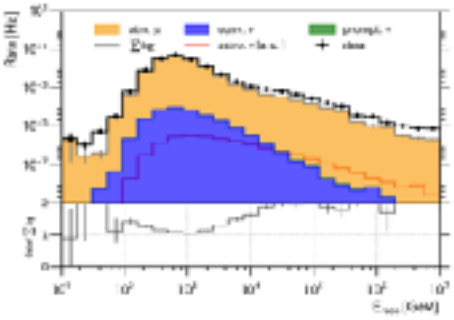
Feedback:

1. Please clarify what “baseline analysis based on pass 2 MC” means
    - Sorry I just mean I did some analysis without considering systematic uncertainty. These include data mc agreement check, study on the change/update of following components: 1. baseline atmospheric neutrino flux model, 2. ice model for monopod reconstruction, 3. calculation of self-veto effect. They are described in section 6 more detailedly.
2. For the “self veto”: please give a few line explanation of the main changes.
    - The main changes include: 1. the new self veto allows for including systematic uncertainties and detector characteristics. 2. the new self veto considers some cases which is approximated by the old one. For example, the new self veto takes into account fluctuations from stochastic losses of muons propagating through the Earth toward the detector and includes muons from other branches of the shower in which the neutrino was produced.
3. On the “Level 3” section: what are the backgrounds that the “Cascade L3 processing” is trying to reduce?
    - Cascade L3 processing reduce both muon and atmospheric neutrino background. For example containment cut reduces the muon background because muon is track like signal which is less likely to be contained in the detector region. In addition, coincident events (coincident muons and coincident muon+neutrino) are reduced.
4. What are “single events”? Do you mean events w/o coincidence atmospheric muon?
    - yes, single events means events w/o coincidence.
5. How is the 2.55/3 decided on? It is not the ratio of 79/86 (which I also do not think is the right thing to do).
    - 2.55/3 is the ratio of event rate of 2010 and other years. I just rescaled ic79 data to compare the shape of other variables with other years.
6. Was this followed up on? This is a very clear difference for 2010 and 2020, separated by almost 10 years?
    - I will answer this question later.
7. Sorry, this does not make sense to me. Can you link to further explanation about why the cascade passing rate changed? Does the new filter just have a different passing rate?
    - We looked up related document. And here is the conclusion:
    - This is the proposal for CascadeFilter_13 (https://user-web.icecube.wisc.edu/~mlesiak-bzdak/Documents/Proposals/cscd-filter-l2-ic86-2013-v07.pdf). It listed the cuts and changes comparing with CascadeFilter_12. The main changes are that a new cut on NString > 1 is proposed and the cutting parameters are changed. For pass1 CascadeFilter_13 is used starting from 2013 while for pass2 it is used for all the years. According to the proposal (Table 1 on page 6), the difference between the level2 event rate for 2012 filter and 2013 filter does not explain the difference shown in our figure. The table shows that 2013 filter has a higher rate (30.0) than 2012 filter (28.5), while our figure shows a higher rate for

2012. Then we looked up Hans' thesis (https://www.stonybrook.edu/commcms/grad-physics-astronomy/_theses/niederhausen-hans-may-2018.pdf). The table 3.1 on page 50 agrees with our plots. Year 2012 has a higher level3 single contained rate (193.9 mHZ) than other years (~155 mHZ). The reason is unknown yet **but the data agrees with corsika well for pass1 (2012 corsika also has a higher rate than 2011 corsika).** Besides, for pass2 analysis, we don't necessarily require the agreement with pass1, especially for the very early years when the filters and software are not stable yet. The uniform across years and MC/data agreement is more critical.

8. Sorry, this is also not clear to me. Can you add more documentation about why this is true? This otherwise seems a clear change in the high-energy cascade rate.
    - The plot at right hand side is from Hans' thesis. It shows each component of level3 single contained events. So at this level, the sample is still muon background dominated. And we can see in the later part that in the final cascade sample, the high energy events look identical. Then this difference should not influence our fitting result.



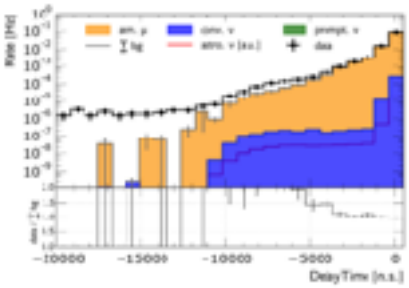
9. I am confused. In previous sections, you gave comparisons for 2010—the 79 string version of IceCube. Why are you now saying that you only select IC86 configurations? Also, does this means that runs with strings dropped (does this ever happen?) are not included in the analysis? Also, can you please clarify if deepcore is included in the analysis? Or is this only based on non-DeepCore DOMs?
    - Sorry, I did not make it clear here and the wiki is updated. We request the runs with all strings are good. Which means for years 2011-2020, we request the runs with 86 strings are active and for year 2010 we request the runs with all 79 strings are active. What we dropped is the runs with bad strings, for example the active strings of run 122299 in 2013 is 85. (https://convey.icecube.wisc.edu/data/exp/IceCube/2013/filtered/level2/IC86_2013_GoodRunInfo.txt).

	2010	2011-2020
active strings	79	86

    - And deepcore is included in the analysis. We did monopod reconstruction both with and without deepcore. The cuts depend on both of them and the final energy and direction is from the reconstruction with deepcore.
10. How are the outliers defined quantitatively? Can you also add some small comment about why the rate various throughout the year?
    - The outliers is defined by 20% deviation from rolling average of the neighborhood. And the rate variation throughout out the year due to the change

of atmospheric density throughout the year, lead by the change of atmospheric temperature.

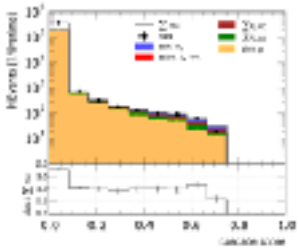
11. Please add more narrative. E.g. “After we select events passing the L3 filter, we apply a further selection. The selection is different for high…”. Also, please try to make it clearer the sets of cuts being pursued. I might suggest that you re-order the wiki page so that it goes “level 3 cascade selections” and then immediately “higher level cascade selection.” And then you can put all the pass1/pass2 checks in to the “appendix” so to speak. Otherwise the cuts are interrupted by this big section on pass1/pass2 comparison, and on the year-to-year comparison.
    - I re-ordered my wiki, thank you.
12. Please explain the purpose of this cut.
    - The background is dominated. This cut is inherited from the previous cascade analysis. Previously it is used for getting a better data/mc agreement. As the figure shows there is a large discrepancy below -5000.



13. Like For point 11, please add more narrative. Also, explain why the L5 cut is needed. I guess it’s because the muon background in the cascade L3 sample is still too high?
    - Yes. You are right. From the plot above, after level4D, atmospheric muon still dominants.
14. Can you please add more explanation of why applying the pass 1 BDT to pass 2 data is okay? BDTs, or ML techniques in general, can be extremely sensitive to e.g. data/MC mismatches. So I think justification needs to be added here.
    - We compared the bdt variable distributions on level5a. Pass1 corsika, which was used for training bdt, matches with pass2 data. And we re-trained the bdt for pass2 and burn sample below 10 TeV. The new model does not performs better. This is added to wiki, see my wiki for detailed information.
15. What is the intuitive meaning of the hybrid score? If the hybrid score is >0.75, does that mean it is 75% chance (or better) of being hybrid event? Or is it really some arbitrary discriminant variable?
    - Add the definition of the bdt score. It does not mean it is 75% chance of being hybrid event.
16. Why is the sample split up this way? And which sample is the signal sample which will be used in the fit? I guess it’s the “cascade sample”?
    - The reason we split the neutrino sample into cascade sample and hybrid sample is that we want to split and fix the conventional atmospheric neutrino background. The shape of hybrid sample is cascade + track. Numu charge current decay is the event with that kind of shape. Numu charge current decay produces hadrons (looks like cascade to detector) and muon (looks like track). So this sample constrains the flux numu events then constrains the atmospheric neutrino flux. And cascade sample is the signal sample because most of

astrophysical neutrinos fall into this sample. But we used all three samples (cascade, hybrid, muon) in the fit.

17. Again, I don't know what this cut/coordinate means? For muons, why would you want a large cascade.score? Perhaps some plots need to be added to this section to better explain the selections.
    - According to pass1 wiki (https://wiki.icecube.wisc.edu/index.php/Multi_Year_Cascades/Low_Energy), this cut is applied because of the discrepancy in the first bin.



18. It would be nice if something like this was added to the low-energy section. In fact, I suspect this answers some of the questions I had about e.g. "Delay Time". See if you can rework these sections so the explanations come in the right order, etc.
    - Sure. I have added that.
19. Link to the software which makes this cut.
    - I have added to wiki.
20. What is the motivation for wanting to recover this event? As stated, this is non-scientific, and must be explained.
    - This was proposed at the beginning of the review process but finally we decided to not change it in the analysis. I updated the wiki to make it clear.
21. Which one did you go with at the end? And now that you say that, please specify which version of xgboost you used, and why you chose that version.
    - The main point is that we should keep bdt used for training and classification the same version. If we use previous bdt model, we should keep the same software version. If we need to retrain the bdt, we are free to use any version we want but still need to keep the version same for training and running.
22. This is not quite my job, but this is an unacceptable location for permanent storage of analysis code (/data/user isn't even backed up!!). This should be moved to GitHub/SVN, and reproducibility checks ensured, before the analysis can proceed. Please discuss with WG leads and WG/technical reviewer.
    - I will upload it to the github.
23. See question 20. Why should this event be rescued? This seems like the definition of cherry picking, and I don't understand why you do it. Also, can you please provide the energy and zenith of this event? A steamshovel view might also be nice.
    - See answer to question 20.
24. The distribution for cscdSBU_MonopodFit4energy is clearly *very* different. This must be explained.
    - Sorry, the labels are incorrect and I updated the wiki. There are three plots. From left to right, they are astrophysical+conventional+prompt neutrino flux, conventional components only and prompt components only. In this analysis we propose to switch from berss(blue line) to prompt from sibyll2.3c. See

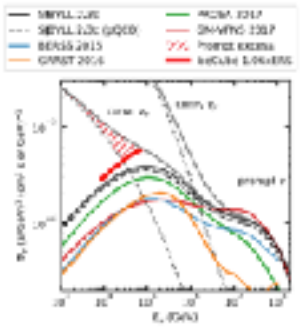


figure 6.4 in https://journals.aps.org/prd/abstract/10.1103/PhysRevD.100.103018 (attached here). The prompt flux difference is up to one order of magnitude for different model but we have not seen prompt flux in any analysis.

25. What is meant by "lower the muon energy veto"? Is this a parameter of the software? This needs to be explained better. It seems we also really need data/MC comparison to know if the new self-veto calculation is working?
    - "muon energy threshold" is a parameter in self veto software that we tune to match down going atmospheric neutrinos in Monte Carlo with our data. Yes you are right, we performed a data/mc comparison and updated the self veto study with the muon energy threshold selection result to my wiki.
26. Much more details are needed (imo) about how the fit is done. Can you please write down the likelihood? I would also like to see the 2D histograms for data for all three samples.
    - I updated my wiki page.
27. Can you specify clearly which are to be solved for, and which are nuisance? I think everything that is not the 2 astrophysical neutrino flux parameters are nuisance, right?
    - We have 5 parameters to be fitted. They are astro_norm, astro_index, conv_norm, prompt_norm and muon_norm. And the rest are nuisance parameters. See table 2 from (https://arxiv.org/pdf/2001.09520.pdf).
28. Please add a Q&A section to your wiki.
    - I have added it.
29. I'm afraid I don't see any discussion of data/MC agreement. I think this needs to be added.
    - I have added it.
30. I think post-unblinding checks and plans needed to be added as well.
    - I will add it.
31. From a high-level perspective: Have you considered testing beyond the single power law? Cascades should, in principle, give you the ability to probe e.g. broken power laws, piece-wise unfoldings, log-parabolas, etc. This would be very interesting. I think the previous analysis made such tests?
    - Yes, these will be parts of the test after the unblinding.
32. I expected some plots of e.g. the expected sensitivity, and examples of performance on burn sample only. For example, what result does your fit give if run only the burn sample of the 10 years? What is the expected sensitivity on the 10 year sample, and what is the impact of the systematics on the sensitivity? Because this analysis is (essentially) an extension of a prior selection, I think it would be good for you to quantify in what ways the sensitivity improves and have those projections in the wiki before moving to unblinding request. I assume Shigeru as your WG reviewer is also thinking about this.
    - For sensitivity part, this can and will be done. For burn sample fit, we can do it to test background because currently we are only allowed to look events below 10 TeV.