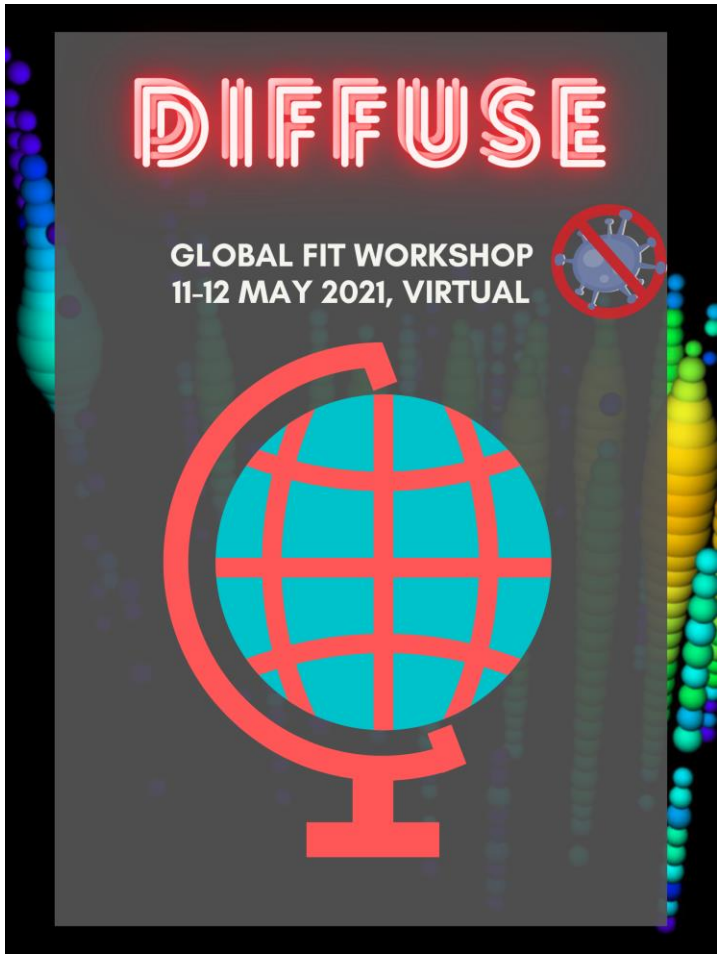
SnowStorm

A Summary of the status of SnowStorm Simulations for a GlobalFit of IceCubes Neutrino Data

Physics Institute III B
RWTH AACHEN UNIVERSITY

Outline

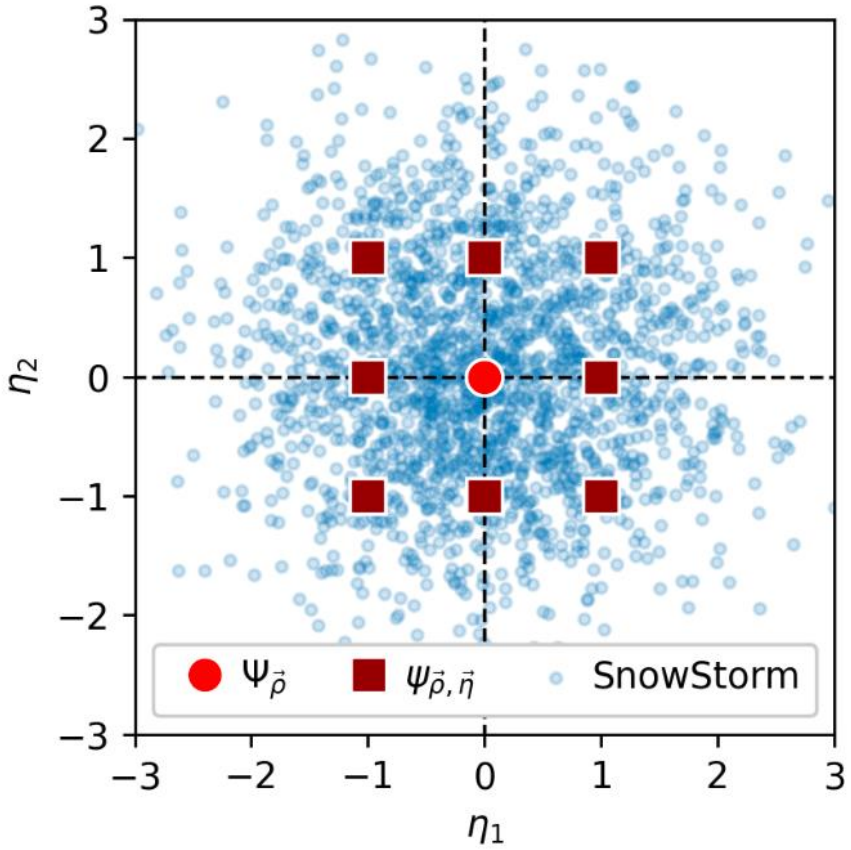
- Reminder: What is SnowStorm?
  - Why SnowStorm?
- How do SnowStorm simulations work?
  - The SnwoStorm simulation chain
- What simulations/MC sets were produced since the Tokyo workshop?
- How to use the simualtions?
  - Processing SnowStorm MC to final analysis level
  - SnowStorm parameters/frame objects
- Recent developments
  - New SnowStorm simulations?
- Summary & Outlook



Reminder: What is SnowStorm?

- Continuous variation of nuisance parameters
  - Instead of generating multiple simulation sets for some specific choices/combinations of nuisance parameters, generate a SnwoStorm event ensemble
  - In the ensemble, a different combination of the nuisance parameters is chosen based on their allowed phase space
  - More detailed method overview:
    - SnowStorm paper
    - Ben's talk at the Spring Collaboration Meeting 2020 (slides)

- Multiple nuisance parameters/detector systematics can be included within a single SnowStorm set at the same time



SnowStorm for the GlobalFit

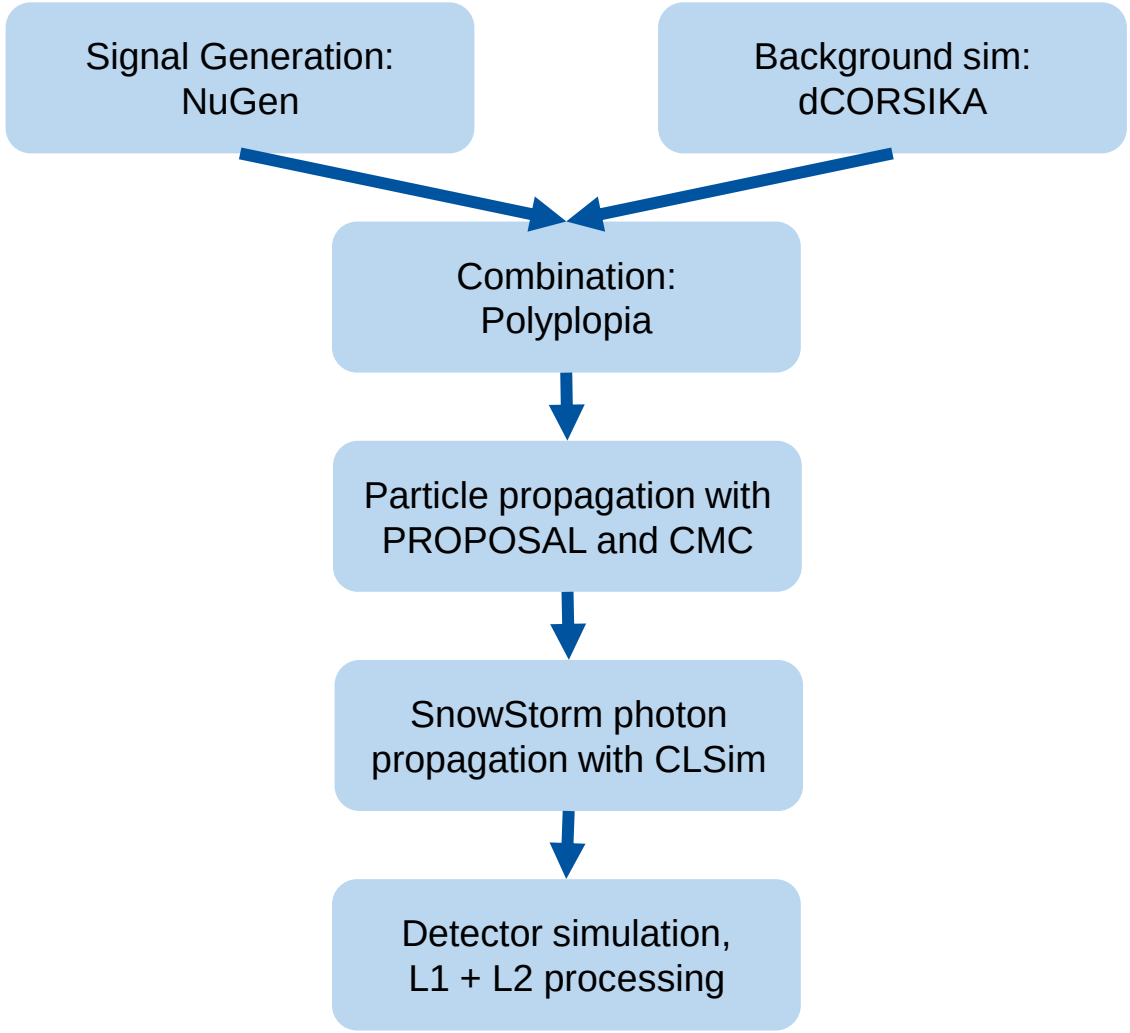
- GlobalFit: Combination of multiple analyses/event selections into a single “global” fit of IceCube’s diffuse neutrinos
  - Consistent treatment of systematics uncertainties for all contributing analyses/event selections is crucial

- New, “up-to-date” simulations have been a major collaborative work throughout 2020
  - SnowStorm MC + “Manuel’s” (ESTES) standard MC

- SnowStorm advantages:
  - No need to deal with multiple simulations sets:
    - 1x baseline + X discrete systematic sets
    - ➢ 1x SnowStorm sets which includes all systematics
  - Each (sub) analyses can “pick” the nuisance parameters/detector systematics it needs
    - ➢ Marginalize other ones if they are not important

SnowStorm Simulation Chain

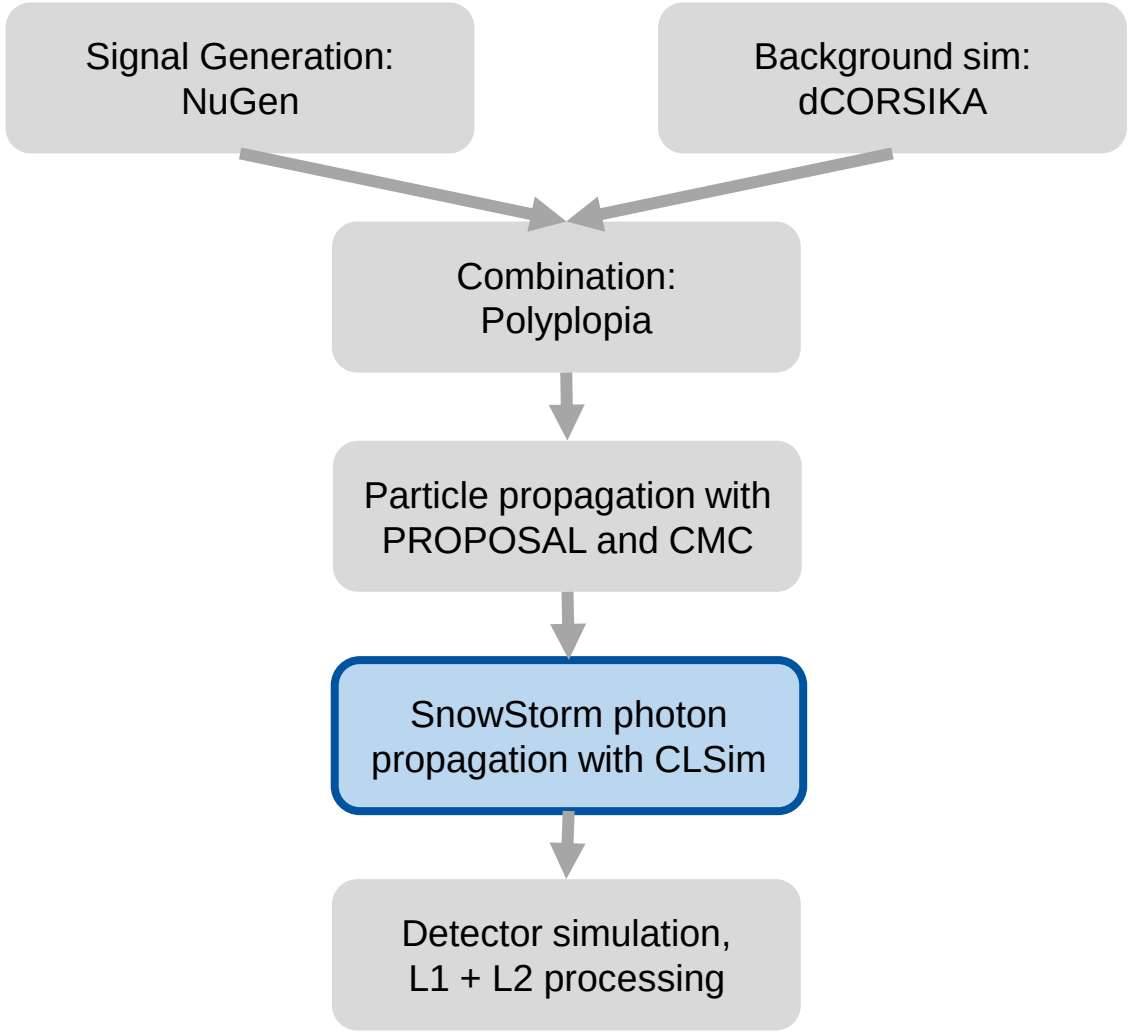
- Merge of signal + background particles into a single I3MCTree, before propagation
- “SnowStorm magic” happens during photon propagation:
  - Application of SnowStorm perturbes for varying the ice model parameters
- Software locations:
  - `snowstorm` software project in icetray/main
  - `SnowSuite` script collection in simprod-scripts
  - SnowStorm software documentation
- Github repo for iceprod simulation configs



SnowStorm Simulation Chain – Photon Propagation

- Standard photon propagation:
  1. Load some ice-model including a choice of e.g. the DOM efficiency and the HoleIce parametrization parameters
  2. Run CLSim using this fixed ice-model for all frames

- SnowStorm photon propagation:
  1. Define all SnowStorm parameters to use and their sampling distributions in a config file
  2. Load a baseline ice-model
  3. Dice SnowStorm ice-model parameters and update CLSim's photon propagation kernel
  4. Process a bunch of frames with these ice-model settings (~ 100 – 1000)
  5. Dice + load new SnowStorm ice-model parameters
  6. Repeat steps 4 + 5



SnowStorm Simulation Chain – Perturbers/Parametrizations

- Currently there are 6 “snowstormable” systematics implemented (table on the right)
- Baseline ice-model Spice3.2.1
- Basline HoleIce model: unified, p0 = p1 = 0
- IceWavePlusModes: icewave ice-model using Fourier decomposition to model depth depended scattering and absorption (SnowStorm paper)

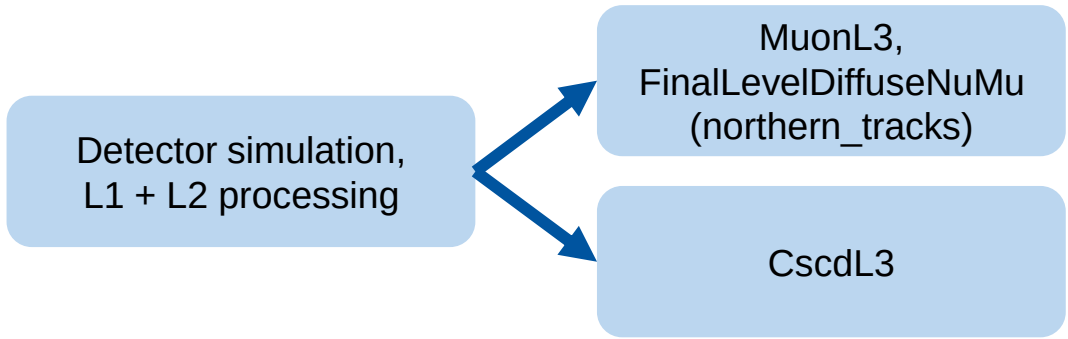
Systematic	Sampling Distribution	Range
IceWavePlusModes	2x 12 Gaussians	- - -
Scattering	uniform	
Absorption	uniform	
AnisotropyScale	uniform	
DOMEfficiency	uniform	
HoleIceForward_Unified	uniform	

Table: Overview of all available systematic perturbations for SnowStorm

SnowStorm Simulations – What happened since the Tokyo Workshop

- Produced several small-scale benchmark sets for testing validating the simulation chain and software
  - This process included several bugfixes like upside down DOMs, 1.35^2 more sensitive DeepCore DOMs, introduction of new style I3MCTrees, …
- Merge of the snobo branch into combo
- Production of a first set of large-scale SnowStorm MC production:
  - Perturbation of 5 detector systematics (table on the right)
    - Conservative uniform sampling distributions (+ranges) after a lot of discussions (diffuse + calibration group)
  - Added level3 (muon + cascade) and final level (DiffuseNuMu/northern_tracks) processing to the chain
    - Idea: have ready-to-use files from iceprod

Systematic	Sampling Distribution	Range
Scattering	uniform	[0.9, 1.1]
Absorption	uniform	[0.9, 1.1]
AnisotropyScale	uniform	[0.0, 2.0] (= 0-15%)
DOMEfficiency	uniform	[0.9, 1.1]
HoleIceForward_Unified	uniform	p0 [-1.0, +1.0] p1 [-0.2, +0.2]



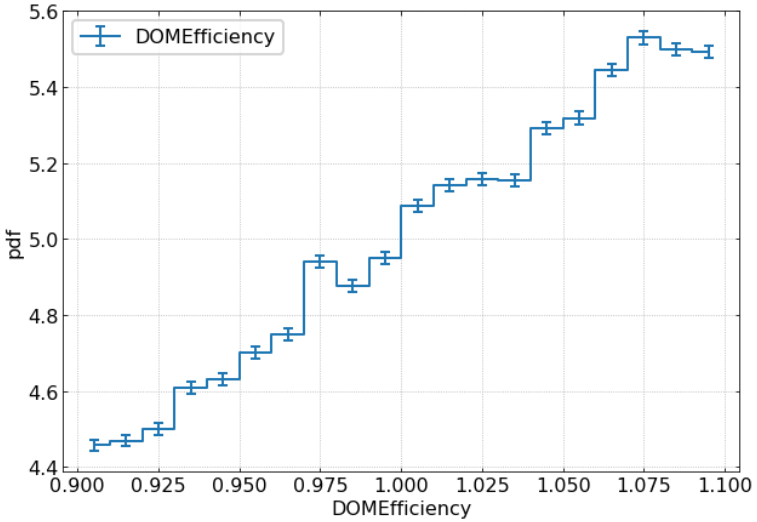
SnowStorm – Simulation Sets

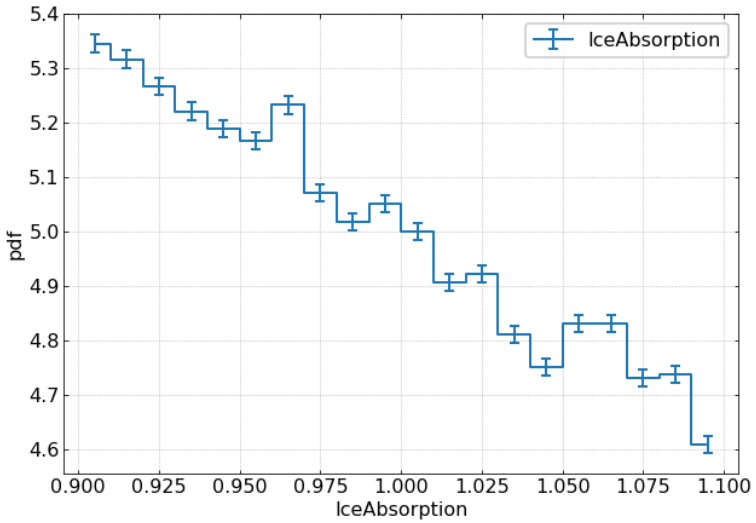
- “Final” all flavor production sets
- Energy split for better resource utilization during production
- Available files:
  - Generation (incl. polyplopia + muon prop)
  - Level2
  - Level3: Muon + Cascade
  - FinalLevel: DiffuseNuMu (northern_tracks)
- Wiki page with list of all sets and links to the files
- `MCPrescale = 1`
  - MC equivalent of the MinBias_Filter
  - Driven by C. Haack for #low-level-ml

Dataset ID	Flavour	Energy Range	Spectrum	Notes
21430	NuMu	1e2 – 1e4 GeV	1.5	
21431	NuMu	1e4 – 1e6 GeV	1.5	
21432	NuMu	1e6 – 1e8 GeV	1.0	DOM oversizing 3
21468	NuE	1e2 – 1e4 GeV	1.5	
21469	NuE	1e4 – 1e6 GeV	1.5	
21470	NuE	1e6 – 1e8 GeV	1.0	DOM oversizing 3
21471	NuE	1e2 – 1e4 GeV	1.5	
21472	NuE	1e4 – 1e6 GeV	1.5	
21473	NuE	1e6 – 1e8 GeV	1.0	DOM oversizing 3

SnowStorm Simulation Sets – Example

- SnowStorm MC yields a continuous function of observables/events with respect to the systematic parameters
  - “old-style”, discrete systematic sets would have given you only ~handful of points
- DOM efficiency/ice absorption distribution of events on FinalLevelDiffuseNuMu
- One can re-weight the sampling distribution to any other distribution on analysis level





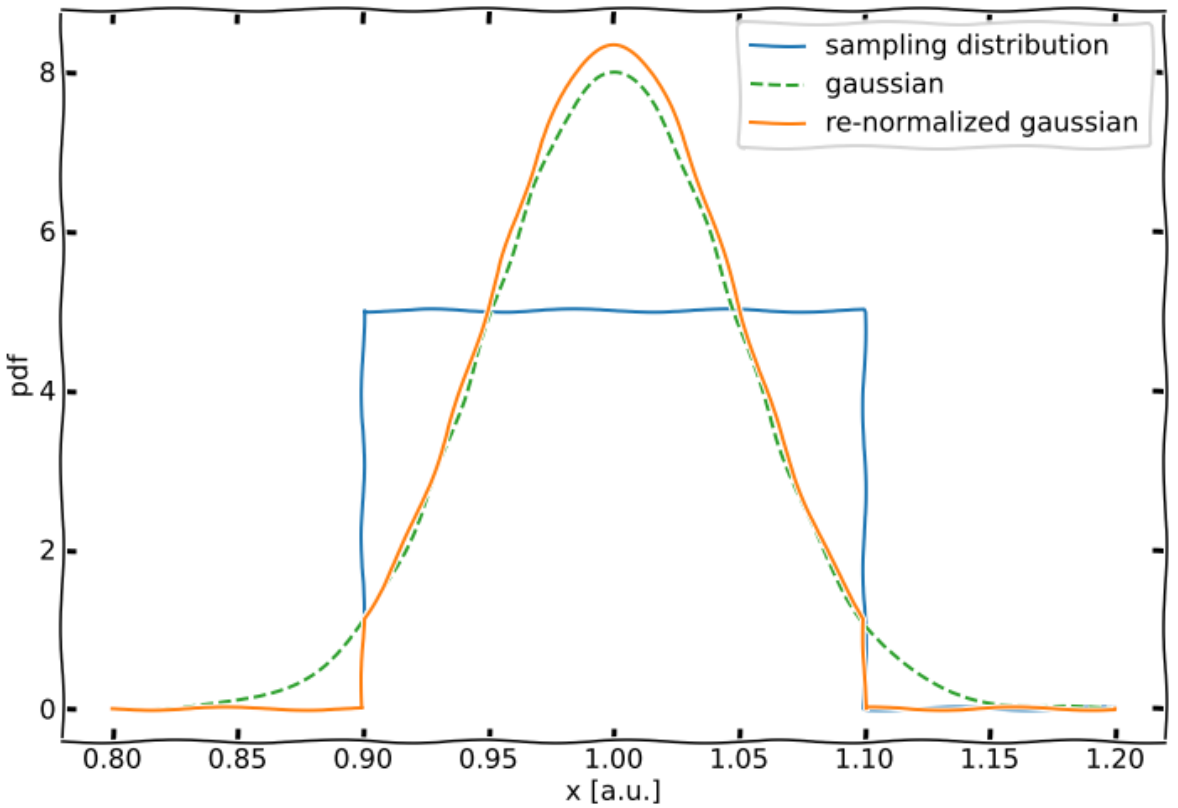
SnowStorm – Reweighting Method

- Re-weighting of the SnowStorm parameters within the MC set to some (arbitrary) distribution on analysis level (e.g. normal dist.)
  - Directly yields a MC prediction for a specific choice of nuisance parameters, i.e. the current hypothesis, in the fit

- Can easily be achieved by adding an additional weight to each MC event:

$$w_i = \frac{p_{sys}\,(sys_i, \xi_i)}{p_{sys}^{sim}(sys_i)} \cdot \ldots$$

p_{sys} : reweighting distribution
p_{sys}^{sim} : sampling distribution
sys_i : event's systematic value
ξ_i : nuisance parameter



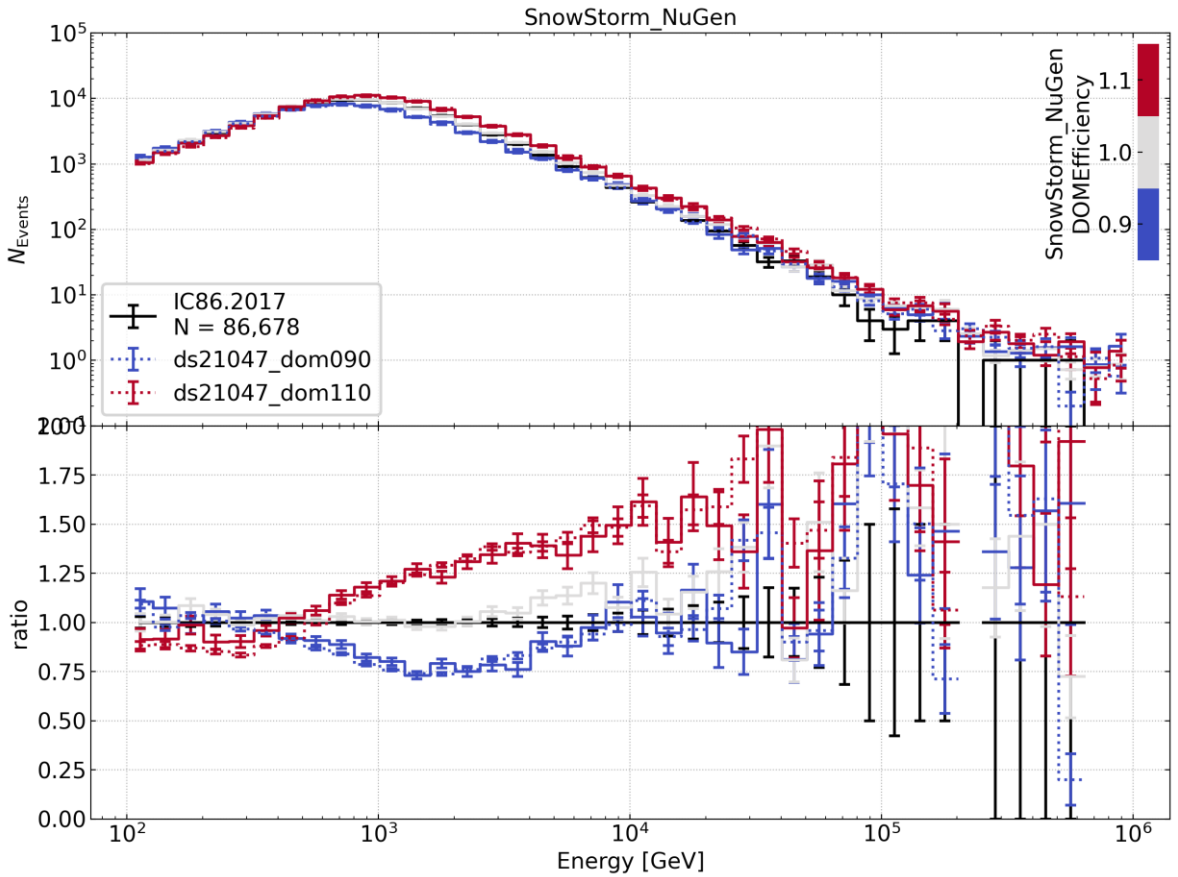
SnowStorm Simulation Sets – Example

- Truncated energy for FinalLevelDiffuseNuMu for different DOM efficiencies

- Solid lines: SnowStorm ensemble reweighted to $\pm$ 10% DOM efficiency
- Dashed lines: prediction of discrete simulation sets for the same ±10% variations
- Black line/ratio baseline: IC86.2017 data

➢ By reweighting the SnowStorm “event ensemble” one can reproduce the expectation from the “old style” discrete simulation sets

➢ How to make this plot? How to process SnowStorm simulations?



SnowStorm Simulation Sets – How To use them

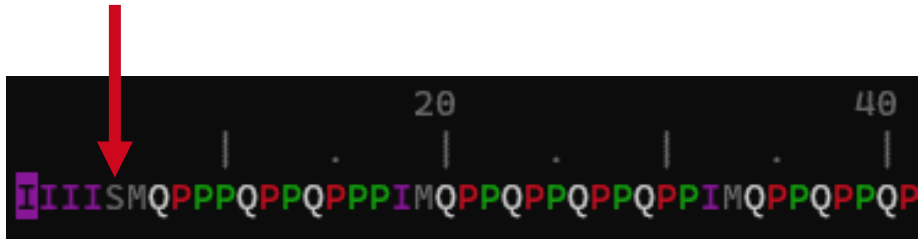
- All SnowStorm parameters, ice-model settings, etc. are stored in M and S frames
  - All SnowStorm simulations can be processed/used in the same way as previous/other NuGen simulations
  - ➢ Make sure to not drop M+S frames during further processing of the files

- You can use any recent combo/icetray version for processing that can deal with custom frame types
  - combo/V01-00-02 was used for production which is available in /cvmfs/icecube.opensciencegrid.org/…

SnowStorm Simulation Sets – How To use them

- All SnowStorm parameters, ice-model settings, etc. are stored in (new) M and S frames
  - All SnowStorm simulations can be processed/used in the same way as previous/other NuGen simulations
  - ➢ Make sure to not drop M+S frames during further processing of the files

- You can use any recent combo/icetray version for processing that can deal with custom frame types
  - combo/V01-00-02 was used for production which is available in /cvmfs/icecube.opensciencegrid.org/…



- S-Frame:
  - Only one per file
  - Simulation frame for bookkeeping:
    - initialized SnowStorm parametrizations (parameters)
    - SnowStorm sampling distributions

SnowStorm Simulation Sets – How To use them

- All SnowStorm parameters, ice-model settings, etc. are stored in (new) M and S frames
  - All SnowStorm simulations can be processed/used in the same way as previous/other NuGen simulations
  - ➢ Make sure to not drop M+S frames during further processing of the files

- You can use any recent combo/icetray version for processing that can deal with custom frame types
  - combo/V01-00-02 was used for production which is available in /cvmfs/icecube.opensciencegrid.org/…

- S-Frame:
  - Only one per file
  - Simulation frame for bookkeeping:
    - initialized SnowStorm parametrizations (parameters)
    - SnowStorm sampling distributions

- M-Frame:
  - One per every “bunch” of events
  - Sampled SnowStorm parameters
    - It also tells you how many events/frames the ice-model was applied to during simulation (before triggering)

SnowStorm Simulation Sets – How To use them

- All SnowStorm parameters, ice-model settings, etc. are stored in (new) M and S frames
  - All SnowStorm simulations can be processed/used in the same way as previous/other NuGen simulations
  - ➢ **Make sure to not drop M+S frames during further processing of the files**
- You can use any recent combo/icetray version for processing that can deal with custom frame types
  - combo/V01-00-02 was used for production which is available in /cvmfs/icecube.opensciencegrid.org/…

- S-Frame:
  - Only one per file
  - Simulation frame for bookkeeping:
    - initialized SnowStorm parametrizations (parameters)
    - SnowStorm sampling distributions
- M-Frame:
  - One per every “bunch” of events
  - Sampled SnowStorm parameters
    - It also tells you how many events/frames the ice-model was applied to during simulation (before triggering)

➢ How to extract parameters from the S and M-frame?

SnowStorm – Reading the Frame Objects

S(imulation)-Frame

- SnowstormParametrizations
  - List of used SnowStorm systematics/parametrizations
- SnowstormParameterRanges
  - List of tuples mapping [start : end] elements of the SnowStormParameters vector to the systematics
- SnowstormProposalDistribution
  - Vector containing the original sampling distributions in serialized form

M(odel)-Frame

- SnowstormEventsPerModel
  - Number of events this ice model was originally being applied to (on generation level, i.e. before triggering)
- SnowstormParameters
  - Vector of sampled ice-model parameters used for all following events until the next M frame occurs

SnowStorm – Reading the Frame Objects

S(imulation)-Frame

- SnowstormParametrizations
  - List of used SnowStorm systematics/parametrizations
- SnowstormParameterRanges
  - List of tuples mapping [start : end] elements of the SnowStormParameters vector to the systematics
- SnowstormProposalDistribution
  - Vector containing the original sampling distributions in serialized form

M(odel)-Frame

- SnowstormEventsPerModel
  - Number of events this ice model was originally being applied to (on generation level, i.e. before triggering)
- SnowstormParameters
  - Vector of sampled ice-model parameters used for all following events until the next M frame occurs

Complicated and not directly usable with the I3HDFWriter

SnowStorm – Reading the Frame Objects

- ✓ A „snowstorm parameter wrapper“ exists taking care of all the different frame objects
- ✓ It creates a “SnowstormParameterDict” with a simple mapping of parameter name → value in the M-Frame
  - Like the I3MCWeightDict from NuGen
  - Single frame object with all important information for the user
  - Works with I3HDFWriter
- ➢ This was not applied for Level2 during initial production but on later analysis level only…
- ➢ Will add this for the second round of SnowStorm MC production

M(odel)-Frame

- SnowstormEventsPerModel
  - Number of events this ice model was originally being applied to (on generation level, i.e. before triggering)
- SnowstormParameterDict

```
SnowstormParameterDict [I3Map<__cxx11::string, double>]:
[Absorption => 0.965954,
AnisotropyScale => 1.41439,
DOMEfficiency => 1.07651,
HoleIceForward_Unified_p0 => 0.844656,
HoleIceForward_Unified_p1 => 0.176947,
Scattering => 1.08156]
```

SnowStorm – Preliminary Summary

- ✓ Setup python 3 simulation chain for SnowStorm
- ✓ Production of several benchmark sets
  - ✓ Test, cross-checks, verification, etc.
- ✓ Merge of snowstorm software project + scripts into combo (icetray) for V01-00-00 release
- ✓ Production of a first batch of larger all flavor SnowStorm simulation sets
- ✓ Start investigating different concepts of using/dealing with SnowStorm MC in an analysis
- ✓ Slowly growing userbase of people working with those new SnowStorm sets

- ➢ Include the SnowStorm parameter wrapper in the simulation chain
  - ➢ Ideally the dictionary should be created during application of the SnowStorm perturber
  - ➢ Hands on session?

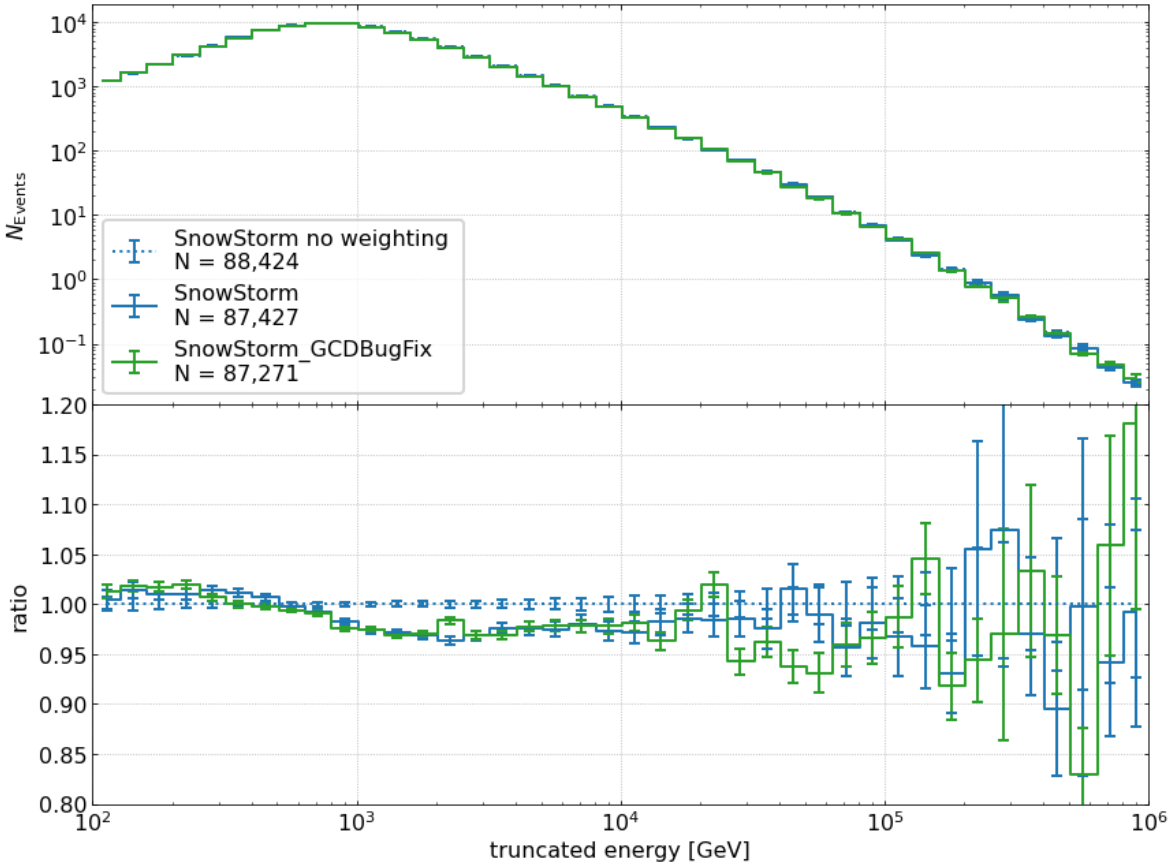
SnowStorm – Recent Developments

- ✓ Setup python 3 simulation chain for SnowStorm
- ✓ Production of several benchmark sets
  - ✓ Test, cross-checks, verification, etc.
- ✓ Merge of snowstorm software project + scripts into combo (icetray) for V01-00-00 release
- ✓ Production of a first batch of larger all flavor SnowStorm simulation sets
- ✓ Start investigating different concepts of using/dealing with SnowStorm MC in an analysis
- ✓ Slowly growing userbase of people working with those new SnowStorm sets

- ➢ Include the SnowStorm parameter wrapper in the simulation chain
  - ➢ Ideally the dictionary should be created during application of the SnowStorm perturber
  - ➢ Hands on session?
- ➢ During more detailed comparisons of SnowStorm and latest non-SnowStorm standard simulations using the full statistic available, some issues/differences showed up
- ➢ In December/January two bugs were discoverd almost in parallel
  - ➢ GCDFileBug (notlimited to SnowStorm MC)
  - ➢ Polyplopia bug (affecting only SnowStorm simulations)

SnowStorm – GCD File Bug

- The GCD file used in all 2020 simulations had some (all?) NaN values for the SPE peak position for multiple DOMs
- This bug affects almost all simulation sets produced in 2020, i.e. it is not limited to SnowStorm
- A new, fixed GCD file was generated by Juan Carlos
- Initiated production of new SnowStorm sets with this bugfix
  - Only NuMu fully completed, NuE and NuTau at ~50%
- Investigated the impact on FinalLevelDiffuseNuMu
  - Effect can be fully covered by a ca. -1% shift in DOM efficiency



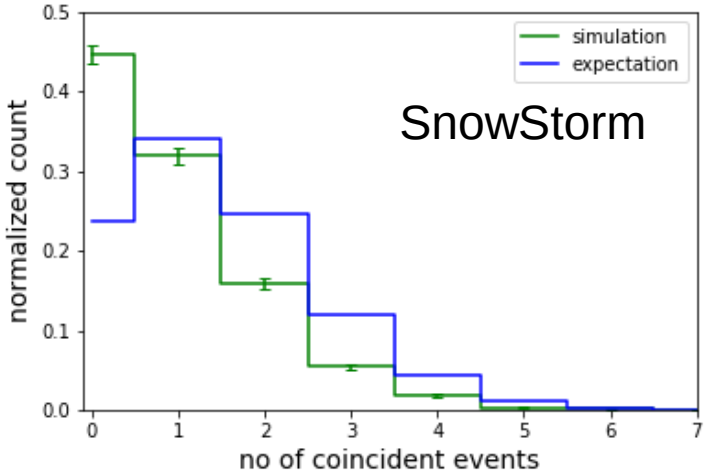
SnowStorm – Polyplopia Bug

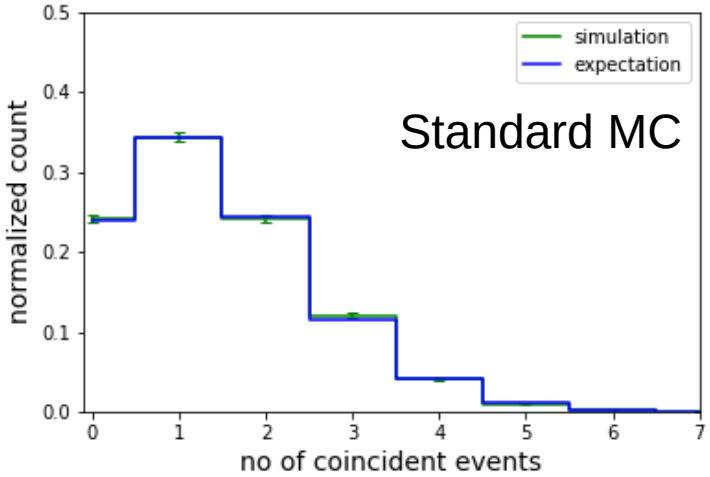
- However, there was still a difference between SnowStorm/standard simulations

- Found an issue with the combination/merge of signal and background simulation with Polyplopia:
  - For SnowStorm, less CR coincident events were merged (simulation) than expected from the plain CORSIKA rate

- Issue due to a misconfiguration of the polyplopia segment:
  - Polyplopia assumed a CORSIKA primary simulation and merged only N-1 coincidences…

➢ Fix applied/pushed to SnowSuite scripts
➢ Production of new benchmark sets





SnowStorm – Latest Benchmark sets

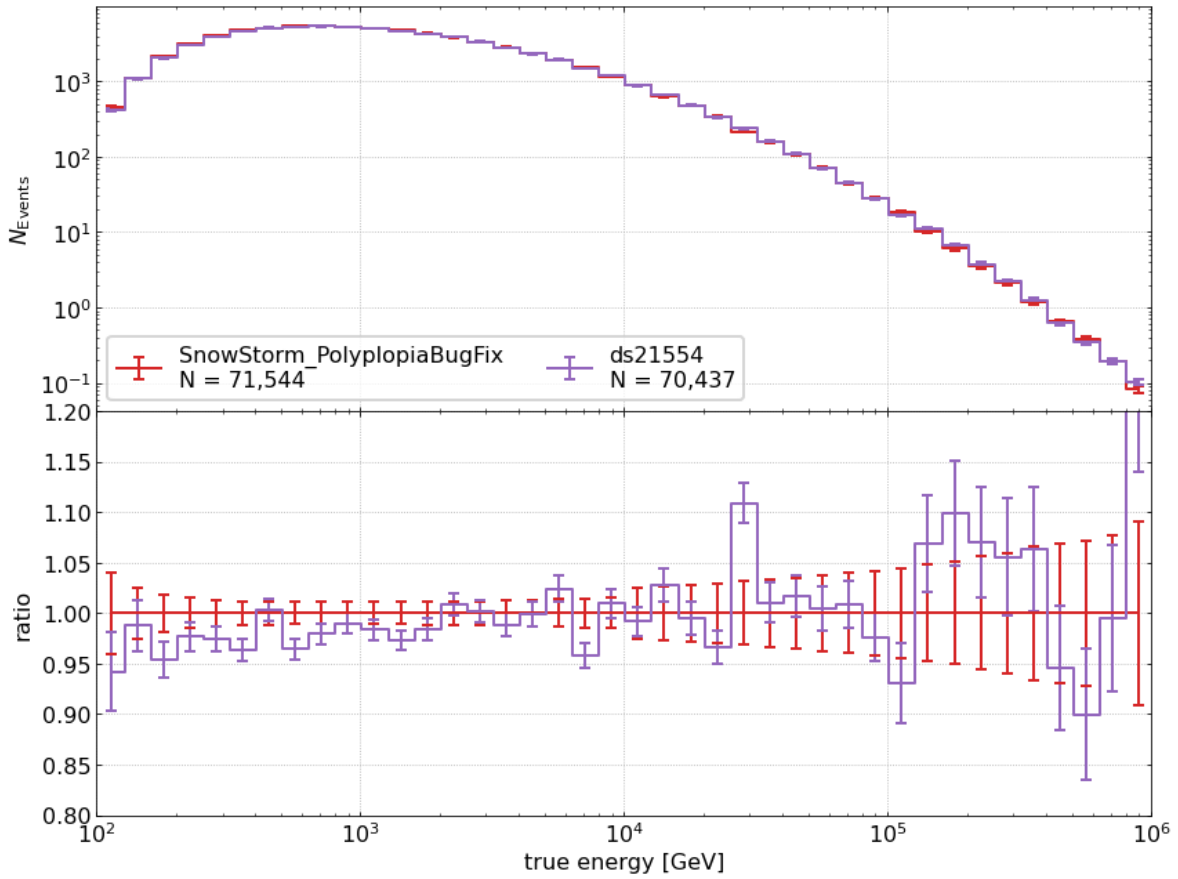
✓ Found and fixed two bugs after initial production

- Red: “latest” SnowStorm (GCD + Polyplopia bugfix)
- Purple: “latest” standard sim (GCD bugfix)

➢ Still looking suspicious below ~1e4 GeV ?

➢ Further investigate this? Sets are available:
Wiki page

➢ Produce new large scale SnowStorm simulations with our current “best knowledge”?
  ➢ Include additional improvements/updates?
  → *next talk/discussion item*



Summary and Outlook

- ✓ Merge of snowstorm software project + scripts into combo (icetray)
  - ✓ Included for V01-00-00 and later releases
- ✓ Production of a first batch of larger all flavor SnowStorm simulation sets
- ✓ Start investigating different concepts of using/dealing with SnowStorm MC in an analysis
- ✓ Slowly growing userbase of people working with those new SnowStorm sets
- ✓ Tool/framework development for using SnowStorm MC
- ✓ Found and fixed GCD file + Polyplopia bug
  - ✓ Bugfixes available only in smaller scale benchmark sets
- ➢ Second round of SnowStorm MC production?
  - ➢ Include GCD + Polyplopia bugfix
  - ➢ Wavedeform: Recently it was found that different versions producing different pulses are used for data and MC…
  - ➢ Update baseline ice model/adjust sampling distributions?
    - ➢ SnowStorm is based on Spice3.2.1, update to SpiceBFRv1 (v2 is not available in CLSim yet)
    - ➢ Next talk/discussion

Appendix

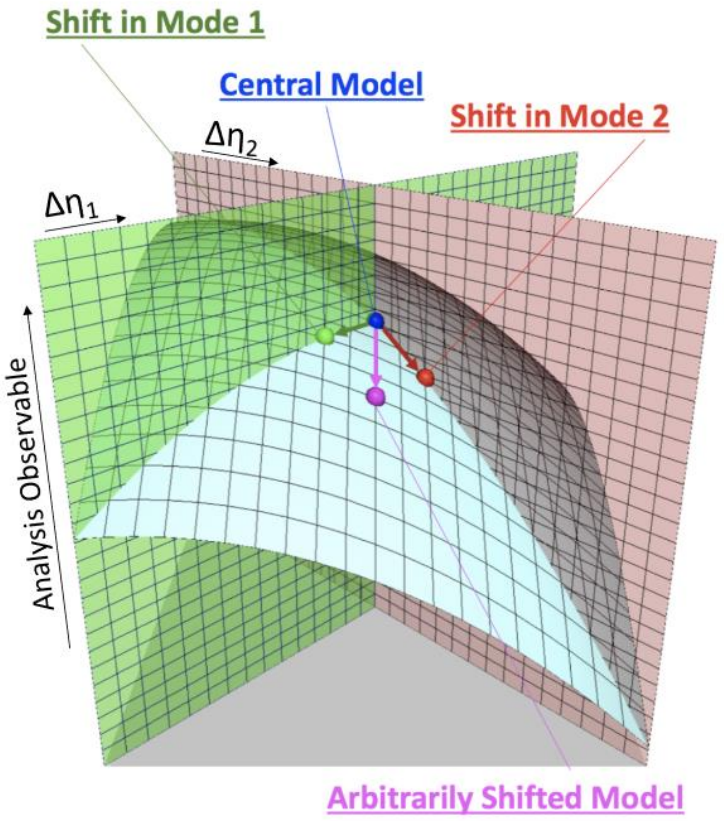
SnowStorm Aplication – Gradient Method

- (1) Use the Snowstorm MC set to extract the gradient in analysis space (chapter 2 in the paper)

- Assumptions:
  1. “effects of systematic uncertainties on analysis variables are sufficiently small that they can be treated perturbatively”
  2. “statistical uncertainty on Monte Carlo event counts is very small compared to that on the data”

- “If the effects of the nuisance $\vec{\eta}$ parameters are perturbative, this implies that the distribution function at any $\vec{\eta}$ can be written as a Taylor expansion around the central distribution:”

$$\psi_{\vec{\rho},\vec{\eta}} = \Psi_{\vec{\rho}} + \vec{\eta}.\vec{\nabla}_{\eta} \left[\psi_{\vec{\rho},\vec{\eta}}\right]_{\vec{\eta}=\vec{0}} + \mathcal{O}(\eta^2),$$

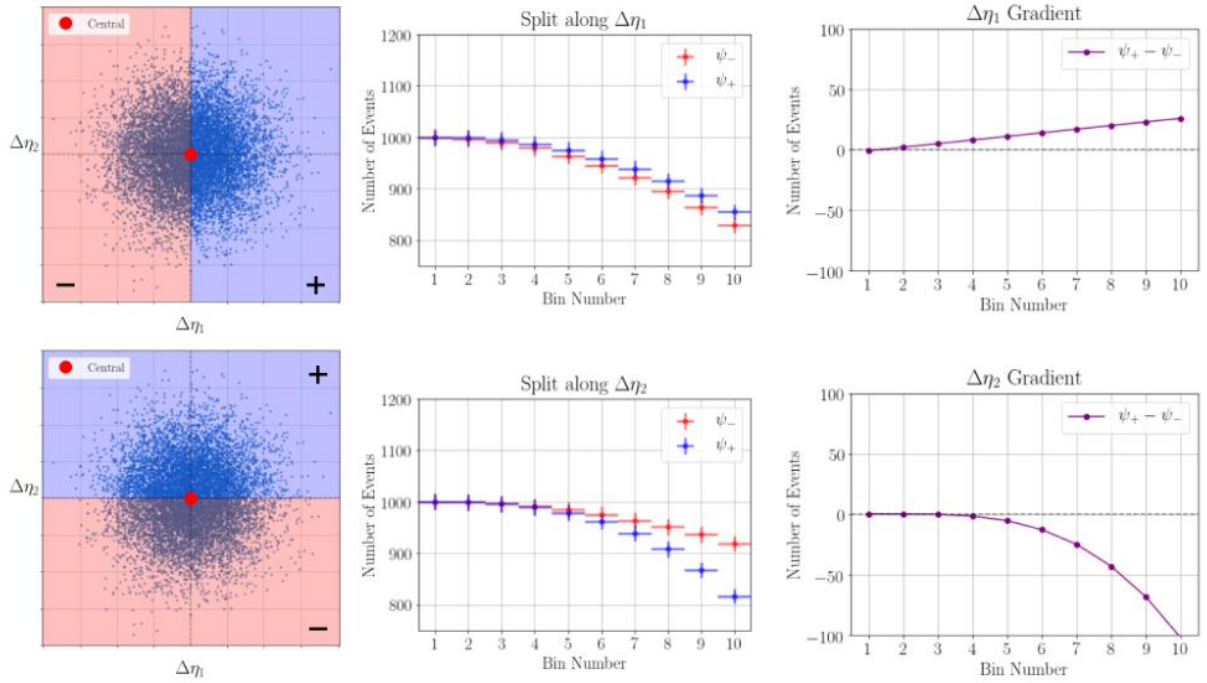


SnowStorm MC – Towards the Use in an Analysis – Gradient Method

- If “the prediction of the SnowStorm ensemble and the central model are identical $\mathcal{O}(\eta^2)$ effects” (and “comparison show[s] they are equivalent within the available statistical uncertainty”), this expression reduces to

$$\psi_{\vec{\rho},\vec{\eta}} = \Psi_{\vec{\rho}} + \vec{\eta}.\vec{G}_{\vec{\rho}}, \qquad \vec{G}_{\vec{\rho}} \equiv \vec{\nabla}_{\eta} \left[\psi_{\vec{\rho},\vec{\eta}}\right]_{\vec{\eta}=\vec{0}}.$$

- The “magic” part is to extract the gradient from the SnowStorm Ensemble:
  - Cutting the ensemble in half
  - By weighting: “consider constructing a prediction where each event in the SnowStorm ensemble is weighted by a factor of η_i”



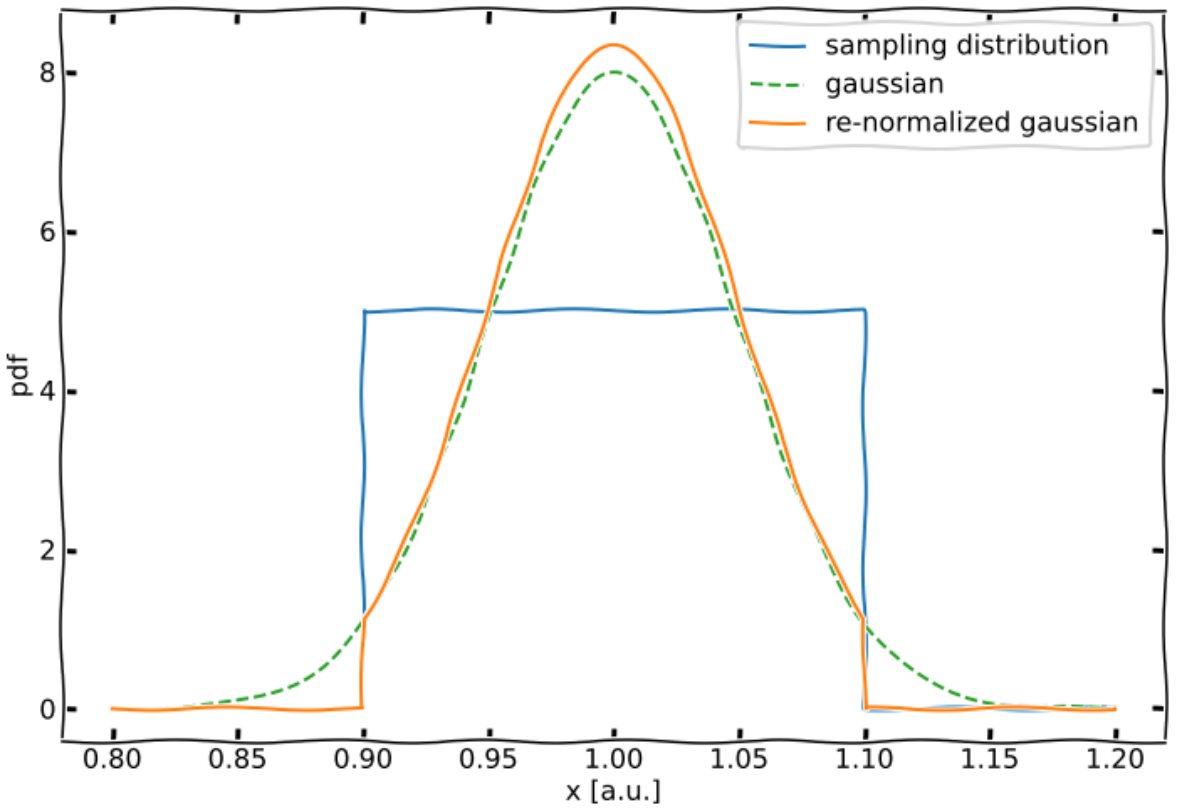
SnowStorm Application – Reweighting Method

- (2) Re-weighting of the SnowStorm parameters within the MC set to some (arbitrary) distribution on analysis level (e.g. normal dist.)
  - Directly yields a MC prediction for a specific choice of nuisance parameters, i.e. the current hypothesis, in the fit

- Can easily be achieved by adding an additional weight to each MC event:

$$w_i = \frac{p_{sys}\,(sys_i, \xi_i)}{p_{sys}^{sim}(sys_i)} \cdot \ldots$$

p_{sys} : reweighting distribution
p_{sys}^{sim} : sampling distribution
sys_i : event's systematic value
ξ_i : nuisance parameter



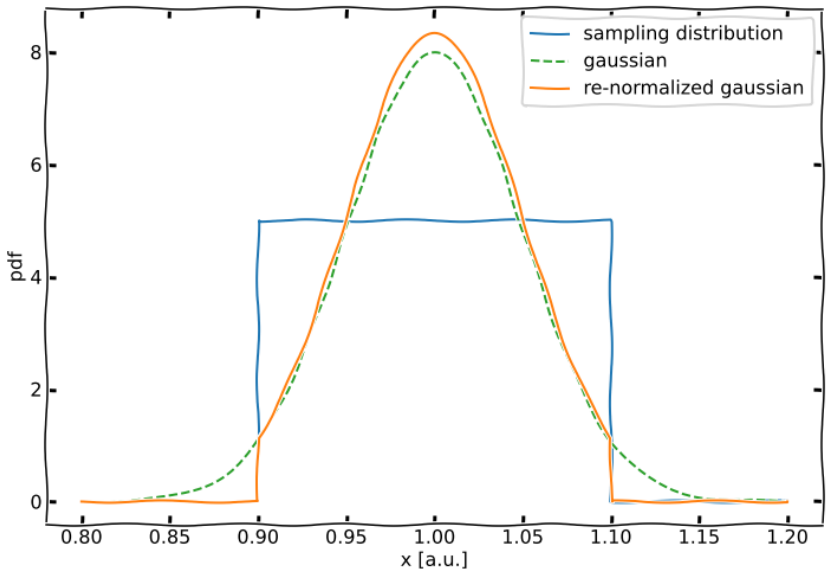
SnowStorm Application – Reweighting Method

- (2) Re-weighting of the SnowStorm parameters within the MC set to some (arbitrary) distribution on analysis level (e.g. normal dist.)
  - Directly yields a MC prediction for a specific choice of nuisance parameters, i.e. the current hypothesis, in the fit

- Can easily be achieved by adding an additional weight to each MC event:

$$w_i = \frac{p_{sys}\,(sys_i, \xi_i)}{p_{sys}^{sim}(sys_i)} \cdot \ldots$$

p_{sys} : reweighting distribution
p_{sys}^{sim} : sampling distribution
sys_i : event's systematic value
ξ_i : nuisance parameter



- Choice of the re-weighting distribution?

➢ For a normal distribution: One can show that in the limit of small widths the prediction is equivalent to a delta-function centered at $x_0 = \mu$

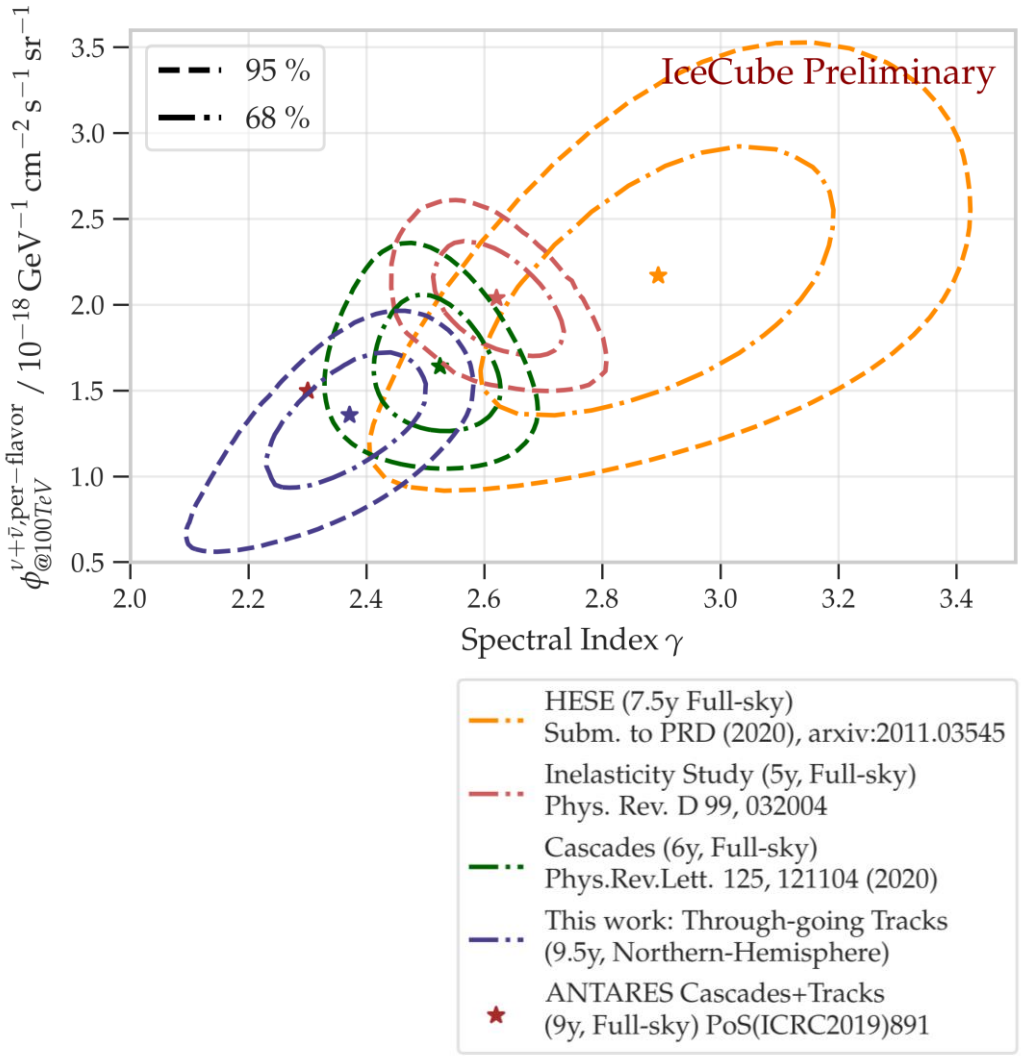
(sketch: width increased for better visualization)

Why SnowStorm for the GlobalFit?

- GlobalFit: Combination of multiple analyses/event selections into a single “global” fit of IceCube’s diffuse neutrinos

- For this, a consistent treatment of systematics uncertainties for all contributing analyses/event selections is crucial

- Advantages of SnowStorm:
  - No need to deal with multiple simulations sets:
    - 1x baseline + X discrete systematic sets
    - ➢ 1x SnowStorm sets which includes all systematics
  - Each (sub) analyses can “pick” the nuisance parameters/detector systematics it needs
    - ➢ Marginalize other ones if they are not important



SnowStorm – Polyplopia Bug

- However, there was still a difference between SnowStorm/standard simulations