Tau Neutrinos in IceCube

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

im Fach: Physik
Spezialisierung: Experimentalphysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Neha Lad M. Sc.
geboren am 18. April 1995

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr. Julia von Blumenthal

Dekanin der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Caren Tischendorf

Colophon

This document was typeset with the help of KOMA-Script and LaTeX using the open-source kaobook template class.

The source code of this thesis is available at:
https://github.com/LeanderFischer/phd_thesis

Abstract

In icecube, we have many neutrinos, select some very high energy ones, spend 1 year with them to group them in three flavour categories. I guess we will learn something about where they came from by doing this. Pretty normal stuff, not at all racist.

Zusammenfassung

Im IceCube haben wir viele Neutrinos, von denen wir einige mit sehr hoher Energie auswählen, verbringen 1 Jahr mit ihnen, um sie in drei Geschmackskategorien einzuteilen. Ich vermute, dass wir auf diese Weise etwas darüber erfahren, woher sie kommen. Ziemlich normales Zeug, ganz und gar nicht rassistisch.

Contents

Results | 1

The 7.5 years of HESE data (2010-2017) was previously used to measure the composition of astrophysical neutrino flavors [1] (in particular to search for ν_τ events) and energy spectrum [2]. This dataset included 102 events, (of which 60 events were above 60 TeV), that passed the HESE selection criteria, as outlined in section **??**. Two events were identified as Double Cascade candidates using the particle identifier described in section **??**.

For this iteration, the analysis has undergone significant changes compared to previous versions while maintaining consistent selection criteria and particle identification methods. The most notable difference lies in the ice model, specifically its properties and influence on the reconstruction of tau decay length, detailed in section **??**. The SPICE-Bfr may have impacts on the overall energy estimates, as the number of photons collected in a given time window have significantly changed in this model of ice depending on the alignment of DOMs with respect to the iceflow axis [3]. Furthermore, the treatment of detector systematics has evolved, utilizing the SnowStorm method (see section **??**) instead of discrete Monte Carlo simulation sets used previously. The update also includes revised reconstruction tables [4] for the maximum likelihood reconstruction method described in section **??**. Additionally, various corrections have been applied to monte carlo simulations, encompassing reweighting to incorporate corrections due to tau polarization and initial state radiation corrections to the Glashow cross-sections (see section **??**). Lastly, some of the nuisance parameters and analysis software are also different (see section **??**). As a first step, the 7.5 years of HESE data were re-unblinded due to these changes.

This chapter presents flavor measurements made using 12 years (11.69 to be exact) of HESE data. It begins by discussing the re-unblinding of 7.5 years of data, followed by results from the 12-year fit, including Data-Monte Carlo agreement and detailed post-unblinding checks. Finally, the flavor measurement results are presented and interpreted in the last sectinons.

[1]: Abbasi et al. (2021), *Detection of astrophysical tau neutrino candidates in IceCube*

[2]: Abbasi et al. (2021), *IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data*

[3]: Abbasi et al. (2024), *In situ estimation of ice crystal properties at the South Pole using LED calibration data from the IceCube Neutrino Observatory*

[4]: Yuan (2023), *Detecting neutrinos in IceCube with Cherenkov light in the South Pole ice*

1.1 (Re)Unblinding of 7.5 years of HESE Data

The re-unblinding of the HESE-7.5 data provided new insights, revealing that 64 events met the HESE selection criteria, each with a deposited energy exceeding 60 TeV. It included 6 Double Cascade events*, which is a significant increase from the previous analysis that identified only two Double Cascade events. Notably, 4 of the additional events had initially been classified as single cascades. The reclassification was largely driven by the application of the energy asymmetry cut, which proved to be a crucial factor in differentiating between single and double cascades

---

* Unless explicitly specified, all mentioned classifications are by using SPICEBfr, the default icemodel of the analysis presented in this thesis. At some places in this section, specific comparisons are shown using SPICE-3.2.1

(see section **??**). A key difference in this re-unblinding was the icemodel used for reconstruction, as mentioned before. Hence, the data was processed using both SPICE-3.2.1 (used in previous iteration) and SPICE-Bfr to make direct comparisons with the previous results. Using SPICE-3.2.1, reunblinded data contained total of 62 HESE events above 60 TeV deposited energy, of which 7 events were classified as Double Cascades. Moreover, of these 6 and 7 observed Double Cascades, using different ice models, only 3 are common, landing 7 new candidate events. Despite the changes in the total number of classified events, the two common Double Cascade events identified in both iterations exhibited nearly identical reconstructed properties, as outlined in Table 1.1 (for SPICE-3.2.1).

should there be a similar table for bfr??

Table 1.1: Comparison of Reconstructed quantities of events classified as Double Cascades upon re-unbling 7.5 years of HESE data using **SPICE-3.2.1 icemodel** with previous results (grey cells). Shown in the table are (from left), MJD (Modified Julian Date), reconstructed length, reconstructed Energy of first (E_1) and second (E_2) cascades, Energy asymmetry (E_A), Energy Confinement (E_C) and classified morphology (as per previous anlaysis) of the event. The two common events, 57134 (*Double Double*) and 56265 (*Big Bird*) have nearly identical reconstructed quantities. The most striking differences here are specifically reconstructed length and the fact that E_1 and E_2 seem to be almost flipped for some events, even though reconstructed directions (zenith and azimuth, not shown in the table) remains almost idnetical. The change in E_1 and E_2 changes the E_A, which is the discrimation cut between single and double cascades.

MJD	Length		E_1		E_2		E_A		E_C		morphology
57677	27 m	37 m	3.1 TeV	128 TeV	148 TeV	4 TeV	-0.96	0.94	1	0.99	Single
57134	16 m	17 m	18 TeV	8.7 TeV	92 TeV	79 TeV	-0.89	-0.80	0.99	0.99	Double
56763	10.2 m	84 m	63 TeV	107 TeV	87 TeV	17 TeV	-0.16	0.72	0.99	0.99	Single
55477	14 m	279 m	70 TeV	184 TeV	84 TeV	67 TeV	-0.09	0.46	0.99	0.98	Single
55800	12 m	20 m	105 TeV	133 TeV	89 TeV	38 TeV	0.08	0.55	1	0.99	Single
56221	12 m	12 m	209 TeV	102 TeV	132 TeV	237 TeV	0.22	-0.4	0.99	0.97	Track
56265	17 m	16 m	0.8 PeV	1.2 PeV	1 PeV	0.6 PeV	-0.08	0.29	0.99	1	Double

No matter what icemodel is used, in reunblinding an excess of double cascade events was observed, suggesting the presence of unnoticed changes not already accounted for. Careful search of each step in the particle identification process revealed a significant change in the reconstruction method from the previous iteration. Notably, this change involved the incorporation of high quantum efficiency Digital Optical Modules (DOMs) from DeepCore. In earlier analyses, these DOMs had been excluded from the reconstruction of high-energy neutrino events, in millipede-based reconstructions outlined in **??**. The rationale for their exclusion stemmed from the smaller statistical uncertainties of digitized waveforms, in compared to the systematic uncertainties. These systematic uncertainties, which were not well-characterized for individual DOMs, could not be factored into the table-based likelihood fitting method. Nonetheless, due to advancements in simulation, including enhanced reconstruction tables and detector simulations, the analysis presented in this thesis included the DeepCore DOMs, hence the next step was to explore why Monte Carlo predictions had underestimated the number of Double Cascade events.

Figure 1.1 shows the distribution of classified Double Cascade, both with and without the inclusion of DeepCore DOMs. The monte carlo simulations predicts only $\sim$ 2 Double Cascade events, yet the data revealed 6 (7) Double Cascade events using SPICE-Bfr (SPICE-3.2.1) when DeepCore DOMs were included and only 3 (4) when they were excluded. It remains clear that no matter what icemodel is used, the distribution of reconstructed energy and overall expectation remains same irrespective of whether DeepCore DOMs are included or not. This discrepancy pointed
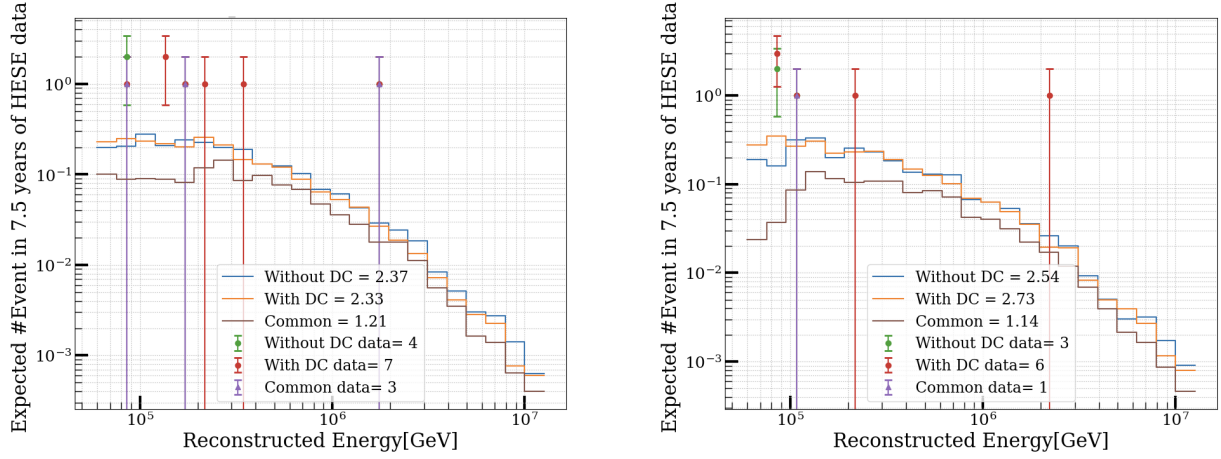
Figure 1.1: Distribution of number of expected classified as double cascacdes in 7.5 years of HESE data usign SPICE-3.2.1 (left) and SPICE-Bfr(right) icemodel with (*With DC*) and without (*Without DC*) DeepCore DOMs, along with data events for the respective configurations.

to potential issues in either the simulation or reconstruction processes involving DeepCore DOMs, indicating that further investigation at the monte carlo level is necessary. Such investigations requires efforts that were beyond the timeline of this thesis work and considering the historical exclusion of DeepCore DOMs (as well as other "bad" DOMs like bright or saturated modules) from reconstruction chains in previous iterations of HESE analyses, this analysis ultimately decided not to include DeepCore DOMs in the full sample unblinding. Ultimately, the re-unblinding of the HESE-7.5 data resulted in 62 events with deposited energies above 60 TeV. Of these, 45 were classified as single cascades, 3 as Double Cascades, and 14 as track events. A detailed comparison between these reunblinded results and previous results, including classified morphologies (with different iterations of icemodels and DeepCore inclusion/exclusion) is shown in Table 1.2.

> In one of the appendix, a short discussion about SPE charge distribution checks done with jvs shall be included

Table 1.2: Event classification of 7.5 years of HESE data (Previous) compared with reunblinded sample outcome (all events have Reconstructed total energy > 60 TeV). The comparison is shown for all 2 different icemodels, each of which further broken down into with and without the deepcore doms. The final outcome of this reunblinding, in order to move forward with the full sample is the last column (without DeepCore using SPICEBfr).

Morphology	Previous	Reunblinded			
		SPICE-3.2.1		SPICE-Bfr	
		DeepCore	No DeepCore	DeepCore	No DeepCore
Cascades	41	41	44	42	45
Double Cascades	2	7	4	6	3
Tracks	17	14	14	16	14
Total	60	62	62	64	62

1.2 Unblinding of 12 years of HESE data

The High-Energy Starting Events (HESE) sample, covering approximately 12 years of IceCube detector live time (4268.7 days) from 2010 to August 2022, was unblinded in January 2024 for the analysis outlined in Section 4.3. This sample includes 167 events that passed the HESE selection criteria, which involved the veto and total charge conditions as detailed

in Section 4.2. Of these, 3 were coincident events that had previously been removed by manual inspection, but were retained this time since the current Monte Carlo simulations now account for such coincidences. Despite this, the energy threshold of 60 TeV effectively filtered them out, resulting in a final sample of 97 events above the 60 TeV deposited energy threshold. Using the ternary topology identification method from this thesis, 64 events were classified as single cascades, 5 as double cascades, and 28 as tracks.

1.2.1 Fit results

Blind Fit results

In IceCube, all the analysis are done in a blind fashion, meaning reconstruction and other fit details are applied to simulation to produce expectations in form of Monte Carlo PDFs and later data is fit to this expectation to make the desired measurement. Details of such a fit are given in **??**. For such fits, before looking at the full data results, sometimes a middle step is considered where only nuisance parameters are revealed first (and/or only background region is looked at) to see any striking disagreement between the data and monte carlo or unexplainable behaviour of any nuisance parameters (e.g. if they are hitting the fit range boundary etc). Stopping criterions are set-up beforehand in case this step shows any troublesome results. For HESE sample, since overall expectation is quite low, looking at a subsample would certainly be unyielding. Thus, blindfit for this analysis included observing the nuisance parameters behaviour and goodness of the fit. The stopping criterions included, minimum p-value of 5% for the goodness of fit test (see **??** for details) and applying priors on systematics in case they hit the boundary. These priors are derived from the so far gathered best knowledge of the ice and detector, and were tested by running pseudo trials (see **??**) to test their effects on signal distributions. As a post-unblinding check, they are to be removed to see if any significant changes are observed in signal parameters.

The bestfir values of nuisance parameters are summarized in Table 1.3 along with their gaussian priors. Initially, the holeice parameters, $\eta_{\text{h.ice}-\text{p0}}$ and $\eta_{\text{h.ice}-\text{p1}}$ were hitting the boundary, but upon applying the tested priors this effect was resolved. As a post unblinding check, these priors were removed to check any changes in signal parameters, no such significant changes were observed. The likelihood includes a penalty term (see **??**) which was checked for the 2D scan of flavour ratio triangle to see the effect of pulls due to this prior.

> perhaps the triangle plot showing llh penalty due to these priors can be shown here? or should this discussion have its separate appendix?

Most of the nuisance parameters are primarily fitted to their priors with minimal deviations from the modeled uncertainties. The detector systematics are set away from default values, although they still remain within one standard deviation of the applied prior width. This can be explained as follows: The method used to apply systematic uncertainties, known as the "gradient method" (see **??**), in the fit assumes linearity whether moving away from or towards the central value. In an ideal scenario, the range within which the parameter is allowed to vary (and the simulation range) should be narrow enough that the central value can always be recovered. However, these parameters are considered only as
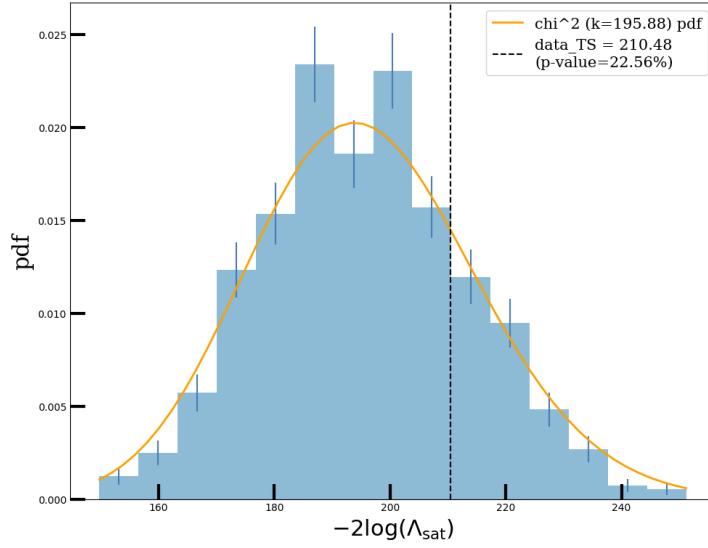
Figure 1.2: Distribution of the test statistics for 1000 pseudotrials injected at best-fit (signal parameters kept blind). Veritcal line shows TS of data, which matches quite well with degrees of freedom derived by fitting a χ^2 to the distribution.

nuisance in the fit, and they are not meant to be measured. Furthermore, the detector is not so precisely calibrated as to know the true value of these parameters. Therefore, in practice, the range is kept quite wide. In the analysis presented here, where the overall expectation is low and the data is binned in histograms, such things can happen, causing the nuisance parameters to shift. In order to save computational power, the gradients are calculated once for a given spectrum assumption. If the underlying spectrum of data is far from this assumed spectrum, small changes in these signal parameters can cause larger deviations. Most of these systematics thus ended up having a gaussian prior in the fit, as listed in the Table 1.3.

Lastly, the goodness of the fit (GOF) is tested by using the pseudo data. This data is genereated by injecting the Best Fit Parameter (BFP) values as the true parameters and then fluctuating the expected bin counts to consider MC uncertainty and Poisson fluctuations in the data. A distribution of Test Statistic (TS) values is then observed, which is derived from the ratio of the fitted likelihood to the saturated Poisson likelihood. The saturated likelihood represents the alternative hypothesis, where the data perfectly matches the assumed model. The saturated Poisson likelihood assumes the most ideal scenario, where each analysis bin observes statistically significant data, following a chi-squared distribution with degrees of freedom equal to the difference between the number of analysis bins and the number of free parameters in the fit. However, in reality, not all bins observe data, so some deviations are expected. By comparing the distribution of TS values from these pseudo-data trials to the distribution derived from the real data, a p-value can be calculated. The p-value represents the probability of finding a TS value from the trials that is at least as large as the one from the data fit. Figure 1.2 illustrates this distribution, where the data TS is compared with the distribution of trial TS values. The obtained p-value is 22.56%. Based on this test, it is concluded that the fit result is consistent with the expected outcome from the pseudo-data trials.

Table 1.3: The best-fit parameter values of the nuisance parameters. The uncertainties are calculated at the 68% confidence level through a profile likelihood scan assuming Wilks' theorem, in case of a flat likelihood space, fit boundaries are given as limits. Last column states the gaussian priors on the parameters (if applicable) in terms of the mean (μ) and width (σ). The table is divided in terms of type of the nuisance parameters, above part includes all the parameters that affects the atmosphgeric neutrino components and the lower part consists if parameters stemming through various detector components. For details, see **??**

	Parameter	Best-Fit value	Prior (μ, σ)
Φ_{conv}	Conventional Flux normalisation	$0.99^{+0.19}_{-0.2}$	(1.0,0.2)
Φ_{prompt}	Prompt Flux Normalisation	$0.0^{+2.25}_{-0.0}$	-
ξ_{CR}	Interpolation etween Cosmic Ray Models	0.042^{+2}_{-1}	(0,1)
$\Delta\gamma_{CR}$	Cosmic Ray Spectral Index Shift	-0.00^{+1}_{-1}	(0,0.05)
H_{Barr}	Barr-parameter modifying the pion-contribution	$0.0^{+0.5}_{-0.5}$	(0,0.15)
W_{Barr}	Barr-parameter modifying the kaon-contribution	$0.0^{+0.5}_{-0.5}$	(0,0.40)
Y_{Barr}	Barr-parameter modifying the pion-contribution	$0.0^{+0.5}_{-0.5}$	(0,0.30)
Z_{Barr}	Barr-parameter modifying the pion-contribution	$0.0^{+0.5}_{-0.5}$	(0,0.12)
$\Phi_{muongun}$	Muon Flux Normalisation	$1.16^{+0.42}_{-0.43}$	(1,0.5)
I_{scale}	scale factor for Neutrino Nucleon Inelasticity weight	$0.99^{+0.1}_{-0.09}$	(1,0.1)
Detector Systematics			
η_{domeff}	Optical Efficiency of DOMs	$1.04^{+0.06}_{-0.04}$	(1.0,0.1)
η_{abs}	Ice Absorption Scaling	$0.99^{+0.04}_{-0.04}$	(1.0,0.05)
η_{scat}	Ice Scattering Scaling	$0.98^{+0.04}_{-0.04}$	(1.0,0.05)
$\eta_{h.ice-p0}$	parametrization for refrozen icecolumn	$-0.27^{+0.28}_{-0.38}$	(-0.27,0.5)
$\eta_{h.ice-p1}$	parametrization for refrozen icecolumn	$-0.08^{+0.04}_{-0.05}$	(-0.042,0.05)
η_{aniso}	Ice Anisotropy Scaling	$0.99^{+0.54}_{-0.63}$	-

Full fit results

The bestfit values of the signal parameters for the 97 HESE events are summarized in Table 1.4. As explained in **??**, the flavour fractions within `NNMFit` framework are fitted as scaling factors s_{ν_e} and s_{ν_τ}, which modifies the flux of electron and tau neutrinos relative to the muon neutrino flux. Which is why, the signal parameters listed in Table 1.4 needs to be converted to their corresponding flavour fractions (), giving us measured flavoured composition of astrophysical neutrinos as, $\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$. Since the measurement is done for the entire spectrum, this value that of best-fit flavour composition which is expected on earth in case neutrinos at high energy sources are produced with muon damped scenario, hints that diffuse neutrino spectrum at high energies (sensitivity range of this analysis is   100TeV-10PeV) may be dominated by sources with such production mechanisms. Nevertheless, the limits derived (see 1.3) cannot reject either of these scenarios with reasonable significance.

reference fraction conversion formula here

The best-fit value of astrophysical index (γ_{astro}), assuming a single power law is $2.84^{+0.19}_{-0.18}$. This value is in well agreement with the previous results [2]. As for the normalization, as per the formulation of the flavour fit within the `NNMFit` the best-fit value shown in the table is astrophysical ν_μ normalization, which when converted using appropriate transformations gives the total flux normalization (for all flavours) $\Phi^{astro}_{\nu+\bar{\nu}} = 5.94^{+5.64}_{-4.28}$. The

[2]: Abbasi et al. (2021), *IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data*

check this number again!!

Table 1.4: The best-fit parameter values for the signal parameters for a single power-law model at 100 TeV muon neutrino energy for both particle and antiparticle. The uncertainties are calculated at the 68% confidence level through a profile likelihood scan assuming Wilks' theorem.

Parameter		Best-Fit value
ϕ_{ν_μ}	astro. ν_μ normalisation [$10^{-18}\text{GeV}^{-1}\text{cm}^{-2}\text{sr}^{-1}\text{s}^{-1}$]	$2.53^{+1.78}_{-1.49}$
γ_{astro}	astro. spectral index	$2.84^{+0.19}_{-0.18}$
s_{ν_e}	astro. ν_e scaling factor to modify total flux norm ($\Phi^{\text{astro}}_{\nu+\bar{\nu}}$) relative to ϕ_{ν_μ}	$0.45^{+0.68}_{-0.40}$
s_{ν_τ}	astro. ν_τ scaling factor to modify total flux norm ($\Phi^{\text{astro}}_{\nu+\bar{\nu}}$) relative to ϕ_{ν_μ}	$0.896^{+1.34}_{-0.89}$

larger uncertainity on normalization is expected for such a fit where each flavour components are allowed to be free, that in turn have larger uncertainty on them (see Table 1.4).

Lastly, the table in Table 1.5 provides the breakdown of expected events classified into specific morphologies and their individual components. Additionally, the comparison includes the expectation based on a $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ flavor ratio, which was determined by running the fit with fixed flavor ratios. It is important to note that the expectation aligns more closely with the data when the flavor ratios are independently fitted.

Table 1.5: The expected number of HESE events, classified into three morphologies, assuming a fixed flavour ratio of 1:1:1 and at the best-fit flavour ratio. The total expected event counts are further broken down into each of the flux components, Astrophysical (Astro), Conventional (Conv), Atmospheric Muons (Muon). Prompt component is not shown here as best-fit value of the prompt norm (Φ_{prompt}) is 0.

Reconstructed Morphology	$\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$(fixed)				$\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$ (free)				Data
	Astro	Conv	Muon	Total	Astro	Conv	Muon	Total	
Cascades	58 ± 2	6.8 ± 0.9	-	65 ± 2.7	57 ± 2	6 ± 0.79	-	63.4 ± 2.4	**64**
Tracks	11.8 ± 0.7	6.3 ± 1	2 ± 3	20 ± 3.6	16 ± 0.8	5.7 ± 0.9	1.84 ± 2.7	23.4 ± 3.4	**28**
Double Cascades	3.2 ± 0.3	0.4 ± 0.2	-	3.5 ± 0.5	3.8 ± 0.3	0.3 ± 0.2	-	4.1 ± 0.4	**5**

1.2.2 Data/Monte Carlo Agreement

Even with a reasonably good fit, any significant mismodeling can only be discerned through a comparison with the Monte Carlo expectations. Nonetheless, the subsequent figures demonstrate the distributions of the analysis observables for all three morphologies, for HESE events with deposited energies above 60 TeV. These plots are generated using the best-fit values described in Table 1.4 and Table 1.3. Each figure displays individual components as well as the sum of all components under the label **MC sum**. Additionally, to highlight any noteworthy features, a ratio plot is included for each distribution, illustrating the ratio of data events to each bin of the MC sum. Although there are observed slight fluctuations, the presence of considerable statistical uncertainties makes it challenging to pinpoint any specific spectral features.

Figure 1.3 and Figure 1.4 shows the distribution of reconstructed energy (left) and cosine of reconstructed zenith (right) for events classified as single cascades and tracks resepctively in the 12 years of HESE data. The zenith plot clearly illustrates the all-sky selection feature of the HESE sample by overall being flat throughout the zenith distribution. It also

highlights the decrease of the conventional component in the downgoing region (cos(zenith>0.25)), indicating the self-veto effect. Additionally, the muongun component is exclusively visible in the track histogram due to fact that the ternary classifier predominantly classifies muons as tracks. However, it's important to note that the expectation exhibits substantial statistical uncertainties due to the lack of sufficient `muongun` monte carlo. To mitigate the impact of the noisy distribution caused by extensive Monte Carlo uncertainties, a kernel density estimation (KDE) is employed to smoothen out the fluctuations present. While both morphology distributions reveal some minor structures, they do not demonstrate any statistically significant characteristics.
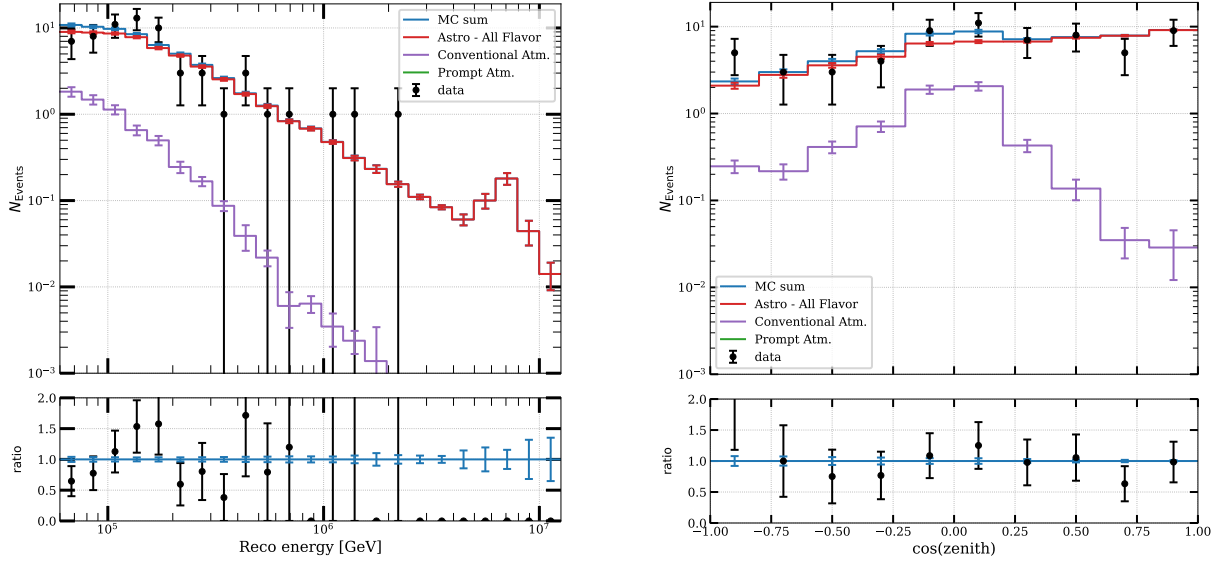
Figure 1.3: Observable distributions of the total deposited energy (left) and the zenith angle (right) for events classified as **single cascades** along with data point positions. Individual components of the fits are produced using bestfit values given in Table 1.4 and Table 1.3 and MC sum is sum of all of these components. Prompt component is missing on account of bestfit value of 0 for the Φ_{prompt}.
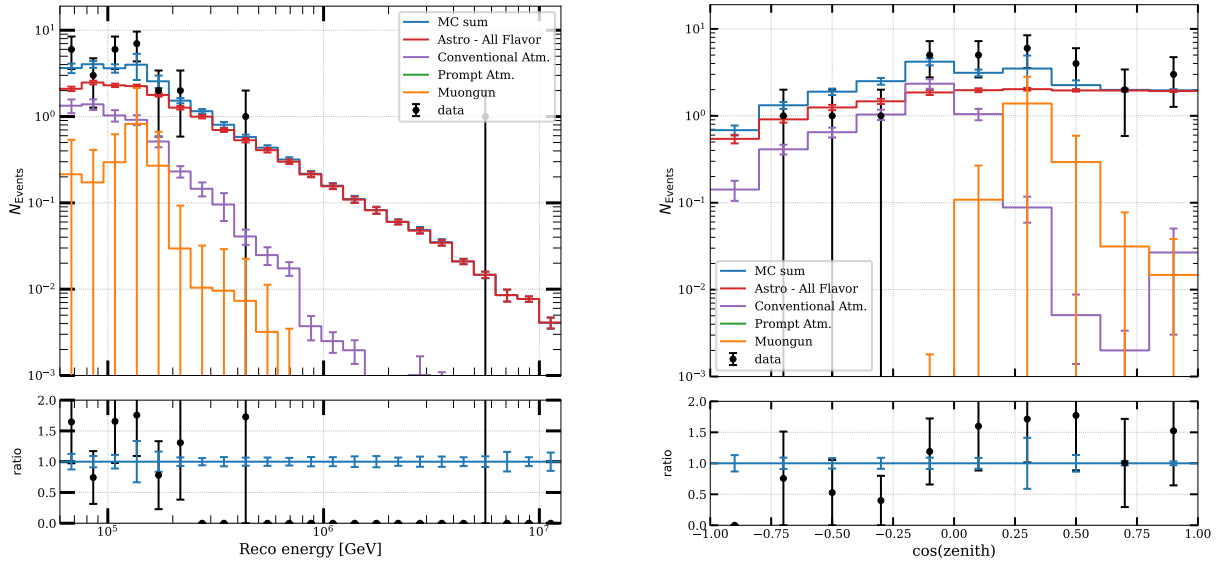
Figure 1.4: Observable distributions of the total deposited energy (left) and the zenith angle (right) for events classified as **tracks** along with data point positions. Individual components of the fits are produced using bestfit values given in Table 1.4 and Table 1.3 and MC sum is sum of all of these components. Prompt component is missing on account of bestfit value of 0 for the Φ_{prompt}.

Figure 1.5 illustrates the distribution of reconstructed energy (left) and reconstructed length (right) for events classified as double cascades in the 12 years of HESE data. In the second energy bin, there is an overfluctuation, but it's challenging to incorporate this as a feature due to the large uncertainties in both the monte carlo and data. An interesting feature of the fit can be seen in the conventional component distribution. The application of gradient correction as an additive factor per component indicates that systematics are trying to adjust this component to account for the excess of double cascades at low energy and high length. This suggests that due to fewer statistics compared to tracks, there is more variability in this parameter across the double cascade bins. Another noteworthy observation is that one event, although close to 100 TeV, has a large length, suggesting that it is a misclassified track. This particular event demonstrates Energy Confinement near the edge (0.97), a cut that is supposed to differentiate between a double cascade and a track, but not quite above the threshold to be a track (0.99). And hence, the event passed all the cuts and ended up being classified as a double cascade. This becomes more evident in 2D distribution of energy and length in Figure 1.6, where this event clearly ends up in a background dominated region of the PDF (see **??** for signal and background pdfs of double cascades).
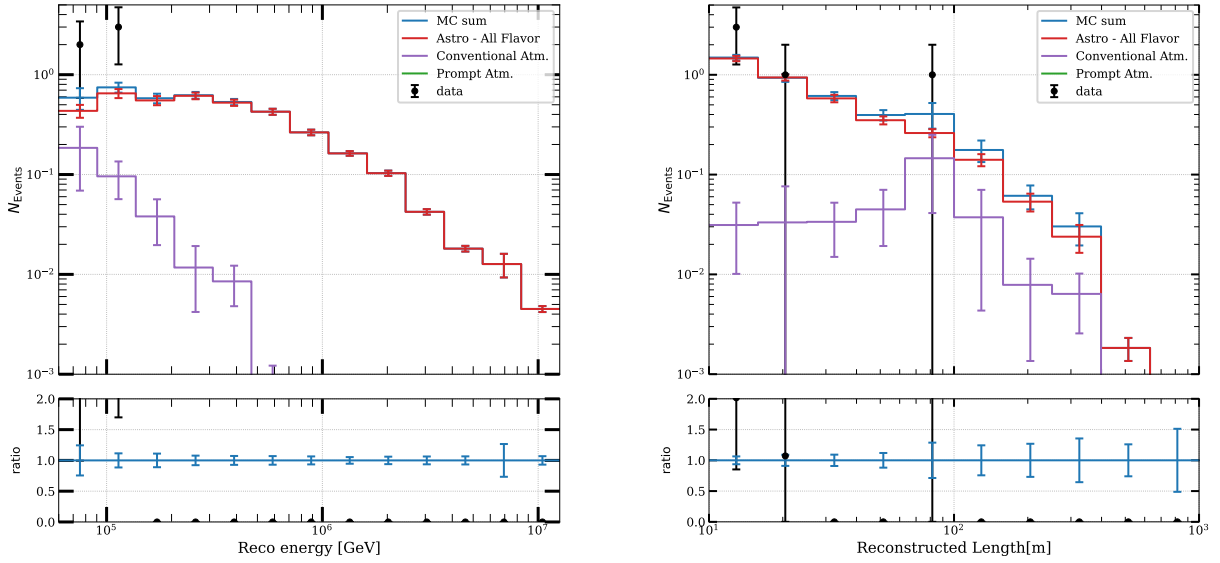
Figure 1.5: Observable distributions of the total deposited energy (left) and the reconstructed length (right) for events classified as **double cascades** along with data point positions. Individual components of the fits are produced using bestfit values given in Table 1.4 and Table 1.3 and MC sum is sum of all of these components. Prompt component is missing on account of bestfit value of 0 for the Φ_{prompt}.

The figure Figure 1.6 illustrates a two-dimensional PID pdf of reconstructed double cascade events, based on the best-fit values described in the tables. The vertical lines demonstrate how quickly the single-cascade background decreases with length. For instance, 68% of misclassified single cascades have reconstructed double-cascade lengths of less than ∼ 14 m, 90% have lengths below ∼ 20 m, and only 1% have lengths exceeding ∼ 25 m. Two of the data events are outside the 68% region, but one of these events is still within the signal region indicated by white lines. The event with a reconstructed length of approximately 95 m but less than 80 TeV energy is clearly in the background region. Another important observation from this figure is the lack of smoothness expected from

PDFs when performing forward folding fits. The gradients are applied as an additive correction per bin, to allow detector systematic variations in the fit. This in turn can significantly change the per bin expectations, making the distributions more fluctuating.
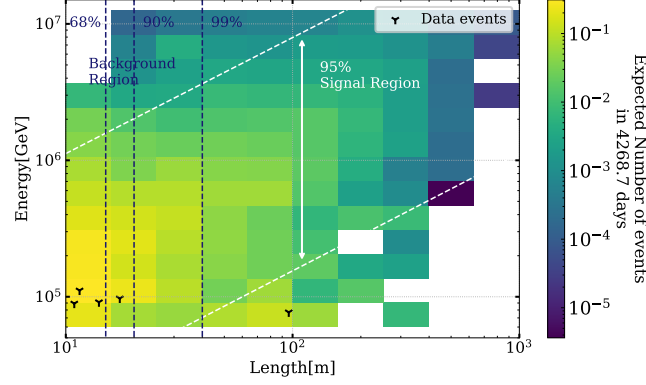
Figure 1.6: Two-dimensional distribution of reconstructed energy vs reconstructed double cascade length of the monte carlo events classified as double cascades. Monte carlo sum is produced using all the bestfit parameters from the Table 1.4 and Table 1.3. Position of the data events are marked with Υ . Signal (white) and Background (dark blue) dominated regions are marked with their respective percentiles.
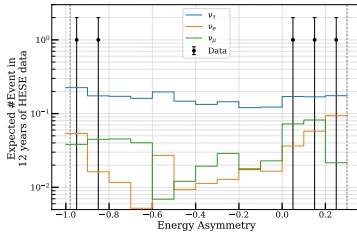
Figure 1.7: Distribution of reconstructed energy asymmetry, defined as $E_A = \frac{E_1 - E_2}{E_1 + E_2}$, (where E_1 and E_2 are energies of first and second cascades, respectively) at best-fit, along with data events. Distribution is shown for all flavours separately, to emphasis domination of ν_τ double cascades in the signal region (verticla grey lines).

[5]: Wilks (1938), *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*

Figure 1.7 presents the distribution of the reconstructed energy asymmetry for simulated events in the double-cascade sample, using the best-fit astrophysical and atmospheric spectra. The two vertical lines represent the selection cuts applied to choose the double cascade events, as described in section 4.3. Although two of the 5 events are near the cut edge, they all still fall comfortably within the signal-dominated region of the sample.

1.3 Flavour Composition of Diffuse Astrophysical Neutrinos

The flavor composition of astrophysical neutrinos is measured using 97 high-energy starting event (HESE) events, which are divided among three morphologies. The profile likelihood scans for the flavor scale factors, specifically s_{ν_e} and s_{ν_τ}, are shown in Figure 1.9. The results indicate that the tau scale factor (s_{ν_τ}) is only able to reject the possibility of no ν_τ flux by 1σ ($-2\Delta \log L = 1$), as reflected in the wide contours in the 2D scan shown in Figure 1.8. The 1D scans are used to derive 1σ (68%) confidence regions, shown in Table 1.4. The test statistic $-2\Delta \log L$ compares the global best-fit values of the unconstrained fit to the conditional best-fit values of all remaining model parameters at a fixed scan point. The confidence regions depicted in Figure 1.8 are calculated using Wilk's theorem [5] , as a full Feldman-Cousins construction of the entire flavor composition phase space is computationally demanding (see section **??** for details). The agreement of Wilk's theorem is verified by comparing the coverage of the test statistic distribution from Monte Carlo pseudo-experiments to the coverage of a χ^2-distribution. A large fraction of the flavor composition phase space is found to be slightly over-covered, indicating that Wilk's theorem yields a conservative confidence region. Although a part of the 90% confidence region seems to suffer from slight under-coverage, the
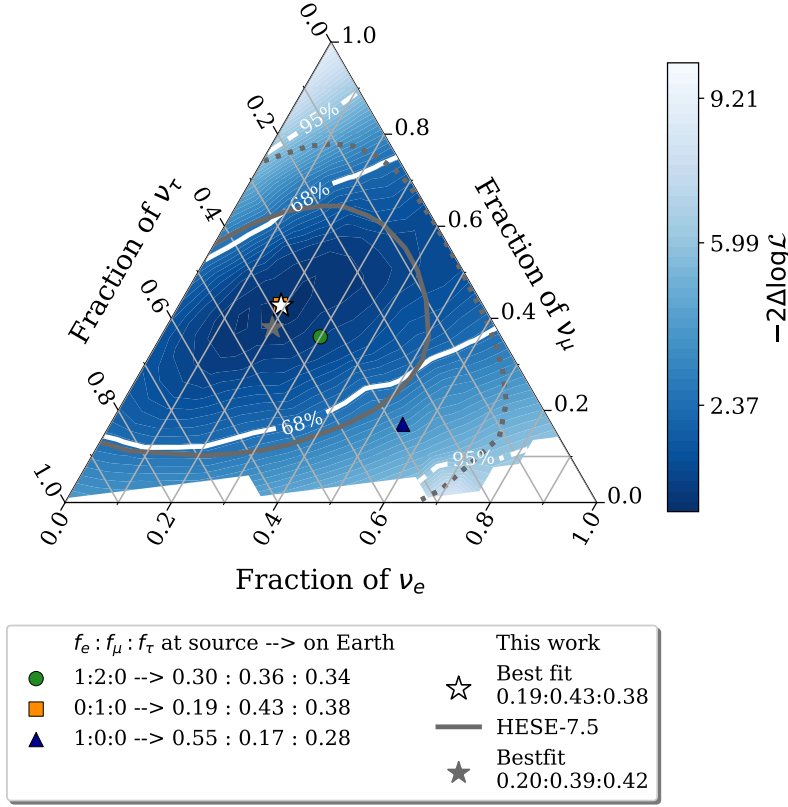
Figure 1.8: A 2 dimensional profile likelihood scan of the astrophysical neutrino flavor composition at Earth using 12 years of HESE data, classified in three event morphologies (see text for details). Each point on the triangle corresponds to a flavor composition of $\nu_e : \nu_\mu : \nu_\tau$ which can be read off the axes along the tick directions specified. The best-fit flavor composition of 0.19 : 0.43 : 0.38 is indicated with a white star. The white solid and dashed lines represent the 68% and 95% confidence regions, respectively, obtained from the χ^2-approximation using Wilk's theorem. Three flavor compositions expected at Earth from different source scenarios are also marked (**??**). The best-fit flavor composition of a previous measurement that used 7.5 years of HESE data is indicated in grey star, with the 68% and 95% confidence regions represented by the grey solid and dotted lines, respectively [1].

χ^2-approximation is deemed sufficient for presenting the measurement result in Figure 1.8.
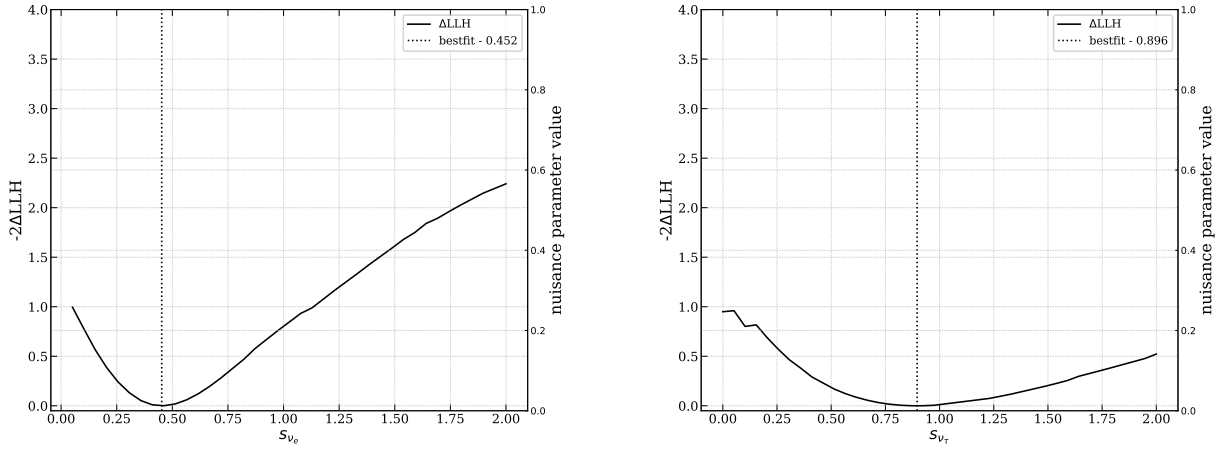
Figure 1.9: 1 dimensional profile likelihood scan of flavour scale factors s_{ν_e} (left) and s_{ν_τ} (right). Solid black line corresponds to the profile likelihood, defined by the likelihood ratio $-2\Delta\log\mathcal{L}$ comparing a fixed value to the best-fit value (denoted by dotted line).

Why are the limits worse than the previous measurements?

The best-fit flavor composition $\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$ aligns well with the previous measurement [1] of $\nu_e : \nu_\mu : \nu_\tau = 0.20 : 0.39 : 0.42$ (grey lines in Figure 1.8). However, with 4 more years of data, which includes more double cascade events, it raises the question of why the limits are not as good. The reason is quite simple, the grey contours are derived using an extended likelihood, based on density estimates by producing a

[1]: Abbasi et al. (2022), *Detection of astrophysical tau neutrino candidates in IceCube*

[6]: Stachurska (2020), *Astrophysical Tau Neutrinos in IceCube*

tauness

"Tauness" refers to the Bayesian posterior probability that an event is derived from a ν_τ interaction. This can be approximated by analyzing the fraction of events in close prox-

imity to the reconstructed properties of the data events, which are anticipated to result from ν_τ interactions. In essence, tauness provides insight into the probability of each double cascade originating from a ν_τ inter-

action compared to any other type of interaction.

large number ($\sim 10^6$ events per particle) of resimulations of 2 classified double cascades to assess the **tauness** . A dedicated algorithm (RODEO) was developed and used to compute the density estimate for sparse datasets produced from these resimulations in multiple dimensions (see [6] for details). This updated likelihood can be seen as the unbinned, higher-dimensional version of evaluating the contributions of signal-like and background-like events, which was carried out in a two-dimensional binned way (just as done for the analysis presented in this thesis). Because of this, the limits one should actually compare with the previous measurement should be the one derived using *the non-extended likelihood*, which is the same likelihood and PDF setup used for this iteration of the analysis. Figure 1.10 shows such a comparison. The best-fit flavor composition $\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$ measured using this analysis (black lines) aligns well with the previous measurement of $\nu_e : \nu_\mu : \nu_\tau = 0.29 : 0.43 : 0.28$ (blue lines) using the same likelihood. Although not significantly, the limits derived from this iteration do get better along the ν_μ fraction. Only 68% limits are shown, without the likelihood space to keep the plot easy to read.

Figure 1.10: The best-fit flavor composition of 0.19 : 0.43 : 0.38 (black), using 12 years of HESE data (this work) compared with the best-fit flavor composition of a previous measurement that used 7.5 years of HESE data [1]. Comparison is shown by usingth esame likelihood formulation (none extended likelihood without resimulations, see text for details). The solid and dashed lines represent the 68% and 95% confidence regions, respectively, obtained from the χ^2-approximation using Wilk's theorem. Three flavor compositions expected at Earth from different source scenarios are also marked (**??**).

Sensitivity at BestFit

Please remember the following text:

The setup and all the relevant updates to the sample and selection chain for the analysis presented in this thesis were done using the best-fit values from the previous analysis. The two signal parameters that affect the flavor measurement, especially the ν_τ fraction and hence the double cascade

events, are the index of the primary neutrino energy spectrum γ_{astro} (assuming a single power law) and the normalization. A harder spectrum (low γ_{astro}) leads to a higher expected flux, resulting in more expected events, while a softer spectrum (high γ_{astro}) leads to a lower flux for samples with a similar effective area. This point was discussed in section **??** while discussing the sensitivity of the analysis. The analysis setup and sensitivity were shown using a spectrum with an index of 2.87, which is almost identical to the best-fit value for this analysis (2.84) assuming equal partition of the neutrino flavor ($\nu_e : \nu_\mu : \nu_\tau = 0.33 : 0.33 : 0.33$). Even though the measured flavor ratio is slightly different, the index remains quite close to what the data agrees with, and the estimated sensitivity projected much tighter constraints on 2D flavor measurements, specifically along the ν_τ axis . To understand the lack of improvement to reduce degeneracy along the ν_e / ν_τ axis, the sensitivity of the flavor measurement was first tested at the best fit.

cite sensitivity figure here

The sensitivity of the bestfit analysis is demonstrated in Figure 1.13. It is determined by using an asimov dataset [7] created with the best-fit values of all signal and nuisance parameters from Table 1.4 and Table 1.3. For the sake of simplicity in comparison, a Poisson likelihood was utilized, as SAY likelihood tends to yield more conservative limits (refer to the appendix for some comparisons). The objective here is to examine the most ideal scenario. Moreover, constructing pseudotrials where data is drawn from SAY likelihood is not straightforward, therefore, to maintain consistency in comparisons, all fits depicted in Figure 1.13 are computed using Poisson Likelihood. Consequently, the 68% data limits displayed on the same plot (black line) are narrower compared to those shown in Figure 1.8. Additionally, there is a slight shift in the bestfit point. The key point here is that, given these signal parameters, the analysis is sensitive not only to measure a non-zero ν_τ fraction, but also to reject it with better significance (approximately 3σ as shown in Figure 1.11 - the 1D profile likelihood scan of s_{ν_τ} for the asimov dataset used to derive the 2D limits shown in Figure 1.13). The sensitivity results give rise to three questions:

Figure 1.11: 1 dimensional profile likelihood asimov scan of ν_τ scale factor, s_{ν_τ} using **simulation** injected at bestfit point. Solid black line corresponds to the profile likelihood, defined by the likelihood ratio $-2\Delta\log\mathscr{L}$ comparing a fixed value to the best-fit value (denoted by dotted line).

[7]: Cowan et al. (2011), *Asymptotic formulae for likelihood-based tests of new physics*

Are data limits reliable?
The contours shown in Figure 1.8 are based on the assumption that Wilk's theorem holds. This theorem states that the $-2\Delta\log\mathscr{L}$ approximately follows a χ^2-distribution with $k = \mathrm{dof}(\hat{\theta}, \hat{\xi}) - \mathrm{dof}(\theta_t, \xi_t)$, where $\mathrm{dof}(\theta, \xi)$ represents the number of free parameters in the fit. In the case of the flavour fit, there are 2 free parameters - s_{ν_e} and s_{ν_τ}. It is important to note that Wilk's theorem is only valid if the sample size is large and the model parameters are not bounded. However, the HESE sample is relatively small with only a few events passing all selection cuts. Additionally, the available Monte Carlo statistics are not large enough to provide sufficient statistics in every bin of the analysis histograms. Furthermore, the parameters are bounded as the flavour fractions cannot be negative. Therefore, it is essential to verify the validity of Wilk's theorem given these conditions.

Is deriving sensitivity using an asimov dataset a good choice?
The Asimov dataset is a convenient alternative to generating a large number of pseudo trials for testing a specific realization. Computing the test statistic distribution of both conditional and global best-fit values for each point in the triangle is quite expensive. It should not be assumed

that the test statistic value that maximizes the likelihood of the Asimov dataset is always equal to the median of the test statistics derived from the full distribution of various pseudo experiments. Therefore, the next step is to calculate sensitivity using Monte Carlo pseudo experiments.

Is the 2 dimesnional monte carlo PDF ideal for the fit to differentiate between signal and background?

Forward-folding fits, as described in section **??**, heavily relies on the expected values for each bin of the analysis histograms. It is crucial to choose reliable variables so that the likelihood space looks different for signal and background. For this reason, the analysis variables for double cascade bins are different from those for cascades and tracks. This is because the zenith distribution of astrophysical ν_τ flux shows no significant variation between signal and background. The PDF used in this case was selected due to the correlation of the Tau decay length (L_{dc}) with the energy of the second cascade (E_2) for double cascades resulting from ν_τ-cc interactions, contributing to the population of signal events along this diagonal . The 2D Monte Carlo PDF at best fit, shown in Figure 1.6, contains Monte Carlo events populating the bins along the diagonal. However, the data events are in regions where the background population is expected with nearly equal probability. While drawing data using poisson distribution, fluctuations around the best-fit are considered which in principle should give all possible realisations of the data events ending up in bins with high expectations [1]. Since data events are observed in bins with not so high expectations as well, it may happen that most pseudo trial histograms would end up not replicating data event distributions. Therefore, in addition to the pseudotrials described earlier, the histograms produced for each trial were checked to see how often the events ended up in a region mostly occupied by background-like events.

Both of these tests were conducted and their detailed descriptions are provided in Appendix A. The validity of Wilk's theorem was tested for a few points on the 68% and 95% contours. This was achieved by generating pseudo datasets (500 trials per dataset), each injected with a specific flavor composition (a point on the contour), while keeping the rest of the fit parameters at their best-fit values. Each trial was fitted twice: once by keeping the parameters fixed at the injection point and once with them free. A likelihood ratio test was performed for each trial by taking the ratio of free to fixed likelihood values. The results are represented in a TS distribution shown in Figure A (for points on the 68% line). For comparison, a χ^2-distribution with 2 degrees of freedom is also shown (black line). The TS distribution of some points shows a minor shift from the χ^2 (k=2) case, but nevertheless, shows no significant violation of Wilk's theorem.

The second test was conducted in a similar way, but this time the fixed point for all trials was kept at the data best-fit point. In Figure B, each point denoted with a marker "+" represents the points for which pseudo datasets were generated. Each dataset was generated by injecting the point denoted with "+" and fitted twice: once by keeping s_{ν_e} and s_{ν_τ} fixed at the data best-fit values (from Table 1.4) and once with them free. This was done to test if the true flavor fraction is what the data fit returns and with what confidence the injected point can be rejected. The result is

reference PDFs from the analysis chapter

1: This was also the motivation beg=hind the resimulations that were performed for the previous iteration of the analysis to measure the flavour composition

Figure 1.12: Test statistic distribution of the pseudo datasets. All datasets are injected at set of points on 68% contour (solid white line) of Figure 1.8 (also shown in legend). a χ^2-distribution with 2 degrees of freedom is also shown (black line) for refernece.

Figure 1.13: Comparison of Measured flavoured ratio (black line) with asimov sensitivity(maroon line) and pseudo trials (marked with '+'). Each + represents a pseudo dataset, drawn from flavour composition of that very point. Colorbar shows confidence intervals of each of these points, to reject the bestfit flavour composition of $\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$. All other parameters of the fit are injected at their bestfit values.

shown in Figure B. The color scheme shows the confidence level each point has to reject the null hypothesis (data best-fit in this case). As can be seen, the pseudo-trial distribution matches quite well with the Asimov sensitivity. In fact, it predicts even tighter constraints compared to the Asimov case, indicating that neither data limits nor Asimov limits are unreliable.

Lastly, the aforementioned third point was checked. Ideally, one would generate pseudo trials until all the Poisson-distributed events fall within the same region of the PDF as shown in Figure 1.6. However, this approach was found to be computationally expensive. For instance, each dataset displayed (around 400) in Figure 1.13 was produced from 500 trials, with each trial fitted twice. On average, each fit took approximately one hour to complete. Therefore, before proceeding, the proportion of PDFs that did not meet the cut criteria was estimated to determine whether using computational resources for further trials was justified. Several tested datasets from different regions of the triangle revealed that, for most datasets, only about 20% of the 500 trials resulted in final histograms with events concentrated in the lower energy bins. There was no feasible way to generate a sufficient number of trials to reduce statistical errors and set definitive limits. As a workaround, only the small proportion of trials that met the necessary criteria were used to generate the TS distributions, allowing an update of the sensitivity shown in Figure 1.13. This updated result is presented in Figure 1.14. As evident from the vanished points, not only were most of the trials excluded, but the confidence intervals also became significantly larger—especially along the ν_τ axis—aligning more closely with the data limits. This concludes why both asimov sensitivity and pseudo trials in Figure 1.13 showed such tight constraints. In an ideal case, one can perform resimulations of the

5 double cascade events as it was done for the previous iteration, but this would in turn increase computational resources and was beyond the timeline of work presented in this thesis, hence it was not performed.

Figure 1.14: Comparison of Measured flavoured ratio (black line) with asimov sensitivity(maroon line) and pseudo trials (marked with '+'). Each + represents a pseudo dataset, drawn from flavour composition of that very point. Colorbar shows confidence intervals of each of these points, to reject the bestfit flavour composition of $\nu_e : \nu_\mu : \nu_\tau = 0.19 : 0.43 : 0.38$. Only trials with final histogram in three lower energy bins are kept in TS distribution. All other parameters of the fit are injected at their bestfit values.
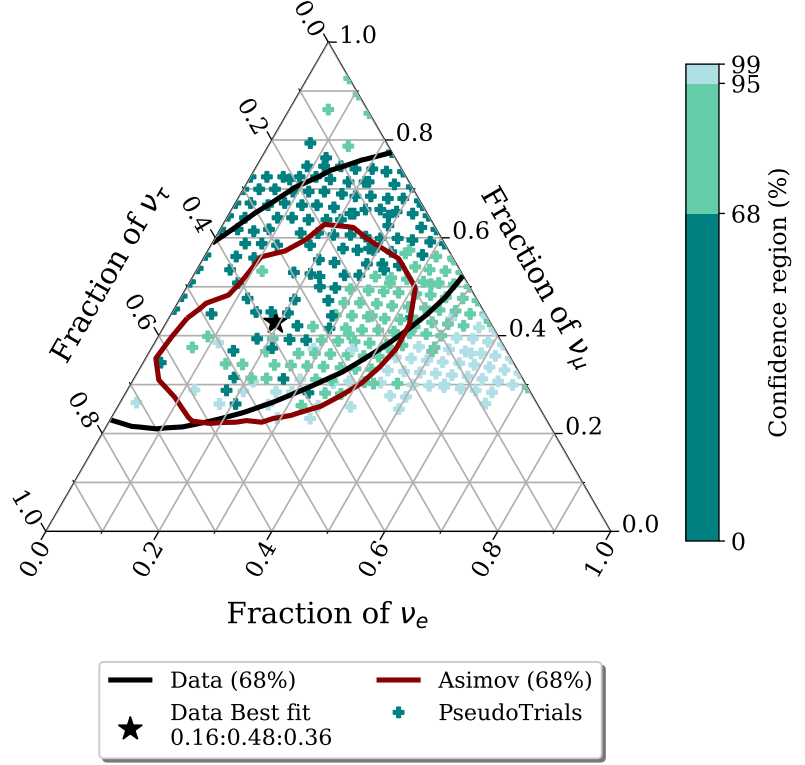
1.4 Discussion

The analysis setup and particle identification methods developed and refined throughout this thesis demonstrate that IceCube is sensitive enough to measure the flavor ratios of the astrophysical neutrino spectrum and set tight constraints. However, more robust techniques, particularly for reconstructing double-cascade events, are needed to produce more reliable Monte Carlo PDFs for these fits. Regarding the measured flavor ratio, as shown in Figure **??**, although the best fit suggests neutrinos are produced by the muon-damped scenario at high-energy sources, no other source scenarios can be excluded with high significance. The neutron beam scenario ($\nu_e : \nu_\mu : \nu_\tau = 1 : 0 : 0$ at the source, evolving to 0.55:0.17:0.28 on Earth) is disfavored by approximately 1σ but remains consistent at 2σ.

A specific measurement of the neutrino energy spectrum was not part of the analysis presented here, as the focus was primarily on flavor measurement, with efforts directed toward improving particle identification. Nonetheless, the HESE sample could also be used to measure, and potentially search for, spectral features. Recent independent studies have discovered a spectral break in the neutrino spectrum at around 30 TeV, softening the spectrum at higher energies, which is within the range of this analysis [8, 9]. This finding offers insight into why previous measurements targeting different energy ranges produced significantly

[8]: Naab et al. (2023), *Measurement of the astrophysical diffuse neutrino flux in a combined fit of IceCube's high energy neutrino data*

[9]: Basu et al. (2023), *From PeV to TeV: Astrophysical Neutrinos with Contained Vertices in 10 years of IceCube Data*

different spectral indices for a single power-law spectrum [2, 10, 11]. Since the HESE sample focuses on high-energy events, where no such spectral features have been observed, and has much lower statistical power than those other samples, no spectral feature searches were attempted during this thesis work.

A less model-dependent approach to describing the astrophysical neutrino flux is the differential unfolding of the energy spectrum. This method divides the flux into energy bins, where each bin is fit independently with a constant $\sim E^{-2}$ spectrum, rather than assuming a continuous power-law across the entire energy range. Ideally, this unfolding would be performed separately for each neutrino flavor to avoid assuming a uniform spectral shape for all flavors. Measuring the flavor composition as a function of energy would also be valuable, as neutrino production processes are energy-dependent. However, due to the limited statistical power of the current dataset, meaningful results could not yet be obtained.

Finally, improvements in reconstruction methods and more data are necessary for future experiments to shed more light on flavor measurements. Double-cascade reconstruction currently has an upper energy limit, as higher-energy events are only partially contained within the detector due to their geometry. In addition, improved detection hardware could provide better information on photon charge and direction, enhancing the overall reconstruction process. These points, along with the potential for a proposed new generation of neutrino detectors, will be discussed in detail in the next chapter.

[2]: Abbasi et al. (2021), *IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data*

[10]: Abbasi et al. (2022), *Improved Characterization of the Astrophysical Muon–neutrino Flux with 9.5 Years of IceCube Data*

[11]: Aartsen et al. (2020), *Characteristics of the Diffuse Astrophysical Electron and Tau Neutrino Flux with Six Years of IceCube High Energy Cascade Data*

APPENDIX

List of Figures

List of Tables

Bibliography

Here are the references in citation order.

[1] R. Abbasi et al. "Detection of astrophysical tau neutrino candidates in IceCube". In: *The European Physical Journal C* 82.11 (2022). DOI: `10.1140/epjc/s10052-022-10795-y` (cited on pages 1, 11, 12).

[2] R. Abbasi et al. "IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data". In: *Physical Review D* 104.2 (2021). DOI: `10.1103/physrevd.104.022002` (cited on pages 1, 6, 17).

[3] R. Abbasi et al. "In situ estimation of ice crystal properties at the South Pole using LED calibration data from the IceCube Neutrino Observatory". In: *The Cryosphere* 18.1 (2024), pp. 75–102. DOI: `10.5194/tc-18-75-2024` (cited on page 1).

[4] T. Yuan. "Detecting neutrinos in IceCube with Cherenkov light in the South Pole ice". In: *Nucl. Instrum. Meth. A* 1054 (2023), p. 168440. DOI: `10.1016/j.nima.2023.168440` (cited on page 1).

[5] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: `10.1214/aoms/1177732360` (cited on page 10).

[6] J. Stachurska. "Astrophysical Tau Neutrinos in IceCube". PhD thesis. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2020. DOI: `http://dx.doi.org/10.18452/21611` (cited on page 12).

[7] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: `10.1140/epjc/s10052-011-1554-0` (cited on page 13).

[8] R. Naab et al. "Measurement of the astrophysical diffuse neutrino flux in a combined fit of IceCube's high energy neutrino data". In: *PoS* ICRC2023 (2023), p. 1064. DOI: `10.22323/1.444.1064` (cited on page 16).

[9] V. Basu et al. "From PeV to TeV: Astrophysical Neutrinos with Contained Vertices in 10 years of IceCube Data". In: *PoS* ICRC2023 (2023), p. 1007. DOI: `10.22323/1.444.1007` (cited on page 16).

[10] R. Abbasi et al. "Improved Characterization of the Astrophysical Muon–neutrino Flux with 9.5 Years of IceCube Data". In: 928.1 (2022), p. 50. DOI: `10.3847/1538-4357/ac4d29` (cited on page 17).

[11] M. G. Aartsen et al. "Characteristics of the Diffuse Astrophysical Electron and Tau Neutrino Flux with Six Years of IceCube High Energy Cascade Data". In: *Phys. Rev. Lett.* 125 (12 Sept. 2020), p. 121104. DOI: `10.1103/PhysRevLett.125.121104` (cited on page 17).