Energy and $X_{\text{max}}$ reconstruction of Cosmic Rays using the Surface Radio Array of IceCube

Roxanne Turcotte for the IceCube Collaboration
Prototype station deployed in January 2020 over the IceCube detector.
Operating continuously (background and event data).
Air showers recorded!
Outline

• Implement a reconstruction method inspired by the technique developed by LOFAR \( \text{(Buitink, S., et al. Physical Review D 90.8 (2014): 082003.)} \)
• Try a similar model on our prototype with higher frequency band
• Work towards a log-likelihood
  • Start by a PDF study
• Results & Improvement prospects
The technique

- Consist of comparing data to a bunch of simulations
- The simulations intrinsically differ because of the shower-to-shower fluctuations which affect the $X_{\text{max}}$
- This method gives us an estimate on:
  - $X_{\text{max}}$ via the shape of the footprint
  - the energy via « scaling » the footprint
- Expected better reconstruction $\rightarrow$ higher frequency band 70-350MHz (other experiments use ~30-80MHz)
- Test on prototype station $\rightarrow$ Only 3 antennas!

We cover only a very small area of the footprint
Event selection

Subset of events

- 16 selected events
- From ~100 PeV to ~350 PeV
- Recorded between
  - 24.10.2020 - 26.01.2021
  - With associated scintillator & IceTop data
- No IceTop quality cut and most of the events are outside the calibrated space
Creation of a simulation set per event

- Create a set of air shower Monte-Carlo simulations for each event
- Use IceTop reconstruction as input for CORSIKA/CoREAS
- IceTop accuracy for our events:
  - Energy -> not characterised
  - Azimuth ~ 0.3º
  - Zenith ~ 0.2º-0.4º
  - Core ~ 2m-8m
- Each set has 100 simulations (50 H/50 Fe)
- The electronic response is added to the simulations —> Forward folding
- Use the star shape pattern (Coleman, A. arXiv: 2205.02258) and interpolate for the 3 antennas of the prototype station
- Waveforms filtered between 80-300MHz
Reconstruction of one simulated event

- Pick one simulation in the specific simulation set of one event
- Inject noise (real noise from prototype station) for the creation of the mock-up waveforms
  - Noisy waveforms
- Compare it to all the other simulations (with different $X_{\text{max}}$) to find the best fit
  - Quiet waveforms

Simulation set:
All simulations of the same event without noise

Inject noise on one simulation

Compare mock-up event with quiet waveforms
Finding the pulse and defining the observable

- Find pulses - same for quiet/noisy:
  - Define a 200ns window around where the pulse is expected
  - For each antenna, add the two channels together
  - Find maximum
  - Make a symmetrical window around it
  - Calculate average power of the Hilbert envelope of each channels individually
  - $\rightarrow$ reconstruction in the "data-space"!

$$P = \frac{1}{\Omega \cdot t_{tot}} \int_{-\Pi}^{\Pi} H^2(t)dt$$

CoREAS simulation — noisy waveform

CoREAS simulation — quiet waveform
Compare LDF - Minimisation

<table>
<thead>
<tr>
<th>Event Reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compare LDF to Simulation Set</td>
</tr>
<tr>
<td>Xmax reco.</td>
</tr>
<tr>
<td>Energy reco.</td>
</tr>
</tbody>
</table>

- Treat the two channels of the three antennas individually —> 6 - 1 d.o.f
- Minimise using $\chi^2$
- Find optimal $f$ - scaling factor for energy

\[ \chi^2 \text{ - Power Hilbert envelope} \]

\[ \chi^2_{\text{power}} = \sum_{i}^{\text{nCh}=6} \left( \frac{P_i - \sigma_i}{\sigma_i} - f^2 \cdot P_{\text{sim},i} \right)^2 \]

where,
- $\sigma_i$ = noise power in waveform $i$
- $P_i$ = average power in the pulse of noisy waveform $i$
- $P_{\text{sim},i}$ = average power in the pulse of quiet waveform $i$
- $f$ = free variable to fit

\[ E_{\text{CRreco}} \approx f \cdot E_{\text{sim}} \]

Example with mock-up event

**Xmax reconstructed value**

Fit a parabola on the 50 smallest $\chi^2$ values

**Energy reconstructed value**

Take $f$ from the lowest $\chi^2$ value
Estimation of the accuracy of reconstruction per event

The previous step is redone for all the simulations in the set -> estimated resolution per event
We do it for all the events of the subset
Bias in lower energy
Characteristics of event well reconstructed under investigation
- No clear dependance on zenith/azimuth/energy
- Probably lying on the Cherenkov-like ring
- Those having a core outside the IceTop array are in general badly reconstructed
Reminder that most of the events are outside the optimal IceTop reconstruction
- --> Better quality cut needed in the future!
Real measurements - 50ns window $\chi^2$

We see a similar trend in the real data.
About half of the events are reconstructable or very close.
The rest have most likely and initial reconstruction too far from the truth or the technique is not optimal (?) also the $\chi^2$ values are too high.
Investigating the PDF

- Studying the effect of the noise on the power
- Take all the pulses of the antennas in a star shaped layout —> many different pulses strength
- Inject 1000x some ideal noise on each pulse
  - ideal noise = cane expectation + thermal noise
- Each histogram represent one pulse

Increasing SNR
Behaviour of the Gaussian distributions

Gaussian mean with ideal noise (Cane + thermal from LNA)

\[ SNR = \frac{H_{\text{max}}^2}{\langle \text{RMS}_{\text{noise}} \rangle^2} \]

- Where,
  - \( \langle \text{noise}_{\text{RMS}} \rangle^2 \rightarrow \) RMS of the noise averaged over all waveforms
  - \( \langle \text{noise}_p \rangle^2 \rightarrow \) same using power instead of RMS
  - Calculated using a 128ns sub-window method

\[ P'_{\text{sim}} = P_{\text{sim}} + c \cdot (1 - d \cdot e^{-\beta x^a}) \]
Behaviour of the standard deviations

- The standard deviation of the noisy pulses is constant below $SNR \lesssim 1$
- Then increase as a function of the pulse strength
- Parametrised with a broken power law

$$\sigma(P_{\text{sim}}) = a \cdot \left(1 + \left(\frac{P_{\text{sim}}}{P_0}\right)^k\right)^{-\frac{1}{2}}$$
Migrating from $\chi^2$ to LLH

- Log-likelihood with the average power of Hilbert Envelope
  - Assuming Gaussian distribution everywhere
  - Using the functions $\sigma(f, P_{\text{sim}})$ and $P'_{\text{sim}}$ derived earlier

\[
\chi^2_{\text{power}} = \sum_i \left( \frac{(P_i - \sigma_i) - f_i^2 \cdot P_{\text{sim},i}}{\sigma_i} \right)^2
\]

From $\chi^2$ - Power Hilbert envelope

To LLH - Power Hilbert envelope

\[
-2 \ln \mathcal{L} = \sum_{i=0}^{n\text{Ch}=6} \left( 2\sqrt{2\pi} \ln \sigma(f, P_{\text{sim},i}) + \left( \frac{P_i - f^2 P'_{\text{sim},i}}{\sigma(f, P_{\text{sim},i})^2} \right)^2 \right)
\]

where,

\[
\sigma(f, P_{\text{sim},i}) = a \cdot \left( 1 + \left( \frac{f^2 P_{\text{sim},i}}{P_0} \right)^{k_f} \right)^{\frac{1}{2}}
\]

\[
P'_{\text{sim}} = P_{\text{sim}} + c \cdot \left( 1 - d \cdot e^{-\beta x^{\alpha}} \right)
\]
Results

• Hard to compare because...
  • There is still events with badly reconstructed in Xmax.
  • Also points towards better cuts
  • Need more and better events, nevertheless bias seems reduced even overshot

$\chi^2$ - real noise

LLH - real noise
Why is the parametrisation not optimal?

- Real noise
  - Bimodal distribution
  - Non-gaussian for small pulses
  - $\rightarrow$ 2d parametrisation for LLH!
- Reducing the noise
  - Better filtering (median filter)
  - Denoiser ML - Abdul
  - Hardware improvements

Real noise distribution $\rightarrow$ does not follow Gaussian

LLH - ideal noise (Cane + thermal from LNA)
Conclusion

- We implemented a technique based on the state-of-the-art analysis
  - which works in this higher frequency band and with 3 antennas
- LLH represent better the system with this observable
- It’s a prototype station, we are still learning about our data
- More events soon —> better quality cuts

Outlook

- Going for a 2D PDF parametrisation of noise
- Incorporating information about the spectrum and moving the core position
- Using information from the scintillation detectors as first input
- More stations !!
Back up slides
Methodology

Event Selection
- Recorded Data
- IceTop coinc.
  - yes
- Other Criterions
  - yes
- Radio Event
  - runID, eventID

Method Verification
- Pick one Simulated Event from Simulation Set
- Add Noise
  - Mock-up Data for runID, eventID

Creation Simulations Set
- IceTop Reconstruction
  - Energy
  - Azimuth
  - Zenith
  - Core
- 50x Proton
- Air Shower Simulation Set
- Preprocessing and Detector Response
- Radio Simulation Set
- Simulated Event

Event Reconstruction
- Compare LDF to Simulation Set
  - Energy reco.

Uncertainties Estimation
- Redo for all simulations in the set
  - Xmax true - Xmax reco.
  - E true - E reco. / E true
Maximum of the Hilbert envelope - Bias and standard deviation
2D PDF of noisy waveforms