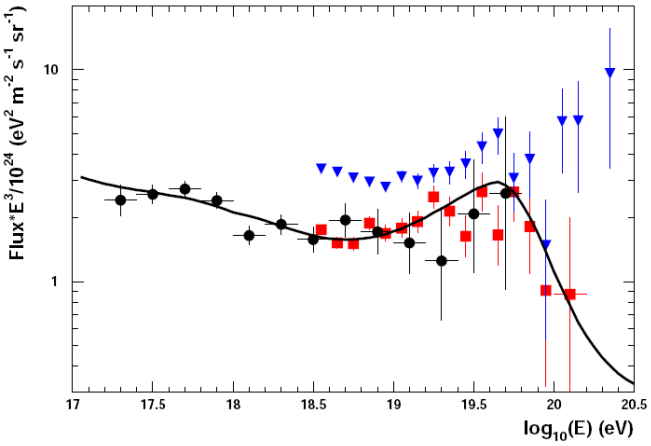


Figure 1.3: *Ultra High Energy Cosmic ray E^{-3} fluxes measured with HiRES-I, HiRES-II and AGASA experiments.[Aa04a])*

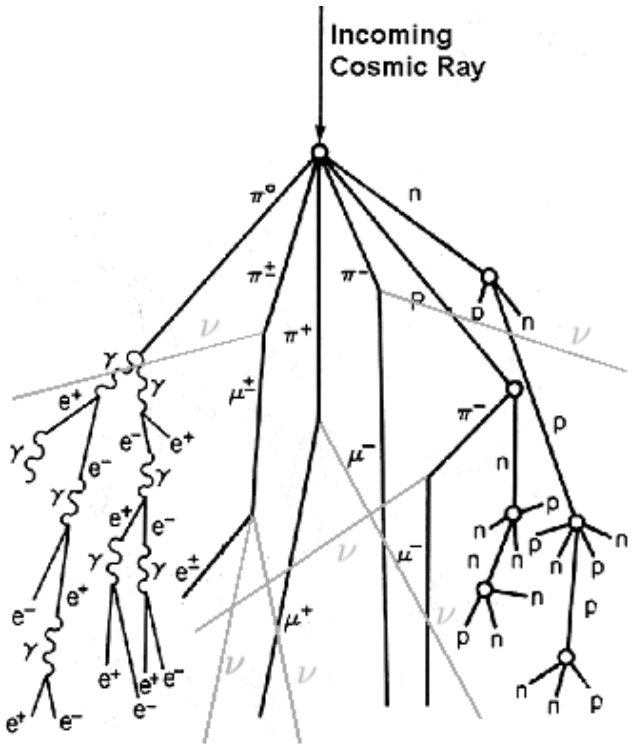


Figure 1.4: *Electromagnetic (left) and hadronic (right) cascades produced by the interaction of primary cosmic rayd in the Earth's atmosphere. Atmospheric muons and neutrinos are produced by charged pions. Analogous decays also take place from kaons.*

However Greisen [Gre66], Zatsepin and Kuzmin [ZK66] predict a strong attenuation at $5\ 10^{19}$ eV, termed the GZK cutoff. Above this point, protons have enough energy to interact with the cosmic background radiation through a delta resonance. Due to their extremely rare occurrence, the reliability of the measure (Fig.1.3) as well as the origin of these events [Yam08] are very discussed topics.

Though great progress has been made in the last decades, the questions of cosmic rays are far from solved and their origin, their nature or their acceleration to the highest energies still have to be determined conclusively.

1.2.2 Cosmic ray interaction in the atmosphere

Primary cosmic rays initiate cascades (Fig.1.4) of new high energy particles when penetrating and interacting in the atmosphere. Among them, muons and neutrinos constitute an important background for the search of cosmic neutrinos.

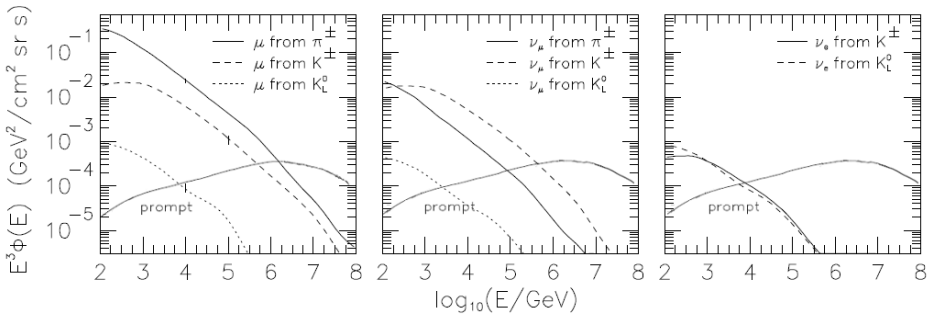


Figure 1.5: *Contribution of π, K and charm (prompt) decays to atmospheric neutrino and muon fluxes. Modified from [TIG96]*

From a general point of view, particles can generate secondaries either by interaction or by decay, depending on the particle lifetime (τ_m) and energy, as well as the medium. The density of the interstellar medium is very low. Interaction length λ_m is there much higher than decay length ($d_m = \beta\gamma c\tau_m$), meaning that particles preferentially decay before interacting with matter. It is rather different in a much denser environment such as the atmosphere. For particles whose energy exceeds the critical energy ϵ_m (where decay and interaction lengths equal), interaction becomes more likely than decay.

Therefore, the muon spectrum below $\epsilon_\mu = 1$ GeV is rather flat. It then steepens gradually to follow the primary cosmic ray spectrum until 100 GeV - 1 TeV, where the contribution from charged pions ($\epsilon_\pi = 115$ GeV) and kaons ($\epsilon_K \sim$ 800 GeV) decay (Fig.1.5) is shaded off by the increasing probability of pion interaction. Beyond this point, the muon spectrum still steepens to asymptotically follow a power spectrum $\gamma = -3.7$.

The decay of pions and kaons, as well as the decay of muons, also produce neutrinos. Due to helicity suppression, charged pions decay almost exclusively (>99.9%) in the muonic channel. Neglecting the muon contribution to neutrino production, atmospheric muons and neutrinos have a similar spectral shape. At the highest energies ($>$ 1 PeV), contribution of *prompt* flux (Fig.1.5) from the decay of charmed mesons becomes non negligible and flattens the muon and neutrino spectra(Fig.1.6).

Time and zenith angle dependencies are expected. The first is due to seasonal effects from the variation in the density of the atmosphere: atmospheric production increases in summer because of the lower density increasing interaction length and therefore favoring decay. The latter is due to the penetration angle in the atmosphere, leading to longer path for larger angles and therefore giving same result.

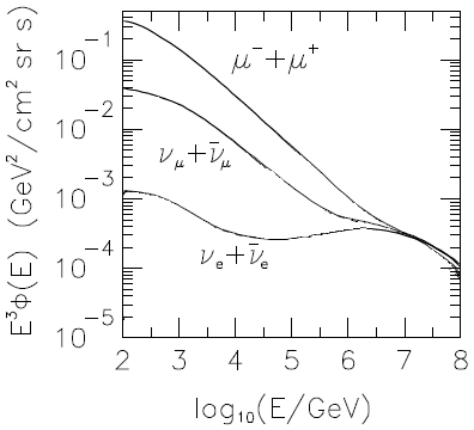


Figure 1.6: *Total atmospheric neutrino and muon fluxes. Modified from [TIG96]*

1.3 Neutrinos

The hypothesis of the existence of the neutrino in 1930 slowly permitted the understanding of measurements made fifteen years earlier which the theory could not explain. In December 1930, Wolfgang Pauli sent a letter to a Congress of Physics taking place in Tübingen, Germany. In his letter, the Austrian physicist suggests a "desperate remedy" [1] that might account for the missing energy and explain the continuous energy spectra of the electron produced in the beta decay. He predicts several characteristics of the new particle,which he names neutron, after its neutral electric charge. Fourteen month later, the neutron is discovered by Chadwick. Yet the Chadwick's particle is too heavy and hence does not match the one Pauli suggested. In October 1933, during the Solvay Congress in Brussels, Fermi suggested to rename Pauli's mysterious particle the neutrino (small neutron). During the same period, the mass of neutrino was supposed to be much smaller than the electron's and Bethe predicted its cross section to be small enough to allow the neutrino to travel through the earth without being absorbed. In the fifties, Fred Reines and Clyde Cowan used nuclear reactor to prove the existence of the neutrino[Cow56].

1.3.1 Neutrino Production

Observations of high energy electromagnetic radiation have permitted the discovery of various astrophysical objects discussed further in this section. These

[1] verzweifelte Ausweg

observations can still be used for the quest of neutrinos. On one hand, γ-rays are produced in several leptonic mechanisms as e^+e^- annihilations, synchrotron emission or inverse compton scattering. On the other hand, hadronic processes involve neutral pions which also decay resulting into γ-rays. Neutrinos, however, provide direct proof for hadronic processes since they can only be produced through the decay of charged mesons created simulteanously with neutral ones. The detection of neutrinos associated with γ-rays sources would thus discriminate between leptonic and hadronic models for acceleration mechanisms.

Mesons come either from photomeson interactions between relativistic protons and ambient photons, or interaction of accelerated nuclei with protons.

$$p+\gamma \rightarrow \Delta^+ \rightarrow \begin{cases} n+\pi^+ \rightarrow n+\mu^+ +\nu_\mu \\ \qquad\qquad\quad \hookrightarrow e^+ +\nu_e +\bar{\nu}_\mu \\ p+\pi^0 \rightarrow p+2\gamma \end{cases}$$

$$p+N \rightarrow \pi^0\pi^\pm K^0 K^\pm ...$$

γ-rays and neutrinos are then produced simultaneously. However high energy γ-rays travelling over cosmological distances suffer from their interaction with photons of the extragalactic background field, making the Universe opaque to γ-rays emitters located at extra-galactic distances (Fig.1.7).

One can thus have unknown sources of neutrinos and γ-rays hidden at long distances from the Earth. Neutrinos will then be the privileged messengers from these most distant regions of the Universe.

Fermi acceleration

One of the most commonly admitted acceleration mechanisms considered to be responsible for the origin of galactic cosmic rays is Fermi acceleration [Fer49]. It involves charged particles being reflected by the moving interstellar magnetic field and gaining energy. A detailled treatment can be found in [Gai90].

We first consider the situation illustrated in Figure.1.8. A plane shock front is assumed to move with velocity u, larger than the speed of sound c_s in the medium, while the velocity of the shocked gas (behind the shock front) is $v = \beta c = \frac{3}{4}u$. We consider the collision between the front and a relativistic particle with energy E_0 at an angle θ_0 between its direction and the normal to the wave front. The center of mass frame coincides with the shocked gas frame. In this frame, the energy of the particle is

$$E_0' = \gamma\, E_0(1 - \beta\cos\theta_0) \tag{1.1}$$

with $\gamma = \left(1-\beta^2\right)^{-1/2}$.

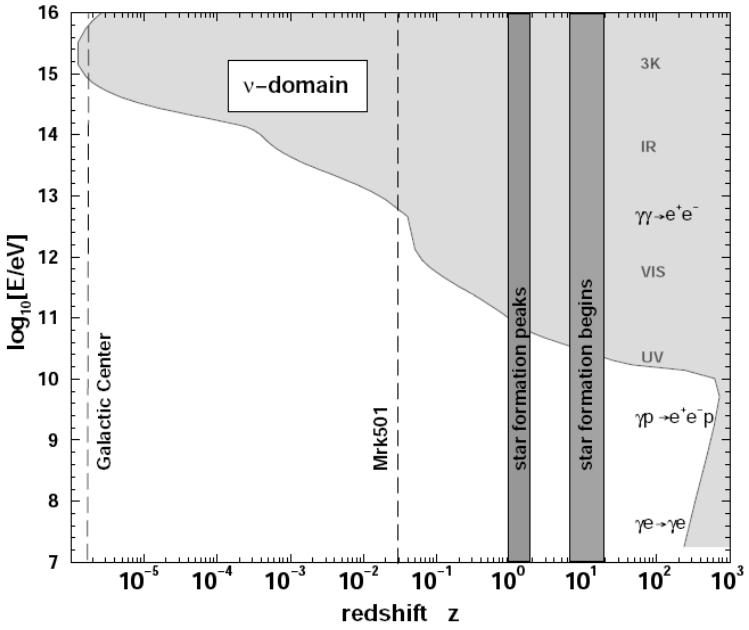


Figure 1.7: *Limits of γ-ray observation as a function of the redshift of the emitter (from [LM00]) : $z = \frac{1}{c} H_0\, r$, where H_0 is the Hubble constant and r the distance of the object. The most distant regions of the Universe are opaque (shaded region) to high energy γ-rays due to various absorption processes (indicated on the right of the figure). These regions are therefore visible only thanks to neutrinos.*

The scattering in the magnetic field is supposed to be elastic, therefore the particle escapes the gas with the same energy $E_1' = E_0'$ and with an angle θ_1. Back to the observer's frame, the energy after collision is :

$$E_1 = \gamma\, E_1'(1 + \beta \cos\theta_1) \tag{1.2}$$

Substituting Eq.1.1 in Eq.1.2, the average energy gain can be found as :

$$\frac{\Delta E}{E_0} = \frac{E_1 - E_0}{E_0} = \gamma^2 \left(\beta^2(1 - \cos\theta_0 \cos\theta_1) - \beta(\cos\theta_0 - \cos\theta_1)\right) \tag{1.3}$$

The next step is to average over angles. The flux of particles is isotropic, so the probability for a particle to reach the shock wave is proportional to its velocity perpendicular to it :

$$\frac{dn}{d\cos\theta} = 2\cos\theta \tag{1.4}$$

Then one finds $< \cos\theta_1 >= \int_0^1 \cos\theta_1 dn = 2/3$ and similarly $< \cos\theta_0 >=$

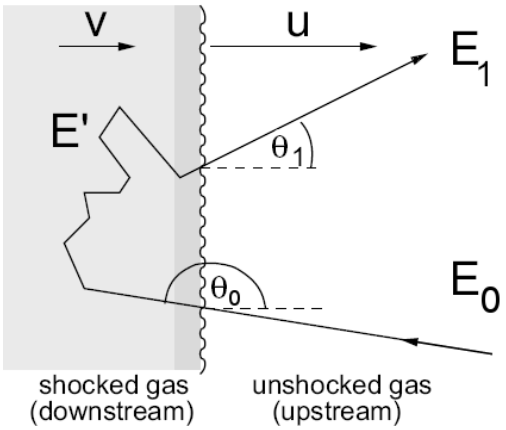


Figure 1.8: *Illustration of the first order Fermi acceleration principle. A plane shock front is assumed to move with velocity u, while the velocity of the shocked gas (downstream) is v. A particle enters from the unshocked region (upstream) with energy E_0 and undergoes multiple scattering into the shocked region before to leave the shock front with an energy E_1.*

$-2/3$. Injecting this in Eq.1.3, one gets

$$\frac{< \Delta E >}{E_0} = \gamma^2 \left(\frac{4}{3}\beta + \frac{13}{9}\beta^2 \right) \tag{1.5}$$

At first order approximation, $\gamma^2 \sim 1$ and $< \frac{\Delta_E}{E} > \approx \frac{4}{3}\beta = \frac{u}{c}$. This result is known as first order Fermi acceleration.

However, in his original paper [Fer49], Fermi proposed the acceleration mechanism to occur by elastic scattering of relativistic charged particle in the irregularities of magnetic fields within plasma clouds. The equations 1.1 - 1.2 remain valid with v the velocity of the cloud. Nevertheless changes in the geometry induce modifications in the average angle calculations. The particle gets out of the gas cloud isotropically and then

$$< \cos\theta_1 > = \int_{-1}^{1} \cos\theta_1 \, d\cos\theta_1 = 0$$

Access to the cloud depends on the relative velocity between the cloud and the particle:

$$< \cos\theta_0 > = \int_{-1}^{1} \cos\theta_0 \frac{1}{2}(1 - \beta\cos\theta_0) \, d\cos\theta_0 = -\beta/3$$

Eventually substituting these results in Eq.1.2, one obtains the second order

Fermi acceleration:

$$\frac{< \Delta E >}{E_0} = \gamma^2 \frac{4}{3}\beta^2 \approx \frac{4}{3}\beta^2 \tag{1.6}$$

Since the acceleration area has a finite size, a particle either has a probability P_{esc} to escape it after an acceleration cycle, or is re-injected and the acceleration process is resumed. Hence, the energy of the particle escaping the acceleration are after n cycles is

$$E_n = E_0(1+\xi)^n \tag{1.7}$$

with $\xi = \frac{\Delta E}{E_0}$ defined in Eq.1.3. Then reaching an energy E requires $n = \frac{\ln(E/E_0)}{\ln(1+\xi)}$ cycles. The probability for a particle to be still trapped into the region after n shocks is $(1-P_{esc})^n$. So the proportion of particles accelerated to an energy higher than E is :

$$\begin{aligned} N(>E) &= \sum_{i=n}^{\infty}(1-P_{esc})^i \\ &= \frac{(1-P_{esc})^n}{P_{esc}} \\ &= \frac{(E/E_0)^{-\gamma}}{P_{esc}} \end{aligned} \tag{1.8}$$

with the spectral index $\gamma = \frac{\ln((1-P_{esc})^{-1})}{\ln(1+\xi)} \sim P_{esc}/\xi$.

Using Eq.1.5 for first-order acceleration, one gets $\gamma \approx P_{esc}c/u$. A typical differential spectrum for strong shocks is

$$\frac{dN}{dE} \propto E^{-2.1} \tag{1.9}$$

Incorporation of non-diffusive transport properties allows the spectra index to reach up to 2.5.

Second order Fermi acceleration is usually expected to lead to particle spectra that are typically flatter than those produced in standard shock accelerations.

1.3.2 High Energy Neutrinos Sources

This section introduces a review of galactic and extragalactic objects already identified as sources of high energy γ-rays, and believed to be possible sources of high energy neutrinos. One has to take note that presently their inner structure or acceleration and production mechanisms of high energy particles are mostly uncertain. The detection of neutrino emissions from these objects could allow a better understanding of these mechanisms and therefore improve the theoretical

models. A total of 26 of these objects (Table 8.2) have been listed to be credible neutrino candidates and then used for a dedicated analysis in this thesis. A comparison of the theoretical emission fluxes for several candidates is given in Figure 1.12.

Active Galactic Nuclei

Active Galactic Nuclei (AGN) are extra-galactic objects thought to be credible candidates for high energy neutrino emission combined with already observed high energy γ-rays. The general picture of an AGN is a dust torus (r $\sim$ 100 pc) falling on and feeding an accretion disk (r $\sim 10^{-3}$ pc) around a supermassive black hole ($> 10^8 M_\odot$). Two relativistic jets are emitted perpendicular to the accretion disk and are powerful accelerators of relativistic particles driben by energy from the infalling matter. Large scale jets occur only for 10 per cent of AGN, called radio-loud AGN. The remaining are called radio-quiet AGN.

Radio-loud and radio-quiet AGNs are both subdivided into several classes of objects whose characteristics and properties depend on the angle of view of the object from the Earth (Fig.1.9).

Categories of radio-quiet AGN are radio-quiet quasars and Seyfert galaxies (named after Carl Seyfert, who was the first to observe these objects). He noticed peculiar galaxies with an unusually bright center. The core of such galaxies is expected to be very small and shows highly variable brightness. Seyfert galaxies are strong emitters in the radio, infrared , ultraviolet, and X-ray. Seyfert galaxies are classified into two categories characterizing the broadness of emission lines from ionized gas (H,He,N,O). Seyfert-1 account for both narrow (mainly hydrogen) and broad spectra while Seyfert-2 account for narrow ones only. This distinction is now dying out and replaced by a more specific nomenclature depending on the relative strengths of the narrow and broad components. Usually, the distinction between a Seyfert-1 galaxy and a radio-quiet quasar is unclear and generally depends on the luminosity of the object, the quasar being of higher luminosity than the Seyfert.

Classification of radio-loud AGN is mainly based on spectral properties. The first category concerns radio-galaxies. As the appellation suggests, these objects are characterized by a strong emission of electromagnetic radiation in the radio band emitted by synchrotron processes originating from high energy electrons in intense magnetic fields localized in the radio lobes along the jets perpendicular to the galaxy. Synchrotron radiation can be extended to more energetic wavebands if the acceleration processes are powerful enough. Radio-galaxies correspond to a view of the AGN's when the observer is roughly coplanar with the galactic plan.

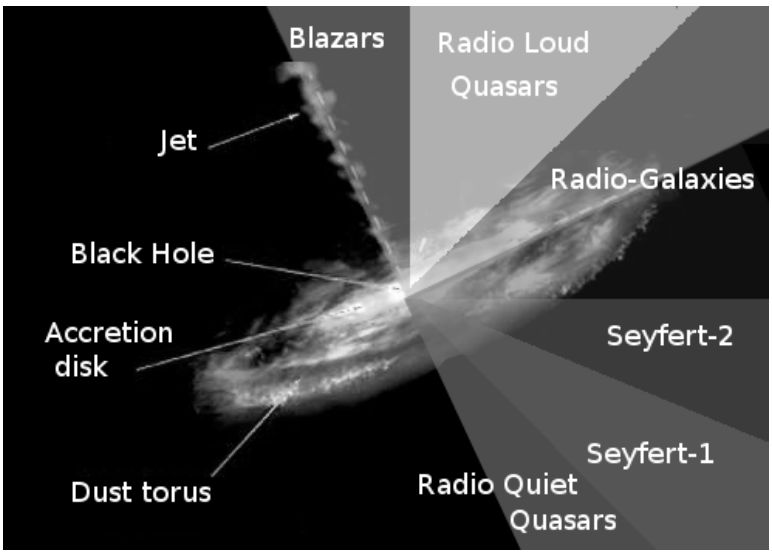


Figure 1.9: *Schematic view of the general structure of an Active Galactic Nuclei. AGN can lead to very different observations depending of the vantage point of the observer. Radio-loud AGN is shown on the upper part. Lower part illustrates Radio-quiet AGN.*

The width of emission lines actually depends on the viewing angle offered for the object and results in the division of the nucleus in two distinct parts named *Narrow Line Region* (NLR) and *Broad Line Region* (BLR). Hence, for the orientation of NLRG (or Seyfert-2 galaxies in the radio-quiet conditions) only slow motion clouds (10^3 km s^{-1}) from the first region are visible, explaining the narrowness of the emission lines observed from these objects. At larger angles (BLRG, Seyfert-2) the view of the central core is widing and permits the observation of the fast moving (10^5 km s^{-1}) cloud region close to the black hole, responsible for the appearance of broader lines in the spectrum.

Increasing the viewing angle, the next category of objects is quasars. They were first observed in the sixties, playing an important role in the identification of very distant objects. Ther are typified by strong radio emission from a point source, justifying the "quasi-stellar" denomination. Moreover, their emission spectrum does not correspond to any known chemical composition. Their origin was thus a mystery, until the observation of quasar 3C273 highlighted the fact that the observed spectrum was actually caused by a significant redshift of the original spectrum. It therefore led to the conclusion that quasars were the most distant objects ever observed in the Universe.

Finally, the last category of AGNs concerns blazars. They can be observed when the jets of the AGN are aligned with the line of view. Blazars display very strong and variable emission in the radio to X-ray range. The absence of lines

in the UV-visible region leads to the sub-classification of BL Lac objects. The alignment of the axis of the relativistic jets in the direction of the Earth makes blazars the best candidates for the observation of neutrino emission correlated with high energy photons. Typical spectra of blazars generally show a double-peaked structure. The lower peak in the soft X-ray region is thought to be due to synchrotron radiation. The higher one, ranging in the MeV-TeV region, has still an unknown and disputed origin. The search for high energy neutrinos is then motivated by the possibility to favor hadronic processes over leptonic processes in case of neutrino observation correlated to TeV γ-rays.

Leptonic models involve relativistic electrons interacting with ambient photons from synchrotron radiation or external contribution, through inverse Compton scattering. Hadronic models ([Ner02],[NR04]) assume the production of pions resulting from the collision of accelerated protons (through Fermi acceleration in the jets or in the accretion disk) with the radiative environment or gas clouds. Fluxes of γ-rays and neutrinos are related by :

$$\int_{E_\gamma^{min}}^{E_\gamma^{max}} E_\gamma \frac{dN_\gamma}{dE_\gamma} dE_\gamma = K \int_{E_\nu^{min}}^{E_\nu^{max}} E_\nu \frac{dN_\nu}{dE_\nu} dE_\nu \tag{1.10}$$

with K=1 for proton-proton process [AMH02] and K=4 for photomesons interactions . These coefficient can be determined following relativistic kinematics, as it is detailed in Table 1.1.

pp										$p\gamma$					
π^0		π^+				π^-				π^0		π^+			
$1/3$		$1/3$				$1/3$				$2/3$		$1/3$			
γ	γ	e^+	ν_e	$\bar{\nu}_\mu$	ν_μ	e^-	$\bar{\nu}_e$	$\bar{\nu}_\mu$	ν_μ	γ	γ	e^+	ν_e	$\bar{\nu}_\mu$	ν_μ
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Table 1.1: *Relation between γ-rays and neutrinos, deduced from the kinematics involving proton-proton and photomeson processes:*
For pp: $E_\gamma = 2 \times \frac{1}{6} E_p = \frac{1}{3} E_p$ and $E_{\nu_\mu} = 4 \times \frac{1}{12} = \frac{1}{3} E_p = E_\gamma$
For $p\gamma$: $E_\gamma \langle x_{p\to\pi} \rangle = 2 \times \frac{1}{3} E_p \langle x_{p\to\pi} \rangle = \frac{2}{3} E_p \langle x_{p\to\pi} \rangle$
and $E_{\nu_\mu} = 2 \times \frac{1}{12} E_p \langle x_{p\to\pi} \rangle = \frac{1}{6} E_p \langle x_{p\to\pi} \rangle = 4E_\gamma$, with $\langle x_{p\to\pi} \rangle \simeq 0.2$. is the average fraction of proton's energy transferred to the pion. [HH05]

It is natural to think that sources of highest γ-rays are regular candidates for detectable neutrinos. However, it is shown in [NS02] that TeV-Blazars are not expected to be sources of significant neutrino fluxes and less energetic photon emitters are preferred. The reason is an incompatibility between the mean free path of protons ($\lambda_p \propto \sigma_{p\gamma}$) and photons ($\lambda_\gamma \propto \sigma_{\gamma\gamma}$), compared to the size of

the AGN core. The radius of the core has to be large enough to permit the production of high energy photons from the interaction of protons with soft photons in the central region . On the other hand, it has to be small enough to allow these high energy photons to escape out of the core. It infers $\sigma_{p\gamma} > \sigma_{\gamma\gamma}$ which is not observed since $\sigma_{p\gamma} \sim 10^{-28}$ cm^2 and $\sigma_{\gamma\gamma} \sim 10^{-25}$ cm^2.

Hence, several possible MeV-TeV blazars as neutrinos sources candidates will be considered in this analysis.

Microquasars

Microquasars are binary stars emitting X-rays. Unlike other binary star systems, relativistic jets are observed in the radio band. They then are very similar to AGNs with the main difference being the composition of the central core. The compact object at the center of a microquasar is mainly a neutron star or a black hole of much smaller mass than these expected in AGN models. It is surrounded by a accretion disk composed of matter coming from the companion, usually a regular star. Due to the smaller mass of the inner black hole, the radiation of the accretion disk falls in the range of the X-ray domain in place of the UV radiation observed in quasars. Moreover, because of the compact size of the object, luminosity variations observed in microquasars are much faster than those experienced by quasars.

Several models have been proposed to explain the acceleration mechanisms taking place in these objects and finally leading to the production of γ-rays and possibly high energy neutrinos. Assuming the acceleration of non-thermal protons to high energy by shock waves in the jets [LW01] implies the production of TeV neutrinos by photo-meson processes. The neutrino burst is then believed to precede the associated radio flare by several hours. An alternative model [TRM05], requiring a massive companion star, assumes interactions between ions from the stellar wind and high energy protons ejected by the compact object, leading to the production of neutrinos together with the γ-ray emission. Constraints on hadronic models [TRM05] and neutrino flux predictions [DGWL02] have been proposed.

Microquasar SS433 is an interesting source [Mar84] due to its particular features. It is binary system composed of a black hole accompanied with a supergiant star located within a supernova remnant (see below) most probably at the origin of the central black hole. It is also characterized by a continuous jet emission, precessing in period of about 162 days and thus producing Doppler blueshifts and redshifts in the observed spectrum. Among this spectrum is the clear H_α emission line at 656.6nm, indicating proton acceleration in the jets.

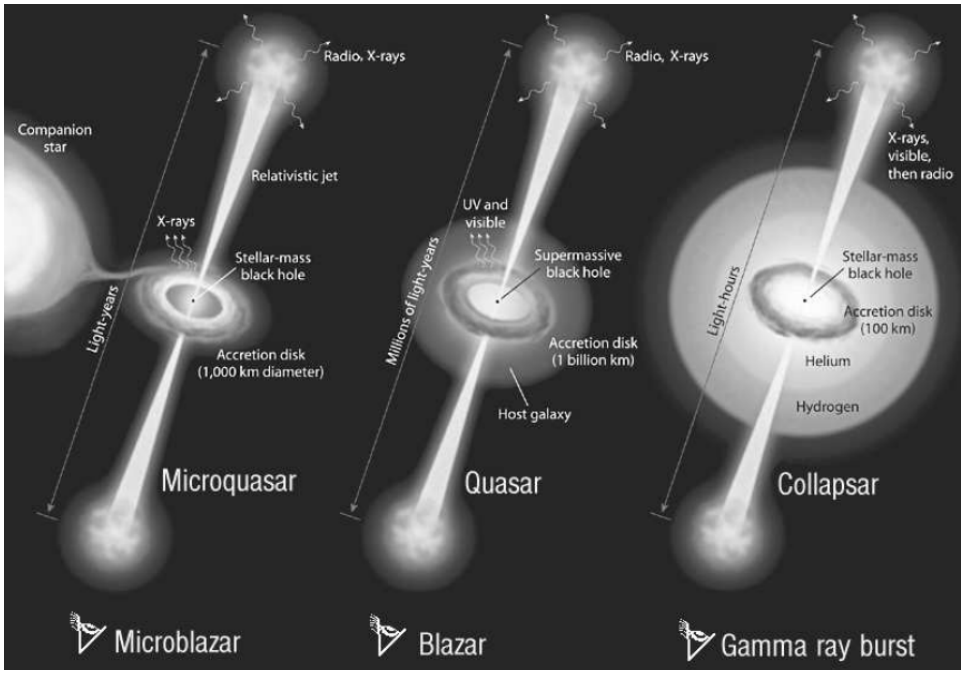


Figure 1.10: *Basic illustrations (not at scale) of the similarities between quasar, microquasar and gamma-ray burst. It is worth noticing that respective sizes of objects are very different. Jets extension is of the order of light-hour for GRB (right), light-year for microquasar (left, with its companion star) and millions of light-years for AGN (center). Picture from [Mir02]*

Supernovae and Supernova Remnants

A supernova is a very rare event, occuring only about once every 50 years in a galaxy similar to the Milky Way. There exist two basic types of supernovae. Type Ia Supernova results from the thermonuclear explosion of a carbon-oxygen white dwarf close to the Chandrasekhar mass ($1.44 M_{\odot}$) [HF60], possibly doubled in a binary system [YL98]. Type II are characterized by a prominent hydrogen contribution in their emission spectra [Fil05]. They are the energetic explosion of a massive star which can no longer sustains the hydrostatic equilibrium between radiation and gravitational pressure and its core collapse [Bet90].

During the star's life, the equilibrium is maintained by the combustion of increasingly heavier elements (H,He,Ca,...) balancing the gravitational pressure within the star. Depending on the star's original mass, those successive nuclear fusions might produce all known elements up to iron. Iron has the maximum binding energy for nucleons, implying that no more heat production can occur for fusion for any heavier element. The star undergoes gravitational collapse as

the electrons of iron atoms are absorbed by protons through inverse beta decay, producing a prompt burst of neutrinos :

$$e^- + p \rightarrow n + \nu_e$$

The star density increases forming a neutron star together with a massive release of thermal energy in the form of neutrino-antineutrino pairs with energy of order of $\sim$ 20 MeV. When the critical density of nuclei is reached, the collapse stops and a rebound phase occurs producing a shock wave that propagates outward and the remaining mass of the star is ejected, forming a Type-II supernova.

The expending shells of material formed from the outer layers of the former star after its explosion is called supernova remnant (SNR).

Observations of SNRs in the X-ray and radio wavelength range can account for synchrotron radiation from electrons accelerated to relativistic energies [Bal06]. Interactions of these electrons with ambient photons lead to high energy γ-ray emission, but too weak in quantity to match the observed flux of γ-ray. A fraction of detected γ-rays might be produced via decay of π^0 from pγ-interactions. At the same time, high energy neutrinos are expected from the decay of charged pions.

Pulsars

A pulsar is a fast spinning neutron star generating intense magnetic fields along its axis of rotation. Though neutron stars were theoretically envisaged by Baade and Zwicky in 1933 [BZ34], the first observation of a pulsar was made accidentally by Bell and Hewish in 1967. Similar sources featuring quick and regular pulses were rapidly found. Neutron stars are very small and compact objects able to spin at periods down to milliseconds, generating the intense magnetic and electric fields. The rotating magnetic field induces huge electric potential differences which accelerate charged particles to high energy [BP97].

The interactions of these accelerated particles with the surrounding ambient matter can lead to the production of γ-rays and neutrinos through mesons production and decay. Production mechanisms [Nag04] involving interactions between hadronic components assumed to be predominant in the *pulsar wind*. Neutrino fluxes have been estimated leading to the conclusion [Nag04] that pulsars with magnetic fields of 10^{12} G and a period of 1 ms should be detected by neutrino detectors. It has also been calculated for the Crab Nebulae pulsar under the hypothesis [BP97] of heavy nuclei acceleration in the magnetosphere. The neutrino flux is expected to exceed the atmospheric neutrino background for neutrino energy from 3.5 - 5 TeV, depending on several parameters (period, magnetic field intensity,...).

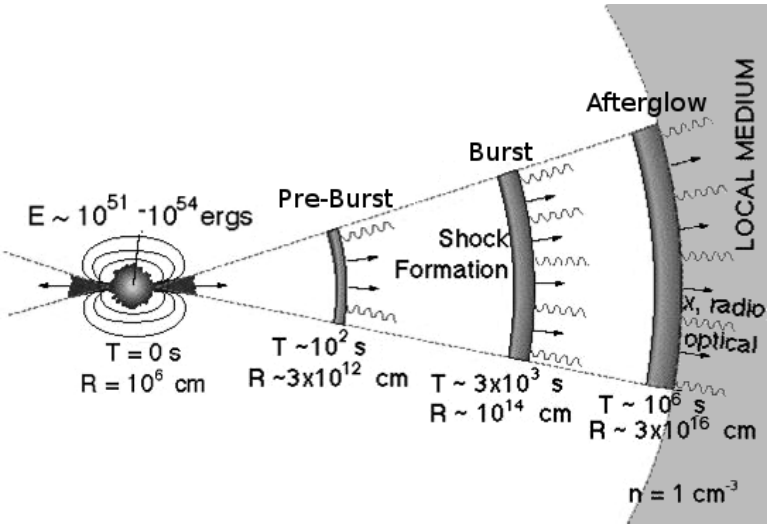


Figure 1.11: *Schematic view of the fireball model in Gamma Ray Bursts.*

Gamma Ray Bursts

Gamma Rays Bursts (GRBs) are short time pulses of soft γ-rays. Accidentally discovered in the 60's, they are one of the brightest and most energetic events in the Universe. In spite of numerous observations, their origin and their mechanisms are still uncertain and lots of theoretical conjectures have been made.

Among all existing models, the *fireball model* (Fig.1.11) is the most popular . It suggests that GRBs originate from the internal dissipation of the kinetic energy of an ultra-relativistic flow (burst) of protons into photons [RM94][BDD09], or in external blast waves (afterglow) shocks [DM99]. An even deeper mystery concerns the inner (thus hidden to any observation) object which bring to the initial matter setting off a GRB. Several theories exist as well : black holes, pulsars, massive stars, supernovae, etc. A complete review of GRB's physics can be found in [Pir04].

GRBs could be detectable sources of high energy neutrinos [WB97] of energy range $10^{14} - 10^{17}$ eV. These neutrinos are produced by internal or external shocks [Vie98] of the GRB process itself and are therefore independent of the inner engine.

Predictions for diffuse flux of neutrinos

In the case the neutrino flux from point sources taken individually is too small to be detected, the combined effects of sources isotropically distributed in the

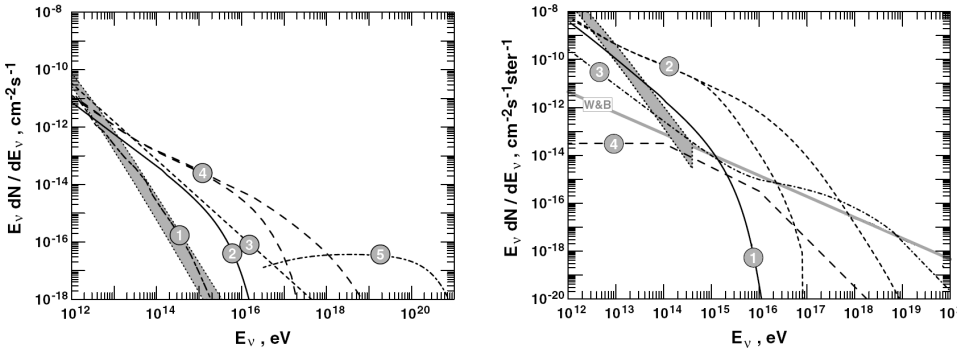


Figure 1.12: $\nu_\mu + \bar{\nu}_\mu$ *flux predictions [Sta06] for*
(a) neutrino sources candidates. The shaded area accounts for the horizontal and vertical atmospheric neutrino fluxes expected within 1° around the source.
(b) diffuse cosmic fluxes. The shaded area accounts for the horizontal and vertical atmospheric neutrino fluxes. The W&B curve is the Waxman-Bahcall bound.

sky can nevertheless be sufficient to provide a detectable signal. A detailed review of expected sources and diffuse fluxes is given in [Sta06], and illustrated in Figure.1.12.

Individual flux predictions for typical neutrino sources candidates are proposed in Figure 1.12-(a): fluxes from the sun (1), supernova remnant IC443 in case of pure hadronic origin of γ-rays observed by EGRET (2), BL Lac Mrk 501 (3), quasar 3C273 (4)[SP94] and quasar 3C279 jet(5) [Man93].

In addition to these fluxes, Figure.1.12-(b) shows expected diffuse cosmic fluxes from different general types of emitters: neutrino from the center of the Galaxy (1), AGN cores (2) [SP94] and jets (3), GRB (4). The limit proposed by Waxmann and Bahcall on the neutrino production [WB99], if the total energy of accelerated protons in sources of ultra high energy cosmic rays is converted to neutrinos, is also given.

Telescopes are in some ways like time machines. They reveal galaxies so far away that their light has taken billions of years to reach us. We in astronomy have an advantage in studying the universe, in that we can actually see the past. [...] A few centuries ago, the pioneer navigators learnt the size and shape of our Earth, and the layout of the continents. We are now just learning the dimensions and ingredients of our entire Cosmos, and can at last make some sense of our cosmic habitat.

Sir Martin Rees

2

The IceCube Neutrino Observatory

The IceCube Neutrino Observatory is a neutrino telescope under construction in the South Pole glacier. The detector is based on a three-dimensional array of optical sensors deployed deep into the polar ice in order to collect the Cherenkov radiation. This radiation can be emitted by the charged lepton resulting from a charged current interaction of the neutrino in the matter surrounding the detector or from a hadronic shower produced by a charged or neutral current interaction of the neutrino inside the detector. This chapter covers the different physical processes involved in the detection of a neutrino. We will then develop the successive steps for the detection and the reconstruction the kinematic parameters of the neutrino. Finally the future extensions of the detector are described in the last section.

2.1 Neutrino Detection

2.1.1 Neutrino interactions

Neutrinos and anti-neutrinos interact only through weak interactions with nuclei and electrons when travelling through matter. These very rare interactions occur either via the exchange of Z (neutral current) or W (charged current) bosons :

$$\begin{aligned} \nu_l + N &\rightarrow l + X \qquad \text{(CC)} \\ \nu_l + N &\rightarrow \nu_l + X \qquad \text{(NC)} \end{aligned} \tag{2.1}$$

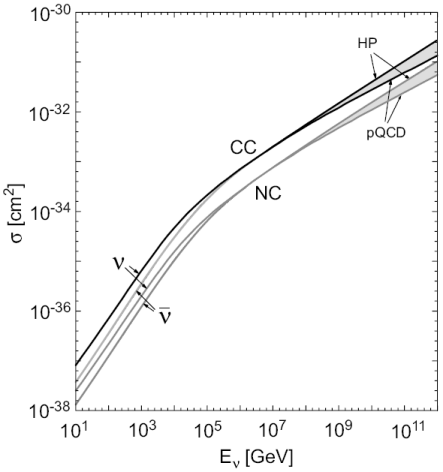


Figure 2.1: *Cross section for charged current (CC) and neutral current (NC) neutrino and anti-neutrino interaction in function of the energy (from [Ack06]). Extrapolations for high energy are shown, following* hard pomeron *(HP) and perturbative QCD (pQCD) models.[GY01]*

where l is the lepton flavor (e,μ,τ), N is the target nucleon and X is a combination of hadronic final state compatible with conservation laws.

The topology of cascade events produced by neutral current interactions in the detector does not allow the determination of the original direction of the neutrino. Thus, only charged current interactions will be considered in this work. The differential cross section for the CC-interaction can be written as :

$$\frac{d^2\sigma}{dx\,dy} = \frac{2G_F^2 M\,E_\nu}{\pi}\left(\frac{M_W^2}{Q^2+M_W^2}\right)^2 \left[x\,q(x,Q^2) + x\,\bar{q}(x,Q^2)\,(1-y^2)\right] \quad (2.2)$$

where G_F is the Fermi coupling constant for weak interaction, M and M_W are the masses of the nucleon and the W boson, and $-Q^2$ is the invariant squared momentum transfer between the incident neutrino and the outgoing lepton. q and $\bar{q}$ represent the parton distribution function of the nucleon. Finally, x and y account for the Bjorken scaling variables, respectively the fraction of the 4-momentum of the nucleon, carried by the interacting quark, i.e. $x = \frac{Q^2}{2M\,(E_\nu - E_l)}$ and the fraction of the neutrino energy carried away by the W boson $y = 1 - E_l/E_\nu$.

At low energies, cross-sections (Fig.2.1) increase linearly with the energy. The neutrino cross-section is found to be four time greater than that of the anti-

neutrinos one.[1] However, the difference decreases with the energy and at about 100 TeV, their cross sections become equal, since they now interact mainly with sea quarks of the nucleon. Within the same energy range grow the cross-sections more slowly ($\sigma \propto E_\nu^{0.36}$).

2.1.2 Propagation of charged leptons

Following eq.2.1, the neutrino interaction with a nuclei produces one of three leptons, each of which leaves a different signature in the detector. Electron and tau neutrino events, which generate respectively electromagnetic and hadronic cascades, mainly have a short range and look like a point source of light within the detector. The accuracy of their direction measurement is therefore very limited.

Unlike cascade-induced events, muons can travel long distances while continuously losing energy through ionization and radiative processes such as pair production, photonuclear interaction or bremsstrahlung. The average rate of muon energy loss can be written as :

$$-\frac{dE}{dx} = a(E) \, + \, b(E)\, E \tag{2.3}$$

where $a(E)$ is the ionization energy loss, given by the Bethe-Bloch equation (see. e.g. [Lon92] for details) and $b(E)$ accounts for the radiative processes.

Below the muon critical energy, where $a(E_c) = b(E_c)E_c$, $E_c \approx 650$ GeV in the ice, ionization dominates the energy loss, resulting in a continuous energy deposit of about 0.2 GeV m^{-1}.

Above this energy, stochastic losses start to dominate and the muon energy can be estimated via the energy-dependent secondary Cherenkov radiation along the track.

2.1.3 Cherenkov radiation

A relativistic charged particle moving through a transparent medium of refractive index n with a speed $v = \beta\, c$ faster than the speed of light c/n in that medium, produces Cherenkov radiation. Cherenkov photons are emitted with an angle

$$\cos\theta_C = \frac{1}{\beta\, n} \tag{2.4}$$

[1] interaction with u and d for ν and $\bar{\nu}$ respectively

From Eq.2.4, one can see that Cherenkov radiation is only emitted if $\beta > \frac{1}{n}$. Moreover, the Cherenkov angle increases with β to reach its maximum $\theta_{C_{\max}} = \arccos(n^{-1})$. Since the refractive index of ice is $n = 1.34$ and the muons effectively have $\beta \sim 1$, the Cherenkov angle in IceCube is :

$$\theta_c = 41^\circ \tag{2.5}$$

It is worth noting that the energy loss due to Cherenkov radiation is small compared to the total ionization loss (see previous section), even for minimally ionizing particles :

$$\frac{d^2U}{dxd\omega} = \frac{\omega e^2}{4\pi\epsilon_0\, c^3}\left(1 - \frac{c^2}{n^2v^2}\right) \tag{2.6}$$

where U is the energy of the particle, ω is the frequency of the photon. The number of photons emitted per unit track length with wavelength between λ_1 and λ_2 is given by :

$$\frac{dN}{dx} = 2\pi\,\alpha\,z^2\int_{\lambda_1}^{\lambda_2}\left(1 - \frac{1}{(n(\lambda)\beta)^2}\right)\frac{d\lambda}{\lambda^2} \tag{2.7}$$

where α the fine structure constant and z the charge of the particle. Hence, one gets that 170 photons with wavelengths[2] between 350 and 500 nm are emitted per centimeter track length by a muon travelling through the detector. As it will be seen in the next section, these photons can suffer scattering and absorption before reaching the optical sensors.

2.2 The AMANDA Neutrino Telescope

The previous section has described the various physical processes involved in the indirect detection of a neutrino with a Cherenkov detector. The basic scheme consists of the charged current interaction of a neutrino with a nucleus within or around the instrumented volume leading to the production of a charged lepton. High energy muons are particularly interesting for point source searches since they can travel several kilometers in the ice while emitting Cherenkov radiation along their paths. Holes has been drilled deep in the ice in order to bury strings of optical modules (OMs) equipped with photomultipliers tubes, along with signal and power supply cables allowing the detection of Cherenkov light. Although Cherenkov photons can be scattered and absorbed, the record of their arrival time and intensity at each sensor composing the detector array allows the reconstruction of the direction and the energy of the neutrino induced muon.

[2] sensitive area detection of the detector photomultipliers

However, the detection of cosmic neutrinos have to cope with a large background of atmospheric events from the interaction of primary cosmic rays with the Earth's atmosphere (described in section 1.2.2). On one hand, atmospheric neutrinos can also induce a relativistic muon within or in the surroundings of the detector and create a irreducible background $\sim$ 1000 times higher than the expected cosmic signal. Unlike neutrinos, atmospheric muons produced in the northern hemisphere are on the other hand absorbed by the Earth on their way south towards the detector. Since only southern, thus downward-going, atmospheric muons can reach the detector, one can get free of the atmospheric muon background by looking exclusively to the northern hemisphere and using the Earth as a filter. The distribution of arrival times during an event detection is therefore crucial to determine whether the event comes from above or bellow.

2.2.1 Detector geometry

The Antarctic Muon And Neutrino Detection Array (AMANDA) was the first neutrino detector deployed in the antarctic glacier. The deployment started in 1993 with the first prototype : AMANDA-A. It consisted of 4 strings of twenty optical modules deployed within the first kilometer of ice. Unfortunately, the effective scattering lengths observed at these depth were badly underestimated, a consequence of air bubbles trapped in the ice. This severely limites track reconstruction. It was originaly thought that these bubbles would have disappeared at 800 m, but due to extreme low temperatures, this is not reached until $\sim$1300 m. Hence, AMANDA-A is not used for data analysis, but had nevertheless been used as a calorimeter. The building of a bigger and deeper detector started in 1995.

The AMANDA-II detector went through several stages (Fig.2.2). First four strings, AMANDA B-4, referred to the inner circle deployed between 1995 and 1996. Then 6 more string were added in 1997 to constitute AMANDA-B10. The detector was extended in 1998 and 2000 with an outer ring of 9 strings to reach eventually a total of 677 optical modules distributed on 19 vertical strings. The global detector shape is a cylinder with a diameter of 200 m buried at depth between 1500 m and 2000 m, with the exception of 3 strings (#11-13) extending from 1150 m to 2350 m and a misdeployed string (#17) stacked between 1000 m and 1530m.

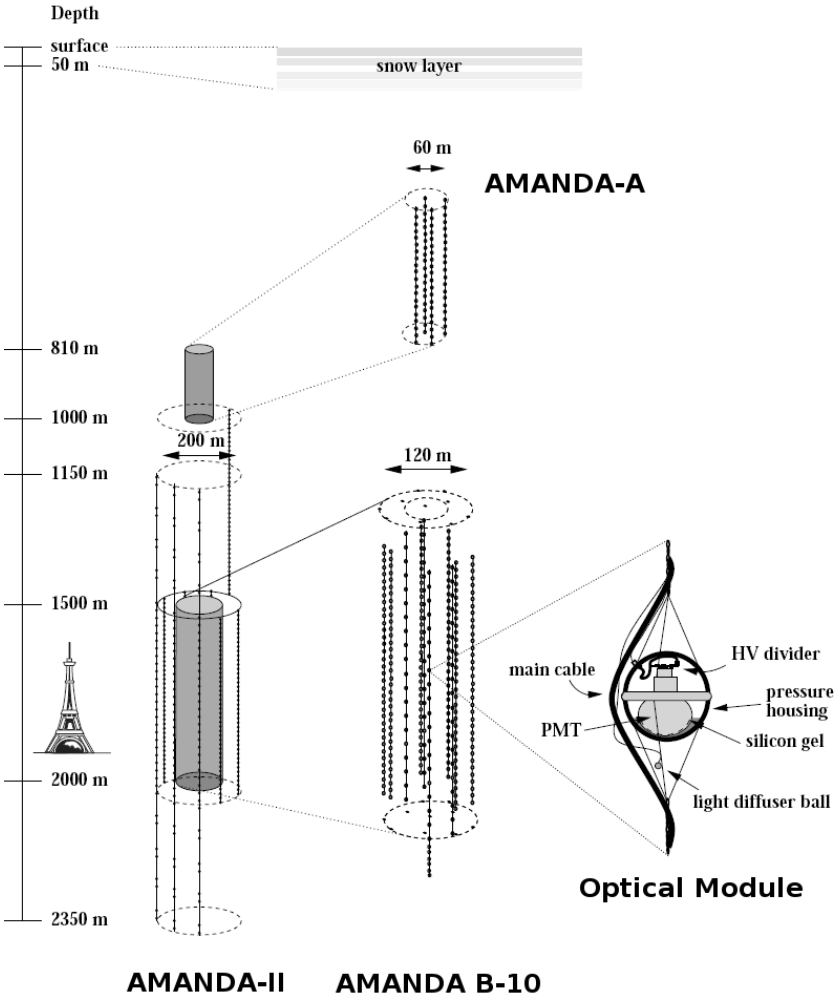


Figure 2.2: *Schematic view of the AMANDA detector, with its successive configurations : AMANDA-A, AMANDA B-10 and AMANDA-II. Main components of a deployed optical Module is shown on the right.*

2.2.2 Technical description

Optical module

The main component (see Fig.2.2-right) of the Optical Modules deployed in AMANDA-II is an 8-inch Hamamatsu photo-multiplier tube (PMT), housed in a thick pressure resistant 13-inch Benthos[3] glass sphere. A transparent silicone gel (with an adapted refraction index avoiding light attenuation) ensures mechanical and optical couplings between the PMT's photo-cathode and the glass of the sphere.

High voltage is supplied from the surface either through a coaxial cable (strings #1-4) or a twisted-pair cable (#5-19). The cable is also used to transmit the PMT pulses back to the surface. Due to the long extension (2 km) of cables, the fast signals suffer from dispersion: a standard single photo-electron PMT pulse of $\sim$ 10 ns is broadened to 200-400 ns with coaxial cables, reduced to 100-200 ns with twisted-pair. In addition, one suffers from fake pulses induced by cross-talk from adjacent cables. On most recent strings, the use of optical fibre solves the dispersion and cross-talk problems. However, optical fibre is much more fragile thus twisted pair cable was still used in parallel in case of failure of the optical fibres, which occurs in about 10% of the deployed channels.

Readout system

At the surface, the electrical pulse is amplified by the SWAMP (SWedish AM-Plifier) while the optical signal is transformed into electrical pulse by the Optical Receiver Board. The signal is then multiplied and sent into two prompt and one delayed channels.

The former are sent respectively to the Transient Waveform Recorder[4] (TWR, see below) and to a discriminator, where a digital output is created anytime the incoming pulse crosses over a predefined threshold and stopped when it crosses below it. Times corresponding to the positive (leading edge, t_{LE}) and negative (trailing edge t_{TE}) crossing are measured in the Time-to-Digital-Converter (TDC). The capacity of the buffer limits the record to 8 pulses within a window of 32 μs. The rectangular pulse is also sent to a trigger logic, which sent back a stop-signal to the TDC 10 μs after a possibly interesting event occurs.

The delayed channel is retarded for 2 μs before being analyzed in the peak sensing Analog-to-Digital-Converter (pADC), where the maximum pulse amplitude (V_{ADC}) in the interval [-2 μs ; 8μs] around the trigger time is stored.

[3] except for AMANDA-A OMs, whose housing was manufactured by the Billings company
[4] since 2003

Eventually, signal from ADC and TDC are gathered in the DAQ system along with their arrival time (measured from the GPS clock) and stored on disk and magnetic tapes. During all the readout through the discrimination and triggering process, the detector is not able to take further data. The *dead-time* of the detector depends on the trigger rate, which can be of the order of 60 Hz.

Transient Waveform Recorder Data Acquisition - The acquisition system has been upgraded in 2003 with the so-called TWR-DAQ[5]. A second prompt channel offers an independent and more efficient data acquisition system, bypassing the dead-time issue. The pulse of each DOM is sampled by the TWR every 10 ns and the integrated charged of the PMT signal is measured. Description and performance of the TWR-DAQ can be found in [Sil05].

Event trigger

The data acquisition is asked to trigger when possibly interesting event occurs in the detector. Trigger conditions have been tuned in view of both detector characteristics and signatures of expected events : storage capacity is limited and triggering on weak events induces longer detector dead time and therefore increases the probability to miss an interesting signal. On the other hand, a too restrictive trigger possibly misses more signal events.

Different triggers are implemented in AMANDA-II. The Multiplicity trigger requires a signal from 24 OMs to be recorded within a time window of 2.5 μs. The String trigger, designed to retain low energy events, requires multiple hits on one string : 6 modules are asked for inner strings (1-4) and 7 modules for outer strings (5-19) within 2.5 μs.

Calibration

Characteristic information of the hit recorded (t_{LE},t_{TE},V_{ADC}) will be used to deduce arrival time of photons (t_{OM}) and the photo-electron multiplicity N_{pe} at the triggered module.

However geometry, time and amplitude calibrations are necessary beforehand to obtain an accurate estimation of the hit characteristics.

An excellent knowledge of the position of each module is important to enable the reconstruction of muon tracks. The exact position (x, y) of a string is determined by triangulation. The vertical position z of each module on the string is first inferred from pressure measurements. The precision is then improved by

[5] μ-DAQ refers to the standard DAQ

the use of laser pulse emitted by *in situ* light sources, leading to an accuracy of about 50 cm.

Defining t_0 as all the delays induced by the acquisition system, the arrival time of photon at an OM is :

$$t_{OM} = t_{LE} - t_0 - \frac{\alpha}{\sqrt{V_{ADC}}} \tag{2.8}$$

where the last term accounts for an amplitude dependent correction on the pulse dispertion. The calibration constants t_0 and α are determined by a frequency doubled YAG-laser (532 nm), located at the surface and pulsing laser light into the ice through optical fibres ending in a diffuser ball. Signals are detected by the nearby module for its calibration.

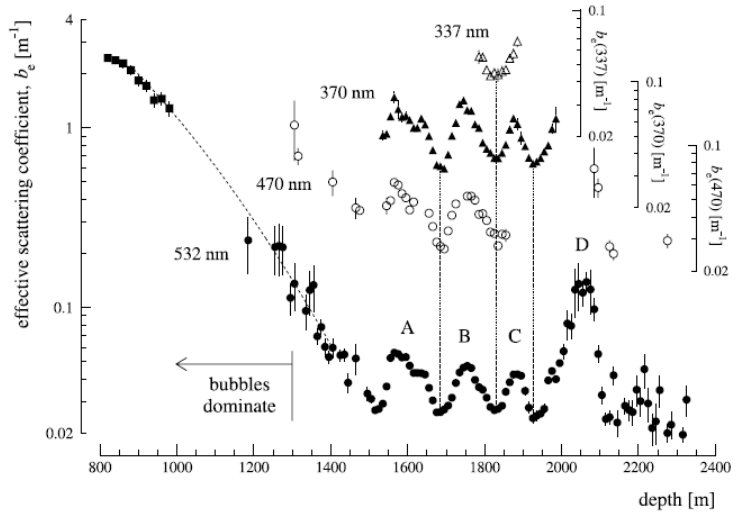
2.2.3 Optical properties of the Antarctic ice

Apart from the opto-electronic detector hardware, the most important component of the AMANDA detector is the ice itself. Its optical properties condition the performances of the detector. Even if the Antarctic ice is an extremely clear and pure environment, snow compaction and climatic changes the Earth has been undergoing for thousands of years nevertheless has caused imperfections and variations in the ice structure, modifying its optical properties. These properties have been measured thanks to *in situ* light sources deployed together with the AMANDA-II Optical Modules and covering several wavelength ranging between 313 nm to 560 nm. This topic has been extensively studied in [Aa06a].

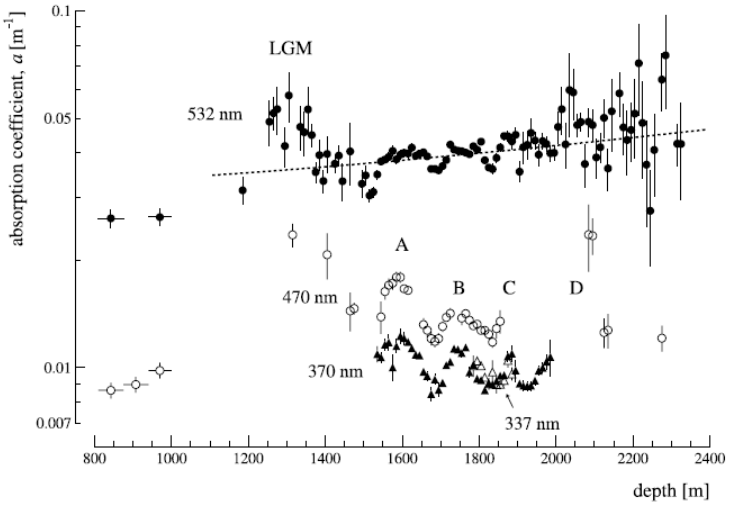
Reconstruction of the track of incoming particles is crucially depending on the efficiency and the time resolution of the Cherenkov light detection. This operation is complicated by the fact that photons undergo scattering due to the presence of dust particles in the ice. The scattering length λ_s is a few meters long. The average deviation [HP98], taking into account the strong anisotropy of the angular distribution of the scattered light, has been determined [Aa06a] to be $< \cos\theta >= 0.94$. Hence, the two parameters allow the definition of the effective scattering length :

$$\lambda_s^{eff} = \frac{\lambda_s}{1 - < \cos\theta >} \tag{2.9}$$

which is equivalent to the geometric scattering λ_s in the case of linear path. At depths of 800 to 1000 m scattering is dominated by residual air bubbles. Deeper in the ice, air bubbles progressively disappeared. Scattering effects are then due to dust layers, corresponding to different geological eras, of several tens of meters from 1400 m.



(a) Scattering



(b) Absorption

Figure 2.3: *Effective scattering coefficient (a) and absorption coefficient (b) of the polar ice vs. depth and at various wavelengths. The four peaks labelled from A to D correspond to dust layers. The peak visible at a depth of 1300 m (LGM) in absorption (b) is due to the Last Glacial Maximum. It is not visible in (a) because of dominant bubble scattering.(From [Aa06a])*

Absorption is the other phenomena which thwarts the arrival of photons to the OM. It is characterized by the absorption length λ_a, which expresses the distance at which the survival probability of a photon in the ice drops to $1/e$. The absorption coefficient $a = \lambda_a^{-1}$ can be described by a "three-component" model which take into account several contributions : the electronic structure of the crystal ice, the presence of dust layer and the molecular absorption of infrared light in the ice. The importance of these coefficients depends on the wavelength of the light. Hence, the effect of dust layer is clearly visible at short wavelength (300-500 nm) and fades away at 532 nm, because of the dominant contribution of molecular absorption at this wavelength. Bubbles do not influence absorption length.

2.3 Reconstruction Method and Event selection

2.3.1 Track reconstruction

First guess reconstruction

The event rate in the detector does not allow a complete reconstruction of the track geometry online. Nevertheless a rough estimation of the event shape is performed to partially distinguish between overwhelming downward going muon events from a possible upward going neutrino signal, and save time and CPU for a better determination of interesting neutrino-like events.

"First guess" reconstruction is based on the *residual time*,defined as the difference $t_{res} = t_{meas} - t_{OM}$ between the actual t_{meas} and expected arrival time t_{OM} of Cherenkov light not undergoing scattering. The former is measured at each triggered OM while the latter is deduced from geometrical hypothesis (Fig.2.4). Suppose a relativistic ($\beta = 1$) muon traveling along the direction $\vec{d}$ from its position $\vec{r}_0$ at time t_0. Its position at time t would be :

$$\vec{r}(t) = \vec{r}_0 + c(t - t_0)\vec{d} \tag{2.10}$$

Cherenkov radiation is emitted continuously along the track with an angle θ_C. The arrival time of the Cherenkov photons at a module located at $\vec{r}_{OM}$ is :

$$t_{OM} = t_0 + \frac{1}{c}(\vec{r}_{OM} - \vec{r}_0)\,\vec{d} + \frac{r_\perp}{c_{ice}\sin\theta_c} \tag{2.11}$$

where $r_\perp = |(\vec{r}_{OM} - \vec{r}_0) \times \vec{d}|$ is the closest distance of the OM from the track.

The approximate track direction is estimated from pattern recognition algorithms. The Direct Walk algorithm [Ste02] is based on the simple principle that

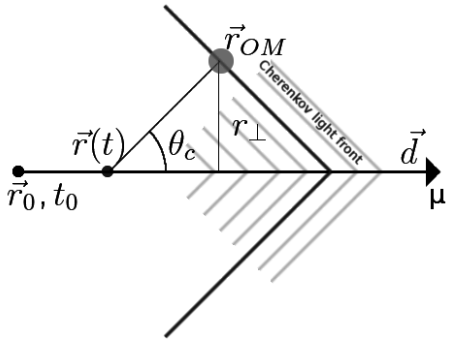


Figure 2.4: *Geometrical characteristics of unscattered photons radiation from the muon track to the OM. A muon track of vertex $(\vec{r}, t)$ and direction $\vec{d}$ emits Cherenkov photons at the moment t. Photons are detected by OM $\vec{r}_OM$ located at a distance $r_\perp$ of the track at t_{OM}.*

optical modules close to the actual muon track would experience almost *direct* hits, that is small time residuals. The method searches all pairs of hits separated by a distance d less than 50 m and with a time difference required to be

$$|\delta t| < \frac{d}{c} + 30 \quad \text{ns} \tag{2.12}$$

All these couples are potential track candidates with a quality Q depending on the number of hits associated with each of the track element within defined distance and time windows. Only best track candidates ($Q > 0.7 \max(Q)$) survive the selection. In case of multiple track, a cluster search is performed by counting the number of tracks within a cone of 15° around the direction of each candidate. The averaged direction of all tracks around the candidate with the largest number is the final track direction.

JAMS is another pattern recognition algorithm (see i.e. [Ack06]) specifically designed to distinguish between coincidental muon events and real upward-going events. It consists of 50 preselected direction search defining planes on which the hit sample is projected. "Good events" would be expected to cluster[6] while multiple coincidental events would be scattered. Basic track reconstruction is applied on each selected cluster and a neural network is ultimately used for choosing the final track.

[6] at least 7 hits are requested

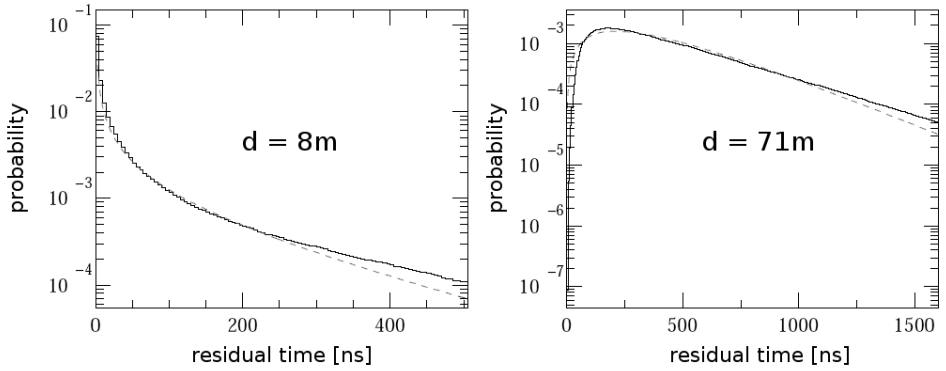


Figure 2.5: *Probability density function of residual time from MonteCarlo (in plain) and the fitted Pandel function (in dashed) for a muon track distance of 8 m (left) and 71 m (right).[Ahr04]*

Likelihood based reconstruction algorithm

The final output of the first guess algorithms can be used as a seed for the full reconstruction of event direction. It is done following a likelihood based procedure (see 3.2.1). The parameters of the muon track, $\vec{\Omega} = (\vec{r}, \theta, \phi)$ with $\vec{r}$ the vertex and θ and ϕ, respectively, the zenith and azimuth angle, are determined by maximizing the likelihood function :

$$\mathcal{L} = \prod_i \mathcal{P}(t_{res,i}|\vec{\Omega}) \tag{2.13}$$

The time residual $t_{res,i}$ can then be calculated for each hit i with respect to the track hypothesis $\vec{\Omega}$.

The time residual probability density function $P(t_{res}|\vec{\Omega})$ in function of the distance $d(\vec{\Omega})$ from the muon track is described analytically by the *Pandel function* [Pan96] :

$$\mathcal{P}(t_{res}|d) = \frac{\tau^{-d/\lambda}\, t_{res}^{d/\lambda-1}}{N(d)\,\Gamma(d/\lambda)}\, e^{-\left[t_{res}\left(\frac{1}{\tau}+\frac{c_{ice}}{\lambda_a}\right)+\frac{d}{\lambda_a}\right]} \tag{2.14}$$

with the normalization factor

$$N(d) = e^{-d/\lambda_a}\left(1+\frac{\tau\ c_{ice}}{\lambda_a}\right)^{-\frac{d}{\lambda}} \tag{2.15}$$

λ_a is interpreted as a typical absorption length in the ice. c_{ice} is the speed of light in ice. λ and τ are free parameters depending on the geometry.

The best fit from MonteCarlo simulations (Fig.2.5) of the propagation of photons in the ice leads to the following values : $\lambda = 33.3$ m, $\tau = 557$ ns and $\lambda_a =$

98 m. However the true arrival time of photons suffers several uncertainties. In addition to the noisiness of PMT, causing extra hits with a rate of 1 kHz, unscattered photons are normally distributed around $t_{res} = 0$ with an uncertainty $\sigma = 15$ ns. Furthermore scattering processes induce late arrival photons. These uncertainties have to be convoluted with the Pandel function Eq.2.14 to obtain a accurate description of the expected photon arrival time.

In order to be robust against possible local minima, the procedure is repeated 32 times with $\vec{\Omega}_n$ of the preceding iteration as a starting value for the next one.

Paraboloid reconstruction

The quality of the angular reconstruction of the neutrino trajectory is a crucial parameter to correlate its detection to a possible cosmic source. It is natural to think that longer tracks or larger number of triggered OMs would imply a better resolution of the reconstructed direction. An estimator of the angular resolution is computed by a local paraboloid approximation of the likelihood ratio around its determined minimum $(\hat{\theta}, \hat{\phi})$.

$$-2 \log \left(\frac{\mathcal{L}(\theta, \phi)}{\mathcal{L}(\hat{\theta}, \hat{\phi})} \right) = \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \tag{2.16}$$

where x and y are the error ellipse axes[7]. σ_x and σ_y are determinated by taking the best adjustment of Eq.2.16. The paraboloid error is defined as the one-dimensional resolution obtained by taking the square root of the area of the ellipse:

$$\sigma_{par} = \sqrt{\sigma_x \sigma_y} \tag{2.17}$$

Hit cleaning

A hit recorded by a optical module is not necessarily due to Cherenkov radiation in the ice. Interesting signal is contaminated by several types of noise that generates artefacts in the detector. Some are directly due to the electronics : thermal emission of photomultiplier, dark noise in the cable or cross-talk effect from signal echo in nearby cables. Furthermore some modules suffer from technical defects (noisy or unstable rates) or are simply dead and ought not be taken into account. Hence a preliminary hit cleaning is important to identify and remove fake signal and thereof reduce biases in the reconstruction of muon tracks.

[7] not necessarily oriented along the zenith and azimuth directions

Hit cleaning depends on the performed analysis. The cut procedure for muon tracks removes hits on base of :

- Time window : hits outside a time range [4.5 μs ; 11.5 μs] are removed
- Isolated hit : hits with no partner within 100 m and 500 ns are removed.
- *time over threshold* (TOT): only hits with $TOT = t_{TE} - t_{LE}$ (cf.sec.2.2.2) over a given threshold are kept.
- Amplitude. Low amplitudes result in large uncertainties.
- OM noise rate : temporary (or permanent) unusual high (or low) rates entails the rejection of the affected OMs.
- cross-talk : Large amplitude associated with small Time Over Threshold are removed by a cut on ADC vs. TOT map (individually for each OM).

2.3.2 Simulations

The entire chain of processes, from neutrino production to detector response, has been simulated in order to understand and optimize as much as possible the capabilities of the AMANDA-II detector.

The simulation chain begins with the generation of neutrino induced events. It concerns not only the cosmic signal we are looking for, but also the neutrino background produced by the interaction of cosmic rays in the atmosphere. All possible energy-dependent source spectra as well as the atmospheric neutrino energy spectrum are produced in a single simulation of spectrum E^{-1} with equally distributed neutrino-nucleon interactions within the interaction volume. A re-weighting depending on the energy and the zenith angle is therefore applied to generate events fluxes following the desired point source signal. A parametrization proposed by Lipari [Lip93] for atmospheric neutrinos is applied.

Cosmic rays air showers from primary hadrons of energies up to 10^{20} eV interacting in the atmosphere, are simulated with CORSIKA [Haa]. The propagation of both atmospheric and neutrino induced muons is described in Muon Monte-Carlo (MMC) [CR04]. The travel of Cherenkov photons in the ice from the muon track to the AMANDA modules is simulated using Photonics tables [LMW$^+$07]. The detector response is described by AMASIM [Hab]. It simulates individual OMs response to photo-electrons and transforms the number and time of each pulse into information furnished by the data acquisition.

2.3.3 Event selection

The events recorded by the AMANDA detector can basically be divided into three categories:

- direct atmospheric muon background. The huge majority ($\sim 10^9$ yr^{-1}) of events are muons produced in the atmosphere. These events are down-going in the detector[8], and reliable reconstruction methods exist to filter away these events.

- an irreducible background of muons ($\sim 10^3$ yr^{-1}) induced by the interaction of atmospheric, mainly up-going neutrinos with the earth matter.

- signal events. Only a few of them ($\sim 1 - 10$ yr^{-1}) concern high energy cosmic neutrinos

The selection of the final data set used in this analysis is throughly detailled in [Bra09]. Raw data recorded by AMANDA-II between 2000 and 2006 contains 1.29×10^{10} events for a lifetime of 1387 days[9].

A retriggering on the event sample is applied after hit cleaning. This is justified by the fact that an event with some withdrawn hits may no longer contain the required minimum of 24 hits for the Multiplicity Trigger or satisfy the String Trigger. The procedure cleans roughly 50$\%$ of the event sample.

The remaining events are reprocessed through the first guess Direct Walk (DW) and JAMS algorithms (sec.2.3.1) and cuts are applied to the zenith angles [10]: $\theta_{DW} > 70^\circ$ and $\theta_{JAMS} > 80^\circ$. About 1% of the original set of events survive the cuts and are fully reconstructed.

The final event selection is done by applying cuts on event parameters in order to reduce the sample size by another factor of 1000. The criteria used for the final selection are : the Bayesian Likelihood Ratio, the paraboloid reconstruction (Eq.2.17), the smoothness [11][Ahr04]and the output of a support vector machine [SS] trained on the given parameters. The final sample is composed of 6595 neutrinos candidates. 6100 are originating from the northern hemisphere and will be used for this analysis.

[8] muons generated in the northern atmosphere are stopped by the earth before reaching the detector

[9] after accounting for the detector dead time (about 15$\%$) as well as detector shutdowns.

[10] values are interchanged for 2005 and 2006 data

[11] characterising the homogeneity of the hits along the track

2.4 Present and Future of the IceCube Neutrino Observatory

2.4.1 The IceCube Neutrino Telescope

IceCube is the successor of AMANDA. Its construction began in 2004 and its completion is planned for 2011. When achieved, the detector will consist of 80 strings of 60 Digital Optical Modules (DOMs) deployed at a depth ranging from 1500 to 2500 m. The distance between each string is about 120 m, forming a hexagonal instrumented volume of one cubic kilometer. During its construction, the detector was successively composed of 9 (IC-9, 2006), 22 (IC-22, 2007) and 40 (IC-40,2008) strings. At the end of Polar Season 2008-2009, 59 strings have already been deployed.

The IceCube technology offers several enhancements with respect to the AMANDA detector. The main improvement is the presence of the electronic components within the module. A mainboard enables the acquisition, the discrimination and the digitization of the PMT signal before sending it toward more than two kilometer-long twisted pair cables to the surface. DOMs also contain a larger photomultiplier tube of 10 inch, an LED flasher board used for calibration, and are deployed in much clearer ice, offering higher quality data and an enhanced time resolution of 3-4 ns.

AMANDA was integrated into the IceCube array until its shutdown in May 2009.

2.4.2 IceTop

The IceTop air shower array is the surface extension of each IceCube string. It will be composed of 160 Cherenkov detector - two per string - each containing 2 DOMs frozen into an ice tank of 3 m^3. The aims of the surface detector are the study of the cosmic ray energy spectrum, as well as to provide a calibration and veto tool on cosmic ray induced muon for IceCube.

2.4.3 DeepCore

The DeepCore extension of IceCube will be composed of 6 additional strings at the center of the IceCube array. Its purpose is to extend the sensitivity of IceCube to lower energies ($< 10^{11}$ eV), thanks to a more compact array of 60 DOMS deployed in very clear ice. 10 DOMs will be above the dust layer, in a depth range between 1750 and 1860 and the remaining 50 will be below it between 2100 and 2450 m. DOMs will be equipped with high quantum

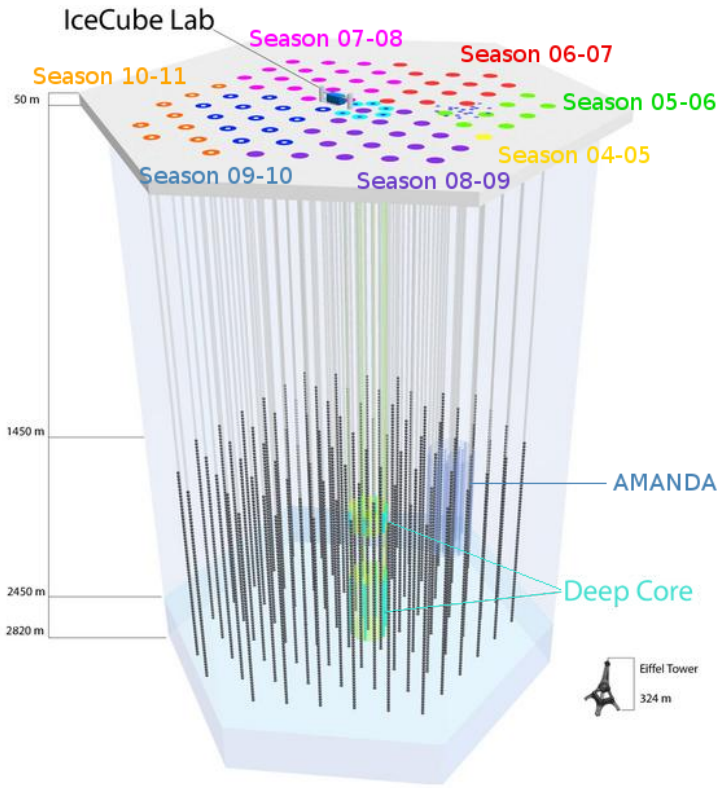


Figure 2.6: *Schematic view of the IceCube Neutrino Observatory. IceCube successive deployments are shown at the surface level (IceTop Air Shower array is not represented). The position of the former AMANDA-II detector (in blue) and the future DeepCore array (in turquoise) are shown within the IceCube array.*

efficiency PMTs. They will be spaced at a distance of 7 m on a string[12]. The very first string of DeepCore has been successfully deployed in January 2009. The 5 remaining are planned to be installed during the Polar Season 2010.

[12] 10 m for DOMS of the top layer

Part II

False Discovery Rate Controlling Procedure : A new approach on Point Source Search

A long terme, la réalisation d'un évènement improbable tend vers la certitude.

Jean Largeault

3

Statistical tools and concepts

This chapter begins with a short recall of useful statistical concepts (see for example [Jam06] for further information) as an introduction to the second section, which is dedicated to the description of a classical approach of decision theory, based on the Likelihood method. This approach nevertheless suffers several limitations motivating the new approach developed in this thesis. Hence, the third part covers the general description of the multiple testing method controlling the False Discovery Rate, as well as its extension to multivariate hypotheses. Finally this chapter concludes with the construction of limits and sensitivity.

3.1 Basics

Two different approaches are usually followed in order to check experimentally a hypothesis. Basically, the frequentist definition is based on the repetability of the measurement, while the bayesian approach implies the degree of belief, thus the subjectivity, of the user.

Frequentist approach

The frequentist approach of statistics was initialy developed by von Mises [vM57]. As suggested by the name, it consists of taking the probability of a variable x_i as the limit of the frequency at which the result is observed after an infinity (or sufficient number) of identical experiments. This comes to ask

"How often will this event have occurred after N trials? " In this case, the probability stabilizes gradually when the number of experiments increases.

$$\mathcal{P}(x_i) = \lim_{N\to\infty} \frac{N_i}{N} \tag{3.1}$$

This approach has two main limitations. First, the experiment must be repeatable under identical conditions. Second, the probability is not only a property of the experiment but also a property of the sample used to draw out this probability.

Bayesian approach

The Bayesian approach, named after the British mathematician Thomas Bayes, is based on an estimate of a *prior* distribution, which reflects the *degree of belief* of the user and depends on the information available. It can be summarized by the simple question *"How much are you ready to bet for this event to happen?"*. The basis of this approach is the so-called Bayes' Theorem and conditional probabilities.

Bayes' Theorem states that given the conditional probability of A given B, written $\mathcal{P}(A|B)$ one has

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)\,\mathcal{P}(A)}{P(B)} \tag{3.2}$$

One can use Bayes' Theorem (eq.3.2) in order to quantify the degree of belief in the hypothesis θ_i out of the set of hypotheses $\{\theta_i\}$, given the available set of data X^0 :

$$\mathcal{P}(\theta_i|X^0) = \frac{\mathcal{P}(X^0|\theta_i)\,\mathcal{P}(\theta_i)}{\mathcal{P}(X^0)} \tag{3.3}$$

The probability $\mathcal{P}(X^0|\theta_i)$ of obtaining the observed measurements X^0 given the hypothesis θ_i is supposed to be known since it reflects the user's knowledge of his experimental set-up.

The probability $\mathcal{P}(X^0)$ to obtain the results X^0 whatever the hypothesis maybe is not known but can be nevertheless considered as a normalization constant since $\sum_i \mathcal{P}(\theta_i|X^0) = 1$. One can even get rid of this by taking the ratio of probabilities between two hypotheses :

$$\frac{\mathcal{P}(\theta_i|X^0)}{\mathcal{P}(\theta_j|X^0)} = \frac{\mathcal{P}(X^0|\theta_i)\,\mathcal{P}(\theta_i)}{\mathcal{P}(X^0|\theta_j)\,\mathcal{P}(\theta_j)}$$

The prior probability $\mathcal{P}(\theta_i)$ reflects the degree of belief in the hypothesis θ_i before any experimentation. Though it is not supposed to depend on data, the prior probability is influenced by the user's beliefs since he has to put a guess on the probability (as functions of models, theories or whatever information thought to be relevant to put the prior) before making the measure. Even defining each hypothesis equally probable is a trivial example of a prior.

This supposition makes the whole method subjective since the probability can then be modified by a change in the user's knowledge.

Both approaches have advantages and disadvantages in their construction of a reliable definition of probability. The choice to prefer either one or the other is decided by the meaning the user wants to give to the probability.

3.1.1 Decision theory and test of hypothesis

Minimax decisions

Experimental data are usually analyzed in order to take a decision about the validity of an hypothesis. On one hand, it is natural to think that the decision $d = \delta(X^0)$ is actually based on the set of data X^0 measured. On the other hand, the loss $L(\theta, d)$ inferred by the fact to have taken the decision d depends on the parameter θ of the hypothesis. Taking the loss as a function of the decision rule $\delta(X^0)$, one can define the risk function as the average loss over all possible observations, described by the pdf $f(X^0|\theta)$:

$$\mathcal{R}(\theta, \delta) = \int L(\theta, d)\ f(X^0|\theta)\ dX^0 \tag{3.4}$$

Many procedures have been developed, dedicated to the most suitable choice of decision rules. However, the most important is the *minimax* method proposed by von Neumann. This method aims to choose, among all the possible decisions δ which can be taken about the parameter θ, the decision rule δ_M which minimizes the maximum risk expressed by the risk function :

$$\sup_{\theta} \mathcal{R}(\theta, \delta_M) = \inf_{\delta}\ \sup_{\theta} \mathcal{R}(\theta, \delta) \tag{3.5}$$

So the *minimax* decision might not be the most efficient one but is nevertheless the safer one.

Tests of hypothesis

Suppose that one has to decide whether an hypothesis H_0, called the *null hypothesis*, is plausible on the basis of experimental observations. One defines

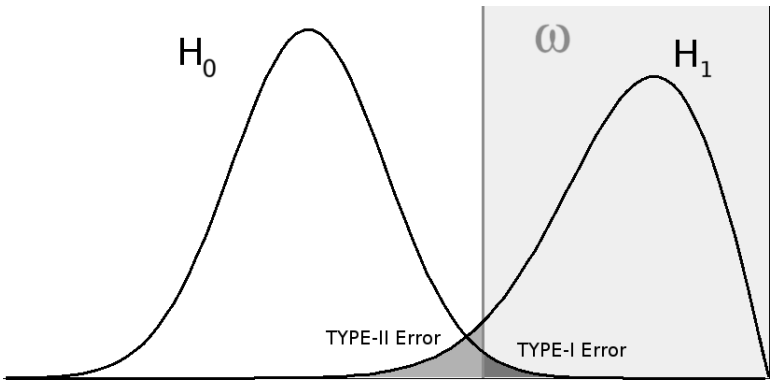


Figure 3.1: *Illustration of type-I and type-II errors related to the Null Hypothesis (H_0) and the alternative (H_1) for a given rejection region ω.*

the *test statistic* S, a function of the observations, that will allow us to take a decision as for the correctness of the hypothesis. The probability for an event, characterized by the value $s \in S$, to be compatible with the null hypothesis is called the *p-value* of the event, and is defined as :

$$p(s) = \int_s^\infty H_0(y)dy \tag{3.6}$$

One divide the space Ω of all possible values of S into a critical region ω and a region of acceptance $\Omega - \omega$. Observations s that belong to ω suggest the null hypothesis H_0 not to be true. Then choosing a test of H_0 involves choosing a test statistic S and a critical region ω.

The size of the critical region is habitually adjusted to obtain a desired *confidence level* α, defined as the probability of S to belong in the critical region ω when H_0 is true :

$$\mathcal{P}(S \in \omega | H_0) = \alpha \tag{3.7}$$

The aim of a test is to discriminate the null hypothesis H_0 against the alternative hypothesis H_1. A measure of this capability is refered as the *power* of the test $1 - \beta$, defined as the probability of X falling into the critical region ω while H_1 is true :

$$\mathcal{P}(S \in \omega | H_1) = 1 - \beta \tag{3.8}$$

In other words α is the probability that H_0 would be rejected even if H_0 was indeed true, while β is the probability that S will be in the acceptance region if

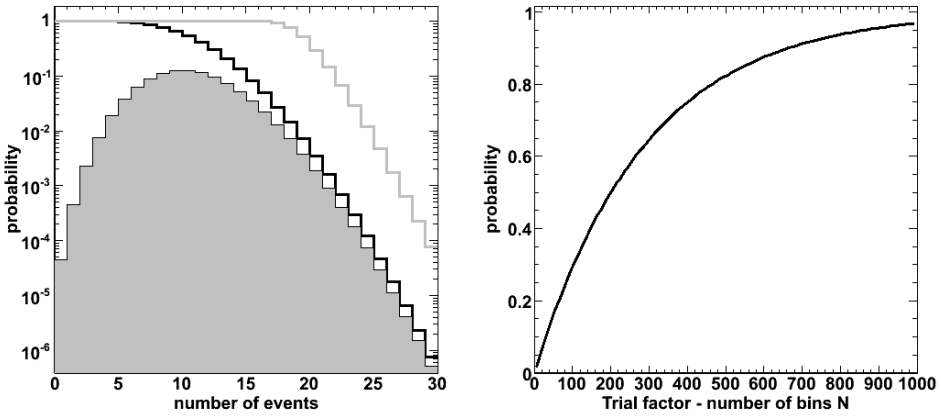


Figure 3.2: *(a) Poisson p.d.f. $\mathcal{P}(n, \mu = 10)$ (histo) and the cdf giving the probability to observe at least n events when 10 events are expected in a bin (black). However this probability must be corrected with the number of bins (*i.e. *the number of trials) for which the measurement is done. Hence the (corrected) probability to observe n events somewhere in the N =100 bins is given in grey.*
(b) The probability to observe at least $n = 20$ events in a bin, as a function of the number of bins N, for a constant Poisson mean $\mu = 10$.

the alternative hypothesis is true. There are also refered as respectively *type-I error* and *type-II error*. An general illustration is given on Fig.3.1.

Trial factor

Following the adopted point of view, a probability can be seen as a value expressing the degree of belief, or the frequency, for one event to happen. If one wants to evaluate the probability of an event to occur in a whole set of events, the size of this set has to be taken into account. That is what is called the trial factor.

Let's illustrate this notion with a simple example. We consider a sample of T events randomly distributed in N independent bins. An average of $\mu = T/N$ is expected in each bin. It is well known that the number of events actually observed in a bin follows the Poisson distribution $\mathcal{P}(n; \mu)$. The probability to observe at least such a number n of events is

$$p_1(n) = \sum_{i=n}^{T} \mathcal{P}(i; \mu) = 1 - \sum_{i=0}^{n-1} \mathcal{P}(i; \mu)$$

However, if the estimation is repeated for every bin, the probability to eventually

observe n events in at least one of the bins is :

$$p_N(n) = 1 - (1 - p_1(n))^N$$

These 3 probabilities $\mathcal{P}(n;\mu)$, $p_1(n)$ and $p_N(n)$ are illustrated on Fig.3.2-(a) for T=1000 and N=100. As an example, let's suppose that the maximal number of events counted in a bin is $n = 20$. The probability to observe exactly 20 events in a bin is $\mathcal{P}(n = 20; \mu = T/N = 10) = 1.87\ 10^{-3}$, while the probability to observe at least 20 events is $p_1(20) = 3.45\ 10^{-3}$.

However, this probability has to be corrected by the number of measurements which have been performed, to obtain the real odds to actually see such an number of events. Taking the trial factor into account, the probability to observe at least one bin with at least 20 events among 100 observations is $p_{100}(20) = 0.29$.

Obviously, this probability increases with the number of trials, as can be seen in Figure.3.2-(b).

3.2 The classical approach and its limitations

3.2.1 Likelihood method

Neyman-Pearson test and Likelihood ratio

When H_0 and H_1 are both known, the most powerful procedure is the Neyman-Pearson test [NE33]. This test relies on likelihood ratio, defined as :

$$\lambda = \frac{\mathcal{L}(x_1, ..., x_N|H_0)}{\mathcal{L}(x_1, ..., x_N|H_1)} \tag{3.9}$$

where $\mathcal{L}$ is the likelihood function for the N outcomes x_i of the tested random variable X following the distribution $f(x|H_j)$ under the hypothesis H_j with $j \in \{0, 1\}$:

$$\mathcal{L}(\vec{x}|H_j) = \prod_{i=1}^{N} f(x_i|H_j) \tag{3.10}$$

Under this procedure, the critical region will be defined by $\lambda \leq c_\alpha$ where c_α guarentees the confidence level to be at least $1 - \alpha$:

$$\mathcal{P}(\lambda(x) \leq c_\alpha|H_0) = \alpha \tag{3.11}$$

LLH Ratio Methodology for Point Source Search

This section summarizes a typical analysis procedure developed following the likelihood ratio for the Point Source search with AMANDA.

An unbinned method dedicated to the search for neutrino sources over a set of events has been proposed in [BDDP$^+$08] and adapted in [Bra08] to take account of the neutrino energy. The contribution of an event located in $\vec{x}_i$ from a source in $\vec{x}_s$ is estimated via the source pdf:

$$\mathcal{S}_i = \frac{1}{2\pi\sigma^2} e^{-\frac{|\vec{x}_i - \vec{x}_s|^2}{2\sigma^2}} \int P(N_{ch}|E_\nu) P(E_\nu|\gamma) dE_\nu \tag{3.12}$$

where $|\vec{x}_i - \vec{x}_s|$ is the angular distance between the event and the source, σ is the event angular uncertainty, N_{ch} the event energy estimator (see sec.6.2) and γ the source index spectrum.

The background pdf is given by

$$\mathcal{B}_i = \frac{P(N_{ch}|\phi_{atm})}{\Omega_{band}} \tag{3.13}$$

where ϕ_{atm} is the atmospheric neutrino flux and Ω_{band} the background contribution in the declination band.

Combination of source and background densities leads to the likelihood:

$$\mathcal{L}(\vec{x}_s, n_s, \gamma) = \prod_{i=1}^{N} \Big(\frac{n_s}{N} \mathcal{S}_i + (1 - \frac{n_s}{N}) \mathcal{B}_i \Big) \tag{3.14}$$

with n_s the number of signal events in the band. The likelihood is minimized with respect to n_s and γ. Further details and results are available in [Bra09].

3.2.2 Drawbacks

Source model

The main difference between the LLH and the FDR approaches lies in the fact that the former requires an assumption on the source behavior. This has implications at two separate levels of the LLH methodology.

First, *prior* hypotheses on the energy with respect to the source neutrino spectrum have to be determined. All these hypotheses (one per spectral index) are drawn out of NUSIM generator [BDDP$^+$08]. It then might introduce bias in the analysis if unexpected source behavior eventually occurs.

Second is the assumption made on the expected spectral index. It is assumed that the γ estimator preferentially lies in the range [2.0;2.7] and a penalty is applied for other spectra during the minimization.

Such hypotheses, either on the energy distribution or expected power law spectrum is totally unneeded in the FDR methodology.

Trial Factor

Each position $\vec{x}_i$ tested is one measurement associated with a probability p_b,namely the *pre-trial factor probability*, then a deviation with respect to the expected background. However, it has been seen in section 3.1.1 that the number of measurements (trials) increases the probability to actually find an event originally expected with a probability p_b.

Hence, identifying a 3 or 5 σ deviation is more and more probable as the number of measurements increases. Thus, it does not reflect the real probability of such a deviation to be a hint for a non-background fluctuation.

Therefore, the probability corrected from the trial factor effect is determined *a posteriori* by repeated generation of pure background skies with randomized data. This procedure is repeated for several declination bands and an interpolation is needed to obtain the corrected significance for a given location.

3.3 False Discovery Rate Controlling Procedure

The milestone of this thesis is a rejection method controlling the false discovery rate on simultaneous testing of a set of null hypotheses. This FDR procedure has been first developed in [BH95], also know as the *Benjamini-Hochberg procedure*, in the restricted context of independent test statistics. It has been extended to take account of dependencies between tests in [BY01]. A multivariate approach has been developed by [Chi08].

3.3.1 False Discovery Rate

Considering simultaneous testing of m null hypotheses, of which m_0 are true, one can build several estimators to quantify the quality of the testing procedure.

First is the *type-I error* (Fig.3.1) by the number U of false rejections of true nulls among the total number R of hypotheses rejected. It is linked to the size of the test which is the probability of making such an error at each test. The most widely used related quality indicator of a testing procedure is the *Family*

	H_0	H_1	Total
rejected	U	T	R
not rejected	V	S	$m-R$
Total	m_0	$m-m_0$	m

Table 3.1: *Summary of the different outcomes counts of multiple hypothesis tests.*

Wise Error Rate (FWER) which is defined as the probability of making at least one type-I error:

$$\text{FWER} = \mathcal{P}(U > 0) \tag{3.15}$$

This indicator might appear inadequate in case a source is made up of several rejected events where it will lead to overly conservative tests. Another indicator that we will use in the following is the *False Discovery Rate* (FDR) defined as the expectation value of the rate of false rejection among all rejections:

$$\text{FDR} = E\left(\frac{U}{R \vee 1}\right) \tag{3.16}$$

A simple interpretation is that the expectation value of the confidence level of a discovery will be 1 - FDR. It is worth noticing that FDR $\leq$ FWER . The equality states when all null hypotheses are true ($m = m_0$, $R = U$).

3.3.2 Multiple testing under dependency

The FDR procedure developed in [BY01] is a generalization of the initial procedure available for independent tests. It has been applied to various astrophysical and cosmological searches (*i.e.*[MGN$^+$]). The idea is to have an adaptive threshold on p-values (eq.3.6) which guarantees both high detection power and control of the FDR and hence the confidence level of a hypothetical detection. The procedure itself is rather simple. Given a set of N p-values and a maximum FDR input value α:

- sort the p-values in increasing order, to get the ordered set $\{p_i\}_{i\in 1,..,N}$.
- find the last index i_c so that

$$i_c = \left\{\max i | p_i \leq \frac{1}{\chi}.\alpha.\frac{i}{N}\right\} \tag{3.17}$$

  with χ a coefficient that has to be introduced to account for dependency between the tests. It is 1 under independence and the theoretical value in the worst case is $\sum_{i=1}^{N}\frac{1}{i}$. The most probable worst value is 1+1/8.[Ben]

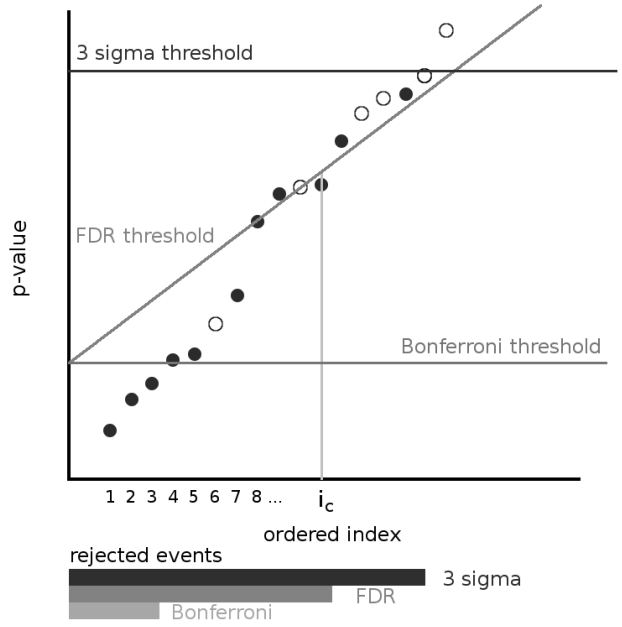


Figure 3.3: *Basic illustration of the False Discovery Rate rejection method on multiple test hypotheses. White dots correspond to background events (following H_0) while black ones account for signal events (following H_1).*
The FDR threshold is compared with two other different methods. The 3-sigma rejection (equivalent to the rejectoin area ω discussed in Figure 3.1) does not take into account the multiple character of the p-value distribution, that is the trial factor N. Therefore, it suffers many false rejections and the FDR nor FWER are controlled.
A very conservative way to control the FWER is the Bonferonni approach [Abd07], consisting in dividing the rejection threshold by the number of trials. Its counterpart is the strong decrease of the detection power.

- all p-values p_i such that $i \leq i_c$ will be rejected and considered as sources with a confidence level of (1-α).

The procedure is proven to be "minimax" (see 3.1.1) for Gaussian, Poissonian and exponential distributions [DJ].

3.3.3 Multivariate extension

An extension of the FDR procedure from univariate to multivariate p-values is proposed in [Chi08]. The purpose of this section is to stress the main results concerning FDR control using nested regions of p-values, which will be of utility in the first attempt to apply the Point Source dedicated method to bivariate

p-values.

Multivariate decision regions

One defines the multivariate p-value under a null hypothesis H_i as :

$$\vec{p_i} = (p_{i1}, ..., p_{iK}) \tag{3.18}$$

where K is the dimension of the multivariate p-value and $p_{i1}, ..., p_{iK}$ are the individual p-values under H_i. The condition that these p-values are independent under true H_i must be fulfilled. One associates the variable $\theta_i = 1$ when H_i is false, $\theta_i = 0$ otherwise. Like in univariate decision theory, one defines the critical region

$$\omega_t \subset [0,1]^K \qquad 0 \leq t \leq 1 \tag{3.19}$$

such that

$$\ell(\omega_t) = t \tag{3.20}$$

is the Lebesgue measure of the region, and :

$$\omega_s \subset \omega_t,\ 0 \leq s \leq t \leq 1 \tag{3.21}$$

One also defines

$$\begin{aligned} R(t) &= \sum_{i=1}^{n} \vec{1}\{\vec{p_i} \in \omega_t\} \\ U(t) &= \sum_{i=1}^{n} (1-\theta_i)\vec{1}\{\vec{p_i} \in \omega_t\} \end{aligned}$$

respectively the number of nulls and true nulls which p-values are in ω_t.

Finally the rejection procedure consists of rejecting H_i if and only if $\vec{p_i} \in \omega_\tau$ for the value τ of t defined as :

$$\tau = \sup \left\{ t \in [0,1] : \frac{\tau}{\alpha} \leq \frac{R(t) \vee 1}{n} \right\} \tag{3.22}$$

Definition of decision regions associated with a Null Hypothesis

The question is to find a set of ω_t satisfying the conditions given before. It is shown that the procedure can be applied to an arbitrary nested family of Borel sets 3.19. The procedure as it is expressed in Eq.3.22 requires $R(t)$ to be

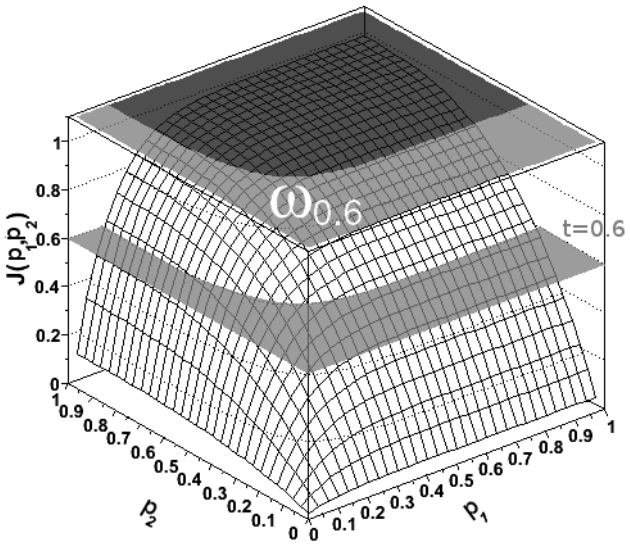


Figure 3.4: *Illustration of decision region $t = 0.6$ for bivariate test hypothesis. Here, $J(p_1, p_2) = f_1(p_1)\, f_2(p_2)$*

evaluated for each t. Then sets associated with a hypothesis test have to be parameterized with some function $J(\vec{x})$ such that :

$$\omega_t = \left\{\vec{x} \in [0,1]^K : J(\vec{x}) \leq t\right\} \tag{3.23}$$

An illustration is shown on Fig.3.4 for K=2.

One obtains an equivalent description of the procedure 3.22, consisting in applying the FDR univariate procedure (Sec.3.3.2) to the set of p-values s_i defined as :

$$s_i = J(\vec{p}_i) \tag{3.24}$$

3.4 Limits and Sensitivity

3.4.1 Flux Limits

The method used to derive the limits has been developed by Feldman and Cousins [FC98]. This approach aims at solving recurrent problems encountered when constructing traditional classical (*e.g.* empty set intervals) [Ney37] and

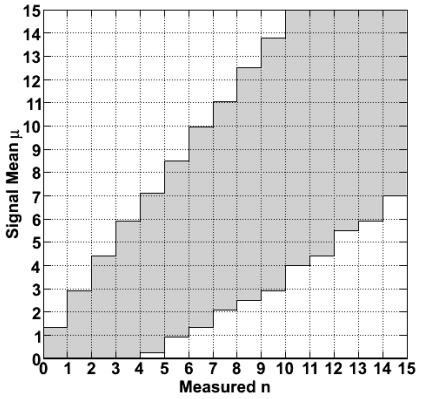


Figure 3.5: *Feldman-Cousins confidence belt for 90%CL confidence interval for unknown Poisson signal mean μ in the presence of a Poisson background with known mean b =1.5.*

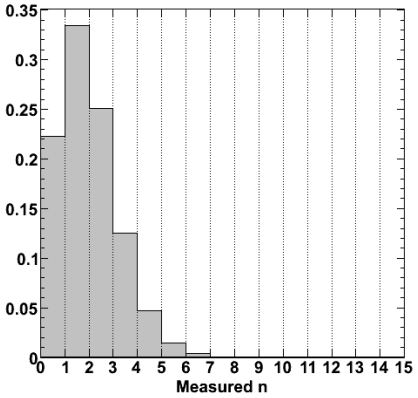


Figure 3.6: *Number of events expected under the hypothesis of pure Poisson background contribution (b =1.5)*

Bayesian (*e.g.* prior hypothesis)[Pee68] intervals. The construction of Feldman-Cousins Limits is briefly summarized in the following.

One wants to build the confidence belt $\mathcal{I}(n) = [\mu^- \,;\, \mu^+]$ of the (unknown) signal contribution μ depending on the measured number of events n. The method consists of calculating the acceptance interval

$$\mathcal{I}(\mu) = \left[n^-(\mu) \,;\, n^+(\mu)\right]$$

for each possible signal contribution μ. One supposes that $\mathcal{P}(n|\mu)$ is known. For each n, one computes $\mathcal{P}(n|\mu_{best})$ with μ_{best} being the physically allowed value of the signal which maximizes $\mathcal{P}(n|\mu)$. One then defines the ratio of likelihoods :

$$R(n|\mu) = \frac{\mathcal{P}(n|\mu)}{\mathcal{P}(n|\mu_{best})} \tag{3.25}$$

The milestone of the procedure is to take the entire set of $R(n|\mu)$ for the given μ and to order its elements in decreasing order. n is added to the acceptance region until the sum of $\mathcal{P}(n|\mu)$ meets or exceeds the desired confidence level. The procedure is repeated for every μ.

If statistical and systematic errors are negligible, the upper limit on the number of signal events at a fixed confidence level μ^+_{CL} can be related to the flux limit by :

$$\Phi^+_{CL} = \frac{\mu^+_{CL}}{R_\nu(\Phi)\; t_{obs}}\Phi \tag{3.26}$$

where R_ν is the event rate expected for a flux Φ during a observation time t_{obs}.

3.4.2 Sensitivity

In case of pure background contribution at a constant mean b, the number of events n observed varies obeying Poissonian probabilities $\mathcal{P}_b(n)$. Following this statement, one defines the sensitivity as the expectation value of the upper limit in absence of signal, averaging over the possible experimental outcomes :

$$\langle \mu_{CL}^{sens}(b) \rangle = \sum_{n=0}^{\infty} \mathcal{P}_b(n) \mu^+(n) \tag{3.27}$$

This limit (as well as μ_{CL}^+) is independent of any assumption on the signal source. However one can calculate the sensitivity for a given neutrino flux in the same way it has been done for upper limit (Eq.3.28) :

$$\Phi_{CL}^{sens} = \frac{\mu_{CL}^{sens}}{R_\nu(\Phi)\, t_{obs}} \Phi \tag{3.28}$$

In these matters the only certainty is that nothing is certain.

Pliny the Elder

4

Fixed Binning FDR Procedure applied on Point Source Search

This chapter aims at describing the construction of the complete False Discovery Rate controlling procedure dedicated to the Point Source neutrino search grounded on the statistical rejection method described on section 3.3. The first part covers the first attempt in building the method with a general binned framework. Results and a quantitative comparison with a classical method are presented in the second and third section. Finally, a first improvement taking into account the resolution of the detector is proposed in the last part of this chapter.

The question of the determination of the background behavior is a central matter that will be described in a later chapter.

4.1 General principles

This section covers the very first approach aimed at proving the feasibility of a general method which allow control of the FDR in the frame of a point source search for high energy neutrinos. The philosophy of the method is based on a simple model consisting of an even pixelization of the sky and a hypothesis based solely on the information of the reconstructed neutrino position in the sky. Thanks to the detector cylindrical symmetry and its special geographical position along the Earth's rotation axis, neutrinos are assumed to be uniformly distributed in right ascension.

So, the main idea of the neutrino point source search is to find a local tiny excess of events over the expected background by studying their position distribution. Such excess will show up as a significant discrepancy in the right ascension distribution compared to the expected flat distribution around a given declination.

4.1.1 Null Hypothesis

The first step in building a multiple hypothesis test is to define the observable which will be used for drawing probabilities out of our sample. In this case, one partitions the sky with equal size pixels, with the extra condition of requiring declination bands structure in order to make good use of the detector symmetry.

This approach has two advantages: First the counting of the number of events in a cell gives the simplest estimates of the neutrino density. Second the distribution of this number of events in the case of a pure background hypothesis is well known and follows a Poisson distribution. Moreover, exploiting the detector symmetry, the expected value of the distribution can be easily calculated. Let us define $n_{obs}^{(i,j)}$ as the number of events in the i^{th} bin of the j^{th} declination band and N^j the number of bins in the band. Then the expected background for the bin (i,j) is

$$n_{bg}^{(i,j)} = \frac{1}{N^j - 1} \sum_{k \neq i} n_{obs}^{(k,j)} \tag{4.1}$$

Hence, the Null Hypothesis is defined for each bin, associated with the probability (p-value) to only contain pure atmospheric neutrinos. The p-value of the bin $\gamma = (i, j)$ is defined as

$$P^{\gamma}(n_{obs}^{\gamma} | n_{bg}^{\gamma}) = \sum_{k=n_{obs}^{\gamma}}^{\infty} \frac{e^{n_{bg}^{\gamma}} (n_{bg}^{\gamma})^k}{k!} \tag{4.2}$$

Finally one applies the rejection procedure developed in section 3.3 on the sample $\{P^{\gamma}\}$.

4.1.2 Pixelizations

The main challenge here is to choose a configuration which allows on one hand, the control of the False Discovery Rate, and on the other hand, the best discovery potential of a non-atmospheric neutrino flux.

The analysis is limited to three families of pixelisations illustrated in Figure.4.1: rectangular, sin-rectangular and rhombus. These configurations are of particular

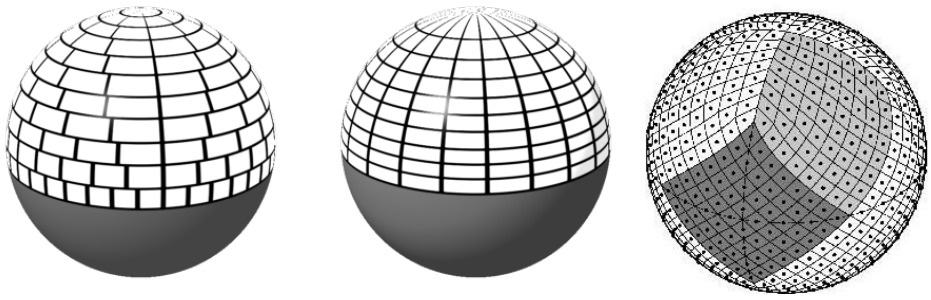


Figure 4.1: *Illustration of the three famillies of sky partition geometries used in this analysis.*
(a) Rectangular geometry has declination bands with equal angular height.
(b) Sin-Rectangular geometry has declination bands with a constant sine-projection on the z-axis.
(c) Rhombus geometry with split level N =3 (from [Hea])

interest due to their common geometric characteristics : they all are structured in declination bands and allow the construction of pixels with equal area. The two first consist of a partition in rectangular cells. Rectangular declination bands (Fig.4.1-a) have equal angular height while sin-rectangular ones (Fig.4.1-b) have constant sine-projection on the z-axis, the Earth's axis of rotation. The parametrisation of both families is defined by the number of declination bands and the solid angle of a cell. The third one (Fig.4.1-c) is an adaptation out of the HEALPIX (Hierarchical Equal Area isoLatitude Pixelization)[Hea] pixelisation developed by the COBE experiment. It consists of a sky partition with equal rhombus-shaped cells. The pixel size is entirely characterized by the unique parameter N, accounting for the level of division of the original array, composed of 12 rhombuses (for the entire sphere). Each splitting divides a pixel in 4 equal parts.

4.1.3 Simulations

The number of events expected in AMANDA is of the order of one thousand per year. The number of false discoveries tolerated to ensure a high confidence level is about 1 in 100 discoveries. Therefore, the search for the optimal pixelisation needs the generation of thousands of skies for each pixel configuration in order to have sufficiently low statistical errors to verify the control of the false discovery rate.

The MonteCarlo generation of atmospheric neutrinos is based on the declina-

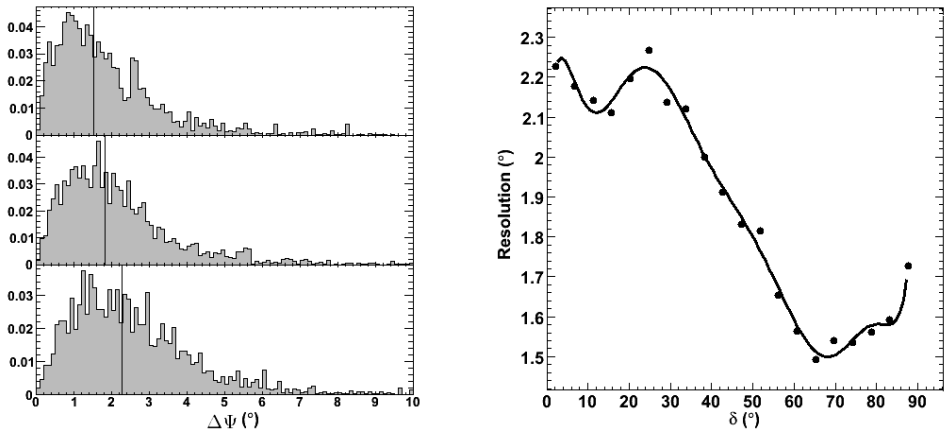


Figure 4.2: *Detector point spread function and resolution.*
Left *: Detector point spread function (PSF) for declination $\delta = 25\,^\circ$ (bottom),$\delta = 47.25\,^\circ$ (middle) and $\delta = 70\,^\circ$ (top) from MonteCarlo simulations. The resolution is defined as the median of the PSF.*
Right : *Resolution of the AMANDA-II detector as a function of the declination*

tion profile observed with 2000-2003 AMANDA data after the whole process of quality cuts (sec.2.3). A polynomial fit has been performed to have an analytical expression (Fig.6.1) and allows the simulation of event declinations statistically consistent with data without risking a bias from the repetition of samples. For similar reasons, the right ascension is generated uniformly because of the axial symmetry of the detector.

Simulated cosmic neutrinos are generated following a distribution accounting for the detector point spread function (Fig.4.2), centered on a theoretical source. The detector point spread function (PSF) is deduced from MonteCarlo simulations of the detector response, presented in section 2.3.2. It is defined as the distribution of the angular difference between the true and the reconstructed trajectory direction of events in the detector.

The number of neutrinos composing the source - referred in the following as the source luminosity - is fixed beforehand. Source positions are generated uniformly (*i.e.* constant density per steradian) over the northern sky.

4.1.4 Statistical Estimators

This analysis aims at three main objectives :

- a proof for the feasibility of a method controlling the FDR in the context

of a point source search

- a estimation of the discovery potential of the method

- the optimal pixelisation configuration reaching the best discovery potential while controlling the FDR

So as to quantify both the detection efficiency and the false discovery rate for a given configuration, one needs to take into account the relative position of the simulated source (if it exists) and the rejected cell(s).

It is worth stressing that *rejected* means here a p-value that has been found to be statistically not consistent with the Null Hypothesis. In other words, it is compatible with the Alternative Hypothesis and thus thought to be signal. Such events are those of greatest interest and are kept to carry out the analysis.

For each cell γ, three different possibilities may arise in case of rejection, associated with three random variables SRC^γ,FDR^γ,MIS^γ, illustrated in Figure 4.3 :

- One considers a **true detection** if a cell rejected by the method contains the real source position. The cell gets $SRC^\gamma = 1$.

- A **false detection** is counted for each cell which has been rejected by the method and for which no sources are located either in the cell itself nor in cells directly adjacent. These cells get $FDR^\gamma = 1$.

- The special case of **misplaced detection** occurs when a cell rejected by the method is an adjacent cell to the source-containing cell. The source detection is not counted ($SRC^\gamma = 0$). The rejection is nevertheless not considered false ($FDR^\gamma = 0$), since triggering might not be entirely due to background fluctuation, considering the proximity of the source. The use of this variable is especially important to avoid an overestimation of the FDR in skies with a high luminosity source and/or when the source is located along a pixel boundary. The cell is tagged with $MIS^\gamma = 1$.

For each sky, equivalent variables summarize the possible kinds of rejection one can observe :

- Source detected. Identify whether the source has been correctly located (1) or not (0), i.e.

$$\widetilde{SRC} = \sum_\gamma SRC^\gamma \tag{4.3}$$

- False detections. Identify the ratio of cells which are wrongly rejected over the total amount of cells rejected. This is a direct application of the FDR definition (see 3.3).

$$\widetilde{FDR} = \frac{\sum_\gamma FDR^\gamma}{(\sum_\gamma FDR^\gamma + SRC^\gamma + MIS^\gamma) \vee 1} \tag{4.4}$$

- Misplaced detections. Identify the ratio of misplaced cells over the total amount of rejected ones :

$$\widetilde{MIS} = \frac{\sum_\gamma MIS^\gamma}{(\sum_\gamma FDR^\gamma + SRC^\gamma + MIS^\gamma) \vee 1} \tag{4.5}$$

A sky corresponds to a statistical realization. It might be interesting to stress two important features. Firstly, $\widetilde{FDR}$ (Eq.4.4) and $\widetilde{MIS}$ (Eq.4.5) depend on the total amount of rejection in the sample. Conversely the true detection, $\widetilde{SRC}$ (Eq.4.3), only signifies either the discovery or non-discovery of the real source. Then $\widetilde{SRC}$ can only be either 1 or 0. Secondly, the control of the $\widetilde{FDR}$ is not required for every simulated sky. However, this requirement must be fulfilled over a large number of equivalent realizations of a studied configuration.

Hence, the false discovery rate ($\widehat{FDR}$) and the discovery potential ($\widehat{SRC}$) of a configuration are simply the average of, respectively, the false and true detections over all skies simulated. In other words, a configuration has $\widehat{FDR} =< \widetilde{FDR} >$ and $\widehat{SRC} =< \widetilde{SRC} >$.

Given the binary nature of the estimated random variables (a discovery is either true or false[1]), the statistical error ε on the estimator x after N realizations is :

$$\varepsilon = \sqrt{\frac{x\,(1-x)}{N-1}} \tag{4.6}$$

In order to obtain a reasonable statistic, 5000 sky realizations have been generated. Each sky contains the 3369 events recorded by AMANDA-II between 2000 and 2003, for which right ascensions have been scrambled. The analysis has been repeated for each binning configuration for a total of 87 configurations.

The first part determines which configurations allow the control of the False Discovery Rate. The second point shows, on one hand, whether the FDR is still controlled when a signal is added over the background, and on the other hand which configuration offers the best power of detection while controlling the FDR. Finally, this optimal configuration is used to make an extensive analysis of the properties of the FDR method in this framework, and is compared to a classical approach.

[1] it is shown in section 4.3.4 that contribution from $\widehat{MIS}$ is negligible

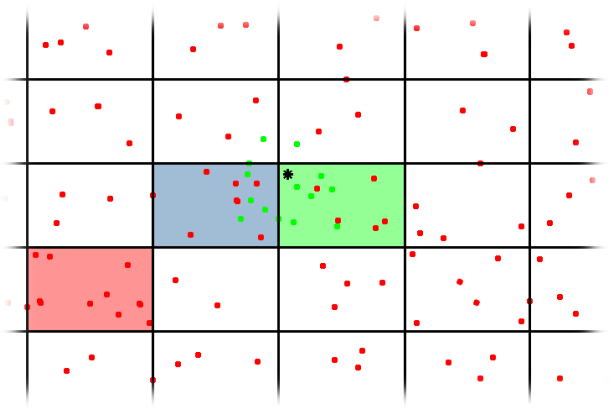


Figure 4.3: *Illustration of the types of cells in case of rejection. Dots account for source (green) and background (red) events. Black dots account for the source true positions. Rejected (colored) cells can either be a true ($SRC^\gamma = 1$ in green), a misplaced ($MIS^\gamma = 1$ in blue) or a false ($FDR^\gamma = 1$ in red) detection. See text for details.*

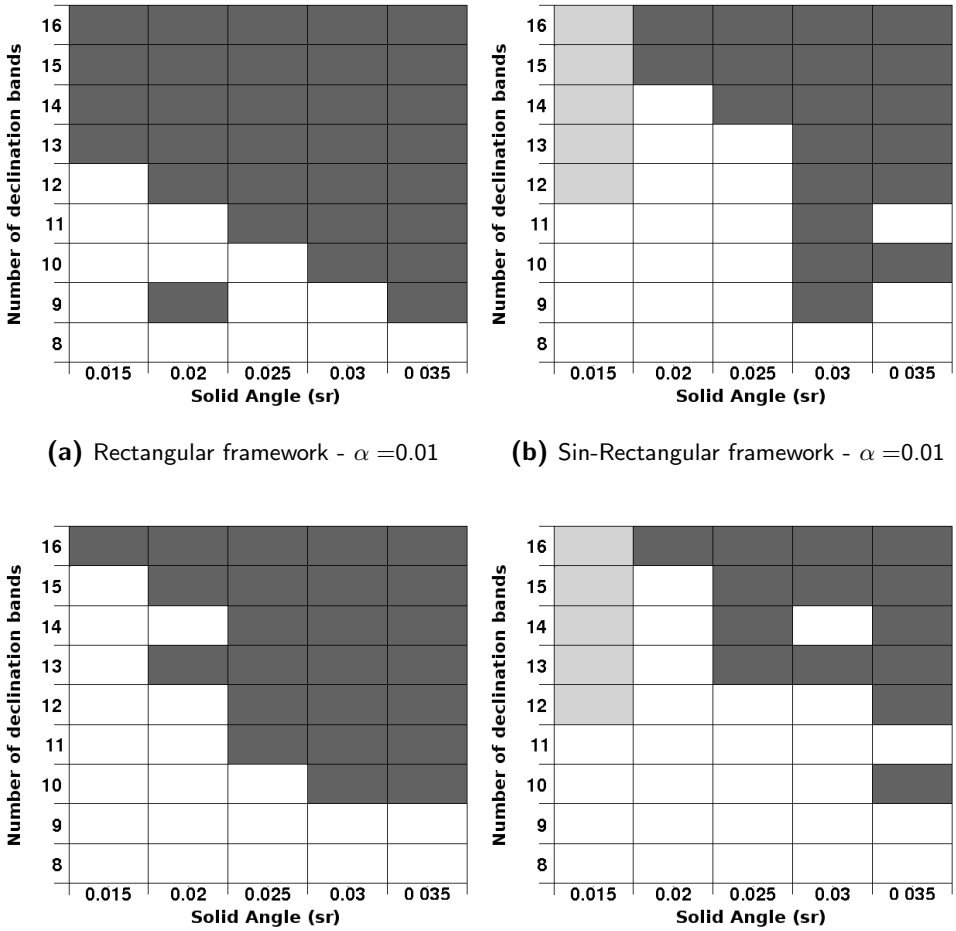
4.1.5 Control of the FDR

Only pure background skies have been used for this part. The control of the FDR has been checked twice with two requested rates of $\alpha = 0.01$ and $\alpha = 0.05$, corresponding to confidence levels of respectively 99% and 95%. The coefficient of correlation used in this analysis is the standard coefficient $(1 + \frac{1}{8})$ assumed to be optimal [Ben]. Since no source had been added, SRC^γ and MIS^γ are always zero while FDR^γ can only be 1 in case of rejection of at least one cell, 0 otherwise.

The rhombus configuration is characterized by the split level N of the minimal configuration. The control of the FDR has been verified for $N = 2$ and $N = 3$, corresponding to a partition of respectively 104 and 400 cells. The FDR is controlled within the requested confidence levels for $N = 3$. The configuration $N = 2$ overestimates the FDR, due to oversized cells.

A detailed analysis of rectangular and sin-rectangular partitions shows that the latter offers a better stability of the FDR control. Configurations have been tested for cell sizes between 15 and 35 msr and a number of declination bands between 8 and 16. Several observations can be drawn out of the results summarized in Figure.4.4.

A general interpretation of results confirms the importance of the configuration for the validity of the method. It is also seen that, like in the Rhombus partitioning, the larger the bin, the higher the FDR. This effect is enforced by the



(a) Rectangular framework - α =0.01

(b) Sin-Rectangular framework - α =0.01

(c) Rectangular framework - α =0.05

(d) Sin-Rectangular framework - α =0.05

Figure 4.4: *Status of the FDR control for Rectangular (a,c) and Sin-Rectangular (b,d) frameworks. These results have been obtained with 5000 generations of background skies of 3500 neutrinos. The white boxes accounts for controlled FDR with a requested confidence level of 99% (a,b) and 95% (c,d), black otherwise. The grey boxes are unstudied configurations.*

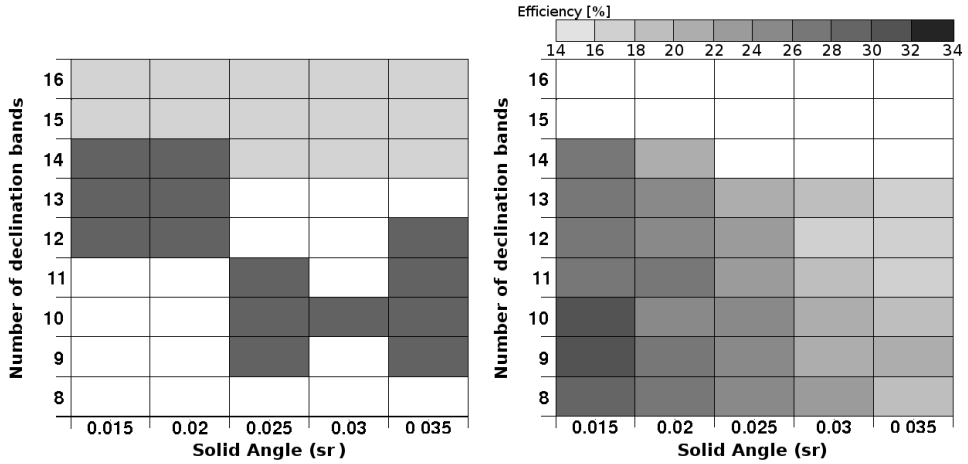


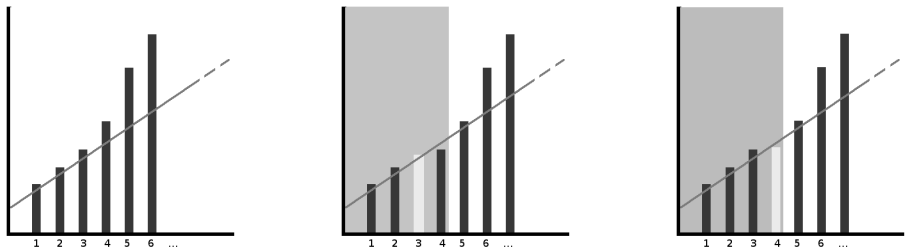
Figure 4.5: *FDR (a) and efficiency (b) for the sin-rectangular configuration on 3000 simulated skies with 3500 background events and a source with a luminosity of 1 to 30 neutrinos. In (a) the white boxes accounts for controlled FDR with a requested confidence level of 99%, black otherwise. The grey boxes are unstudied configurations.*

number of declinations bands, indicating that narrow cells are more sensitive to false detection than square ones. This is most probably due to the combination of two effects. The first effect concerns the background calculation. It has been seen in Eq.4.1 that the background estimation for a cell takes into account the number of events counted in bins of the same declination band. Increasing the number of declination bands will lead to narrower bins. Since the angular size of the bin is fixed, narrower bands are compensated by broader cells, which in turn yield few bin per band. This eventually produces a estimation of the expected background value prone to larger fluctuations. Following the statistical Law of Large Numbers [Jam06], this effect is negligible when the expected number of events and the number of cells are sufficiently high.

The second effect concerns the cell shape and the event distribution within a cell. Broader cells increase the probability of gathering background fluctuations from more distant parts of the sky and suffer from an effect of stacking since the relative position of the events within a bin is unknown. This point is of particular importance and will be discussed further (Sec.5.2) in this thesis.

4.1.6 Detection power

Following the previous results, the analysis of the detection capabilitiy of the method is focused on the sin-rectangle pixelisation. A series of 100 randomly



(a) **(b)** **(c)**

Figure 4.6: *Sketches illustrating how a source cell triggers the rejection of background cells, which have not triggered with pure background sky. Background cell p-values are in black, the source cell p-value is in white. The rejection area is highlighted in grey :*
(a) Pure background (or weak source) configuration. All p-values are over the threshold, no rejection ($\widetilde{FDR} = 0$)
(b) The source cell p-value induces a shift in the ordering, and the background p-value no longer crosses the threshold: 3 false rejections over 4. ($\widetilde{FDR} = 0.75$)
(c) The source cell p-value is under the threshold: 3 false rejections over 4. ($\widetilde{FDR} = 0.75$)

located sources are added over the background skies. Each source is composed of 1 to 30 neutrinos distributed according to a 2D-gaussian corresponding to the PSF (Fig.4.2) of the detector. Each sky contains only one source. Results are shown (Fig.4.5) for configurations with a cell size between 15 and 35 msr and a number of declination bands between 8 and 14.

The control of the FDR for these configurations is shown in Figure 4.5-(a) and is compared with Figure 4.4-(b), while the efficiencies are shown in Figure 4.5-(b). The most striking feature is that the presence of a source modifies the control of the FDR. This can be understood by analyzing the rejection method in details. The source cell usually presents one of the smallest p-value. This can play either a positive or a negative role on the FDR estimator.

When ordering the p-values in order to apply the rejection, it sometimes happens that the source cell shifts the ordering one step to the right in such a way that a background cell's p-value, which would have been over the threshold (and then not rejected) in a pure background sky (Fig.4.6-(a)), finds itself slightly under the threshold due to the source cell's p-value (Fig.4.6-(b)).

It can also happens that the source cell's p-value, though not being the lowest one, does not cross the threshold and then induces the rejection of all the (background) p-values with even lower p-values (Fig.4.6-(c)).

It might be interesting to emphasize the fact that in any of the cases of Figure 4.6, the threshold line remains unchanged since the number of cells is identical. What is changed is the distribution of p-values.

However, configurations ensuring the control of the false discoveries are those offering the best source identification. This is not surprising since the identification of a true source together with the rejection of a false event reduces the contribution to the FDR estimator. Indeed the absence of detection of the true source implies $\widetilde{FDR} = 1$ while $\widetilde{FDR} < 1$ otherwise. Moreover, in the case of the most efficient configurations, the source gets the lowest p-value and is in this case usually the only rejected cell.

These results illustrate the importance of the adaptive threshold in the rejection algorithm, as well as the importance of the whole sample of p-values, in place of taking them individually.

Eventually, the configuration that guarantees the best detection power while controlling the FDR is the sin-rectangular frame composed by cells of 15 msr, over 10 declination bands, for a total of 420 bins of 8.6° width. This partition will be used for a detailed analysis of the method's properties.

These preliminary results have been obtained by taking the 3369 events of 2000-2003 AMANDA-II data[2] with scrambled right ascension . It is interesting to study the performance of the method by quantifying the efficiency of the procedure as a function of the number of background neutrinos, i.e. the lifetime of the detector, and the number of source neutrinos, related to the source expected flux.

Analyses of the FDR control and the source detection efficiency with the chosen optimal configuration have been done for skies composed with a background varying between 1000 and 10000 events and a source with a luminosity between 1 and 40 neutrinos.

Results synthesized in Figure 4.7 show that the false discovery rate is uniformly controlled within statistical errors. This proves the robustness and the data independence of the procedure. With no surprise, the detection power increases with the source flux and decreases with the duration of data-taking (hence the number of background event recorded) for a given source luminosity. The latter does not nevertheless mean a decrease of the detection power with longer data-taking, since the flux is relative to the ratio between the total number of source neutrinos and the number of background events. Hence, for a source with a constant flux, twice as many source events are expected when doubling the lifetime of the detector. As an example, 30% of sources with 15 neutrinos

[2]equivalent tests have been performed on background skies where the event declinations are uniformly distributed. The results are similar.

have been detected over a total background of 2000 events. So doubling the acquisition time means that 30 sources events are expected over 4000 events, leading to a detection probability enhanced to 60% for an equivalent flux.

4.2 Comparison with a fixed threshold method

The purpose of the present section is to give a qualitative comparison of the efficiency of the FDR controlled method with a classical "fixed threshold" method. A common illustration is the traditional 5 sigmas rejection requested in most analyses to claim a discovery.

4.2.1 Threshold determination

The first step of the fixed threshold procedure is to find the rejection threshold which will be applied to the p-value sample. This part is done on background skies only. For each realization, the p-value of every cell is calculated rigorously the same way as it was for the previous procedure. However, only the maximal excess parameter, i.e. the p-value of the cell which has the weakest probability, is taken for the estimation of the threshold. Looking at the distribution obtained for a statistically sufficient number of equivalent skies, the rejection threshold ξ_t can be set so that the rate of maximal p-values above the threshold is the requested type-I error rate (Fig.4.9).

By construction, the control of type-I errors on pure background skies is guaranteed with this approach.

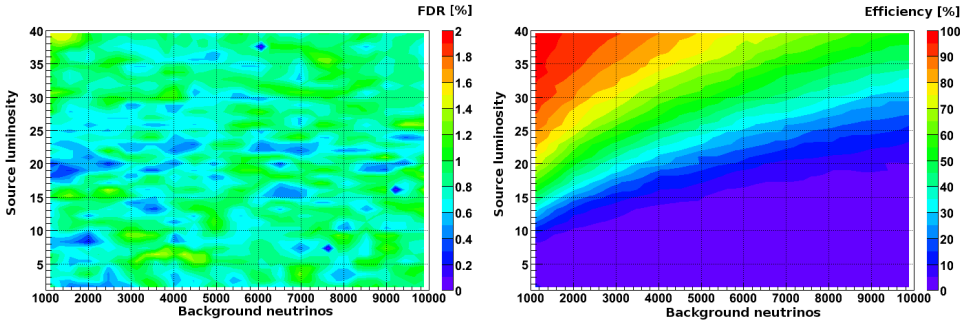
4.2.2 Type-I error and Discovery potential

On one hand, we are interested to know whether the error control is still ensured when signal is added over the background, and on the other hand, the detection power of the method compared to the FDR controlling procedure.

In order to answer these two questions, the procedure consists of determining the p-value of each cell in the now usual way, and simply rejecting the p-values exceeding the threshold. The sorting of rejected cells into true or false discoveries is done in the same way as for the FDR procedure (Fig.4.3).

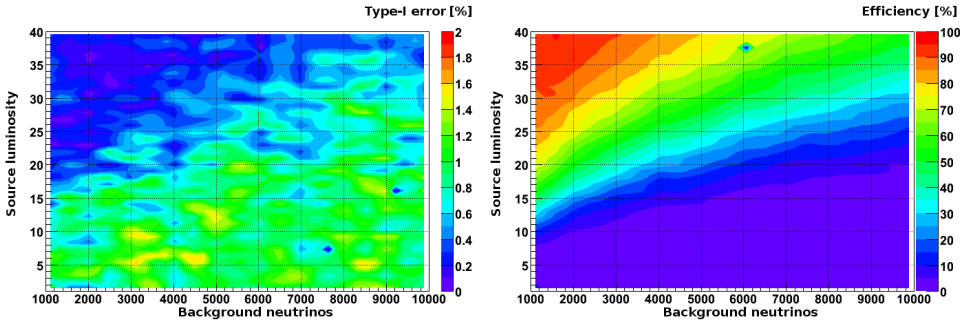
4.2.3 Results

This basic method has been analyzed for the same cell parameters used for the FDR analysis. Background studies show that the threshold, determined for



(a) False Discovery Rates **(b)** Efficiencies

Figure 4.7: *FDR (a) and efficiency (b) for the sin-rectangular configuration, 10 declination bands and a bin size of 15 msr, with the FDR procedure. FDR requested is 1%.*



(a) Type-I errors **(b)** Efficiencies

Figure 4.8: *Type-I error (a) and efficiency (b) for the sin-rectangular configuration, with 10 declinations band and a bin size of 15 msr, with a fixed threshold method . FDR requested is 1%.*

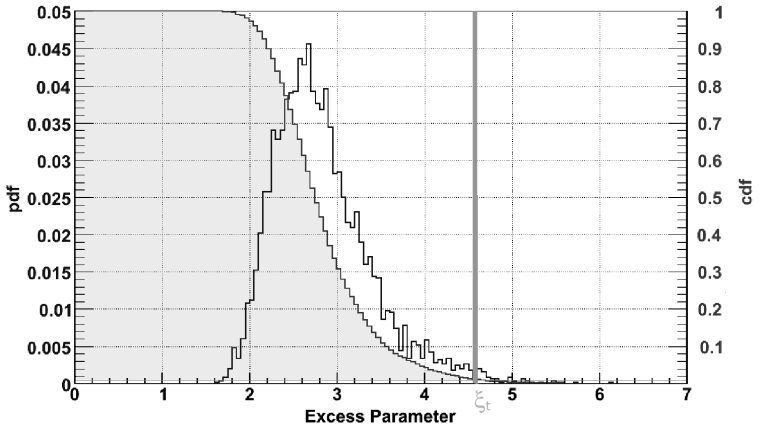


Figure 4.9: *Fixed threshold method : distribution (white histo) of the maximal excess parameter, i.e. the p-value of the cell which has the weakest probability, and the probability (grey histo) to reach this excess. The threshold (black line) is the value of the parameter where the probability is 0.01. This threshold guarantees the confidence level on background only.*

each configuration, is quite stable and is about 4.6. This result can easily be understood by making a rough calculation on the expected threshold while considering independent p-values. The studied partition is composed of 420 cells. Then each realization constitutes a sample of 420 p-values. The probability to have at least one p-value over the critical p-value p_t is (see section 3.1.1 for detailed illustration) :

$$\mathcal{P}(\xi^\gamma > \xi_t) = 1 - (1 - p_t)^{420}$$

When this probability is requested to be equal to 0.01, one finds the critical significance, defined as $\xi_t = -\log_{10} p_t = 4.62$, which is very close to the value numerically found.

Trials were repeated to ensure that possible discrepancies observed between both methods are only due to the different approaches and not to statistical fluctuations.

No major discrepancy have been observed for the detection power . However, type-I errors (Fig.4.8-(a)) are not controlled for all configurations. More precisely, high signal over background ratios underestimate the fixed threshold method, while very low signals overestimate it. In other words, this means that the confidence level of this method does depend on the data. This unwanted phenomenon is completely unobserved with the FDR controlling procedure (Fig.4.5-(a)). Although the fixed threshold method used in this comparison

might seem simplistic, this result gives a hint of the potential of the FDR procedure.

4.3 Smeared neutrinos

This section decribes a first improvement of the procedure by taking into account the resolution of the detector. The idea is to consider the neutrino events in the detector not as point-like events, but to define a local probability of presence around the reconstructed neutrino position. The smearing of the event depends on the detector point spread function, which itself depends on the declination.

The main consequence of this approach is that any event might influence the p-value of several cells, proportionally to its probability of presence within them. It is nevertheless obvious that the sum of an event's contribution must be 1. The motivation behind this endeavour is to significantly reduce the effects observed for events located on or near pixel boundaries. In the first attempt developing a FDR controlling procedure, it has been shown that a neutrino being on one side or another of a pixel boundary had a non-negligible impact on the p-value calculation, since it would have been counted entirely for a unique cell and then have an identical weight as the neutrino which is located near the center of the cell. It is clear that taking the position uncertainty may significantly reduce this biais.

4.3.1 Probability Distribution Functions

In the first procedure, the number of neutrinos counted in a cell was an integer, expected to follow Poisson's distribution.

Due to the *smearing* of the neutrinos around their reconstructed positions, this statement will obviously not be valid anymore. It is then necessary to redefine both the variable and its probability distribution function which will be used to estimate the background.

The main difference with respect to the previous procedure is the continuous behavior of the variable used to determine the p-value, since a neutrino contribution can now be any value between 0 and 1. Important correlations between connected cells are introduced due to the possibility for a event to contribute to several of them. Then, it can be easily understood that the expected distribution of the estimator is much more complex than a Poisson distribution. An analytical description ends up being very tricky, if not impossible. Furthermore, as the detector resolution depends on the declination, the p.d.f. of each declination band has been inferred from a numerical representation of simulated

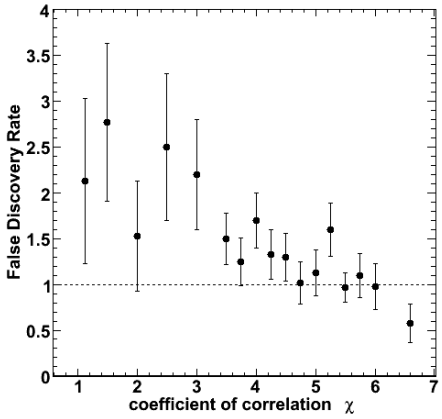


Figure 4.10: *False Discovery Rate in function of the coefficient of correlation. The requested FDR is 0.01.*

background events.

4.3.2 Correlation effects in the FDR procedure

Since the smearing introduces new strong correlations between the cells and subsequently between the p-values, the coefficient of correlation applied in the method for the point-like approach is not valid and would induce an enhanced FDR when applied to smeared events. Thus, tuning of the coefficient of correlation (χ) has to be done in order to recover control of the FDR. Analysis of the FDR as a function of the coefficient has been performed on background skies. The requested FDR is $\alpha = 0.01$. Results are shown on Figure.4.10. As expected, the FDR is overestimated with low coefficients. However, the rate decreases with higher coefficients before reaching the requested rate from $\chi \sim 5$. As a result, the new coefficient of correlation is fixed at 6.0.

4.3.3 Detection power

It has been seen through Figures 4.4 and 4.5 that the control of the FDR on background skies does not automatically infer the control of the FDR when a source is inserted. Moreover, the increase of the coefficient of correlation could possibly lead to a loss in terms of detection power. Thus the false discovery rate and detection efficiency have been estimated with the smearing approach and the results have been compared with the point-like approach. The same configuration (sin-rectangular with 10 declination bands and a bin size of 15

msr) and the same sample of background skies have been used in order to expressly see the impact of the smearing on the two criteria.

A total of 12500 background skies containing 3500 background neutrinos and sources with luminosity between 1 and 20 neutrinos have been generated and analyzed. Despite of the stronger coefficient, *i.e.* a lower threshold, the smearing approach gives a total detection efficiency of $21.7\% \pm 0.5$, compared to 14.5 $\%\pm$ 0.3 with the point-like approach. The false discovery rate is controlled within statistical errors (1.08 $\%\pm$ 0.09) for the smearing. As expected from the last results the FDR is controled on this sample for the point-like approach (0.78 $\%\pm$ 0.09).

4.3.4 Effects of smearing

Another interesting point of this comparison is the knowledge of the contribution of the smearing in the (non-)detection of true and false sources. The number of rejected cells identified as false, true or misplaced sources following the variable defined in section 4.1.4 are listed in the Table 4.1. The first and second rows correspond to the cell rejections due to respectively the point-like and smearing method only. The third row accounts for the rejections made with both methods.

It appears from these results that the smearing brings a real enhancement compared to the point-like method, since it allowed the detection of 933 new sources which the point-like approach did not spot. The counterpart of this is an increase of the false discoveries, though compatible with the requested rate. On the other hand, it is clear that the contribution of the point-like method is negligible since it allows the detection of 31 new sources (i.e. less than 2%) while infering even more false detections.

There is therefore a clear evidence for the supremacy of the smearing approach with respect to the point-like one.

It is worth noticing the importance of separating the misplaced identifications from FDR estimates. It is not surprising to observe that mislocations are mainly

	cell type			
	FDR	SRC	MIS	TOTAL
pointlike	53	31	1	85
smearing	102	933	64	1099
both	63	1779	14	1856
total	218	2743	79	3040

Table 4.1: *Comparaison between point-like and smeared neutrinos method.*

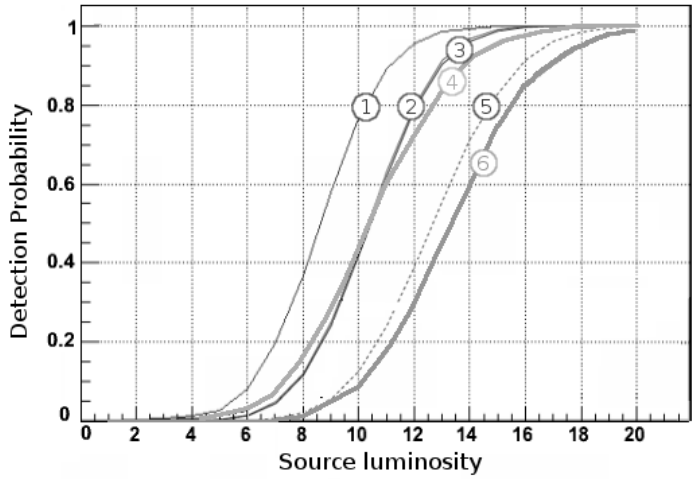


Figure 4.11: *Detection probabilities of a source located at 22.5° of declination, over 1000 background neutrinos, for :*
(1) Smearing-FDR 99% CL (if mislocations do not contribute to the FDR estimator)
(2) Unbinned Likelihood 5σ paraboloid
(3) Unbinned Likelihood 5σ gaussian
(4) Point-like-FDR 99% CL
(5) Binned 5σ
(6) Fixed threshold method 99% CL

occurred with the smearing procedure. Even if their occurrence is small ($\leq 0.5\%$), taking them as a false discovery would have lead to the non control of the FDR, with regard to the very low rate requested.

4.3.5 Comparison with other methods

FDR procedures have been compared with other statistical methods used for searching neutrinos point sources. Several are based on likelihood ratio ([NK06], [BDDP$^+$08]) or on classical binned method ([Hau04], [Ack06]). To fulfill a fair comparison, the characteristics of simulated skies in these analyzes have been reproduced. Hence 5000 realizations composed with 1000 background neutrinos per sky, and sources located at declination 22.5°, with a luminosity in the range between 1 and 20 neutrinos, have been performed and analyzed, with a confidence level of 99%, by the three procedures described through this chapter. Results are summarized on Figure 4.11.

The best procedure appears to be the Smearing-FDR method. Obviously, the

smearing leads to a significant enhancement in the detection power[3], compared to Unbinned Likelihood and point-like-FDR methods. It is interesting to note that the latter seems to be slightly better than the likelihood-based methods for tiny sources, and progressively loses efficiency when the luminosity increases. Finally the binned classical method [Ack06] and the fixed threshold method (section 4.2) are much less efficient.

[3]considering mislocations as a false discoveries would increase the FDR over the requested value. Therefore a more stringent coefficient of correlation would be required, leading to a decrease of the power

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Bragg

5

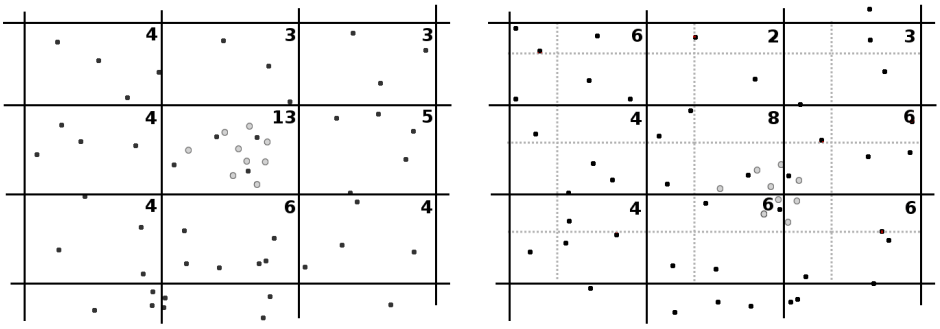
Local Grid FDR Procedure

This chapter details the successive modifications applied to the gridded FDR procedure to get free of several limitations, enumerated in the first section. The second part covers the substitution of the fixed partition by a local binning. The choice of a suitable observable to recover information about the energy of events is made in the third section. The two last sections are dedicated to the build up of p-values with the energy estimator. Finally, a summary of the procedure is proposed at the end of the chapter.

5.1 Drawbacks of the fixed binning

The signal from a source centered on a bin is mainly concentrated inside that single bin whilst a similar source located on the boundary can be spread over contiguous pixels, therefore diluting the accumulation of events indispensable to the detection of weak signal (Fig.5.1). In rarer cases, this situation can even lead to a misplaced source (see section 4.1 for definition and Table 4.1 in section 4.3.4 for results) decreasing the power of the method. This issue can partially be reduced by repeating the procedure with several shifted frames (see *i.e.* [Hau04]). The counterpart of this approach is that the probability has to be corrected with some *trial factor* and at the same time one loses one main advantage of the FDR procedure.

The use of a fixed partition scheme also introduces a lack of accuracy in the determination of the source position. Indeed, upon getting the final result of



(a) Favorable position **(b)** Unfavorable position

Figure 5.1: *Illustration of a favorable (a) and unfavorable (b) geometry for the detection of the source with identical binning and events disposition. Background and sources neutrinos are represented respectively in black and grey. The only difference is the grid relative position.*

the procedure, the user has no information about the real location of the events within the rejected cells. The source can be located anywhere in the bin. By default, the only choice is to identify the source position with the triggered bin center.Therefore the precision of the method is limited to the size of the bin, which is about 8° width, as it has seen in section 4.1.6.

Moreover the method's output does not allow one to know whether the registered excess is due to events stacked locally or scattered randomly in the cell. It has been seen in the results presented in the previous chapter, that this uncertainty induces an increase of the false detections in less favorable configurations. Looking at Figure 5.1-(b) for illustration, there is no difference between the upper left pixel and the bottom center one, both containing 6 events. However, the former gets its events randomly scattered in the bin whereas an accumulation in a very limited sector (upper right corner) can be seen in the latter.

In the following we look for an alternative way to determine the p-values without the need of a partitioning of the events.

5.2 Local Grid FDR procedure

5.2.1 Building the local bin framework

The main idea is to define a local circular search region around each neutrino. The radius of the search region depends on the resolution of the detector measured at the event declination (Fig.4.2). For the generality of the discussion,

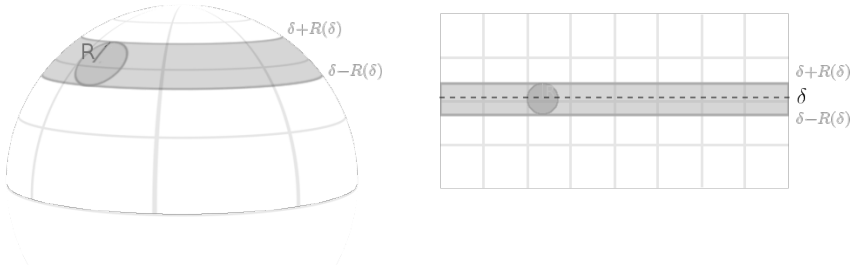


Figure 5.2: *Calculation of the expected background in the search region. For the clarity of the figure, the declination band and the search area have been oversized.*

one defines the radius as $R(\delta)$, depending only of the neutrino's declination δ. The question of the optimization of the search region will be discussed in section 5.5.2.

Individual search regions obviously solve the problem of the position of the neutrino in the bin. Since each neutrino is now in the center of its own region, their positions are all equivalent. The p-value will simply be determined by the local density of events within the search region. An important difference with respect to the previous approach is that a p-value is now associated to each neutrino in place of each sky bin. The question one previously tried to answer with the FDR rejection method was whether a given part of the sky contained a point source. Due to the individual determination of the p-value, the question becomes whether a given event is a source neutrino. With the fixed binning, the whole northern hemisphere was covered and associated with a probability. This statement is not valid with the present configuration.

5.2.2 p-value

The p-value is calculated in the same way as was done in the fixed binning approach : the observable is the number of events n_{obs} within the search region. The probability is calculated by comparing this number with the expected number of events n_{bg} in the region with respect to the declination distribution of $\mathcal{D}(\delta)$ of the total number of events N in the sample.

Hence for the neutrino ν_i located at a given declination δ, the circular search region of angular radius $R(\delta)$ covers a area (Fig.5.2) of :

$$2\pi(1 - \cos R(\delta)) \tag{5.1}$$

while the area of the declination band containing the search region is :

$$4\pi \ \cos\delta \ \sin R(\delta) \tag{5.2}$$

The number of expected background event is :

$$n_{bg}^{i} = N \frac{1-\cos R(\delta)}{2\cos\delta \ \sin R(\delta)} \int_{\delta - R(\delta)}^{\delta + R(\delta)} \mathcal{D}(\theta) d\theta \tag{5.3}$$

The p-value is equivalently defined as Eq.4.2. With this technique a p-value is associated with each event. The set of p-values is then submitted to the FDR rejection method presented in section 3.3.

5.2.3 Interpretation of the rejected sample

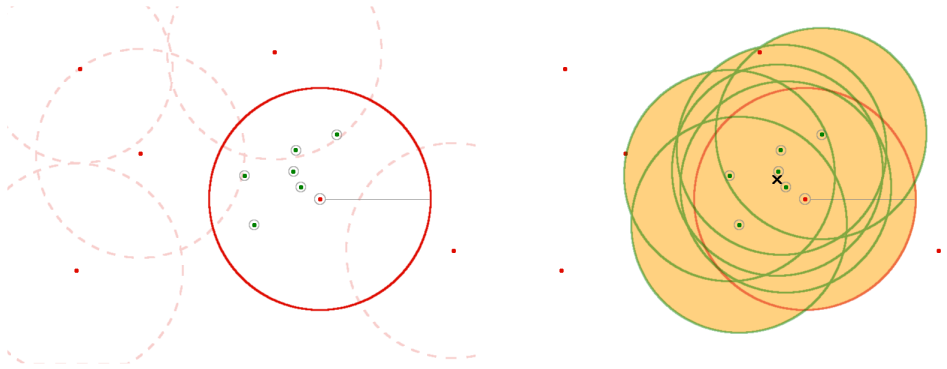
Following the original interpretation of the method, the events rejected are expected to be source neutrinos with a confidence level of 1-α. The natural estimate of the false discovery rate $\widehat{\alpha} = \widehat{FDR}$ is the ratio of background neutrinos rejected over the total number of rejections. Nevertheless, this statement needs to be modified.

At this point, it is important to keep in mind that a point source will show up as an excess of events on top of an expected background. Therefore it is most probable to find background neutrinos among source events, and it is *a priori* impossible to distinguish between the two types on based of their position. It implies that, in the case of a real source being identified, part of the events rejected by the method can nevertheless be *false rejections* (Fig.5.3-(a)). However these rejections are actually the consequence of the presence of source events in the search bin of rejected background events. Such rejections would not have occurred otherwise. This kind of issue is similar to the shift in the p-value ordering induced by the source's cell, discussed in Figure.4.6. However, in this case the p-values are not only shifted but many are also strongly modified by the presence of source events[1].

Most of the true detections will then be accompanied by one or more false discoveries. These *irreducible false rejections* unfairly increase the false discovery rate and accounting for them would dramatically penalize the power of the method.

Moreover it is worth noticing that these events are of necessity since they contribute to the local excess, thus to the detection.

[1] though such effects exist in the fixed binned approach, they are much weaker.



(a) bins **(b)** cluster

Figure 5.3: *Illustration of the local grid rejection procedure (a) and clustering (b).*
(a) Search regions are shown for background events (red) only. Regions are not represented for source events (green) for the sake of clarity. Rejected events are encircled. The highlighted background event is an irreducible false rejection *since it is blended in the source events.*
(b) Only triggered events (circled) are used for the clustering. If their regions overlap (which is the case on this example), they belong to the same cluster (shaded). The cluster center (cross) is defined as the center of gravity of the cluster (see text for details)

In conclusion, the estimation of the FDR has to be re-thought in order to accurately take into account and control the real effects of background fluctuations only.

5.2.4 Clustering

As decribed, up to now the method can lead to multiple counting of signal events as the regions attached to each neutrino can overlap. It is worth to remember here that the goal of the method is to localize cosmic sources of neutrinos rather than to individually draw out cosmic neutrinos out of background events. A practical way to overcome the problems of irreducible false discoveries and multiple counting is to gather the selected events from a common cause under one single identification. Moreover, it naturally permits one to extend the method to multiple source detection. The most suitable scheme is a hierarchical minimal distance clustering [JMF99]. One nevertheless has to determine when two rejected events belong to the same cluster, hence considered to have the same cause. Since neutrino position has an uncertainty related to the detector resolution, the criteria will be that an event belongs to a cluster if its distance relative to the closest event of the cluster is smaller than the sum of the resolution of the two considered events. From a geometrical point of view, and

in the particular scheme of bin sizes equal to the resolution, this condition is equivalent to group the overlapping regions (Fig.5.3).

From a technical point of view, the clustering algorithm can be summarized like this :

- Step 0: each neutrino of the rejected set N_r belongs to a different cluster C_i.
  As long as the number of clusters at step $n+2$ is smaller than at step n:
- Step $n+1$: Compute distances between two clusters, that it the minimal distance between two events belonging to two different clusters :

$$d_{ij} = \{\text{Min}(\text{Distance}(\nu_a, \nu_b)) \quad | \quad \nu_a \in C_i \,, \nu_b \in C_j\}$$

- Step $n+2$: if the distance between two clusters is lower than the sum of angular resolutions on the direction of the the corresponding neutrinos $R(a)$ and $R(b)$, the two clusters they belong to will be merged

$$\text{if} \;\; d_{ij} < R(a) + R(b) \quad \longrightarrow \quad C_i = C_j = C_i \cup C_j$$

The centroid of the cluster accounts for the source location. Unlike fixed bin method, this definition has the advantage to account for the (rejected) events repartition. The precision of the reconstructed position is enhanced and does not depend on the search methodology. A calculation variant, by weighting events with their p-value, has been tested with similar results.

So at the end of the clustering, the original set of neutrinos events has been replaced by a set of clusters. Each cluster is considered as a possible source. This implies that the control of the FDR is required on clusters. This is similar to the first method, where the FDR control was actually requested on cells.

5.2.5 Results

The efficiency of detection of sources located at 42.5° with a luminosity varying between 0 and 20 neutrinos over a background of 1000 events has been determined. The background is roughly the equivalent of one year of data-taking with the AMANDA detector. A total of 500 realizations have been simulated by using the declination distribution of events from 2000-2004 AMANDA-II data, allowing the check of the stability of the FDR at a level of 1% . Results are summarized on Figure 5.8 (p. 92).

Though being more efficient than the point-like binned method, the unbinned approach shows itself less sensitive than the enhanced smearing procedure.

N	δ_s	$n_{bg}(\delta_s)$	DP 50%	DP 90%
1000	42.5°	0.68	3.5 10^{-10}	4.6 10^{-10}
4282	30.0°	2.91	1.2 10^{-10}	$>$ 1.5 10^{-10}

Table 5.1: *Detectable fluxes (in TeV cm^{-2}s^{-1}) with the unbinned approach using p-values determined from the number of events in the search region. N gives the total number of backgroud events in a sky. δ_s is the declination of the simulated source. $n_{bg}(\delta_s)$ gives the expected number of events in the search region defined around the source. The two last columns give the fluxes needed for the source to be detected with 50% and 90% chance, respectively.*

Hence, almost 10 cosmic neutrinos are requested to have 50% chance to identify the source over an average background of $n_{bg} = 0.68$ events. One needs 13 events in order to reach 90% chance.

The False Discovery Rate is nevertheless thoroughly controlled at the requested level of 1%. More, it seems that the rate is even too conservative since FDR estimators oscillate around an average value of 0.0022 $\pm$ 0.0005. It could possibly mean that the coefficient of correlation (fixed at the previously determined value of 6.0) is overestimated and therefore can be tuned down somewhat. By adjusting it, one can expect to increase the discovery potential of the method.

A similar analysis has been performed under the same simulation conditions for background skies composed of 4282 atmospherics events corresponding to the 5 years of data taking with AMANDA-II between 2000 and 2004, and a source at 30°. Results are shown on Figure 5.9 (p. 92). Of course the discovery potential sensibly decreases for the same number of neutrinos, with respect to the 1 year results, as the average background is now 2.91 events. This result must be expressed in terms of fluxes in order to have a compatible comparison (Table.5.1). About 16 sources neutrinos are needed to detect the excess with a probability of 50%, but it corresponds to a flux 3 times fainter.

However, the stability of the FDR request is not as good as it was for the shorter lifetime. One can observe a slight excess of false discoveries at higher luminosity. But the small statistics at which the analysis has been performed do not exclude the possibility for these excesses to be due to statistical fluctuations, since the FDR estimator $\widehat{\alpha}$ suffers large statistical uncertainties. The average FDR on the whole simulated skies is $\widehat{\alpha} = 0.0075 \pm 0.0008$, which confirms the general control since a maximum rate of $\alpha = 0.01$ is tolerated. A global increase of the FDR estimator is nevertheless noticed with the number of background and source events. This point will be explained further in section 5.4.4.

5.3 The energy information

Up to now the point source detection based on the FDR method only takes into account the angular position of the neutrino on the sky map by looking for an excess of events over the local expected background.

Since energy spectra of cosmic neutrinos are thought to be generically harder than atmospheric ones, the energy of neutrinos is another piece of information that can be used to distinguish between atmospheric and cosmic origin. Cosmic neutrino energies typically are supposed to have a spectral index of -2 while atmospheric secondaries follow a measured index of -3.7.

This property can be exploited to increase the detection efficiency. Since most energetic neutrinos will be expected to be preferentially of cosmic origin, increasing the contribution of energetic neutrinos will facilitate their detection.

Nevertheless, introducing the energy information within the procedure encounters two obstacles. First, the energy of an event is poorly reconstructed in the detector (*cf.* Fig.6.4). Second, one must introduce this new information into probabilities, without disrupting the FDR control nor decreasing the detection efficiency.

5.3.1 Energy estimator

The most robust and simple variable available to evaluate the neutrino energy is the number of hits N_{ch}, defined as the number of optical modules triggered by the muon passing through the detector. The purpose is not to estimate precisely the energy but to use a variable linked to the energy and which p.d.f. under signal and background hypotheses are significantly different (Fig. 5.4).

5.4 Combining position and energy p-values

This section presents a first attempt in inserting the energy information into the FDR rejection procedure to enhance the distinction between atmospheric and cosmic neutrinos. The p-value related to the position information p_p is calculated in the way described previously. The energy distribution $\mathcal{E}$ (see Chapter 6) is used to deduce the p-value related to the energy of the neutrino :

$$p_e(E) = \int_E^\infty \mathcal{E}(x)dx \tag{5.4}$$

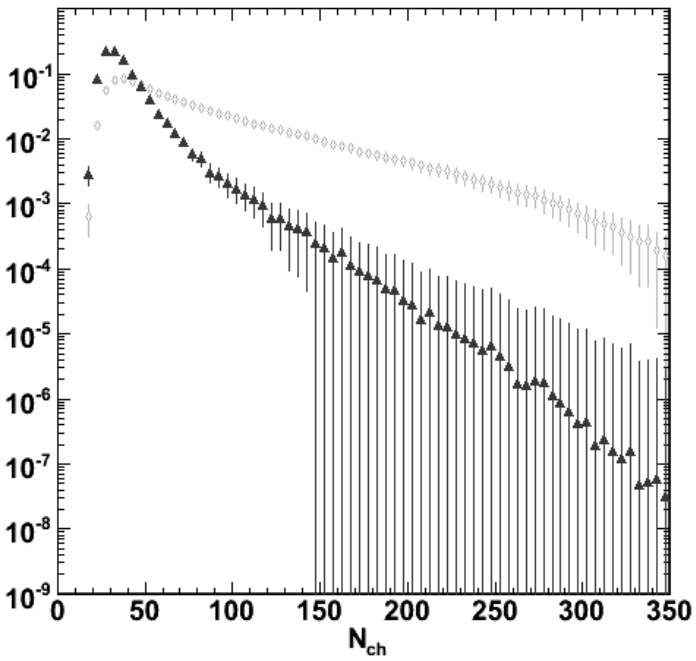


Figure 5.4: *Atmospheric (▲) and E^{-2} cosmic (♢) spectra from simulation*

5.4.1 Minimal p-value

The simplest way to combine both p-values p_e and p_p for each neutrino is to take the smallest one as general p-value and apply the FDR algorithm on the sample.

However, the FDR is not controlled due to the definition of the "p-value". Energy and angular position are supposed to be independent. Therefore the two p-values p_e and p_p are uncorrelated and $\vec{p}$ is uniformly distributed in the two-dimensional space $[0,1] \times [0,1]$. However $p \equiv \mathrm{Min}(p_e, p_p)$ does not fulfill the condition of uniform distribution ($\mathcal{P}(p) = 1$) but $\mathcal{P}(p) = 2p$ for $p \in [0,1]$.

The control of the FDR (with CL 1-α) under Null Hypothesis H_0^p has been proven for p_p in section 5.2. Identically, the control of the FDR is guaranteed under Null Hypothesis H_0^e from the definition of p_e. At first approximation, the Null hypothesis H_0 on the remaining p-value is to take the less favorable condition for each event and then reject the event if either H_0^e or H_0^p is rejected. Should the two rejection areas do not overlap, which is not the case in the

practice, the confidence level is decreased to 1 -2α. Hence, such an approach fails in controlling the FDR and another procedure has to be established.

5.4.2 Multivariate p-value approach

The FDR procedure has originaly been developed for multiple testing on one single variable. A theoretical extension, presented in section 3.3.3, has been developed to generalize the method to multivariate p-values.

This requires the construction of a new *p-value* s, a combination of the two p-values p_e and p_p, respectively, from the energy and the angular position information.

Following notations introduced in section 3.3.3, K=2 and $\vec{p_i} = (p_{ei}, p_{pi})$. The aim is to determine the best suitable function $J(\vec{p})$.

The rejection region associated with threshold value u is defined as :

$$\Gamma_u = \left\{ \vec{p} \in [0,1]^2 : p_p^{\eta} p_e^{2-\eta} \leq u \right\} \tag{5.5}$$

with η the weight accounting for the importance of the density in the rejection process [2] . Then one obtains :

$$J(\vec{p}) = \begin{cases} s(1 + \ln s^{-1}) & \text{if } \eta = 1 \\ \frac{\eta}{2(\eta-1)} s^{1/\eta} + \frac{2-\eta}{2(1-\eta)} s^{1/(2-\eta)} & \text{otherwise} \end{cases} \tag{5.6}$$

with $s = p_p^{\eta} p_e^{2-\eta}$

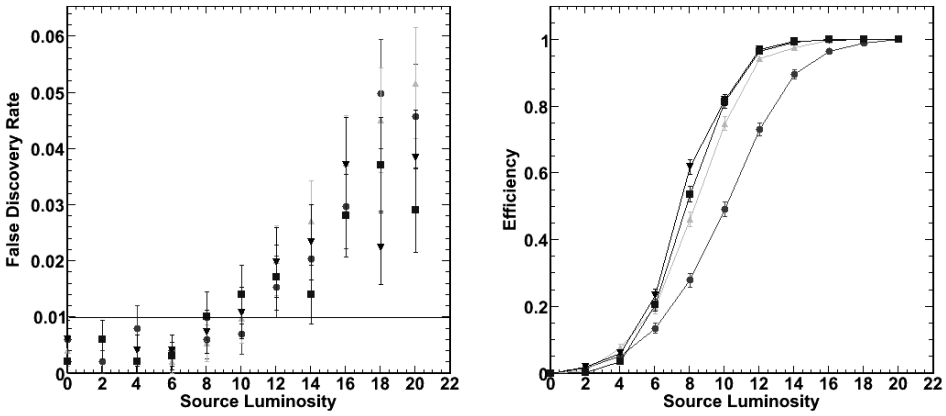
Preliminary results

Analyses have been done for several value of $\eta \in]0, 2[$. A sample of 500 skies of 1000 background neutrinos have been generated. A simulated cosmic source has been added at 22.5° of declination [3] on top of the background. Results are shown in Figure.5.5.

Even if the detection power is best with $\eta = 1$, no value of η ensures the stability of the False Discovery Rate as soon as the source luminosity becomes large: typically, higher than 8 or 10 cosmic neutrinos over an expected local background of 0.68 events. This preliminary result raises a couple of remarks. First, the energy contributes to a better efficiency for the detection of faint sources. Moreover, the fact that $\eta \geq 1$ induces that *in fine* the energy has more

[2] counter-intuitively , the higher the exponent, the lower the resulting impact since $p \leq 1$

[3] in order to compare results with likelihood method.



(a) False Discovery Rate **(b)** Efficiency

Figure 5.5: *FDR (a) and efficiency (b) of the multivariate approach for weigth on position p-value $\eta = 0.6$ (•), 0.8 (▲), 1 (▼) and 1.2 (■).*

importance than the local density. Nevertheless it also thwarts the stability of the FDR at higher luminosities when sources are more visible and where the use of the FDR method is less compelling.

5.4.3 Multivariate with adaptative weighting

A way to circumvent the problem at high luminosity would be to adapt the coefficient η as a function of the source luminosity, but such information is not *a priori* known. However one can use the information given by the FDR procedure described in section 5.2 as a first guess about the presence of a source. The bigger the cluster, the brighter the hypothetical source. A tiny source or a fluctuation of background would induce no or few rejected events and the energy would therefore enable to discriminate between source and background. On the other hand, a bright source would have already triggered the original method and the energy is here of little use.

Investigations have been done in order to determine the relation between the applied weight $\eta(n)$ and the preliminary cluster size n. Eventually, a sigmoid (Fig.5.6) gives a good detection efficiency while guaranteeing FDR control :

$$\eta(n) = \lambda_0 + \frac{\lambda_\infty - \lambda_0}{1 + e^{(n_t - n)}} \tag{5.7}$$

A basic interpretation of the parameters can be made by looking at the size of real and false source clusters obtained with the FDR procedure applied on posi-

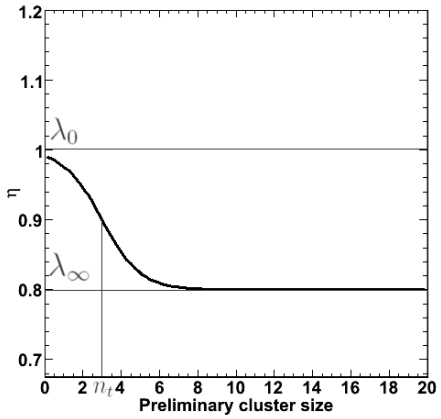


Figure 5.6: *Weight applied on the position p-value p_p as a function of the preliminary cluster size n. A weight 2-η is applied on p_e.*

tional p-values, and compare it when applied on the energy and the multivariate technique. It can be seen in Figure.5.7 that the size of false clusters seldom exceeds 3 neutrinos, and the contribution of energy (2nd row) and position (1st row) are comparable. False detections do not depend on the source luminosity, neither with the number of events nor the energy. However, it can happen that a exceptionally high number of events gives from time to time a false cluster whose size can reach 6 or 7 events.

Talking about true discoveries, the procedure applied to the angular position is clearly much more efficient than on the energy. Moreover, the latter shows that the maximal size of the cluster with the energy information is 2. It means that finding more than 2 ultra energetic neutrinos within a very localized region[4] is very improbable for a typical detector lifetime of 1 year. On the other hand, the number of events only does not allow any true discovery for a source weaker than 6 events. Beyond this limit, position cluster size can reach the source luminosity. Therefore, the energy information is important in enhancing the detection of tiny sources, that is small clusters. Its impact wanes with bigger clusters, and must then be reduces to avoid an increase of the false discovery rate.

This justifies that, below a cluster size of $n_t = 3$ events, position and energy have to be equally taken into account and equivalently weighted ($\lambda \sim \lambda_0 = 1$). After, the impact of the energy must be reduced, and its corresponding

[4] Technically, within the detector resolution

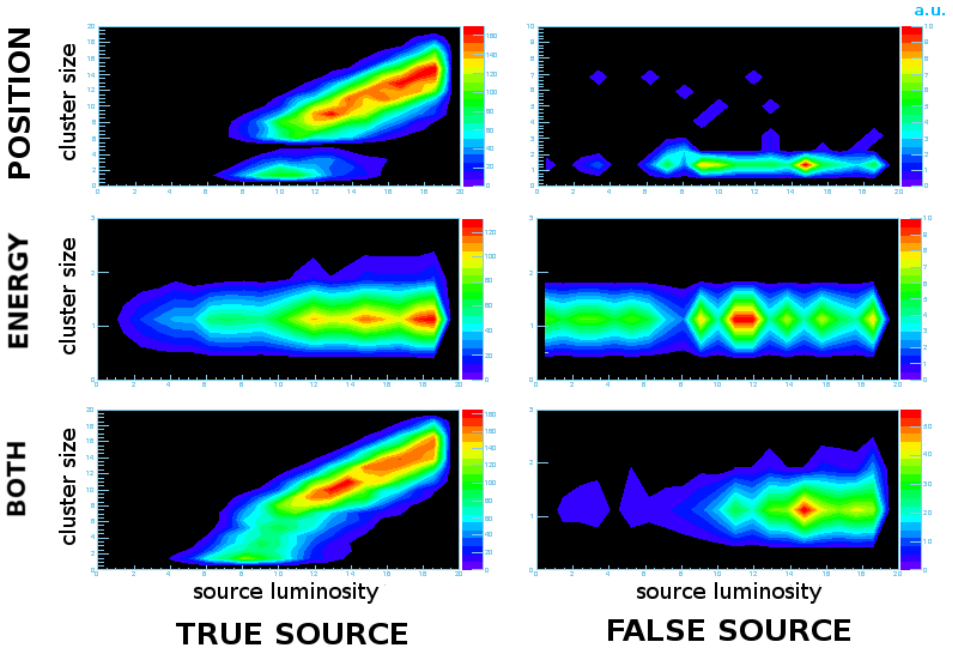


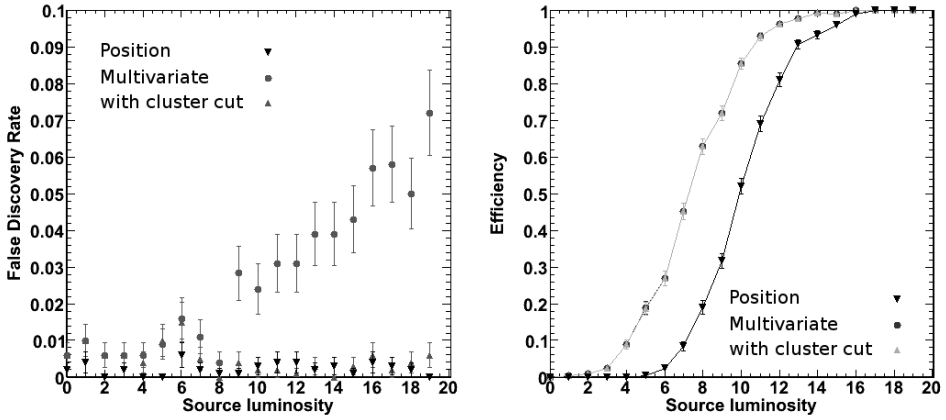
Figure 5.7: *Size of true and false clusters as a function of the source luminosity following the rejection method using p-values deduced respectively from the angular position (1st row), the energy (2nd row) and the multivariate approach using the sigmoid with parameters* $\lambda_0 = 1$ *,* $\lambda_\infty = 0.8$ *and* $n_t = 3$ *.*

exponent $(2-\eta)$ is then bigger than 1. Conversely, the weight for the angular position is decreasing with the luminosity to quickly tend to $\eta \to \lambda_\infty = 0.8$.

Eventually, combining both effects gives a better efficiency at low luminosities thanks to the energy, while the false clusters size barely exceeds 3 events.

5.4.4 Final cut on single event clusters

FDR is still not stable for source luminosities above 10 cosmic neutrinos over 0.68 expected atmospheric events (1 year), though the probability to have such a bright source is very weak. Anyway should such event occur will it be quite obvious, and the method developed here would not be really needed to point it out. Nevertheless, one desires and expects the rejection procedure to be throughly conservative. So, a cut will be applied at the end of the procedure, in order to recover the control of false discoveries while not to penalize the efficiency of the method. Although the adaptive weight does not fix the issue encountered when using a fixed weight, it has been seen from the results



(a) False Discovery Rate **(b)** Efficiency

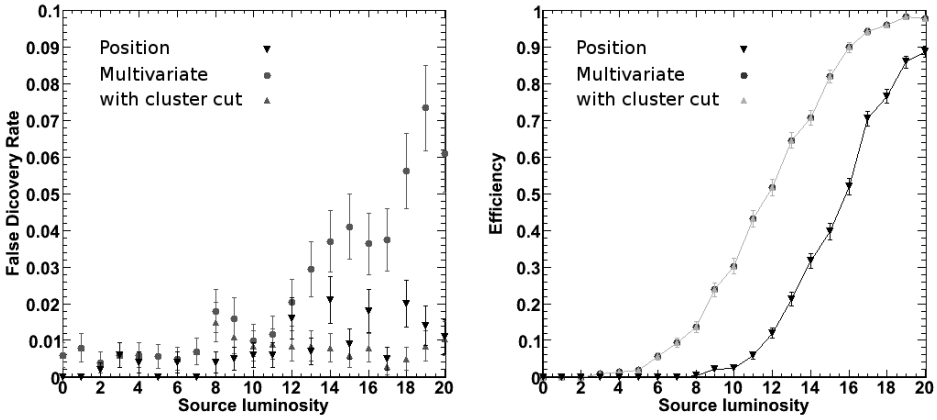
Figure 5.8: *FDR (a) and Efficiency (b) of the unbinned approaches as a function of the luminosity (over one year) of a source located at 42.5°. 500 realizations of 1000 background events (∼ 1 year of data-taking) have been simulated. Results obtained with the information on angular position only, the multivariate approach without and with the cut on single clusters.*

discussed that most of the falsely detected clusters are single event clusters. Therefore, it seems natural that a powerful cut will be not to take into account such clusters. Nevertheless, it would also severly reduce the power of the procedure for the lowest luminosities, where single true clusters occur preferentially. Then, a compromise must be found and one eventually sets the final condition of getting rid of single clusters if and only if a bigger cluster occurs within the same sky.

5.4.5 Comparison between angular position and multivariate approaches

Analyzes have been performed to estimate the contribution of the energy compared to the angular position criteria. The same simulated event samples have been used to draw out the performances of the method using the angular position (discussed in section 5.2.5) and the multivariate position + energy approach.

Results are summarized in Figures 5.8 and 5.9. A clear enhancement is noticed when the energy is taken into account. However, as discussed in section 5.4.4, the FDR diverges at high luminosity. The cut on single clusters is necessary to recover the desired FDR. Looking at Fig.5.8-(a), it is seen that the cut starts being useful for source luminosities over 6 neutrinos when the average expected



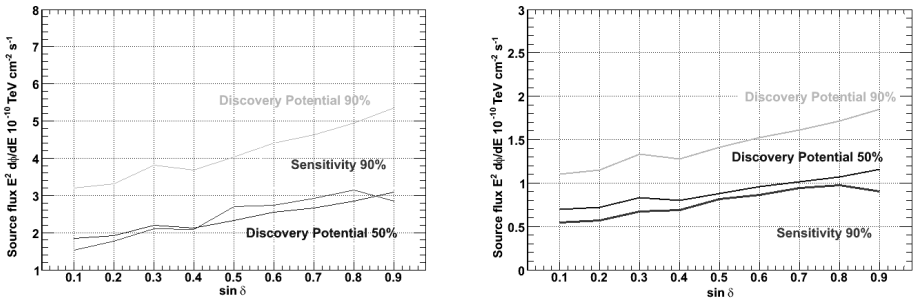
(a) False Discovery Rate **(b)** Efficiency

Figure 5.9: *FDR (a) and Efficiency (b) of the unbinned approaches as a function of the luminosity (over 5 years) of a source located at 30°. 500 skies of 4282 background events (equivalent to 2000-2004 data taking) have been simulated. Results obtained with the angular position only, the multivariate approach without and with the cut on single clusters.*

background is 2.91 events (for a total of 1000 background events in the sky). It can be understood by looking at the true cluster size with respect to the source luminosity (bottom-left in Fig.5.7). A single neutrino *cluster* is removed if a multi-events cluster, most probably a real source cluster, occurs in the same sky. Such clusters start appearing from a source luminosity of 6 neutrinos.

Another effect on false discoveries is the sudden break in the rate observed around a luminosity of 6-8 (Fig.5.8-(a)) and 8-10 (Fig.5.9-(a)) neutrinos for an equivalence of respectively 1 year and 4 years of data taking. It correspond to a transition where the use of the energy starts loosing importance in the detection of real sources, since excesses are progressively becoming big enough to be detected by using the angular position only. On the other hand, in this region, energy is nevertheless useful to discriminate between true and false detections and to help get rid of wrong rejections. Of course, for higher luminosities, the cut on single clusters is needed because of the increasing obviousness of the detection, inferring numerous unusually low p-values, leading to similar shift effects on p-values already discussed and illustrated in Figure.4.6.

It is also important to emphasize that the power of the method is not affected by the cut on clusters. This point is explained by the fact that a true detection would be deleted only if one neutrino had been detected and a false detection with at least 2 clustered events had simultaneously occurred. The combination



(a) 1000 background events **(b)** 4282 background events

Figure 5.10: *Discovery potentials at 50% and 90% with 99% Confidence Level for the multivariate method applied over 1000 (a) and 4282 (b) background events. Also shown here the sensitivity of the method from Feldman-Cousins limits at 90% Confidence Level.*

of these two criteria is very improbable.

The sensitivity and discovery potential of the multivariate method are shown in Figure.5.10.

5.5 Univariate p-value approach

Using energy in addition to the angular position information enhances the discovery potential. Unfortunately, the method developed step by step in the previous section requires the introduction of parameters which need to be fine tuned beforehand to allow the method to work properly. This complexity constrasts with the general simplicity of the original method. As it has been seen, the main complications comes from the multivariate character of the rejection criteria.

The aim of this section is to develop an efficient rejection procedure where the energy and the angular position are naturally merged into one single p-value.

5.5.1 p-value

The simplest way to combine angular position and energy (estimated by the number of hits N_{ch}, which is an integer) into a unique variable is to look at the total energy collected within a search region. This new observable reflects the local energy accumulation around an event. The corresponding p.d.f. can be theoretically determined rather easily.

If $\mathcal{E}_0$ is the p.d.f. of the energy distribution of an atmospheric neutrino, it would

also be the p.d.f. of the total energy collected in a search region containing no other neutrino than the central one. If a neutrino has only one other neutrino belonging to the central neutrino's search region then the total energy distribution will be :

$$\mathcal{E}_1(x) = \sum_{a=0}^{x} \mathcal{E}_0(a)\, \mathcal{E}_0(x-a)$$

By recurrence, the total accumulated energy within a search area containing n extra neutrinos is distributed as :

$$\mathcal{E}_n(x) = \sum_{a=0}^{x} \mathcal{E}_0(a)\, \mathcal{E}_{n-1}(x-a)$$

Finally, the probability to get n neutrinos within a search region where a background of $n_{bg}(\delta)$ neutrinos is expected follows the Poisson distribution $\mathcal{P}(n; n_{bg}(\delta))$. The total energy distribution, i.e. the Null Hypothesis, is:

$$\mathcal{E}_{TOT}(N_{ch}; \delta) = \sum_{n=0}^{\infty} \mathcal{P}(n; n_{bg}(\delta))\, \mathcal{E}_n(N_{ch}) \tag{5.8}$$

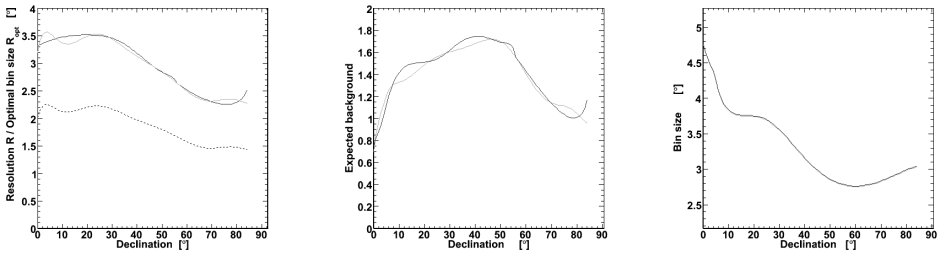
leading to the usual definition of the p-value of an event located at declination δ, with total energy N_{TOT} in its search region :

$$p(N_{TOT}, \delta) = \sum_{N=N_{TOT}}^{\infty} \mathcal{E}_{TOT}(N) \tag{5.9}$$

Though taking the energy density literally is the most obvious manner to gather in one variable the energy and the position of events, it nevertheless brings a major difference in the way to interpret the subsequent p-value. The contribution of the central event to the p-value is totally equivalent to the contribution of any other event in the search region. Unlike in the multivariate approach, where the energy of the central event only was taken into account, all the energies counted in the search region will identically participate to the estimation of the p-value.

5.5.2 Discussion about definition of the search region's size

An important matter to be discussed is the definition and the optimization of the search region's size. It has been said in section 5.2 that the width only depends on the declination of the event defining its center.



(a) **(b)** **(c)**

Figure 5.11: *The original 3-step procedure for the calculation of the search region. (a) Minimal size R_{75} to cover 75% of the source (black), and the optimal resolution $R_{opt} = 1.585\ R$ (grey). The dashed line accounts for the resolution R. (b) The background expected n_{bg} in the search region with size defined in (a). (c) The search region ensuring $\mu_{\max}$ for R_{75} (black) and R_{opt} (grey) are identical.*

Constant background and maximization of the source coverage

The first construction of the search region is based on the particular requirement that the expected background is a constant for all events. It means that the whole set of events will follow the same p.d.f. since the distribution only depends on the expected background (Eq.5.8). The calculation of the p.d.f. is time and CPU consuming because of the convolution of $\mathcal{E}_n$ with the probability to have n events in the search region.

Asking for a common expected background speeds up the p-value calculation. The value of the expected background is determined beforehand by asking that the area covers at least 75% of the events of a hypothetical source for all declinations (Fig.5.11). It is assumed that the source events are spread around the true position following a two-dimensional gaussian with a variance equal to the mean value of the detector resolution. The first step of the expected background calculation is to calculate the size $R_{75}(\delta)$ of the search region as a function of the declination. Knowing the declination distribution of events, it is straightforward to determine the equivalent background expected in such region :

$$n_{bg} = \frac{1 - \cos R_{75}(\delta)}{2 \cos\delta \ \sin R_{75}(\delta)} \int_{\delta - R_{75}(\delta)}^{\delta + R_{75}(\delta)} \mathcal{D}(\theta) d\theta \tag{5.10}$$

The final expected background is the maximal value of this function :

$$\mu_{\max} = \max_{\delta} n_{bg}(\delta) \tag{5.11}$$

Eventually, the calculation of the required size to cover such background is

performed for every declination.

$$S(\delta) = \arccos\left(1 - \frac{\mu_{\max}}{n_{bg}(\delta)}(1 - \cos R_{75}(\delta))\right) \tag{5.12}$$

Optimal search region

It appears from complementary studies that the preliminary radius R_{75} is equivalent to the optimal resolution $R_{opt} = 1.585 \times R$, R_{opt} corresponding to the theoretical width minimizing $\sqrt{bg}/\text{signal}$ [BvC02]. This direct connection between the hypothetical source coverage and the resolution gives two main advantages. First the definition of the search region was the last parametrization of the method which required a hypothesis on the source model. To substitute it by the resolution of the detector makes the method completely model independent. Second the determination of the size directly from the resolution simplifies and speeds up considerably the procedure. The procedure is therefore optimized by discarding the condition of constant expected background (Eq.5.11) and subsequently the adjustment of the search region (Eq.5.12). the radius of the search area will then directly be defined by R_{opt}.

5.5.3 Remarks on multiple sources detection

The method has been checked in the more realistic case of multiple sources. The purpose of this section is to see whether the presence of a source would infer the discovery of another one, or the control of the confidence level.

The analysis has been performed with a stronger requested FDR $\alpha = 0.005$, on simulated skies containing 2 sources: A first source with a variable luminosity between 0 and 20 events is generated at different declinations (see Table 5.2) over a total background of 1000 atmospheric-like events. Furthermore, each sky contains a second source, referenced here as the *fixed* source, which consists in 6 cosmic events spread around the source location at a declination of 22.0° and a right ascension of 83.6°.

Figure.5.12 summarizes the FDR stability (a) and the detection efficiency (b) of both sources, the first one being located at 22.5° of declination. A corrected detection power is presented. It takes into account the fact that the number of event observed n follows the Poisson distribution of mean strength $\bar{\nu}_{src}$ related to the source flux. The detection power is therefore computed by summing the efficiency $DP_{ev}(n)$ of detecting n neutrinos weighted by the Poisson probability

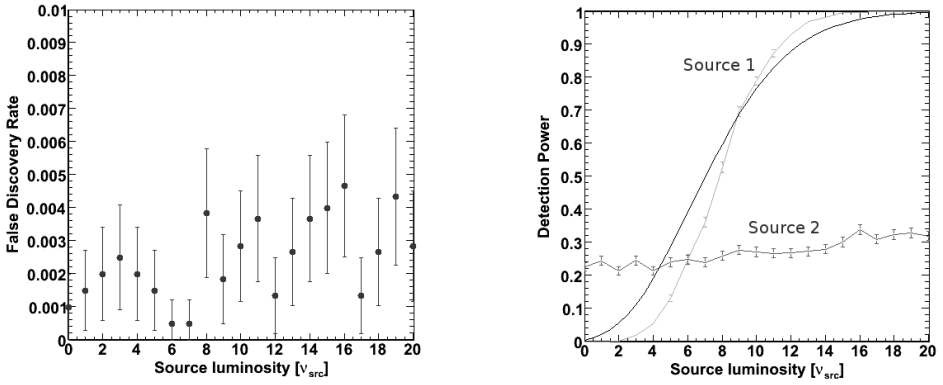


Figure 5.12: *Multisource results for source nr.1 at (180° ; 22.5°), with luminosity between 0 and 20 neutrinos, and source nr.2 at (80° ; 22.0°) with a fixed luminosity of 6 cosmic events, on top of a background of 1000 atmospheric events.*
(a) False Discovery Rate is fulfilled with a CL99.5%
(b) Discovery potential of source nr.1 ($DP_{ev}(n)$ in gray, and $DP_{src}(\bar{\nu}_{src})$ in black) and source nr.2 as a function of the luminosity of the first source.

$\mathcal{P}(n|\bar{\nu}_{src})$:

$$DP_{src}(\bar{\nu}_{src}) = \sum_{n=0}^{+\infty} \mathcal{P}(n|\bar{\nu}_{src}) \times DP_{ev}(n). \tag{5.13}$$

It can be seen from results presented in Table 5.2 that the presence of a (not very bright) source does not significantly influence the discovery potential of another source. Moreover, Fig.5.13 shows that the relative position of sources have no effect on their discovery.

However, the discovery of a second weak source slightly increases with the luminosity of the first one. A simultaneous increase of the FDR can be observed, though the effect is weak and does not disrupt the stability of the estimator. This increase is due to the combination of very small p-values and a cumulative effect, finally leading to a behavior similar to the problem faced in the gridded method and explained in section 4.1.6.

The explanation can be done in two steps. First by considering the events around the bright source; high density of events not only infer smaller p-values but also a higher quantity of such small p-values. Hence the shift effect observed (cf.Fig.4.6) in the ordering of p-values is accentuated since many (sources) events are pushing other p-values to the right. As an illustration, the largest background fluctuation or a tinier source event, which would have been indexed by $i = 1$ in the case of no bright events, is shifted to position – let's say –

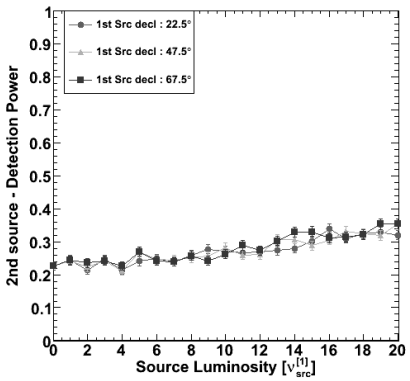


Figure 5.13: *Discovery potential of the source nr.2 as a function of the luminosity and declinations of the source nr.1. It is clear that the efficiency does not depends on the position of the other source.*

$i \sim 20$. Thus it implies a tolerated threshold[5] equivalently multiplied by 20. As a consequence, the probability for such an event to be rejected is slightly increased. However a difference exists between background and faint source event. The former will mostly be an isolated event, and then will be thrown away by the cut on single clusters without affecting the false discovery rate, while the latter could have several other close-by events rejected altogether and therefore be part of a multi-event cluster surviving the cluster cut, thus increasing the discovery rate of the source.

[5] given by $i \times \frac{\alpha}{N}$, see eq.3.17

Variable source declination	Flux DP 50%		Flux DP 90%	
	w/o fixed src.	with fixed src	w/o fixed src.	with fixed src.
17.5	1.99	1.99	3.54	3.55
22.5	2.05	2.04	3.64	3.63
27.5	2.23	2.24	3.96	3.97
37.5	2.58	2.58	4.58	4.56
42.5	2.61	2.62	4.62	4.64
47.5	2.69	2.67	4.62	4.78
57.5	2.82	2.78	5.00	5.01
67.5	3.20	3.14	5.71	5.63

Table 5.2: *Detectable fluxes (units: 10^{-10} TeV cm^{-2} s^{-1}) to observe a source with 50% and 90% chance, with and without the fixed source.*

5.5.4 Final procedure

Figure 5.14 provides a complete review of the final FDR procedure applied to point source searches with the AMANDA-II detector. The procedure can be basically divided in 3 blocks thoroughly described in this thesis in dedicated sections referenced on the diagram. The first part concerns the construction of the Null Hypothesis, detailed in the Chapter 6. The second part concerns the optimal procedure to determine the p-value. The final part refers to the processing of non-background events in case of rejection of the Null Hypothesis by the FDR controlling method.

The analysis of the response of the method to simulated events from the Data Challenge (Chapter 7) has permitted us the fine tuning of the procedure to its final version. AMANDA-II data presented in Chapter 8 have been analyzed with this version.

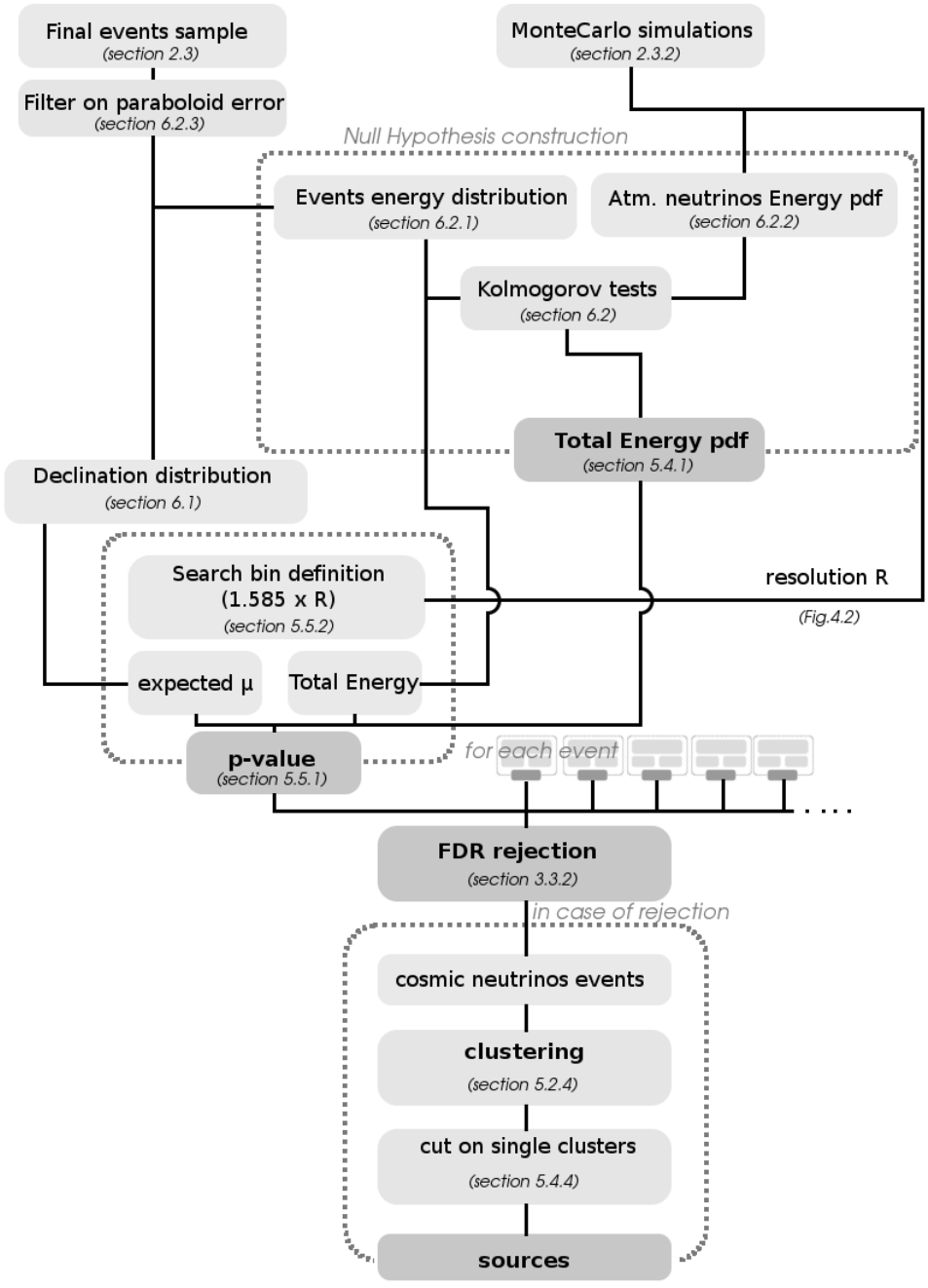


Figure 5.14: *Summary of the final False Discovery Rate procedure applied to point source search.*

When evaluating a model, at least two broad standards are relevant. One is whether the model is consistent with the data. The other is whether the model is consistent with the 'real world'.

Kenneth A.Bollen

6

Probability density functions

The sensitivity of a signal identification depends on the precise knowledge of the distribution of the corresponding background for the retaining parameters. This point is a crucial matter in the False Discovery Rate controlling procedure. Hence, unlike likelihood ratio-based methods where a hypothesis is made on both the background and the source (Eq.3.14), the expected behavior of the background is the only information one authorizes to introduce in the rejection process.

This chapter is therefore dedicated to the determination of the probability distribution functions for the 2 parameters (angular position and energy) used in this analysis. The first part concerns the establishment of the angular distribution of events. The second part covers the construction and the verification of the validity of the energy pdf.

6.1 Density pdf

It has been seen that the angular distribution of the background $\mathcal{D}(\delta)$ is only dependent of the declination and can be seen as uniform as a function of the right ascension for a given declination band.

This distribution is used to determine the expected number of background events in the bin (for the fixed binned approaches, Eq.4.2) or in the individual search region (for the unbinned procedures, Eq.5.3 and Eq.5.8).

A general estimation of the event distribution from the data would be sufficient.

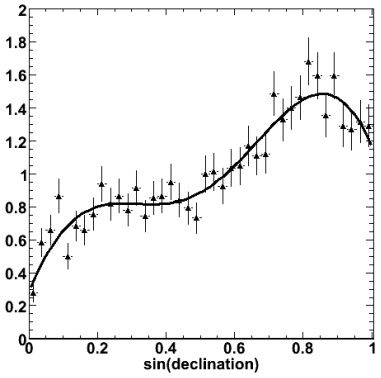


Figure 6.1: *Events distribution as a function of the declination $\mathcal{D}(\delta)$ from the AMANDA-II 2000-2006 data sample. A polynomial fit is performed to have a generic expression of the background behavior and facilitate simulations.*

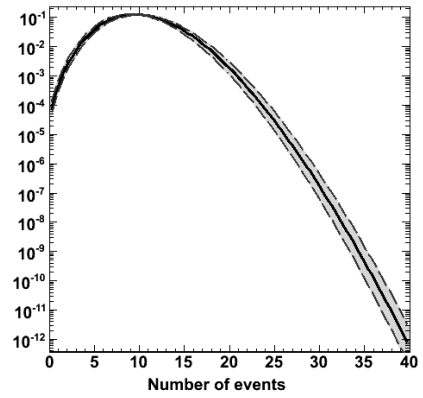


Figure 6.2: *Poisson distribution of mean $\mu = 10$. The dashed lines account for an error of 5% on the estimation of the mean, that is $\mu = 9.5$ and $\mu = 10.5$.*

Indeed the analytic expression of the pdf stating the density of events is well known since it follows a Poisson distribution. Only the Poisson mean μ needs to be determined vs. the declination *via* the events density distribution (for example the AMANDA-II 2000-2003 data shown in Figure 6.1). The distribution has been fit with a polynomial function to facilitate both the background estimation and simulations of large numbers of equivalent realizations without the introduction of systematics due to repetitive declinations. The same procedure is repeated for ther other data samples (AMANDA 97-99, AMANDA-II 2000-2006) used in this work.

The mean does not need to be known with an extended precision. An illustration is given in Figure 6.2, where the Poisson distribution for a typical mean ($\mu = 10.0$) is shown and sided with the equivalent distributions with an uncertainty of 5% on the mean value. Hence, such uncertainty on the mean gives an uncertainty on the p-value for 10 and 25 observed events :

$$p(10) = \sum_{n=10}^{\infty} \mathcal{P}(n|10.0 \pm 0.5) = 0.54^{+0.06}_{-0.06}$$

$$p(25) = \sum_{n=25}^{\infty} \mathcal{P}(n|10.0 \pm 0.5) = 4.69^{+5.24}_{-2.61}\ 10^{-5}$$

A 5% uncertainty barely influences the rejection of the event.

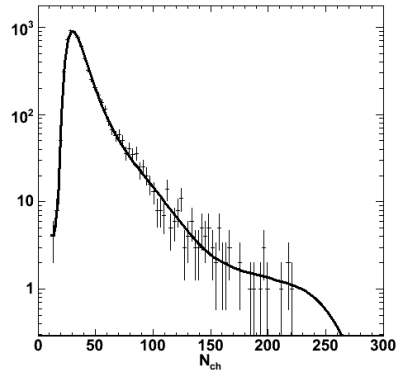


Figure 6.3: *Energy distribution (N_{ch}) from the AMANDA-II 2000-2003 data. The fitted distribution $\mathcal{E}_0$ (here non-normalized) is shown in plain.*

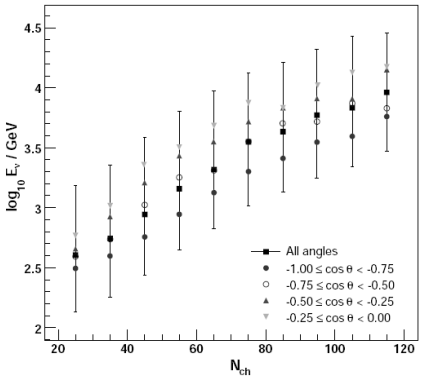


Figure 6.4: *Neutrino energy vs. number of hits channels. Black squares account for the average energy on all zenith-angles below the horizon. [Aa09]*

6.2 Energy pdf

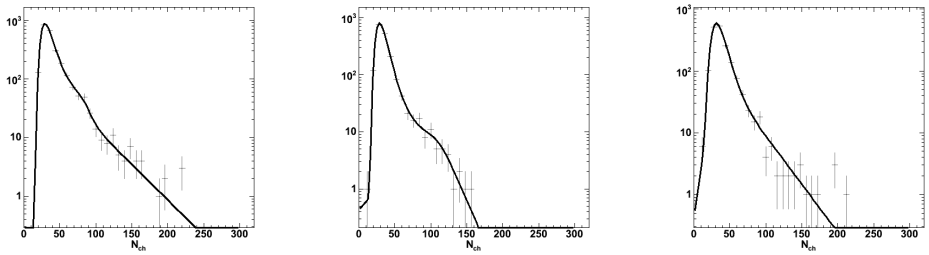
The energy is estimated by looking at the number of hits (N_{ch}) recorded among the DOMs during the travel of the muon through the detector. In the following, this observable will be used as an estimator of the energy. Their correlation can be seen in Figure.6.4.

6.2.1 PDF from data

The distribution of the energy results of the convolution of the muon energy with the detector response. As for the angular distribution, it could be determined from the experimental data (Fig.6.3).

Declination dependency

The geometry of AMANDA-II detector implies variations in the energy distribution of events as a function of the declination. An event at high declination passes through the detector along its longest extension. The length of the track in the instrumented volume is important and a high number of optical modules can then be reached by Cherenkov photons. When the declination is decreasing, the path of the particle in the detector is progressively shortened, and a lower number of hits are expected, leading to an energy underestimation. However,



(a) $\delta < 32.5^\circ$ **(b)** $32.5^\circ \leq \delta \leq 52.5^\circ$ **(c)** $\delta > 52.5^\circ$

Figure 6.5: *Energy distributions (N_{ch}) from data sample, for low (a), mid (b) and high (c) declinations. The fitted distributions E_0 (non-normalized) are shown in solid curve.*

this effect is counterbalanced by the contamination of prompt muons. Muon events at low declination are expected to be more energetic, yielding an increase in the observed number of hits. This residual muon contribution will be confirmed later in section 6.2.3.

The energy distribution is therefore sub-divided in three declination bands containing approximatively the same number of events.

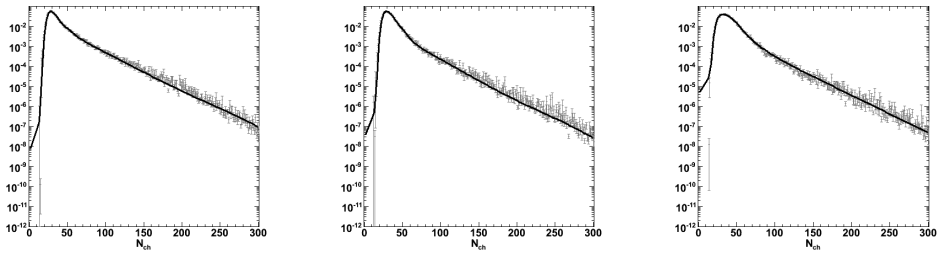
- low declinations : $\delta < 32.5^\circ$
- mid declinations : $32.5^\circ \leq \delta \leq 52.5^\circ$
- high declinations : $\delta > 52.5^\circ$

Figure 6.5 shows that non-negligible discrepencies are qctually to be seen in the energy distribution as a function of the declination.

Uncertainties on the distribution

Due to a lack of statistics, the high energy part ($N_{ch} > 120$, *i.e.*$E_\nu > 10$ TeV) of the spectrum is poorly known, as it can be seen in Figures 6.3 and 6.5. Tiny variations in the adjustment of the curve fitting at high energy induce changes on several orders of magnitude on the p-value of the less probable events. However, the discrimination power of the method depends on the determination of the high energy region of the spectrum. Therefore a precise description of the behavior at high energy is therefore mandatory to ensure the reliability of the procedure.

Moreover it has to be kept in mind that data may contain a couple of non-atmospheric neutrino events, which are precisely the ones we are looking for.



(a) $\delta < 32.5^\circ$ **(b)** $32.5^\circ \leq \delta \leq 52.5^\circ$ **(c)** $\delta > 52.5^\circ$

Figure 6.6: *Energy Distributions (N_{ch}) from MonteCarlo, for low (a), mid (b) and high (c) declinations. The fitted distributions E_0 (non-normalized) are shown in plain.*

Unlike for the density distribution, where cosmic events are in any case overwhelmed by true background events at all declinations, a single contribution at very high energy will lead to an overestimation of the background and results in a strong decrease of the efficiency of the method. The main consequence is to inhibit the rejection of this very event.

These arguments justify the desire to dissociate as much as possible the estimation of the distribution from the data sample on which the analysis is done.

6.2.2 PDF from simulations

A way to solve this problem is to draw the energy pdf out of MonteCarlo simulations (Fig.6.6) of the detector response to atmospheric events, which can describe the energy behavior at the highest energies with enough statistics and without being biased by possible non-atmospheric neutrino events. However it is necessary to ensure beforehand that the MonteCarlo describe correctly the expected background events. This is the purpose of Kolmogorov tests described in the following.

Kolmogorov tests

The compatibility of the MonteCarlo with the data for the three declination bands was performed by mean of Kolmogorov tests. The results of the tests are shown on Figure 6.7 and are summarized in Table 6.1.

Only the mid-declination band passed the test condition $\alpha = 0.01$. Moreover the discrepancy between data and MonteCarlo at low declinations is particularly important. MonteCarlo overestimates small N_{ch} compared to data. However this phenomenon tends to reverse at higher declinations. The shape of the cdf's

seems nevertheless to be similar.

The MonteCarlo has been adjusted in order to minimize the observed shift effect by applying a shift on N_{ch} value. A forward shift of one unit for low declinations and a backward shift of two units for high declinations are the best adjustment. No shift has been applied on mid-declination since it already satisfied the test. Kolmogorov tests have been performed again with the corrected MonteCarlo distributions. Results are shown in Figure 6.8 and still summarized in Table 6.1. An enhancement in the agreement can be seen, and the resulting MonteCarlo distribution now matches the real events distribution at high declination. Even reduced, there is still an incompatibility at low declinations.

6.2.3 Filtering on muon events

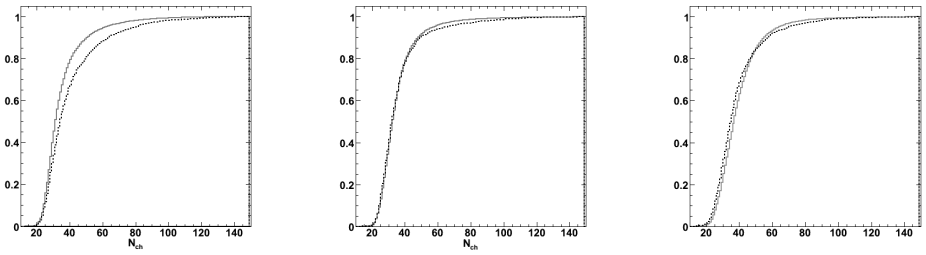
The reason of the disagreement between data and MonteCarlo noticed in the previous section manifests itself by an excess in the range between 60 and 120 N_{ch}, that is events with energy between 1 and 10 TeV. It has been shown by J.Kelley [Kel] that the excess results from misreconstructed muons. He also demonstrated that it could be eliminated by cuts on paraboloid error (σ_{par}) as a function of N_{ch}: events with

- $\sigma_{par} > 5.033 - 0.036 N_{ch}$ if $N_{ch} < 80$
- $\sigma_{par} > 2.1$ otherwise

are eliminated. The outcome of the cut is shown in Figure 6.9. The histograms and dots account for the data sample respectively before and after filtering. It can be seen that the cut mainly affects the low declination band where the excess of events with N_{ch} between 60 and 120 completely disappear. This brings the data sample back to better agreement with the MonteCarlo distributions (solid black line). It is worth noticing that, unlike for highest declinations, the filtering effect on mid-declination events is non-negligible.

Kolmogorov tests on filtered data

The same procedures were applied on filtered data, resulting in an improvement for low and high declinations while the mid declination distribution remains stable. Figures 6.10 and 6.11 illustrate the test respectively for non shifted and shifted procedures. Although statistical compatibility is not obtained for low declination, one can see that the agreement between the MonteCarlo and the data distribution was largely improved. MonteCarlo will nevertheless be used in



(a) $\delta < 32.5^\circ$ **(b)** $32.5^\circ \leq \delta \leq 52.5^\circ$ **(c)** $\delta > 52.5^\circ$

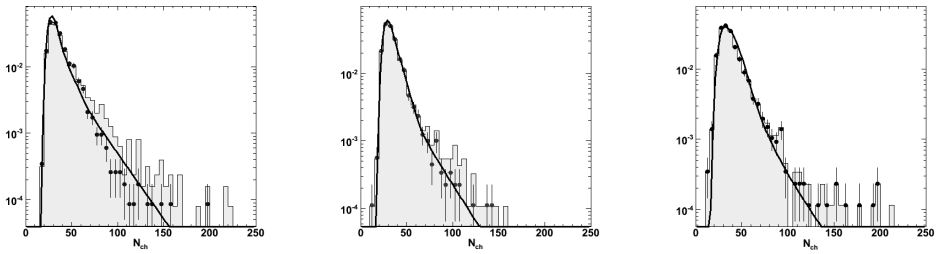
Figure 6.7: *Kolmogorov test on non filtered data sample (color plain) and non-shifted MonteCarlo (black dashed), for low (a), mid (b) and high (c) declinations.*

(a) $\delta < 32.5^\circ$ **(b)** $\delta > 52.5^\circ$

Figure 6.8: *Kolmogorov test on unfiltered data sample (color solid) and shifted Monte-Carlo (black dashed), for low (a) and high (b) declinations. Mid declination is equivalent to 6.7-(b)*

			w/o readjustment		with readjustment	
Decl.band	events	Limit 99%	D_{max}	Concl.	D_{max}	Concl.
$\delta < 32.5^\circ$	2507	0.033	0.130	NO	0.105	NO
$32.5^\circ \leq \delta \leq 52.5^\circ$	1827	0.038	0.027	YES	unneeded	
$\delta > 52.5^\circ$	1742	0.039	0.081	NO	0.033	YES

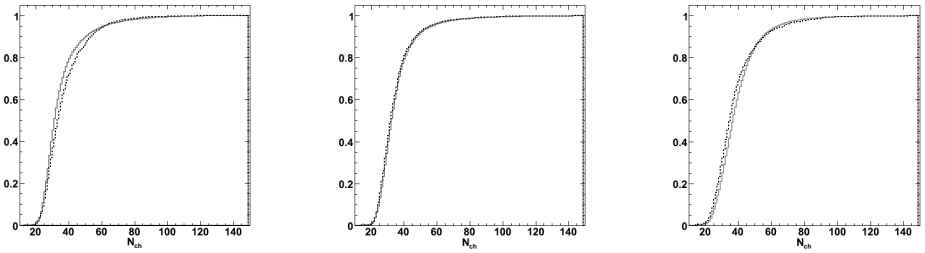
Table 6.1: *Kolmogorov tests outcomes with non-filtered data. D_{max} accounts for the maximal distance between data and MonteCarlo cdf's.*



(a) $\delta < 32.5^\circ$ **(b)** $32.5^\circ \leq \delta \leq 52.5^\circ$ **(c)** $\delta > 52.5^\circ$

Figure 6.9: *Energy Distributions (N_{ch}) from data sample before (histo) and after (points with error bars) filtering , for low (a), mid (b) and high (c) declinations. Distributions from MonteCarlo simulations are shown in solid line.*

the analysis as the remaining discrepancy appears at rather low values of N_{ch} (< 60) corresponding to energies below 1 TeV.



(a) $\delta < 32.5^\circ$ **(b)** $32.5^\circ \leq \delta \leq 52.5^\circ$ **(c)** $\delta > 52.5^\circ$

Figure 6.10: *Kolmogorov test on filtered data sample (color plain) and non-shifted MonteCarlo (black dashed), for low (a), mid (b) and high (c) declinations.*

(a) $\delta < 32.5^\circ$ **(b)** $\delta > 52.5^\circ$

Figure 6.11: *Kolmogorov test on filtered data sample (color plain) and shifted Monte-Carlo (black dashed), for low (a) and high (b) declinations. Mid declination is equivalent to 6.10-(b)*

			w/o readjustment		with readjustment	
Decl.band	events	Limit 99%	D_{max}	Concl.	D_{max}	Concl.
$\delta < 32.5^\circ$	2330	0.034	0.096	NO	0.053	NO
$32.5^\circ \leq \delta \leq 52.5^\circ$	1781	0.039	0.037	YES	-	-
$\delta > 52.5^\circ$	1728	0.039	0.084	NO	0.026	YES

Table 6.2: *Kolmogorov tests outcomes with filtered data*

There's no challenge more challenging than the challenge to improve yourself.

Michael F. Staley

7 Data Challenge

This chapter describes the verification of the performances of the FDR procedure before being applied on real data. The final check is done, for the first time in the collaboration, with a large set of simulated skies, so-called *Data Challenge*. In order to simulate as much as possible the real conditions, the existence and the characteristics of the possible search are not communicated to the team performing the analysis.

The chapter starts with a description of the Data Challenge rules. Then several analyses will be performed following the successive distributions, discussed in Chapter 6, of the energy background. These analyzes will validate the rejection mechanism and give the detection potential and the sensitivity of the FDR procedure.

7.1 Rules

The Data Challenge consists in a *blind* analysis of 17442 sky samples in order to check the reliability and the efficiency of the False Discovery Rate procedure. Each sky is composed by a set of events from the AMANDA 2000-2006 data sample used as background on top of which a simulated source has possibly been added. Neither the angular position nor the luminosity of the simulated source is known, if such a source exists. The information available is :

- the angular position (right ascension and declination)

- the energy (in terms of N_{ch})
- the paraboloid error σ_{par} from the reconstruction of the muon track (see section 2.3.1)

of each event.

The procedure, summarized on Fig.5.14, is applied on each sky. The Null Hypothesis is a combination of the angular position and the energy of the events. All the 17443 skies are statistically equivalent, with the exception of the negligible addition of maximum 20 events (over a background of more than 6000 events). Therefore the angular position distribution will always be the same (Fig.6.1). Since possibly added source events will preferentially contribute to the high N_{ch} region, where very few background events are present, the energy spectrum could differ from an analysis to another.

For each sky, the results of the metho are, *a posteriori*, compared to the simulated imput. A source is considered to be well identified if its reconstructed angular position matches the simulated one within 5°. The Confidence Level was set to 99%.

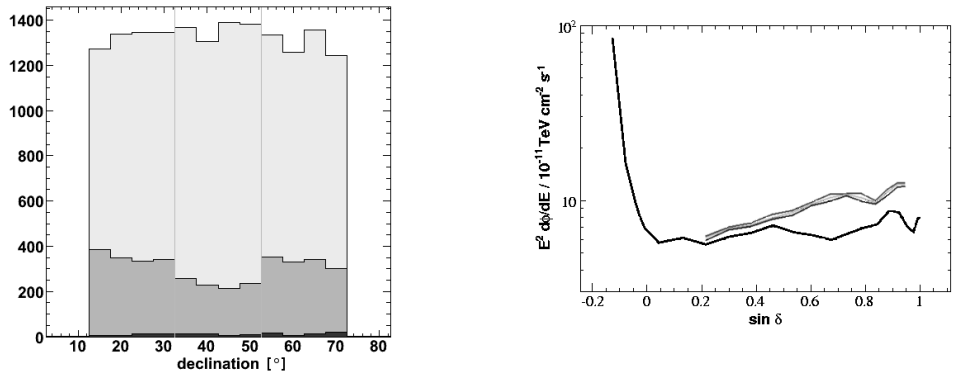
To ensure the total blindness of the data challenge, the sources position and luminosity were unknown at the time of the analysis. Sources have been generated with a luminosity ranging between 0 (pure background) and 20 neutrinos by step of 2, and are uniformly distributed between 12.5° and 72.5° of declination. These ranges were unknown during the verification of the procedure.

7.2 Null Hypothesis determination from the experimental data

7.2.1 Single hypothesis on the energy distribution

The FDR analysis based on angular position convoluted with the energy measurements was first applied using the N_{ch} distribution drawn out from the experimental data (Fig.6.3), as described in section 6.2.1. The results are summarized in Figure.7.1.

In order to quantify the importance of the requested false discovery rate on the efficiency of the procedure, the analysis has been repeated with three FDR of $\alpha = 0.003$, $\alpha = 0.006$ and $\alpha = 0.01$. The FDR is controlled in all cases. Figure 7.1-(b) illustrates fluxes needed to reach a discovery potential of 50%. More restrictive FDR conditions naturally give a faint decrease of the efficiency. It can also be seen on this figure the negative effect of the unique pdf on required



(a) **(b)**

Figure 7.1: *Results of the analysis performed on Data Challenge sample with the data distribution of the energy as the energy pdf.*
(a) number of true (dark grey) and false (black) detections vs. source true declination. The total number of generated sources is shown in light grey for a requested FDR of 1%.
(b) Flux detectable with a discovery potential of 50% for a FDR of 0.3% (upper grey curve), 0.6% (middle grey curve) and 1% (lower grey curve). The black curve is the likelihood method efficiency with 99%CL.

flux from mid declination sources ($\sin\delta$ between 0.5 and 0.8), compared with the results from the Likelihood method [Bra].

Fig.7.1-(a) represents the number of true and false detections in function of the declination of the source. These numbers are expected to be independant of the source declination. However it is clear that the present method suffers a decrease of efficiency for detection located between 30° and 50° of declination. This is explained by looking at Figure 6.5, representing the energy distribution for events located at declination regions below 32.5° (a), above 52.5° (c) and between these two declinations (b), referred in the following as low, high and mid declination bands.

It can be seen that unlike for low and high declination bands, the N_{ch} distribution for the mid declinations bands has no high energy tail and obviously differ from the global N_{ch} distribution (Fig.6.3) used for the determination of the p-value for all events. The energy does not exceed $\sim$ 150 N_{ch} for mid declinations bands, compared to other bands events where the distribution extends up to 220 N_{ch}. The N_{ch} estimator leads to an energy determination systematically lower in the mid declination band. This variation is not taken into account in the present analysis since the same average pdf is used to deduce p-values whatever the event location. The background contribution in the mid declination band is

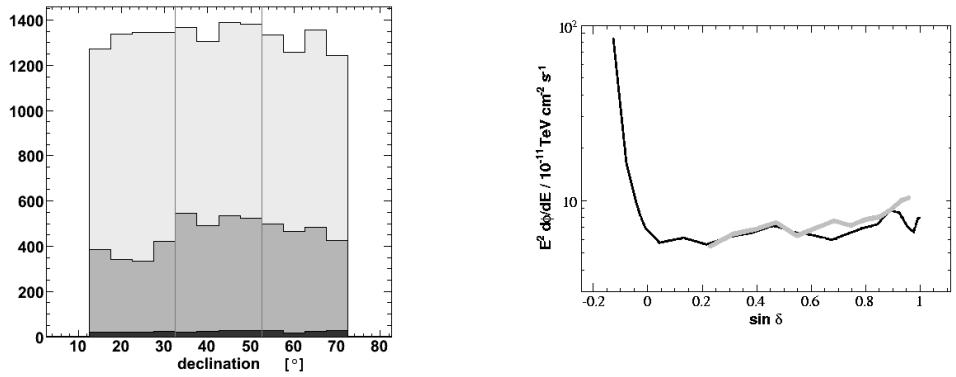


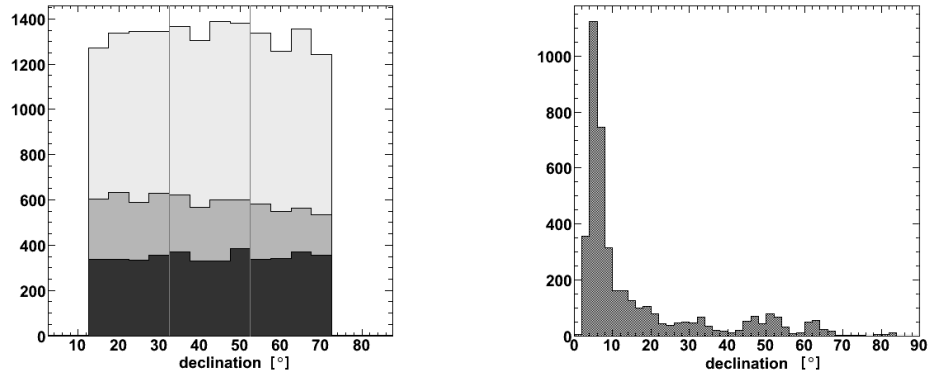
Figure 7.2: *Results of the analysis performed on Data Challenge sample with the data distributions (for low, mid and high declinations) of the energy as the energy pdf.*
(a) number of true (dark grey) and false (black) detections vs. source true declination. The total number of generated sources is shown in light grey. The requested FDR is 1%.
(b) Flux detectable with a discovery potential of 50% for a FDR of 1% (grey). The likelihood method efficiency with 99%CL is shown in black.

then overestimated, leading to the observed loss of efficiency.

In conclusion this first analysis definitely confirms the ability of the method to control the confidence level of a detection. Hence the efficiency is far to be optimal : a unique, averaged pdf is not the best assumption given the dependence of the energy estimation with the declination.

7.2.2 Declination dependency

The procedure has been repeated using three distributions reflecting the energy behavior at low, mid and high declinations (Fig.6.5) instead of the general energy distribution (Fig.6.3). The false discovery rate obtained on the data sample is $\widehat{\alpha} = 1.01\% \pm 0.07$ which is compatible with the requested 99%CL. A general enhancement of the discovery potential of the method is globally observed (Fig.7.2). The efficiency is improved in the mid-declination band, thanks to the better adjustment of the background energy distribution. An improvement is also observed at high declination while the low declination band efficiency is stable compared to the previous results. Nevertheless the efficiency for this band is now below the discovery rate of the other bands. This might be explained by harder N_{ch} distribution at low declination leading to an overestimation of the neutrino energy for background events at low declination (Fig.6.5-(a)). Though



(a) (b)

Figure 7.3: *Results of the analysis performed on Data Challenge sample with the energy distribution drawn out from MonteCarlo simulations.*
(a) number of true (dark grey) and false (black) detections vs. source true declination. The total number of generated sources is shown in light grey for a requested FDR of 1%.
(b) Declination of false discoveries. An large unexpected excess is observed at very low declinations, yielding the non-control of the procedure.

the method gives an efficiency similar with the likelihood method, a systematic effect is still observed linked to the poor estimation of the high part of the energy spectrum. The results comparison in this section proves once again the sensitivity of the method on the right description of the background behavior.

7.3 PDF from simulations

A particularly interesting illustration of the rejection mechanism of the method is observed when analyzing the results obtained with a null hypothesis distributions build on the basis of simulation data.

In section 6.2.2 it is shown that the MonteCarlo does not match the data for the low declination band. This discrepancy has a dramatic impact on the control of the false discovery rate, as it can be seen in Fig.7.3-(a). Indeed the FDR is measured at 14%, although the error rate was requested to 1%. Moreover, the observation of a false rejections is independent of the true source position. However, 85 % of falses discoveries is observed at low declination, with a peak around $\delta = 8\ ^\circ$ (Fig.7.3-(b)). This confirms the fact that false discoveries are due to the lack of knowledge of the energy distribution at low declinations.

The disagreement could result from unknown background events contributing

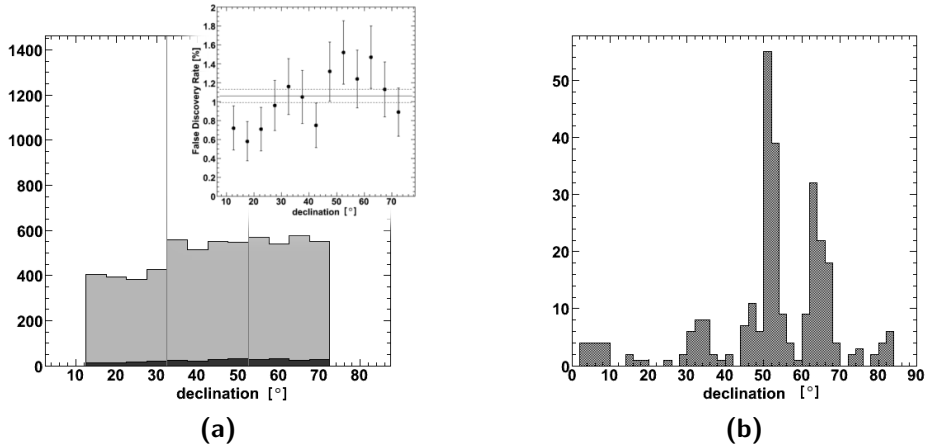


Figure 7.4: *Results of the analysis performed on Data Challenge sample with the energy distribution drawn out from MonteCarlo simulations after filtering.*
(a) number of true (grey) and false (black) detections vs. source true declination. The insert shows the false discovery rate as a function of the true source declination.
(b) Declination of the observed false discoveries.

mainly at low declination, or from source events in the data sample.

This background was identified as badly reconstructed atmospheric prompt muons. These events must then be removed beforehand to obtain the desired compatibility between data and simulations, by applying an additional cut on the paraboloid error, as described in section 6.2.3.

From the method point of view, the events tagged as false discoveries are not actual false rejections. The method actually perfectly plays its role since it permits to identify an additional background contribution which was originally not taken into account in the energy distribution.

7.4 Data Challenge results

The analysis performed up to now allowed to identify a new background source. The final results presented in this section were obtained after the removal of this background and keeping the null hypothesis dependance on declination as described in section 6.2.3.

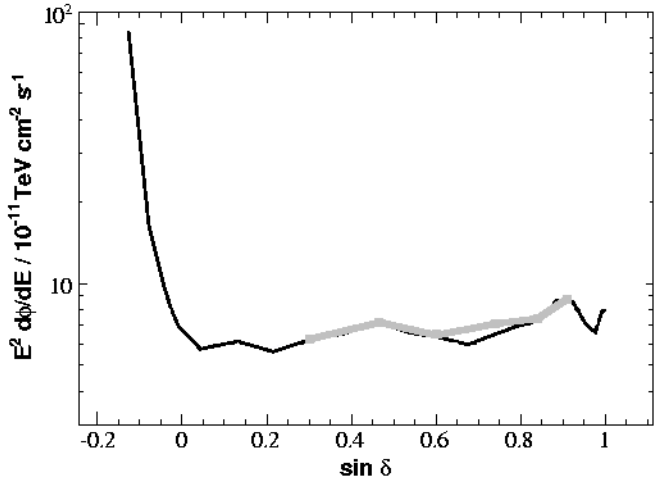


Figure 7.5: *Discovery Potential 50% and FDR 1% for the FDR controlling procedure (grey).The flux is compared with corresponding LLH discovery potential with 99%CL (black).*

7.4.1 Results of the analysis

The results are presented in Fig.7.4. The control of the false discovery rate is presently recover within statistical errors. With a confidence level of 99%, it is obtained a rate of $\widehat{\alpha} = 1.06 \pm 0.07$ on the 17443 samples. The observed fluctuations (insert in Fig.7.4-(a)) could be attributed to the difference of the background description in the three declination bands.

Comparing Fig.7.3-(b) and Fig.7.4-(b) accounting for the position of the false discoveries respectively with and without the preliminary filtering on data, helps explain the bahavior of the FDR in function of the declination.The striking feature is the disappearance of the bunch of falsely rejected events at low declinations.

This means that although it is not perfect at the sense of Kolmogorov test (cf.section 6.2.3, the MonteCarlo distributions are now close enough to data to allow a reliable description of the background.

It can be even seen in Figure 6.11-(a) that the MonteCarlo overestimates the background contribution for low declination band, resulting in a lack of discoveries (Fig.7.4-(a)) with respect to the discovery rate observed in the other

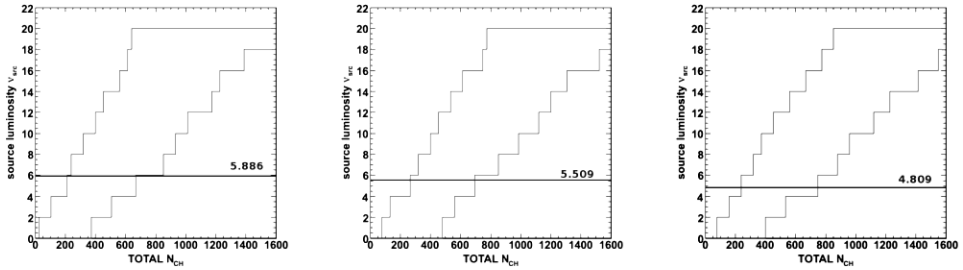


Figure 7.6: *Feldman-Cousins limits and sensitivity as a function of the total energy collected in a search bin, for low (a), mid (b) and high (c) declination band.*

declinations.

Coming back to Figure 7.4-(b), a bump and two peaks can be seen respectively around 32°, 52° and 65°. The bump around 32° is most probably due to edge effects between the mid and low declination bands. The energy distribution changes abruptly around these two regions and the p-value of an event depends on which side of the border it falls in.

Another factor can explain the two main peaks observed at 52° (on top of a possible edge effect between mid and high declination bands) and 65°. A slightly higher number, namely 2 or 3, of high energy events than expected occurs around these declinations, in the real data sample used to create the background samples for this analysis: all background samples have identical declinations and energy while right ascensions are scrambled. There is nevertheless a probability that part of these higher energy events from time to time gather by chance. This case induces lower p-values for these events and then increase their probability to be rejected by the method. However these variations ($\sim$ 50 occurrences on $\sim$ 17000 samples) are much weaker compared to the peak previously observed at 8° in case of non-filtered data ($>$ 1000 occurrences). They therefore do not thread the control of the FDR.

7.4.2 Performances

Global Discovery Potential and Sensitivity

Figure 7.5 shows the discovery potential at 50% and with a FDR of 1% as a function of the declination for point sources of ν_μ generated assuming an energy spectrum of E^{-2}. The comparison is made with the corresponding discovery potential obtained with the likelihood method at a confidence level of

99%. Performances of both methods are compatible. It has however to be stressed that the likelihood method is based on an hypothesis on the signal energy spectrum which exactly match the generated energy spectrum of the sources whilst the FDR method is completely independant of this hypothesis. Furthermore, it has been seen in section 5.5.3 that the FDR method is efficient in the simultaneous detection of multiple sources. This characteristic does not appear in this analysis since Data Challenge sample only contains one source per sky.

In Figure 7.6 can be found the Feldman-Cousins limit intervals and the sensitivity (see section 3.4) of the FDR method for the three different declination bands considered in the analysis. Hence, the total energy measured in the event region gives an estimation of the contribution of a possibly unseen source. The average sensitivities on a band at low, mid and high declination are respectively of 2.1 10^{-11}, 3.0 10^{-11} and 3.1 10^{-11} TeV cm^{-2} s^{-1}.

Around a source

Figure 7.7 gives the accuracy of the reconstructed source angular position in case of detection. The position of the source is defined as the center of gravity of the rejected events constituting the cluster. The precision of the source localization naturally increases with the number of detected events. It depends also of the true declination of the source (Fig.7.8), since the accuracy is strongly correlated with the detector resolution. However, the choice of a common energy distribution dedicated to each declination band tends to make the precision uniform inside a declination band. Hence, the three-part structure in the pointing accuracy is clearly visible.

The global source pointing accuracy is shown in Figure 7.7 without any distinction about luminosity or declination. It appears from Figure 7.7-(a) that a average bias of 0.43° is observed between the reconstructed declination of the source and its true declination.

It can be seen in Figure.7.7-(b) that 95% of the identified sources are reconstructed at an angular distance less than 2° around the true position. However some extreme case show slightly poorer precision up to 5°. Sources hardly leads to a detectable manifestation beyond this limit. Hence, in the Data Challenge, clusters detected further than 5° of the generate sources have been counted as false rejections.

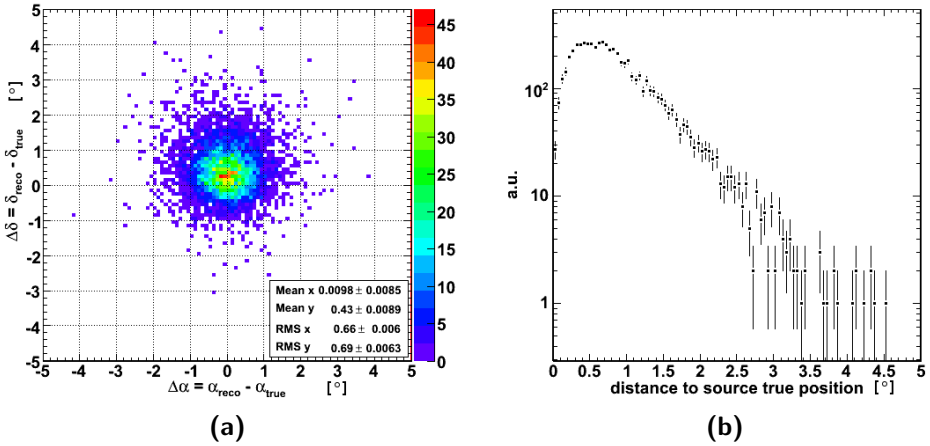


Figure 7.7: *The Source Pointing Accuracy (SPA) is defined as the distance between the reconstructed position (COG of the cluster) and the source true position.*
(a) Error between the reconstructed position and the true source position.
(b) Angular distance to the source true position.

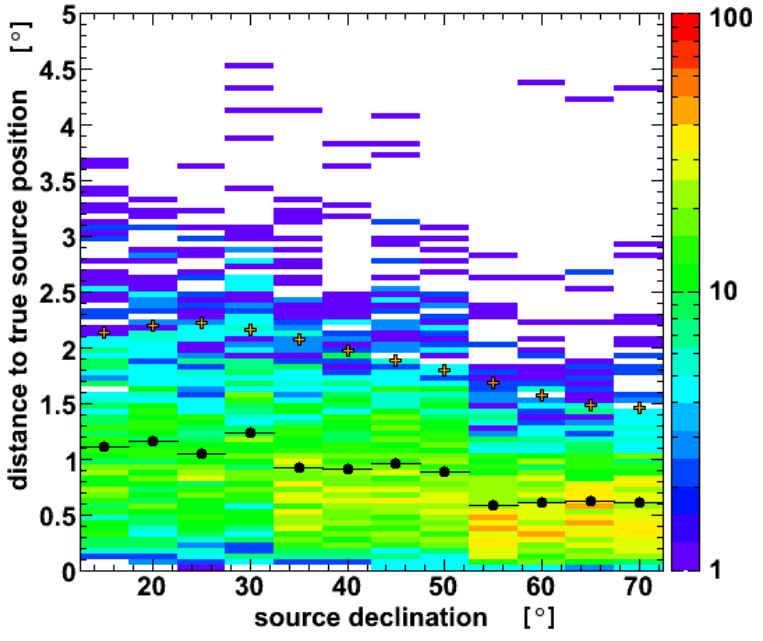


Figure 7.8: *SPA in function of the declination of the true source. The mean is represented by black dots. The crosses account for the detector resolution.*

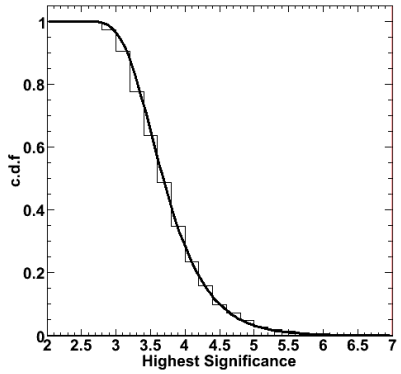


Figure 7.9: *Post Trial factor distribution of the lowest p-value in pure background skies from Data Challenge sample. The distribution is fitted with Eq.7.2*

Post-Trial factor equivalent probability

The procedure described in this analysis naturally controls the false discovery rate, thus the confidence level of a discovery. However, it could be interesting to have the expression of the highest excess $\xi_{\max} = -\log_{10}$ p-value$_{\min}$, observed in a sample in terms of *probability of occurrence* in pure background skies, as it is done *a posteriori* for likelihood based searches. For a set of independent p-values, this *post-trial factor* probability (see section 3.1.1) to see at least such an excess due to background fluctuation is given by :

$$p_{tf}(\xi_{\max}) = 1 - \left(1 - 10^{-\xi_{\max}}\right)^N \tag{7.1}$$

where N is the size of the sample.

However, p-values are strongly correlated from the way they are calculated, and Eq.7.1 must be modified with the coefficient η, accounting for the correlation effect :

$$p_{tf}(\xi_{\max}) = 1 - \left(1 - 10^{-\xi_{\max}}\right)^{N/\eta} \tag{7.2}$$

The coefficient $\eta \in [0; 1]$ with the extreme cases $\eta = 1$ and $\eta =$ N for respectively totally independent and dependent p-values. All p-values are the same in the latter case. Therefore p_{tf} is then independent of the trial factor since $p_{tf}(\xi_{\max}) = 10^{-\xi_{\max}} = p - value_{\min}$.

The distribution of maximal significances with the 17443 sky samples from Data Challenge is fitted (Fig.7.9-(a)) with Eq.7.2. The best fit is obtained for $\eta = 2.14\ ^{+0.15}_{-0.29}$.

There's two possible outcomes : if the result confirms the hypothesis, then you've made a discovery; if the result is contrary to the hypothesis, then you've made a discovery.

Enrico Fermi

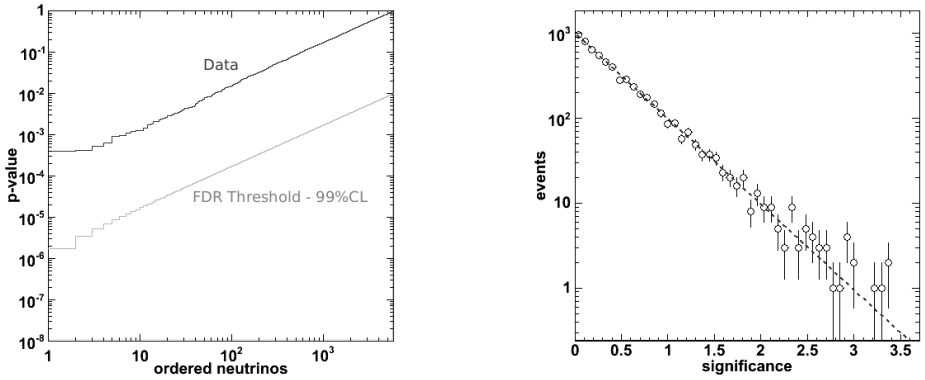
8

Analysis on AMANDA-II 2000-2006 data

The results obtained with the False Discovery Rate controlling procedure applied on the final set of events collected by the AMANDA-II detector between 2000 and 2006 are presented in this chapter. Although these events have been already handle to settle and verify the validity of different points of the method, they were used in a blinded way by a randomization of the right ascension of the events. For this final analysis, the data sample are *unblinded*, that is recovering the complete knowledge about the reconstructed direction of each event. The first part of the analysis concerns a full survey of the northern sky. The second part consists in a detailed scan of 26 predefined regions where sources of high energy γ-rays have been identified. These sources are considered as potential neutrinos emitters.

8.1 All Sky Survey

A total sample of 6100 events survives the cut selection described in section 2.3. However it has been seen thanks to the Data Challenge analysis (section 7.4) that a complementary cut has to be applied to remove a residual background of downgoing muons. Hence a final set of 5848 events survives the final filtering procedure. Figure 8.5 shows the reconstructed position of each event in linear projection. Following the procedure, each neutrino position is the center of a search region and will then be associated with the p-value accounting for the probability for the neutrino to be of atmospheric origin.



(a) FDR Threshold **(b)** p-value distribution

Figure 8.1: *Results of the sky survey with AMANDA-II between 2000 and 2006*
(a) Illustration of the FDR rejection threshold applied on data
(b) Distribution of the p-values. The dashed line is the exponential corresponding to a uniform distribution of p-values required by the FDR rejection method.

The set of p-values are ordered in increasing order and the FDR rejection threshold is calculated with a confidence level of 99%. The results of the rejection method are given in Figure 8.1. It can be seen that no p-value actually crosses the threshold. Therefore no rejection of the Null Hypothesis occurs and no discovery can be claimed at the chosen confidence level.

In absence of rejection, no clustering was performed (see Figure 5.14. The following discussion will then be made on the entire set of events. The maximal significance ξ= 3.39, corresponding to a p-value p = 4.05 10^{-4} is observed at δ= 4.74°, α=190.9°. This probability is more than 2 order of magnitude higher than the FDR threshold ($p_1 = \frac{0.01}{5848}$ = 1.71 10^{-6}). Prior analysis on Data Challenge skies (section 7.4) allows to deduce that the probability (Eq.7.2) to observe such a maximal excess in background skies is 74%.

Table 8.1 lists the 11 lowest p-values. With the exception of the event indexed #9, the events stem from two separate regions, referred in the following as the first and second *hot spot*, respectively around δ= 4°, α=190° and δ= 54°, α=170°. The first hot spot is located less than 5° away from quasar 3C273, which has been listed as one of the 26 possible source candidates (see next section). The second one corresponds to the location of the maximal significance found with the likelihood ratio method [Bra08].

The reasons for the events located in these two areas to have such high significances are rather different. The high significances from the first spot are due to a higher density of events, while the second comes from the contribution of

very high energy events. These regions are zoomed in Figures 8.3 and 8.4. The main events composing the first hot spot are indexed with the number 1,3,5,6 in Table 8.1 and highlighted on the left part of Figure 8.3. The right part of the figure indicates the three most energetic events (with respectively 83, 63 and 60 N_{ch}). Despite their higher energy, these events are not the ones with higher significance, due to a lower number of close-by events.

Globally, the energy of the considered events are low. However, the search bins are larger ($R_{opt} \sim 3.5^{\circ}$) and the expected background is lower ($n_{bg} \sim 7.5$). Finally the number of counted events is actually higher if compared with the second spot (labelled with numbers 2,7,8,10 and 11). As an example the probability to observe at least 19 events when one expects 7.69 events is $p_p = 4.1\ 10^{-4}$.

Unlike for the first spot, Figure 8.4 shows that the energy distribution in the case of the second spot is characterized by high values of N_{ch}. Hence, it is interesting to notice in the right plot that the 3 most central events in Figure 8.4 have respectively an energy of 175, 91 and 75 N_{ch}. The probability p_e to observe an event with 175 N_{ch} is $p_e = 2.5\ 10^{-4}$. However, it can be seen on the left plot that none of these high energy events correspond to the neutrino with the highest significance.

Figure 8.2 gives an illustration of the distribution of the significances (top) and

i	RA $^{\circ}$	Dec $^{\circ}$	R_{opt} $^{\circ}$	n_{obs}/n_{bg}	N_{ch}	$\sum N_{ch}$	ξ	Φ_0	region
1	190.9	4.74	3.54	19 / 7.69	31	791	3.39	9.26	1st HS
2	168.3	55.08	2.67	17 / 8.99	24	906	3.37	13.45	2nd HS
3	189.5	3.90	3.56	20 / 7.57	28	770	3.27	9.22	1st HS
4	172.7	54.66	2.68	17 / 9.06	34	892	3.19	13.42	2nd HS
5	190.4	3.36	3.56	18 / 7.46	53	726	3.08	9.20	1st HS
6	190.3	2.64	3.55	17 / 7.32	38	717	3.03	9.17	1st HS
7	171.9	53.46	2.73	17 / 9.25	75	872	2.96	12.00	2nd HS
8	168.9	53.60	2.72	16 / 9.22	54	865	2.91	12.02	2nd HS
9	29.2	17.31	3.43	17 / 7.97	22	750	2.90	9.71	-
10	168.4	54.32	2.70	16 / 9.11	30	856	2.89	12.05	2nd HS
11	172.3	54.61	2.69	16 / 9.07	175	846	2.84	10.73	2nd HS

Table 8.1: *List of the 11 events with the lowest p-values. RA and Dec are the event angular coordinates. R_{opt} accounts for the radius of the search region. n_{obs} and n_{bg} are respectively the number of observed and expected event in the search region. N_{ch} is the energy estimator of the event. $\sum N_{ch}$ is the sum of the energies for the total number of events observed in the search region. ξ gives the significance and Φ_0 is the neutrino flux limit on $E^{-2}\ d\Phi^{+}_{90}/dE$ 10^{-11} TeV cm^{-2} s^{-1}. These events are mainly gathered into two local regions in the sky, referred as the 1st (Fig.8.3) and 2nd (Fig.8.4) Hot Spot.*

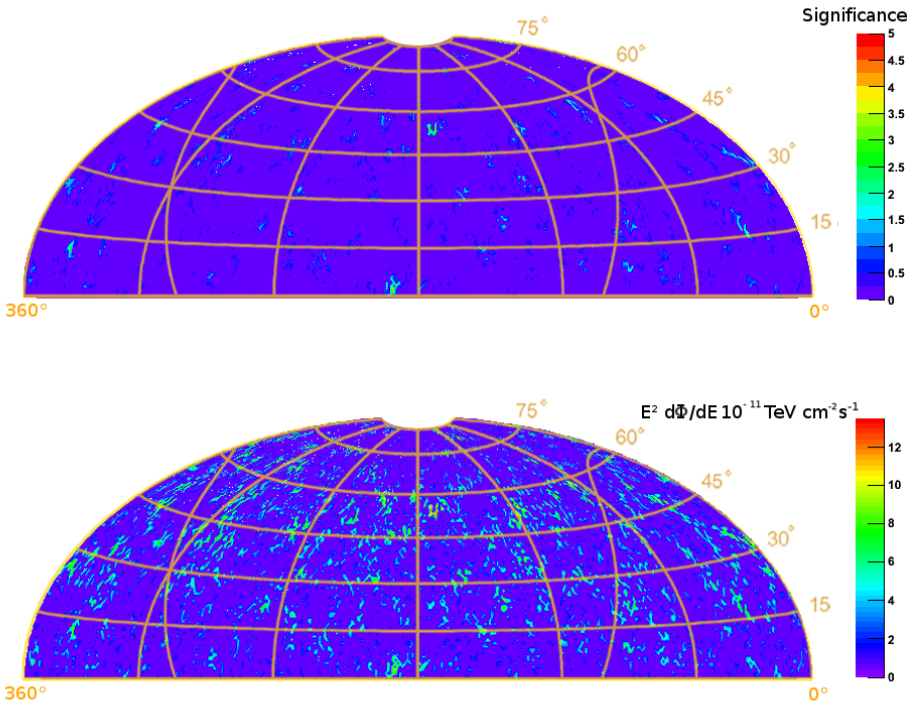


Figure 8.2: *Top : Map of significances ($\xi = -\log_{10}$p-value) obtained with the all sky survey in Hammer-Aitoff projection.*
Bottom : Sky map of flux upper limits deduced from Feldman-Cousins intervals.

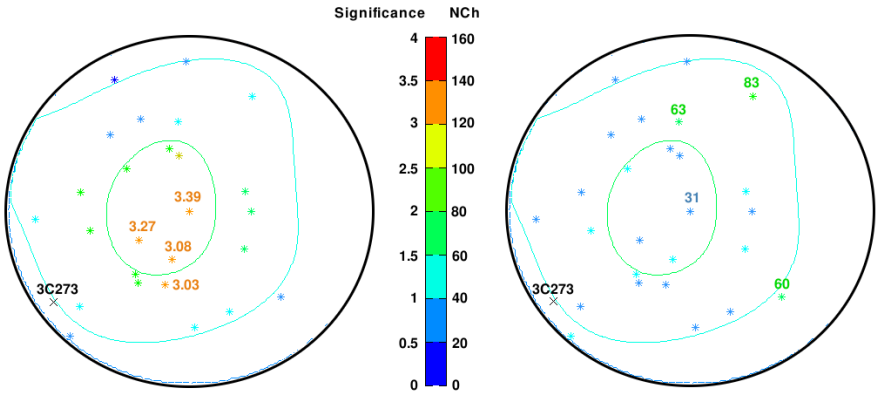


Figure 8.3: *Significance (left) and energy (right) distribution within 5° around the FDR Hot Spot (δ =4.74 °; α =190.9 °). The colored lines reflect the energy density. The position of the quasar 3C 273 is indicated. On the left plot are shown the events with the four highest significance in the region. Identically, the three most energetic events are labelled on the right plot. See text for details.*

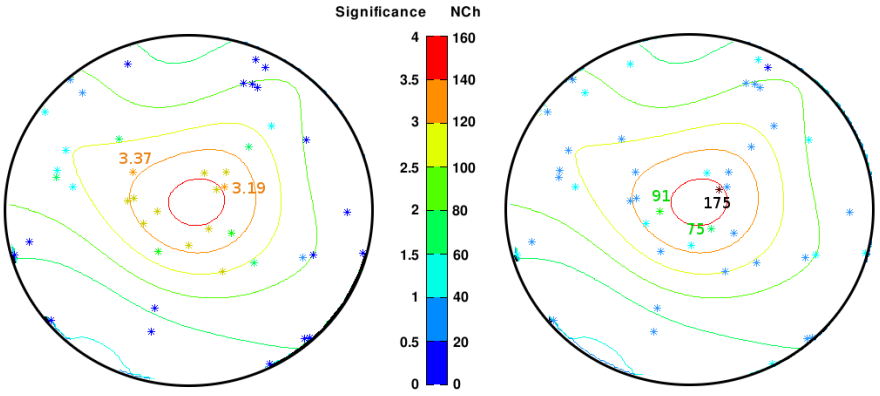


Figure 8.4: *Significance (left) and energy (right) distribution within 5° around the second hottest spot. This location (δ = 54 ° ; α = 170 °) corresponds to the maximal significance obtained with the likelihood ratio based analysis. On the left plot are shown the events with the highest significance in the region. The three most energetic events are labelled on the right plot. See text for details.*

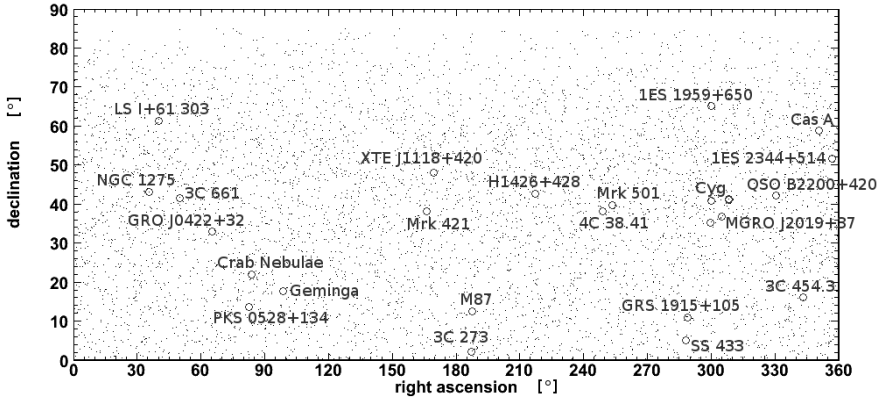


Figure 8.5: *Map of the final 5848 events used in this analysis of AMANDA-II 2000-2006 data. In red are superposed the 26 neutrinos source candidates for which a local scan has been performed.*

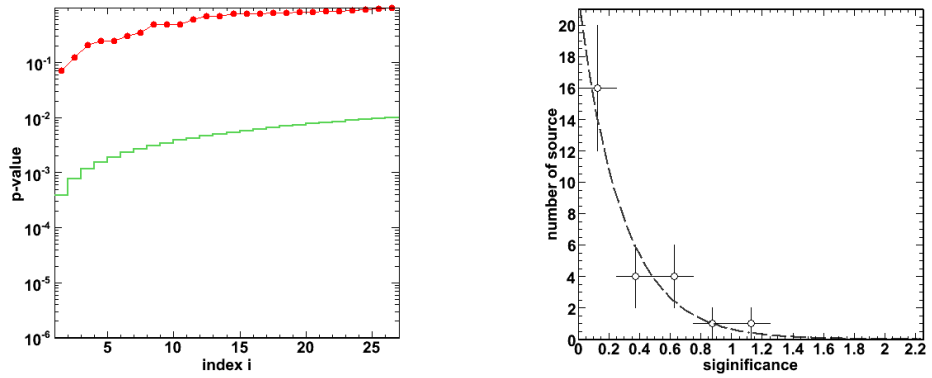
fluxes (bottom) deduced from the Feldman-Cousins limits on total N_{ch} shown in Figure 7.6. It is worth to remind that flux calculation is not available for regions with no reconstructed event. The flux limit around the first hot spot is $\sim$9.2 10^{-11} TeV cm^{-2} s^{-1}. The flux limit around the second hot spot is 10.7 - 13.5 10^{-11} TeV cm^{-2} s^{-1}. The average flux limit on the nothern sky is 4.504 10^{-11} TeV cm^{-2} s^{-1}.

8.2 Source Search

A particular attention have been payed on the location of 26 preselected high energy γ-ray sources to which neutrino production could be associated. The angular positions have been taken as the center of a search region and a set of 26 p-values has been defined which the FDR rejection procedure can be applied on. As it can be seen in Figure 8.6, no p-value crosses the threshold and therefore no discovery can be claimed within a confidence level of 99$\%$. An estimate of the upper limit on the expected flux from the sources is put following the collected energy and the Feldman-Cousins intervals defined in Figure 7.6. Results are summarized in the Table 8.2.

Further information are provided in the Table.8.3. For each source location, the maximal excess $\xi_{5^\circ}^{\max}$ within a 5° range has been identified, as well as the significance and the angular distance of the closest neutrino.

The highest excess found among the 26 positions is the Crab Nebulae. The significance at this position is rather low (ξ=1.15). However, the neutrino with



(a) FDR Threshold **(b)** p-value distribution

Figure 8.6: *(a) Illustration of the FDR rejection threshold (green) applied on sources catalog (red). No p-value crosses the threshold. (b) Distribution of the p-values.*

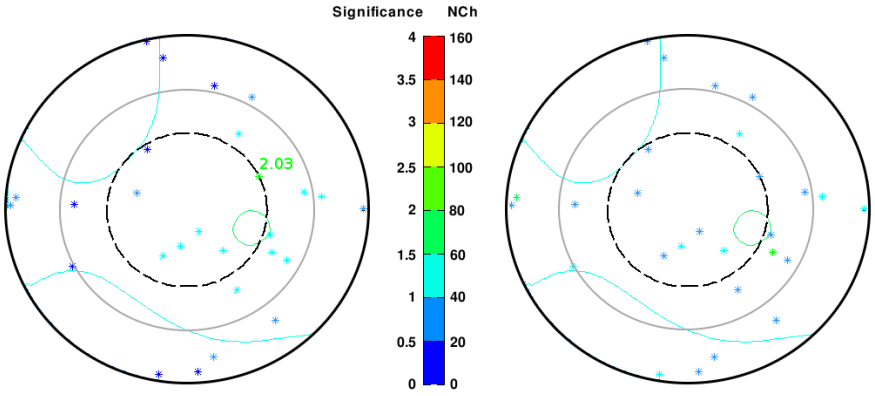


Figure 8.7: *Events significance (left) and energy (right) within 5° (plain black circle) around Crab nebulae (δ =22.01 °;α =83.6 °). The dashed circle accounts for the resolution of the detector at this declination. The grey plain circle is the search region of angular radius R_{opt}. The colored lines reflect the energy density.*

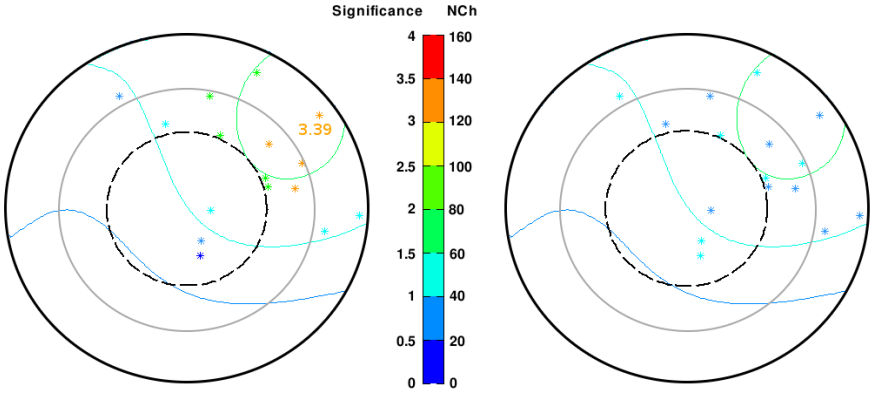


Figure 8.8: *Events significance (left) and energy (right) within 5° (plain black circle) around the quasar 3C273 (δ =2.05 °;α =187.3 °). The dashed circle accounts for the resolution of the detector at this declination. The grey plain circle is the search region of angular radius R_{opt}. The colored lines reflect the energy density.*

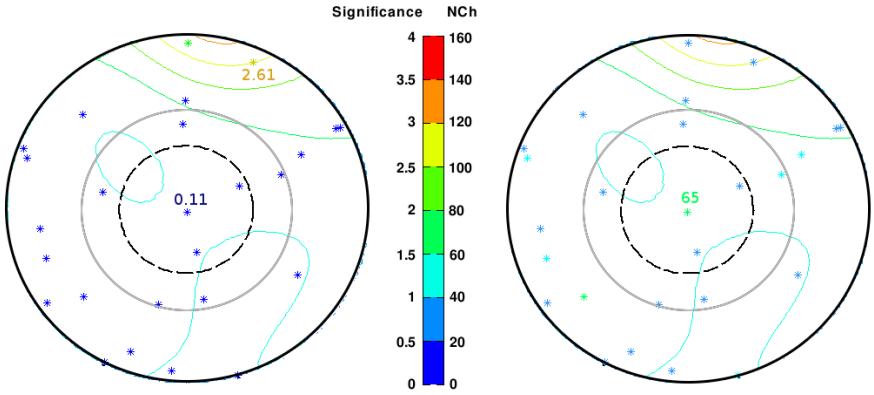


Figure 8.9: *Events significance (left) and energy (right) within 5° (plain black circle) around XTE J1118+480 (δ =48.08 °;α =169.6 °). The dashed circle accounts for the resolution of the detector at this declination. The grey plain circle is the search region of angular radius R_{opt}. The colored lines reflect the energy density.*

the highest excess in the region has a significance of 2.03. The distribution of events is shown in Figure 8.7.

Figure 8.8 shows the event distribution within an angle of 5° around the quasar 3C273. The search region (of angular radius R_{opt}) and the resolution of the detector are shown respectively in plain and dashed circles. It has been mentioned in previous section that several events with high significances (cf Fig.8.3) are located $\Delta\Psi \sim 4$ °northwards from the source. The low significance calculated when the search region is centered on the source position is explained by the fact that all this excess is located outside the region. The local excess [1] might nevertheless still be caused by a neutrino flux from the quasar. The deviation to the north might be explained by the cut-off introduced by the horizon, since the analysis is limited to the northern sky. All events in this region are of low energy. The low p-values are therefore entirely due to a density effect. The source being located close to the horizon, the contribution from additional events reconstructed slightly below the horizon due to the resolution of the detector is totally wiped out. The horizon cut-off also explains the difference of flux calculated between the hot spot position (Table 8.1) and the source position (Table 8.2).

The 5° area around XTE J1118+480 is shown in Figure 8.9. Here again the majority of detected events are of low energy. However it can be interesting to note that the most energetic event in the region is located at the position of the source. The accumulation that can be guessed on the upper part of the plot is actually the second sky highest *Hot Spot* centered in Figure 8.4.

[1] non significantly rejected by the FDR procedure

Object	RA $^\circ$	Dec $^\circ$	R_{opt} $^\circ$	n_{ev}	n_{bg}	$\sum N_{ch}$	ξ	Φ^+_{90}
Crab Nebulae	83.63	22.01	3.51	14	8.64	540	1.15	6.91
3C 273	187.28	2.05	3.52	11	7.11	427	0.91	5.64
Cyg A	299.87	40.73	3.10	12	9.73	471	0.68	7.23
Cyg X-3	308.11	40.96	3.09	15	9.74	456	0.61	7.24
Cyg OB2	308.29	41.32	3.08	15	9.74	456	0.61	7.25
GRS 1915+105	288.80	10.95	3.35	11	7.47	364	0.52	4.39
Mrk 501	253.47	39.76	3.13	12	9.70	419	0.46	6.00
1ES 2344+514	356.77	51.71	2.79	10	9.48	364	0.30	5.26
1ES 1959+650	300.00	65.15	2.35	9	7.12	317	0.31	6.29
QSO B2200+420	330.68	42.28	3.05	10	9.77	376	0.31	4.85
GRO J0422+32	65.43	32.91	3.35	9	9.49	330	0.21	4.53
3C 454.3	343.49	16.45	3.40	7	7.81	257	0.15	3.79
MGRO J2019+37	304.83	36.83	3.22	10	9.61	308	0.15	3.55
XTE J1118+480	169.55	48.04	2.90	8	9.76	289	0.11	3.73
4C 38.41	248.82	38.14	3.17	9	9.65	287	0.11	3.58
NGC 1275	49.95	41.51	3.08	9	9.75	290	0.11	3.62
H 1426+428	217.14	42.68	3.04	8	9.78	285	0.10	3.64
Cyg X-1	299.59	35.2	3.27	9	9.57	278	0.10	3.48
Cas A	350.85	58.82	2.53	7	8.25	248	0.08	4.20
3C 661	35.67	43.04	3.03	8	9.79	273	0.08	3.65
Mrk 421	166.11	33.21	3.18	9	9.66	260	0.07	2.27
LS I +61 303	40.13	61.23	2.45	4	7.77	216	0.06	2.94
PKS 0528+134	82.74	13.52	3.35	5	7.55	187	0.05	1.75
Geminga	98.48	17.77	3.44	5	8.03	170	0.03	1.95
M87	187.71	12.39	3.35	4	7.49	136	0.02	1.71
SS 433	287.96	4.98	3.54	2	7.71	76	0.001	0.92

Table 8.2: *Summary of the results with the search region centered on the 26 candidate source positions. The source name is given with its position.* R_{opt} *is the angular radius of the search region.* n_{obs} *and* n_{bg} *are respectively the number of observed events and expected event in the search region.* $\sum N_{ch}$ *is the total energy in the search region.* ξ *gives the significance and* Φ^+_{90} *is the neutrino flux limit on* E^{-2} $d\Phi^+_{90}/dE$ 10^{-11} *TeV cm*$^{-2}$ *s*$^{-1}$

Object	$\xi_{5^\circ}^{\max}$	ξ_{cl}	$d_{\xi_{\mathrm{cl}}}$ [°]	
Crab Nebulae	2.03	1.34	0.70	Pulsar
3C 273	3.39	1.11	0.65	Quasar
Cyg A	1.34	0.007	0.82	Radio-galaxy
Cyg X-3	0.85	0.61	0.51	Binary X-ray src.
Cyg OB2	0.85	0.61	0.14	Cluster
GRS 1915+105	0.54	0.53	0.46	Microquasar
Mrk 501	0.92	0.37	0.97	TeV Blazar
1ES 2344+514	0.93	0.31	0.31	BL Lac.
1ES 1959+650	1.20	0.57	0.50	Blazar
QSO B2200+420	2.31	0.29	0.62	BL Lac.
GRO J0422+32	0.59	0.15	0.99	X-Ray Nova
3C 454.3	0.50	0.15	1.03	Quasar
MGRO J2019+37	0.84	0.15	0.18	VHE γ-ray src.
XTE J1118+480	2.61	0.11	0.07	X-Ray Nova
4C 38.41	0.92	0.11	1.20	Quasar
NGC 1275	0.68	0.11	0.19	Seyfert Galaxy
H 1426+428	0.65	0.10	1.42	Blazar
Cyg X-1	1.34	0.007	0.53	Binary X-ray src.
Cas A	1.07	0.04	0.91	SN Remnant
3C 661	0.27	0.08	0.11	BL Lac.
Mrk 421	0.28	0.001	1.56	AGN
LS I +61 303	0.63	0.06	0.48	Microquasar
PKS 0528+134	0.72	0.09	2.42	Quasar
Geminga	1.08	0.13	2.16	Pulsar
M87	0.73	0.05	1.75	AGN
SS 433	0.52	0.001	1.43	Microquasar

Table 8.3: *Summary of the results with the search region centered on the 26 candidate source positions (cont'd).* $\xi_{5^\circ}^{\max}$ *is the maximal event p-value found within 5° around the source position.* ξ_{cl} *and* $d_{\xi_{cl}}$ *are the significance and angular distance of the closest event to the source.*

9

Conclusions and perspectives

This work described an original statistical approach for the search of astrophysical neutrinos point sources with the AMANDA-II neutrino detector. The method has been eventually applied on the data collected between 2000 and 2006.

The False Discovery Rate controlling procedure is based on the statistical multiple tests rejection algorithm mathematically develop by Benjamin and Hochberg. It takes naturally into account the trial factor encountered when working on a set of multiple null hypotheses. Moreover, the rejection procedure is only based on the knowledge of the background distribution. In other words, no information about any alternative signal has to be known, making the method particularly suitable for the detection of unknown sources. Finally, the rate of false discoveries accepted for the analysis is fixed beforehand by the user.

This mathematical algorithm has been customized for the search of neutrino point sources.The thesis thoroughly covered the development of the method, from a simple sky pixelisation which permitted to verify the faisability of the method, to the multiple enhancements eventually leading to the final procedure. On one hand, a performant observable combining available information about the angular position and the energy of the observed neutrino has been defined. This observable allowed to build one dimensional probability density functions for the null hypothesis. On the other hand, a clustering algorithm has been developed to avoid multiple counting of the same signal and reconstruct the angular position of the source.

The comparison of the FDR procedure with the likelihood ratio method shows

that the both approaches gives similar performances, with the advantages that the FDR method is independent of the source model and naturally dedicated to multiple source detection. No significant excess has triggered the rejection method applied to the AMANDA-II data collected between 2000 and 2006. As no signal was found, an upper limit on the neutrino flux from a source on the northern sky is calculated: E^2 dΦ/dE = 4.5 10^{-11} TeV cm^{-2}s^{-1} at a confidence level of 99%.

The two regions corresponding to the lowest null hypothesis probability are located on one hand, 4° away from the quasar 3C 273, and on the other hand at the position corresponding to the best likelihood ratio. Both are possible interesting regions for future analysis with the IceCube detector.

An analysis has been performed on a list of 26 positions corresponding to as many sources of high energy γ-rays and thought to be possible emitters of high energy neutrinos. No significant excess were found among the corresponding 26 studied angular positions.

During the last four years, the IceCube detector has progressively been brought to an ever bigger instrumented volume, allowing to reach a sensitivity of $E_\nu^2 dN_\nu/dE_\nu = 5.5\ 10^{-12}$ TeV cm^{-2} s^{-1} sr^{-1} after one year. After five years the sensitivity approaches $E_\nu^2 dN_\nu/dE_\nu = 1.7\ 10^{-12}$ cm^{-2} s^{-1} sr^{-1} TeV [Aa04b]. Following neutrino flux predictions, several sources searched in this present analysis, such as LS I+61 303 [NR04] or Cyg-X3 [Bed05], would be detectable within one year of data taking with the complete IceCube detector. Up to know, data gathered by the partially build detector configurations are available. Analysis of these new events with the current search procedure could soon be considered. However the null hypothesis probability density function has to be reevaluated to take into account the specific characteristics of the new detector.

The procedure can be envisaged to be extend to a better coverage of the northern sky by a new definition of the search regions non-centered on each neutrino, in order to generalize the setting of flux limits and maybe enhance the discovery potential of the method.

A possible extension of the FDR controlling procedure will be to take into account the arrival time of an event as a third observable to identify cosmic neutrino flares. The arrival time of a neutrino could be a major hint in the detection of neutrino flares since the probability to observe uncorrelated neutrinos in a limited angular region and time window is very low. This extension could then be especially usefull for the search of transient sources like Gamma Ray Bursts or Active Galactic Nuclei, if related to simultaneous γ flares. Preliminary studies have been started on this topic with encouraging perspectives.

List of Figures

List of Tables

Bibliography

[Aa04a] R. U. Abbasi and al. Measurement of the flux of ultrahigh energy cosmic rays from monocular observations by the high resolution fly's eye experiment. *Phys. Rev. Lett.*, 92(15):151101, 2004.

[Aa04b] J. Ahrens and al. Status of the icecube neutrino observatory. *New Astronomy Reviews*, 48(5-6):519 – 525, 2004. 2nd VERITAS Symposium on the Astrophysics of Extragalactic Sources.

[Aa06a] M. Ackermann and al. Optical properties of deep glacial ice at the south pole. *Journal of Geophysical Research*, 111:13203–+, 2006.

[Aa06b] V. Aynutdinov and al. Baikal experiment: Main results obtained with the neutrino telescope nt200. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 567(2):423 – 427, 2006.

[Aa09] R. Abbasi and al. Determination of the atmospheric neutrino flux and searches for new physics with amanda-ii. *Physical Review D (Particles, Fields, Gravitation, and Cosmology)*, 79(10):102005, 2009.

[Abd07] H. Abdi. Bonferroni and šidák corrections for multiple comparisons. *Encyclopedia of Measurement and Statistics*, 2007.

[Ac07] J. A. Aguilar and ANTARES collaboration. The data acquisition system for the antares neutrino telescope. *NUCL.INSTRUM.METH.A*, 570:107, 2007.

[Ack06] M. Ackermann. *Search for signals from cosmic point-like sources of high energy neutrinos in 5 years of AMANDA-II data*. PhD thesis, Humboldt-Universität, Berlin, 2006.

[Ahr04] J. et al. Ahrens. Muon track reconstruction and data selection with the amanda detector. *Nucl. Instrum. Meth.*, 2004.

[AMH02] J. Alvarez-Muniz and F. Halzen. Possible high-energy neutrinos from the cosmic accelerator rx j1713.7-3946. *The Astrophysical Journal Letters*, 576(1):L33–L36, 2002.

[Ant09] Status and first results of the antares neutrino telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 604(1-2, Supplement 1):S136 – S142, 2009.

[Ba99] V.A. Balkanov and al. The lake baikal neutrino telescope nt-200: status, results, future. *Nuclear Physics B - Proceedings Supplements*, 75(1-2):409 – 411, 1999.

[Bal06] J. Ballet. X-ray synchrotron emission from supernova remnants. *Advances in Space Research*, 37(10):1902 – 1908, 2006.

[BDD09] . Bonjak, F. Daigne, and G. Dubus. Prompt high-energy emission from gamma-ray bursts in the internal shock model. *Astronomy and Astrophysics*, 498(3):677–703, may 2009.

[BDDP+08] J. Braun, J. Dumm, F. De Palma, C. Finley, A. Karle, and T. Montaruli. Methods for point source analysis in high energy neutrino telescopes. *Astroparticle Physics*, 29:299–305, 2008.

[Bed05] W. Bednarek. High energy neutrinos from microquasars. In *International Cosmic Ray Conference*, volume 5 of *International Cosmic Ray Conference*, pages 67–+, 2005.

[Ben] Y. Benjamini. private conversation.

[Bet90] H. A. Bethe. Supernova mechanisms. *Rev. Mod. Phys.*, 62(4):801–866, 1990.

[BH95] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J.R.Statist.Soc.B*, 57(1):289–300, 1995.

[BP97] W. Bednarek and R. J. Protheroe. Gamma rays and neutrinos from the crab nebula produced by pulsar accelerated nuclei. *Phys. Rev. Lett.*, 79(14):2616–2619, Oct 1997.

[Bra] J. Braun. http://www.icecube.wisc.edu/ jbraun/combinedpointsource

[Bra08] J. Braun. Amanda 2000-2006 point source search methodology. *Icecube internal 200804001*, 2008.

[Bra09] J. Braun. *A Maximum-likelihood search for neutrino Point Sources with the AMANDA-II detector*. PhD thesis, University Wisconsin, 2009.

[BvC02] A. Biron von Curland. *Search for Atmospheric Muon-Neutrinos and Extraterrestric Neutrino Point Sources in the 1997 AMANDA-B10 Data*. PhD thesis, Humboldt-Universität, 2002.

[BY01] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

[BZ34] W. Baade and F. Zwicky. Remarks on super-novae and cosmic rays. *Phys. Rev.*, 46(1):76–77, 1934.

[Chi08] Zhiyi Chi. False discovery rate control with multivariate p-values. *Electronic Journal of Statistics*, 2:368, 2008.

[Cow56] F. Reines F. B. Harrison Cowan, Jr. Detection of the free neutrino: A confirmation. *Science*, 124(103), 1956.

[CR04] Dmitry Chirkin and Wolfgang Rhode. Muon Monte Carlo: A high-precision tool for muon propagation through matter. 2004.

[DGWL02] C. Distefano, D. Guetta, E. Waxman, and A. Levinson. Neutrino flux predictions for known galactic microquasars. *The Astrophysical Journal*, 575:378, 2002.

[DJ] D Donoho and J. Jin. Asymptotic minimaxility of false discovery rate thresholding for sparse exponential data. *to appear in Annals of Statistics*.

[DM99] Charles D. Dermer and Kurt E. Mitman. Short-timescale variability in the external shock model of gamma-ray bursts. *The Astrophysical Journal Letters*, 513(1):L5–L8, 1999.

[EW01] A. D. Erlykin and A. W. Wolfendale. Structure in the cosmic ray spectrum: an update. *Journal of Physics G Nuclear Physics*, 27:1005–1030, 2001.

[FC98] Gary Feldman and Robert Cousins. Unified approach to the classical statistical analyses of small signals. *Physical Review D*, 57(7):3873–3888, 1998.

[Fer49] E. Fermi. On the origin of the cosmic radiation. *Phys. Rev.*, 75(8):1169–1174, 1949.

[Fil05] A. V. Filippenko. Supernovae and their massive star progenitors. In *The Fate of the Most Massive Stars*, volume 332 of *Astronomical Society of the Pacific Conference Series*, pages 33–+, 2005.

[Gai90] Thomas K Gaisser. *Cosmic rays and particle physics*. Cambridge Univ. Press, Cambridge, 1990.

[GP03] S.B. Gabriel and G.J. Patrick. Solar energetic particle events:phenomenology and prediction. *Space Science Reviews*, 107:55–62, 2003.

[Gre66] K. Greisen. End to the cosmic-ray spectrum? *Phys.Rev.Lett.*, 16(17):748–750, 1966.

[GY01] A. Z. Gazizov and S. I. Yanush. Small-x growth of nucleon structure functions and its manifestations in ultrahigh-energy neutrino astrophysics. *astro-ph/0112244*, 2001.

[Haa] D. Heck and al. Corsika: A monte carlo code to simulate extensive air showers.

[Hab] S. Hundertmark and al. Amasim — simulation tool for amanda. http://www.physto.se/ amanda/amasim/amasim.html.

[HA03] A. Heijboer and ANTARES Collaboration. Point Source Searches with the ANTARES Neutrino Telescope. In *International Cosmic Ray Conference*, volume 3 of *International Cosmic Ray Conference*, pages 1321–+, July 2003.

[Han] W. Hanlon. http://www.physics.utah.edu/ whanlon/spectrum.html.

[Hau04] T. Hauschildt. *Search for Cosmic Point Sources of High Energy Neutrinos with the AMANDA-II Detector*. PhD thesis, Humboldt-Universität, Berlin, 2004.

[Hea] NASA Healpix. http://healpix.jpl.nasa.gov.

[Hes12] Victor Hess. Über beobachtungen der durchdringenden strahlung bei seiben freiballonfahrten. *Physics Z.*, 13:1084–1091, 1912.

[HF60] F. Hoyle and W. A. Fowler. Nucleosynthesis in Supernovae. *Astrophysical Journal*, 132:565–+, November 1960.

[HH05] Francis Halzen and Dan Hooper. High energy neutrinos from the tev blazar 1es 1959+650. *Astroparticle Physics*, 23(6):537 – 542, 2005.

[HP98] Y. D. He and P. B. Price. Remote sensing of dust in deep ice at south pole. *Journal of Geophysical Research*, 103:17401–+, 1998.

[Jam06] F. James. *Statistical Methods in Experimental Physics*. World Scientific, 2nd edition, 2006.

[JMF99] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[Kam08] K.-H. Kampert. Ultra high-energy cosmic ray observations. *Journal of Physics Conference Series*, 120(6):062002–+, 2008.

[Kel] J. Kelley. http://www.icecube.wisc.edu/ jkelley/atm/excess.html.

[Lip93] Paolo Lipari. Lepton spectra in the earth's atmosphere. *Astroparticle Physics*, 1(2):195 – 227, 1993.

[LM00] John G. Learned and Karl Mannheim. High-energy neutrino astrophysics. *Annual Review of Nuclear and Particle Science*, 50(1):679–749, 2000.

[LMW+07] J. Lundberg, P. Miocinovic, K. Woschnagg, T. Burgess, J. Adams, S. Hundertmark, P. Desiati, and P. Niessen. Light tracking through ice and water–scattering and absorption in heterogeneous media with photonics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581(3):619 – 631, 2007.

[Lon92] M.S. Longair. *High Energy Astrophysics: Volume 1, Particles, Photons And Their Detection*. Cambridge University Press, 1992.

[LS88] I. Lerche and R. Schlickeiser. On the energy spectrum of galactic primary cosmic rays - results from the transport equation. *Astrophysics and space science*, 145:319–354, 1988.

[LW01] Amir Levinson and Eli Waxman. Probing micro-quasars with tev neutrinos. *Physical Review Letters*, 87:171101, 2001.

[Man93] K. Mannheim. gamma-rays and neutrinos from a powerful cosmic accelerator. *Phys. Rev. D*, 48(6):2408–2414, Sep 1993.

[Mar84] B. Margon. Observations of SS 433. *Annual Review of Astronomy and Astrophysics*, 22:507–536, 1984.

[MGN+] Christopher J. Miller, Christopher Genovese, Robert C. Nichol, Larry Wasserman, Andrew Connolly, Daniel Reichart, Andrew Hopkins, Jeff Schneider, and Andrew Moore. Controlling the false discovery rate in astrophysical data analysis. *The Astronomical Journal*.

[Mir02] I. F. Mirabel. Microquasars as sources of high energy phenomena. *arXiv(astro-ph/0211085)*, 2002.

[Mue91] D. Mueller. Energy spectra and composition of primary cosmic rays. *Astrophysical Journal*, 374:356–365, 1991.

[NA37] Seth H. Neddermeyer and Carl D. Anderson. Note on the nature of cosmic-ray particles. *Phys. Rev.*, 51(10):884–886, 1937.

[Nag04] S. Nagataki. High-energy neutrinos produced by interactions of relativistic protons in shocked pulsar winds. *The Astrophysical Journal*, 600:883, 2004.

[NE33] J. Neyman and Pearson E. On the problem of the most efficient test of statistical hypotheses. *Philosophical Transactions A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.

[Ner02] A. et al. Neronov. Large scale extragalactical jets powered by very-hight energy gamma rays. *Phys. Rev. Letters*, 89:051101, 2002.

[Ney37] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions A*, 236:333–380, 1937.

[NK06] T. Neunhöffer and L. Köpke. Searching for localized cosmic particle sources with an unbinned maximum likelihood approach. *Nuclear Instruments and Methods in Physics Research*, A 558:561–568, 2006.

[NR04] A. Neronov and M. Ribordy. Neutrino signal from γ-ray-loud binaries powered by high energy protons. *Phys. Rev.*, D79:043013, 2004.

[NS02] A. Neronov and D. V. Semikoz. Which blazars are neutrino loud? *Phys. Rev.*, D66:123003, 2002.

[Pan96] D. Pandel. *Bestimmung von Wasser- and Detektorparametern und Rekonstruction von Myonen bis 100 TeV mit dem Baikal-Neutrinoteleskop NT-72*. PhD thesis, Humboldt-Universität, Berlin, 1996.

[Pee68] H. W. Peers. Confidence properties of bayesian interval estimates. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3):535–544, 1968.

[Pir04] Tsvi Piran. The physics of gamma-ray bursts. *Rev.Mod.Phys*, 76:1143–1210, 2004.

[Ptu81] V. S. Ptuskin. On the Acceleration of Cosmic Rays in the Interstellar Medium. *Soviet Astronomy Letters*, 7:374–+, 1981.

[PW65] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophysical Journal*, 142:419–421, 1965.

[RM94] M. J. Rees and P. Meszaros. Unsteady outflow models for cosmological gamma-ray bursts. *Astrophysical Journal*, 430:L93–L96, August 1994.

[Sil05] A. Silvestri. Performance of amanda-ii using transient waveform recorders. *29th International Cosmic Ray Conference*, 5:431–434, 2005.

[SP94] A. P. Szabo and R. J. Protheroe. Implications of particle acceleration in active galactic nuclei for cosmic rays and high energy neutrino astronomy. *Astroparticle Physics*, 2(4):375 – 392, 1994.

[SS] B Schölkopf and A.J. Smola. *Learning with Kernels – Support Vector Machines, Regularization, Optimization, and Beyond*.

[Sta06] T. Stanev. High energy neutrinos: sources and fluxes. *Journal of Physics: Conference Series*, 39:386–392, 2006.

[Ste02] P. Steffen. Direct walk ii (improved version of direct walk). *AMANDA internal report*, 2002.

[TIG96] M. Thunman, G. Ingelman, and P. Gondolo. Charm production and high energy atmospheric muon and neutrino fluxes. *Astroparticle Physics*, 5(3-4):309 – 332, 1996.

[TRM05] Diego F. Torres, Gustavo E. Romero, and Felix Mirabel. Neutrinos from microquasars. *CHIN.J.ASTRON.ASTROPHYS.*, 5:S133, 2005.

[Vie98] M. Vietri. On the energy of neutrinos from gamma-ray bursts. *Astrophysical Journal*, 507:40–45, 1998.

[vM57] R. von Mises. *Probability, Statistics and Truth*. Dover Publications, 2nd edition, 1957.

[WB97] Eli Waxman and John Bahcall. High energy neutrinos from cosmological gamma-ray burst fireball. *Phys. Rev. Letters*, 78(12):2292–2295, 1997.

[WB99] Eli Waxman and John Bahcall. High energy neutrinos from astrophysical sources: An upper bound. *Phys. Rev.*, D(59), 1999.

[WPH65] R. W. Wilson, A. A. Penzias, and D. C. Hogg. Measurement of the Flux Density of CAS A at 4080 Mc/sec. *Astrophysical Journal*, 70:150–+, 1965.

[WW90] J. Wdowczyk and A. W. Wolfendale. Implications of an extragalactic origin for the highest energy cosmic rays. *Astrophysical Journal*, 349:35–40, 1990.

[Yam08] T. Yamamoto. The UHECR spectrum measured at the Pierre Auger Observatory and its astrophysical implications. In *International Cosmic Ray Conference*, volume 4 of *International Cosmic Ray Conference*, pages 335–338, 2008.

[YL98] L. Yungelson and M. Livio. Type ia supernovae: An examination of potential progenitors and the redshift distribution. *The Astrophysical Journal*, 497(1):168–177, 1998.

[ZK66] G. Zatsepin and V Kuzmin. Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4(53), 1966.

[ZV06] V.N. Zirakashvili and H.J. Völk. Cosmic ray reacceleration on the galactic wind termination shock. *Advances in Space Research*, 37(10):1923 – 1927, 2006. Astrophysics.