

# **First Search for Heavy Neutral Leptons with IceCube DeepCore**

**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  
**Leander Fischer M. Sc.**  
geboren am 24. Oktober 1992  
in Heidelberg

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

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

### **Copyright Notice**

This book is released into the public domain using the CC-BY-4.0 code.

To view a copy of the CC-BY-4.0 code, visit:

<https://creativecommons.org/licenses/by/4.0/>

### **Colophon**

This document was typeset with the help of KOMA-Script and L<sup>A</sup>T<sub>E</sub>X using the open-source kaobook template class.

The source code of this thesis is available at:

[https://github.com/LeanderFischer/phd\\_thesis](https://github.com/LeanderFischer/phd_thesis)

## **Zusammenfassung**

Zusammenfassung ...

## **Abstract**

Abstract ...



# Todo list

maybe I want a figure for this, or not so important? (YELLOW) . . . . .	1
Elaborate whether this is the case (show it in a plot?). Discuss directionality of cascades in general. (ORANGE) . . . . .	1
fix caption of this figure (RED) . . . . .	2
fix caption of this figure (RED) . . . . .	3
one half sentence on why this number was chosen? (ORANGE) . . . . .	4
make on figure, where the two selected regions are highlighted? (RED) . . . . .	7
fix caption (RED) . . . . .	7
re-make normalized? (ORANGE) . . . . .	7
fix caption (RED) . . . . .	7
re-make normalized? (ORANGE) . . . . .	7
fix caption (RED) . . . . .	7
fix caption (RED) . . . . .	7
I think from these plots it's probably enough if I state the median energy of first and second cascade at final level, as a benchmark for the energies of the two cascades (ORANGE) . . . . .	7
fix caption of this figure (RED) . . . . .	9
describe the "good fit" selection when I first show plots that use it (ORANGE) . . . . .	9
add some 1D resolution plots (ORANGE) . . . . .	9
fix caption of this figure (RED) . . . . .	10
fix caption of this figure (RED) . . . . .	10
fix caption of this figure (RED) . . . . .	11
fix caption of this figure (RED) . . . . .	11
fix caption of this figure (RED) . . . . .	11
Make plot to show efficiency of the OscNext selection for HNL events. (ORANGE) . . . . .	11
add information about the matter profile used (ORANGE) . . . . .	15
add information about the oscillation probability calculation and the software used for it (RED) . . . . .	15
Should I adapt the total numbers to match the sum of the rounded individual parts? (YELLOW) . . . . .	15
add 3D expectation and/or S/sqrt(B) plots . . . . .	16
Add fractions of the different particle types in the bins for benchmark mass/mixing (another table?) (ORANGE) . . . . .	16
I feel like I have to be a bit more precise on what is the fit metric (e.g. the mod chi2) and what is the TS, as in the mod chi2 difference, which is the actual TS, right? (RED) . . . . .	16
Do I want/need to include the description of the KDE muon estimation? (YELLOW) . . . . .	17
Blow up labels/legend/title and make it more readable in the margin or move it into the main text? (RED)	17

I don't like this formulation, but don't know better right now.. (YELLOW) . . . . .	17
elaborate why this is also done to cover the whole energy range for the pion production, referencing the Barr Block plot that I haven't included yet :D (RED) . . . . .	17
I truly dislike this sentence, too, better ideas? (YELLOW) . . . . .	17
I'm just writing out the data from the table, but I need to mention/motivate the included priors here and maybe just point to the table for the ranges/nominal values? (Not quite sure about this) (RED) . . . . .	17
say something about their priors (RED) . . . . .	18
say something about their priors (RED) . . . . .	18
cite?! (ORANGE) . . . . .	18
I should add some final level effects of some systematics on the 3D binning and maybe discuss how they are different from the signal shape, or so? (ORANGE) . . . . .	18
Do I want more information about the different pipelines and stages? Could link back to the extra stage I wrote and add the earth model and oscillation calculation information here, I guess?! (ORANGE) . . . . .	18
again, I think fit metric and TS are mixed up a bit (RED) . . . . .	19
Do I want additional plots for this (fit diff, LLH distr, minim. stats, param. fits)? (YELLOW) . . . . .	19
here again, this is just the fit metric, right? (RED) . . . . .	19
Add bin-wise TS distribution? Add 3D TS maps? (RED) . . . . .	19
fix dimension to fit them in one row! (RED) . . . . .	20
sort the variables also by type, same as in the table "best_fit_parameters"? (ORANGE) . . . . .	21
Show best fit hole ice angular acceptance compared to nominal and flasher/in-situ fits, maybe? (YELLOW) . . . . .	21
Discuss what it means that the parameters are at these values? Here, or somewhere else? (RED) . . . . .	21
fix table caption (RED) . . . . .	21
fix once I have them produced (RED) . . . . .	21
fix once I have the brazil bands (RED) . . . . .	22
fix caption for this figure (RED) . . . . .	22
make plot with BFPs and limit, comparing to upper limits from other experiments, think about how to present it (RED) . . . . .	22
specify which they are, once I have them (RED) . . . . .	22
add 1-d data/mc agreement for example mass sample (0.6?) and all 3 analysis variables (RED) . . . . .	22
add table with reduced chi2 for all 1-d distributions (RED) . . . . .	22
Re-make plot with x,y for horizontal set one plot! . . . . .	25
Re-make plot with x, y, z for both cascades in one. . . . .	25
Re-arrange plots in a more sensible way. . . . .	25
sort these by type of nuisance parameter? . . . . .	29

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>vii</b>
<b>1 Detecting Low Energetic Double Cascades</b>	<b>1</b>
1.1 Reconstruction . . . . .	1
1.1.1 Table-Based Minimum Likelihood Algorithms . . . . .	1
1.1.2 Optimization for Low Energy Events . . . . .	2
1.1.3 Performance . . . . .	4
1.2 Double Cascade Classification . . . . .	8
1.3 Generic Double Cascade Performance . . . . .	8
1.3.1 Idealistic Performance . . . . .	9
1.3.2 Realistic Performance . . . . .	9
1.3.3 Low Energy Event Selection Efficiency . . . . .	11
<b>2 Search for an Excess of Heavy Neutral Lepton Events</b>	<b>15</b>
2.1 Final Level Sample . . . . .	15
2.1.1 Expected Rates/Events . . . . .	15
2.1.2 Analysis Binning . . . . .	16
2.2 Statistical Analysis . . . . .	16
2.2.1 Test Statistic . . . . .	16
2.2.2 Physics Parameters . . . . .	17
2.2.3 Nuisance Parameters . . . . .	17
2.2.4 Low Energy Analysis Framework . . . . .	18
2.3 Analysis Checks . . . . .	19
2.3.1 Minimization Robustness . . . . .	19
2.3.2 Goodness of Fit . . . . .	19
2.4 Results . . . . .	20
2.4.1 Best Fit Nuisance Parameters . . . . .	20
2.4.2 Best Fit Parameters and Limits . . . . .	21
2.4.3 Data/MC Agreement . . . . .	22
<b>APPENDIX</b>	<b>23</b>
<b>A Heavy Neutral Lepton Signal Simulation</b>	<b>25</b>
A.1 Model Independent Simulation Distributions . . . . .	25
A.2 Model Dependent Simulation Distributions . . . . .	26
<b>B Analysis Checks</b>	<b>27</b>
B.1 Minimization Robustness . . . . .	27
<b>C Analysis Results</b>	<b>29</b>
C.1 Best Fit Nuisance Parameters . . . . .	29
<b>Bibliography</b>	<b>31</b>



# List of Figures

1.2 short . . . . .	3
1.3 Preliminary 2-d reconstructed versus true cascade energy resolutions . . . . .	5
1.4 Preliminary 2-d reconstructed versus true decay length resolutions . . . . .	6
2.1 xx . . . . .	17
2.2 Asimov inject/recover test (0.6 GeV) . . . . .	20
2.3 Pseudo-data trials TS distribution (0.6 GeV) . . . . .	20
2.4 Best fit nuisance parameter distances to nominal . . . . .	21
2.5 xx . . . . .	22
A.1 Simplified model independent simulation generation level distributions . . . . .	25
A.2 Realistic model independent simulation generation level distributions . . . . .	26
A.3 Model dependent simulation generation level distributions . . . . .	26
B.1 Asimov inject/recover test (0.3 GeV, 1.0 GeV) . . . . .	27



# List of Tables

2.1	Final level background event/rate expectation . . . . .	15
2.2	Final level signal event/rate expectation . . . . .	16
2.3	Analysis binning . . . . .	16
2.4	Nuisance parameter nominal values and fit ranges . . . . .	18
2.5	Staged minimization routine settings . . . . .	19
2.6	xx . . . . .	21
C.1	xx . . . . .	29



# Detecting Low Energetic Double Cascades

# 1

## 1.1 Reconstruction

All existing reconstruction algorithms applied for low energetic atmospheric neutrino events mentioned in Section ?? are either assuming a single cascade hypothesis or a track and cascade hypothesis, which are the two SM morphologies observable at these energies, as was described in Section ?? . A HNL being produced and decaying inside the IceCube detector however, will produce two cascade like light depositions. The morphology and how the cascade properties and their spatial separation depend on the model parameters was introduced in Section ?? . To investigate the performance of the detector to observe these events, a low energetic double cascade reconstruction algorithm was developed, based on a pre-existing algorithm used to search for double cascades produced from high energetic astrophysical tau neutrinos [1] that was established in [2], but first mentioned in [3].

### 1.1.1 Table-Based Minimum Likelihood Algorithms

The reconstruction is relying on a maximum likelihood algorithm, which is the *classical* approach to IceCube event reconstructions, as opposed to ML based methods. A Poissonian likelihood is constructed, which compares the observed photon numbers,  $n$ , with their arrival times to the expected light depositions,  $\mu$ , for a given even hypothesis as

$$\ln(L) = \sum_j \sum_t n_{j,t} \cdot \ln(\mu_{j,t}(\Theta) + \rho_{j,t}) - (\mu_{j,t}(\Theta) + \rho_{j,t}) - \ln(n_{j,t}!) , \quad (1.1)$$

where  $\rho$  are the number of expected photons from noise,  $\Theta$  are the parameters governing the source hypothesis, and the likelihood is calculated summing over all DOMs  $j$  splitting observed photons into time bins  $t$ . The light expectations are calculated using look-up tables [4] that contain the results from MC simulations of reference cascade events or track segments. By varying the parameters defining the event hypothesis, the likelihood of describing the observed light pattern by the expected light depositions is maximized to find the reconstructed event. Algorithms of this kind used in IceCube are described in great detail in [5]. For the table production a specific choice of ice model has to be made, while the calibrated DOM information is taken from the measurement itself.

Based on the tabulated light expectations for cascades and track segments, various event hypothesis can be constructed, like the common cascade only or the track and cascade hypotheses. The hypothesis describing the double cascade signature of the HNL is using two reference cascades that are separated by a certain distance. The whole hypothesis is defined by 9 parameters and assumes that the two cascades are aligned with each other, which is a safe assumption for strongly forward boosted interactions. The parameters are the position of the first cascade  $x, y, z$ , the direction of both cascades  $\phi, \theta$ , and its time  $t$  as well as the decay length  $L$  between the

1.1	Reconstruction	1
1.2	Double Cascade Classification	8
1.3	Generic Double Cascade Performance	8

[1]: Abbasi et al. (2020), "Measurement of Astrophysical Tau Neutrinos in IceCube's High-Energy Starting Events"

[2]: Usner (2018), "Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector"

[3]: Hallen (2013), "On the Measurement of High-Energy Tau Neutrinos with IceCube"

maybe I want a figure for this, or not so important? (YELLOW)

[4]: Whitehorn et al. (2013), "Penalized splines for smooth representation of high-dimensional Monte Carlo datasets"

[5]: Aartsen et al. (2014), "Energy Reconstruction Methods in the IceCube Neutrino Telescope"

Elaborate whether this is the case (show it in a plot?). Discuss directionality of cascades in general. (ORANGE)

two cascades. Assuming the speed of the HNL to be the speed of light,  $c$ , this already defines the full signature. The HNL particle does not produce any light while traveling, as it is electrically neutral. The full 9 parameters describing the event are  $\Theta = (x, y, z, t, \theta, \phi, E_0, E_1, L)$ . To compute the full likelihood the term in Equation 1.1 is summed over both cascade parts,  $i$ , as  $\sum_i \ln(L_i)$ .

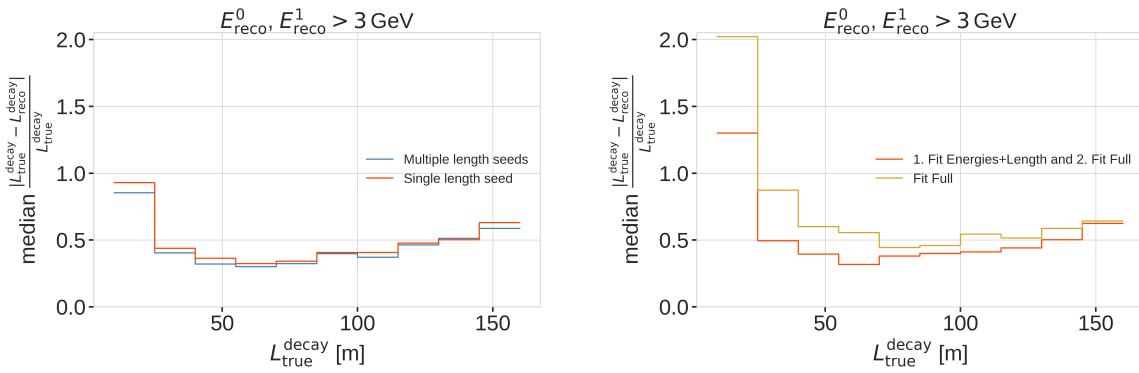
### 1.1.2 Optimization for Low Energy Events

Optimizing the double cascade reconstruction for low energetic events was done in parallel to the development of the model dependent simulation generator introduced in Section ???. A preliminary sample of HNL events was used, containing a continuum of masses between 0.1 GeV and 1.0 GeV and lab frame decay lengths sampled uniformly in the range from 5 m to 500 m. Even though this sample is not representative of a physically correct model and therefore not useful to predict the event expectation, it can still be used to optimize the reconstruction. The double cascade nature of the individual events and the evenly spaced decay length distribution are especially useful for this purpose.

[6]: Abbasi et al. (2022), "Low energy event reconstruction in IceCube DeepCore"

The simulation is processed up to Level 5 of the selection chain described in Section ?? and one of the reconstructions from [6] is applied to the events, fitting a cascade and a track and cascade hypothesis. The results from this reconstruction are used as an input for the double cascade reconstruction, where the position of the vertex, the direction of the event, and its interaction time are used as the input quantities for the first cascade, and the length of the track reconstruction is used as a seed for the distance between the two cascades.

#### Decay Length Seeds



**Figure 1.1**

fix caption of this figure  
(RED)

The full 9 dimensional likelihood space is very complex and can have many local minima, depending on the specific event and its location in the detector. Especially the seed value of the length between the two cascades was found to have a very strong impact on whether the global minimum was found during the minimization. To mitigate this effect, multiple fits are performed, seeding with variations of the input length at different orders of magnitude.

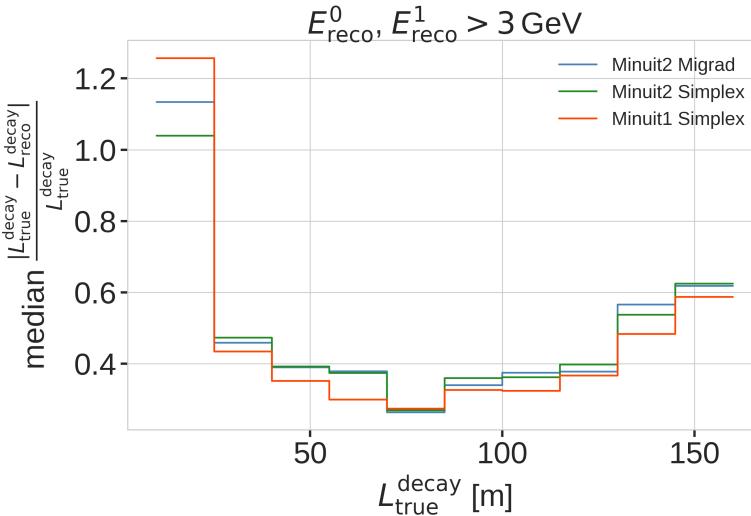
The best result is used, selected based on the total likelihood value of the best fit parameter set. A small improvement in the decay length resolution can be found by using this approach as compared to a single length seed. The effect can be seen in the left part of Figure 1.1, which shows the median, absolute, fractional decay length resolution.

### Fit Routine

Because the length seed showed to have such a large impact on the reconstruction performance, a more sophisticated fit routine, than just fitting all 9 parameters at once, was tested. In a first fit iteration, some parameters are fixed and the resulting best fit point is used to fit all 9 parameters in a second iteration. In the right part of Figure 1.1 it can be seen how a fit split into two consecutive steps, where the first step fits only both cascade energies and the decay length and the second step fits the full 9 parameters, performs better as compared to a single, full 9 parameter fit. The initial seed for both routines is the same.

### Minimizer Settings

To investigate the effect of the minimizer used to find the best fit parameters, the reconstruction was performed using three different minimizers, which were easily accessible within the reconstruction framework. The minimizers used were Minuit1 Simplex, Minuit2 Simplex, and Minuit2 Migrad. The results can be seen in Figure 1.2, where the Minuit1 Simplex minimizer performs best. The initial idea was to test a global minimizer, or a routine that can find the rough position of the global minimum first and then a local minimizer to find the exact minimum, but unfortunately this was not possible with the minimizers available in the framework. From the three tested minimizers, Minuit1 Simplex performed best and was chosen as the default for the reconstruction. The comparison of the decay length resolutions can be seen in Figure 1.2.



fix caption of this figure  
Figure 1.2: title

### 1.1.3 Performance

The chosen reconstruction chain used to test the performance of the detector to observe low energetic double cascades is the following; Minuit1 Simplex is used as the minimizer, the decay length is seeded with 3 different values, 0.5x, 1.0x, and 1.5x the length of the track reconstruction, and the fit routine is split into two steps, where the first step fits the energies and the decay length and the second step fits the full 9 parameters. In the first step, the number of time bins in Equation 1.1 is set to 1, so just the number of photons and their spatial information is used. The second step is seeded with the best results from the first fit and here the number of time bins is chosen such that each photon falls into a separate time bin, which means all time information is used. The average runtime per event is  $\sim 16$  s on a single CPU core, but is very dependent on the number of photons observed in the event, since the likelihood calculation in the second step scales with this number and a table lookup has to be performed for each photon.

To get a more realistic estimate of the reconstruction performance, it is run on a second preliminary sample of HNL events, containing masses between 0.1 GeV and 3.0 GeV and the lab frame decay length is sampled from an inverse distribution in the range from 1 m to 1000 m, which is a better approximation of the expected exponential decay distribution of the HNL. The performance is shown for events where the reconstruction chain was successfully run, the event selection criteria up to level 7 are fulfilled, and the reconstructed energy of both cascades is above 3.0 GeV. This is done to only investigate well reconstructed events with two significant light depositions at a usual final selection level of the oscillation analyses.

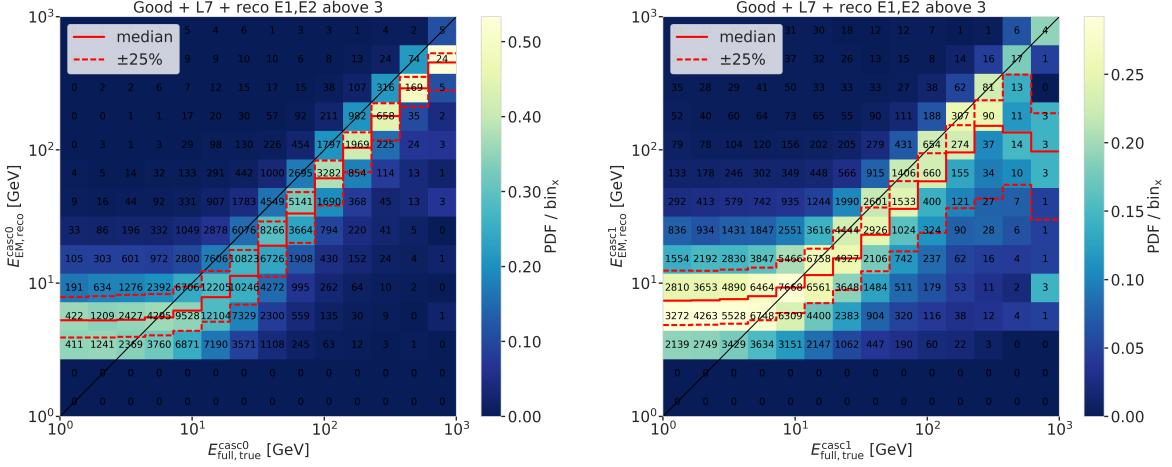
one half sentence on why  
this number was chosen?  
(ORANGE)

### Energy Resolutions

The energy resolution is inspected by looking at the 2-dimensional distribution of reconstructed energy versus the true energy as shown in Figure 1.3. The bin entries are shown as well as the median and  $\pm 25\%$  calculated per vertical column, to get an idea of the distribution for a given energy slice. The color scale is showing the PDF along in each true energy slice, which is the full information combined into the median $\pm 25\%$  lines. The reconstructed energy is only the energy that is observable from photons, while the true energy is the total cascade energy, including the parts that go into EM neutral particles that do not produce light. It is therefore not expected that the median lines up with the axis diagonal, but rather the reconstructed energy is going to be lower.

The histogram for the first cascade energy is shown on the left and above an energy of  $\sim 10$  GeV the reconstruction performs well, with the median being parallel to the diagonal and the spread in the  $\pm 25\%$  quantile being small. Below this energy the reconstruction is over-estimating the true energy, which is a known effect in IceCube, where the reconstruction is biased towards higher energies around the energy detection threshold, because events that enter the sample are events with an over fluctuation in their light deposition, which makes them pass into the selection and being reconstructible in the first place.

For the second cascade the overall behavior is similar, only that the energy where the reconstruction starts to perform good is higher around  $\sim 20$  GeV.



**Figure 1.3:** Reconstructed (EM) energy versus true energy (full) energy for the first cascade (left) and second cascade (right). The color scale is according to the PDF in each vertical true energy slice, with the solid and dashed lines showing the median $\pm 25\%$  quantiles. The bin entries are shown as numbers.

The spread around the median is also larger and starts to expand a lot above 200 GeV, where the statistics are lower as can be seen from the bin counts. It is also very apparent that the majority of energies of the second cascade are at lower true energy values between 1 GeV and 20 GeV.

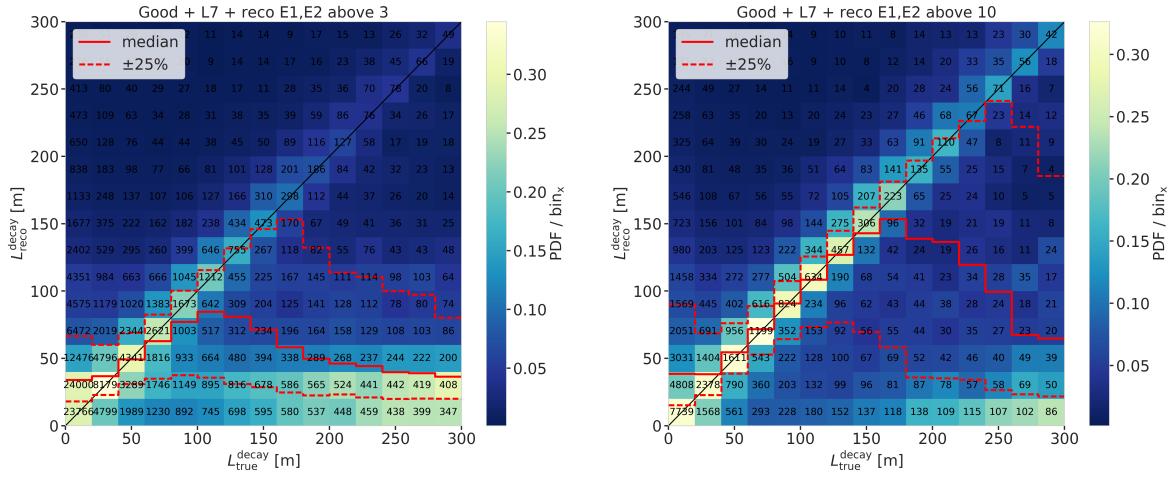
For both cascade resolutions the effect of the reconstruction being biased towards lower values, due to the comparison of the full true energy to the reconstructed EM energy can be seen.

### Length Resolutions

The decay length resolution is also investigated by looking at a similar style of 2-d histograms than for the energies, where the reconstructed decay length is plotted versus the true decay length. The left part of Figure 1.4 shows the distributions after the same selection criteria from Section 1.1.3 are applied. It can be observed that for short true lengths the reconstruction is over-estimating the length, while for long true lengths the reconstruction is strongly under-estimating the length. There is a region between true lengths of 20 m and 80 m where the median reconstruction is almost unbiased, but the 50 % interquartile range is large and increasing from  $\sim 50$  m to  $\sim 70$  m with true decay lengths.

The over-estimation at small true lengths can be explained for multiple reasons, one being that the shortest DOM spacing is  $\sim 7$  m, vertically for DeepCore strings, but mostly larger than that, so resolving lengths below this is very complicated, and the reconstruction tends to be biased towards estimating the length around where the light was observed. Another reason is a similar argument to why the energies are over-estimated at small true values, namely that events that passed the selection and were reconstructed in those cases, probably have an over fluctuation in light deposition, extending further out from the vertices, so the reconstructed length is larger. Additionally, approaching a length of 0.0, the reconstructed length will of course always be a one-sided distribution, because the lengths have to be positive.

The under-estimation at large true lengths is more puzzling, and it seems like the distribution becomes bimodal in the reconstructed lengths, with



**Figure 1.4:** Reconstructed decay length versus true decay length for  $\sim 3\text{ GeV}$  (left) and  $\sim 10\text{ GeV}$  (right) minimum reconstructed cascade energies. The color scale is according to the PDF in each vertical true length slice, with the solid and dashed lines showing the median  $\pm 25\%$  quantiles. The bin entries are shown as numbers.

on population around the diagonal, meaning that they are properly reconstructed, and one population at very short reconstructed lengths, which are badly reconstructed. Above 150 m the badly reconstructed population starts to dominate, and the median resolution drops off strongly. The assumption is that for these events, only one cascade was observed with enough light to be reconstructed, and the reconstruction describes the one observed cascade in two parts, separated by a short distance, driven by similar factors as mentioned before. A check to confirm whether this is the case was to increase the minimum cascade energies to 10 GeV, which is shown in the right part of Figure 1.4. It can be seen that the median resolution is already much better, aligning with the expectation between 40 m and 160 m. Judging from the median resolution and the  $+25\%$  quantile in this range, there is very few events with an over-expectation in the energy, since both of them are alidgning with the diagonal. The spread towards lower reconstructed lengths, on the other hand, is still very large and above 200 m the badly reconstructed population starts to dominate again.

### Badly Reconstructed Cascade Population

#### Ideas to write about:

- ▶ Show/highlight 2d decay length resolution again, basically 2 populations at higher energies
- ▶ only looking at events with  $>10\text{ GeV}$  instead of  $>3\text{ GeV}$  already improves the resolution a lot (low energy and therefore low light is a problem)
- ▶ take a closer look at those populations to find out what's going on
- ▶ multiple things were checked (bad direction (possibly due to seed), cascade position in DC, energies)
- ▶ the direction is worse for the badly reconstructed population, which could be due to a bad seed direction
- ▶ the events are a little further out radially, where the string/DOM spacing is not so tight, which also leads to less light being observed
- ▶ the energy of the second cascade is much lower than the first

- conclusion: it's a combination of all of them, but the main problem is the low energy
- true energies vs true length shows that first cascade energies are much larger than second (pick benchmark lengths and state the values!)

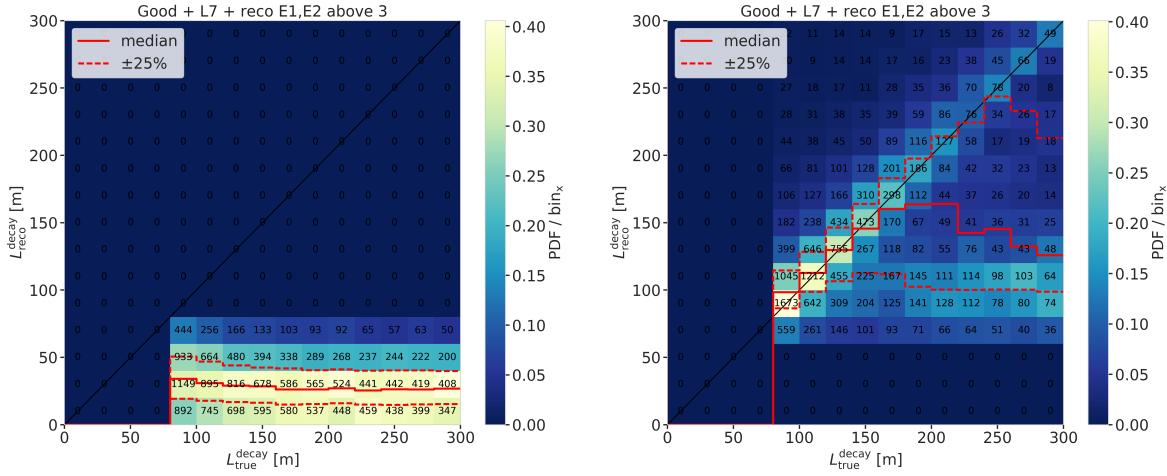


Figure 1.5

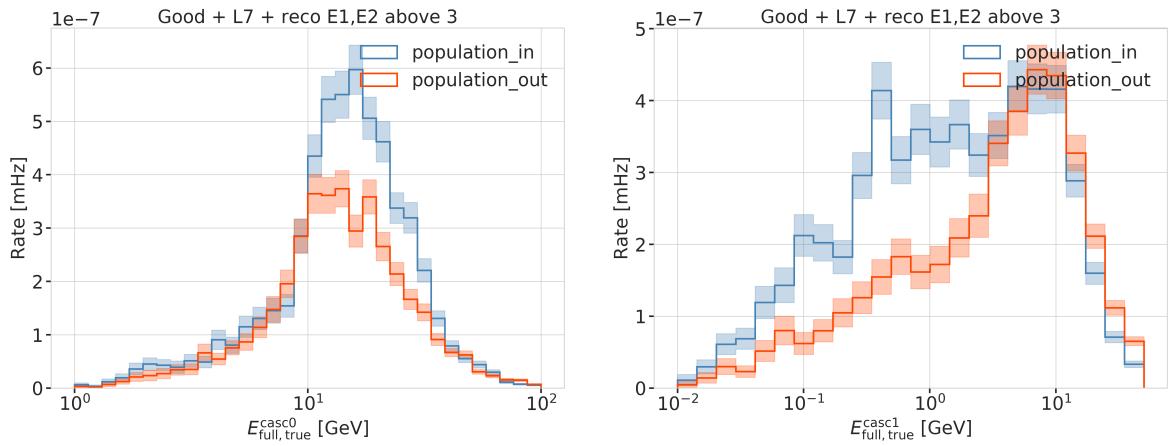


Figure 1.6

make on figure, where  
 re-make normalized?  
 (ORANGE)  
 fix caption (RED)  
 re-make normalized?  
 (ORANGE)  
 fix caption (RED)  
 fix caption (RED)  
 I think from these plots  
 it's probably enough if I  
 state the median energy  
 of first and second cas-  
 cade at final level, as a  
 benchmark for the ener-  
 gies of the two cascades  
 (ORANGE)

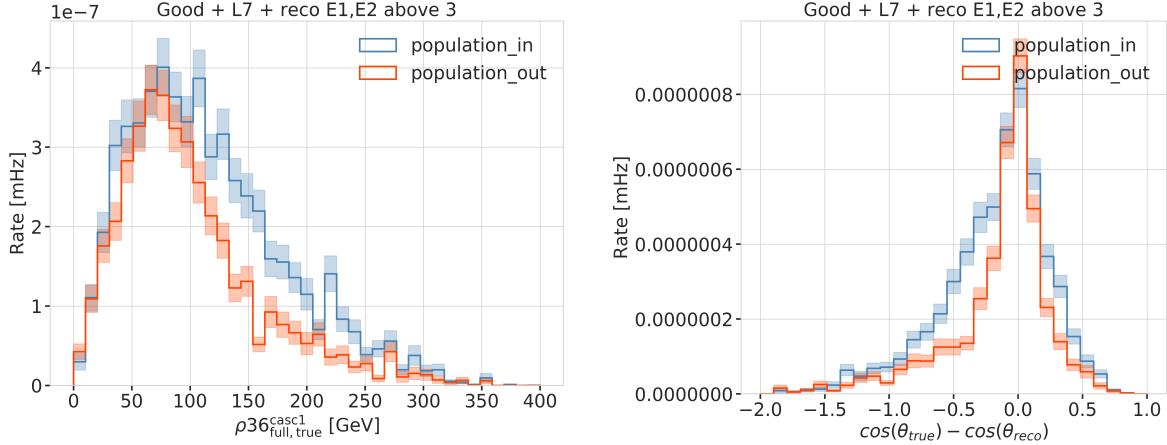
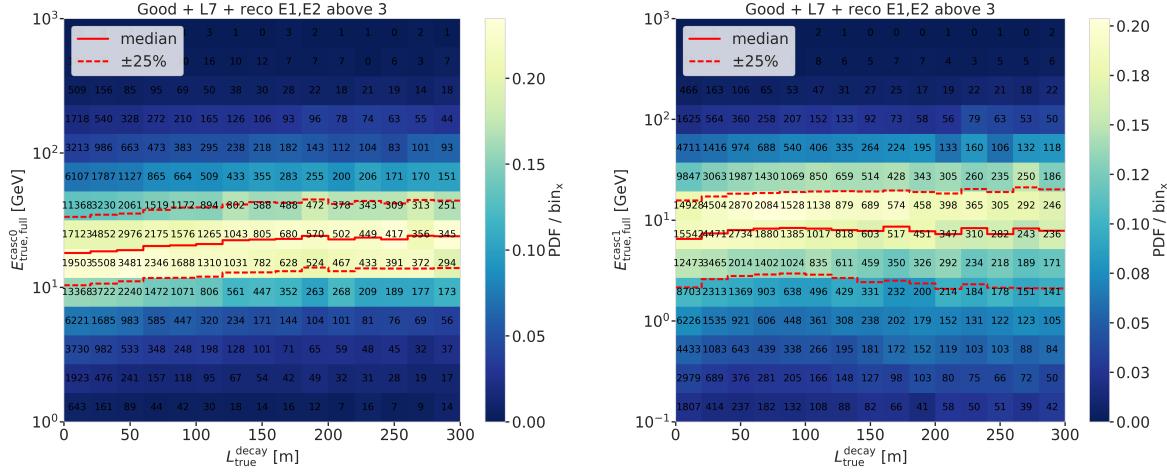


Figure 1.7



## 1.2 Double Cascade Classification

### Missing Points

- ▶ Briefly mention MuMillipede "reconstruction" (depending on how deep I want to go into the classifier I tested?)
- ▶ Calculated variables to input the classifier (distributions?) I guess some examples might do?
- ▶ Cuts applied to make sure the classifier is trained on well reconstructed events (energy, length, true, reco..)
- ▶ Tested classifier (BDT) versions and their performance
- ▶ takeaway? (not sure how to deal with the caveat of the weight being kind of wrong for these..)

## 1.3 Generic Double Cascade Performance

### Ideas:

- ▶ why do these checks again with the model independent samples? (controllable parameter space and distributions, especially energy and

length)

- ▶ benchmark some edge cases (e.g. "best case" of up-going string centered, purely horizontal, and realistic case)
- ▶ investigate where the reconstructions breaks down (in terms of cascade energies, which are the main factor as shown before)

### 1.3.1 Idealistic Performance

- ▶ for the perfect edge case of the event being directly on top of a DeepCore string with an up-going direction the length can be very well reconstructed above a true length of around 20 m
- ▶ 2-d histogram shows that up to a true decay length of ~210 m, there is no under-estimation of the length
- ▶ this means that the under-estimation observed before is only possible if there is no DOMs in between the two cascades
- ▶ this makes a lot of sense, since DOMs being present, but not observing any light is affecting the light expectation that goes into the reconstruction likelihood and therefore makes these hypotheses less likely and therefore incompatible with the data
- ▶ do I want more detailed plots here? I feel like this is enough..

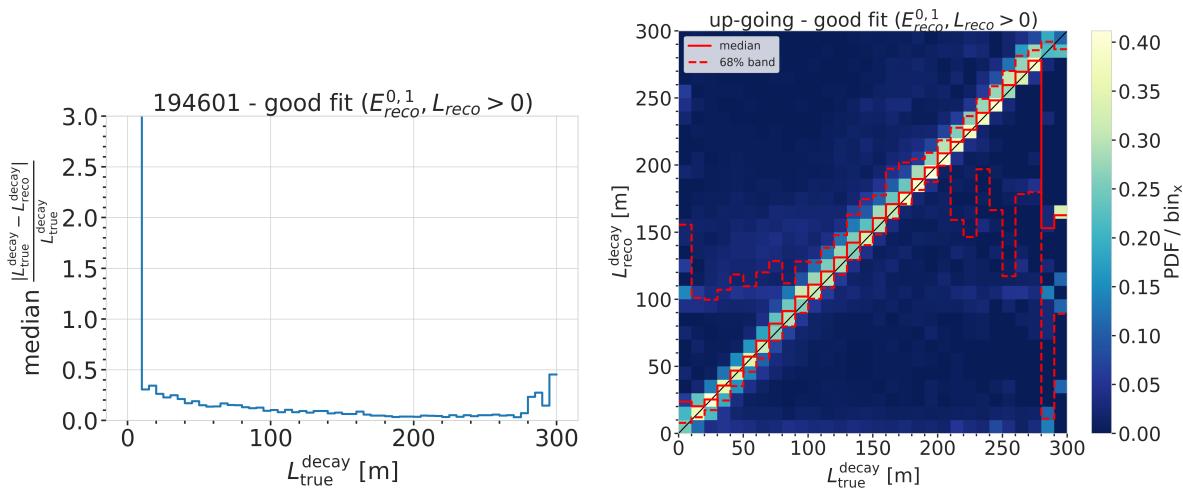


Figure 1.8

fix caption of this figure  
(RED)

### 1.3.2 Realistic Performance

#### Energy/Decay Length Resolution

describe the "good fit" selection when I first show plots that use it (ORANGE)

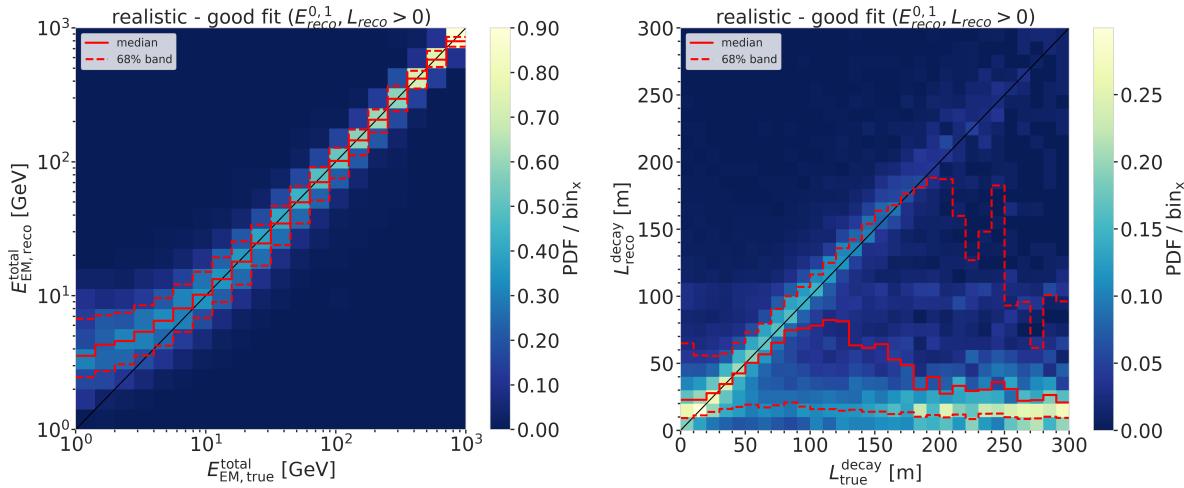
#### 2-D Histograms

add some 1D resolution plots (ORANGE)

#### Things to mention about the 2d-hists:

- ▶ total energy resolution looks very good, above 10 GeV it's almost unbiased and the 1-sigma resolution band is below 20 %

- ▶ individual cascade resolutions mirror this behavior, but are starting to stabilize in energy at lower energies around 5 GeV to 6 GeV with a broader resolution band of 50 %, but reducing drastically with increasing energy (down to 20 % at 100 GeV)
- ▶ interestingly, the second cascade energy reconstruction performs slightly worse, although they have the same energy ranges. This could hint at an asymmetry in the reconstruction process (might relate to how the two cascades are parameterized) or be due to the different positions and the dominantly up-going direction used in the sampling combined with the DOMs looking down (relate this to the sampling distributions explained/shown in the previous chapter)
- ▶ the decay length resolution looks much worse. In the region between 20 m and 80 m it's roughly unbiased, but 1-sigma resolution band is quite wide with a lot of outliers towards short reconstructed lengths. Below 20 m the reconstructed lengths are always over-estimating the true and above 80 m a population of events start to dominate where the decay lengths isn't getting reconstructed at all, which might indicate that one of the cascades wasn't observed. (Relate to the fact that this marginalizes over all energies, meaning also all events which have one cascade with very low energy are included here.)
- ▶ another interesting feature is the band of reconstructed lengths around 100 m, which is probably related to the spacing between most of the strings, which favors the reconstruction to be around this value, because that's the distance at which light can be observed, just from the fact that the DOMs are spaced at this distance (for low energetic cascades, this can dominate the reconstruction)

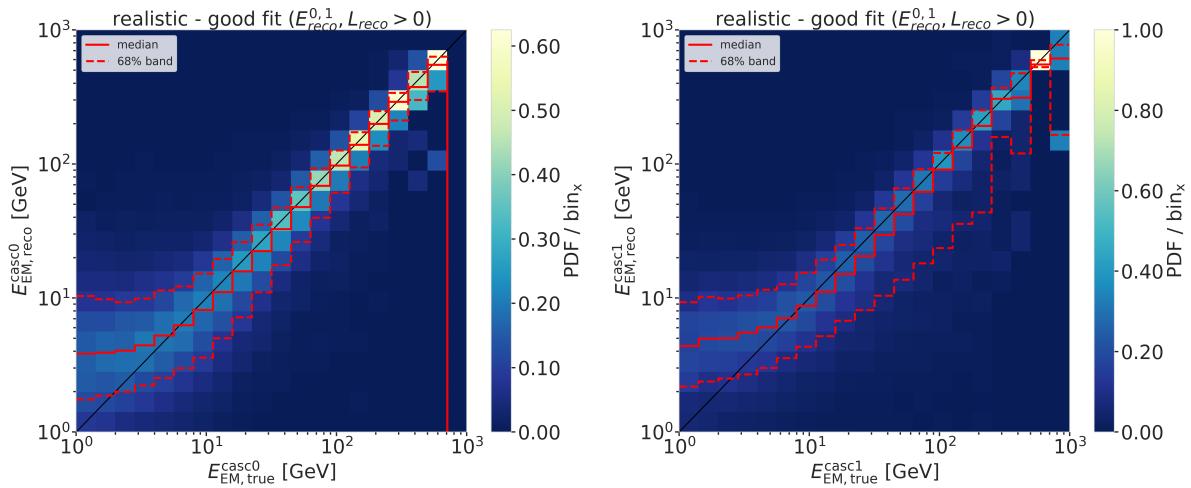
**Figure 1.9**

fix caption of this figure  
(RED)

fix caption of this figure  
(RED)

#### Energy Resolution Things to mention about the energy resolution:

- ▶ Here it can be seen more clearly how the median total energy resolution starts to stabilize around 0.0 at 10 GeV, while for lower energies the reconstruction is over-estimating the true energy. This is a known behavior of energy reconstructions in IceCube, which is mainly due to a selection effect. Only events with a certain amount of light can be reconstructed, which means that the ones with true small energies that are still in the sample are events with over average light production



**Figure 1.10**

due to fluctuations or other effects?

fix caption of this figure  
(RED)

**Decay Length Resolution** Things to mention about the decay length resolution:

- As already mentioned before, the decay length resolution is much worse than the energy resolutions. Figure 1.12 also shows that the median is below 0.0 for short true length and above 0.0 and approaching 1.0 for long true lengths.
- To investigate whether this is really due to the fact that one of the cascades is not observed, the decay length resolution was plotted against the total energy of the event and the minimum energy of the two cascades Figure 1.13.
- It can be seen that the median of the decay length resolution stabilizes at 0.0 for a total energy above 20 GeV, but the spread of the distribution is still quite large with a 1-sigma band of 80 % to 100 %.
- From the plot against the minimum energy it can be seen that the decay length resolution starts to be unbiased for a minimum energy of the cascades of 7 GeV, with an equivalently large spread.
- A preliminary takeaway from this is that the decay length reconstruction is not reliable at all for events with a total energy below 20 GeV or a minimum cascade energy below 7 GeV.

fix caption of this figure  
(RED)

fix caption of this figure  
(RED)

Make plot to show efficiency of the OscNext selection for HNL events.  
(ORANGE)

### 1.3.3 Low Energy Event Selection Efficiency

**Discussion ideas:**

- Show energy, length distribution across the different levels?
- Show efficiency as table across the different levels (MC events + fraction?)
- Compare this to BG efficiency? (maybe rather for the discussion)
- At which level does the selection reduce the HNL the most?
- Is there a place to improve the HNL selection? (Might have to factor in the BG efficiency, as well..)

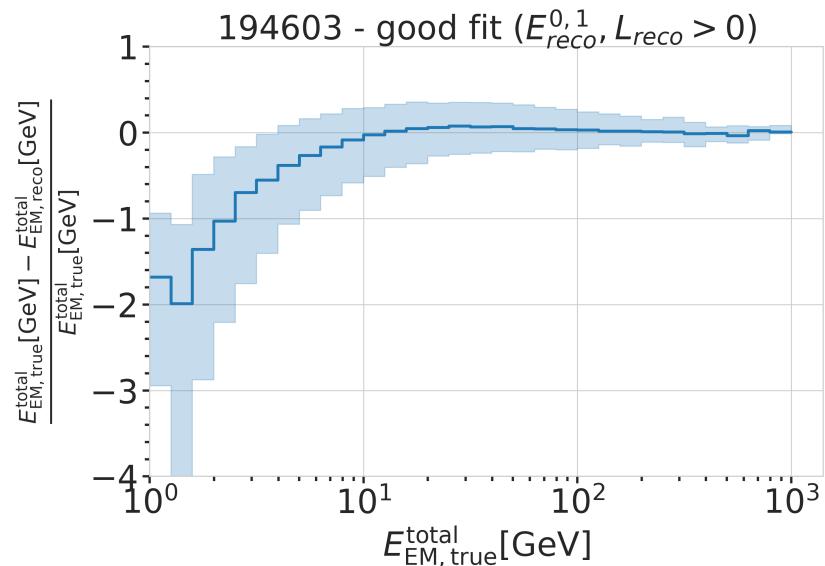


Figure 1.11

- What of this might change with Upgrade? (maybe rather for the discussion)

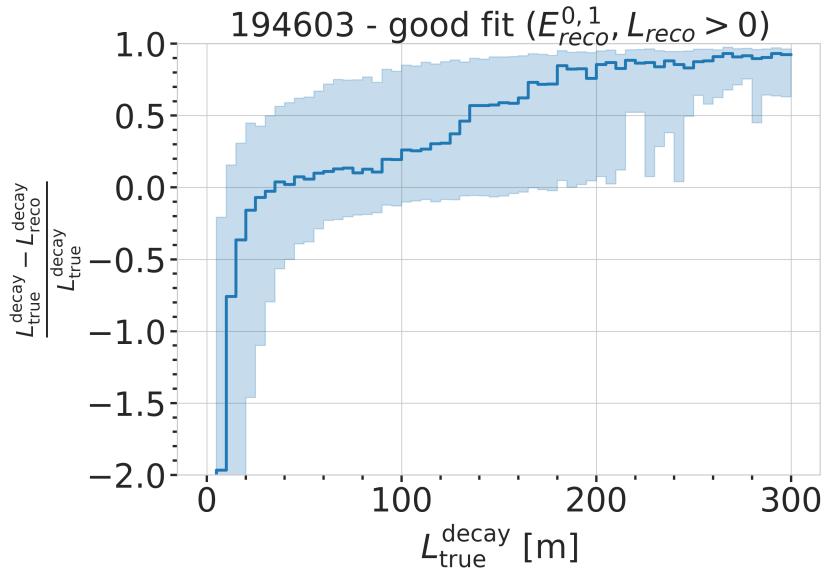


Figure 1.12

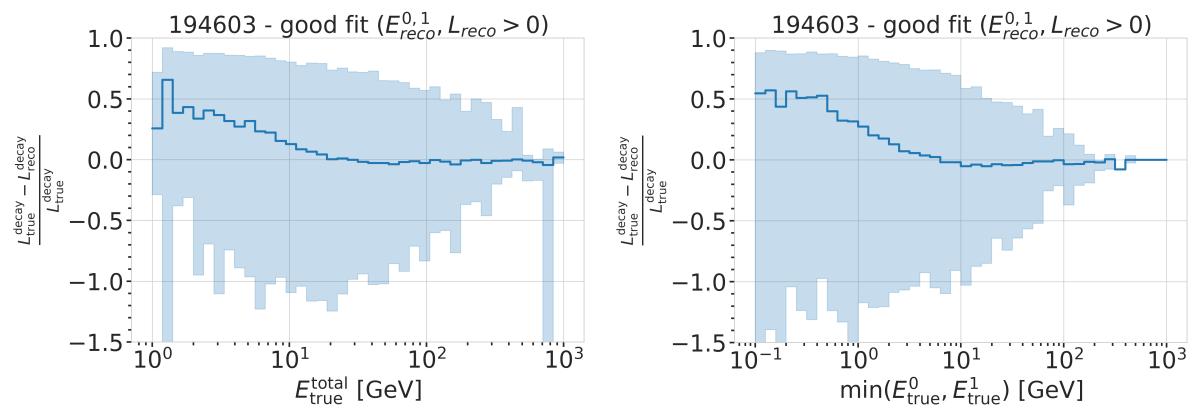


Figure 1.13



# Search for an Excess of Heavy Neutral Lepton Events

# 2

This chapter describes the search for HNL events using 10 years of IceCube DeepCore data. In principle the two physics parameters to be probed are the mass of the HNL,  $m_4$ , and the mixing between the fourth heavy mass state and the SM  $\tau$  sector,  $|U_{\tau 4}|^2$ . Since the mass itself influences the production and decay kinematics of the event and the accessible decay modes, individual mass samples were produced as described in Section ?? . The mass influences the energy distribution, while the mixing both changes the overall scale of the HNL events and the shape in energy and PID. The search is performed for the three mass samples individually, while the mixing is the parameter that can be varied continuously and is measured in the fit.

## 2.1 Final Level Sample

The final level sample of this analysis consists of the neutrino and muon MC introduced in Section ?? and one of the three HNL samples explained in Section ?? . All simulation and the 10 years of IceCube DeepCore data are processed through the full processing and event selection chain described in Section ?? and Section ?? leading to the final level sample. As described in Section ?? , event triggers consisting purely of random coincidences induced by noise in the DOMs have been reduced to a negligible rate, and will not be discussed further.

### 2.1.1 Expected Rates/Events

The rates and the expected number of events for the SM background are shown in Table 2.1 with around 175000 total events expected in the 10 years. The explicit, good detector livetime in this data taking period is 9.28 years. The rates are calculated by summing the weights of all events in the final level sample, while the uncertainties are calculated by taking the square root of the sum of the weights squared. The expected number of events is calculated by multiplying the rate with the livetime. The individual fractions show that this sample is neutrino dominated where the majority of events are  $\nu_\mu$ -CC events.

Type	Rate [mHz]	Events (in 9.28 years)	Fraction [%]
$\nu_\mu^{\text{CC}}$	0.3531	$103321 \pm 113$	58.9
$\nu_e^{\text{CC}}$	0.1418	$41490 \pm 69$	23.7
$\nu_{\text{NC}}$	0.0666	$19491 \pm 47$	11.1
$\nu_\tau^{\text{CC}}$	0.0345	$10094 \pm 22$	5.8
$\mu$	0.0032	$936 \pm 15$	0.5
total	0.5991	$175336 \pm 143$	100.0

Table 2.2 shows the rates and expected number of events for the HNL signal simulation. The expectation depends on the mass and the mixing and shown here are two example mixings for all the three masses that are being tested in this work. A mixing of 0.0 would result in no HNL events at all. It can

2.1 Final Level Sample . . . . .	15
2.2 Statistical Analysis . . . . .	16
2.3 Analysis Checks . . . . .	19
2.4 Results . . . . .	20

add information about the matter profile used (ORANGE)

add information about the oscillation probability calculation and the software used for it (RED)

**Table 2.1:** Final level rates and event expectation of the SM background particle types.

Should I adapt the total numbers to match the sum of the rounded individual parts? (YELLOW)

already be seen that for the smaller mixing of  $|U_{\tau 4}|^2 = 10^{-3}$  the expected number of events is very low, while at the larger mixing of  $|U_{\tau 4}|^2 = 10^{-1}$  the number is comparable to the amount of muons in the background sample.

**Table 2.2:** Final level rates and event expectations of the HNL signal for all three masses and two example mixing values.

HNL mass	Rate [μHz]	Events (in 9.28 years)
$ U_{\tau 4} ^2 = 10^{-1}$		
0.3 GeV	$3.3298 \pm 0.0053$	$974.5 \pm 1.6$
0.6 GeV	$3.0583 \pm 0.0058$	$895.0 \pm 1.7$
1.0 GeV	$2.4988 \pm 0.0059$	$731.3 \pm 1.7$
$ U_{\tau 4} ^2 = 10^{-3}$		
0.3 GeV	0.0057	$1.67 \pm 0.01$
0.6 GeV	0.0220	$6.44 \pm 0.01$
1.0 GeV	0.0248	$7.27 \pm 0.01$

## 2.1.2 Analysis Binning

[7]: Yu et al. (2023), "Recent neutrino oscillation result with the IceCube experiment"

An identical binning to the analysis performed in [7] is used. It was chosen such that the track-like bin has the largest  $\nu_\mu$ -CC fraction. Extending the binning towards lower energies or increasing the number of bins in energy or cosine of the zenith angle did not improve the HNL sensitivities significantly. It also has to be considered that sufficient data events need to end up in the individual bins to result in a good fit, which was already investigated in the previous analysis. To mitigate the low data statistics, a few bins were not taken into account in the analysis. There are three bins in PID (cascade-like, mixed, and track-like), 12 bins in reconstructed energy, and 8 bins in cosine of the reconstructed zenith angle as specified in Table 2.3. Originally, there were two more bins in  $\cos(\theta)$ , which were removed to reduce muons coming from the horizon and some low energy bins in the cascade-like bin are removed due to the low event statistics.

**Table 2.3:** Three dimensional binning used in the analysis. All variables are from the FLERCNN reconstruction explained in Section ??.

Variable	N <sub>bins</sub>	Edges	Step
$P_\nu$	3	[0.00, 0.25, 0.55, 1.00]	linear
$E$	12	[5.00, 100.00]	logarithmic
$\cos(\theta)$	8	[-1.00, 0.04]	linear

add 3D expectation and/or S/sqrt(B) plots

Add fractions of the different particle types in the bins for benchmark mass/mixing (another table?) (ORANGE)

I feel like I have to be a bit more precise on what is the fit metric (e.g. the mod chi2) and what is the TS, as in the mod chi2 difference, which is the actual TS, right? (RED)

## 2.2 Statistical Analysis

### 2.2.1 Test Statistic

The measurements are performed by comparing the weighted MC to the data. Through variation of the nuisance and physics parameters that govern the weights, the best matching set of parameters can be found. The comparison is done using a modified  $\chi^2$  defined as

$$\chi^2_{\text{mod}} = \sum_{i \in \text{bins}} \frac{(N_i^{\text{exp}} - N_i^{\text{obs}})^2}{N_i^{\text{exp}} + (\sigma_i^\nu)^2 + (\sigma_i^\mu)^2 + (\sigma_i^{\text{HNL}})^2} + \sum_{j \in \text{syst}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2}, \quad (2.1)$$

as the *test statistic* (TS). The total even expectation is  $N_i^{\text{exp}} = N_i^\nu + N_i^\mu + N_i^{\text{HNL}}$ , where  $N_i^\nu$ ,  $N_i^\mu$ , and  $N_i^{\text{HNL}}$  are the expected number of events in bin  $i$  from

neutrinos, atmospheric muons, and HNLs, while  $N_i^{\text{obs}}$  is the observed number of events in bin  $i$ . The expected number of events from each particle type is calculated by summing the weights of all events in the bin  $N_i^{\text{type}} = \sum_i^{\text{type}} \omega_i$ , with the statistical uncertainty being  $(\sigma_i^{\text{type}})^2 = \sum_i^{\text{type}} \omega_i^2$ . The additional term in Equation 2.1 is included to apply a penalty term for prior knowledge of the systematic uncertainties of the parameters where they are known.  $s_j$  are the systematic parameters that are varied in the fit, while  $\hat{s}_j$  are their nominal values and  $\sigma_{s_j}$  are the known uncertainties.

## 2.2.2 Physics Parameters

The variable physics parameter in this analysis is the mixing between the HNL and the SM  $\tau$  sector,  $|U_{\tau 4}|^2$ . It can be changed continuously in the range of  $[0.0, 1.0]$  by applying the weighting scheme described in Section ???. The fit is initialized at an off-nominal value of 0.1. The other physics parameter, the mass  $m_4$  of the HNL, is fixed to one of the three discrete masses to be tested, by using the corresponding sample of the HNL simulation described in Section ??.

## 2.2.3 Nuisance Parameters

To decide which systematic uncertainties should be included in the fit, we test the potential impact they have on the TS if they are neglected. The test is performed by creating Asimov data using the BG simulation and the HNL simulation of the 1.0 GeV mass sample at a mixing value of 0.1, which is chosen as a benchmark physics parameter and does not have a significant impact on the test. The systematic parameter of interest is set to a value above its nominal expectation, either pulled up by  $+1\sigma$  or by an educated estimate for parameters without a well-defined uncertainty. A fit is performed fixing the systematic parameter of interest and leaving all additional parameters free. The resulting TS is the mis-modeling significance between this fit and a fit with all parameters free, which would result in a TS of 0.0 for this Asimov test. Parameters below a significance of  $0.1\sigma$  are fixed and the test is performed in an iterative manner until the final set of free parameters is found. Figure 2.1 shows the resulting significances of one of these tests. In the final selection of free parameters the Barr  $h_{\pi^+}$  parameter was also left free and the ice absorption is still kept free, despite showing a small significance. This is done because the bulk ice parameters are not well constrained and are known to have a large impact, which might be concealed in the test, due to correlations with the other parameters.

The scaling parameter  $N_\nu$  is included to account for the unknown overall normalization of the neutrino rate. It has the identical effect on the SM neutrino events and the BSM HNL events and its nominal value is set to 1.0 with a wide range of  $[0.1, 2.0]$ .

Concerning the atmospheric neutrino flux, the CR power law flux correction factor  $\Delta\gamma_\nu$  is included with nominal value of 0.0 and a range of  $[-0.5, 0.5]$ . The nominal value corresponds to a CR power law of  $E^{-2}$  and a slightly conservative prior of 0.1 is applied to the parameter, while latest measurements have an uncertainty of 0.05 [8].

Do I want/need to include the description of the KDE muon estimation? (YELLOW)

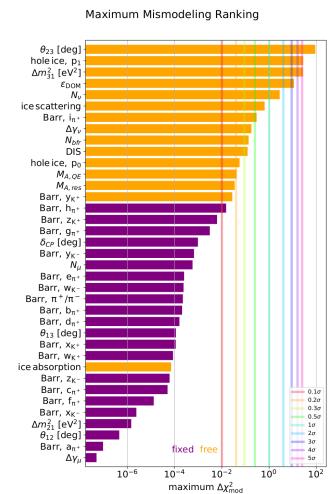


Figure 2.1: "calculated at a mixing of 0.1 and for the 1.0 GeV sample"

Blow up labels/legend/title and make it more readable in the margin or move it into the main text? (RED)

I don't like this formulation, but don't know better right now.. (YELLOW)

elaborate why this is also done to cover the whole energy range for the pion production, referencing the Barr Block plot that I haven't included yet :D (RED)

I truly dislike this sentence, too, better ideas? (YELLOW)

I'm just writing out the data from the table, but I need to mention/motivate the included priors here and maybe just point to the table for the ranges/nominal values? (Not quite sure about this) (RED)

**Table 2.4:** Systematic uncertainty parameters that are left free to float in the fit. Their allowed fit ranges are shown with the nominal value and the Gaussian prior width if applicable.

Parameter	Nominal	Range	Prior
$\Delta\gamma_\nu$	0.0	[-0.5, 0.5]	0.1
Barr $h_{\pi^+}$	0.0	[-0.75, 0.75]	0.15
Barr $i_{\pi^+}$	0.0	[-3.05, 3.05]	0.61
Barr $y_{K^+}$	0.0	[-1.5, 1.5]	0.3
$\theta_{23}[\circ]$	47.5047	[0.0, 90.0]	-
$\Delta m_{31}^2 [\text{eV}^2]$	0.002475	[0.001, 0.004]	-
DIS	0.0	[-0.5, 1.5]	1.0
$N_\nu$	1.0	[0.1, 2.0]	-
$\epsilon_{\text{DOM}}$	1.0	[0.8, 1.2]	0.1
hole ice $p_0$	0.101569	[-0.6, 0.5]	-
hole ice $p_1$	-0.049344	[-0.2, 0.2]	-
bulk ice absorption	1.0	[0.85, 1.15]	-
bulk ice scattering	1.05	[0.9, 1.2]	-
$N_{\text{bfr}}$	0.0	[-0.2, 1.2]	-
$M_{A,\text{QE}}$	0.0	[-2.0, 2.0]	1.0
$M_{A,\text{res}}$	0.0	[-2.0, 2.0]	1.0

[8]: Evans et al. (2017), "Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries"

say something about their priors (RED)

Additionally, the Barr  $h_{\pi^+}$ , Barr  $i_{\pi^+}$ , and Barr  $y_{K^+}$  parameters of the pion and kaon production uncertainties are included with nominal values of 0.0 and ranges of [-0.75, 0.75], [-3.05, 3.05], and [-1.5, 1.5], respectively.

From the cross-section uncertainties introduced in Section ??, all three parameters, DIS,  $M_{A,\text{QE}}$ , and  $M_{A,\text{res}}$  are included in the fit with nominal values of 0.0 for all of them and range [-0.5, 1.5] for DIS and [-2.0, 2.0] for the axial mass parameters  $M_{A,\text{QE}}$ , and  $M_{A,\text{res}}$ .

[9]: Feintzeig (2014), "Searches for Point-like Sources of Astrophysical Neutrinos with the IceCube Neutrino Observatory"

[10]: Kulacz (2019), "In Situ Measurement of the IceCube DOM Efficiency Factor Using Atmospheric Minimum Ionizing Muons"

cite?! (ORANGE)

All the detector systematic uncertainties are included in the fit. The DOM efficiency  $\epsilon_{\text{DOM}}$  has a nominal value of 1.0 and a range of [0.8, 1.2]. It is constrained by a Gaussian prior with a width of 0.1, which is a conservative estimate based on the studies of the optical efficiency using minimum ionizing muons from [9, 10]. The hole ice model parameters  $p_0$  and  $p_1$  are included with nominal values of 0.101569 and -0.049344, respectively, and ranges of [-0.6, 0.5] and [-0.2, 0.2]. The bulk ice absorption and scattering parameters are included with nominal values of 1.0 and 1.05, respectively, and ranges of [0.85, 1.15] and [0.9, 1.2]. They are unconstrained in the fit and the ranges are set to be conservative determined from calibration data

The two atmospheric neutrino oscillation parameters  $\theta_{23}$  and  $\Delta m_{31}^2$  are also included in the fit with nominal values of 47.5047° and  $2.475 \times 10^{-3} \text{ eV}^2$ , respectively. Since they govern the shape and the strength of the tau neutrino flux, by defining the oscillation from  $\nu_\mu$  to  $\nu_\tau$ , they are also relevant for the HNL signal shape. Their ranges are set to [0.0°, 90.0°] and [0.001 eV<sup>2</sup>, 0.004 eV<sup>2</sup>].

I should add some final level effects of some systematics on the 3D binning and maybe discuss how they are different from the signal shape, or so? (ORANGE)

[11]: Aartsen et al. (2020), "Computational techniques for the analysis of small signals in high-statistics neutrino oscillation experiments"

## 2.2.4 Low Energy Analysis Framework

The analysis is performed using the PISA [11] [12] software framework, which was developed to perform analyses "of small signals in high-statistics neutrino oscillation experiments". It is used to generate the expected event distributions from several MC samples, which can then be compared to the observed data. The expectation for each sample is calculated in parallel, applying physics and nuisance parameter effects in a stage-wise manner, before combining the final expectation from all the samples.

## 2.3 Analysis Checks

Fitting to data is performed in a *blind* manner, where the analyzer does not immediately see the fitted physics and nuisance parameter values, but first checks that a set of pre-defined *goodness of fit* (GOF) criteria are fulfilled. This is done to circumvent the so-called *confirmation bias* [13], where the analyzer might be tempted to construct the analysis in a way that confirms their expectation. After the GOF criteria are met to satisfaction, the fit results are unblinded and the full result can be revealed. Before these blind fits to data are performed, the robustness of the analysis method is tested using pseudo-data that is generated from the MC.

Do I want more information about the different pipelines and stages?  
Could link back to the extra stage I wrote and add the earth model and oscillation calculation information here, I guess?! (ORANGE)

[13]: Nickerson (1998), "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises"

### 2.3.1 Minimization Robustness

To find the set of parameters that best describes the data, a staged minimization routine is used. In the first stage, a fit with coarse minimizer settings is performed to find a rough estimate of the *best fit point* (BFP). In the second stage, the fit is performed again in both octants<sup>1</sup> of  $\theta_{23}$ , starting from the BFP of the coarse fit. For each individual fit the *MIGRAD* routine of *IMINUIT* [14] is used to minimize the  $\chi^2_{\text{mod}}$  TS defined in Equation 2.1. *Iminuit* is a fast, python compatible minimizer based on the *MNUT2* C++ library [15]. The individual minimizer settings for both stages are shown in Table 2.5.

To test the minimization routine and to make sure it consistently recovers any physics parameters, pseudo-data sets are produced from the MC by choosing the nominal nuisance parameters and specific physics parameters, without adding any statistical or systematic fluctuations to it. These so-called *Asimov*<sup>2</sup> data sets are then fit back with the full analysis chain. This type of test is called *Asimov inject/recover test*. A set of mixing values between  $10^{-3}$  and  $10^0$  is injected and fit back. Without fluctuations the fit is expected to always recover the injected parameters (both physics and nuisance parameters). The fitted mixing values from the Asimov inject/recover tests are compared to the true injected values in Figure 2.2 for the 0.6 GeV sample. As expected, the fit is always able to recover the injected physcis parameter and the nuisance paramters. The same is true for the other mass samples and the additional plots for the other mass samples can be found in Section B.1.

1: There is a degeneracy between the lower octant ( $\theta_{23} < 45^\circ$ ) and the upper octant ( $\theta_{23} > 45^\circ$ ), which can lead to TS minima (local and global) at two positions that are mirrored around  $45^\circ$  in  $\theta_{23}$ .

[14]: Dembinski et al. (2022), *scikit-hep/iminuit*: v2.17.0

again, I think fit metric and TS are mixed up a bit (RED)

[15]: James et al. (1975), "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations"

Fit	Err.	Prec.	Tol.
Coarse	1e-1	1e-8	1e-1
Fine	1e-5	1e-14	1e-5

Table 2.5: Migrad settings for the two stages in the minimization routine. Err. are the step size for the numerical gradient estimation, Prec. is the precision with which the LLH is calculated, and Tol. is the tolerance for the minimization.

2: A pseudo-data set without statistical fluctuations is called Asimov data set.

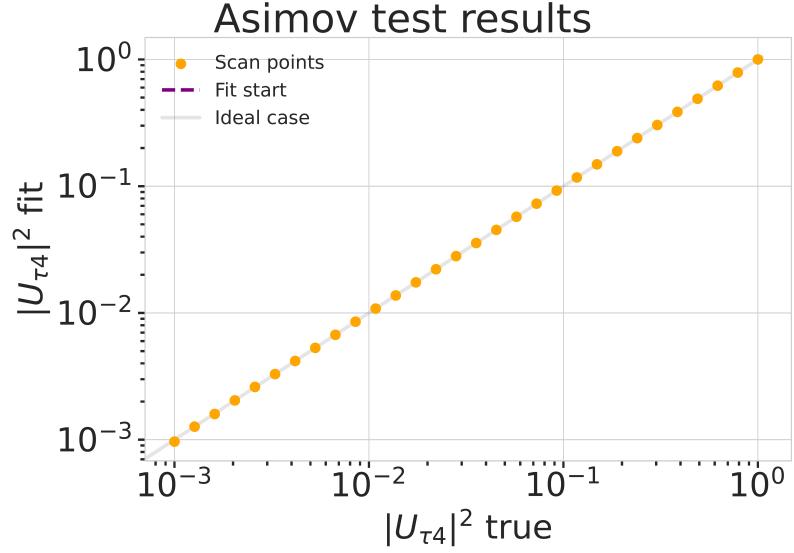
Do I want additional plots for this (fit diff, LLH distr, minim. stats, param. fits)? (YELLOW)

here again, this is just the fit metric, right? (RED)

Add bin-wise TS distribution? Add 3D TS maps? (RED)

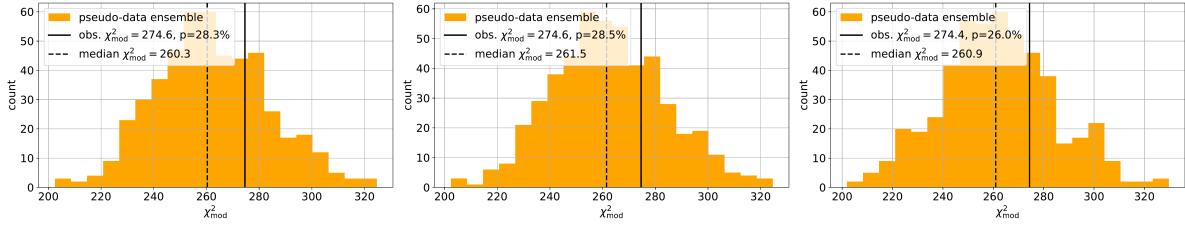
### 2.3.2 Goodness of Fit

To estimate the GOF, pseudo-data is generated from the MC by injecting the BFP parameters as true parameters and then fluctuating the expected bin counts to account for MC uncertainty and Poisson fluctuations in data. First, the expectation value of each bin is drawn from a Gaussian distribution centered at the nominal expectation value with a standard deviation corresponding to the MC uncertainty of the bin. Based on this sampled expectation value, each bin count is drawn from a Poisson distribution, independently, to get the final pseudo-data set. These pseudo-data sets are then fit back with the analysis chain. By comparing the distribution of TS values from this *ensemble* of pseudo-data trials to the TS of the fit to real data, a p-value can be calculated. The p-value is the probability of finding a TS value at least as large as the one from the data fit. Figure 2.3 shows the TS distribution from the ensemble tests for the 0.6 GeV mass sample



**Figure 2.2:** Asimov inject/recover test for the 0.6 GeV mass sample. Mixing values between  $10^{-3}$  and  $10^0$  are injected and fit back with the full analysis chain. The injected parameter is always recovered within the statistical uncertainty.

and the observed TS value from the fit, resulting in a p-value of 28.5 %. The p-values for the 0.3 GeV and 1.0 GeV are 28.3 % and 26.0 %, respectively, and the corresponding plots are shown in Section ???. Based on this test, it is concluded that the fit result is compatible with the expectation from the ensemble of pseudo-data trials.



**Figure 2.3:** Observed fit TS and TS distribution from pseudo-data trials for the 0.6 GeV mass sample.

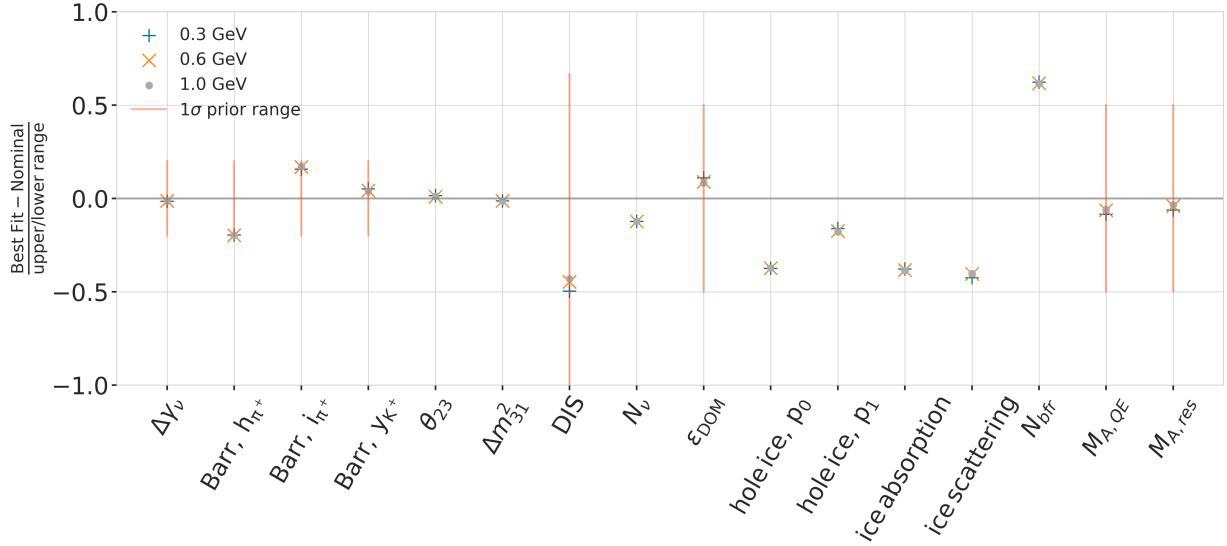
fix dimension to fit them  
in one row! (RED)

## 2.4 Results

### 2.4.1 Best Fit Nuisance Parameters

The resulting nuisance parameter values from the fits are illustrated in Figure 2.4, where the differences to the nominal values are shown, normalized by the distance to the closest boundary. The results from all three fits are shown in the same plot and the fits prefer values of the same size for all three mass samples. For parameters that had a Gaussian prior, the  $1\sigma$  range is also displayed. As was already confirmed during the blind fit procedure, all fitted parameters are within this range, but the Barr  $h_{\pi^+}$  parameter is smaller and the Barr  $i_{\pi^+}$  is larger than expected, both being very close within the  $+1\sigma$  and the  $-1\sigma$  range, respectively. The DIS parameter fits to a smaller value than the nominal and all ice parameters, both hole ice  $p_0$ , and  $p_1$  as well as bulk ice absorption, and scattering are found at values lower than the nominal. The effective ice model parameter,  $N_{\text{bfr}}$ , prefers a value of  $\sim 0.74$ , indicating that the data is more *BFR*-like (value of 1.0) than *Spice 3.2.1*-like (value of 0.0). For completeness's sake, the explicit results are listed in Table

C.1. There, the nominal values and the absolute differences to the best fit value are also presented.



**Figure 2.4:** Best fit nuisance parameter distances to the nominal values, normalized by the distance to the closest boundary. For parameters with a Gaussian prior, the  $+1\sigma$  range is also shown.

## 2.4.2 Best Fit Parameters and Limits

The fitted mixing values are

$$\begin{aligned}|U_{\tau 4}|^2(0.3 \text{ GeV}) &= 0.003^{+0.084}, \\|U_{\tau 4}|^2(0.6 \text{ GeV}) &= 0.080^{+0.134}, \text{ and} \\|U_{\tau 4}|^2(1.0 \text{ GeV}) &= 0.106^{+0.132},\end{aligned}$$

with their  $+1\sigma$  uncertainty. All of them are compatible with the null hypothesis of 0.0 mixing, although the 0.6 GeV and 1.0 GeV fits indicate a mixing value around 0.09. The best fit mixing values and the corresponding upper limits at 68 % and 90 % confidence level (CL) are listed in Table 2.6, also showing the CL at the null hypothesis, which is the probability of excluding the null hypothesis with this fit. The CLs are estimated by assuming that Wilks' theorem [16] holds, meaning that the TS follows a  $\chi^2$  distribution with one degree of freedom.

HNL mass	$ U_{\tau 4} ^2$	68 % CL	90 % CL	NH p-value
0.3 GeV	0.003	0.087	0.194	0.97
0.6 GeV	0.080	0.214	0.355	0.79
1.0 GeV	0.106	0.238	0.396	0.63

sort the variables also by type, same as in the table "best\_fit\_parameters"? (ORANGE)

Show best fit hole ice angular acceptance compared to nominal and flasher/in-situ fits, maybe? (YELLOW)

Discuss what it means that the parameters are at these values? Here, or somewhere else? (RED)

[16]: Wilks (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses"  
**Table 2.6:** xx

fix table caption (RED)

Figure 2.5 shows the observed likelihood profile for the 0.6 GeV sample, which is the difference in  $\chi^2_{\text{mod}}$  between the best fit and each scan point in  $|U_{\tau 4}|^2$ . Also shown is the expected likelihood profile, based on a scan over Asimov data produced at the BFP. The observed CLs are slightly larger/smaller than the expected CL. To ensure this is compatible with random fluctuations, the expected likelihood is also profiled for 100 pseudo-data sets, which are generated at the BFP and then fluctuated using both Poisson and Gaussian fluctuations, to include the data and the MC uncertainty, as

fix once I have them produced (RED)

fix once I have the brazil bands (RED)

was already done for the ensemble tests. The resulting CLs are shown as the colored areas and the observed contour is well within the 68 % band, confirming that it is compatible with data fluctuations. Figure ?? shows the same likelihood profiles and bands for the other two mass samples. For both of them the observed CLs are also slightly larger/smaller than the expected, but still within the 68 % band of the pseudo-data trials, so they are also compatible with random fluctuations.

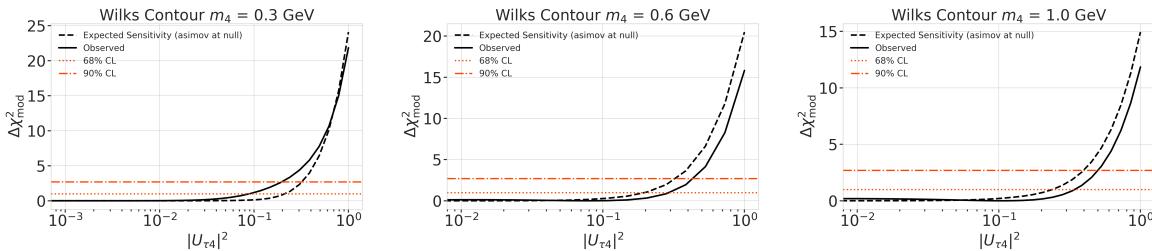


Figure 2.5: xx  
fix caption for this figure (RED)

make plot with BFPs and limit, comparing to upper limits from other experiments, think about how to present it (RED)

specify which they are, once I have them (RED)

add 1-d data/mc agreement for example mass sample (0.6?) and all 3 analysis variables (RED)

add table with reduced chi2 for all 1-d distributions (RED)

### 2.4.3 Data/MC Agreement

At the BFP, the agreement between the data and simulation is probed by comparing the 1-dimensional analysis distributions for PID, energy, and cosine of the zenith angle. As an example, two distributions for the 0.6 GeV mass sample are shown in Figure ?? . The data is compared to the total MC expectation, which is also split up into its composing parts. Good agreement can be observed in the pull distributions and is quantified by a reduced  $\chi^2$ , which is close to 1.0 for all distributions. The reduced  $\chi^2$  for all investigated distributions is listed in Table ?? , while the distributions themselves can be found in Section ?? .

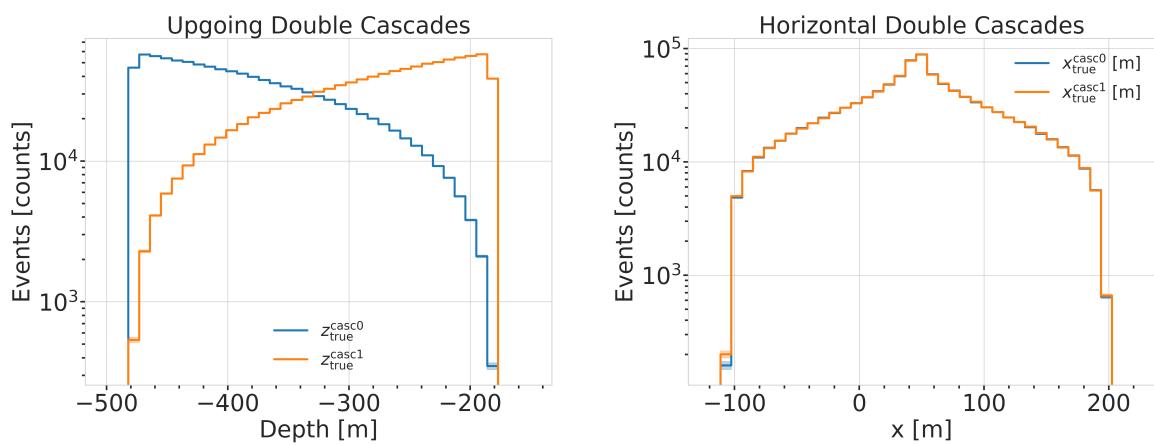
## **APPENDIX**



# A

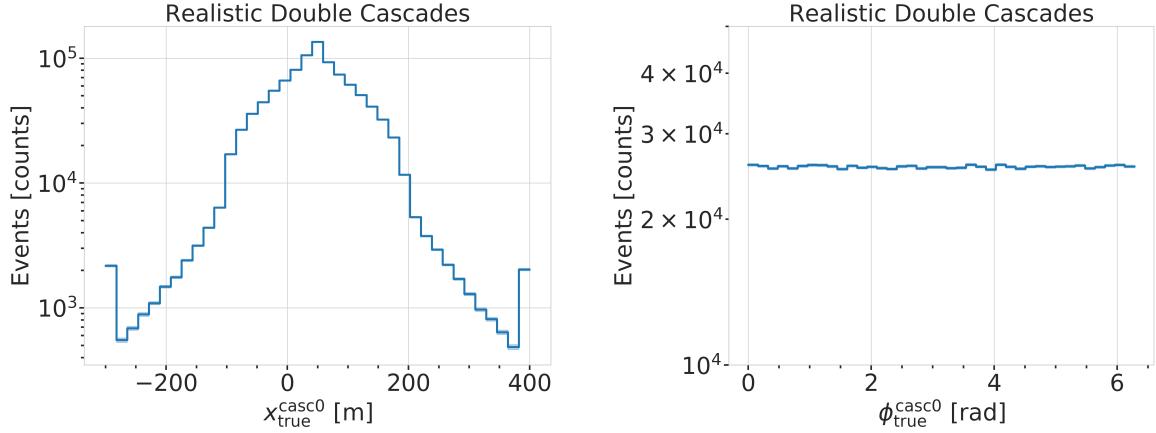
## Heavy Neutral Lepton Signal Simulation

### A.1 Model Independent Simulation Distributions



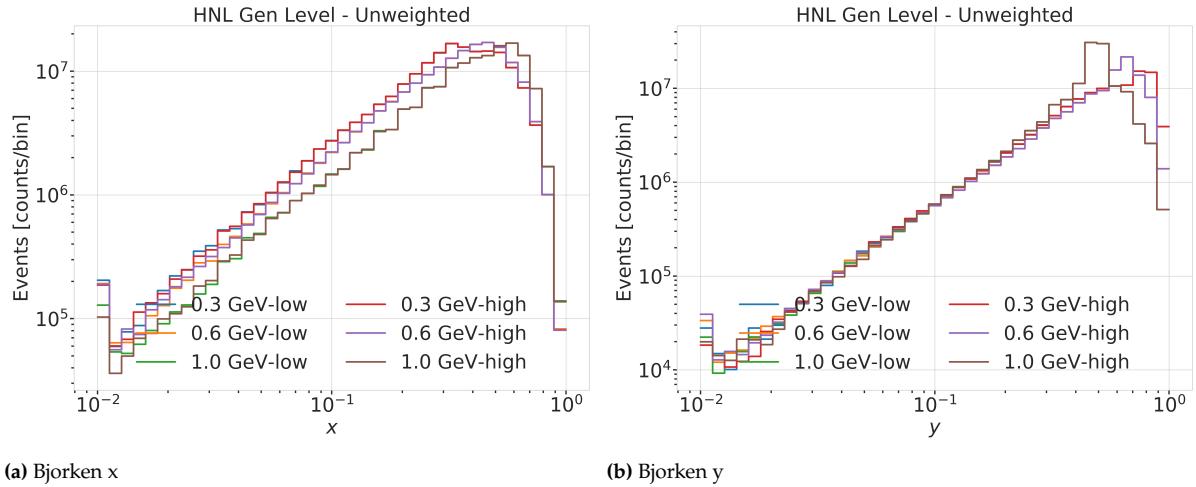
**Figure A.1:** Generation level distributions of the simplistic simulation sets. Vertical positions (left) and horizontal positions (right) of both sets are shown.

- Re-make plot with  $x, y$  for horizontal set one plot!
- Re-make plot with  $x, y, z$  for both cascades in one.
- Re-arrange plots in a more sensible way.



**Figure A.2:** Generation level distributions of the realistic simulation set. Shown are the cascade  $x, y, z$  positions (left) and direction angles (right).

## A.2 Model Dependent Simulation Distributions



**Figure A.3:** Generation level distributions of the model dependent simulation.

# B

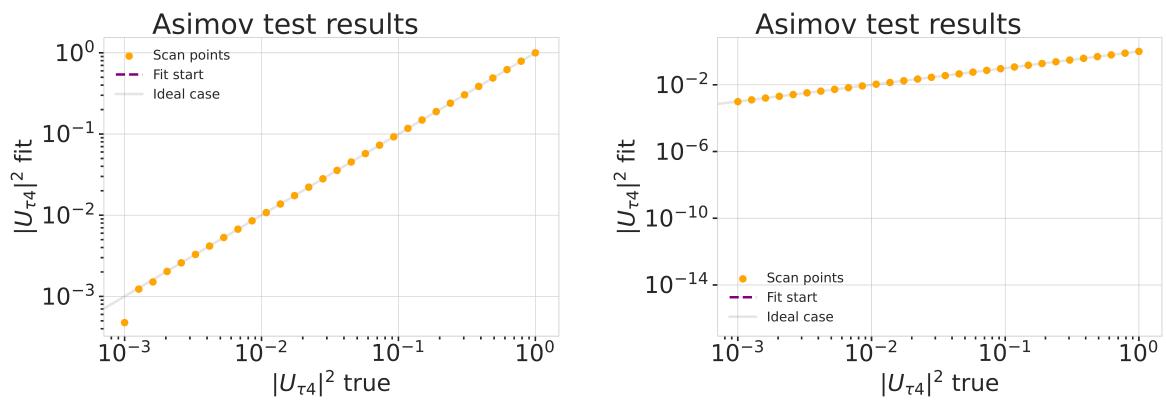
---

## Analysis Checks

---

### B.1 Minimization Robustness

Figure B.1 shows additional Asimov inject/recover tests for the 0.3 GeV and the 1.0 GeV mass sets. The tests were described in Section 2.3.1.



**Figure B.1:** Asimov inject/recover test for the 0.3 GeV (left) and the 1.0 GeV (right) mass sets. Mixing values between  $10^{-3}$  and  $10^0$  are injected and fit back with the full analysis chain. The injected parameter is always recovered within the statistical uncertainty.



# C

---

## Analysis Results

---

### C.1 Best Fit Nuisance Parameters

Table C.1: xx

Parameter	Nominal	Best Fit			Nominal - Best Fit		
		0.3 GeV	0.6 GeV	1.0 GeV	0.3 GeV	0.6 GeV	1.0 GeV
$\Delta\gamma_\nu$	0.000000	-0.007926	-0.006692	-0.006596	0.007926	0.006692	0.006596
Barr $h_{\pi^+}$	0.000000	-0.147475	-0.148481	-0.148059	0.147475	0.148481	0.148059
Barr $i_{\pi^+}$	0.000000	0.475448	0.513393	0.521626	-0.475448	-0.513393	-0.521626
Barr $y_{K^+}$	0.000000	0.076176	0.062893	0.057548	-0.076176	-0.062893	-0.057548
$\theta_{23}[\circ]$	47.504700	48.117185	47.918758	48.010986	-0.612485	-0.414058	-0.506286
$\Delta m_{31}^2 [\text{eV}^2]$	0.002475	0.002454	0.002454	0.002455	0.000020	0.000021	0.000019
DIS	0.000000	-0.248709	-0.223302	-0.215666	0.248709	0.223302	0.215666
$N_\nu$	1.000000	0.889149	0.889055	0.889559	0.110851	0.110945	0.110441
$ U_{\tau 4} ^2$	0.100000	0.003019	0.080494	0.106141	0.096981	0.019506	-0.006141
$\epsilon_{\text{DOM}}$	1.000000	1.021984	1.017789	1.016689	-0.021984	-0.017789	-0.016689
hole ice $p_0$	0.101569	-0.161341	-0.161051	-0.160129	0.262910	0.262620	0.261698
hole ice $p_1$	-0.049344	-0.073701	-0.075596	-0.076261	0.024357	0.026252	0.026917
ice absorption	1.000000	0.943261	0.942463	0.942000	0.056739	0.057537	0.058000
ice scattering	1.050000	0.986152	0.989289	0.989438	0.063848	0.060711	0.060562
$N_{\text{bfr}}$	0.000000	0.746684	0.740255	0.736215	-0.746684	-0.740255	-0.736215
$M_{A,\text{QE}}$	0.000000	-0.170528	-0.128150	-0.120345	0.170528	0.128150	0.120345
$M_{A,\text{res}}$	0.000000	-0.125855	-0.080875	-0.070716	0.125855	0.080875	0.070716

sort these by type of nuisance parameter?



# Bibliography

Here are the references in citation order.

- [1] R. Abbasi et al. "Measurement of Astrophysical Tau Neutrinos in IceCube's High-Energy Starting Events". In: *arXiv e-prints* (Nov. 2020) (cited on page 1).
- [2] M. Usner. "Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector". PhD thesis. Berlin, Germany: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018. doi: [10.18452/19458](https://doi.org/10.18452/19458) (cited on page 1).
- [3] P. Hallen. "On the Measurement of High-Energy Tau Neutrinos with IceCube". MA thesis. Aachen, Germany: RWTH Aachen University, 2013 (cited on page 1).
- [4] N. Whitehorn, J. van Santen, and S. Lafebre. "Penalized splines for smooth representation of high-dimensional Monte Carlo datasets". In: *Computer Physics Communications* 184.9 (2013), pp. 2214–2220. doi: <https://doi.org/10.1016/j.cpc.2013.04.008> (cited on page 1).
- [5] M. G. Aartsen et al. "Energy Reconstruction Methods in the IceCube Neutrino Telescope". In: *JINST* 9 (2014), P03009. doi: [10.1088/1748-0221/9/03/P03009](https://doi.org/10.1088/1748-0221/9/03/P03009) (cited on page 1).
- [6] R. Abbasi et al. "Low energy event reconstruction in IceCube DeepCore". In: *Eur. Phys. J. C* 82.9 (2022), p. 807. doi: [10.1140/epjc/s10052-022-10721-2](https://doi.org/10.1140/epjc/s10052-022-10721-2) (cited on page 2).
- [7] S. Yu and J. Micallef. "Recent neutrino oscillation result with the IceCube experiment". In: *38th International Cosmic Ray Conference*. July 2023 (cited on page 16).
- [8] J. Evans et al. "Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries". In: *Phys. Rev. D* 95 (2 Jan. 2017), p. 023012. doi: [10.1103/PhysRevD.95.023012](https://doi.org/10.1103/PhysRevD.95.023012) (cited on pages 17, 18).
- [9] J. Feintzeig. "Searches for Point-like Sources of Astrophysical Neutrinos with the IceCube Neutrino Observatory". PhD thesis. University of Wisconsin, Madison, Jan. 2014 (cited on page 18).
- [10] N. Kulacz. "In Situ Measurement of the IceCube DOM Efficiency Factor Using Atmospheric Minimum Ionizing Muons". MA thesis. University of Alberta, 2019 (cited on page 18).
- [11] M. G. Aartsen et al. "Computational techniques for the analysis of small signals in high-statistics neutrino oscillation experiments". In: *Nucl. Instrum. Meth. A* 977 (2020), p. 164332. doi: [10.1016/j.nima.2020.164332](https://doi.org/10.1016/j.nima.2020.164332) (cited on page 18).
- [12] <https://github.com/icecube/pisa> (cited on page 18).
- [13] R. S. Nickerson. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises". In: *Review of General Psychology* 2 (1998), pp. 175–220 (cited on page 19).
- [14] H. Dembinski et al. *scikit-hep/minuit: v2.17.0*. Version v2.17.0. Sept. 2022. doi: [10.5281/zenodo.7115916](https://doi.org/10.5281/zenodo.7115916) (cited on page 19).
- [15] F. James and M. Roos. "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. doi: [10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9) (cited on page 19).
- [16] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. doi: [10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360) (cited on page 21).