Muon Energy Reconstruction in ANTARES using Neural Networks

R. Bruijn

August 2002

Muon Energy Reconstruction in ANTARES using Neural Networks

Ronald Bruijn

Master's thesis

Supervisor

Dr. E. de Wolf
NIKHEF, Amsterdam

August 2002
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam

Abstract

The ANTARES detector is a neutrino telescope located at the bottom of the Mediterranean sea. The aim is to determine the direction and energy of high energetic cosmic neutrinos. This is achieved by detecting the Cherenkov light produced by the muons created in charged-current interactions of the neutrinos with the rock below the detector or the water. The measurement of the neutrino energy is important as it gives us insight into the astrophysical processes in which they are created.

In this thesis I describe the development of a method to reconstruct the energy of the muon using neural networks. The method makes use not only of the light yield measured in the array of photomultipliers but also of the arrival times of the photons. The energy of muons with an energy between 1 TeV and 10 PeV could be determined within a factor 2 to 1.6, improving with energy. The achieved resolution is a considerable improvement compared to the result of the conventional method used in ANTARES. The results are hardly dependent on the position and direction of the muon track in the detector. The method can be easily extended to muon energies below 1 TeV and above 10 PeV.

Contents

Chapter 1

Introduction

The field of astroparticle physics is a relatively young one. It combines the two more classical scientific disciplines of astrophysics and high energy physics. High energy physicists can study particles that are accelerated to energies not achievable by any earth based accelerator, beyond the PeV scale, originating from cosmic sources. The flux of these particles provides a means of studying astrophysical objects and processes other than with their electro-magnetic spectrum. The use of cosmic particles is in this sense the next step in the evolution of astronomy, which started with optical observations.
Particles that play a special role in astroparticle physics are neutrinos. These particles, leptons, are neutral and only interact through the weak force with an extremely small cross-section. Because of these properties, a neutrino is not bent by galactic and extra-galactic magnetic fields and not easily absorbed by matter which makes up the universe. Also, neutrinos are not affected by the micro-wave and infra-red background which limits the maximum range of high-energy gamma rays to 100 Mpc. In short, neutrinos are a good probe for astrophysical research because they point straight to their source and can travel larger distances than high energy photons, as they are not easily absorbed or scattered.
The small cross-section of neutrinos is also a draw-back, as their detection requires a large target volume. The ANTARES collaboration will build a detector placed in the Mediterranean sea which uses sea-water and the earth as targets for neutrino interactions. Charged-current interactions of muon type neutrinos (ν_μ) result in muons which emit light through the *Cherenkov-effect* as they travel through the water. The ANTARES detector is aimed at measuring this light and using it to determine the direction and energy of the original neutrino.
Just as observations of different wavelengths of light (optical, x-ray, gamma-ray) are important for electro-magnetic astronomy, the determination of the energy of the neutrino is important for neutrino-astronomy. The spectra of neutrino energies can be compared to the spectrum of wavelengths emitted by a source. Different models of astrophysical objects such as Active Galactic Nuclei predict different fluxes of neutrino induced muons. The energy of the muon is directly related to the energy of the neutrino. An accurate measurement of the energy can teach us a lot about the inner workings of the neutrino sources and the high-energy processes involved. Also, a good energy resolution can contribute to the search for point sources of neutrinos and can improve cuts to reduce background.
In my effort to develop a method to determine the energy of the muon I have made use of yet another discipline: neural computation. The field of neural computation aims to develop computational techniques based on the most complex information processing unit available : the animal and human nervous system. Inspired by the broad range of complex tasks fulfilled by a collection of simple units, nervous cells, the *neural network* has been developed. The method to reconstruct the energy of the muon in the ANTARES detector I have developed, presented in this thesis, makes use of this technique.

In this thesis I will present my contribution to the ANTARES collaboration which is the devel-

opment of a method to reconstruct the energy of the muon using neural networks. It is described in chapters 4 and 5.

Chapter 2

ANTARES

The goal of a neutrino telescope is to determine the direction and energy of cosmic neutrinos. This endeavour is hampered by a combination of several factors. Firstly, the neutrino being, a neutral particle only interacting through the weak force, has a small cross-section for interaction with matter. Secondly, the predicted fluxes point to a very small event rate, which, thirdly, will be obscured by the very large background from atmospheric neutrinos and muons.
A type of detector that can compensate for these factors is a *water-Cherenkov detector* based on a principle proposed by Markov [1]. This concept entails the use of an array of photomultipliers (PMTs) in a large volume of a transparent medium to detect the light which is emitted by a charged particle, a muon, through the Cherenkov effect. Water and ice are two suitable media which can be found in abundance in the oceans or, in the case of ice, at the Antarctic. The muon is created in a charged-current reaction of a neutrino. The ANTARES collaboration will build a detector which is based on this concept at a depth of 2400 m in the Mediterranean sea. The ANTARES detector is not the first detector based on this principle. The concept was pioneered by the DUMAND [2] collaboration which aimed at building a detector at 4000 m depth near Hawaii. A lot of research has been done and experience has been acquired on designing and operating a underwater neutrino telescope. The project was stopped in 1996. Other experiments include the BAIKAL [3] experiment in the Baikal lake, AMANDA [4] which uses the Antarctic ice as radiating medium. ANTARES is not only intended to pursue its scientific goals, but also as a technology study for a much larger water-Cherenkov detector with an instrumented volume of 1 km^3.

In this chapter I will give a summary of the scientific motivation for building a telescope capable of detecting high-energy neutrinos. This is followed by a description of the concepts and techniques used for neutrino detection and a description of the detector. I will continue with an overview of the software that is used to study the detector and to process data. Finally, the expected fluxes and background measured by the detector will be considered.

2.1 Scientific Motivation

This section summarizes the scientific program of the ANTARES detector. Details can be found in [5], [6],[8] and [9].

2.1.1 Cosmic Rays

The main motivation for the construction of a telescope capable of detecting high-energy neutrinos lies in the observed flux of cosmic rays. Cosmic rays are charged particles that are incident from the sky. As can be seen from figure 2.1, the energy spectrum of cosmic rays extends to

energies higher than 10^{20}eV. The spectrum can be described by a power law

$$\frac{dN}{dE} = c \cdot E^{-\gamma} \tag{2.1}$$

where γ is called the spectral index. The spectral index is not constant. The value up to 10^{15} eV (the 'knee') is 2.7, after which the spectrum becomes steeper with a spectral index of 3.1. At around 10^{19} (the 'ankle') the spectral index again changes value to around 2.7. The question

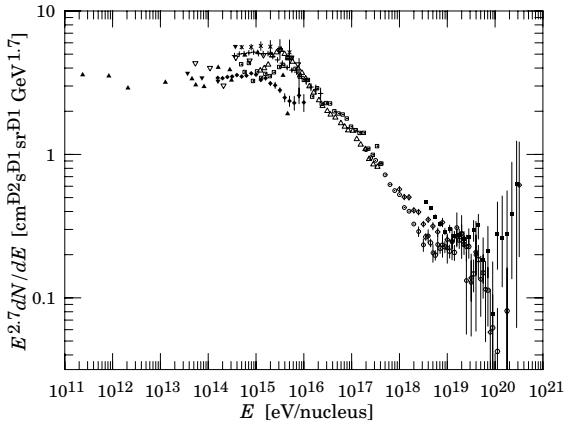

Figure 2.1: The Cosmic Ray flux spectrum, taken from [10]

is about the origin of these cosmic rays. Since the particles are charged, they do not follow a straight path from their point of origin to the Earth, as their paths are bent by galactic and extra-galactic magnetic fields. Therefore it is impossible to trace them back to their source. Only neutral particles retain directional information. Up to the knee the spectrum is believed to be caused by the acceleration of particles by means of the *first-order Fermi* acceleration mechanism [11]. The origin of the remainder of the spectrum is still subject of speculation. While the origin of the cosmic rays in this region is not quite clear, more is known about its composition. The steep part between the knee and the ankle probably consists mainly of heavy nuclei while above the ankle the spectrum is most likely dominated by protons of extragalactic origin [12]. This poses another question as cosmic rays with energies above 10^{19} eV are expected to interact with the cosmic microwave background (Greisen-Zatsepin-Kuz'min (GZK) cutoff) which limits the distance they can travel.
Neutrinos can help us understand the cosmic ray spectrum and the questions about its sources. This is through the association of high-energy neutrino production with the production of high-energy protons. Neutrinos can be generated through the decay of pions and the subsequent decay chain. Pions decay into a muon and a neutrino :

$$\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}/\overline{\nu}_{\mu} \tag{2.2}$$

followed by the decay of the muon:

$$\mu^{\pm} \rightarrow e^{\pm} + \nu_{e}/\overline{\nu}_{e} + \overline{\nu}_{\mu}/\nu_{\mu} \tag{2.3}$$

The pions to begin with can be produced by proton-proton interactions :

$$p + p \rightarrow \pi^{\pm} + X \tag{2.4}$$

The high-energy gamma rays associated with the cosmic rays can also be produced in proton-proton interactions :

$$p + p \rightarrow \pi^{0} + X \rightarrow 2\gamma + X \tag{2.5}$$

In objects with a dense photon gas, pions can also be produced by :

$$p + \gamma \rightarrow \Delta^+ \rightarrow n + \pi^+ \tag{2.6}$$

and

$$p + \gamma \rightarrow \Delta \rightarrow p + \pi^0 \tag{2.7}$$

The first proces contributes to the neutrino flux through the decay of the π^+ and the latter to the gamma-ray flux through the decay of the π^0.
As neutrinos are neutral, their paths are not bent by magnetic fields. They also have a small cross-section for interaction with matter and are not affected by the GZK cutoff. This way neutrinos can be used to study astrophysical objects. However, the neutrino energy spectrum could have dips at energies around $4.2 \cdot 10^{21}$ eV, assuming a neutrino mass of 1 eV due to absorption on the background of relic neutrinos [13].

2.1.2 Neutrino Sources

Supernova Remnants This type of source is thought to be the principal candidate for the production of high-energy cosmic rays. A supernova is the explosion of a massive star. The expanding shell of a supernova can accelerate particles to high energies through first-order Fermi acceleration. The formation of a neutron star can enlarge the acceleration through the effects of its magnetic field.

X-Ray Binary An X-Ray binary is a stellar system composed of a non-compact object and a compact object (neutron star or black hole) rotating around their common center of gravity. An accretion disc is formed around the compact object which draws in matter from the non-compact object. The accretion leads to stochastic acceleration of protons and gives rise to visible jets perpendicular to the accretion disc.

Active Galactic Nuclei Active Galactic Nuclei (AGN) are, averaged over time, the most powerful observed astronomical objects. The standard model of AGN poses a very massive black hole ($10^6 - 10^9$ solar masses) which accretes several solar masses a year. There are several processes that can lead to a neutrino flux from these objects. Protons could be accelerated in the middle of the jets perpendicular to the accretion disc. Also, protons could be accelerated in the discs themselves.

Gamma-Ray Bursts The observation of Gamma-Ray Bursts (GRB) dates back to the 1960's. Since then, after the first identification of their optical and X-Ray afterglow in 1997 it has become clear that they are not only very powerful but also very distant phenomena (redshift of 0.7 to 3.4). The total emitted energy depends on the model. If assumed isotropic, described by the *fireball* model [14], it is between 10^{51} and 10^{54} erg in gamma-rays. This value is a factor of several hundred lower when a beamed emission is assumed, like in the *cannnonbal* model [15]. The fireball model describes the deposition of a solar mass of energy in a 100 km radius resulting in a ultra-relativistic expanding "fireball". Subsequent overtaking of different shells result in the GRB. "Cannonballs" are relativistic chunks of matter originating from an accretion disc that is formed by matter from the parent star that falls back after a supernova. The GRB is produced by interaction of the "cannonballs" with the supernova shell. Neutrinos can be formed by interactions of the hadronic parts of the fireballs or cannonballs.

Relic Sources A possibly very interesting source of cosmic rays with extremely high energy ($> 10^{19}$ GeV) which can evade the GZK cutoff are relic sources. These cosmic rays may originate from the decay of massive particles. These particles could be gauge or Higgs bosons which are formed in the collapse or annihilation of topological defects such as cosmic strings or monopoles.

2.1.3 Other Physics

The ANTARES detector may be able to contribute to the interesting and active field of *neutrino oscillations*. Strong evidence for oscillations between neutrino flavours is found in the SNO [16] and Super-Kamiokande [17] experiments by measurements of the flux of solar neutrinos. These oscillations require a non-zero mass difference between the neutrino flavours. In a two neutrino flavour scheme, the oscillation probability is given by :

$$P = \sin^2(2\theta)\sin^2\left(1.27\frac{L(\mathrm{km})}{E(\mathrm{GeV})}\Delta m^2(\mathrm{eV}^2)\right) \tag{2.8}$$

The parameter space covered by the ANTARES detector provides information on the oscillation probability for $10^{-3}\mathrm{eV}^2 < \Delta m^2 < 2\times10^{-2}\mathrm{eV}^2$.

Luminous matter only accounts for a small fraction of the density of the universe. Several candidates are proposed which can make up the missing mass, the so-called Dark Matter. The lightest supersymmetric particle (LSP) the neutralino is a candidate for non-baryonic Dark Matter. They are predicted to accumulate in heavy bodies such as our sun. Annihilation of a neutralino pair yields a fermion-antifermion pair or two body combinations of W, Z and Higgs bosons. The decay of these particles produces neutrinos.

A sensitivity of the ANTARES detector to τ-neutrinos is expected through their typical signature. The detection of two hadronic showers could be an indication of such a neutrino. The first shower is caused by the ν_τ interaction and the second by the decay of the produced τ lepton which has a very short lifetime.

Other phenomena that are not directly related to (the production of) neutrinos can be detected by their predicted particular signature in the detector. Magnetic monopoles that are predicted by grand unified theories to have been formed shortly after the Big Bang would have cooled down to very low velocities ($\beta < 10^{-3}$). The sequential Cherenkov flashes that this particle would emit could be detected. Another signature could be the production of enourmous amounts of light. This is expected for the passing of a relativistic monopole or q-balls, the dark matter candidates predicted by super-symmetry.

2.2 Neutrino Interactions

Neutrinos are not measured directly. The detector is optimized to detect the (anti-)muon that is created in a charged-current interaction of a muon-(anti-)neutrino with a nucleus in the rock below the detector or in the water surrounding the detector:

$$\nu_\mu(\overline{\nu}_\mu) + N \rightarrow \mu^-(\mu^+) + X \tag{2.9}$$

where N is the struck nucleus and X its hadronic final state. For energies below 10 TeV the cross-section for this reaction (averaged over neutrinos and anti-neutrinos) is [8] :

$$\sigma(\nu_\mu \rightarrow \mu X) \approx 0.50 \times E_\nu[TeV] \times 10^{-35}\mathrm{cm}^2 \tag{2.10}$$

Figure 2.2 shows the neutrino cross-section for a wider energy range. At low energies the linear behaviour can be seen and at higher energies the suppression due to the W propagator.
Important to the neutrino telescope concept is the relation between the incident angles of the neutrino and the muon. Above 1 TeV a safe limit on the average scattering angle is :

$$\langle\theta\rangle < \frac{1.5^o}{\sqrt{E_\nu[TeV]}} \tag{2.11}$$

Hence, with increasing energy the direction of the muon and the neutrino are more and more the same.

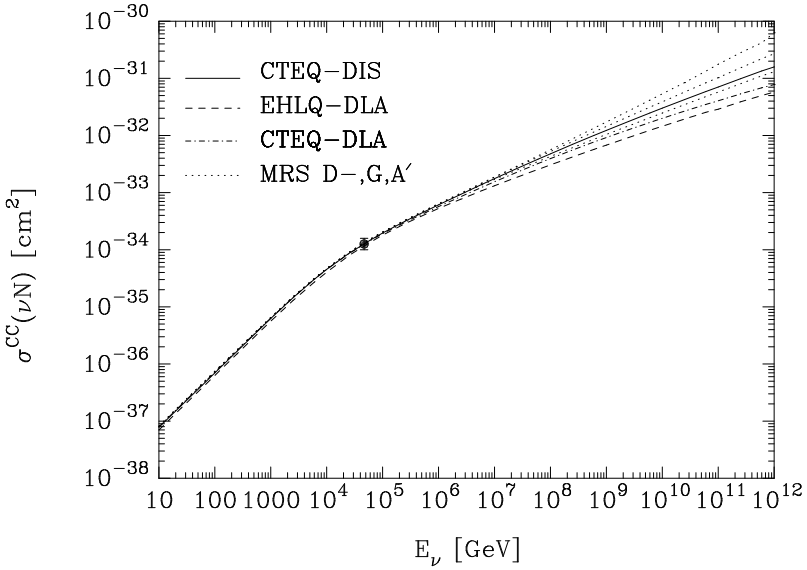

Figure 2.2: Charged current cross section for different parton distributions, taken from [18]. The data point is measured by the ZEUS collaboration.

2.3 Detection Principle

Energetic charged particles traversing a dielectric medium with a velocity $v > c/n$ (with n the refractive index of the medium) emit Cherenkov radiation [19] . The number of photons and the angle at which they are emitted depend on the velocity of the particle and the refractive index. The angle is determined by the relation

$$\cos\theta_c = \frac{1}{\beta n} \tag{2.12}$$

with $\beta > 1/n$. In the energy range in which ANTARES is interested $\beta = 1$ and n is constant. Hence, the Cherenkov angle is independent of the energy of the muon and the light is emitted in a Cherenkov cone with a fixed angle (see figure 2.3). From the condition $\beta > 1/n$, a critical energy E_{crit} can be determined below which no light is emitted. With the refractive index at the ANTARES site of 1.34 the values of θ_c and E_{crit} become 41.7^o and 160 MeV respectively. The emitted photons have a continuous spectrum. The energy loss by emission of Cherenkov photons per unit length and per unit wavelength for a particle with charge Ze (for a muon Z=1) is given by

$$\frac{dE}{dxd\lambda} = 4\pi^2 r_e mc^2 \frac{1}{\lambda^3}\left(1 - \frac{1}{\beta^2 n^2}\right) \tag{2.13}$$

with r_e the classical Bohr radius of the electron. Since the wavelengths of the emitted photons peak at short wavelengths, the light has a blue colour.
The Cherenkov effect is not the only source of light from the muons. With increasing energy other processes contribute to the light yield. Secondary particles which are emitted in these processes can also emit Cherenkov light.

2.4 Muon Propagation

Energy loss of a muon traversing material consists of several parts. Their relative importance depends on the material but mostly on the muon energy (see figure 2.4). The continuous loss due to ionization is dominant at low energies. It rises slowly logarithmically and is practically

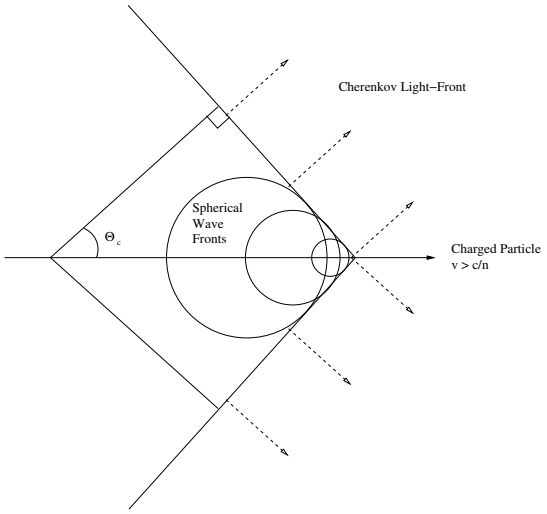

Figure 2.3: Cherenkov light emitted by a charged particle traversing a transparent medium with refractive index n.

constant above 1 GeV. Other losses are due to radiative processes such as Bremsstrahlung, pair-production and photo-nuclear interactions. These processes are stochastical in nature and cause large fluctuations in energy loss. The radiative loss in the high-energy limit increases linearly with the muon energy. Usually, the average energy loss is expressed as

$$-\left\langle \frac{dE_\mu}{dx} \right\rangle = \alpha(E_\mu) + \beta(E_\mu)E_\mu \tag{2.14}$$

where α is the contribution from ionization and β is the sum of the fractional radiation losses. To a first approximation these parameters are independent of the muon energy. At the critical energy $\epsilon_c = \frac{\alpha}{\beta}$ the ionization loss equals the radiative loss. The value of ϵ_c is about 600 GeV in water and 500 GeV in rock. Above this energy, the loss is dominated by the stochastic radiative processes.

2.5 The Detector

The ANTARES detector (figure 2.5) will consist of a three-dimensional array of photomultipliers (PMTs, for details see [21]) to detect the Cherenkov light emitted by passing muons. The array will be realized by grouping the PMTs on strings that are anchored to the sea-bed and held up by buoys. The PMTs are housed in pressure resistant Optical Modules (OMs) (figure 2.6) of which three are arranged in a storey. The PMTs are oriented 45^o downward to maximize detector efficiency for upgoing muons. The storeys are separated vertically by a mechanically resistant electro-optical cable of 12 m in length. Each string will support 30 storeys and thus 90 PMTs, and will have a height of more than 400 meters. The detector will consist of at least 10 strings, totalling to 900 PMTs. The signals from the PMTs are digitized and time-stamped by a custom made integrated circuit, the Analog Ring Sampler (ARS). This chip is able to process complete waveforms or single photo-electron events; the latter returns the number of photo-electrons (NPE) in an integration window.
Data is transported through optical fibres in the cables to a junction box, from where the data is sent through a 40-km long cable to shore, where further processing is done. The data-taking

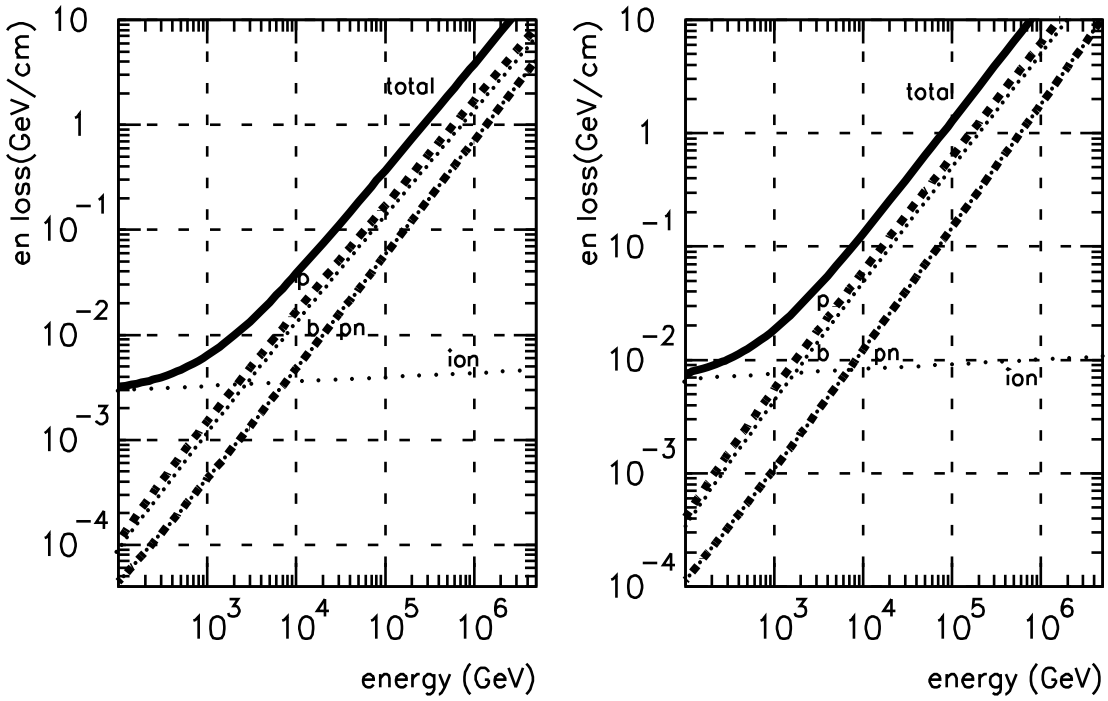

Figure 2.4: Total and partial energy losses for muons in water (left) and standard rock (right) as a function of muon energy. Total energy loss: solid line, Pair production (p): dashed line, Bremsstrahlung (b): dotted line, Photo-nuclear interaction (pn): dot-dashes, Ionization (ion): large dots. Picture taken from [20]

concept of ANTARES is "all-data-to-shore" [22], meaning that no data reduction is done by off-shore hardware triggers. Triggering is done on-shore by software allowing for easy adjustment of trigger conditions depending on circumstances. Such a circumstance could be the alert of the detection of a gamma-ray burst in the electro-magnetic spectrum.

A continuous rate of background light is present due to the β-decay of ^{40}K and bioluminescence. An average rate of 60 kHz is measured for the chosen PMT.
The ANTARES collaboration will build the detector at a depth of 2400 m in the Mediterranean Sea, 37 km off-shore of La-Seyne-sur-Mer, near Toulon (France). This location (42^o 50'N,6^o 10'E, figure 2.7), together with the Earth's rotation will give an annual sky coverage of 3.5π sr (figure

2.8). There will be an instantaneous overlap of about 0.5π sr with the coverage of the AMANDA detector located at the South-Pole which integrates to a common coverage of 1.5π sr. A potentially important source of high-energy neutrinos, the Galactic-Centre, will be visible for 67 % of the time. The reason for placing the detector as deep as possible is to reduce the background of down-going atmospheric muons (figure 2.9).

When referring to positions in the detector volume, I used the following conventions : positive x is North, positive y is West, positive z is up from the sea-bottom. The point $(x, y, z) = (0, 0, 0)$ is the center of gravity of the PMT positions. The angles used are the zenith angle $0 \leq \theta \leq \pi$ with respect to the z-axis and the azimuth angle $0 \leq \phi < 2\pi$ with respect to the x-axis.

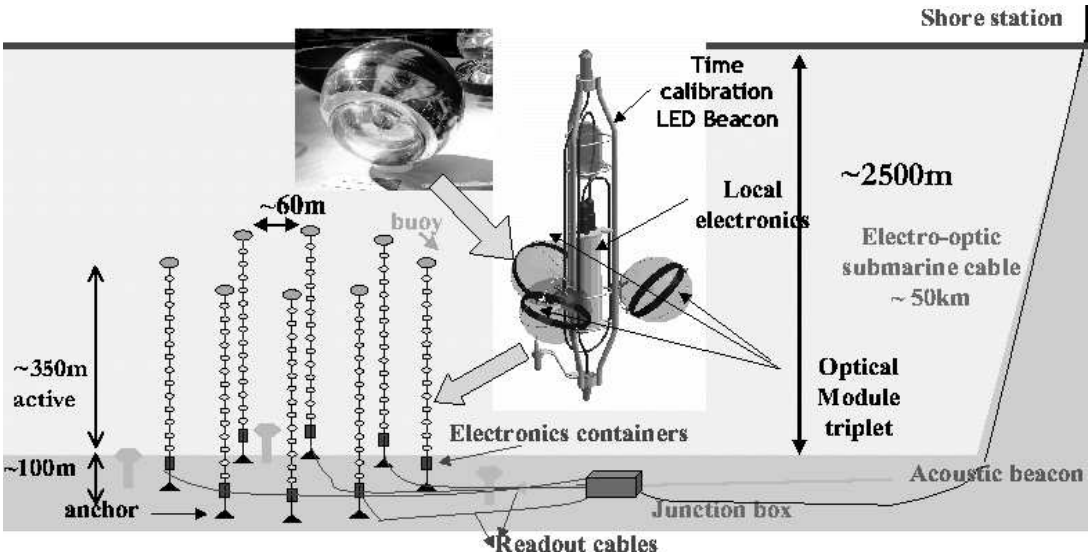

Figure 2.5: Schematic overview of the ANTARES detector and some of its components.

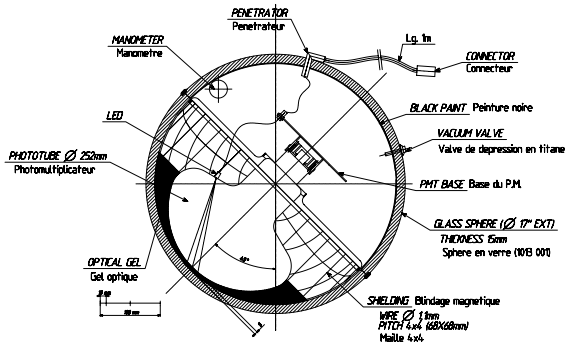

Figure 2.6: The Optical Module.

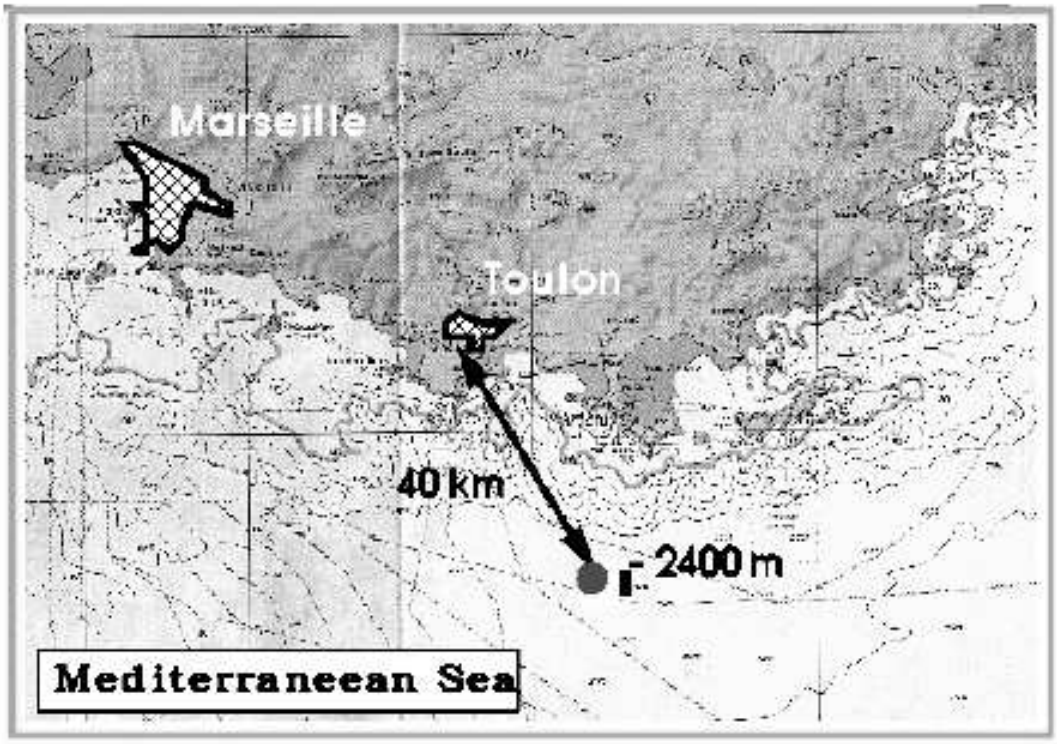

Figure 2.7: The location of the ANTARES detector.

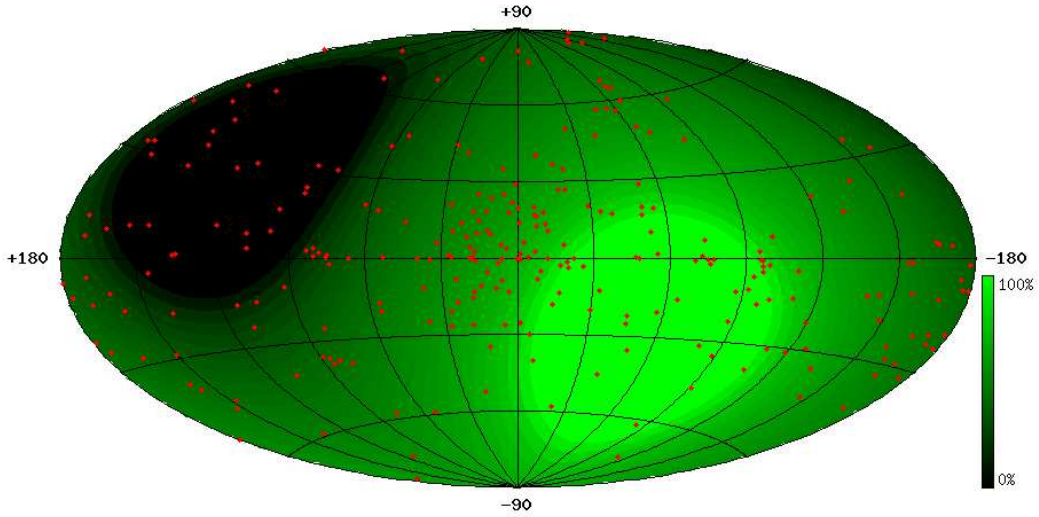

Figure 2.8: A sky-map for a detector at the ANTARES location. The map is in galactic coordinates and is an Aitoff projection. The lightest colour indicates the region which is 100 % visible per natural day and the black colour is a region of the sky that is not visible. The dots indicate objects from the third EGRET catalog [23]

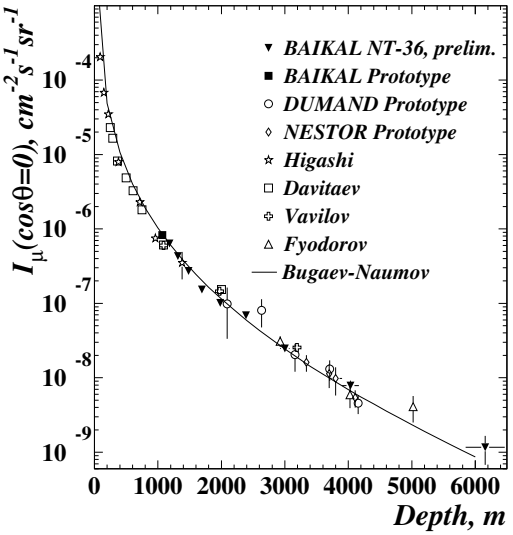

Figure 2.9: Flux of atmospheric muons as function of depth.

2.6 Reconstruction

The task of the reconstruction software is to find estimates for the parameters regarding position and direction of the muon track (x, y, z, θ, ϕ). This is done primarily by using the measured arrival times of Cherenkov photons at the PMTs (hits). The reconstruction algorithm I will describe shortly is developed by A. Heijboer and is referred to as the "Aart-Strategy". It is the algorithm I have used to reconstruct the direction and position of the muon tracks. Below you will find a short summary of the algorithm, as described in detail in [24].

Let $\vec{q}$ be the position of the hit at the PMT and $\vec{p}$ the position of the track at arbitrary time t_0. The track has a normalized direction $\vec{d}$. The vector pointing from $\vec{p}$ to $\vec{q}$ is $\vec{v}$. The component of $\vec{v}$ parallel to $\vec{d}$ is $l = \vec{v} \cdot \vec{d}$ and the perpendicular component is $k = \sqrt{\vec{v}^2 - l^2}$ (see figure 2.10). From this the theoretical arrival time of the light at the PMT can be derived :

$$t^{th} = t_0 + \frac{1}{c}(l - \frac{k}{\tan\theta_c}) + \frac{1}{v_g}(\frac{k}{\sin\theta_c}), \tag{2.15}$$

with v_g the group velocity of light in the sea water.
The reconstruction algorithm consists of four different procedures which are ran consecutively. Each of the procedures improves the results of its predecessor. The first procedure is a linear fit of a straight line through points that are associated with points on the track. This is effectively a least-squares technique. The points of the track are estimated from the likelihood that a track at a certain distance of a PMT will cause a hit of the measured amplitude. For the next two procedures the residual r_i is calculated for each hit as a function of the estimates of the track parameters $(\hat{\vec{p}},\hat{\vec{d}})$:

$$r_i = t_i - t_i^{th}(\hat{\vec{p}}, \hat{\vec{d}}) \tag{2.16}$$

The best estimates are given by the parameters which maximize Q, which is defined as:

$$Q = \sum_{i=1}^{N} g(r_i(\hat{\vec{p}}, \hat{\vec{d}})) \tag{2.17}$$

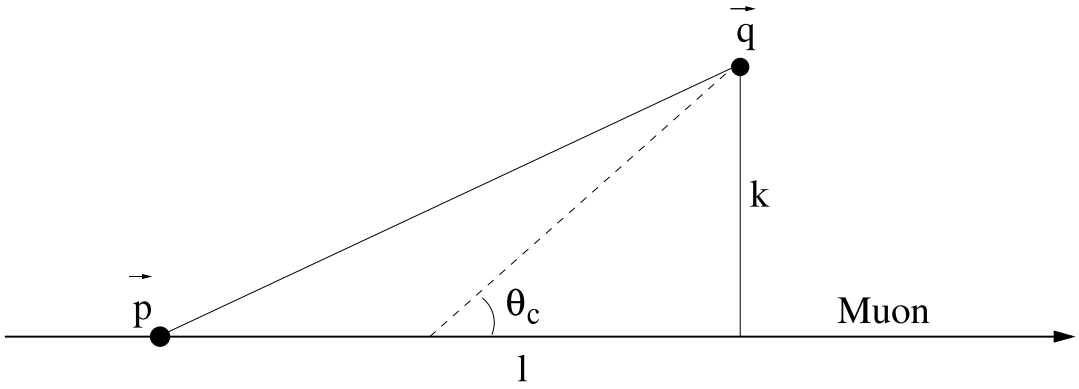

Figure 2.10: Geometrical relation between the muon track and point $\vec{q}$ where the hit is measured. Point $\vec{p}$ is the muon position at time t_0. k and l are the components of the vector $\vec{v} = \vec{q} - \vec{p}$ perpendicular and parallel to the track respectively. The light (dashed line) is emitted at the Cherenkov angle θ_c.

The two algorithms use different functions $g(r)$. Normally, $\ln L(r)$ is used, where $L(r)$ is the probability density function of finding a hit with residual r_i. This is done in the second of the two algorithms. The drawback of using this function is that it is difficult to find the global maximum as it is rather flat for large r. To overcome this problem the first of these two algorithms uses a function that makes it more easy to find a global maximum. This function only describes the data accurately at small values of the residual. The last stage of the parameter estimation uses the method of maximum likelihood. This method looks for track parameters which yields the highest probability of the observed data using the probability density function (PDF). If all hits are uncorrelated the PDF can be written as :

$$P(event|track) = \prod_i P(t_i|t_i^{th}, a_i, b_i, A_i) \tag{2.18}$$

where a_i is the cosine of the incident angle of the photon on the PMT, b_i the distance travelled by the photon and A_i the amplitude of the hit.
A quality cut is implemented in the "Aart-Strategy" based on the quantity $\log(L)/NDOF$ which is the value of the likelihood function at the fitted maximum divided by the number of degrees of freedom. The effect of this cut can be seen in figure 2.11 which shows the error on the angle of the reconstructed muon track. As can be seen it greatly reduces the error of the angular reconstruction.

2.7 Simulation

Currently the ANTARES detector is under construction and there is no data available yet. That is why simulations of the detector response and all of the involved physics are necessary. There are two types of simulations. One type is the simulation of the data-stream coming from the detector. This is used to develop and test the software systems which should handle the data-stream [25]. The other type of simulation involves the simulation of the physics processes and the response of the detector. In this simulation, data is produced that should resemble actual data coming from an operational detector. Subjects that can be studied with the simulations are :

- Optimisation and characterisation of parameters in the detector design. It is important to know how certain choices affect the performance of the detector. These include for ex-

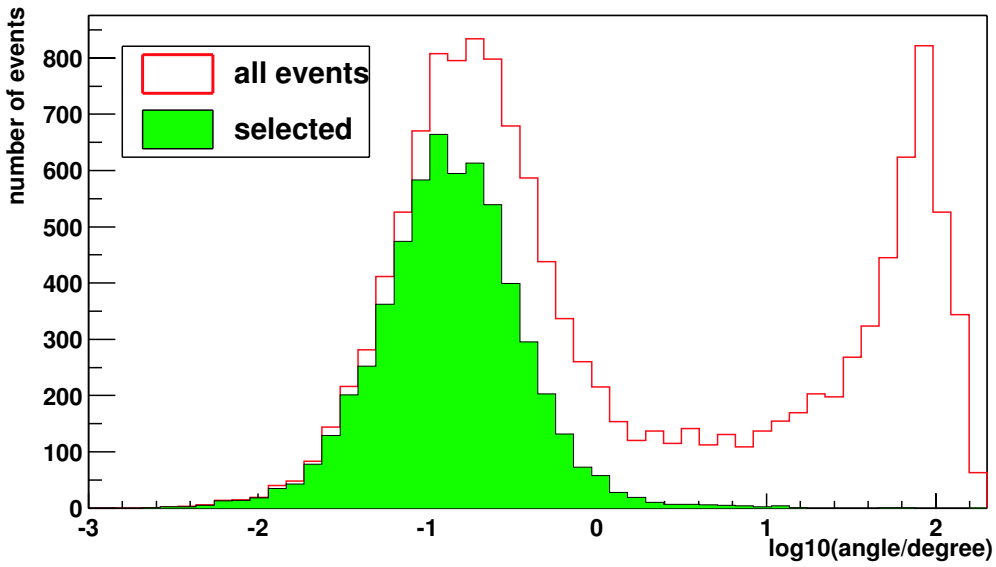

Figure 2.11: Error on the reconstructed angle of the muon track before (all events) and after quality cut (selected). These are up-going muons with an energy between 50 GeV and 100 TeV. Picture taken from [24].

ample the detector geometry, water properties (such as scattering and absorption length), photomultipliers and electronics.

- Detector performance. It is possible to characterize the ability of a given detector to measure signals from sources of different types. This way the effective detection area for neutrinos with a certain energy can be estimated. The part of the sky that is covered can also be determined.

- Development of reconstruction algorithms. Methods to determine the direction of a neutrino and to measure its energy can be tested using simulated data. The advantage of this is that one knows what the output should be. This way the angular resolution and the energy resolution of the detector can be tested and improved.

The usefulnes of the results that are obtained by using the simulations depends on the validity of the physics processes and detector response that is generated. Therefore, the output of the simulation software should be checked with theory and more importantly with measured data, which can originate from other experiments.

2.7.1 Events

In order to reduce the amount of processing time needed to generate events, the production of light of all the events is simulated in a *can* which surrounds the detector; this is illustrated in picture 2.12. The can extends to 2.5 times the maximum absorption length in water around the detector. With an absorption length of 55 m this is approximately 140 m. Outside of this can, the emission of Cherenkov light and light coming from EM-showers is not simulated.
Depending on the type of event being studied, muons are generated in different ways, e.g. :

- Single muons in the detector. In this case it is sufficient to generate muons which have their origin on the surface or in the volume of the can.

- Muons from neutrinos that are produced in point sources in the universe include interactions of the neutrinos with the Earth and the effect of movement of the Earth.
- Muons originating from atmospheric showers require a longer chain of simulation steps. This starts with the parametrisation of the cosmic ray primaries, followed by a simulation of the atmospheric shower and the tracking of the muons through rock and water to the can.
- Specialized topics such as neutrinos produced in neutralino annihilation in the Earth have their own requirements.

Several software packages have been developed by the ANTARES collaboration for the generation of events. Two of these packages have been used in this study. These packages are GENTRA [26] for generating single muon events and GENHEN [27] to produce muons that are induced by astrophysical neutrinos.

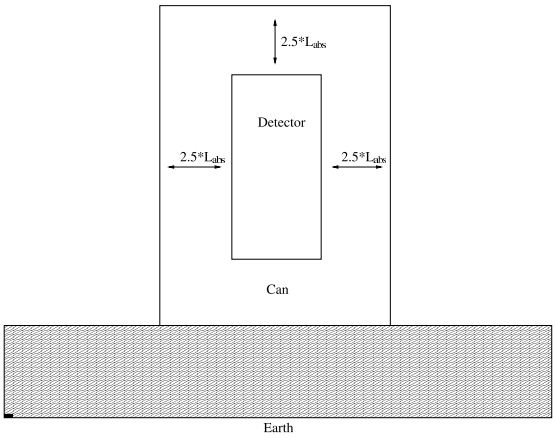

Figure 2.12: Schematic side-view of the detector and the volume (can) in which the simulations take place.

2.7.2 Detector Response

Finally the path of the muons and secondaries through the detector volume and the detector response to the emitted light is simulated. The large number of photons that are produced by a single muon and its secondaries passing through the detector makes it time consuming to perform a simulation in which every Cherenkov photon is generated and propagated individually. Because of the need for high statistics an alternative method is used. This method makes use of 'photon-tables' which tabulate PMT hit arrival time and amplitudes as function of the parameters of a track or an electro-magnetic shower. These tables are generated in a full Monte-Carlo simulation, which does perform the complete simulation including the photon tracking. The simulation takes the following into account:

- The *optical properties* of the sea-water are used to simulate the scattering and absorption of photons in the water.
- The *OM properties* including the quantum-efficiency, the attenuation factor depending on photon incident angle and the transmission through the glass of the pressure-sphere and the optical gel. They are important for the calculation of the hit amplitudes.

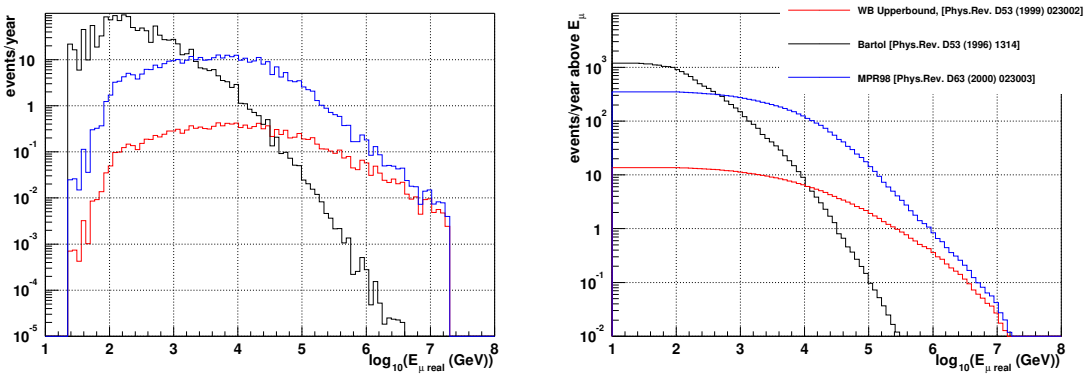

Figure 2.13: Figures showing the predicted measured flux (left) and integrated event rate (right) of neutrino induced muons in ANTARES for different models, assuming perfect energy reconstruction. The "Bartol" flux represents the background of muons induced by atmospheric neutrinos.

Another important part of the detector response is the simulation of the electronics. This is done by smearing the photon arrival times and the amplitudes. Also the functionality of the ARS is simulated by introducing an integration time of 25 ns during which hit amplitudes are summed, followed by a dead time of 250 ns in which no data is taken. To avoid a long dead time, there are two ARSs which alternatively take data.
Above I have described roughly the software package KM3 ([28],[29]) which was used in this study.

2.8 Expected Fluxes

In section 2.1, I have introduced some possible sources for high-energy neutrinos. The expected (diffuse) neutrino flux depends on the model. Figure 2.13 shows the spectrum of reconstructed neutrino induced muons for two different models (and background, see below) and the corresponding predicted event rates, taking into account the effective area and the efficiency of the ANTARES detector and the reconstruction software. Perfect energy reconstruction is assumed, thus the true energy is displayed. The models are upper-limits calculated using the cosmic ray flux by Waxmann and Bahcall (WB) [30] and Mannheim, Protheroe and Rachen (MPR) [31]. Also shown is the predicted background of high-energy atmospheric neutrinos [32]. These are formed in the decay of pions or kaons created by interactions of cosmic rays and cannot be distinguished from astrophysical neutrinos. The muons that are produced in the atmospheric cascades also constitute a background. This flux has a zenith-angle dependence (figure 2.14) and by placing the detector on the bottom of the sea, facing downwards, the background can be greatly reduced. As can be seen from figure 2.13, the background due to atmospheric neutrinos is large at low muon energies and small at high energies where the flux of astrophysical neutrinos is predicted to be higher. In section 5.3 I will discuss the effects of the energy reconstruction method that is developed in this thesis on the reconstruction of the fluxes and the calculated event rates.

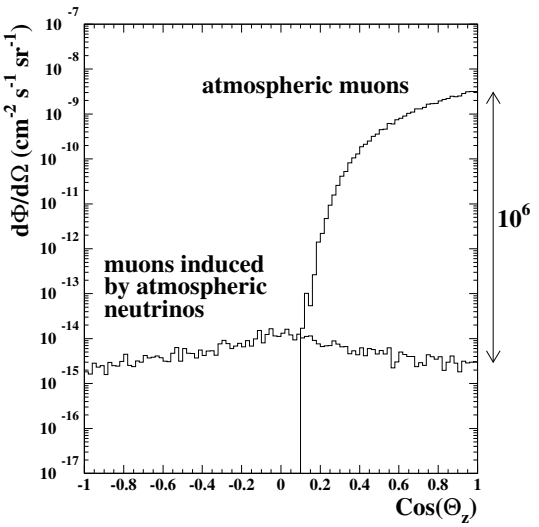

Figure 2.14: Background flux of muons against zenith angle. In this picture the zenith angle points to the direction from which the muons originate, opposed to the definition in section 2.5. (Figure taken from [5])

Chapter 3

Neural Networks

The use of artificial neural networks seems to have earned its place among the analytical tools of the experimental high-energy physicists. Some classical examples are their use in jet tagging [33] and invariant mass reconstruction [34]. Further examples include an analysis of top quark production in the D0 detector [35] and a possible use in the search for the Higgs boson at the Large Hadron Collider [36].
The use of neural networks has several advantages. The parallel structure of a network makes it inherently suitable for parallel distributed processing, speeding up calculation time. This way neural networks can be used for quick online data filtering. An advantage of a more conceptual nature is that neural networks are capable of solving problems for which the logical structure is not entirely clear. A network "learns from examples" and derives the relations between its input parameters and the output parameters "on its own".
This chapter describes the theory behind the type of neural network I used in my study, the *feed-forward layered network*, and the algorithm I used to train the networks (i.e. teaching them the relations between the parameters), the *backpropagation* algorithm. The books by Bishop [37] and Müller, Reinhardt and Strickland [38] provide more detailed information. The first section of this chapter is an introduction to the biology on which the neural network theory is based.

3.1 Biology

As the human and animal central nervous system is the main source of inspiration for the field of neural computation, I will give a short introduction to its function. A more in-depth discussion can be found in [38].
However complex and capable of a broad range of tasks, the cells which make up the central nervous system all have the same basic plan. Except, of course, the cells that form the supporting structures. A neural cell is called a neuron (figure 3.1). The main part is the cell body, or soma, which typically has a size of $10 - 80$ μm. From this several extensions reach out, the dendrites, and one tubular fibre, the axon, which can split into several branches. Both dendrites and axons typically have a diameter of a few μm. The size of neurons varies between 0.1 mm (inter-neurons in the human brain) and 1 m (neurons in the limbs). While the dendrites mainly function as receptors for signals from other neurons, the axon transmits outgoing signals. The end of an axionic branch connects to another neuron at a dendrite or the soma. The joint is called a synapse. At the synapse, the two cells are separated by a mere 200 nm, which permits us to speak of the pre-synaptic (axonic) side and post-synaptic sides.
Transmission of signals between neurons, at the synapse, is chemical and the transmission within a neuron, along the axon, is electrical. The electrical signal originates from a discharge that starts on the soma and travels along the axon. The interior of the neuron, the protoplasm, is negatively charged with respect to the surrounding neural liquid. The potential difference is about 70 mV and is supported by active pumping of the cell membrane and its impermeability to

Na^+ ions, which causes a deficiency of positive ions in the protoplasm. Chemical substances called neurotransmitters, that are released from the pre-synapse, alter the permeability of the cell membrane, which causes a transient weakening or depolarization of the resting potential. When at some point the membrane potential drops below a certain threshold, usually about -60 mV, the membrane looses its impermeability to Na^+ and depolarises completely. The discharge starts out from the cell-body and propagates along the axon. Because the discharge in each previous section is complete, the pulse moves in one direction and does not diminish in strength. When the pulse reaches the synapses at the end of the axon, it in turn can release neurotransmitters. Signals can be encoded in the modulation of the pulsing frequency. Different types of neurotransmitters have different post-synaptic effects, they can even polarize the cell. Not every threshold is the same. Learning is basically the changing of connections between neurons, their coming and going and their change in strength, defined as their ability to polarize or depolarize the post-synaptic neuron. This and more leads to the complex signal processing of which the central nervous system is capable and enables us to walk, think, listen to music and play go.

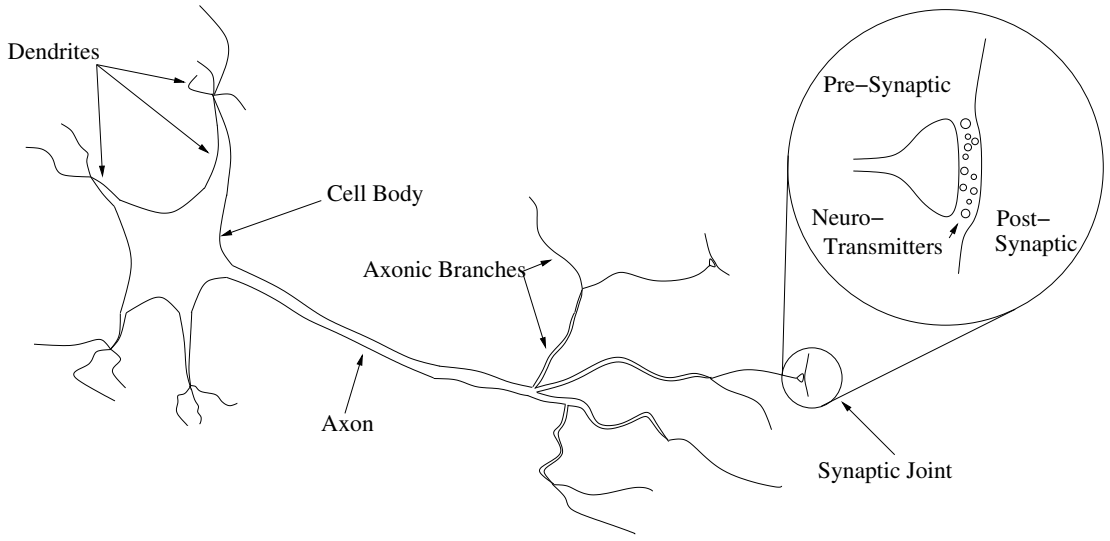

Figure 3.1: Schematic drawing of a neuron.

3.2 Mathematics

The mathematical abstraction of a neural network is defined as (from [38]) *a directed graph with the following properties* :

1. A state variable n_i is associated with each node i.
2. A real-valued weight w_{ik} is associated with each link (ik) between nodes i and k.
3. A real-valued bias θ_i is associated with each node i.
4. A transfer function $f_i[n_k, w_{ik}, \theta_i, (k \neq i)]$ is defined, for each node i, which determines the state of the node as a function of its bias, of the weights of its incoming links, and of the states of the nodes connected to it by these links.

3.2.1 Feed Forward Layered Networks

In the following I will restrict myself to one subclass of neural-networks, the feed-forward layered network. In this type of network the nodes are grouped in layers and information flows in

one direction, see figure 3.2. This means that the connection weights w_{ik} are maximally asymmetric. At one end there is the *input* layer which feeds the next layer. The next layer can be one or more *hidden* layers or directly the *output* layer. In mathematical terms the input-output relation defines a *mapping* and the network provides a *representation* of this mapping.

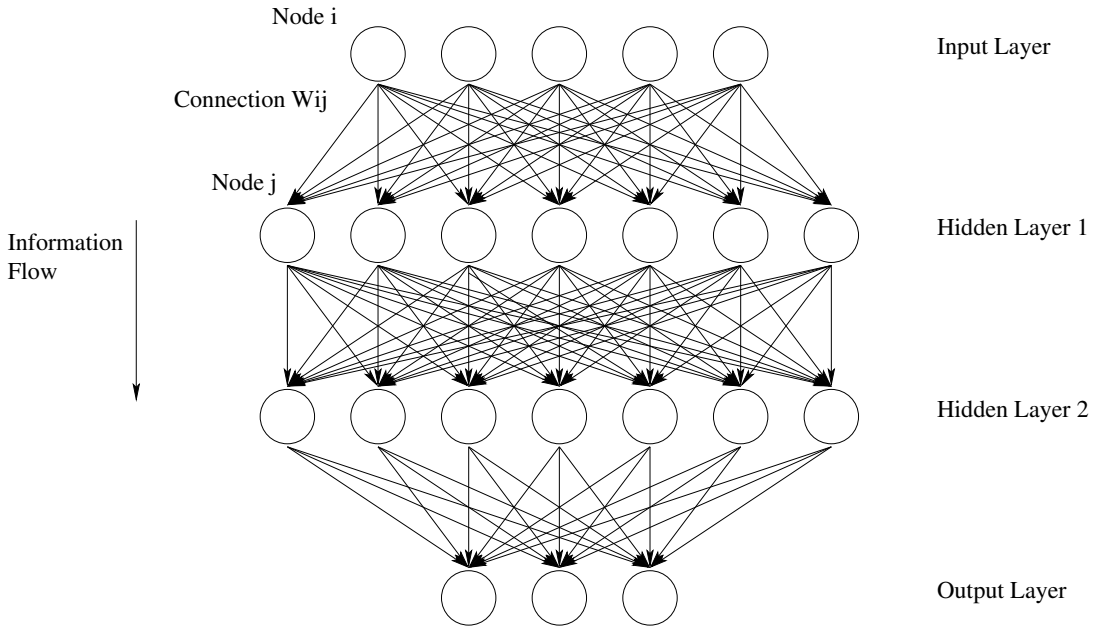

Figure 3.2: A feed forward network with two hidden layers.

How is the state or output of an individual node (neuron) calculated? It is almost equivalent to the biological case. First the sum of the outputs of the nodes that feed into it multiplied by their respective connection weights is calculated. Note that in the case of a feed-forward layered network these nodes are the ones in the previous layer. After subtraction of the bias, the sum is fed into the transfer function of the node in question and this results in the output of the node. If one takes as bias the weight of an extra node that always has unit output, the output of node j can be written as :

$$z_j = g_j(a_j) \tag{3.1}$$

with

$$a_j = \sum_{i=1}^{d} w_{ji} z_i + w_{j0} = \sum_{i=0}^{d} w_{ji} z_i \tag{3.2}$$

Where d is the number of nodes in the previous layer, z_i the activation of node i, w_{ji} $(i > 0)$ the connection weight with node i, g_j the activation function belonging to node j and w_{j0} the bias $(z_0 = 1)$.

3.2.2 Network Topology and Activation Functions

The choice for a network topology, i.e. the number of layers and nodes, and the choice for the activation functions depend on the task that the network has to perform.
The activation function commonly used for the hidden nodes in the network is the logistic sigmoidal function :

$$g(a) = \frac{1}{1 + \exp(-a)} \tag{3.3}$$

The choice for the activation function for the output nodes depends on the task at hand. The logistic sigmoidal function maps the interval $(-\infty, \infty)$ onto $(0, 1)$. Although this is not suitable for functional mapping, the function has properties that make it a good choice for classification problems ([37]). For functional mapping, the linear activation function is used :

$$g(a) = a \tag{3.4}$$

With one hidden layer of nodes and the choices stated above, any multi-variable function can be approximated to arbitrary accuracy (for proof see [37]). The error scales with $O(1/M)$, where M is the number of hidden nodes. Although one hidden layer is enough, the use of more tends to improve the efficiency of the network [34].

3.2.3 Training

Training of a neural network is basically minimizing an error-function by adjusting the free parameters in the model. The free parameters in a neural network are the connection strengths in the form of the matrix w_{ji} and the biases θ_i. The biases can be replaced by the connection strengths w_{j0} of a node that fires with unit strength.
For the training one uses a training set which contains n training patterns. A training pattern consists of the values of the input nodes $\overline{x}^n$ and the target output $\overline{t}^n$. Each pattern is propagated through the network as described above, see equation 3.1 and further. Generally this will yield an output $\overline{y}^n$ which differs from $\overline{t}^n$.
We suppose that the network error can be expressed as a differentiable function of the output variables so that

$$E^n = E^n(y_1, ..., y_c) \tag{3.5}$$

for c output nodes.
The total error function is the sum of the error functions of the separate patterns :

$$E = \sum_n E^n \tag{3.6}$$

An error function typically used is the sum-of-squares error which, for a single input vector n, is :

$$E^n = \frac{1}{2} \sum_{k=1}^{c} (y_k - t_k)^2 \tag{3.7}$$

3.2.4 Backpropagation

There are numerous methods to minimize the error function. The method described in the following is the error backpropagation algorithm.
In the following I will omit the superscript n for input and activation variables. The error function is minimized by using the derivatives of E^n with respect to the weights w_{ji}. Note that E^n depends on the weight w_{ji} only via the summed input a_j to node j. Using the chain rule for partial derivatives yields

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \tag{3.8}$$

we define

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} \tag{3.9}$$

Using (3.2) we can write

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \tag{3.10}$$

Substituting (3.9) and (3.10) into (3.8) we obtain

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j z_i \tag{3.11}$$

The values of z_i are known by the propagation of the pattern through the network. In order to get the derivative it is only necessary to calculate δ_j for each hidden node and each output node in the network. For the output nodes this is straightforward.

$$\delta_k \equiv \frac{\partial E^n}{\partial a_k} = g^{'}(a_k)\frac{\partial E^n}{\partial y_k} \tag{3.12}$$

where z_k from equation (3.1) is denoted by the output activation y_k. Expressions for $g^{'}(a)$ and $\frac{\partial E^n}{\partial y_k}$ can, for example, be found from equations (3.3), (3.4) and (3.7).
To obtain the δ's for hidden nodes we use the chain rule

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k}\frac{\partial a_k}{\partial a_j} \tag{3.13}$$

The sum runs over all nodes k to which node j has output connections. Using definition (3.9) and formulae (3.1) and (3.2) we get the *back-propagation* formula

$$\delta_j = g^{'}(a_j)\sum_k w_{kj}\delta_k \tag{3.14}$$

By applying this formula recursively we can obtain all the δ's for the hidden nodes and input nodes by propagating the error backwards through the network. The weights are updated in the following way

$$\Delta w_{ji} = -\mu\frac{\partial E}{\partial w_{ji}} \tag{3.15}$$

where μ is the step size and typically has a value between 0.1 and 1. The gradients are then recalculated with the new values of w_{ji} over many iterations and the error converges to the smallest possible value. This searching for the minimum in weight space has several problems. The local gradient does not always point to the minimum and the search could end in a local minimum. A method to overcome this problem is to add a momentum term to the gradient descent formula. This effectively adds inertia to the motion through weight space. It diminishes oscillations and helps to overcome local minima. The modified update formula is

$$\Delta w_{ji}(t+1) = -\mu\frac{\partial E}{\partial w_{ji}} + \eta\Delta w_{ji}(t) \tag{3.16}$$

Where η is called the momentum parameter.

3.3 Data Samples

To train and test a neural network, several data samples are needed. A data sample consists of *patterns*. A pattern consists of data that are the network input and intended output variables.

- The *training set* is used for the actual training of the network. The error on this set is minimized during the training process.
- The *validation set* is needed to monitor the performance of the net on patterns that are not used during training. This is to verify the ability of the network to generalize and to avoid *overfitting* (see figure 3.3) on the patterns from the training set.
- A *test set* is used after training to measure the performance of the network.

It is important that the training set is different from the test and validation sets. The last two can be the same.

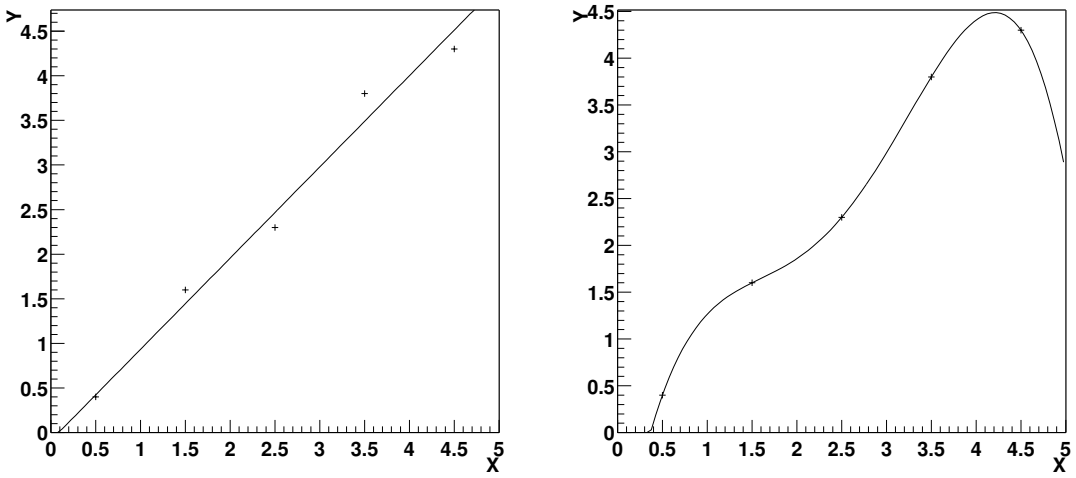

Figure 3.3: Example of overfitting with polynomials. The data points have a linear relation and are smeared. Left shows a linear fit. Right is a fit with a 4th order polynomial. One sees that the 4th order polynomial, though it fits perfectly to the data, does not generalize it correctly. The training process of a neural network should be monitored as it can overtrain on the data set. This shows up as an increase in the error of the validation set, while the error of the training set keeps decreasing. The number of patterns in a training set should be larger than the number of free parameters in the network.

Chapter 4

Energy Reconstruction

The reconstruction of energy is an important part of neutrino astronomy, comparable to the observation of different wavelenghts of light in optical astronomy. It can teach us about the inner workings of the sources which produce neutrinos among other particles. Models of astrophysical objects such as Active Galactic Nuclei can be tested as different models predict different fluxes of neutrinos. Questions on acceleration mechanisms of cosmic rays can be answered.
An accurate energy reconstruction can also aid in reducing the background of atmospheric neutrinos. As shown in section 2.8 this background dominates at low energies. To exclude contamination of the data by atmospheric neutrinos, a cut can be applied on the reconstructed energy of the muon. The energy cut depends on the quality of the reconstruction. The better the resolution of the energy reconstruction, the closer the cut can be placed at the ideal value, determined assuming perfect reconstruction.
When determining the direction of neutrinos in, for example, a search for point sources in the sky, the relation between the direction of the muon and the direction that the neutrino had plays a role. Formula 2.11 shows the relation between the angles of the muon and the neutrino at high energies: a small error on the reconstructed energy leads to a small error on the direction of the neutrino and the search for point sources can be more accurate.

In this chapter I will first introduce the conventional way of reconstructing the energy of the muon in ANTARES. Then, the chapter will be devoted to the energy reconstruction method I developed using the technique of neural networks. Parameters significant to the energy determination are introduced. Then the data samples and the networks will be described. After some remarks on the training of the networks, a section devoted to the results will follow. In this section the performance of the network and its dependence on some parameters will be discussed.

4.1 Conventional Strategy

The conventional energy reconstruction algorithm in ANTARES is the one developed by F. Hubaut [6] and reviewed by A. Oppelt [39]. The algorithm takes into account the increased energy loss of the muon with energies above 600 GeV. Above this energy the radiation losses due to Bremsstrahlung, pair production and nuclear processes become dominant over the energy loss due to ionization (see section 2.4). From formula 2.14 we can deduce for the amplitudes a_i registered by a PMT:

$$a_i \propto (a_0)_i \cdot (1 + \frac{E_\mu}{\epsilon_c}) \tag{4.1}$$

$(a_0)_i$ are amplitudes caused by a minimum ionizing particle and ϵ_c is the critical energy (section 2.4). The energy loss increases with the muon energy, and so does the amplitude ratio. The

excess brilliance that photomodule i registers is :

$$\left(\frac{a_i}{(a_0)_i} - 1\right) \propto \frac{E_\mu}{\epsilon_c} \tag{4.2}$$

In order to reduce the statistical fluctuation and retain as much information as possible, it is necessary to average the quantities a_i and $(a_0)_i$ over the hits which are in the Cherenkov cone. The estimator $x \propto E$ of the energy becomes :

$$x \propto N_{hits} \cdot \left(\frac{\sum_i a_i}{\sum_i (a_0)_i} - 1\right) \tag{4.3}$$

The multiplication with the number of hits, N_{hits}, improves the resolution.
A linear fit is performed on Monte-Carlo data to relate the estimator to the true energy :

$$\log_{10}(E) = a + b \cdot \log_{10}(x) \tag{4.4}$$

Two fits are made, one for the high-energy regime and one for the low-energy regime. The distinction between the two regimes is made by calculating the fractions of hits which fall in one of the categories :

Low Energy : $0.1 < \frac{a_i}{(a_0)_i} < 100$

High Energy : $10 < \frac{a_i}{(a_0)_i} < 1000$

If the low-energy criterium rejects less than 90 % of the hits, the low-energy estimator is used, otherwise the high energy estimator is used. This choice is motivated in figure 4.1 which shows the fraction $a_i/(a_0)_i$ for various energies.
Furthermore, the hits have to be compatible with the hypothesis of a Cherenkov photon originating directly from the reconstructed track. This requirement is implemented by cutting on photons that arrive with a differencte greater than 4 ns with the expected arrival time of a direct photon:

$$|t_i - (t_0)_i| < 4\text{ns} \tag{4.5}$$

with t_i the photon-arrival time and $(t_0)_i$ the expected arrival time of a Cherenkov photon. The motivation for this choice can be seen in figure 4.2, which shows a non-Gaussian tail in photon arrival times caused by secondary particles.

4.2 Extra Parameters

In the method described above, the distribution of hits in time is not used. When speaking of time, the time is relative to the expected arrival times assuming the Cherenkov hypothesis unless stated otherwise. With increasing energy, more signal hits on the PMTs are caused by non-Cherenkov photons[1]. Figure 4.3 shows the ratio of the number of photoelectrons induced in the PMT by pure Cherenkov photons over the total number of photoelectrons (Cherenkov and EM shower electrons) as function of the energy of the incoming muon. The ratio approaches one at low energies, since in the MIP regime approximately all induced photoelectrons originate from Cherenkov photons. At high energies the ratio drops strongly and approaches zero because of the dominating role of photons from the EM shower. The effect of this on the registered photoelectrons can be seen in figure 4.4. It illustrates the increasing photon yield with energy. One can also see the increasing tail of photoelectrons registered at later times. This tail is caused by the fact that photons from EM showers in general do not arrive at the PMTs at the time calculated for the Cherenkov photons. The relative growth of the tail of photoelectrons can be seen

[1]The photons induced by e.g. an EM shower *do* result from the Cherenkov effect of relativistic electrons. I use the term Cherenkov photons for photons directly originating from the muon track.

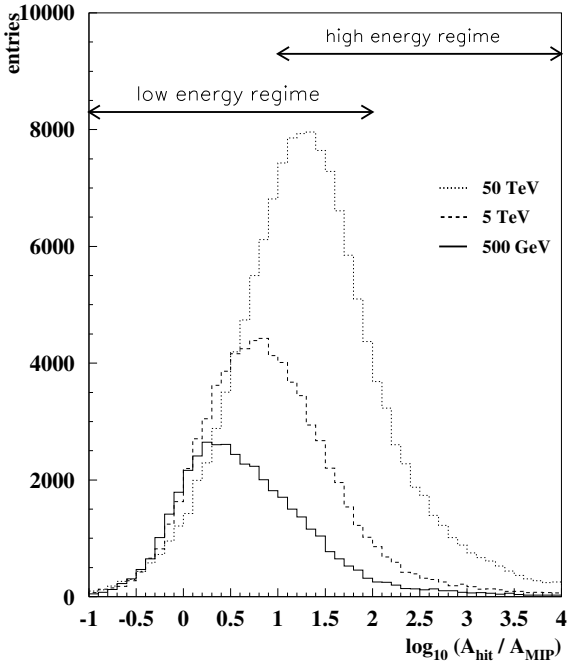

Figure 4.1: $\frac{a_i}{(a_0)_i}$ distributions for different muon energies. The high and low energy regimes are indicated. Picture taken from [39]

in figure 4.2. This way additional information about the muon energy can be obtained from the distribution of photoelectrons with time.
I used this information in the following way. Two unequal sized intervals in time were defined. The first time interval contained the integrated photoelectrons associated with the peak of Cherenkov photons. Only hits with $|t_i - (t_0)_i| < T_1$ ns were allowed. The other time interval $T_1 < t_i - (t_0)_i < T_2$ ns contained the integral of photoelectrons in the tail of the distribution. Figure 4.5 shows these integrated values for both intervals as function of energy. Values for T_1 and T_2 were set at 15 ns and 250 ns. In figure 4.2 one can see that a smaller value for T_1 could be chosen as 15 ns is already past the beginning of the tail. However, the drawback of a smaller T_1 is that the integrated value of photoelectrons becomes smaller and due to the imprecision of a software package that is used ([40]) systematic errors are introduced, especially for muons in the lower energy part of the spectrum. The interval with $T_1 = 15$ ns and $T_2 = 250$ ns contains a reasonable fraction of the tail of the distribution and roughly corresponds 1.5 times the attenuation lenght of the photons. In figure 4.5 the values of both parameters are shown together with their ratio (interval1/interval2) which increases with energy.
The number of photons that arrive at a time close to the predicted Cherenkov time increases more rapidly than the number of photons that arrive later. This is illustrated in figure 4.4. The reason for this is that 63% of the EM shower is contained in a small cylindrical volume around the muon track. Above 400 MeV the radius of the cylinder remains constant at a value of 3.2 cm and the majority of the particles is directed parallel to the direction of the muon [41]. Therefore, the Cherenkov photons from the parallel secondary particles arrive reasonably in time.

Another measure of the muon energy is the number of PMTs that are hit in both time intervals. This is illustrated in figure 4.6. The dependence on the total number of photons is through the increase of the probability that a PMT will produce a signal. As the tail in the time distribution of hits from individual photons increases more rapidly than the Cherenkov peak, figure

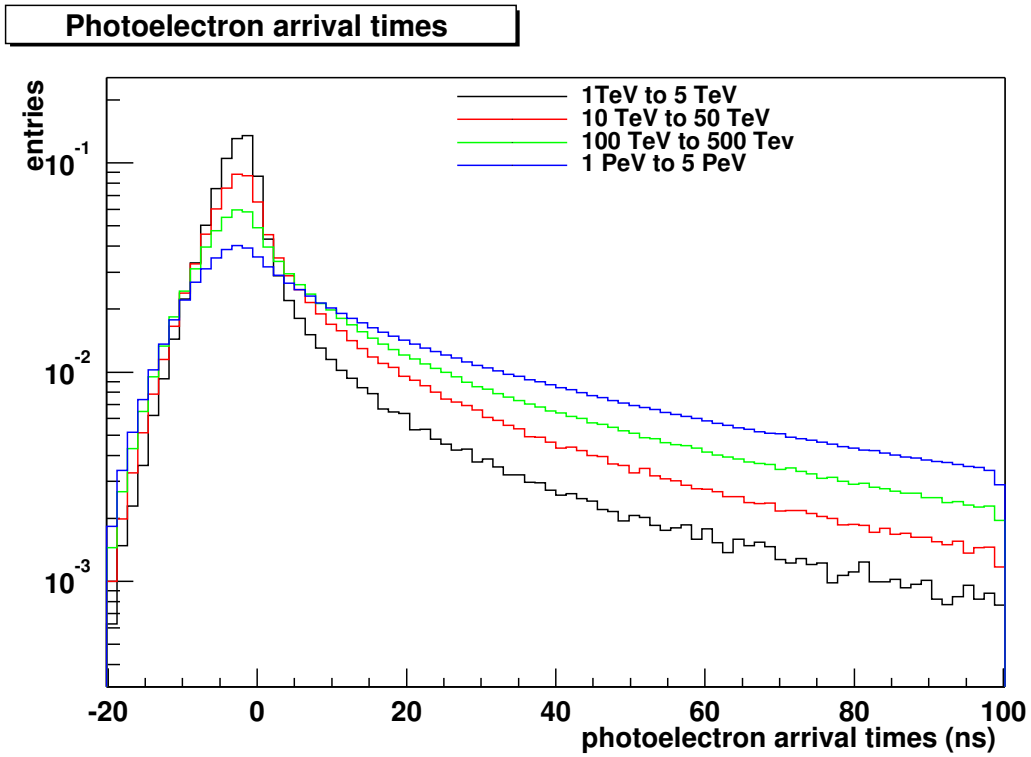

Figure 4.2: Difference between Cherenkov arrival time and real arrival time $(t_i - (t_0)_i)$ for several muon energy ranges.

4.2, the number of PMTs that are hit at a later time increases relatively more than the number of PMTs that are hit in time.

The variables introduced above, i.e. the integrated numbers of photoelectrons and PMT hits in the two time intervals, carry information on the dependence of photon arrival time on the energy. The integrated numbers of photoelectrons contain extra information about the actual number of emitted photons. To investigate the degree of redundancy of these four variables their dependence on each other should be checked. For different muon energies these variables or derived quantities should populate different regions of parameter space. This is visualized in figure 4.7 where a scatterplot can be seen of the ratio of PMTs hit against the ratio of the integrated number of photoelectrons. As can be seen in this plot the different ranges of muon energy represent different populations. The populations overlap for low values of these derived quantities and this could affect the performance of the energy reconstruction. The properties of the muons in these regions have yet to be uncovered.

Even an extremely energetic muon can look like a MIP muon in terms of light detected when it passes the detector at large distance. It is obvious that the position and angle of the track are relevant to the reconstruction of the muon energy. To account for this, the closest distance of the muon to the center of the detector, defined by the center of gravity of the PMTs, has been calculated for each of the three coordinates x,y,z. Figure 4.8 shows the number of photoelectrons from upgoing muons ($0 \leq \cos\theta \leq 1$) averaged over all energies in the range 100 GeV to 100 PeV as a function of the distance to the center of the detector. All distributions fall off at what are approximately the edges of the detector. In the x and y distributions the bumps at the positions of detector strings are caused by the fact that muons pass very closely to the PMTs and cause a relatively high photon yield.

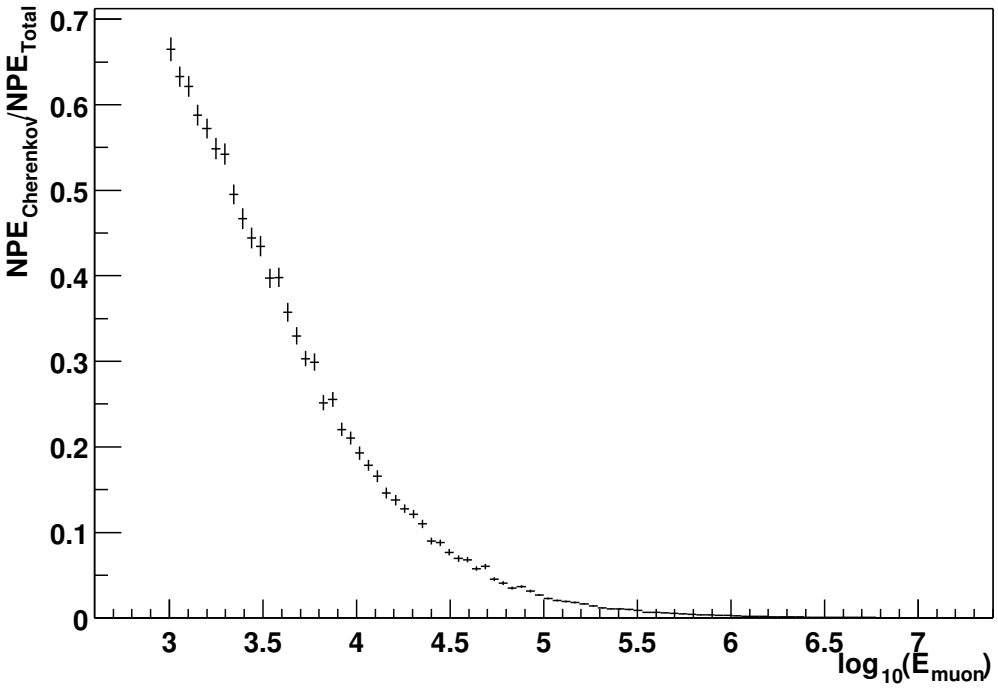

Figure 4.3: The fraction of Cherenkov photoelectrons as function of muon energy. The contribution of ^{40}K has been filtered out.

The ANTARES detector is optimised to detect upgoing muons, e.g. the PMTs "look" downward at an angle of 45^o. This means that for a muon of a certain energy the amount of light that is collected depends also on its direction. Especially the projection of the direction on the z-axis makes a difference, as can be seen in figure 4.9. As the cosine of the angle of the track with the z-axis approaches zero, more light is seen. Although the direction of the track can be described by two parameters (zenith and azimuth angle) I chose to use the three projections of the normalized directions on the x, y and z axes. This is done to simplify the dependence of the light yield on the parameters.

4.3 Network Implementation

4.3.1 Parameters

In the previous section I introduced the variables which I used for the energy reconstruction. Although these variables can be combined into other variables to parametrise the energy of the muon, I choose not to do so. The motivation for this choice is that when the variables are combined into other variables, I restrict the network to use only these and no other possible combinations.
Here follows a summary of the used input parameters :

- NPE_{avg} Average number of photoelectrons. This is the average of all amplitudes of the photoelectrons which are induced within 250 ns of the expected arrival time of the Cherenkov photons.
- NPE_{15} Total number of photoelectrons within 15 ns of the expected arrival time of the

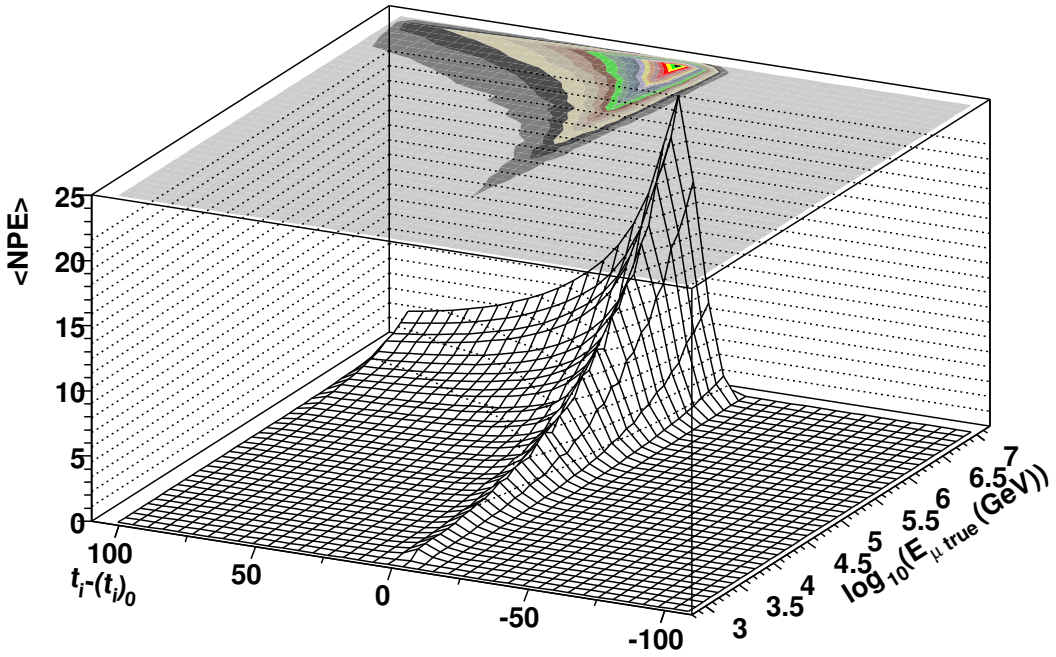

Figure 4.4: Number of photoelectrons as function of arrival time and muon energy (see text).

Cherenkov photons. This should contain the largest fraction of the Cherenkov photons while accounting for scattering.

- NPE_{250} Total number of photoelectrons between 15 ns and 250 ns after the expected arrival time of the Cherenkov photons. This should contain the largest fraction of EM shower photons.
- OM_{15} Total number of PMTs hit within 15 ns of the expected arrival time of the Cherenkov photons.
- OM_{250} Total number of PMTs hit between 15 ns and 250 ns after the expected arrival time of the Cherenkov photons.
- x, y, z The coordinates of the point of closest approach of the muon track to the center of gravity of all the PMTs.
- cx, cy, cz Direction cosines of the muon track.

This gives a total of 11 input parameters. The only output parameter is the energy of the muon track.

4.3.2 Transformation

In order to avoid large absolute gradients and overtraining on one parameter the inputs were transformed in such a way that most values are between 0 and 1. The cosines were not scaled, as their values are inherently suitable, between -1 and 1. The transformations used are the following :

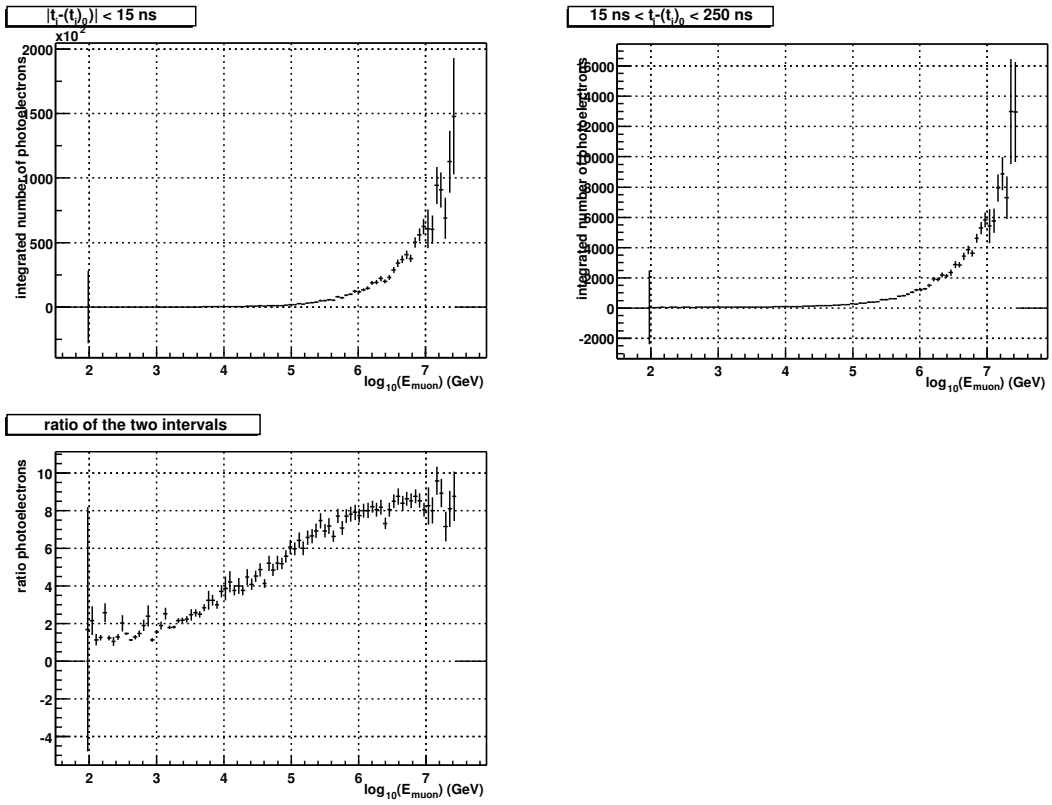

Figure 4.5: The number of photoelectrons in the two time intervals and their ratio as function of muon energy.

- For parameters x, which are restricted to positive values :

$$x^{'} = \frac{2}{\pi} \cdot \arctan\left(\frac{x}{2 \cdot \overline{x}}\right) \tag{4.6}$$

- For parameters x, which can have positive and negative values, the means $\overline{x}_+$ and $\overline{x}_-$ were calculated for the positive and negative values respectively. Then :

$$x^{'} = \frac{1}{\pi} \cdot \arctan\left(\frac{x}{2 \cdot (\overline{x}_+ + |\overline{x}_-|)}\right) + 0.5 \tag{4.7}$$

For the output node, which returns the energy of the muon, a different transformation has been used:

$$e_{node} = \frac{1}{10} \log_{10}(E) \tag{4.8}$$

where E is the energy of the muon and e_{node} is the output of the node.

4.3.3 Data Samples

To train, monitor and test a neural network, different data samples are needed (see section 3.3). The data sample that was used for training covered the full muon energy range on which the network has to operate. As can be seen in figure 2.13 the muon energy spectrum that the ANTARES detector can measure extends to 10^7 GeV, assuming the Bahcall and Waxman upper-limit for the

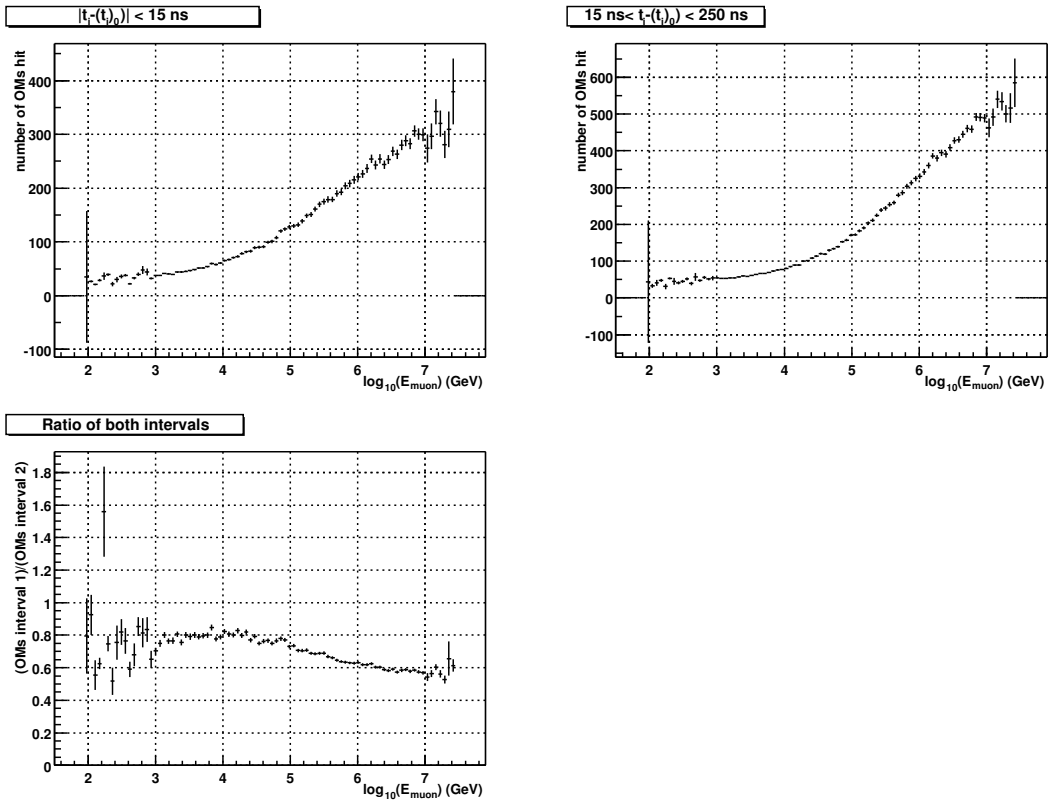

Figure 4.6: The number of PMTs hit in the two time intervals and their ratio as a function of the muon energy.

cosmological neutrino flux. Below 10^3 GeV the neural network method for the energy reconstruction is expected to perform less well because this is the MIP regime and the emitted light is dominated by the Cherenkov photons. I have focussed my efforts for reconstruction of the muon energy in the region between 10^3 and 10^7 GeV.
The network returns the $\log_{10}$ of the muon energy. In order to have a comparable number of events in each decade of energy, a spectrum is needed that is flat in $\log_{10}$, meaning $\frac{\delta N_{muon}}{\delta E} \propto E^{-1}$ To achieve this, I had to take into account the trigger efficiency and the quality cuts on the likelihood of the reconstructed angle and position. Their combined effect is strongest for the lower energy muons. Figure 4.11 shows the effect of the trigger and quality selection on a sample that was generated with a flat energy distribution: relatively more low energy events are removed from the sample. This problem was overcome by generating events using a steeper energy spectrum so that there were more muons generated with a lower energy. When the muons were generated with an energy distribution with a spectral index of 1.17 the effect of the trigger and quality cut are greatly reduced. As a consequence, there were more muons in the range from 10 TeV to 1 PeV. Two additional sets of muons were added to the training sample. The first set contained muons with an energy between 100 GeV and 1 TeV, the other contained muons with energies higher than 10 PeV. This was done to avoid clipping of the network output between $\frac{\log_{10}(10^3)}{10} = 0.3$ and $\frac{\log_{10}(10^7)}{10} = 0.7$ as otherwise the network is only presented with values in this range. The energy spectrum of the training sample can be seen in figure 4.12. The validation and test samples were both generated with an energy spectrum with a spectral index of 1.17 and energies between 1 TeV and 10 PeV.

Below you will find a summary of the characteristics of the samples :

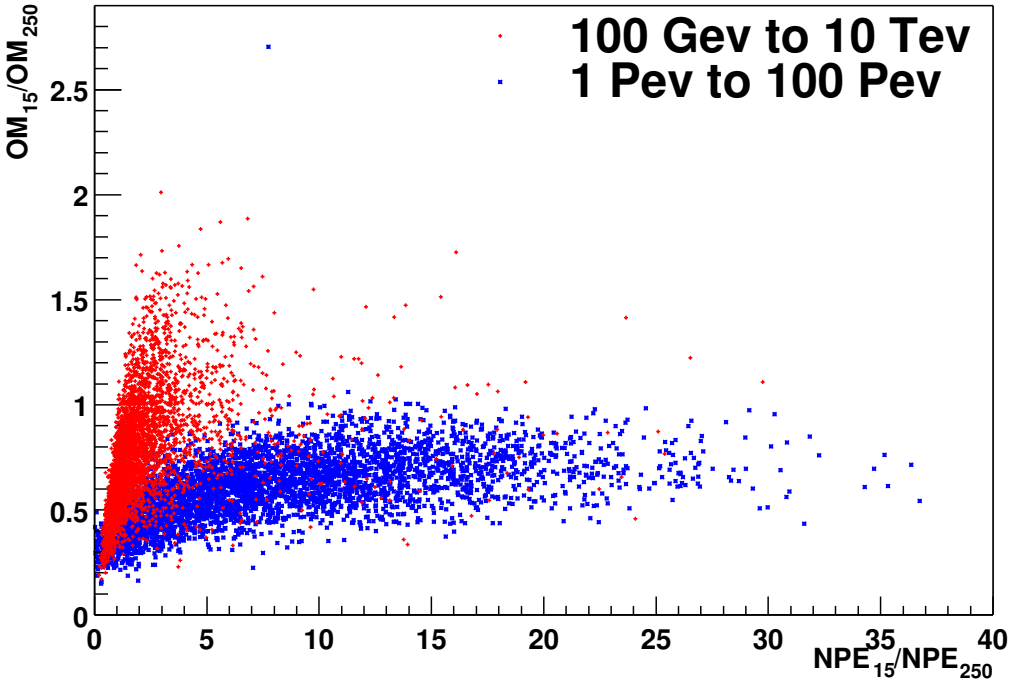

Figure 4.7: Ratio of the number of PMTs hit in two different time intervals versus the ratio of the number of photoelectrons in the same intervals.

Data Set	# events
Training set	18092
Validation set	7289
Test set	14630

Table 4.1: Number of muon tracks in the data samples.

- GENTRA ([26]) was used to generate the muons on the can. The angular distribution was flat and covered the lower 2π of the detector, meaning that all muons are upgoing.
- KM3 ([28],[29]) performed the muon propagation and detector simulation. The simulation included scattering of the light, and the electronics were simulated by assuming 2 ARSs per PMT which operate independently with an integration time of 25 ns and a dead time of 250 ns (see section 2.7.2).
- Reconstruction was done with the "Aart-Strategy" (section 2.6) with the default trigger conditions and only the reconstructed muons that passed the quality cut were used.
- The sizes of the samples can be seen in table 4.1.

4.3.4 Networks

It is not known a-priori which network topology will give the best results (section 3.2.2). A total of 10 different networks were trained. Seven with 1 hidden layer and three with 2 hidden layers. Table 4.2 shows the network topologies in more detail. All networks are fully connected,

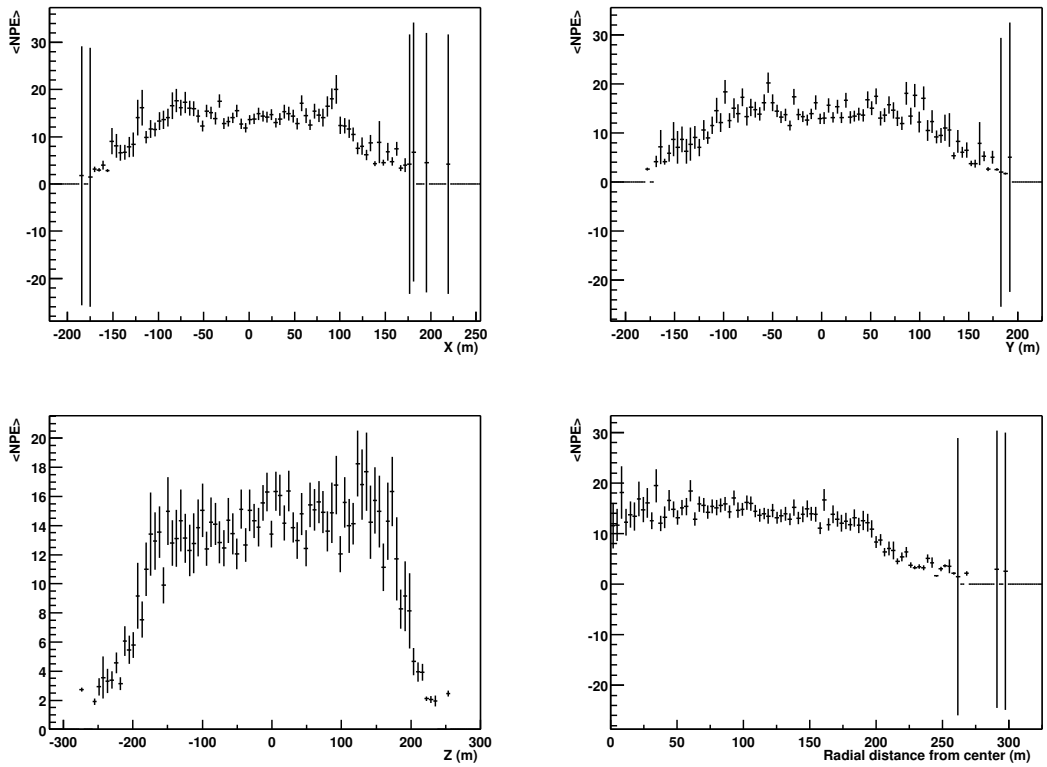

Figure 4.8: Average number of photoelectrons against closest approach of the muon from the center of the detector.

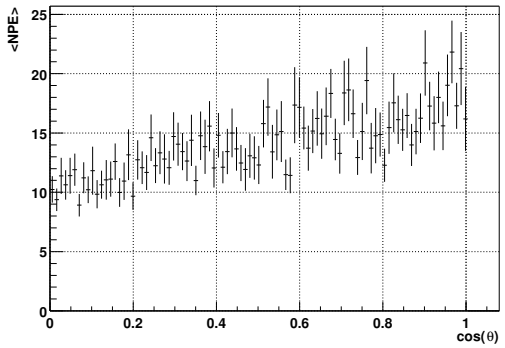

Figure 4.9: Average number of photoelectrons versus cosine of the zenith angle of the muon direction.

meaning that each node in a layer connects to every node in the next layer. The two-layer networks (8,9,10) have approximately the same number of adjustable parameters as the one-layer networks 1,5 and 7 respectively. The sigmoid activation function 3.3 was used for the hidden nodes and the output node had the linear activation function 3.4.

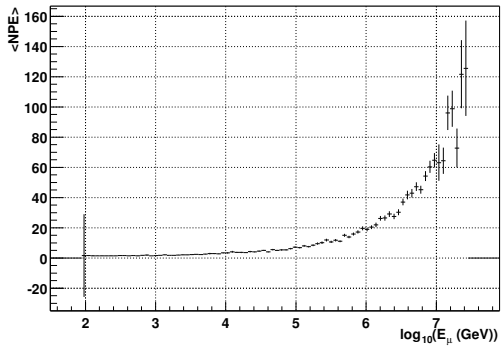

Figure 4.10: Average number of photoelectrons against muon energy.

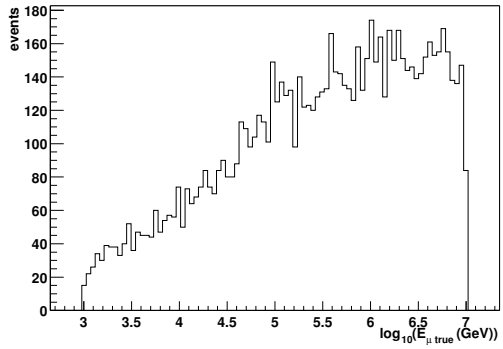

Figure 4.11: Spectrum of a sample of muons that was generated with a flat spectrum after reconstruction and quality selection.

4.3.5 Training

Training, or the adjustment of the weights of the links between the nodes in the networks, was done using the backpropagation algorithm with momentum term (section 3.2.4) with $\mu = 0.2$ and $\eta = 0.5$. Higher values of especially μ led to an increase in oscillations of the sum-of-squares error of the network, which was the network error function. The training was divided in epochs. During an epoch all patterns in the training set are presented sequentially and for each pattern the weights were adjusted. Unless stated otherwise, each training session comprised 2000 epochs. During the training the network error of both the training set and the validation set were monitored. For each epoch the total error was plotted to verify whether the error function became constant and thus a minimum in weight space was reached. Another function of the error of the validation set was to make sure that the network was not over-training on the training set. This is visible by an increase of the error on the validation set while the error on the training set keeps decreasing. Two typical learning curves, which show the network error against the number of epochs can be seen in figures 4.13 and 4.14 for two different networks using the flat training set. The black line is the error of the training set and the red line is the error on the validation set. The offset between the two is due to the different number of patterns in the samples. A common feature of the learning curves of the networks is the plummeting of the error in the first one or two epochs, which unfortunately only can be seen through the steep gradient

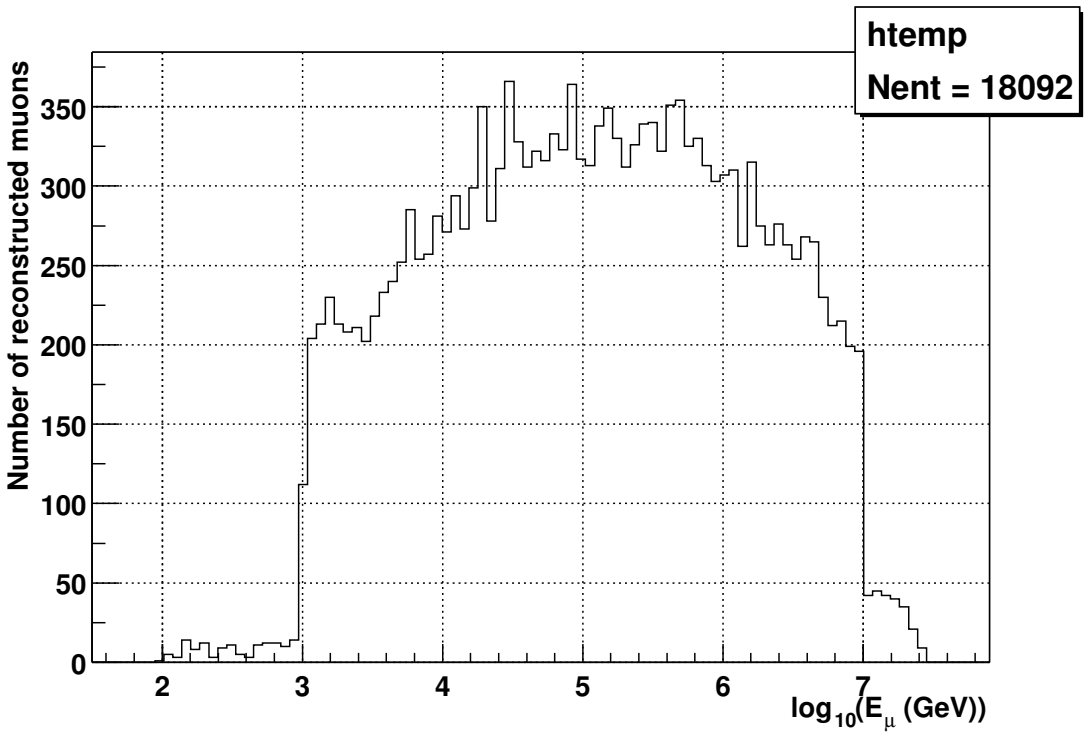

Figure 4.12: The energy spectrum of the muons in the training sample

the lines begin with. The error before training depends on the initialisation of the network by randomizing the weights of the network with a flat distribution between -1 and 1. Although the values vary enormously for the different networks (observed values : network 3, up to $4 \cdot 10^5$; network 7, up to $2.1 \cdot 10^6$;network 9,up to $2 \cdot 10^5$;network 10, up to $3 \cdot 10^5$) repeated observations with different initialization showed that within the first couple of epochs these differences were nihilized. A contribution to these large errors comes from the scaling of the output values and the limited energy range, which limits the output to values typically between 0.3 and 0.7. A network with linear output has a much broader range of output values after initialization. The steep descent of the error in the first few epochs indicates a quick adaptation of the network to these limited output values. All of the networks showed this steep initial descent of the error. In the following decrease of the gradient networks 7 and 8 (for topologies see table 4.2) showed a stepwise behaviour, indicating that the networks had reached local minima and got out of them, see figure 4.14. Although this effect could best be seen in these two networks, it was not uncommon for other networks to show sudden changes in the gradient. The effect was smaller for the validation set error, which decreased more smoothly. Network 4, with 30 hidden nodes, still showed a steep gradient after 2000 epochs and was trained for another 3000 epochs.

4.4 Results

In this section the performance of the trained networks are presentend. First some quantities that were used will be introduced, then the correlation between the reconstructed and simulated energy is shown. The behaviour of the error of the reconstruction will be discussed. This will be followed by a comparison between the different network topologies and the motivation for the

Table 4.2: Network topologies

Network	# units hidden layer 1	# units hidden layer 2
1	11	0
2	15	0
3	20	0
4	30	0
5	57	0
6	75	0
7	246	0
8	7	7
9	20	20
10	50	50

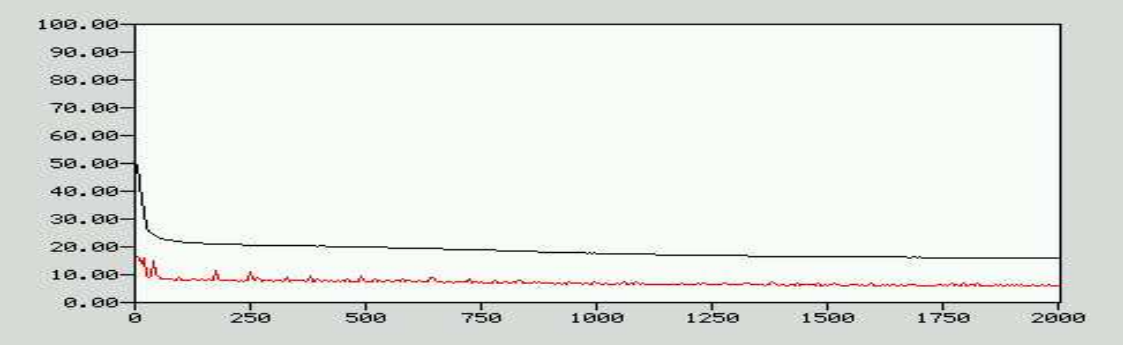

Figure 4.13: Learning curve of the 11/20/1 network (network 3 in table 4.2).

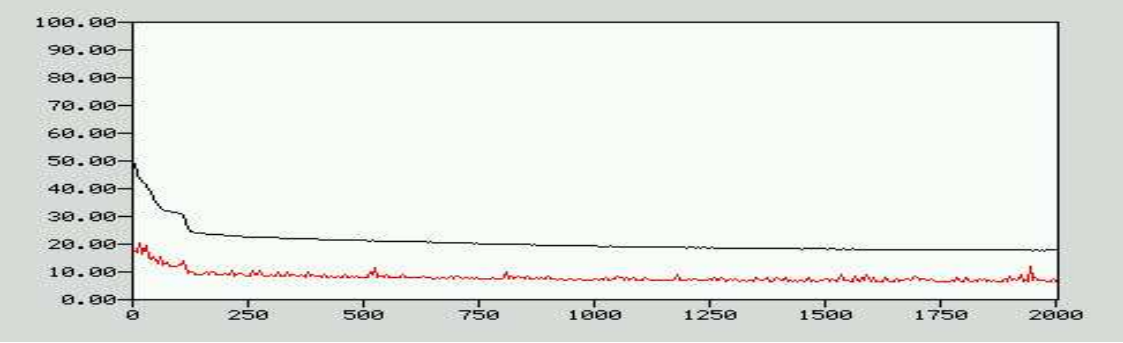

Figure 4.14: Learning curve of the 11/7/7/1 network (network 8 in table 4.2).

choice of networks used in the analyses presented in the following chapter. The dependence of the quality of the reconstruction on the geometrical parameters will be discussed.

4.4.1 Measurement of the Performance

As the output of the network is linear with the 10-base logarithm of the energy (see section 4.3.2) I used the following quantity R to measure the accuracy of the energy reconstruction :

$$R = \log_{10}(E_{network}) - \log_{10}(E_{true}) = \log_{10}(\frac{E_{network}}{E_{true}}) \tag{4.9}$$

with $E_{network}$ the energy as reconstructed by the neural network and E_{true} the true Monte-Carlo energy. The quality of the reconstruction method is measured by the mean μ and more importantly the standard deviation σ of the R distribution. The mean indicates a bias towards either overestimating or underestimating the energy, which are effects that can be corrected for, if necessary. The standard deviation indicates the expected error of the reconstruction of a single muon event, thus the resolution of the reconstruction.

4.4.2 Correlation and Error

Figure 4.15 shows the results of the energy reconstruction for three of the networks: 3,4 and 9. It also shows the means of the distribution per true energy bin and for reference the line $\log_{10}(E_{network}) = \log_{10}(E_{real})$. One can see clearly the correlation between the reconstructed energy and the generated (true) energy. Network 4 (one hidden layer with 30 nodes) shows a bias towards underestimating the energy over almost the full range of muon energies. The correlation extends to the full range of energies. At lower energies the gradient is smaller than at higher energies. This is as expected since at lower energies, when reaching the MIP regime, the energy loss of the muon becomes constant, as do the light yield and the different ratios of the variables such as the integrated number of photoelectrons in the different time bins. By extrapolating the relation between the input variables which are learned for all energies and are true especially for the higher energy range, the network overestimates the energies in the lower energy part of the spectrum. The increase in bias and the standard deviation can be attributed to this effect. Another reason for this, which also applies to the underestimation that occurs at very high energies, is in the data sample used to train the networks: there are very few events below 1 TeV and above 10 PeV, the network has difficulty extrapolating below and above these values respectively and asymmetric errors are introduced in the low and high energy range. This can be solved by adding sufficient events with energies below 1 TeV and above 10 PeV to the training sample. The behaviour of the input variables for events with energies below 1 TeV differs from higher energies. The networks should be able to learn the behaviour of the parameters for these low energies and the reconstruction method can be applied to energies below 1 TeV.

To take a better look at the performance of the networks, the mean and the standard deviation are calculated by plotting the distribution of R in several energy ranges and fitting a Gaussian curve. To determine the energy intervals, the reconstructed value of the energy is used, as this is the only available quantity when dealing with data. Hence, the resolution of the reconstruction as function of reconstructed energy is determined. Figures 4.16 and 4.17 show the results of the fit in four energy ranges. The distributions are not truly Gaussian, as can be seen from the ratio χ^2/ndf of the fits, which correspond to very small χ^2 probabilities [42]. All the other networks show the same pattern, which is that the fit is best in the lowest energy bin and that the distributions are skewed towards lower values with respect to a normal distribution. The same effect occurs when using the conventional method to reconstruct the energy, see figure 5 in [43]. Further comparison between the two methods follows in chapter 5.

For networks 5 and 9 of table 4.2 the fitted mean and standard deviation are shown in figure 4.18 as function of the reconstructed energies. The results for the other networks can be found in appendix A. The effect of the topology of the networks can be seen in figures 4.22 and 4.23.
Several observations can be made :

- The performance of the networks measured in the mean and standard deviation of R does not vary much for the one layer networks, except for the networks which have relatively few (networks 1 and 2, figure A.1) or large numbers (network 10, figure A.2) hidden nodes.
- No clear pattern can be found in the variation of the mean for the one-layer networks.
- The two-layer networks have a smaller spread and absolute value of the mean than the one-layer networks.

- On average the networks perform better with increasing energy.
- The two-layer networks perform equally or better than the corresponding one-layer network. This is best seen in the pairs 1,8 and 7,10. The latter pair is an extreme example.
- For most networks the resolution of the energy reconstruction, measured in the standard deviation of R, varies between 0.3 and 0.2 which correspond to factors 1.6 and 2.0.
- Network 10 performs worst of all networks in terms of mean and standard deviation of R.

4.4.3 Topologies

Now, which network performs best? Networks 3,4,5,6,8,9 and 10 are similar in terms of standard deviation. The one layer networks have a larger bias. Network 9, which has two hidden layers of each 20 nodes, has both low values for the bias and standard deviation and thus would be a good choice. If CPU processing is an important issue in the choice of a network, the one layer networks are best choice. With decreasing number of hidden nodes these networks have increasing biases towards under- or overestimating the energy. The standard deviation in the reconstruction of especially higher energies increases beyond 30 nodes, while this effect on the lower energy reconstruction is only present for larger numbers of hidden units. If the biases would be corrected by multiplying the reconstructed energy with an appropriate factor a network with 20 to 40 hidden nodes in one layer could be a good choice. Further analyses of the networks will be focussed on the two layer network with 20 units in each layer, network 9.

4.4.4 Dependence on Geometrical Parameters

In this section the dependence of the energy reconstruction on the geometrical parameters is studied. This is presented for network 9, an 11/20/20/1 network[2] with network. The other networks show the same behaviour. The distance from the center of the detector and the azimuth and zenith angles are studied.

Dependence on distance

Figure 4.19 shows the mean and the standard deviation of the $\log_{10}(E_{network}/E_{true})$ distribution as a function of the distance of the muon to the center of the detector. At small distances (< 20 m) the reconstructed muon energy is slightly overestimated but $\log_{10}(E_{network}/E_{true})$ is never larger than 0.05 which is a 12 % overestimate. At distances between 175 and 225 m the reconstructed muon energy is slightly underestimated but again not more than 12 %. At moderate distances the quality of energy reconstruction is good. The last bin, at about 245 m shows anomalous behaviour due to small statistics.

Dependence on ϕ and θ

Figure 4.20 shows that there is no dependence of the quality of the reconstruction on the azimuthal angle. Maybe, the asymmetric shape of the detector can be seen in a slight modulation of the mean around zero, but this effect is not present in the distribution of the standard deviation.
There is also no dependence of the quality of the energy reconstruction on the zenith angle (figure 4.21).

The resolution and the bias of the energy reconstruction do not show any clear dependence on the geometrical parameters except for a small decrease in resolution with increasing distance

[2]The terminology "11/20/20/1 network" indicates a network with 11 input nodes, 2 hidden layers with 20 nodes each and 1 output node.

from the detector. This means that the network has learned for example the depence of the light yield on the geometrical parameters and compensates for it.

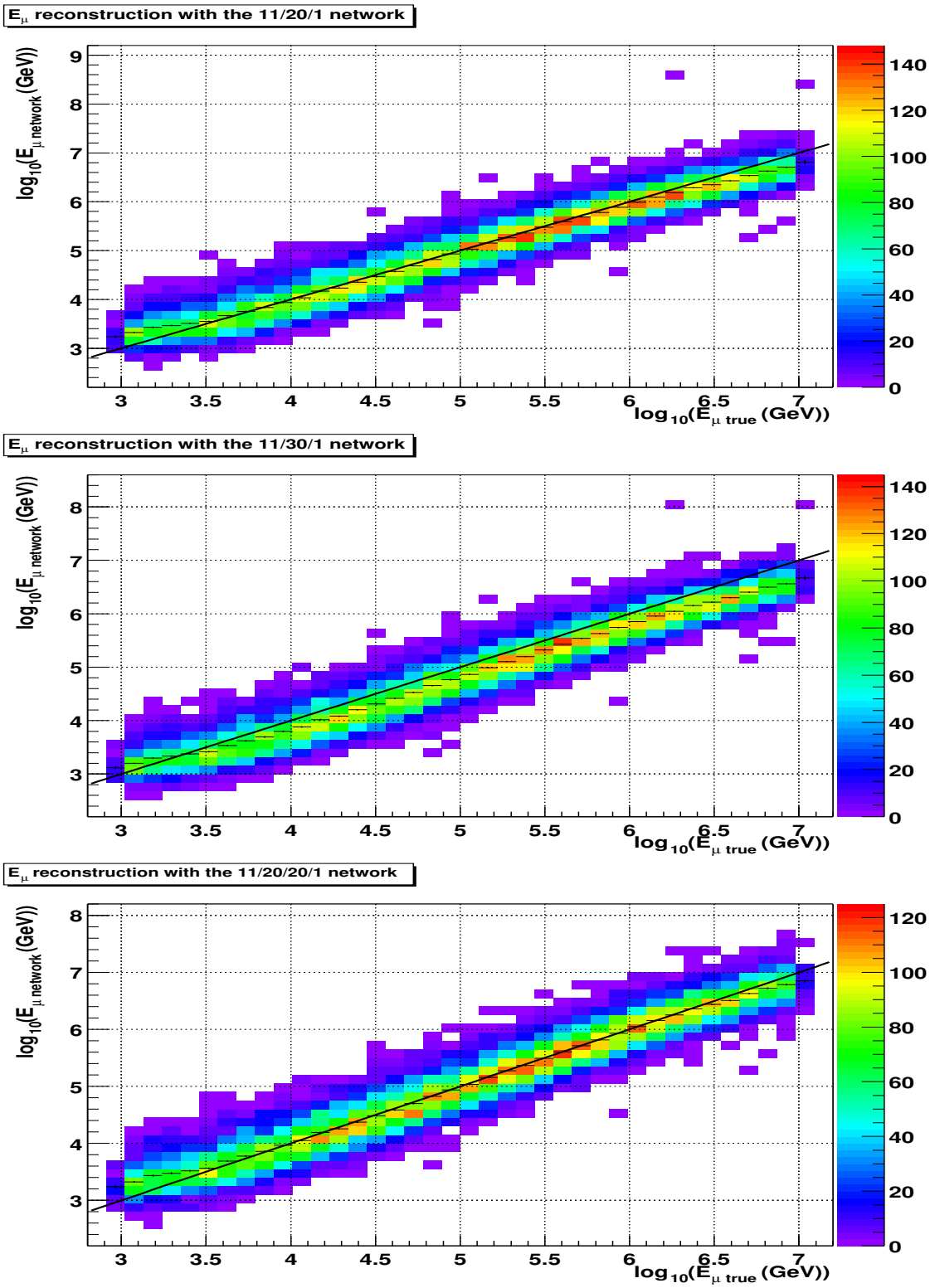

Figure 4.15: Correlation between reconstructed energy and true energy. The average values of the reconstructed energy are shown as is the line $\log_{10}(E_{\mu network}) = \log_{10}(E_{\mu true})$ for reference. Shown are networks 3 (top), 4 (middle) and 9 (bottom)

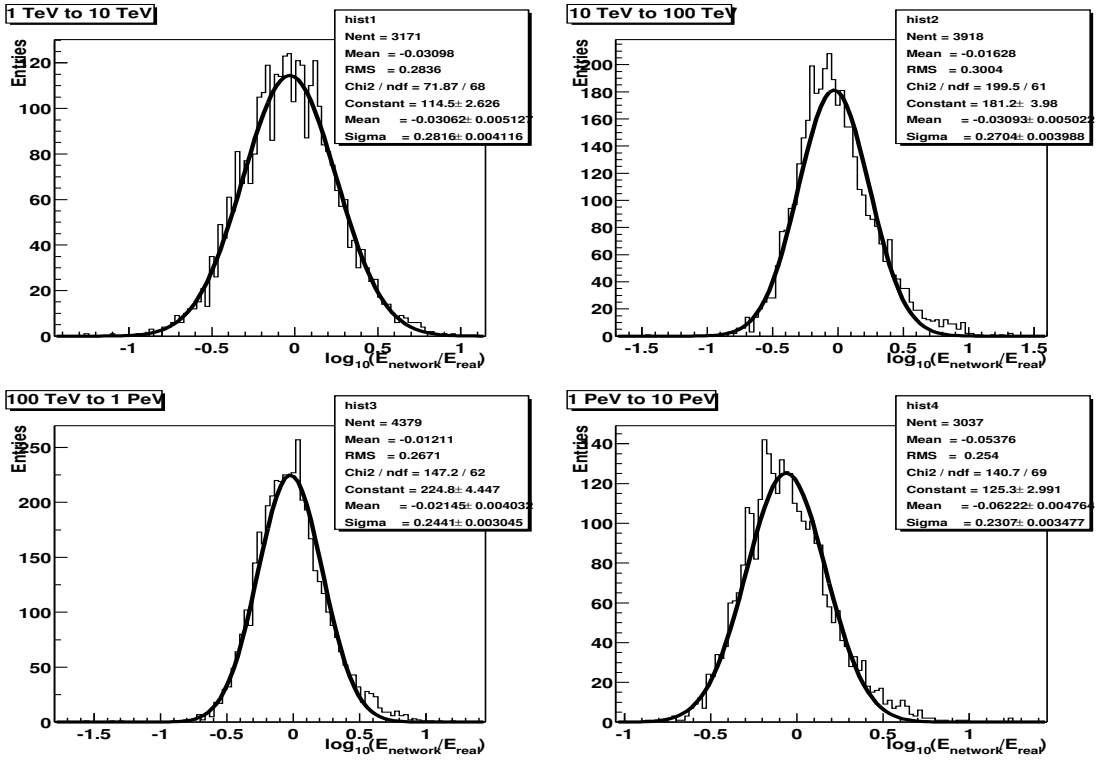

Figure 4.16: Performance of the 11/20/1 network (network 3). The curve is a fit to a Gaussian curve.

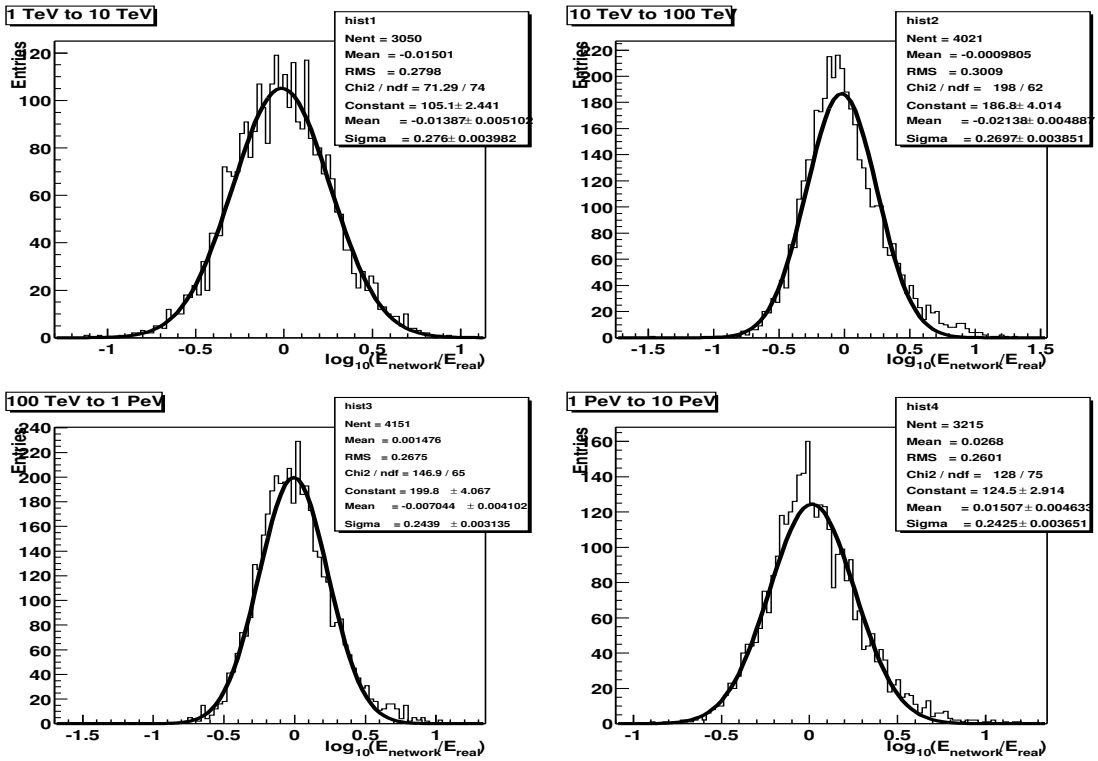

Figure 4.17: Performance of the 11/20/20/1 network (network 9). The curve is a fit to a Gaussian curve.

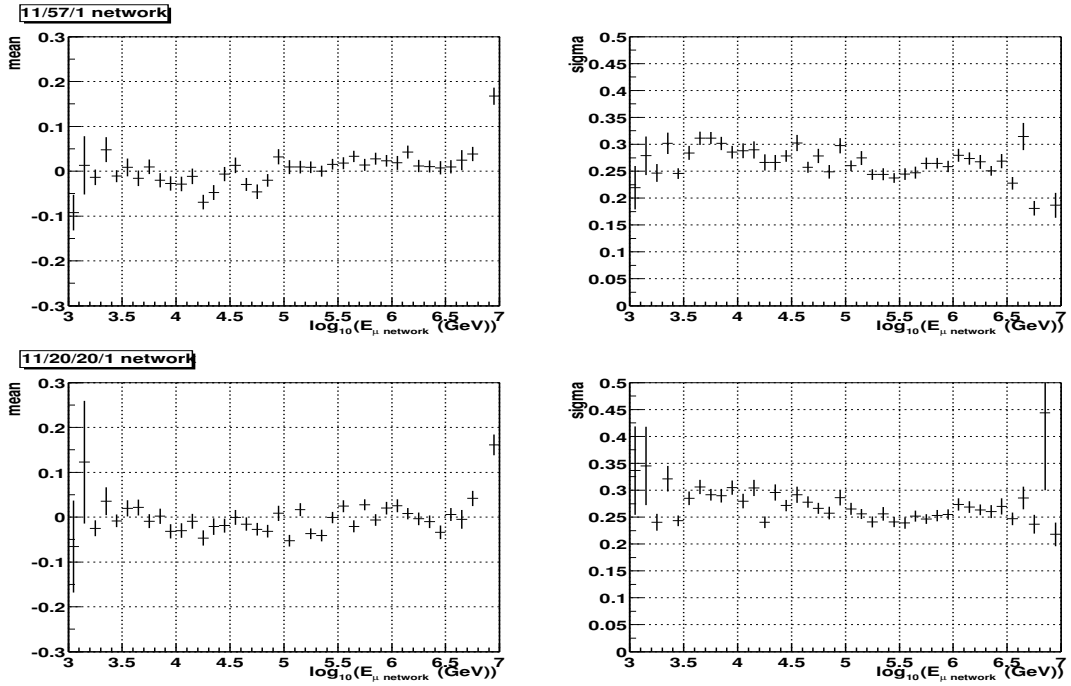

Figure 4.18: Mean and standard deviation of $R = \log_{10}(\frac{E_{network}}{E_{true}})$ for networks 5 and 9 as a function of muon energy. The distributions for the other networks can be found in appendix A

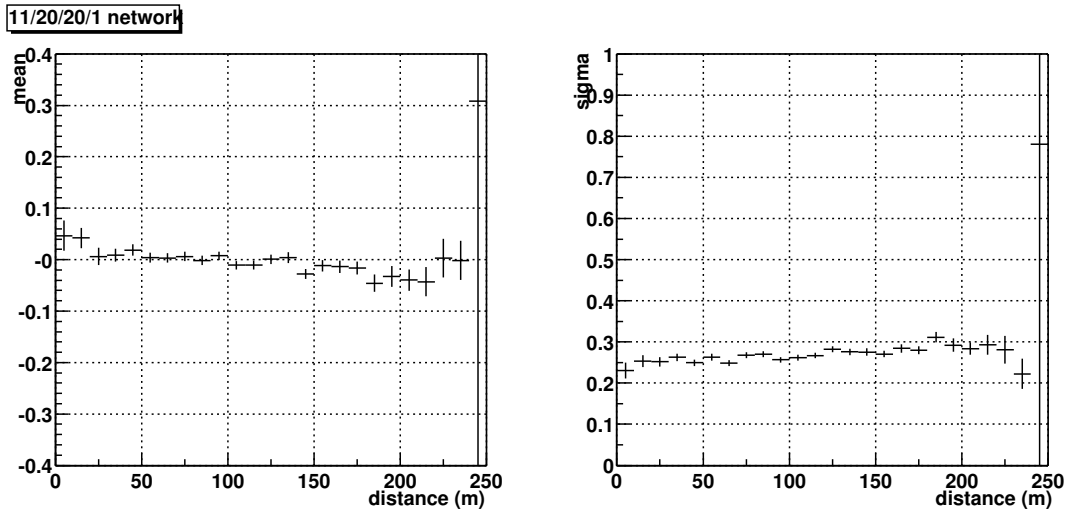

Figure 4.19: Mean and standard deviation of R as a function of the distance from the center of the detector.

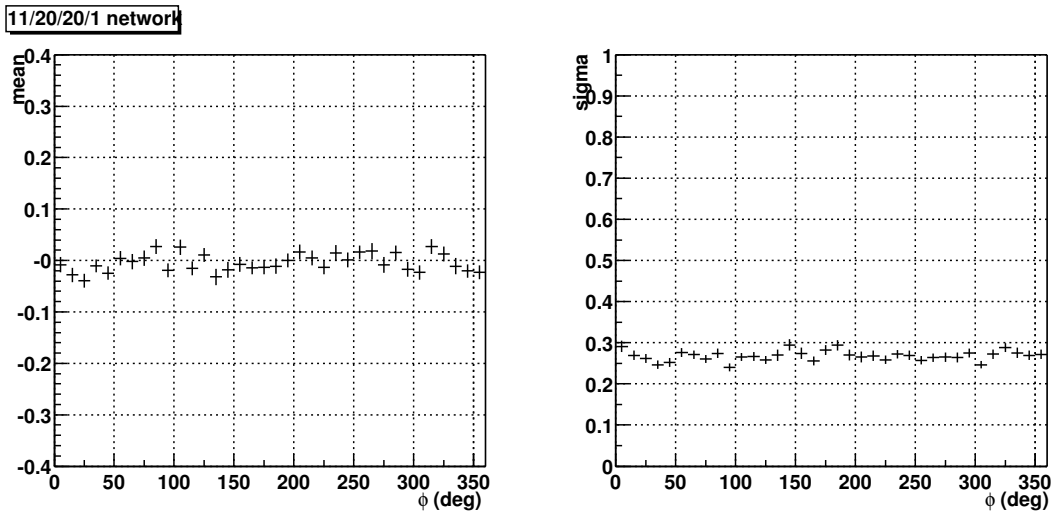

Figure 4.20: Mean and standard deviation of R as a function of the azimuth angle.

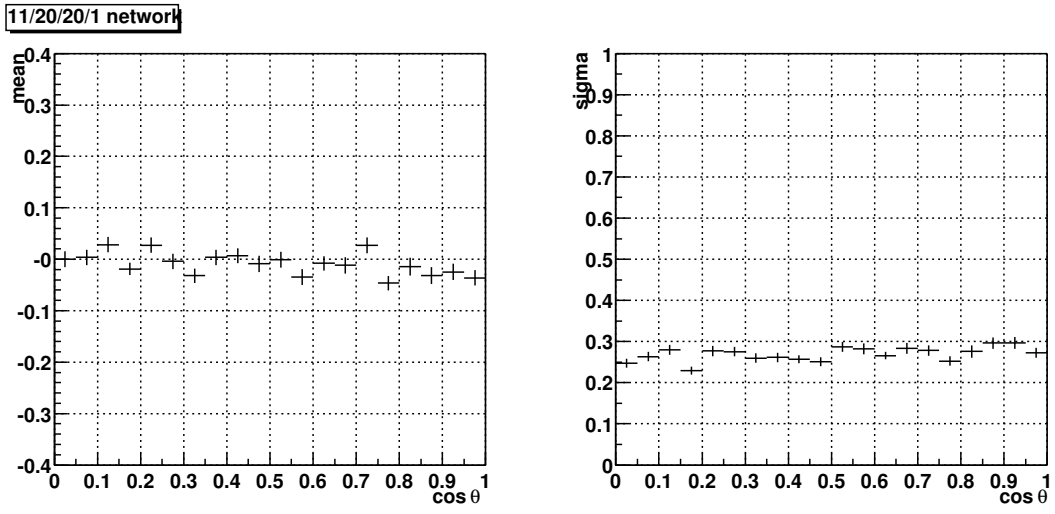

Figure 4.21: Mean and standard deviation of R as a function of the zenith angle.

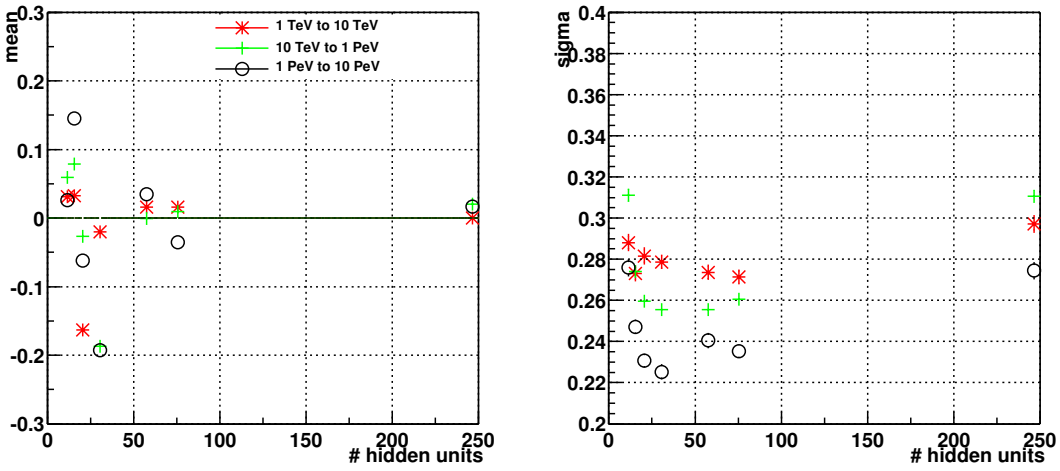

Figure 4.22: Mean and standard deviation of $R = \log_{10}(\frac{E_{network}}{E_{true}})$ for the networks with 1 hidden layer as function of the number of nodes in the layer.

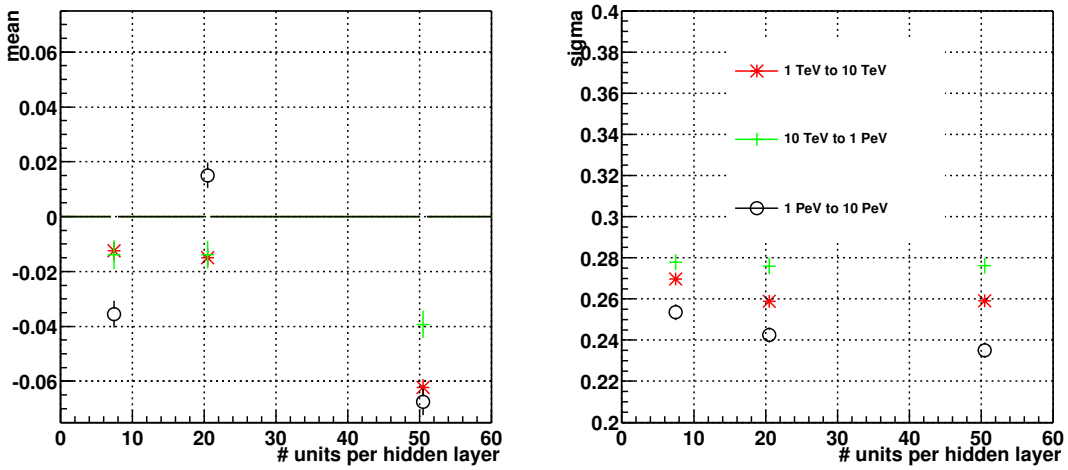

Figure 4.23: Mean and standard deviation of $R = \log_{10}(\frac{E_{network}}{E_{true}})$ for the networks with 2 hidden layers as function of the number of nodes in each layer.

Chapter 5

Discussion

In this chapter I will compare the results of the energy reconstruction with neural networks with the conventional method of reconstruction in ANTARES. Also the neural network reconstruction technique is used to reconstruct fluxes of high energy neutrinos and calculate their integrated event rate in ANTARES. These are compared with the event rates for a perfect energy reconstruction and with the event rate as calculated with the conventional method. The performance of the neural network method using a distorted detector geometry is discussed.

5.1 Conventional Method

The conventional energy reconstruction method as described in 4.1 has recently been improved [43]. The linear fit of equation 4.4 is replaced by a parabolic one

$$\log_{10} x = p_0 + p_1 \log_{10} E + p_2 (\log_{10} E)^2 \tag{5.1}$$

for high and low energy estimators x_{low} and x_{high} and is inverted to get the muon energy as a function of the estimator x. An additional criterium is used to distinguish between high energy and low energy events :

- if $x_{low} < 3.5$ the low energy parameters are used
- if $x_{low} \geq 3.5$ the high energy parameters are used

It is important to note that the results of this energy reconstruction method which I will show below are obtained using a different *angular* and *positional* reconstruction method which is described in [44]. This was necessary because the relevant energy reconstruction software was not available at the time of this study.

5.2 Comparing the Methods

The results of the conventional reconstruction method are shown in figures 5.1 and 5.2. Compared with the neural network reconstruction method (figure 4.15) the conventional method extends to lower muon energies. The networks are tuned for energies between 1 TeV and 10 PeV and, as I will show later, below these energies the performance of the neural network method decreases. The width of the correlation band in figure 5.1 increases with decreasing energy below 100 TeV. This corresponds with the behaviour of the standard deviation of R in figure 5.2 which shows a decrease of the standard deviation that is almost linear with $\log_{10}(E_\mu)$ up to 100 TeV and then remains constant. Note that the true (simulated) energy is used. For comparison, figure 5.3 shows the behaviour of the mean and the standard deviation of R of a 11/20/20/1

network (network 9) as function of true energy. As can be seen, the standard deviation of the neural network method is nearly constant with an average of 0.27 except perhaps in the bin with the smallest energy value. The value of 0.27 is below the lowest value achieved with the conventional method at high energies. The gain of the neural network method over the conventional method is most visible in the energy range from 1 TeV to 100 TeV as with lower energies no comparison is possible. In section 5.3 when calculating fluxes, the behaviour of the neural network method at lower energies will be shown. The mean of both reconstruction methods fluctuates around zero. At low (< 6 TeV) and at high (> 2 PeV) true energies the bias of the neural network reconstruction method starts to diverge. At low energies there is a systematic overestimation and at high energies an underestimation. The reason for this is discussed in 4.4.2. The effect is not visible when looking at the mean as a function of reconstructed energy (figure 4.18) except for the highest energy bin. Therefore the suggested solution to add samples to the training set would be more effective to improve the quality of the reconstruction than by applying a correction factor depending on the reconstructed energy.

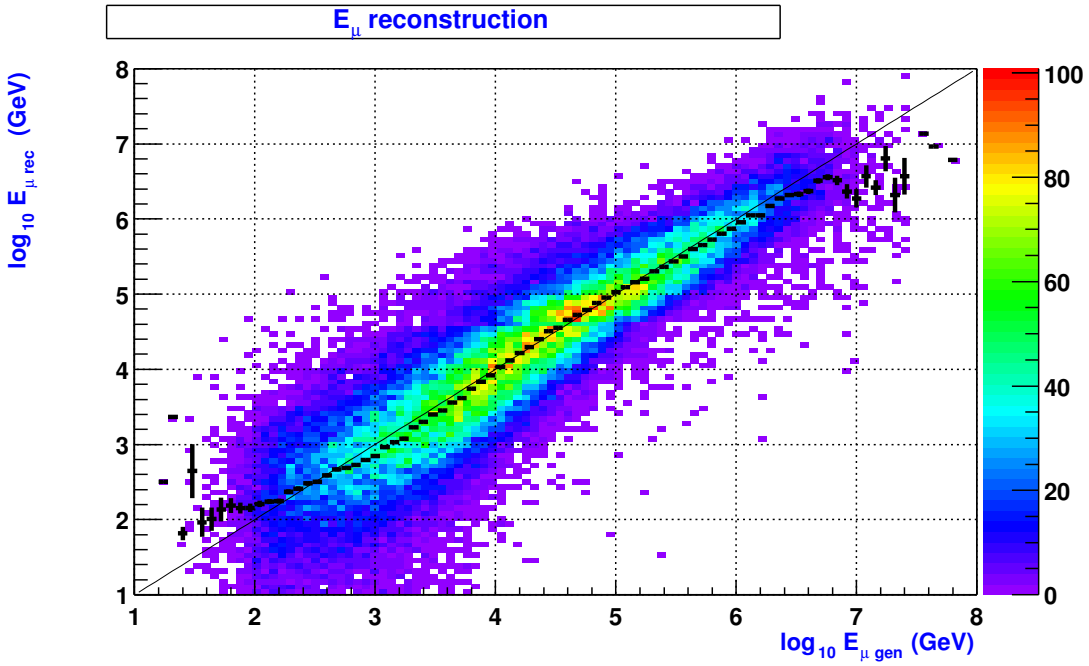

Figure 5.1: Generated muon energy versus reconstructed energy, using an improved version of the conventional energy reconstruction method (from [43]).

5.2.1 Geometrical Parameters

Now the dependence on geometrical factors of both reconstruction methods is compared. The relevant distributions for the neural network method are shown in figures 4.19, 4.20 and 4.21. These are compared with the distributions for the conventional method in figures 5.4 to 5.6 (taken from [43]). The figures show the standard deviation and mean of the distribution of $\log_{10}(E_{reconstructed}/E_{true})$ for the two reconstruction methods.

Dependence on distance

Figures 5.4 shows as a function of the distance to the center of the detector the mean and standard deviation of the R distribution of muon energy reconstruction using the conventional method.

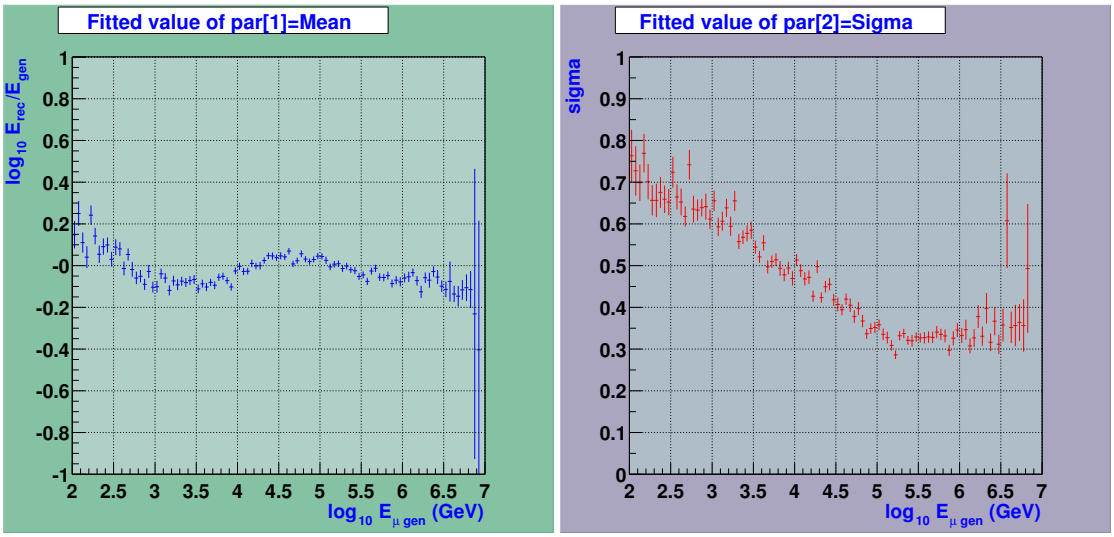

Figure 5.2: Mean and standard deviation of R as function of the generated muon energy for the energy reconstruction with an improved version of the conventional method.

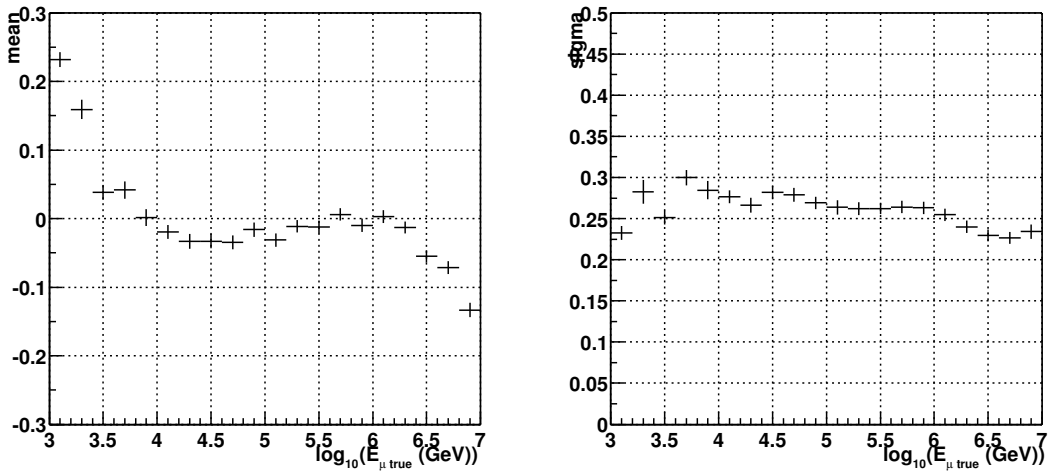

Figure 5.3: Mean and standard deviation of R as function of generated muon energy for the neural network reconstruction method.

Both values drop with increasing distance. This means that the conventional method tends to overestimate the energy of closely passing muons and underestimates the ones that pass further away. The energy of muons passing at larger distance is determined more accurately. This could be an effect of the quality cuts. The effect on the bias can also be seen for the neural network method, but is much smaller, while the effect on the resolution is reversed. It decreases slightly with increasing distance.

Dependence on ϕ and θ

The possible modulation in the azimuthal angle distribution using the conventional method (shown in figure 5.5 and noted in [43]) can at most be seen dimly in the distribution for the neural network method (figure 4.20). The behaviour of the zenith angle dependence (figure 5.6)

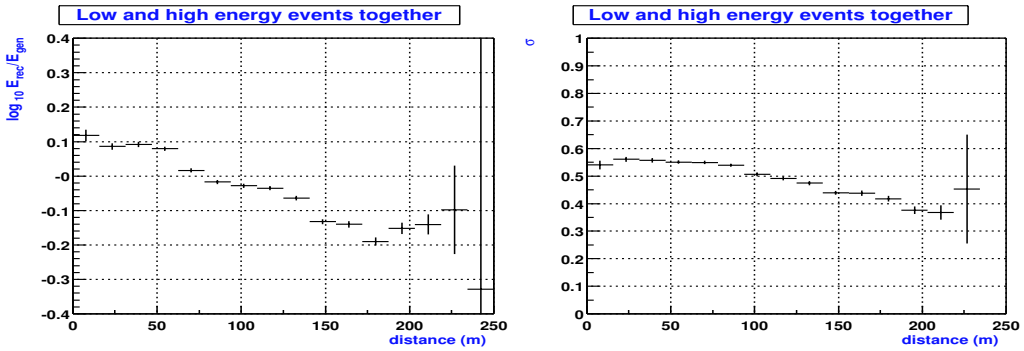

Figure 5.4: Quality of the reconstruction with the conventional method as function of the distance from the center. Left the mean of R and right the standard deviation of R.

is not seen in the distribution for the neural network method (figure 4.21). The conventional method underestimates the energy slightly for horizontal muons while for muons with an azimuthal angle $\leq 37^o$ the energy is increasingly overestimated.

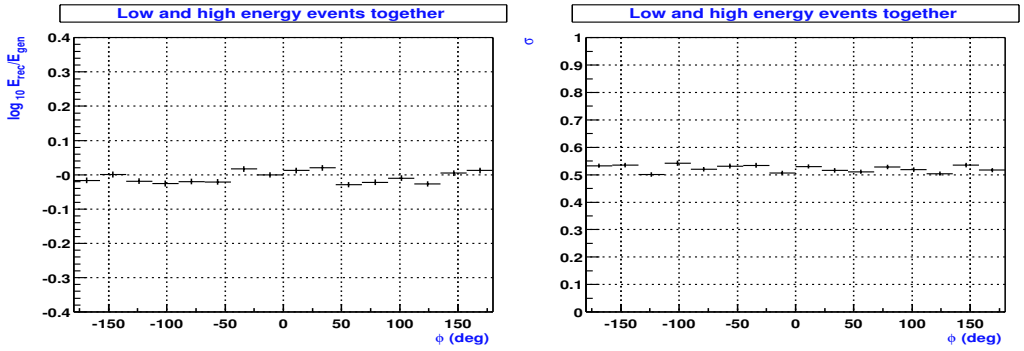

Figure 5.5: Quality of the reconstruction with the conventional method as function of the azimuthal angle. Left the mean of R and right the standard deviation of R.

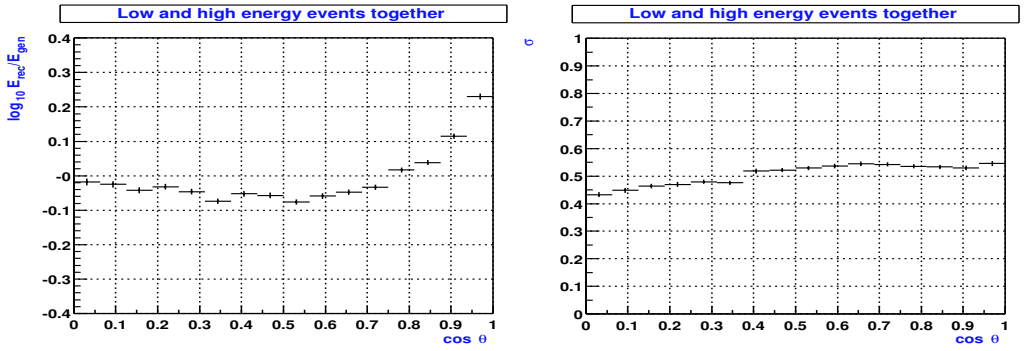

Figure 5.6: Quality of the reconstruction with the conventional method as function of the zenith angle. Left the mean of R and right the standard deviation of R.

Energy reconstruction with the conventional method results in a clear dependence on the geometrical factors which is not seen in the results of the neural network method. The difference can be attributed to the fact that in the neural network method the geometrical factors are used

explicitly as input for the network while the conventional method only uses them implicitly when calculating the expected MIP amplitude.

5.3 Muon Spectra Reconstruction

The neural network energy reconstruction method has been used to reconstruct muon fluxes originating from high-energy neutrinos and predicted by several theoretical models: the Waxman and Bahcall upper bound [30], the MPR98 bound [31] and the Bartol atmospheric neutrino flux [32]. The latter model represents the most important background of high energy neutrinos. Figure 5.7 shows the generated and reconstructed spectra. The generated spectra extend to energies lower than 1 TeV while the reconstruction algorithm is optimized for energies above this value and the neural network is only presented with a limited number of events with energies below this energy. As a result the network can not extrapolate correctly to these low energies and low energy events are reconstructed with a higher energy, mainly between 1 and 10 TeV. Most affected are the two softer spectra (MPR and Bartol), which have a higher spectral index, and thus have more muons with lower energies. Above 10 TeV the effect is gone and the generated and reconstructed spectra are very similar, especially the two hardest (Waxman and Bahcall and MPR).
Also shown in figure 5.7 are the integrated event rates in ANTARES for these spectra. As expected from the reconstruction of the spectra, the event rates that are closest to the event rates with perfect reconstructed energy, are those from the model with the hardest energy spectrum. The Bartol integrated event rate shows a large bump around 1 TeV due to overestimation of lower energies.
Figure 5.8 shows the integrated event rates in ANTARES reconstructed with the conventional method. This figure is made with more statistics than figure 5.7, which is a subset of it. The main difference between the two reconstruction methods can be seen with the Bartol spectrum. With the conventional method no bump appears because this method is more adapted to the reconstruction of energies below 1 TeV. However, at high energies (> 10 TeV) the reconstructed integrated event rates are overestimated more with the conventional method than with the neural network method. For example, the event rate for the Bartol spectrum for muons above 100 TeV is overestimated by a factor 12 using the conventional method and by a factor of 1.6 using the neural network method. The Waxman and Bahcall event rate exceeds the background at about 20 TeV using the network method while this value is around 125 TeV using the conventional method. With the neural network method, a value for a cut in energy which should separate background from signal events can be set closer to the optimal value.

5.4 Detector Distortion

When the ANTARES detector is operating it is subject to environmental conditions. The sea-current can distort the shape of the strings. The buoy is able to restrict the displacement of the top of the string to less than 10 meters with a current of 15 cm/s [7]; the maximum measured current is 19 cm/s. The fluctuations on the current direction and speed are small enough to assume a fixed geometry of the detector within a run.
The neural networks are trained with a particular detector geometry. This affects the geometrical input parameters of the network. When the detector geometry is not the same when reconstructing events this could have an effect on the results. The effects have not yet been studied, but if they affect the resolution of the energy reconstruction there are several solutions to solve this problem. One solution is to train the network for the geometry at hand. The drawback of this is that this should be done for a number of geometries. One could also omit the geometrical input parameters or replace them with parameters that are insensitive to detector distortion.
A study of the effects of the distortion of detector geometry on the reconstruction quality should reveal whether adjustments are needed.

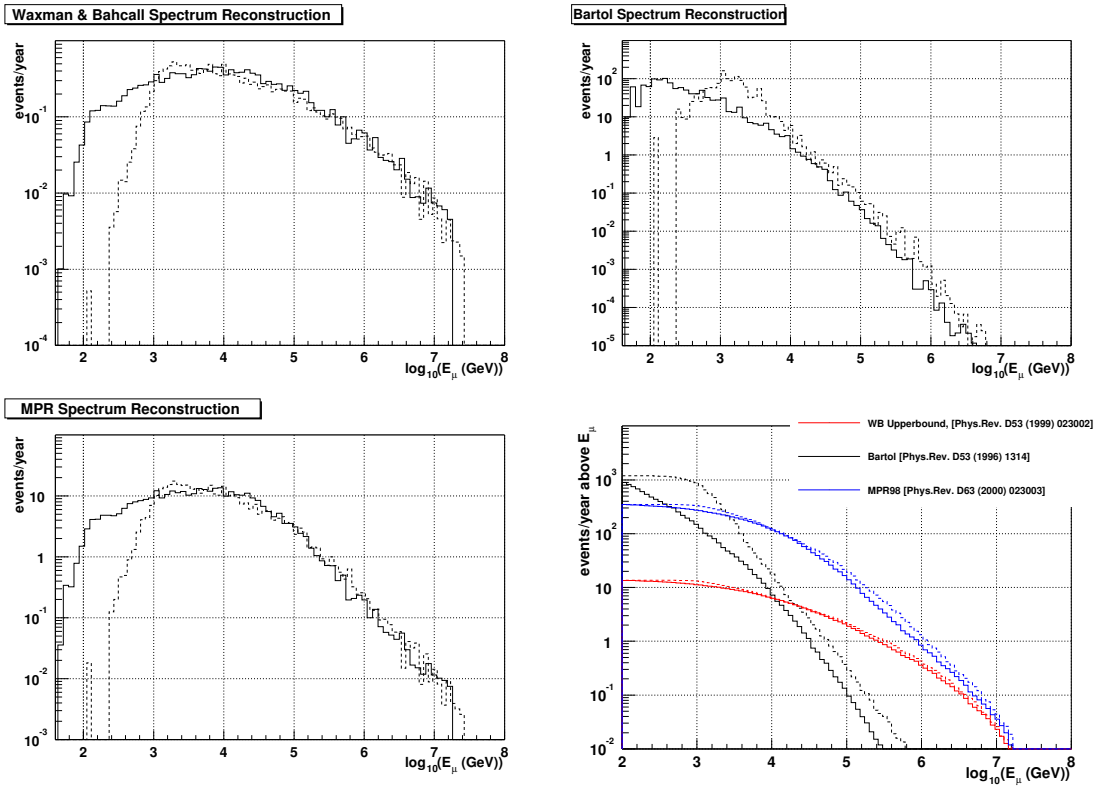

Figure 5.7: Muon flux predicted by three different models. Solid lines represent the simulated energies, dashed lines the reconstructed energies using the neural network method.

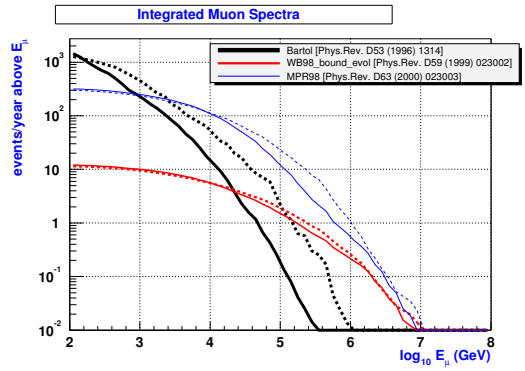

Figure 5.8: Integrated rate of predicted (solid lines) and reconstructed (dashed lines) muons versus simulated and reconstructed muon energy respectively, using the conventional energy reconstruction method.

Chapter 6

Conclusion

In this thesis I set out to develop an alternative method for the reconstruction of the energy of the muon in the ANTARES detector using neural networks. Parameters that carry information on the muon energy other than the light yield were identified in the time distributions of the numbers of photoelectrons induced and the numbers of PMTs that registered hits. Several network topologies were trained and tested. The method was optimized for muon energies between 1 TeV and 10 PeV.

The muon energy can be reconstructed within a factor 2 to 1.6, improving with increasing energy. At low true muon energies ($<$ 6 TeV) there is an overestimation of the energy and at high energies ($>$ 2 PeV) there is an underestimation. This bias can be resolved by extending the training samples of the networks to energies below 1 TeV and above 10 PeV. The dependence on geometrical factors is small.

Comparison with the conventional reconstruction method of ANTARES shows that the neural network method improves the resolution between 1 TeV and 10 PeV of muon energy by a factor 2 at 1 TeV and 1.3 at 10 PeV. The neural network method also performs more stable as function of the distance of the muon track to the center of the detector and its zenith angle.

The neural network method has not been trained for energies below 1 TeV. As a result, muons with energies below 1 TeV are systematically overestimated to energies between 1 TeV and 10 TeV. This is visible in the reconstruction of muon fluxes. As a consequence, harder spectra are reconstructed better than softer spectra. Separation of the flux of atmospheric neutrinos from cosmic neutrinos can be improved using the neural network technique.

Though conceptual improvements such as adding different parameters or tweaking existing ones can be useful, the main improvement of the method should come from training the network for lower and higher energies. Especially training at lower energies is crucial as the flux of low energy muons is higher. Training at higher energies requires the adaption of the Monte-Carlo software which cannot handle these events currently.

The neural network reconstruction method explicitly uses the track position and direction as provided by the algorithm which reconstructs the *position* and *angle* of the muon track. It would be interesting to find parameters which allow the network to reconstruct the energy without the direction and position of the track. As computation with neural networks can be parallelized, speeding up processing, this could perhaps be done real-time in on-line reconstruction.

This thesis shows the usefulness of the neural network technique in data analysis. The technique can also be applied to other off- and online data-processing tasks. Finally, it is my belief that the regular shape of the ANTARES detector is suitable for analyses with pattern recognition techniques using neural networks.

Bibliography

[1] M. A. Markov, *Proc. 10 International Conference on High Energy Physics in Rochester*, (1960) 579

[2] A. Roberts,*The Birth of high-energy neutrino astronomy: A personal history of the DUMAND project*, Reviews of Modern Physics, Vol. 64, No. 1, January 1992

[3] V. A. Balkanov *et al.* [Baikal Collaboration], *Lake Baikal deep underwater neutrino experiment: Status report*, Phys. Atom. Nucl. **61** (1998) 886 [Yad. Fiz. **61** (1998) 978] [arXiv:astro-ph/9709070].

[4] R. Wischnewski *et al.* [AMANDA Collaboration], *The Amanda Neutrino Detector: Status Report*, Nucl. Phys. Proc. Suppl. **85** (2000) 141.

[5] The ANTARES Collaboration, Proposal: *A Deep Sea Telescope for High Energy Neutrinos*,1999

[6] F. Hubaut, *Optimisation et caractérisation des performances d'un télescope sous-marin à neutrinos pour le projet ANTARES*,These de Doctorat, Université de la Méditerranée Aix-Marseille II

[7] The ANTARES Collaboration,*Technical Design Report*,(internal) acessible through

`http://antares.in2p3.fr/internal/tdr/`

[8] J. Engelen et al., *NWO investment proposal: ANTARES, A cosmic neutrino observatory*, NIKHEF

[9] J. Zornoza, *Characterization of two Photomultiplier Models and Study by Monte Carlo Simulation of several Calibration Systems based on Optical Beacons for the ANTARES Detector*, Trabajo de Investigación, Universitat de Valéncia, Departamento de Física Atómica Molecular y Nuclear, March 2001

[10] K. Hagiwara *et al.* [Particle Data Group Collaboration], *Review Of Particle Physics*, Phys. Rev. D **66** (2002) 010001.

[11] M. S. Longair,*High Energy Astrophysics 2nd ed., volume 2*, Cambridge University Press, 1994

[12] D. J. Bird *et al.* [HIRES Collaboration], *Evidence For Correlated Changes In The Spectrum And Composition Of Cosmic Rays At Extremely High-Energies*, Phys. Rev. Lett. **71** (1993) 3401.

[13] Z. Fodor, S. D. Katz and A. Ringwald, *Determination of absolute neutrino masses from Z-bursts*, Phys. Rev. Lett. **88** (2002) 171101 [arXiv:hep-ph/0105064].

[14] T. Piran, *Gamma-Ray Bursts And The Fireball Model*, Phys. Rept. **314** (1999) 575 [arXiv:astro-ph/9810256].

[15] A. Dar and A. De Rujula, *The cannonball model of gamma ray bursts: High-energy neutrinos and gamma-rays*, arXiv:astro-ph/0105094.

[16] Q. R. Ahmad *et al.* [SNO Collaboration], *Measurement of day and night neutrino energy spectra at SNO and constraints on neutrino mixing parameters*, Phys. Rev. Lett. **89** (2002) 011302 [arXiv:nucl-ex/0204009].

[17] S. Fukuda *et al.* [Super-Kamiokande Collaboration], *Constraints on neutrino oscillations using 1258 days of Super-Kamiokande solar neutrino data*, Phys. Rev. Lett. **86** (2001) 5656 [arXiv:hep-ex/0103033].

[18] R. Ghandi et al., hep-ph/9604276

[19] R. Fernow, *Introduction to experimental particle physics*, Cambridge University Press,1986

[20] C. Carloganu, *Muon Interactions at High Energies*, ANTARES-Phys/98-013 (Internal Note)

[21] N. Nijenhuis, *Characterization of the ANTARES Photomultiplier R 7081-20*, Doctoraal scriptie, Universiteit van Amsterdam, Faculteit der Natuurwetenschappen, Wiskunde en Informatica, augustus 2002

[22] R. Donders,*Study of the on-shore Trigger for the ANTARES detector*, Doctoraal scriptie, Universiteit van Amsterdam, Faculteit der Natuurwetenschappen, Wiskunde en Informatica, augustus 2002

[23] *Third EGRET Catalog* accesible through :

`http://cossc.gsfc.nasa.gov/egret/3rd_EGRET_Cat.html`

[24] A. Heijboer, *An algorithm for track reconstruction in ANTARES*, ANTARES-Soft/2002-002 (Internal Note)

[25] J. van Rantwijk, *Data Transmission in the Antares Data Acquisition System*, ANTARES-Soft/2002-005 (Internal Note)

[26] D. J. L. Bailey,*Gentra v4r0 Manual*,(internal) accesible through

`http://www-pnp.physics.ox.ac.uk/~daveb/internal/genhen/manual/gentrav4r0.html`

[27] D. J. L. Bailey,*Genhen v5r1, Software Documentation*, ANTARES-Soft/2002-004 (Internal Note)

[28] S. Navas,L. Thompson, *KM3 User Guide and Reference Manual*, ANTARES-Soft/1999-011 (Internal Note)

[29] D. J. L. Bailey, *KM3 v2r1 : User Guide*, ANTARES-Soft/2002-006 (Internal Note)

[30] E. Waxman and J. N. Bahcall, *High energy neutrinos from astrophysical sources: An upper bound*, Phys. Rev. D **59** (1999) 023002 [arXiv:hep-ph/9807282].

[31] K. Mannheim, R. J. Protheroe and J. P. Rachen, *On the cosmic ray bound for models of extragalactic neutrino production*, Phys. Rev. D **63** (2001) 023003 [arXiv:astro-ph/9812398].

[32] V. Agrawal, T. K. Gaisser, P. Lipari and T. Stanev, *Atmospheric neutrino flux above 1 GeV*, Phys. Rev. D **53** (1996) 1314 [arXiv:hep-ph/9509423].

[33] L. Lönnblad,C. Peterson and T. Rögnvaldsson,*Finding Gluon Jets with a Neural Trigger* Phys. Rev. Lett.65,11 (1990)

[34] L. Lönnblad,C. Peterson and T. Rögnvaldsson, *Mass reconstruction with a neural network* Phys. Lett. B278 (1992) 181-186

[35] The D0 Collaboration, *Neural Network Analysis of Top Quark Production*, Submitted to the *International Europhysics Conference on High Energy Physics*, EPS-HEP99, 15-21 July, 1999, Tampere, Finland

[36] T. Maggipinto et al., *Role of Neural Networks in the seach of the Higgs boson at LHC*, Bari-TH/268-97, May 1997 [arXiv:hep-ex/9705020]

[37] C.M. Bishop,*Neural Networks for Pattern Recognition*, Oxford University Press Inc.,1995

[38] B. Müller, J. Reinhardt and M.T. Strickland, *Neural Networks,An Introduction*, Springer-Verlag Berlin Heidelberg 1990,1995

[39] A. Oppelt,*Energy reconstruction*, ANTARES-Soft/2001-008 (Internal Note)

[40] A. Zell et al., *SNNS, Stuttgart Neural Network Simulator, User Manual, Version 4.2*, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems

[41] R. Mirani, *Parametrisation of EM-showers in the ANTARES detector-volume*, Doctoraal scriptie, Universiteit van Amsterdam, Faculteit der Natuurwetenschappen, Wiskunde en Informatica, January 2002

[42] R. J. Barlow,*Statistics,A guide to the use of Statistical Methods in the Physical Sciences*,John Wiley & Sons Ltd., Chichester, England, 1997

[43] J. Zornoza, *New Improvements on Energy Reconstruction*, ANTARES-Soft/2002-012 (Internal Note)

[44] E. Carmona,J. J. Hernández,*A new reconstruction technique for ANTARES*, ANTARES-Soft/00-11 (Internal Note)

Appendix A

Additional Network Plots

This appendix contains additional plots showing the results of trained networks.

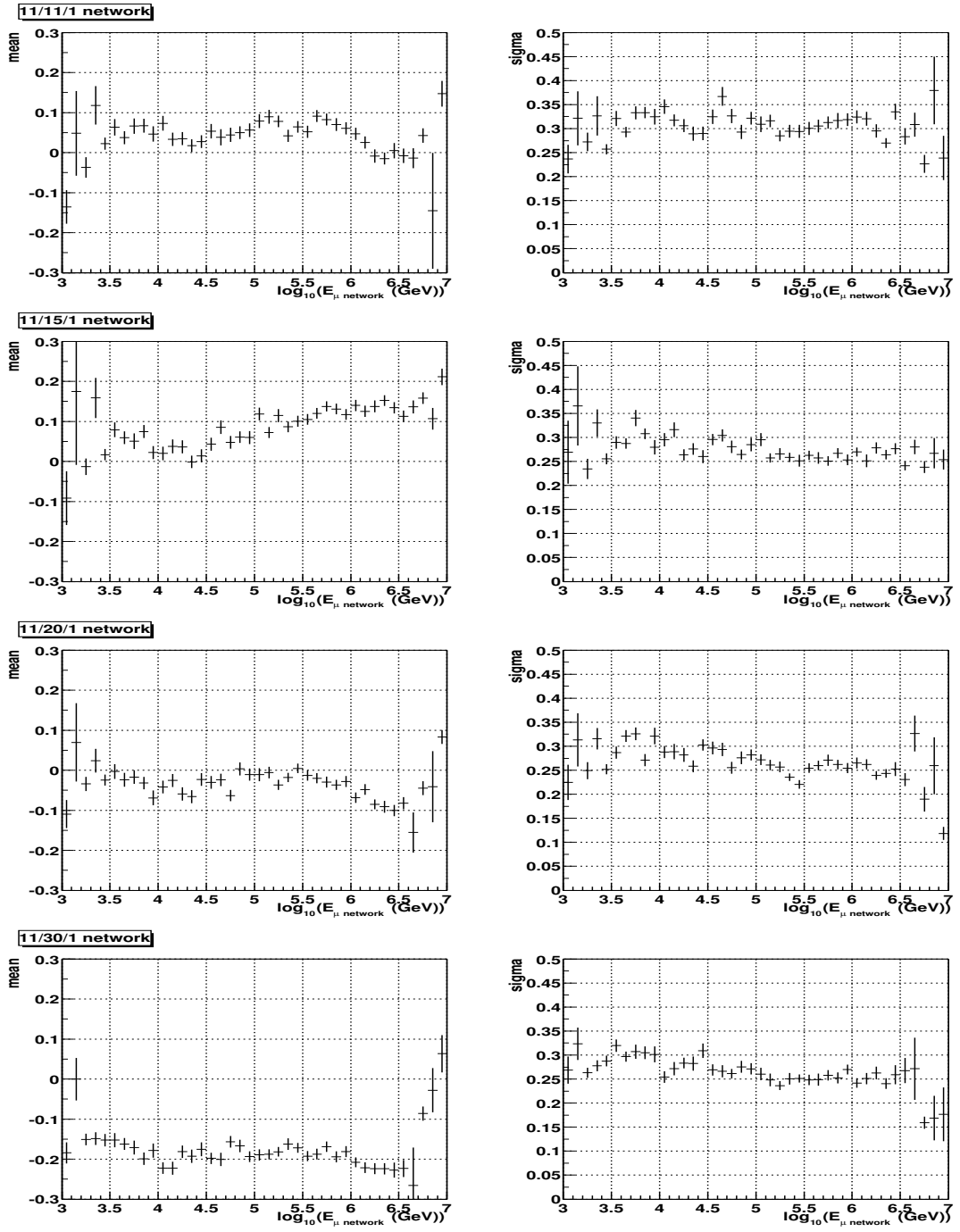

Figure A.1: The means and standard deviations of the energy reconstruction as function of energy for networks 1–4.

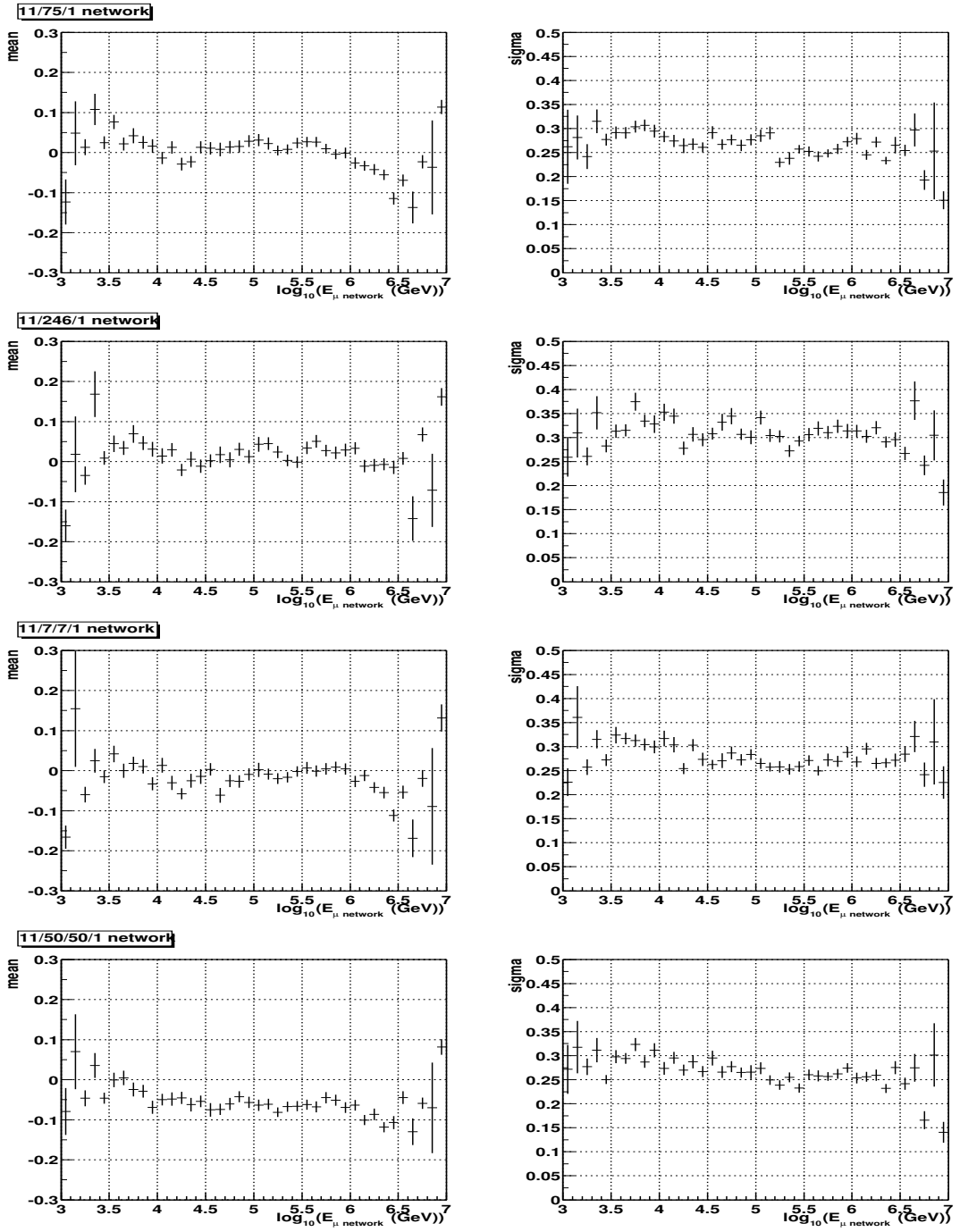

Figure A.2: The means and standard deviations of the energy reconstruction as a function of energy for networks 6,7,8 and 10