

New approach to primary mass composition analysis with simultaneous use of ground and fluorescence detectors data

A. Yushkov*, M. Ambrosio*, C. Aramo*, F. Guarino*[†], D. D'Urso*, L. Valore*

*INFN Sezione di Napoli, via Cintia, Napoli, Italia 80125

[†]Università di Napoli "Federico II", via Cintia, Napoli, Italia 80125

Abstract. We study the possibility to reconstruct primary mass composition with the use of combinations of basic shower characteristics, measured in hybrid experiments, such as depth of shower maximum from fluorescence side and signal in water Cherenkov tanks or in plastic scintillators from the ground side. To optimize discrimination performance of shower observables combinations we apply Fisher's discriminant analysis and give statistical estimates of separation of the obtained distributions on Fisher variables for proton and iron primaries. At the final stage we apply Multiparametric Topological Analysis to these distributions to extract composition from prepared mixtures with known fractions of showers from different primary particles. It is shown, that due to high sensitivity of water tanks to muons, combination of signal in them with X_{\max} looks especially promising for mass composition analysis, provided the energy is determined from longitudinal shower profile.

Keywords: mass composition, hybrid data, Fisher's discriminant

INTRODUCTION

The experimental information on UHECR mass composition coming from different experiments and from different mass reconstruction techniques is quite contradicting [1], [2]. One of the main difficulties is that at these energies mass composition and hadronic interactions properties are both unknown and are deeply entangled. The necessary condition for the solution of this problem is the reconciliation of the results on mass composition obtained from different types of ground and fluorescence data between themselves and with astrophysical predictions on the origin of the anisotropy, 'ankle' and GZK cut-off. Hybrid experiments, like Pierre Auger Observatory [5] or Telescope Array (TA) [6], are perfectly suitable for this purpose, since they provide the opportunity to use combinations of extensive air shower (EAS) parameters to achieve the best possible mass resolution. The key role in this analysis can be played by the muon shower content — the most problematic for hadronic models [3] and the best mass sensitive EAS parameter. The upgrade of Auger with AMIGA scintillator counters array [4] is aimed right at the muon content measurement, but it is easy to show (see Section I) that already the total signal in the Auger water tanks preserves the difference between primaries

in number of muons and can be profitable for primary mass reconstruction, provided the energy is independently determined from the longitudinal shower profile. In the case of TA, which will be in grade to measure only charged particles density, ground data alone will be weekly sensitive to primary particle mass and idea of the use of EAS observables combinations becomes indispensable. Using Auger and TA as examples, in this paper we put forward a strategy allowing to reconstruct primary mass composition from combinations of the fluorescence and ground data keeping in mind the limitations on the affordable simulation statistics of the UHECR showers.

I. GENERAL NOTES ON THE CHOICE OF MASS DISCRIMINATION PARAMETERS

In the following we consider cases of Auger and TA to estimate the expected performance of the proposed mass reconstruction technique. We assume that primary energy can be estimated from the longitudinal shower profile and hence is practically primary mass independent. Briefly speaking, to enhance primary mass resolution of traditional X_{\max} parameter we suggest to use it in linear combinations with other basic shower parameters, such as signal in water tanks or particle density for Auger and TA correspondingly.

The data set used for the analysis was generated with CORSIKA 6.204 [7] (QGSJET 01 [8]/Gheisha [9]) and CORSIKA 6.735 (QGSJET II [10]/Fluka2008.3 [11]) packages and contains 1000 showers for every primary (p, O, Fe) and interaction model at 10 EeV and 37° zenith angle. All longitudinal showers characteristics and charged particles density were taken directly from CORSIKA output files. The calculations of the expected signal in Auger water tanks was performed according to the procedure described in [12], [13] with the use of the same GEANT 4 lookup tables as in [13]. In case of TA we use charged particles density at 1000 m from the axis (D_{1000}) as an example, for densities at another distances the consideration line would be the same. Finally let us note, that qualitatively results for both combinations of interaction models used in the study are very similar, so below we will mostly discuss only results for QGSJET 01/Gheisha, which provides some worse discrimination performance. To characterize the separation of distributions we will use the merit factor $MF = |\bar{x}_{\text{Fe}} - \bar{x}_{\text{p}}| / \sqrt{\sigma_{\text{Fe}}^2 + \sigma_{\text{p}}^2}$, where $\bar{x}_{\text{p, Fe}}$ and

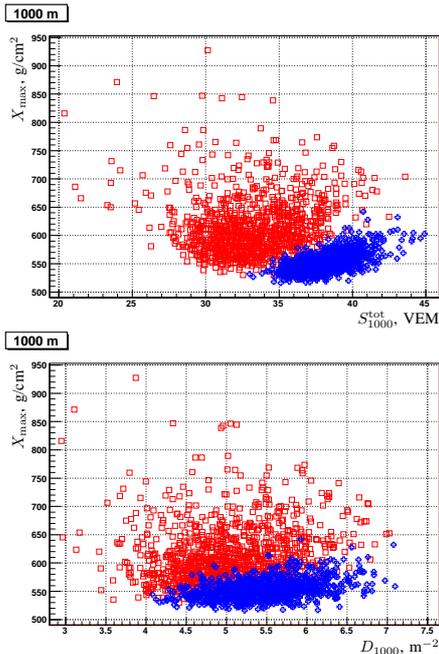


Fig. 1. Depth of shower maximum versus total ground plane signal and particle density for proton (red squares) and iron (blue crosses) showers at 1000 meters from the axis for QGSJET 01 model.

$\sigma_{p, Fe}$ are distributions means and standard deviations correspondingly.

In Fig. 1 we present scatter plots of total ground plane signal and charged particles density vs X_{max} at 1000 m from the shower axis. Good separation of iron and proton showers in $(S_{1000}^{tot}, X_{max})$ plot is both due to discrimination power of X_{max} and to noticeable difference $\sim 13\%$ in the average total signals. In absolute units this difference is the same as the difference between muon signals, but the separation of the total signals (MF=1.4) is surely worse than that of the muon ones (MF=2.5) due to the smearing effect of the electromagnetic component. Despite of change with the distance of the (electromagnetic/muon) signal ratio, a good separation of protons and iron nuclei is kept in a wide range from 700 to 1500 meters, since high sensitivity of Cherenkov tanks to the muon component results in different shifts of p and Fe populations along signal axis in X_{max} vs total signal scatter plots. On the other hand, as expected, the separation between primaries in (D_{1000}, X_{max}) plot is mostly due to discrimination power of X_{max} , since distributions on charged particles density for protons and iron nuclei largely overlap (MF=0.5). In this case one can think of searching for another discrimination parameters combinations, but as it will be shown below, (D_{1000}, X_{max}) pair provides one of the best (among possible within TA conditions) discrimination resolutions.

II. FISHER'S DISCRIMINANT ANALYSIS

The problem of primary particle mass discrimination with the use of combination of two or more shower characteristics falls in the class of standard tasks of

statistical pattern classification analysis (see e.g. [14], [15], [16]) and one of its methods – linear discriminant analysis – was recently applied to study the classification capability of longitudinal profile distribution parameters [17]. Using the Toolkit for Multivariate Data Analysis (TMVA) [16] here we perform a similar study for combinations of different fluorescence and ground data.

As it was already discussed, p and Fe populations are well separated in both examples in Fig. 1, and what's more, they can be separated with high accuracy even by a straight line. Hence, it is opportune to apply in this case just linear discriminant analysis and namely Fisher's method. In this approach one seeks the direction along which two classes will be separated the best, i.e. one looks for the direction in parameters hyperspace, after projection on which the ratio of the distance between distributions means to the sum of their squared variations will be maximized. The evident advantage of this approach is possibility to use any number of parameters avoiding “dimensionality curse” (thus reducing necessary simulations statistics) and to apply easily any further classification tools to the resulting one-dimensional distributions. In addition to Fisher's discriminant the performance of rectangular cut optimization, projective likelihood estimator and function discriminant analysis with quadratic and cubic functions [16] were checked and it was found that none of them outperforms Fisher's approach.

To find the direction, along which the primaries will be separated in the optimal way Fisher's algorithm requires minimum training: already 5–10 events of every primary type can be enough to achieve the same results as in the case of the use of several hundreds events. After application of Fisher's method one gets the new variable which is simply the linear combination of original variables that provides the optimal separation in one-dimensional case. To characterize discrimination capability of different parameters combinations after application of Fisher's technique in Table I we give for them merit factors MF, areas A , separations $\langle S^2 \rangle$ and misclassification rates ξ . Taking protons as ‘signal’ and iron nuclei as ‘background’, one can consider A as area under signal efficiency versus background rejection curve [16] (called also receiver operating characteristics curve [14]), the closer this area to unity, the better the classification is. Separation is defined in [16] as

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy,$$

where $\hat{y}_S(y)$ and $\hat{y}_B(y)$ are the probability density functions for signal and background, $\langle S^2 \rangle = 1$ again means the best separation and corresponds to distributions without overlap. The misclassification rate, used in addition to these statistical variables, is calculated in a very simple way to estimate possible error in event-by-event classification approach. We fit overlapping sides of distributions on Fisher variables for protons and irons

with Gaussian functions and consider all events to the left of intersection point of these Gaussian fits as iron and all other events as protons. In this case some number ξ_p of proton events is recognized as irons and, vice versa, some number ξ_{Fe} of irons is classified as protons.

From Table I one can see, that discrimination for both high energy interaction models is very similar, though in case of QGSJET II the separation of p and Fe is slightly better, especially for combinations of total signal and X_{max} with LDF slope parameter. Certainly, combination of X_{max} with muon signal at 1000 meters provides the best discrimination, but as one can see combinations of depth of shower maximum with the total signal in the tanks in the range 700–1500 meters also provide excellent separation of primaries with misclassification of only $\sim 30 - 50$ events out of 2000. Further addition to this couple of other shower characteristics does not improve significantly the discrimination capability and in the case of the real data can be completely useless due to presence of additional systematic errors, though the situation can change with energy and zenith angle, of course. Taking into account robustness of total signal at 1000 m to LDF reconstruction uncertainties the combination $(X_{max}, S_{1000}^{tot})$ in our view looks as the optimal choice for primary mass composition analysis in Auger experimental conditions. Table I also shows, that despite of weak discrimination power of charged particles density, its use together with X_{max} allows to achieve separation of primaries with MF=1.44 (for QGSJET 01), while for X_{max} distributions alone merit factor is equal to 1.16. At the considered energy and zenith angle (D_{1000}, X_{max}) pair looks like the best choice for primary mass reconstruction with TA.

Certainly, our conclusions on $(S_{1000}^{tot}, X_{max})$ and (D_{1000}, X_{max}) as the best mass discrimination combinations are specific only for the energy and zenith angle discussed, in the sense that for another energies/angles addition of other parameters to these basic pairs may be helpful in optimization of their mass discrimination performance.

III. EXTRACTION OF COMPOSITION FROM TEST SAMPLES WITH MULTIPARAMETRIC TOPOLOGICAL ANALYSIS

The basic idea behind the Multiparametric Topological Analysis (MTA) [18] resides in the classification of showers from different primaries according to their topological distribution in multiparametric space. Considering Fig. 1 one can divide the plane $(X_{max}, S_{1000}^{tot})$ in a number of cells and find probabilities for the showers falling in some particular cell to be initiated by proton or iron. Using only these probabilities on the pure set of proton showers one will erroneously arrive (in case of the overlap of p and Fe populations) to mixed composition. To correct such misclassification it is also necessary to compute mixing probabilities [18], determining the chance of event from one primary mass in the given cell to be misclassified as event of another

TABLE I
DISCRIMINATION PERFORMANCE OF DIFFERENT SHOWER PARAMETERS COMBINATIONS AFTER APPLICATION OF FISHER'S DISCRIMINANT ANALYSIS.

| QGSJET 01 | | | | | |
|--|-------|-----------------------|------|---------|------------|
| Parameters | Area | $\langle S^2 \rangle$ | MF | ξ_p | ξ_{Fe} |
| $[S_{1000}^{\mu}, X_{max}]$ | 1.000 | 0.995 | 2.53 | 10 | 2 |
| $[S_{700}^{tot}, X_{max}]$ | 0.996 | 0.908 | 1.90 | 35 | 15 |
| $[S_{1000}^{tot}, X_{max}]$ | 0.996 | 0.932 | 2.02 | 32 | 14 |
| $[S_{1500}^{tot}, X_{max}]$ | 0.997 | 0.940 | 2.18 | 33 | 14 |
| $[D_{1000}, X_{max}]$ | 0.957 | 0.677 | 1.44 | 139 | 65 |
| $[LDF \beta, X_{max}]$ | 0.925 | 0.578 | 1.29 | 184 | 97 |
| $[S_{1000}^{tot}, LDF \beta]$ | 0.934 | 0.627 | 1.49 | 172 | 78 |
| $[X_{max}, S_{1000}^{tot}, LDF \beta]$ | 0.997 | 0.956 | 2.08 | 20 | 7 |
| $[X_{max}, S_{1000}^{tot}, N_{max}]$ | 0.999 | 0.946 | 2.16 | 25 | 11 |
| QGSJET II | | | | | |
| Parameters | Area | $\langle S^2 \rangle$ | MF | ξ_p | ξ_{Fe} |
| $[S_{1000}^{\mu}, X_{max}]$ | 1.000 | 0.985 | 2.70 | 11 | 1 |
| $[S_{700}^{tot}, X_{max}]$ | 0.999 | 0.961 | 2.18 | 24 | 8 |
| $[S_{1000}^{tot}, X_{max}]$ | 0.996 | 0.942 | 2.25 | 28 | 7 |
| $[S_{1500}^{tot}, X_{max}]$ | 0.994 | 0.937 | 2.32 | 21 | 11 |
| $[D_{1000}, X_{max}]$ | 0.975 | 0.770 | 1.65 | 99 | 54 |
| $[LDF \beta, X_{max}]$ | 0.947 | 0.674 | 1.51 | 135 | 75 |
| $[S_{1000}^{tot}, LDF \beta]$ | 0.952 | 0.718 | 1.64 | 124 | 73 |
| $[X_{max}, S_{1000}^{tot}, LDF \beta]$ | 0.999 | 0.966 | 2.36 | 17 | 4 |
| $[X_{max}, S_{1000}^{tot}, N_{max}]$ | 0.997 | 0.953 | 2.33 | 23 | 5 |

primary mass. Hence, to get both types of probabilities one has to use two independent sets of simulated events. To illustrate classification capability of MTA combined with discrimination power of Fisher's method, we have performed primary composition reconstruction of sample mixtures with known fractions of protons, oxygen and iron nuclei. In Figs. 2, 3 we present the results of MTA application to one-dimensional distributions on Fisher's variables $F(X_{max}, S_{1000}^{tot})$ and $F(X_{max}, D_{1000})$. The composition is very well reproduced when one uses $(X_{max}, S_{1000}^{tot})$ combination, with errors of 2–3% for [p, Fe] and and 3–5% for [p, O] mixtures. The discrimination power of (D_{1000}, X_{max}) couple is surely worse (errors are 3–5% for [p, Fe] and and 8–10% for [p, O] mixtures) and in case of real experimental conditions with additional systematic errors its primary mass classification performance can be of limited use in the case when [p, O] mixture is considered.

IV. CONCLUSIONS

The present study allows to develop a new approach to the mass composition analysis of hybrid data. We propose to use combinations of longitudinal and lateral parameters to achieve maximum primaries separation in multiparametric space. Further application of Fisher's method optimizes discrimination, reducing the problem to one-dimensional case and allowing for lower simulation statistics. At the last stage one can apply different algorithms to extract mass composition from distributions on Fisher variables, which are more mass sensitive in comparison with e.g. traditionally used X_{max} alone. We applied for this purpose MTA technique to the samples

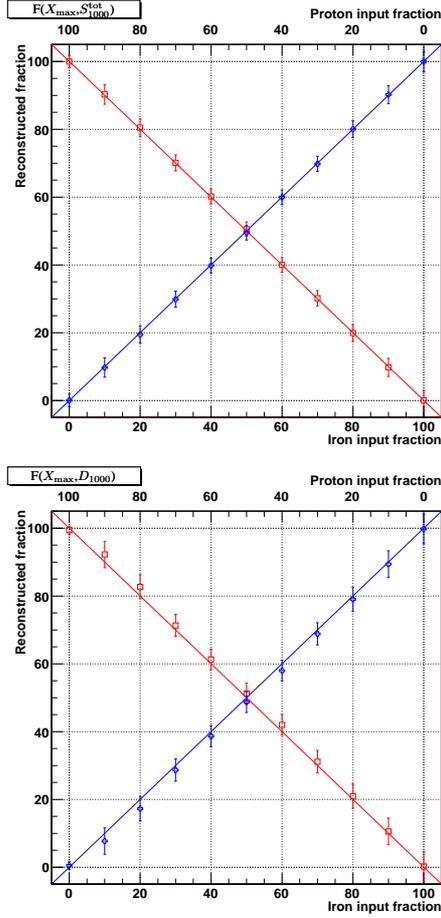


Fig. 2. Reconstructed with MTA on the basis of Fisher's variables $F(X_{\max}, S_{1000}^{\text{tot}})$ and $F(X_{\max}, D_{1000})$ distributions proton (red squares) and iron (blue crosses) abundances in the samples with known primaries content. Lines mark the exact reconstruction results.

with different primaries fractions and retrieved with very good accuracy nuclei abundances from [p, Fe] and [p, O] mixtures.

Regarding the choice of mass sensitive parameters, it was shown, that for Auger the best mass discrimination can be achieved if to use $(X_{\max}, S_{1000}^{\text{tot}})$ pair, provided the primary energy is estimated from the longitudinal shower profile. The charged particles density measured in TA in combination with depth of shower maximum also provides good discrimination of proton and iron showers, though in the real experimental conditions its sensitivity seems to be limited for proton-oxygen mixture case.

ACKNOWLEDGMENTS

We are very grateful to Maximo Ave and Fabian Schmidt for kind permission to use their GEANT 4 lookup tables in our calculations of signal from different particles in Auger water tanks.

REFERENCES

- [1] Maria Giller, *J. Phys.*, vol. G35, pp. 023201, 2008.
- [2] Graciela B. Gelmini, arXiv: 0903.4716 (2009).

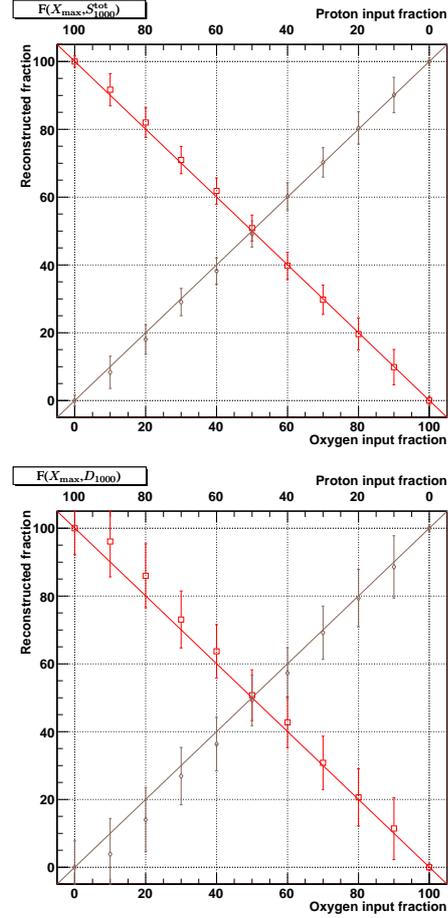


Fig. 3. Same as in Fig. 2, but for proton (red squares) – oxygen (brown diamonds) mixtures.

- [3] Ralph Engel, in *Proc. of the 30th ICRC*, Merida, 2007, vol. 4, pp. 385–388.
- [4] A. Etchegoyen, in *Proc. of the 30th ICRC*, Merida, 2007, vol. 5, pp. 1191–1194.
- [5] J. Abraham, M. Aglietta, I. C. Aguirre, et al., *Nucl. Instrum. Meth.*, vol. A523, pp. 50–95, 2004.
- [6] H. Kawai, S. Yoshida, H. Yoshii, et al., *Nucl. Phys. B (Proc. Suppl.)*, vol. 175–176, pp. 221–226, 2008.
- [7] D. Heck, J. Knapp, J. N. Capdevielle, et al., Forschungszentrum, Karlsruhe, 1998, **FZKA 6019**.
- [8] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, *Nucl. Phys. B (Proc. Suppl.)*, vol. 52, pp. 17–28, 1997.
- [9] H. Fesefeldt, RWTH, Aachen, 1985, **PITHA 85/02 (1985)**.
- [10] S. Ostapchenko, *Nucl. Phys. B (Proc. Suppl.)*, vol. 151, pp. 143–146, 2006.
- [11] G. Battistoni, S. Muraro, P. R. Sala, et al., *AIP Conf. Proc.*, vol. 896, pp. 31–49, 2007; A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, CERN-2005-010.
- [12] Pierre Billoir, *Astropart. Phys.*, vol. 30, pp. 270–285, 2008.
- [13] Fabian Schmidt, Maximo Ave, Lorenzo Cazon, and Aaron S. Chou, *Astropart. Phys.*, vol. 29, pp. 355–365, 2008.
- [14] Andrew R. Webb, *Statistical pattern recognition*, Chichester, England: John Wiley & Sons Inc., 2002.
- [15] L. I. Kuncheva, *Combining pattern classifiers*, Hoboken, USA: John Wiley & Sons Inc., 2004.
- [16] <http://tmva.sourceforge.net/>.
- [17] F. Catalani, J. A. Chinellato, Vitor de Souza, et al., *Astropart. Phys.*, vol. 28, pp. 357–365, 2007.
- [18] M. Ambrosio, C. Aramo, C. Donalek, et al., *Astropart. Phys.*, vol. 24, pp. 355–371, 2005; D. D'Urso, M. Ambrosio, C. Aramo, F. Guarino, and L. Valore, *Nucl. Instrum. Meth.*, vol. A588, pp. 171–175, 2008.