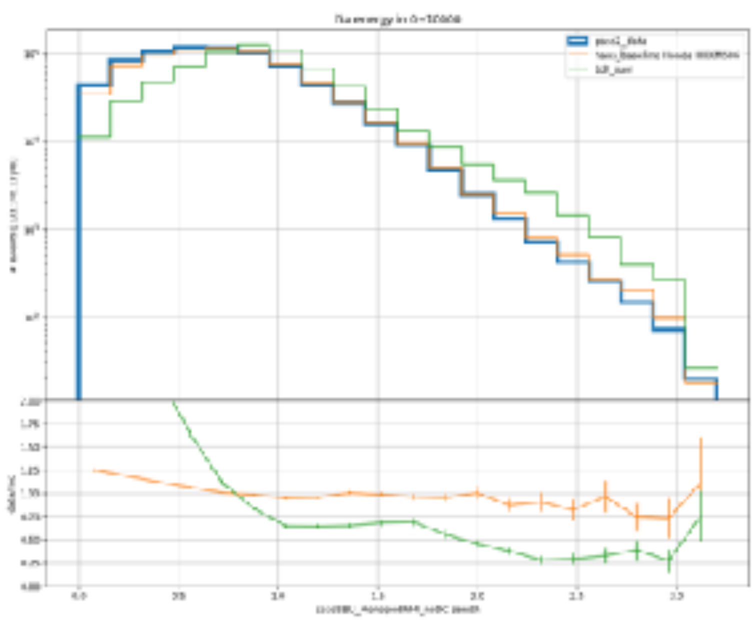
1) Can you explain more about how you retrained the BDT? With what samples? And what is your evidence that retraining the BDT actually helped, aside from the improvement in the goodness of fit? Do you now achieve better data/MC agreement? This was already pretty good, which is why you decided not to retrain a few months ago, right? Can you please add plots to your wiki explaining this?
    1) I used the same bdt variable as pass1 cascade analysis. The difference is the MC datasets used to train the BDT model. I used the Manuel's nugen and muongun datasets, which are the same datasets I used to do the fit. I used part of the MC datasets to train the BDT model and burn them. And I used the rest of the datasets for the fitting. And to fit with new BDT, the bdt score cut is also needed to be adjusted as well. Because BDT score does not have physical meaning, I think the BDT score cut could be changed with the change of bdt model.
    2) The better data/MC agreement could was indicated by the plot on my wiki page (https://wiki.icecube.wisc.edu/index.php/Pass2_Multi_Year_Cascade_Analysis#bdt_model). Generally the new

- Datasets for training BDT (Burned MC):
    - mgun:
        - 21315: 5000/20000,
        - 21316: 10000/40000,
        - 21317: 5000/20000,
        - 21318: 20000/100000,
        - 21319: 20000/100000
    - Nugen:
        - NuMu:
            - 21813: 5000/20000
            - 21814: 5000/20000
            - 21938: 100/1000
        - NuTau:
            - 21867: 100/1000
            - 21868: 100/1000
            - 21939: 100/1000
        - NuE:
            - 21870: 100/1000
            - 21871: 100/1000
            - 21940: 100/1000



input MC dataset (orange line) for training BDT model has a slightly better than the old input MC dataset (green line) for training BDT model over all variables. The largest improvement is shown on variable cscdSBU_Monopod_noDC_zenith. That's because the different icemodel settings between new and old MC datasets and the new MC datasets have a better agreement with data.

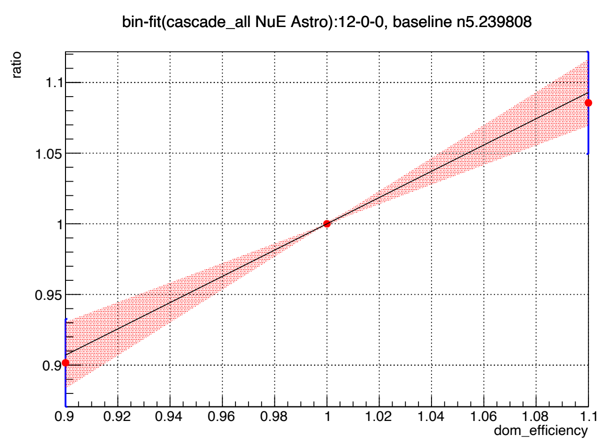
3) Brian: Can you please show data/mc agreement for old BDT model and data/mc agreement for new BDT model. You have shown that the old BDT,

applied to new MC, has better data/mc than old BDT applied to old MC, but I don't think we've seen new DBT on new MC. I think we need to see this too.

1) I prepared a note which includes the data/mc agreement for old BDT model and new BDT model. The conclusion is both old and new BDT model give a good data/mc agreement. For some plots the new BDT model is slightly better but not significant since the old one already have a quite good agreement.

2) I do not understand the "Overview of correction efficiency error" slide in your presentation. Can you please explain it again?

   1) We use the correction efficiency to represent the systematic uncertainties. To get the expected events number from MC dataset with certainty systematic parameter values, we use formula like $N = N_0(\text{baseline parameters})e_1(x_1)e_2(x_2)\dots$. If we use $\sigma^2 = \sum_{i=1}^{m} w_i^2$ to estimate the uncertainty of N, we are actually assuming the uncertainty from efficiency corrections e_1, e_2,… are 0, which is not true. So I took dom_efficiency as an example to show how large the uncertainty of efficiency of corrections are for different bins (red shaded area). After taking these uncertainties into account, the GOF and p-value is improved. This means some part of the previous mismatch between MC and data could be explained by this effect.

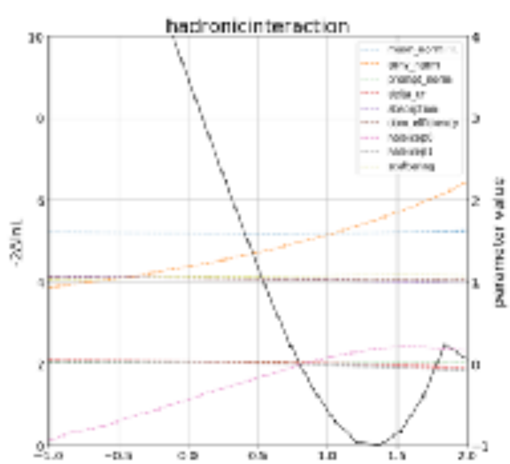


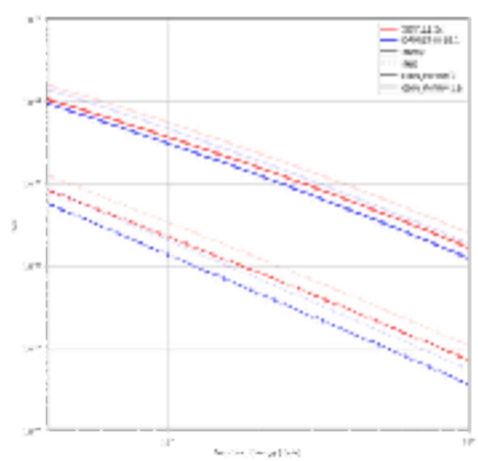
   2) Brian: I'm afraid I'm still confused. So, you are saying that in the case of dom eff, the ratio of systematic MC counts/ nominal MC counts not only increases with energy, but so does the MC uncertainty on that ratio?

      1) I think what the plot shows is that the the uncertainty of the ratio increases if the systematic parameter, dom eff in this case, is far away from the nominal value. In the region of the parameter closed to the nominal value, the uncertainty is well controlled.

3) I still remain surprised by how high the conventional normalization is “pulled” in your fit. That is really different in the global fit framework enhances my confusion. \Delta CR is also different. But it sounds like the talk on Wednesday will explain more from Erik et al, so maybe I should just wait.

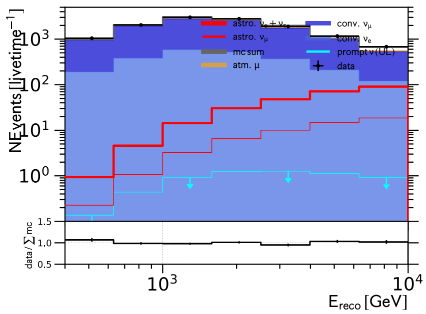
    1) The conventional normalization is highly correlated with the hadronic interaction model. Here is the scan over hadronic interaction model. Conv_norm (orange line) changes a lot with the scan of hadronic interaction model. For our baseline model (SIBYLL2.3c, x=0), the corresponding conv_norm is closed to 1. But our best fit seems prefer a softer hadronic interaction model (x=1.4). In this case the conv_norm increases to 1.7. The interpretation of the correlation between these two parameters is that Hadronic interaction model changes both the normalization and the shape of the neutrino flux. The right plot show the neutrino flux without any filters, in our background region, numu flux from DPMJet with norm=1.5 (solid, blue, light line) is closed to the numu flux from SIBYLL2.3c with norm = 1 (solid, red, dark line).



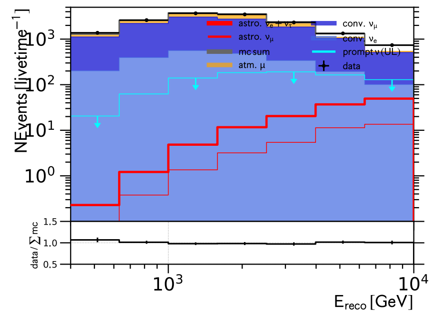


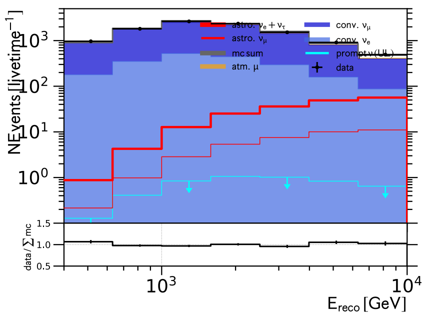
4) On your "GOF before and after the improvements”, slides, I don’t see “before” plots?

    1) Sure, here are the one by one comparison. The plots on the left hand side are “before” and on the right hand side are “after”.

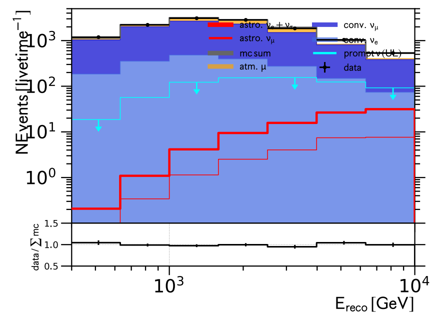


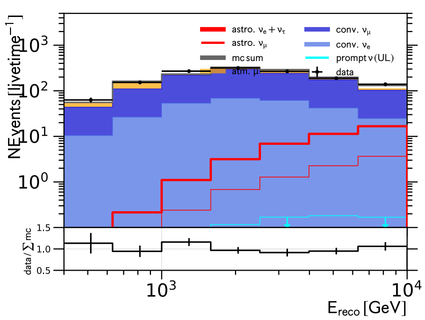
cascade sample all sky



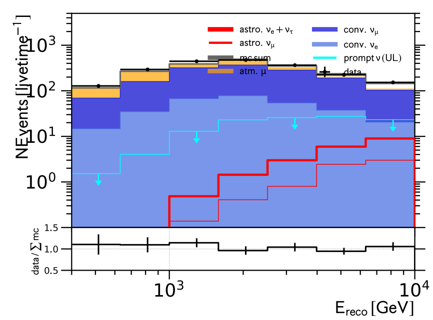


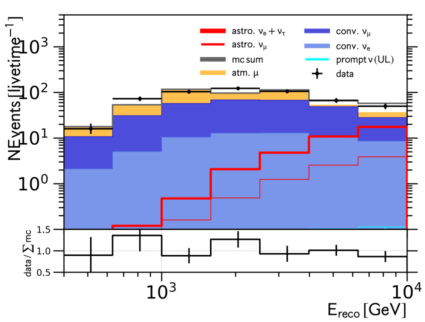
cascade sample up going (coszenith -1~0.2)



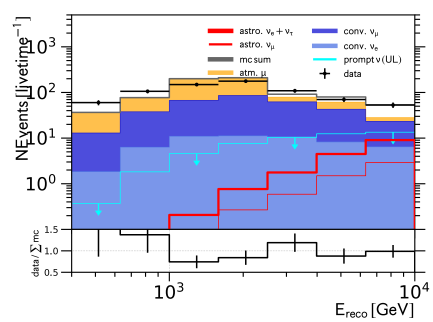


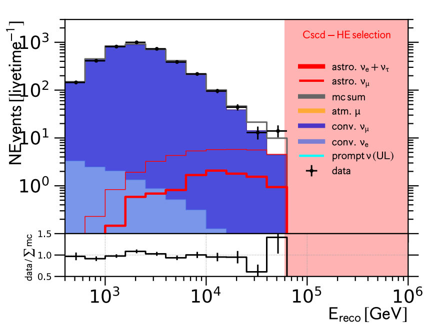
cascade sample down1 (coszenith 0.2~0.6)



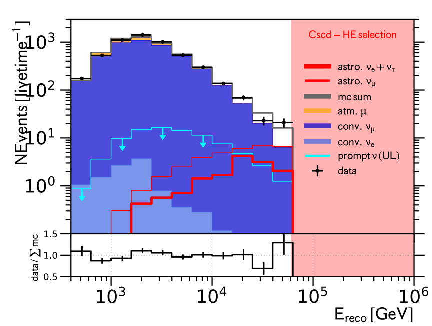


cascade sample down2 (coszenith 0.6~1)





hybrid sample all sky



2) Brian: I might have actually thought things go *worse* in the bins 0.2-0.6 and 0.6 to 1. What am I missing here? I think this is not dissimilar to shigeru's question B
    1) For 0.2-0.6, I think it is pretty hard to say. For 0.6 to 1, I agree that the new model is worse than the old one. But since the down going event rate is order is magnitude lower than the up going event rate. And the error bars for that zenith bin are quite large, I think this could justify the use of new model.

5) Can you go back to your unblinding plan, and remind me of the process you have proposed? It looks like you have already run the fit on the full sample and verified a reasonable GOF. Have you checked that none of your fitted systematics have exceeded their priors? Idk what the proposed steps were, but please update us on them when you come back for unblinding approval in the WG and then when you to go the analysis call.

blind fit result of background data

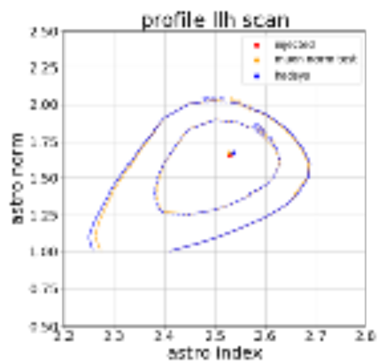
- Fit result new bdt model after correction:
    - muon_norm = 1.60052
    - conv_norm = 1.69852
    - prompt_norm = 3.72025
    - delta_cr = 0.0176558
    - absorption = 1.02315
    - dom_efficiency = 0.974855
    - hadronicinteraction = 1.06778
    - holeicep0 = 0.598135
    - holeicep1 = -0.0201658
    - scattering = 1.07147
    - selfveto = 1443.53
    - likelihood value (gof): 30.3716
    - likelihood value (abs): 319.42

stopping criteria of unblinding plan (https://drive.google.com/drive/folders/1pJuvBwYe-wYIHJRwKJhq3S3kPgCDsZun)

Unblinding Plan
systematic parameters stopping criteria

- muon_norm = [0,4]
- conv_norm = [0,2]
- prompt_norm = [0,10]
- delta_cr = [-0.3,0.3] (prior 0±0.05)
- absorption = [0.79,1.21] (prior 1±0.07)
- dom_efficiency = [0.7,1.3] (prior 1±0.1)
- holeicep0 = [-2, 1]
- holeicep1 = [-0.2, 0.2]
- scattering = [0.79, 1.21] (prior 1±0.07)
- selfveto = [0,5000]
- hadronic interaction model = [-2,3]

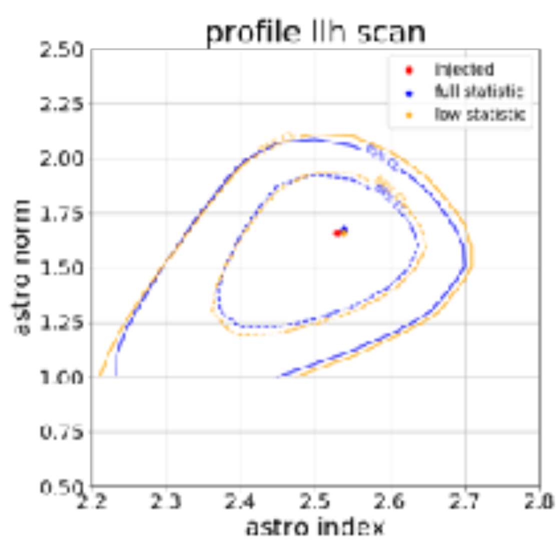
1) As the reminder, here is my slides for the unblinding plan (https://drive.google.com/drive/folders/1pJuvBwYe-wYIHJRwKJhq3S3kPgCDsZun). We have finished the first 2 steps. For the 2nd step, which is the blind fit on background (<10TeV) data. The fitted systematics do not exceed their priors.
2) Those systematics parameters could be explained.
    1) The muon_norm (1.6) is larger than the nominal value (1.0). This was asked by Manuel previously and I did the following check to make sure there is no problem. Since the muon_norm is mostly controlled by muon sample, to understand the influence of muon_norm to astro flux. I did the asimov fit. I used hybrid sample and cascade sample, inject them with muon norm =1 and fit with muon_norm = 1.5. The asimov fit suggests that this does not bias the astrophysics parameters. The red dot is the injected point and the yellow dot and yellow contour are the asimov fit results.

    2) The correlation between conv_norm and hadronic interaction and why it is pulled to a large value is discussed in question 3.
    3) since this is the background data, we don't have constrains to prompt_norm.
    4) delta_cr, absorption, dom_efficiency, holeicep0, holeicep1, scattering and selfveto are within the prior value.
3) I have requested the blinding on the analysis call and was approved from the analysis call. Nathan asked me to give an update to the diffuse working group after I finished the step two. So I think I can go to step three which is run blind fit and check parameters with full datasets once there is no questions from reviewer and working group without go back to analysis call again.

6) "The change of contour" — same question as Nathan. If you're taking a meaningful hit in the size of your contours from lack of MC statistics, the WG can sponsor priority queue processing for you to generate more MC stats. Please let us know if you would like to do this, it seems an obvious "win".

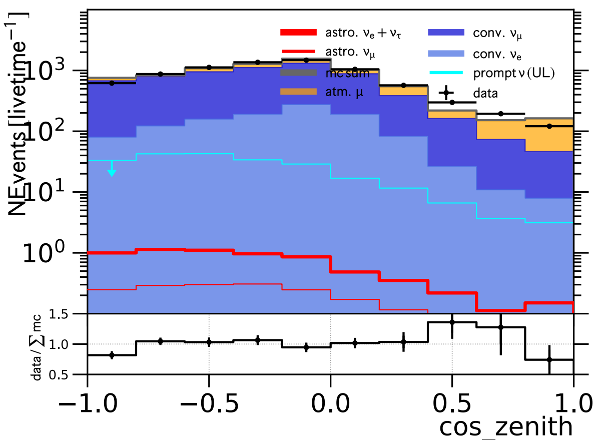
1) This question is asking about page 15 of the slides https://drive.google.com/drive/folders/1mOskTzc0O7UrbqYRTsCZB07aXclTXyXs. Actually I don't think there is a huge change.
2) So there are two plots on this page:
    1) The plot at left hand side shows the change from the uncertainty correction. The difference comes from ideally infinity systematic statistics and current systematic statistics. I don't think generate more systematic datasets will help a lot since the infinity limit is the blue line.
    2) The plot at right hand side shows the change from (the different BDT model + uncertainty correction). The difference between yellow and blue line is more visible. The source of the difference could be the change of the BDT model. But the other important source I guess is that part of the MC datasets was burn during the training of the BDT. I will estimate how large the difference from the loss of statistics due to the burn of MC is. And consider if we want to generate more MC dataset to compensate the burn MC dataset.
3) Brian: have you made a decision yet? Do you want more MC?
    1) As I said before, I used part of the nugen and muongun datasets for new bdt model training. And those MC datasets should not be used in the fitting. The following plot shows the difference between full statistic (all MC datasets are used in fitting) and the low statistic (drop part of MC datasets) for old bdt model. To be explicitly, for example, we have 100,000 files for mgun datasets 21319, 100,000 files are used in old bdt model fitting, while for new bdt model, 20,000 files are used for training and 80,000 files are used for fitting. In the following plot, the blue line is the profile llh scan for old bdt model with 100,000 files from mgun datasets 21319, and the yellow line is the profile llh scan for old bdt model with 80,000 files from mgun datasets 21319. This plot shows how much difference are from burning part of MC datasets for training, and it suggest that the difference is quite small and we don't need to generate more datasets to compensate the burn MC datasets.



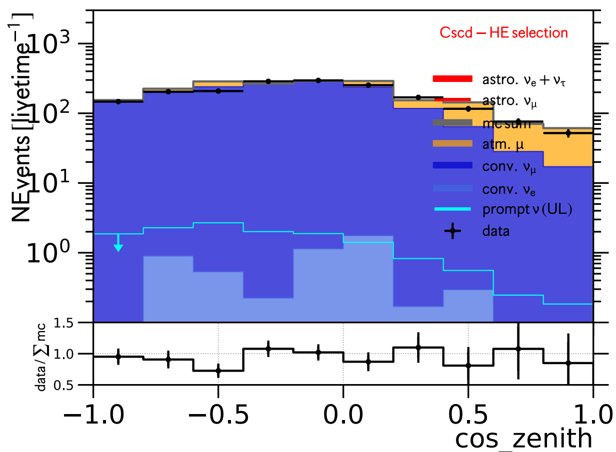
7) Can you show the zenith-dependent data/MC agreement in the background region, with the update fit?

   1) Sure. Here are the zenith distributions. The best fit parameters are from the 3 zenith bins fit (8 energy bins * 3 cos zenith bins for cascade sample, 11 energy bins for hybrid sample and 1 bin for muon sample). But the distributions are plotted in 10 zenith bins. The data and MC show a good agreement.

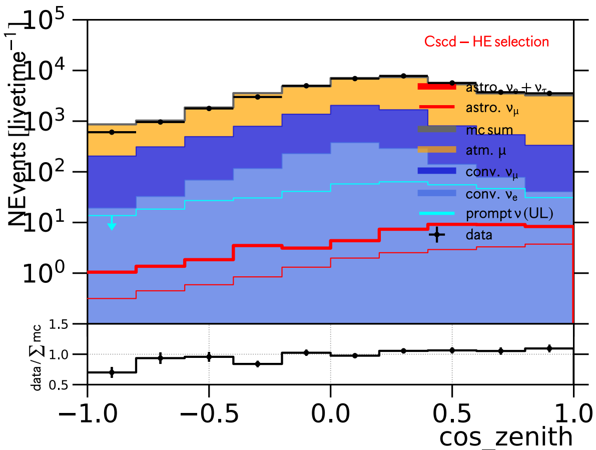
cascade



hybrid



muon



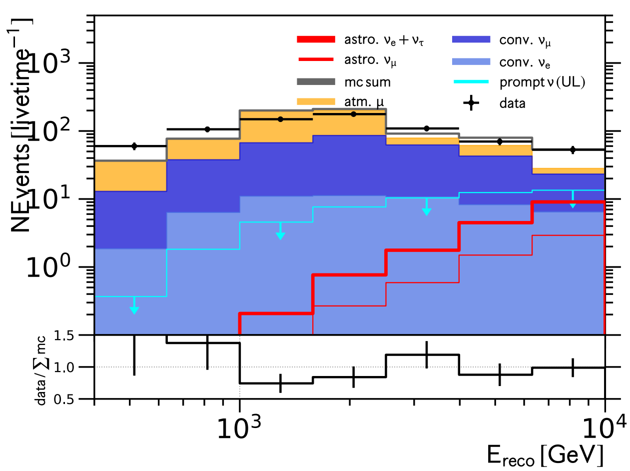
   2) Brian: Is there a “tilt” in the muon distribution? Meaning there is an under-prediction at cos(zen)=-1 trending to an over-prediction at coz(zen)=1?

      1) Yes. I agree that there is a tilt in the muon distribution. I think that’s because we don’t have systematic datasets for muongun events. According the that plot, there is a relatively larger deviation at upgoing

region, while for down-going region, the deviation is much smaller. We do not use the coszenith of muon sample in the fit. And for cascade sample, the muongun events are mostly down-going, where the data/mc is closed to each other. So I think this does not bias our result.

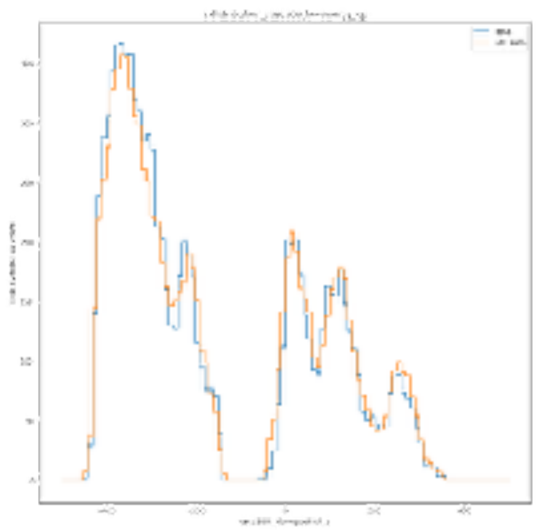
8) Questions from Shigeru: I echoed all the questions asked by Brian. The BDT retraining is a big change, and I need to know what you changed, reasons to justify the changes, why the "retraining" BDT led you to a "better" GOF. Also the event distribution of cosZenith = [0.6 1] i.e. downgoing regime, seems to indicate that MC rate (most likely atm muon penetration rate) underestimates the data. This is my concern. Before jumping to the BDT, please also show us the fundamental plots with the unblinded background data, such as - zenith angle distribution with energy slices,- vertex z position with zenith angle slices

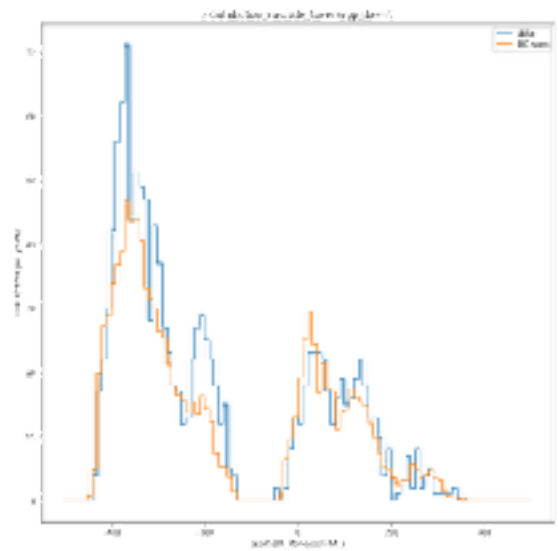
1) Reasons to justify the changes: the old BDT model was trained with old Corsika datasets, which is old, processed with pass1 filters and simulated and reconstructed with old ice model (spice_mie). New MC datasets show a better data/mc agreement before BDT selection (check out the plot for question 1.2). So update the model with new datasets should provide a better data/mc agreement.
2) According to this plot, seems like the MC rate underestimates the data at the first two bins. But we have to say that the error bars of the ratios for the first two bins are quite large. So it is reasonable to claim that it dues to the fluctuation or mis-reconstruction of events to the 3rd and 4th bins. And according to the zenith plot for muon sample in question 7, it is not likely from the atm muon penetration rate.

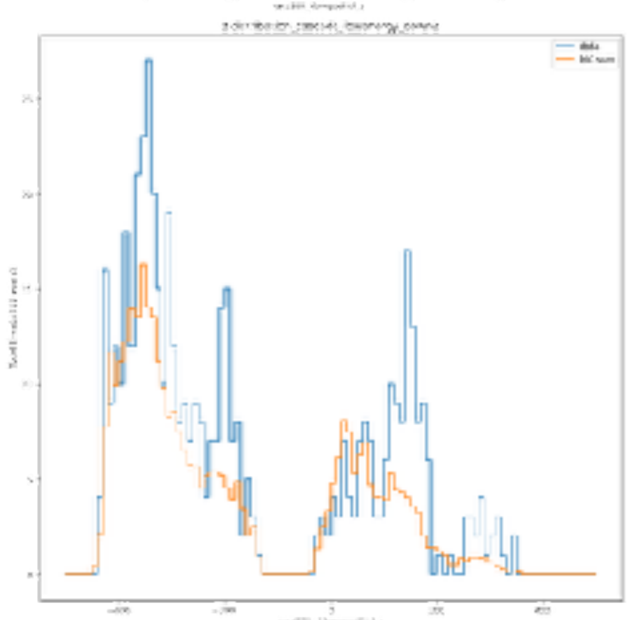


3) The zenith distributions for full energy range (<10 TeV for cascade <60 TeV for hybrid and muon) were shown in question 7.
4) The z distribution with zenith angle slices are shown as following plots. These three plots come from old BDT model. For up going events, we have excellent data/mc agreement. For down going events, MC rate is

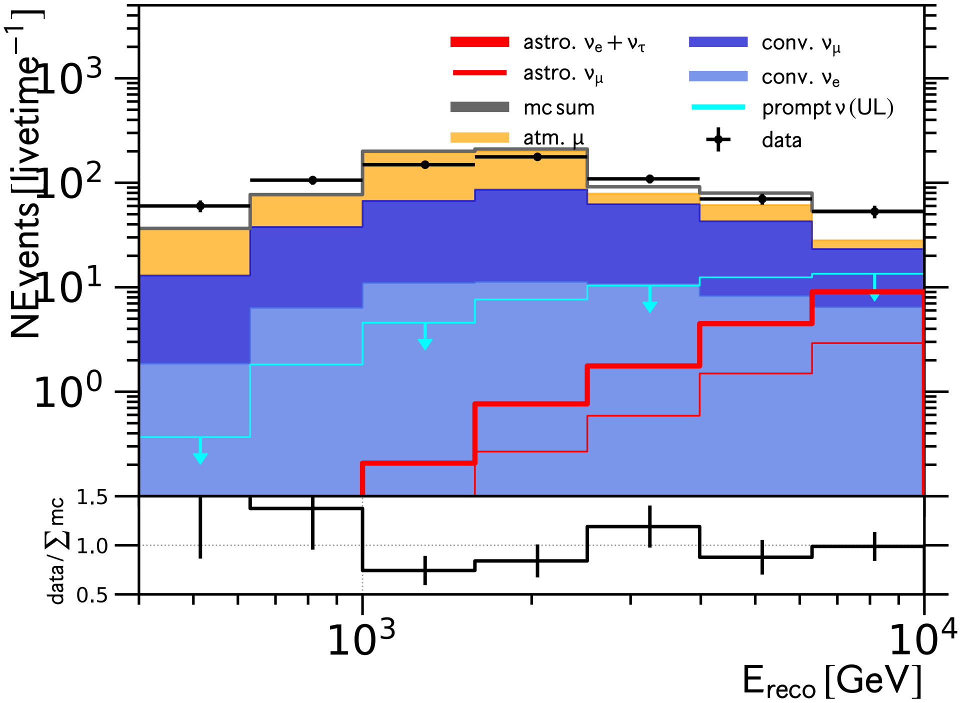
lower than data rate. For down 1 (cos zenith 0.2~0.6) the difference happened at the bottom of the detector. For down 2 (cos zenith 0.6~1) the difference is getting larger and happened at both the bottom and top half.





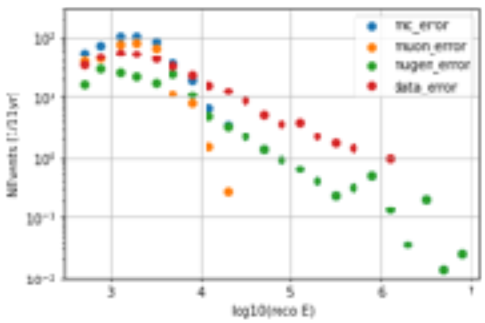


9) A few quick questions:
   1) The energy distribution for cosZenith = [0.6, 1] (i.e., the downgoing region). Are they stacking histogram? why the "mc-sum" line is placed a way higher than the top level of the MC histograms? The "mc-sum" included the prompt mu (UL)?
      1) The conv_numu, conv_nue and atm muon components are stacked. The mc-sum included the pormpt mu, and astro nuetrinos that's why it is higher than the top level of the MC histograms for the high energy bins.

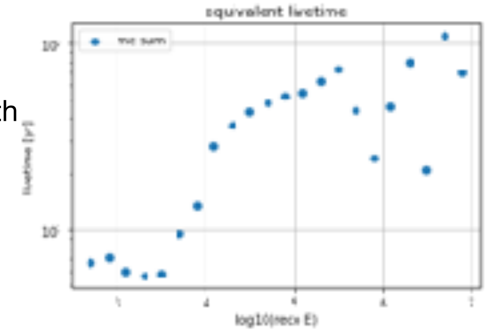


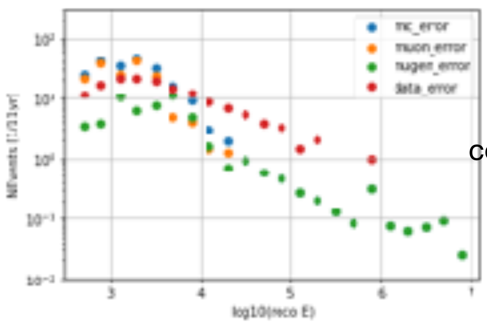
2) About the same plot. You argued that the errors of the data/MC ratio is large so we do not have to worry the apparent mismatch in the low energy region. Where did this error come from? from limited statistics of the MC data? If so, what is the livetime for your MC sample?

   1) The largest part of the MC uncertainty comes from muongun. That's why low energy bins have larger uncertainty than high energy bins. For muongun, the livetime for muongun is about 2 years while the livetime for nugen is > 30 years. But muongun is pretty hard to produce, if we want to make the livetime of muongun 10 years, we need 5x more datasets.
   2) Here are the plots of MC uncertainty and data uncertainty. So the uncertainty from muongun is larger while the uncertainty from nugen is smaller than data.

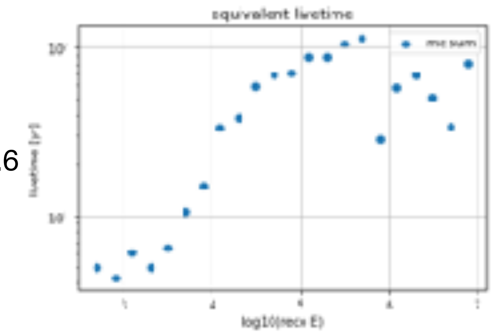


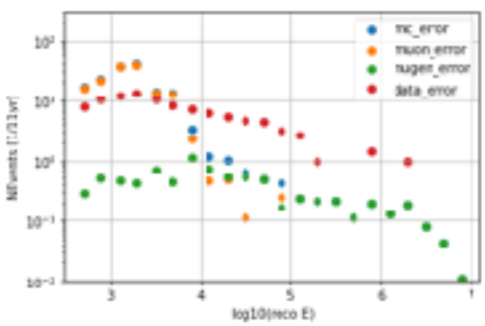
up going cos zenith
-1~0.2



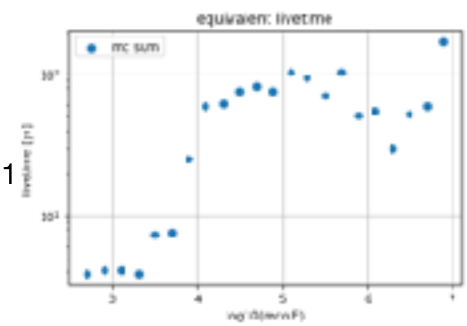


cos zenith 0.2~0.6

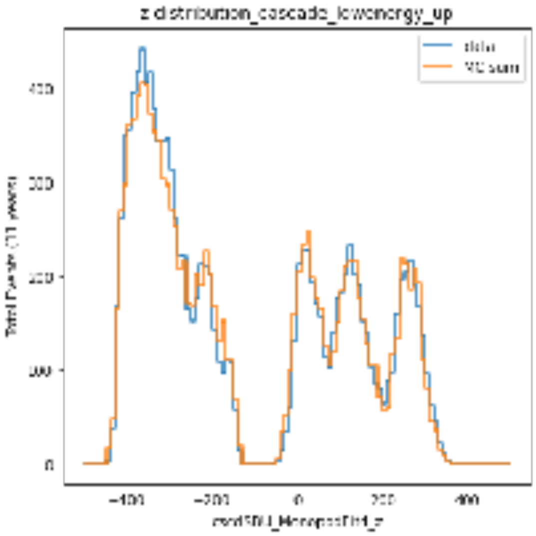


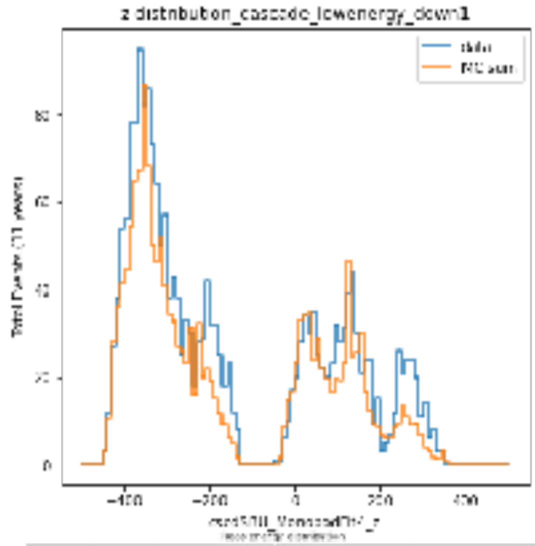


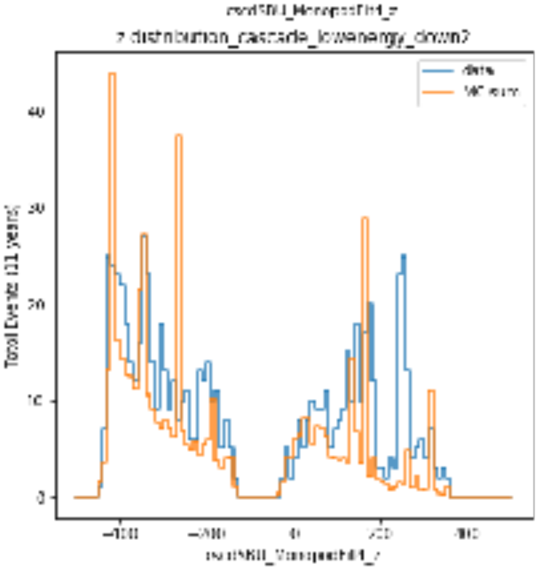
cos zenith 0.6~1

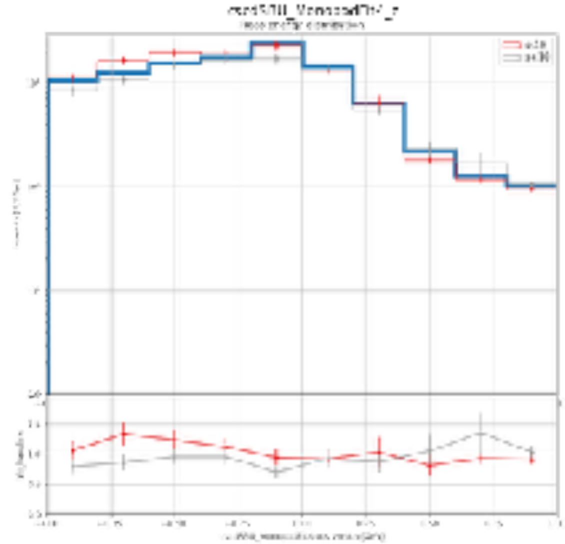


3) the z distribution: Not just the overall rate, but the shallower event rate is obviously underestimated for the cosZennith [0.6,1]. Can you draw the same plot with the updated BDT sample?
   1) Sure. The data still has a higher rate than MC for down going events. But I have to say that these plots does not includes systematic corrections. We know that some systematic parameters such as ice scattering could change the zenith distributions. I think that could explain the higher data rate for down going events.

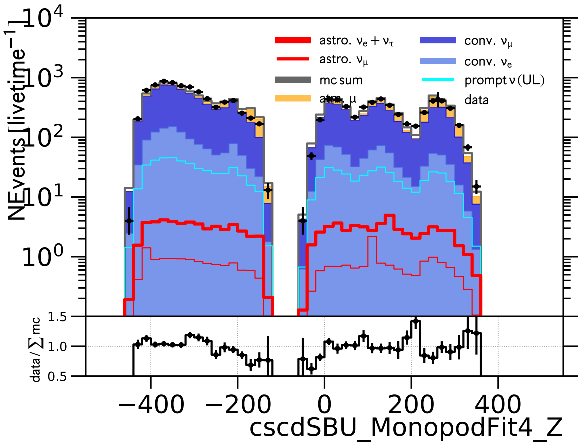


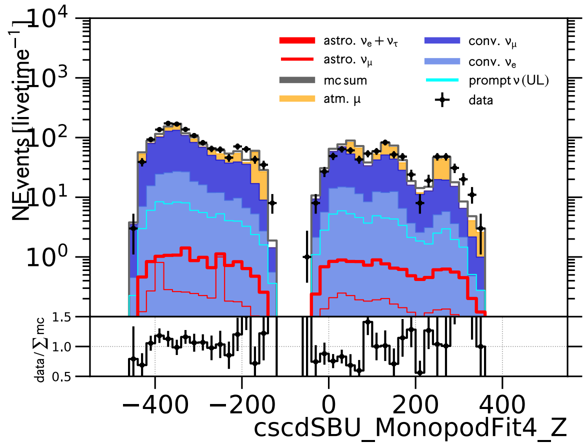


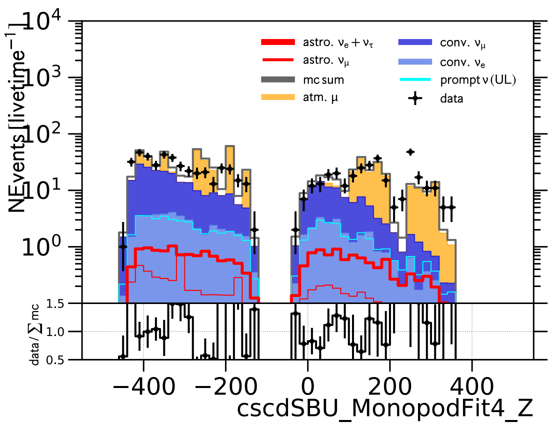




2) As an update, I got the data/mc agreement with systematic parameters included plot. They are up-going (upper left), down1 (upper right), down2 (lower left) correspondingly.)


astro. νe + ντ
astro. νμ
mc sum
atm. μ
conv. νμ
conv. νe
prompt ν (UL)
data
NEvents [livetime⁻¹]
data / Σ mc
cscdSBU_MonopodFit4_Z


astro. νe + ντ
astro. νμ
mc sum
atm. μ
conv. νμ
conv. νe
prompt ν (UL)
data
NEvents [livetime⁻¹]
data / Σ mc
cscdSBU_MonopodFit4_Z


astro. νe + ντ
astro. νμ
mc sum
atm. μ
conv. νμ
conv. νe
prompt ν (UL)
data
NEvents [livetime⁻¹]
data / Σ mc
cscdSBU_MonopodFit4_Z