

A Likelihood Method for Measuring the Ultrahigh Energy Cosmic Ray Composition

B.M. Connolly, S. BenZvi, A.C. O’Neill and S. Westerhoff
for the High Resolution Fly’s Eye (HiRes) Collaboration

Physics Department, Columbia University, New York, NY 10027

Presenter: A.C. O’Neill (oneill@phys.columbia.edu), usa-connolly-B-abs1-he14-poster

We present a method to measure the flux of various cosmic ray compositions by comparing the measured distribution of shower maxima with that of simulated distributions. The advantages of the method include its use of the full shower maximum distribution (rather than just the mean), and its calculation of a flux for various cosmic ray compositions. The method can be expanded to marginalize uncertainties due to choice of spectra, hadronic models, and atmospheric parameters.

1. Introduction

Due to the impracticality of direct observation of cosmic ray primary particles with energies $> 10^{19}$ eV, cosmic ray detectors must rely on indirect methods to understand the chemical composition of the ultra-high energy cosmic ray flux. In an air fluorescence experiment such as the High Resolution Fly’s Eye (HiRes) detector, which records air shower development as a function of atmospheric slant depth, information about primary particle composition can be inferred from observations of X_{max} , the depth at which the air shower particle count is a maximum. Showers induced by protons tend to reach the maximum of the shower development at larger slant depths than showers induced by heavier elements. However, since shower-to-shower fluctuations are large, a determination of the primary type for individual showers is not possible.

In the past, averaged quantities like the elongation rate — the mean height of shower maximum $\langle X_{max} \rangle$ versus the logarithm of primary particle energy — have been used successfully to understand the composition of the cosmic ray flux on a statistical basis [1, 2, 3]. Due to the relative insensitivity to small shifts in composition between selected low-Z or high-Z primaries, these observations have been reported historically as “proton-like” or “iron-like.” Though we expect ~ 100 g/cm² difference between $\langle X_{max} \rangle$ of protons and iron, large shower-to-shower fluctuations (~ 30 g/cm²) cause a significant overlap in their distributions.

Here, we present a method to answer a question not directly addressed by other techniques: in a fluorescence data sample where one has measured X_{max} for each event, what is the best estimate for the expected mean number of protons and iron from that experiment? Once the answer is obtained, the mean number can be converted to a flux for protons and iron.

This method has several advantages over previous approaches. First, the technique can be trivially expanded to evaluate the contribution of any number of compositions. Second, it uses more information than averaged quantities like the elongation rate, utilizing the shape of the X_{max} distribution rather than simply its mean. And third, the method allows the introduction of systematic uncertainties and prior knowledge (or lack thereof) concerning the composition spectra, hadronic generators, parameters concerning the extensive air shower, and atmospheric uncertainties.

In brief, the technique for extracting the cosmic ray composition essentially requires us to find the best normalizations for the “hypothesized” X_{max} distributions for protons and iron. By “best” we mean the sum of the expected mean events for the hypothesized distributions that best match the expected mean for the X_{max} distribution for a data sample D . The chance that the combined hypothesized distributions match the parent

distribution for data is proportional to the likelihood, $l(D|p, f)$, which is maximized to obtain the best estimates of p and f .

We develop the method using simulated stereoscopic data from the HiRes experiment in Utah, whose aspects and reconstruction are described in [4, 5, 6]. The method is tested with many independent data sets originating from a simulated parent set with p_{true} proton events generated with an $E^{-\alpha_p}$ spectrum and f_{true} iron events generated with an $E^{-\alpha_f}$ spectrum. Each bin of this parent distribution is fluctuated according to Poisson statistics, allowing us to build a set of independent simulated data distributions with which we test the predictions p and f . A general description of this technique is given in [7].

2. The Likelihood

To calculate p and f , the mean number of protons and iron, we maximize $l(D|p, f)$, the likelihood of the data sample D given p and f . To this end, Bayes' theorem allows us to convert the probability of a particular data sample D given the composition mixture to the probability of the composition mixture given the data:

$$P(p, f|D) = \frac{P(D|p, f)q(p, f)}{\sum_{p, f} P(D|p, f)q(p, f)}. \quad (1)$$

In this expression, $P(p, f|D)$ is the probability of obtaining some mean number of proton (p) and iron (f) primaries at a HiRes experiment from the sources in D ; $q(p, f)$ is the prior probability of p and f . As a technical aside, we split the prior $q(p, f)$ into two pieces

$$q(p, f) = Q(p, f)q'(p, f). \quad (2)$$

The variable $Q(p, f)$ represents prior information that we may have concerning the distributions for protons and iron. The variable $q'(p, f)$ represents all other prior information. For instance, if the likelihood were evaluated for various shifts in energy scales, atmospheric parameters, or the inelastic cross-sections or inelasticities, q' would be the relative probability of those shifts. In the present analysis, we choose

$$q'(p, f) = \begin{cases} 1, & p, f > 0, \\ 0, & p \text{ or } f < 0 \end{cases} \quad (3)$$

and effectively maximize $P(p, f)$ by maximizing the likelihood $l(D|p, f) \equiv P(D|p, f)q(p, f)$.

To calculate $l(D|p, f)$, we divide the X_{max} distributions into N bins. If p_i (f_i) represents the hypothesized mean proton (iron) count in the i^{th} bin, then the number of hypothesized data events in the i^{th} bin d_i is

$$d_i = p_i + f_i = \epsilon_p \tilde{p}_i + \epsilon_f \tilde{f}_i \quad (4)$$

where ϵ_p (ϵ_f) is the ratio of p (f) to the number of proton (iron) simulated events. Therefore, \tilde{p}_i and \tilde{f}_i are the number of hypothesized proton and iron events scaled to the luminosity of their respective simulated samples. Both \tilde{p}_i and \tilde{f}_i will be marginalized, i.e. integrated out, to eliminate the dependence on those parameters. In practice, $P(D|\epsilon_p, \epsilon_f)$ is calculated instead of $P(D|p, f)$.

The probability for the observed distribution of data events given p, f, \tilde{p}_i and \tilde{f}_i is

$$P(D|p, f, \tilde{p}_i, \tilde{f}_i) = \left(\frac{d_i^{D_i} e^{-d_i}}{D_i!} \right) \left(\frac{\tilde{p}_i^{P_i} e^{-\tilde{p}_i}}{P_i!} \right) \left(\frac{\tilde{f}_i^{F_i} e^{-\tilde{f}_i}}{F_i!} \right) \quad (5)$$

where P_i and F_i are the number of proton and iron events in the i^{th} bin of the simulated samples, respectively. Marginalizing \tilde{p}_i and \tilde{f}_i and multiplying the probabilities for all the bins, we can define the likelihood $l(D|p, f)$:

$$l(D|p, f) = P(D|p, f)q(p, f) \quad (6)$$

$$= \prod_{i=1}^{N \text{ bins}} \int d\tilde{p}_i \int d\tilde{f}_i \left(\frac{d_i^{D_i} e^{-d_i}}{D_i!} \right) \left(\frac{\tilde{p}_i^{P_i} e^{-\tilde{p}_i}}{P_i!} \right) \left(\frac{\tilde{f}_i^{F_i} e^{-\tilde{f}_i}}{F_i!} \right). \quad (7)$$

2.1 Note About Uncertainties

To artificially separate the systematic and statistical uncertainties, one can, for example, remove any prior knowledge concerning the uncertainties in our models for protons and iron (i.e. remove $Q(p, f, \tilde{p}_i, \tilde{f}_i)$ from $l(D|p, f, \tilde{p}_i, \tilde{f}_i)$) and define statistical uncertainty as

$$l_{stat}(D|p, f) \equiv \prod_{i=1}^N \int d\tilde{p}_i \int d\tilde{f}_i \frac{d_i^{D_i} e^{-d_i}}{D_i!} \quad (8)$$

Then, by calculating the total uncertainty (σ_{tot}) from $l(D|p, f)$ and σ_{stat} from $l_{stat}(D|p, f)$, the systematic uncertainty (σ_{sys}) can be deduced.

3. Eliminating Spectral Dependence in the Likelihood

In practice, the energy spectra for protons and iron are unknown. To include our lack of prior knowledge of the energy spectrum, a slightly more sophisticated likelihood is used, such that Eq. (1) becomes

$$P(p, f|D) = \frac{\sum_{\alpha_p=-2,-3} \sum_{\alpha_f=-2,-3} P(D|p, f, \alpha_p, \alpha_f)q(p, f, \alpha_p, \alpha_f)}{\sum_{p, f, \alpha_p, \alpha_f} P(D|p, f, \alpha_p, \alpha_f)q(p, f, \alpha_p, \alpha_f)} \quad (9)$$

Other priors such those concerning atmospheric parameters can be introduced and marginalized in a similar manner. Since we have assumed a flat prior, maximizing the probability $P(p, f|D)$ is to maximize the likelihood

$$l(D|p, f, \alpha_p, \alpha_f) \equiv P(D|p, f, \alpha_p, \alpha_f)q(p, f, \alpha_p, \alpha_f). \quad (10)$$

The likelihood is now

$$l(D|p, f) = \sum_{\alpha_p=-2,-3} \sum_{\alpha_f=-2,-3} \prod_{i=1}^{N \text{ bins}} \int d\tilde{p}_i \int d\tilde{f}_i \left(\frac{d_i^{D_i} e^{-d_i}}{D_i!} \right) \left(\frac{\tilde{p}_i^{P_i} e^{-\tilde{p}_i}}{P_i!} \right) \left(\frac{\tilde{f}_i^{F_i} e^{-\tilde{f}_i}}{F_i!} \right). \quad (11)$$

4. Discussion

It should be noted that this method does not take the place of an elongation rate measurement, but rather answers the specific question: given a set of hypotheses about protons, iron and gamma primaries that includes their interactions with the atmosphere and how those interactions are observed in the detector, what are their relative contributions to the data? In contrast, the elongation rate is unaffected by such assumptions; one is free to interpret what the measured elongation rate might say about the characteristics of air showers or the relative abundances of certain primaries. The results of applications of this method to simulated HiRes stereo data will be presented at the conference. We will also show first results from applications of the method to HiRes stereo data.

5. Acknowledgments

This work is supported by US NSF grants PHY-9321949, PHY-9322298, PHY-9904048, PHY-9974537, PHY-0098826, PHY-0140688, PHY-0245428, PHY-0305516, PHY-0307098, and by the DOE grant FG03-92ER40732. We gratefully acknowledge the contributions from the technical staffs of our home institutions. The cooperation of Colonels E. Fischer and G. Harter, the US Army, and the Dugway Proving Ground staff is greatly appreciated.

References

- [1] D.J. Bird *et al.*, Phys. Rev. Lett. 71 (1993) 3401.
- [2] T. Abu-Zayyad *et al.*, Phys. Rev. Lett. 84 (2000) 4276.
- [3] R.U. Abbasi *et al.*, Astrophys. J., in press (2005) (arXiv:astro-ph/0407622).
- [4] J. Boyer *et al.*, Nucl. Instr. Meth. A 482 (2002) 457.
- [5] P.A. Sadowski *et al.*, Astropart. Phys. 18 (2002) 237.
- [6] J.N. Matthews *et al.*, Proc. of 28th ICRC, Tsukuba, Japan, 350 (2003).
- [7] P. Bhat, H. Prosper, and S. Snyder, Phys. Lett. B 407 (1997) 73.
- [8] R.U. Abbasi *et al.*, Astrophys. J. 610 (2004) L73.
- [9] K. Shinozaki *et al.*, Astrophys. J. 571 (2002) L117.
- [10] T. Abu-Zayyad *et al.*, Phys. Rev. Lett. 84 (2000) 4276.