Likelihood Free Frequentist Inference (LF2I) of atmospheric cosmic-ray showers

Alex Shen[1] Ann Lee[1] Luca Masserano[1] Tommaso Dorigo [2]
Michele Doro [2] [3] Federico Nardi [3]

[1]Department of Statistics and Data Science
Carnegie Mellon University

[2]Istituto Nazionale Fisica Nucleare

[3]Department of Physics and Astronomy
Università di Padova

April 17, 2023

Study of high energy cosmic rays remains challenging

[1]Source: swgo.org (Richard White, MPIK)

Study of high energy cosmic rays remains challenging

Useful for astrophysical research, however

[1]Source: swgo.org (Richard White, MPIK)

Study of high energy cosmic rays remains challenging

Useful for astrophysical research, however

1. Ground-based facilities only partially detect secondary showers

[1] Source: swgo.org (Richard White, MPIK)

Study of high energy cosmic rays remains challenging

Useful for astrophysical research, however

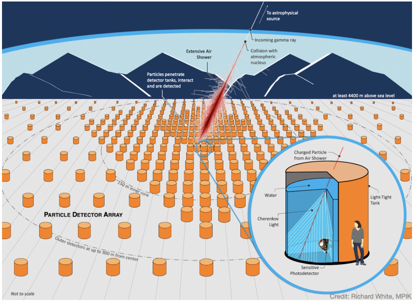
1. Ground-based facilities only partially detect secondary showers
2. Not every cosmic ray is informative (e.g. uncertain trajectory of charged particles)

[1] Source: swgo.org (Richard White, MPIK)

Study of high energy cosmic rays remains challenging

Useful for astrophysical research, however

1. Ground-based facilities only partially detect secondary showers
2. Not every cosmic ray is informative (e.g. uncertain trajectory of charged particles)



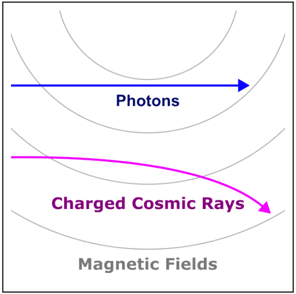
Proposed SWGO[1]

[1] Source: swgo.org (Richard White, MPIK)

Goals of the analysis

Given our observed detections associated with a secondary shower, we hope to

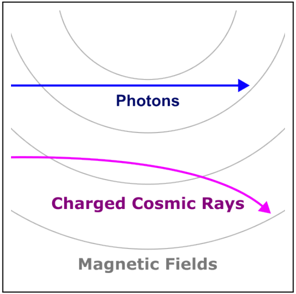
Given our observed detections associated with a secondary shower, we hope to



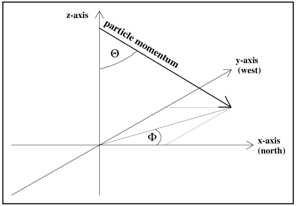
1. Identify *photonic* cosmic rays

Goals of the analysis

Given our observed detections associated with a secondary shower, we hope to



1. Identify *photonic* cosmic rays

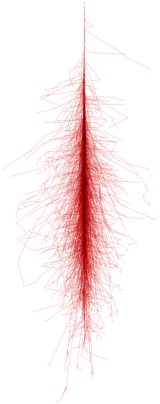


2. Estimate their orientation/energy

Data: CORSIKA Cosmic Ray Simulation Software

Simulation of secondary showers induced by cosmic ray interactions with atmosphere.

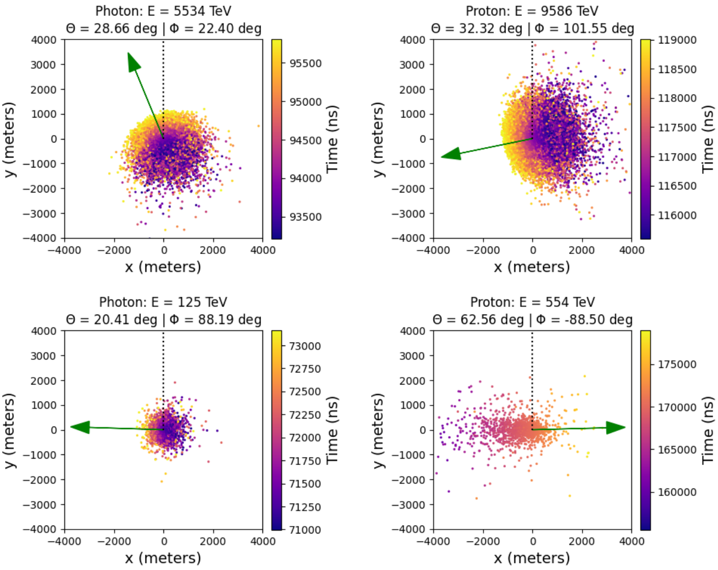
- Input: primary cosmic ray parameters
  - μ : particle identity
  - E : energy
  - Z, A : zenith, azimuthal angles
- Output X: identity, momenta, location, and timing of secondary particles observed at ground level



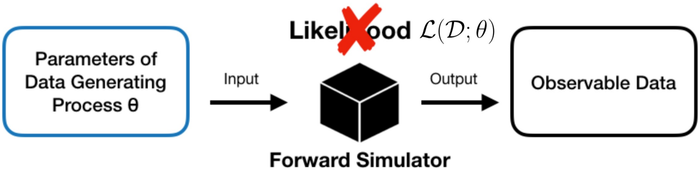
Example Simulation[1]

[1] Source: https://www.iap.kit.edu/corsika/

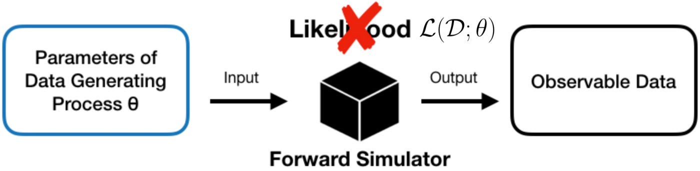
Data: CORSIKA Cosmic Ray Simulation Software



Goal 1: Setup

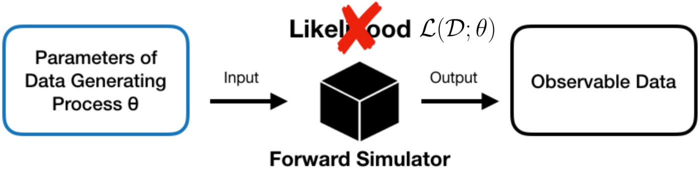

Likelihood $\mathcal{L}(\mathcal{D};\theta)$
Parameters of Data Generating Process θ
Input
Output
Observable Data
Forward Simulator

Goal 1: Setup



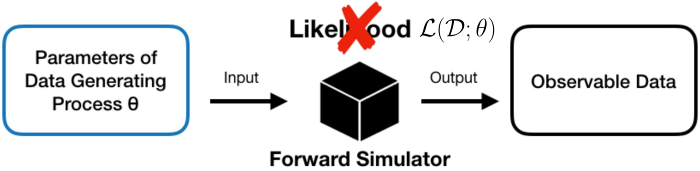
1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$

Goal 1: Setup



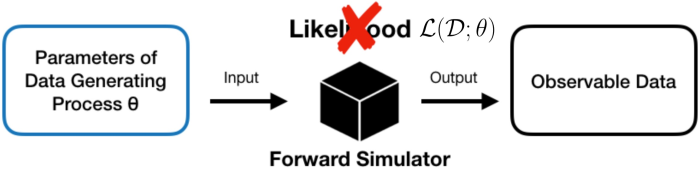
1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$
2. Simulated data $\{(\mu_1, \nu_1, X_1), ..., (\mu_B, \nu_B, X_B)\}$

Goal 1: Setup



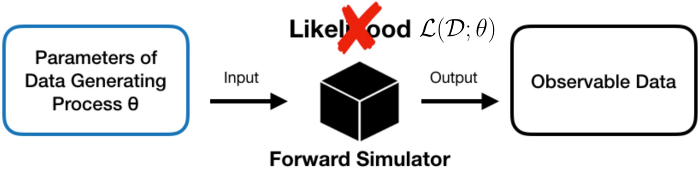
1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$
2. Simulated data $\{(\mu_1, \nu_1, X_1), ..., (\mu_B, \nu_B, X_B)\}$
3. Inference on μ (signal vs background)

Goal 1: Setup



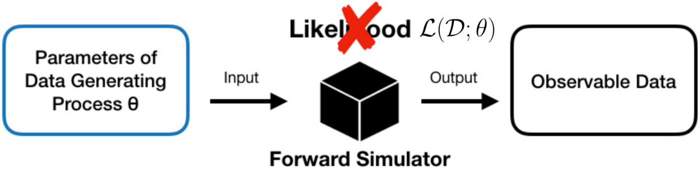
1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$
2. Simulated data $\{(\mu_1, \nu_1, X_1), ..., (\mu_B, \nu_B, X_B)\}$
3. Inference on μ (signal vs background)
   - Photonic CRs are very rare $\approx 1 : 10^5 \rightarrow$ *anomaly detection*

Goal 1: Setup



1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$
2. Simulated data $\{(\mu_1, \nu_1, X_1), ..., (\mu_B, \nu_B, X_B)\}$
3. Inference on μ (signal vs background)
   - Photonic CRs are very rare $\approx 1 : 10^5$ $\rightarrow$ *anomaly detection*
   - Type I error control over all $(\mu, \nu) \in \mathcal{M} \times \mathcal{N}$ $\rightarrow$ *nuisance parameterization*

Goal 1: Setup



1. Likelihood $\mathcal{L}(\theta) = \mathcal{L}(\mu, \nu)$ where $\nu = (E, Z, A)$
2. Simulated data $\{(\mu_1, \nu_1, X_1), ..., (\mu_B, \nu_B, X_B)\}$
3. Inference on μ (signal vs background)
   - Photonic CRs are very rare $\approx 1 : 10^5 \rightarrow$ *anomaly detection*
   - Type I error control over all $(\mu, \nu) \in \mathcal{M} \times \mathcal{N} \rightarrow$ *nuisance parameterization*
4. Directly predicting μ from X is insufficient

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Given an observed footprint X, we test

$$H_0 : \mu = 0 \text{ (proton)} \quad \text{vs} \quad H_a : \mu = 1 \text{ (photon)} \quad (\forall\, \nu \in \mathcal{N})$$

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Given an observed footprint X, we test

$$H_0 : \mu = 0 \text{ (proton)} \quad \text{vs} \quad H_a : \mu = 1 \text{ (photon)} \quad (\forall\, \nu \in \mathcal{N})$$

Test statistic (BFF)

$$\tau_1(X) = \frac{\Pr(X; \mu = 0)}{\Pr(X; \mu = 1)} = \frac{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 0, \nu) d\pi(\nu)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Given an observed footprint X, we test

$$H_0 : \mu = 0 \text{ (proton)} \quad \text{vs} \quad H_a : \mu = 1 \text{ (photon)} \quad (\forall\, \nu \in \mathcal{N})$$

Test statistic (BFF)

$$\tau_1(X) = \frac{\Pr(X; \mu = 0)}{\Pr(X; \mu = 1)} = \frac{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 0, \nu) d\pi(\nu)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ can be estimated via classification on simulated data

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Given an observed footprint X, we test

$$H_0 : \mu = 0 \text{ (proton)} \quad \text{vs} \quad H_a : \mu = 1 \text{ (photon)} \quad (\forall\, \nu \in \mathcal{N})$$

Test statistic (BFF)

$$\tau_1(X) = \frac{\Pr(X; \mu = 0)}{\Pr(X; \mu = 1)} = \frac{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 0, \nu) d\pi(\nu)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ can be estimated via classification on simulated data
2. Use quantile regression to estimate cutoffs $C_\alpha(\nu)$

Goal 1: Nuisance-parameterized Anomaly Detection via LF2I

Given an observed footprint X, we test

$$H_0 : \mu = 0 \text{ (proton)} \quad \text{vs} \quad H_a : \mu = 1 \text{ (photon)} \quad (\forall\, \nu \in \mathcal{N})$$

Test statistic (BFF)

$$\tau_1(X) = \frac{\Pr(X; \mu = 0)}{\Pr(X; \mu = 1)} = \frac{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 0, \nu) d\pi(\nu)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ can be estimated via classification on simulated data
2. Use quantile regression to estimate cutoffs $C_\alpha(\nu)$
3. Reject if $\tau(X) \leq \inf_{\nu \in \mathcal{N}} C_\alpha(\nu)$

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

For photonic cosmic rays, want a $1 - \alpha$ confidence set for $\nu = (E, Z, A)$.
Neyman style test inversion:

$$H_0 : \nu = \nu_0 \quad \text{vs} \quad H_a : \nu \neq \nu_0 \quad \forall\, \nu_0 \in \mathcal{N}$$

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

For photonic cosmic rays, want a $1-\alpha$ confidence set for $\nu = (E, Z, A)$. Neyman style test inversion:

$$H_0 : \nu = \nu_0 \quad \text{vs} \quad H_a : \nu \neq \nu_0 \quad \forall\, \nu_0 \in \mathcal{N}$$

Test Statistic (BFF)

$$\tau_2(X) = \frac{\Pr(X; \mu = 1, \nu = \nu_0)}{\Pr(X; \mu = 1)} = \frac{\mathbb{O}(X; \mu = 1, \nu = \nu_0)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

For photonic cosmic rays, want a $1-\alpha$ confidence set for $\nu = (E, Z, A)$.
Neyman style test inversion:

$$H_0 : \nu = \nu_0 \quad \text{vs} \quad H_a : \nu \neq \nu_0 \quad \forall\, \nu_0 \in \mathcal{N}$$

Test Statistic (BFF)

$$\tau_2(X) = \frac{\Pr(X; \mu = 1, \nu = \nu_0)}{\Pr(X; \mu = 1)} = \frac{\mathbb{O}(X; \mu = 1, \nu = \nu_0)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ is *already estimated from previous step*

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

For photonic cosmic rays, want a $1-\alpha$ confidence set for $\nu = (E, Z, A)$. Neyman style test inversion:

$$H_0 : \nu = \nu_0 \quad \text{vs} \quad H_a : \nu \neq \nu_0 \quad \forall\, \nu_0 \in \mathcal{N}$$

Test Statistic (BFF)

$$\tau_2(X) = \frac{\Pr(X; \mu = 1, \nu = \nu_0)}{\Pr(X; \mu = 1)} = \frac{\mathbb{O}(X; \mu = 1, \nu = \nu_0)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ is *already estimated from previous step*
2. Use quantile regression to estimate cutoffs $C_\alpha(\nu)$

Goal 2: Neyman Confidence Sets for Energy/Direction via LF2I

For photonic cosmic rays, want a $1-\alpha$ confidence set for $\nu = (E, Z, A)$. Neyman style test inversion:

$$H_0 : \nu = \nu_0 \quad \text{vs} \quad H_a : \nu \neq \nu_0 \quad \forall\, \nu_0 \in \mathcal{N}$$

Test Statistic (BFF)

$$\tau_2(X) = \frac{\Pr(X; \mu = 1, \nu = \nu_0)}{\Pr(X; \mu = 1)} = \frac{\mathbb{O}(X; \mu = 1, \nu = \nu_0)}{\int_{\mathcal{N}} \mathbb{O}(X; \mu = 1, \nu) d\pi(\nu)}$$

1. $\mathbb{O}(\cdot)$ is *already estimated from previous step*
2. Use quantile regression to estimate cutoffs $C_\alpha(\nu)$
3. Confidence Set: $\{\nu_0 : \tau(X) \geq C_\alpha(\nu)\}$

Summary

Cosmic ray reconstruction from ground based detector data

Summary

Cosmic ray reconstruction from ground based detector data

1. Identify photonic cosmic rays via LF2I nuisance parameterized anomaly detection

Summary

Cosmic ray reconstruction from ground based detector data

1. Identify photonic cosmic rays via LF2I nuisance parameterized anomaly detection
2. For photons, construct Neyman confidence sets for energy and orientation via LF2I

Summary

Cosmic ray reconstruction from ground based detector data

1. Identify photonic cosmic rays via LF2I nuisance parameterized anomaly detection
2. For photons, construct Neyman confidence sets for energy and orientation via LF2I

Thanks! Any questions?