Data Processing and I3Files

Jake Feintzeig

UW-Madison

Contents

- The Big Picture: Anatomy of an IceCube analysis
- The details
  - What types of data processing is there?
  - What is done at each level of processing?
  - How are the results of data processing used?

Anatomy of an IceCube Analysis

- Goal: **Observe Something**

Usually called "signal"

•Examples:
  - Neutrinos originating from a single point in the sky
  - Any astrophysical neutrinos
  - GZK neutrinos
  - Neutrinos associated with Dark Matter
  - Evidence for neutrino oscillations
  - Spatial distribution of cosmic rays

- Method:
  - Process/reduce data to keep signal and remove background
  - Do some sort of statistical test to decide whether you observe signal

Anatomy of an IceCube Analysis

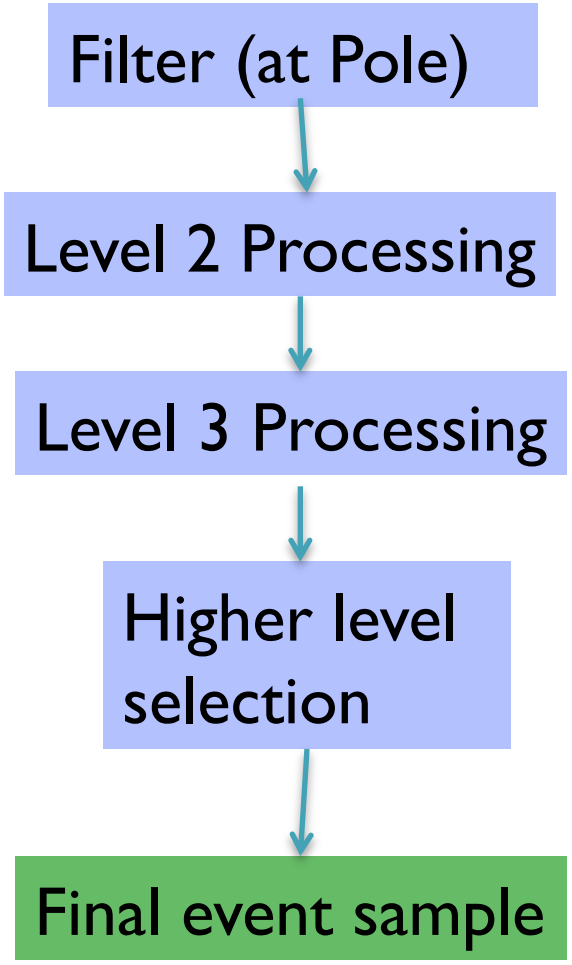
- Goal: **Observe Something**

- Steps:

Separate signal from background

- Apply algorithms to get particle's identity, energy, position direction

- Throw away particles that don't look like signal



Filters select interesting events at Pole to be sent North

- Does my event look like a muon track?
- A spherical cascade?
- Is it really really bright?
- Is it low-energy?
- Is it in the direction of the moon?

Filters must run in real-time and CPU resources at Pole are limited: Filters are a fast first pass through the data to throw away background and trim event rate

Filters select interesting events at Pole to be sent North

Everything you need to know:
http://wiki.icecube.wisc.edu/index.php/Trigger_Filter_Transmission_Board

Level 2 processing applies "higher-level" algorithms to data from all filters

- Once data is in the North, we have more time and computers to analyze data
  - Run more expensive hit cleanings and reconstructions
  - Run newer hit cleanings and reconstructions
- Collaboration-wide
  - Common algorithms – calibration of waveforms, feature extraction of pulses, basic reconstructions
  - All physics working groups (muons, cascades, EHE, etc.) contribute pieces of code that run stream-specific hit-cleanings and reconstructions

Examples of higher-level reconstructions

- Muon track directional reconstructions
  - SPEFit (one or more iterations)
  - MPEFit
- Muon energy reconstructions
  - MuE, MuEX
  - Truncated Energy
  - Millipede
- Cascade reconstructions
  - ACER, CascadeLlh, Credo, Monopod
- Low-energy reconstructions
  - FiniteReco

All these examples use likelihood functions!

Example: The Muon Stream

Muon filter (at Pole):

- Hit cleaning
- LineFit
- Single iteration SPEFit

Level2 processing (in North):

- Hit cleaning
- improvedLineFit
- Single iteration SPEFit
- 2 iteration SPEFit
- MPEFit
- MuE/MuEX/Truncated Energy

Example: The Low-Energy Stream

Low-energy filter (at Pole):	Level2 processing (in North):
• DC Hit cleaning • Linefit • Veto • Etc…	• DC Hit cleaning • improvedLineFit • Single iteration SPEFit • 2 iteration SPEFit • FiniteReco • Etc…

Level2 Resources

- What do the items in the frame mean:
  - http://wiki.icecube.wisc.edu/index.php/2011_Level2_file_structure
- What do the scripts look like?
  - http://code.icecube.wisc.edu/projects/icecube/browser/projects/std-processing/releases/V12-03-00
- Where do I find simulation datasets?
  - http://wiki.icecube.wisc.edu/index.php/Simulation_Production

The Level2a Processing: Because everyone makes mistakes

- In some past years, after the level2 processing was complete, bugs were found in certain steps (calibration, reconstruction settings, etc.), and the data was reprocessed
- IC79: Level2a is the dataset to use
- http://wiki.icecube.wisc.edu/index.php/IC79_Level2a_file_structure

Level 3 Filtering and Processing

- The output of level2 provides new information about the data that can be used to remove background
- There are some reconstructions that are too expensive to run on all the level2 data
- By the time the level2 data is processed, we may have created new reconstructions
- Working groups may want to apply a splitting technique

Level 3 Filtering and Processing

- Level 3 is working group specific
  - Working groups do their own level3
- The data is reduced by a factor of ~30 going from level2 → level3
  - Level3 cuts throw away ~95% of the data, the remaining 5% is processed
- Most analyzers will do further data reductions and event selections to get to their final sample
  - Cuts, Boosted Decision Trees (BDTs), etc.

Looking into some L2 Files

- /data/exp/IceCube/2011/filtered/level2/1111/Level2_IC86.2011_data_Run00118900_Part00000000.i3.bz2
  - Filter Masks
  - Different pulse series
  - Tracks
  - Fit params

Exercise

- Calculate filter rate, make plots of NCh, NString, Qtot, SPEFitSingle zenith and rlogl for one of the following filters
  - Muon
  - Cascade
  - DeepCore
  - Minimum Bias
  - EHE