

UNIVERSIDADE ESTADUAL DE CAMPINAS
SISTEMA DE BIBLIOTECAS DA UNICAMP
REPOSITÓRIO DA PRODUÇÃO CIENTÍFICA E INTELECTUAL DA UNICAMP

Versão do arquivo anexado / Version of attached file:

Versão do Editor / Published Version

Mais informações no site da editora / Further information on publisher's website:

[https://www.epj-](https://www.epj-conferences.org/articles/epjconf/abs/2023/09/epjconf_uhecr2023_03003/epjconf_uhecr2023_03003.html)

[conferences.org/articles/epjconf/abs/2023/09/epjconf_uhecr2023_03003/epjconf_uhecr2023_03003.html](https://www.epj-conferences.org/articles/epjconf/abs/2023/09/epjconf_uhecr2023_03003/epjconf_uhecr2023_03003.html)

DOI: 10.1051/epjconf/202328303003

Direitos autorais / Publisher's copyright statement:

©2023 by EDP Sciences. All rights reserved.

DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo

CEP 13083-970 – Campinas SP

Fone: (19) 3521-6493

<http://www.repositorio.unicamp.br>

Update on the indication of a mass-dependent anisotropy above $10^{18.7}$ eV in the hybrid data of the Pierre Auger Observatory

Eric Mayotte^{1,2,*} and Thomas Fitoussi³ for the Pierre Auger Collaboration^{4,†}

¹Colorado School of Mines, Department of Physics, Golden, CO, USA

²Bergische Universität Wuppertal, Department of Physics, Wuppertal, Germany

³Karlsruhe Institute of Technology (KIT), Institute for Astroparticle Physics, Karlsruhe, Germany

⁴Observatorio Pierre Auger, Av. San Martín Norte 304, 5613, Malargüe, Argentina.

Full author list: https://www.auger.org/archive/authors_2022_10.html

Abstract. We test for an anisotropy in the mass of arriving cosmic-ray primaries associated with the galactic plane. The sensitivity to primary mass is obtained through the depth of shower maximum, X_{\max} , extracted from hybrid events measured over a 14-year period at the Pierre Auger Observatory. The sky is split into distinct on- and off-plane regions using the galactic latitude of each arriving cosmic ray to form two distributions of X_{\max} , which are compared using an Anderson-Darling 2-samples test. A scan over roughly half of the data is used to select a lower threshold energy of $10^{18.7}$ eV and a galactic latitude splitting at $|b| = 30^\circ$, which are set as a prescription for the remaining data. With these thresholds, the distribution of X_{\max} from the on-plane region is found to have a $9.1 \pm 1.6^{+2.1}_{-2.2}$ g/cm² shallower mean and a $5.9 \pm 2.1^{+3.5}_{-2.5}$ g/cm² narrower width than that of the off-plane region and is observed in all telescope sites independently. These differences indicate that the mean mass of primary particles arriving from the on-plane region is greater than that of those from the off-plane region. Monte Carlo studies yield a 5.9×10^{-6} random chance probability for the result in the independent data, lowering to a 6.0×10^{-7} post-penalization random chance probability when the scanned data is included. Accounting for systematic uncertainties leads to an indication for anisotropy in mass composition above $10^{18.7}$ eV with a 3.3σ significance. Furthermore, the result has been newly tested using additional FD data recovered from the selection process. This test independently disfavors the on- and off-plane regions being uniform in composition at the 2.2σ level, which is in good agreement with the expected sensitivity of the dataset used for this test.

1 Introduction

The Pierre Auger Observatory is currently the largest observatory dedicated to studying cosmic rays with energies in the EeV range, so-called ultra-high-energy cosmic rays, UHECR [1]. To do this, it uses both an array of particle detectors on the Earth's surface, the Surface Detector, SD [2], and an array of fluorescence telescopes monitoring the atmosphere above the SD, the Fluorescence Detector, FD [3]. The highest quality data set of the Observatory is made up of UHECR events which have simultaneously been measured by both the FD and SD, so-called *hybrid events*. In a hybrid event reconstruction, the geometry of the shower axis is highly constrained by combining the triggered pixel geometry/timing from the FDs and the high confidence core location/timing provided by the SDs. This results in an angular resolution for the pointing direction of the shower axis of better than 0.5° [4], and a resolution on the location of the shower core of 50 m [5].

The low uncertainty geometric reconstruction provided by the hybrid method allows the evolution of the intensity of UV fluorescence light measured by the FD to be inverted to model the *shower profile*, which is the number

of charged particles in the air shower as a function of the amount of matter it has traversed, the slant depth, X . From the shower profile, the slant depth at which the maximum development of the shower occurs, X_{\max} , can be extracted. X_{\max} is closely related to the mass of the primary cosmic ray which induced the air-shower, but is subject to large fluctuations meaning that it can not be used on a shower-by-shower basis to determine primary mass. However, if collected with sufficient statistics, the first and second moments of distributions of X_{\max} , $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ respectively, can be used to make high certainty estimations of the mean mass of the UHECR events used to form that particular X_{\max} distribution [6].

Up until recently [7], X_{\max} derived from hybrid measurements has been used to study the average composition of the cosmic-ray sky as whole, rather than being used to compare the mean compositions of different subsets of the sky. This choice was likely driven by the relatively sparse statistics available in hybrid studies due to the upper-limit of exposure available to them being set by the relatively low 14 % up-time of FDs. The possibility of splitting the data set into subsamples was therefore limited by the need to maintain sufficient statistics to say something useful about primary composition. However, the Pierre Auger

*e-mail: emayotte@mines.edu

†e-mail: spokespersons@auger.org

Observatory has now collected more than 14-years of FD data, and tens of thousands of high-quality hybrid measurements. With this quantity of data, the sky can be split into different regions and the mean mass of UHECRs arriving from them can be studied. This new reality then prompts two questions:

1. Are the systematic uncertainties which trend with event arrival direction for hybrid reconstruction low enough to allow the mean mass arriving from different regions of the sky to be meaningfully compared?
2. Is there a reason to expect that different regions of the sky may display differing compositions due to astrophysical causes?

Question 1) will be explored in section 4. For question 2) it is clear that the opportunity exists. The flux above the ankle at $\sim 5 \text{ EeV}$ [8] is mixed in composition and has long been thought (now confirmed) to be extragalactic in origin [9]. Furthermore, it definitively displays anisotropy above 8 EeV [10]. Additionally, as was nicely put by Alan Watson in 1990:

“... the Larmor radius of a proton of 10^{18} eV in a $3 \mu\text{G}$ field is about 400 pc , comparable to the thickness of the galactic disk. It follows, therefore, that, if the bulk of cosmic rays are protons, anisotropies associated with the magnetic field structure of the galactic disk might appear as the energy increases.” - [11]

Indeed, there were hints of such a spectral feature starting somewhere around $10^{18.5} \text{ eV}$ for directions within 30° of the galactic plane [11, 12]. Unfortunately, such an excess so far does not appear to be significant in the data of current experiments and therefore has not been given much attention since those initial publications.

Now it is known that the flux above 1 EeV is best described as an evolving mix of light-, intermediate-, and high-mass primaries [13, 14]. Due to the galactic magnetic field, *GMF*, the different mass components present at any given energy will be deflected to different degrees as they travel from their extragalactic sources to Earth. This mass dependent deflection suggests that an anisotropy associated with the structure of the *GMF* would kick in for increasingly heavier components as energies climb. It is therefore distinctly possible that an anisotropy associated with the galactic plane could arise in the higher mass components of the flux at some energy in the *EeV* range.

What follows below are specific tests for such a mass-dependent anisotropy associated with the galactic plane using hybrid data of the Pierre Auger Observatory collected between 2004 and the end of 2018. To avoid repeating the contents of the ICRC 2021 proceedings on this result [7] in its entirety, the contents of this proceeding will aim to include components of the analysis that could not be fit in the eight pages allotted in that publication. Therefore, while this proceeding will cover the details of the analysis, increased space will be given to the cross checks and studies of the systematics uncertainties of the analysis. Additionally, a new test on an independent hybrid data set, recovered from the quality cuts, will also be discussed.

2 Reconstruction and selection

The reconstruction methods, selection cuts, and core analysis of X_{max} distributions used here are largely the same as those described in [15]. A rigorous description of these methods can be found in [6]. Other than the fiducial field-of-view cut, *FidFoV*, which is treated below, a detailed description of the methods will not be provided here.

The important differences from [15] are that the minimum energy for inclusion in the data set has been raised to $E > 10^{18.4} \text{ eV}$, and that, as described below in subsection 2.1, the event acceptance, reconstruction bias, and X_{max} resolution are now treated based on the arrival direction of each event. The lower limit of $10^{18.4} \text{ eV}$ has been chosen as above this energy, the composition is well mixed and expected to be primarily of extragalactic origin [16]. The period over which data has been collected has also been slightly expanded to span events observed between 2004-12-01 and 2018-12-31, yielding 7572 high-quality events. A further subdivision of the data is necessary to test for the hypothesized anisotropy. Following the results of the scan shown in Figure 5, the data is split into the events with a galactic latitude, $|b| \leq 30^\circ$, the *on-plane* sample, and, $|b| > 30^\circ$, the *off-plane* sample.

Fiducial field-of-view selection

To ensure a good reconstruction of X_{max} with the FD, X_{max} itself should be directly observed. The telescopes of the FD have a field-of-view, *FoV*, which is vertically constrained. Additionally, some showers in more vertical geometries will not reach X_{max} before impacting the ground. These factors together lead to a geometric and X_{max} dependence for what events end up in the analyzed data set. If unaccounted for, this X_{max} acceptance, will inevitably bias a composition study based on FD X_{max} data. At the Observatory, this X_{max} acceptance is addressed primarily through mitigation of the effect with the fiducial field-of-view, *FidFoV*, cut. The *FidFoV* cut constrains the FD detector volume to only event geometries where the expected range of X_{max} values would be visible in the FD *FoV*. As can be seen in Figure 1, this changes the natural FD X_{max} acceptance (gray) to one which is unbiased from ~ 600 to $\sim 900 \text{ g/cm}^2$, which spans the typical range of observed event X_{max} values. Remaining effects on rare events outside this range are corrected using parameterizations as a function of X_{max} and primary energy.

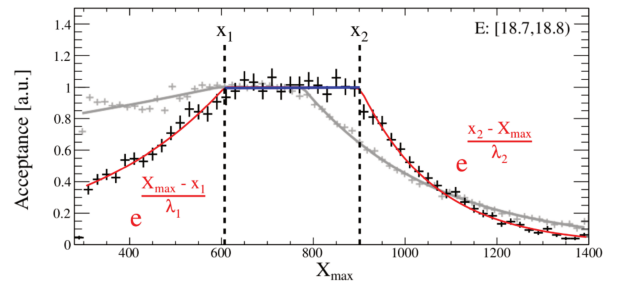


Figure 1: FD X_{max} acceptance before (gray) and after (black) *FidFoV* cuts. A 4-variable parameterization of the post-*FidFoV* acceptance is shown in red. The range of X_{max} values with unbiased sampling is shown in blue.

2.1 Distributions of X_{\max} and arrival direction

After measurement, reconstruction, and selection, the observed X_{\max} distribution does not quite represent the true X_{\max} distribution of all cosmic rays landing within the Observatory. This is due to energy dependent biases on the reconstruction of X_{\max} (B), the resolution on X_{\max} of the hybrid reconstruction method (R), and the residual effects of the X_{\max} acceptance (A). Since the location of an event and its inclination with respect to the observing fluorescence telescope plays a role in the magnitude of A , R , and B , they have an inherent geometric dependence. If unaccounted for, this could bias this study. Distributions of the key geometric relationships between events and the FD can be seen in Figure 2 for on- and off-plane regions. It is clear that there is little difference between the two regions, so A , R , and B are expected to also be similar.

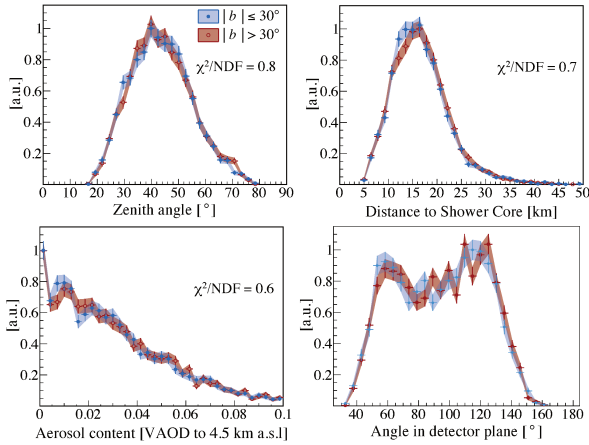


Figure 2: Geometries of the on- and off-plane samples.

To explicitly verify A , R , and B similarity on and off the galactic plane, CONEX [17] is used to generate showers with Sibyll-2.3c [18]. The showers are then isotropically thrown into detector simulations which include the time-dependent state of the FD and SD from 2004 through 2018. These simulations, therefore, mimic the measurement conditions, trigger efficiency, and up-time of the real data, accurately modelling the exposure and geometries of events arriving from all parts of the sky [19]. Two sets of these simulations are produced, one formed from showers generated with a flat sampling of X_{\max} between 300 and 1500 g/cm^2 , the *flat-MC*, and one formed from an equal number of proton, helium, nitrogen, and iron simulations which are then weighted to their abundances observed in data as reported in [20], the *mixed-MC*. These simulated event sets are then subjected to the same reconstruction and selection techniques used on the real data, so that A , R , and B are accurately included in them.

The flat-MC is split into on- and off-plane subsamples which are then used to extract the functional form of A in $0.1 \log_{10}(E/\text{eV})$ energy bins using the method illustrated in Figure 1. The form of A is extracted by leveraging the flatly sampled X_{\max} generation of the Monte Carlo, as, once the plateau of the distribution is normalized to one, the height of each bin represents the acceptance of events with X_{\max} values in that range. The acceptance is then fit

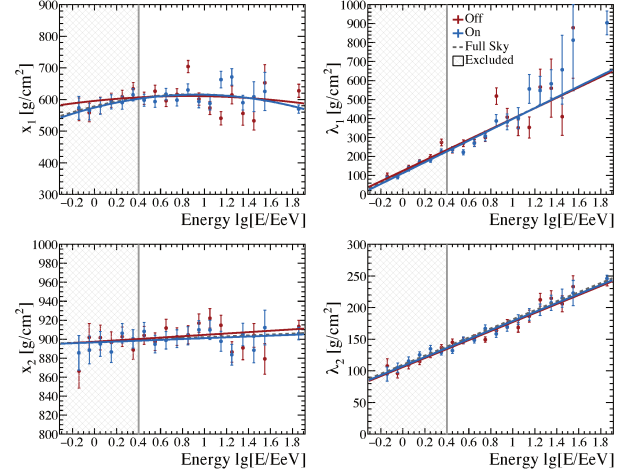


Figure 3: The X_{\max} acceptance parameterizations for the on- and off-plane sky regions from Monte Carlo.

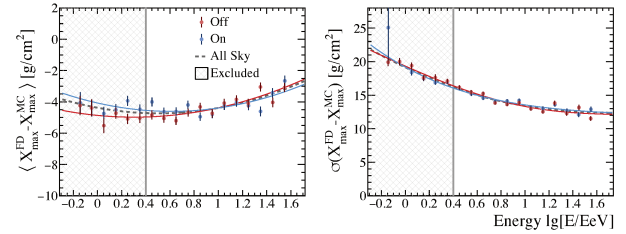


Figure 4: The X_{\max} reconstruction bias and resolution parameterizations for the on- and off-plane sky regions from Monte Carlo. Note: only the detectors and reconstruction X_{\max} resolution is shown. Other effects lowering the resolution are included as specified in [6].

with the 4-component parameterization illustrated in Figure 1. The energy evolution of x_1 , λ_1 , x_2 , and λ_2 is then parameterized separately with 2D polynomials for the on- and off-plane subsamples, resulting in Figure 3. From Figure 3, it is clear that above $10^{18.4}$ eV there is no statistically significant difference in X_{\max} acceptance between the on- and off-plane regions. Even so, the region-specific parameterizations of A are used to correct the 1.4 % of events with partial X_{\max} acceptance using the up-weighting technique outlined in [6]. Uncertainties in these parameterizations result in the systematic uncertainties on the first and second moments specified in Table 1.

The mixed-MC is likewise split into on- and off-plane subsamples which are then used to extract the energy evolution of B and R for each region. B and R are extracted by forming a distribution of the difference between the FD reconstructed value of X_{\max} and the Monte Carlo truth value in $0.1 \log_{10}(E/\text{eV})$ energy bins. From these, the mean reconstruction bias, $B = \langle X_{\max}^{\text{FD}} - X_{\max}^{\text{MC}} \rangle$, and the X_{\max} resolution, $R = \sigma(X_{\max}^{\text{FD}} - X_{\max}^{\text{MC}})$, are extracted in each energy bin. Again, the evolution of each is parameterized with a 2D polynomial. Figure 4 shows that B and R for the two regions are found to agree within errors. These however are also corrected for separately. Fit uncertainties again result in the systematic uncertainties outlined in Table 1. At this point, the X_{\max} distributions and moments from the on- and off-plane regions in each energy bin can be compared without bias from selection and reconstruction.

3 Testing for anisotropy

The specific hypothesis to be tested is whether, above some energy threshold, E_{th} , the mean composition of UHECRs coming from directions near to the galactic plane is significantly higher in mass than those arriving further from it. This is to be tested using X_{max} as a mass sensitive parameter. Typically, X_{max} based composition analyses leverage the first two moments of X_{max} distributions binned in energy, to comment on primary mass. This approach, however, does not lend itself well to quantifying the significance of a result testing the above statement. Instead, a test statistic, TS , which quantifies the degree of dissimilarity between the X_{max} distributions in the two regions in a single value is preferred. For this, the returned value from the Anderson-Darling two-sample homogeneity test [21], *AD-test*, has been selected as it scales with the dissimilarity of the tested distributions. The AD-test has good sensitivity to the full width of a distribution [22], and has more power than the Kolmogorov-Smirnov test while remaining robust against false positives [23].

To use the AD-test and X_{max} for this purpose, two modifications are required. First, a single TS comparing all events in each region above E_{th} is desired. So, all events with $E \geq E_{th}$ in the on- and off-plane samples separately need to be collected into a common on-plane distribution and a common off-plane distribution. To do this, the natural evolution of X_{max} with energy needs to be removed so that spectral features in the flux do not influence the result. Therefore, we define an energy-normalized X_{max} value

$$X'_{max} = X_{max} - \underbrace{(649 + 63.1 Z_{18} + 1.97 Z_{18}^2)}_{\text{EPOS-LHC elongation rate for iron}}, \quad (1)$$

where $Z_{18} = \log_{10}(E_{rec}/\text{EeV})$. The last term in Equation 1 is the natural energy evolution of mean X_{max} for iron primaries as predicted by EPOS-LHC [24]*. Second, the X'_{max} distribution of an on-plane sample populated with primaries which are on average heavier than those in the off-plane sample will display a lower mean and a narrower width than that of the off-plane X'_{max} distribution. Since the null hypothesis is that there is either no composition difference or a heavier off-plane sample, a TS sensitive to the ordering of the X'_{max} distributions is required†. The AD-test is insensitive to ordering, so it is modified to

$$TS = \begin{cases} AD : \langle X'_{max} \rangle^{on} < \langle X'_{max} \rangle^{off} \\ -3 : \text{else} \end{cases}, \quad (2)$$

where AD is the result of the AD-test comparing the on- and off-plane distributions, and -3 is selected as it is well below the minimum of the AD-test.

Scan for energy and galactic latitude thresholds

A scan has been used to select the optimal on/off splitting latitude, b_{split} , and minimum energy, E_{th} , as uncertainties

in GMF models and source distributions make other approaches impractical. In this scan, each trial $[E_{th}, b_{split}]$ pair is used to form on- and off-plane subsets and the TS is extracted. To preserve the statistical strength of the sparse FD data set, a coarse scan of 5° steps in $|b|$ from 20° to 35° and $0.1 \log_{10}(E/\text{eV})$ steps in energy from 18.4 to $19.4 \log_{10}(E/\text{eV})$ is used. The scan is performed on the data set from [6], which includes events through Dec 31st 2012. At the time of writing, this *scan data set* represents 54 % of the analyzed events. The remaining 46 % of events, the *post-scan data set*, is reserved as blind.

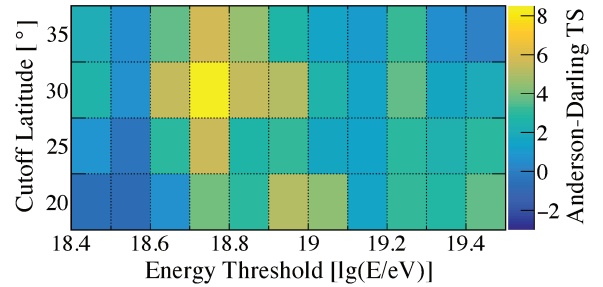


Figure 5: Parameter scan over 54% of the data.

Interestingly, as shown in Figure 5, all tested pairs result in $\langle X'_{max} \rangle^{on} < \langle X'_{max} \rangle^{off}$. An optimal $[E_{th}, b_{split}]$ of $[10^{18.7} \text{ eV}, 30^\circ]$ was found with a $TS = 8.4$. The selected $[E_{th}, b_{split}]$ is applied as a prescription to the post-scan data set, which independently confirms the result with a $TS = 12.6$, for a total $TS = 21.0$ for the full data set.

Statistical significance

The chance probability of the observed TS occurring with in an isotropic sky is tested using Monte Carlo methods on randomized skies derived from the real data. To form each randomized sky, the arrival direction is first decoupled from the energy and X'_{max} values of each event. These are then randomly re-paired to create a new sky which maintains the real X_{max} , energy, and sky exposure distributions, but has a scrambled arrival direction/composition pairing. The above analysis is then used to extract a TS from each sky which is compared to the result in data. Skies which display more extreme on-/off-plane differences than those observed in data are tallied and used to calculate the probability of an isotropic sky generating the observed TS . The results of this procedure are shown in Figure 6.

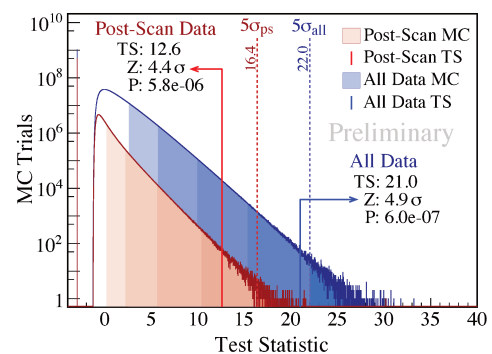


Figure 6: The Monte Carlo determination of the post-scan (red) and all-data (blue) significance with 1 and 10 billion randomized skies, respectively.

*Choice of hadronic interaction model varies result by $\sim 0.02 \text{ g/cm}^2$.

†Modifying the test to also require $\sigma(X'_{max})^{on} < \sigma(X'_{max})^{off}$ would be more restrictive, but conservatively has not been applied.

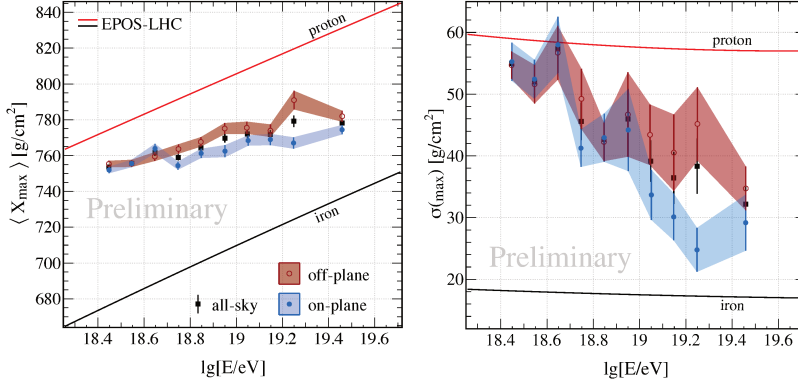


Figure 7: The first (left) and second (right) moments of the X_{\max} distributions from on- and off-plane regions.

For the blind, post-scan data set, the prescribed $[E_{\text{th}}, b_{\text{split}}]$ pair is used to split each randomized sky into on- and off-plane samples and a TS is extracted. In one billion random skies, only 5865 resulted in a more extreme TS than the 12.6 observed in data. This indicates a chance probability of 5.87×10^{-6} which corresponds to 4.4σ .

To calculate the significance of the result when the scan- and post-scan data sets are combined, the entire analysis chain, including the scan, is duplicated. In each random sky, 54 % of the data is used to scan for the $[E_{\text{th}}, b_{\text{split}}]$ pair which results in the most extreme result, fully penalizing for the scan. These values are then used to split all data in the random sky into on- and off-plane subsamples and the TS for the sky is extracted. From 10 billion random skies, only 5964 resulted in a more extreme TS than the 21.0 observed in data. This indicates to a chance probability of 5.96×10^{-7} which corresponds to 4.9σ . The strong penalization of the scanned data is evident as the additional 54 % of the data (with $\Delta\langle X'_{\max} \rangle = 8.5 \text{ g/cm}^2$) only resulted in an 11 % increase of the significance of the observation.

X_{\max} moments and trends

To illustrate the difference in composition on and off the plane, the first two moments of the X_{\max} distribution in each $0.1 \log_{10}(E/\text{eV})$ energy bin has been plotted in Figure 7 for both regions. Above $10^{18.7} \text{ eV}$ there is a clear separation in $\langle X_{\max} \rangle$ for all energy bins. Most energy bins also display a separation in $\sigma(X_{\max})$. Heavier primaries are expected to, on average, have a shallower X_{\max} and lower shower-to-shower fluctuations. Therefore the correlated difference seen here indicates that, for this data sample, primaries from the on-plane region have a higher mean mass than that of the off-plane region above $10^{18.7} \text{ eV}$.

To evaluate the degree to which fluctuation plays a role in the observed result, the growth of the TS over time has been plotted in Figure 8. The time evolution of the signal is consistent with linear growth at a rate of 1.3 TS yr^{-1} . This behavior is in line with expectations for a real difference in mean mass between the subsamples. The shaded region of Figure 8 shows preliminary data from 2019. These reconstructions were not subject to a validated reconstruction chain and may change. Still, when added, a $3.7/4.4\sigma$ (post-scan/all data) statistical significance is expected. The best fit rate of growth of 1.3 TS yr^{-1} remains unchanged.

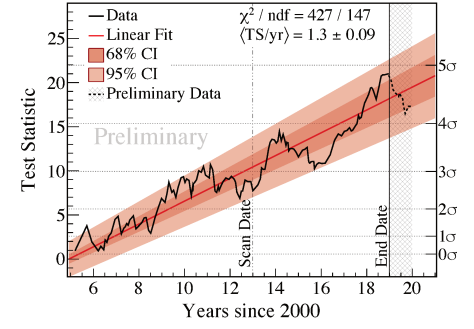


Figure 8: The time evolution of the TS with significance indicated on the right. The shaded region is preliminary data.

4 Systematic uncertainties

Figure 2 shows that there is very little difference between the on- and off-plane regions in the local reference frame of the detector. The data sets from the two regions also consist of events measured with the same instrumentation in the same location, and have been reconstructed with the same methods. Because systematic uncertainties are derived from measurement effects in the local frame, these similarities between the on- and off-plane samples mean that the majority of uncertainty sources outlined in [6] will apply equally to both regions and thereby cancel in a comparison. Furthermore, from the acceptance, resolution, and bias studies in subsection 2.1, the two regions are also free from selection and reconstruction biases.

Table 1: Systematic uncertainties on the difference in the means, $\Delta\langle X_{\max} \rangle$, and widths, $\Delta\sigma(X_{\max})$, of the on- and off-plane X_{\max} distributions.

Source	Uncertainty [g/cm ²] of	
	$\Delta\langle X_{\max} \rangle$	$\Delta\sigma(X_{\max})$
A correction	+1.14 -0.71	+2.37 -1.61
B correction	± 0.36	± 0.01
R correction	0	+1.78 -0.24
Seasonal	+1.00 -1.53	+1.19 -1.23
Instrumentation	± 1.41	± 1.41
Sum in Quadrature	+2.10 -2.23	+3.49 -2.48

To test for potential systematic effects derived from uncertainties in the A , B , and R corrections, possible seasonal effects, and small differences between the instrumentation at different fluorescence telescope sites, FD -sites, several studies were performed. All permutations of the uncertainties in the A , B , and R corrections were evaluated and the maximum changes in $\Delta\langle X'_{\max} \rangle$ and $\Delta\sigma(X'_{\max})$ were recorded. To look for on/off-plane biases associated with instrumentation differences, events seen by two or more FD -sites were used to compare the X_{\max} reconstructions of each site. No significant biases were found, and the instrumentation-derived systematic uncertainties from [6] were adopted. To check for systematics derived from a combination of the seasonal dependence of the FD exposure and normal yearly variation of the atmospheric quality, $\langle X'_{\max} \rangle$ and $\sigma(X'_{\max})$ for the on- and off-plane regions

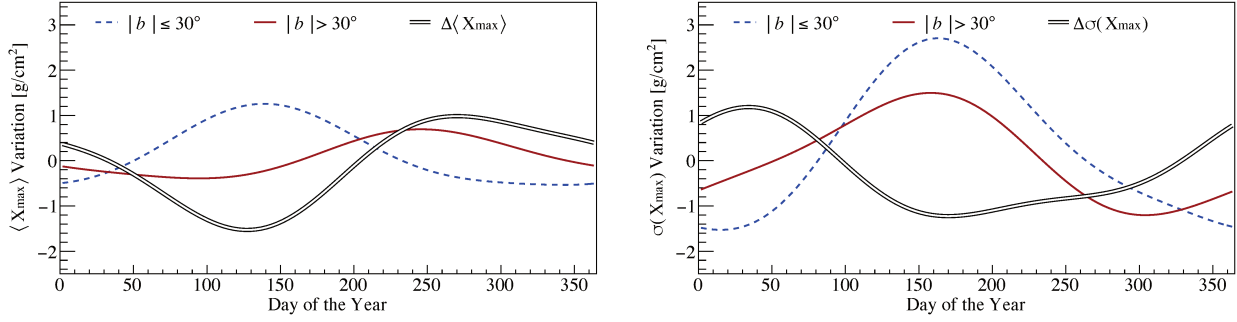


Figure 9: Seasonal fluctuation of the first (left) and second (right) X_{\max} moments for the on-plane (blue) and off-plane (red) samples. The double black line shows the difference between these curves. The maximum and minimum values of these differences are taken as the systematic uncertainty on $\Delta\langle X'_{\max} \rangle$ and $\Delta\sigma(X'_{\max})$.

were tracked over the course of the year, and is shown in Figure 9. The maximum exposure-weighted difference between the two regions at any point of the year has been adopted as the systematic uncertainty due to seasonal effects. These uncertainties and the total systematic uncertainty on $\Delta\langle X'_{\max} \rangle$ and $\Delta\sigma(X'_{\max})$ are listed in Table 1.

Confidence level considering systematic uncertainties

The observed $\Delta\langle X'_{\max} \rangle$ of $9.1 \pm 1.6 \text{ g/cm}^2$ is 4.1 times larger than the 2.2 g/cm^2 systematic uncertainty listed in Table 1. The observed $\Delta\sigma(X'_{\max})$ of $5.9 \pm 2.9 \text{ g/cm}^2$ is 2.4 times larger than its 2.5 g/cm^2 systematic uncertainty. This means it is unlikely that the result could be entirely due to systematic effects. However, the systematic uncertainties in $\Delta\langle X'_{\max} \rangle$ and $\Delta\sigma(X'_{\max})$ may increase the likelihood of an extreme result occurring in data. To quantify the result significance taking possible systematic effects into account, a two step approach is taken. First, the on-/off-plane difference is reduced by adding a value sampled from a Gaussian distribution with $\mu = 2.2 \text{ g/cm}^2$ and $\sigma = 2.5 \text{ g/cm}^2$ to the on-plane sample. Then, the AD-test is applied to the resulting on- and off-plane distributions. Repeating this process 1 million times results in a mean TS of 11.3 ± 0.5 . If these values are converted to significances using the data from Figure 6, this corresponds to at least 3.3σ . Conservatively, to include systematic effects, this lower bound of 3.3σ is adopted as the confidence level of the result.

Cross-check: Results by zenith angle and FD-site

If astrophysical in nature, the difference in composition of UHECR arriving from the on- and off-plane regions should be independently observed by each FD-site and in all zenith angle (θ) ranges. Figure 10 shows that the difference in X_{\max} on and off the plane is indeed present in all zenith angles bins and is also observed by all FD-sites independently. Even more stringently, when the response of each FD-site is split in $\cos^2 \theta$ bins, it appears in 22 out of 28 bins. This independent observation at all sites and zeniths is a strong confirmation of the stereo study described above, showing that detector systematics can not play a large role in the result. Furthermore, because the FD-sites have FoVs differing by 90° on average, each sees

the galactic plane at a different local geometry and time during the year, making it unlikely that some unidentified detector, reconstruction, or atmospheric effect is causing the observed anisotropy.

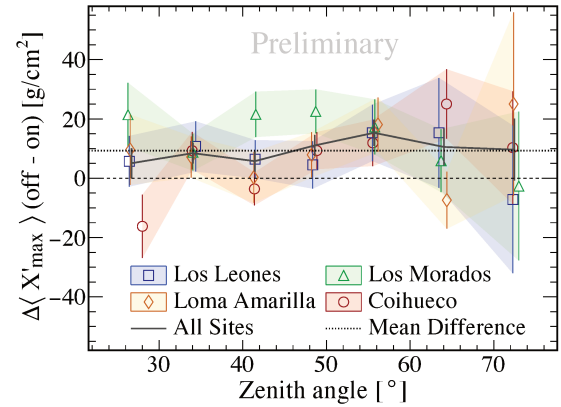


Figure 10: $\Delta\langle X'_{\max} \rangle$ by FD-site and zenith bin.

5 Independent Test of the Result

To cross-check the validity of the previous results, an independent data set is needed. The selection presented in section 2 aims to select only high-quality events. In particular, the *FidFoV* cut is needed to ensure that the distribution of selected events covers the expected range of X_{\max} values with an unbiased acceptance. The events rejected by the *FidFoV* cut are still high-quality and perfectly well reconstructed. In this section, a second data set is built from the events removed in this *FidFoV* cut to form an *out-FidFoV* data set which is 82.9 % as large as the data set already used in this analysis, the *in-FidFoV* data set.

Figure 11 represents the X_{\max} acceptance of this *out-FidFoV* data set in different energy ranges. For all energies, the efficiency maximized for low X_{\max} values and decreases as X_{\max} increases. Indeed, without the *FidFoV* selection, deep showers tend to be under-represented [6]. Thus, the new data set is made out of shallower X_{\max} events. As consequence, the $\langle X_{\max} \rangle$ distribution of the new data set is on average 10 g/cm^2 shallower. Despite this bias, if the difference between on- and off-plane is astrophysical, it should also appear in the *out-FidFoV* data set.

To test this hypothesis, this new data set is divided into on- and off-plane samples using the $E_{\text{th}} = 10^{18.7} \text{ eV}$,

[‡]Treating the other side of the systematic errors in the same way results in a TS of 31.8 ± 1.1 (6.3σ).

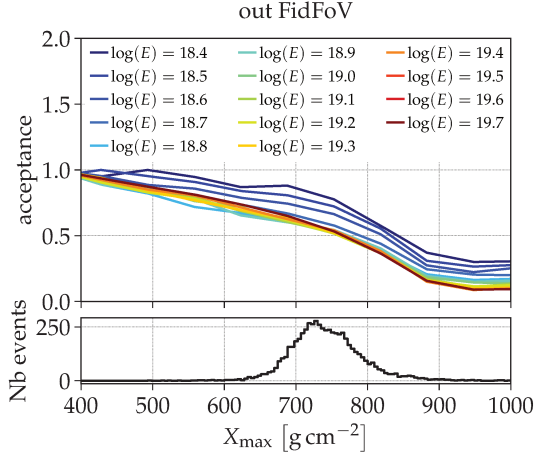


Figure 11: Top panel: acceptance for out-FidFoV data set split in $0.1 \log_{10}(E/\text{eV})$ energy bins. The energy in the legend is the lower energy edge of the bin, excepting 19.7 which corresponds to $\log(E/\text{eV}) \geq 19.7$. Lower panel: the X_{\max} distribution of events for the full energy range.

$b_{\text{split}} = 30^\circ$ splitting determined by the scan. The resulting on/off distributions show a somewhat smaller $\Delta\langle X_{\max} \rangle$ of $\sim 5 \text{ g/cm}^2$, which is only 55 % of what was obtained with the in-FidFoV data set. The AD-test returns a $TS = 1.8$. The post-trial significance of this test is re-evaluated with the method in section 3, using two million randomized MC trials generated from the out-FidFoV data set. The corresponding TS distribution is represented by the orange histogram in Figure 12. The red dashed line depicts the AD-test obtained for the data set. The corresponding significance is $\sim 2.2\sigma$, which gives a probability of 0.03 that this would result from an isotropic sky. The significance seen in this sample is, therefore, much lower than with the in-FidFoV data set.

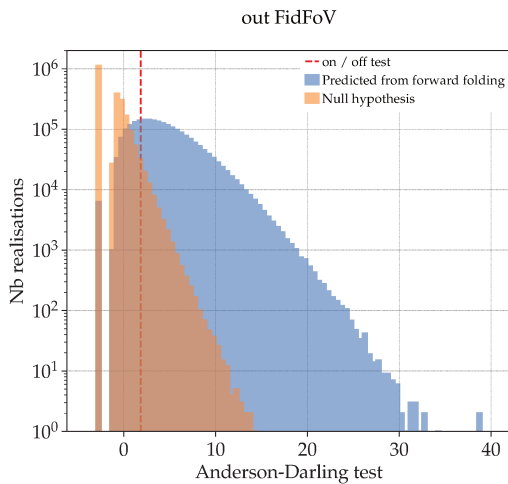


Figure 12: Anderson-Darling test from randomized skies (orange) and from forward folding (blue) for the out-FidFoV data set. The red dashed line indicates the value obtained for data in the on/off test.

The lower significance of the out-FidFoV sample begs the question, why is there such a large difference between the in-FidFoV and out-FidFoV data sets? The bottom panel of Figure 11 shows the distribution of X_{\max} values from events. Clearly the bulk of the distribution lies in a

region where the acceptance is decreasing quickly. It is possible that this could reduce the sensitivity to a difference in composition. To test this hypothesis, two million mock data sets have been generated from the on- and off-plane X_{\max} distributions of the in-FidFoV data set. The non-flat acceptance and lower X_{\max} resolution of the out-FidFoV data set has then been forward folded onto these samples to create out-FidFoV mock data sets which assume the difference in-FidFoV is real. The AD-test is then computed for each mock data set. The blue histogram represents the corresponding TS distribution. It peaks exactly at the value observed with the out-FidFoV data set, showing that the out-FidFoV is indeed less sensitive to the tested anisotropy. Overall, the independent test with the out-FidFoV shows that the on/off separation is present in both data sets, and that the difference seen in the out-FidFoV sample is consistent with the magnitude of the difference seen in the in-FidFoV sample.

6 Mapping the UHECR sky in composition

To aid interpretation of the latitude-dependent difference in composition, a test statistic quantifying the relative difference in X_{\max} between different parts of the sky is mapped in Figure 13, for UHECR primaries with $E \geq 10^{18.7} \text{ eV}$. To produce this map from the in-FidFoV data set, first the requirement $E \geq 10^{18.7} \text{ eV}$ is imposed. Then, because small portions of the sky are analysed, in contrast to the on/off study, each event has its B , R , and A corrected for based on its arriving declination instead of its arriving galactic latitude. This is because local geometry has a time-independent relationship with arrival declination. This means B , R , and A can be corrected equally well for each direction in the sky*.

At this point, a top-hat sampling is used to collect all events with arrival directions within 30° of a point (ℓ, b) into an *in-hat* sample. All other events are placed in an *out-hat* sample. The distributions of X'_{\max} for the in-hat and out-hat samples are then compared using Welch's t-test [25]:

$$TS = \frac{\langle X'_{\max}^{\text{in}} \rangle - \langle X'_{\max}^{\text{out}} \rangle}{\sqrt{\left(\sigma(X'_{\max}^{\text{in}}) / \sqrt{N^{\text{in}}} \right)^2 + \left(\sigma(X'_{\max}^{\text{out}}) / \sqrt{N^{\text{out}}} \right)^2}}, \quad (3)$$

where N^{in} and N^{out} are the event counts for the in- and out-hat samples respectively[§]. This procedure is repeated for top-hats centered on each point in a 5° by 5° galactic latitude and longitude grid. The result is shown in Figure 13, which illustrates the relative composition of UHECRs with $E \geq 10^{18.7} \text{ eV}$ arriving from each point in the sky.

In Figure 13, positive TS values (red) indicate that events within 30° of that point have a lighter mean mass than the rest of the sky. Negative values (blue) indicate that events within 30° of that point have a heavier mean mass than the rest of the sky. An excess of heavy particles

*Using declination-dependent corrections changes the on/off comparison only by $+0.1 \text{ g/cm}^2$ and increases systematic uncertainties.

[§]Because Welch's t-test considers event statistics, the FD arrival direction-dependent exposure is naturally treated through its use.

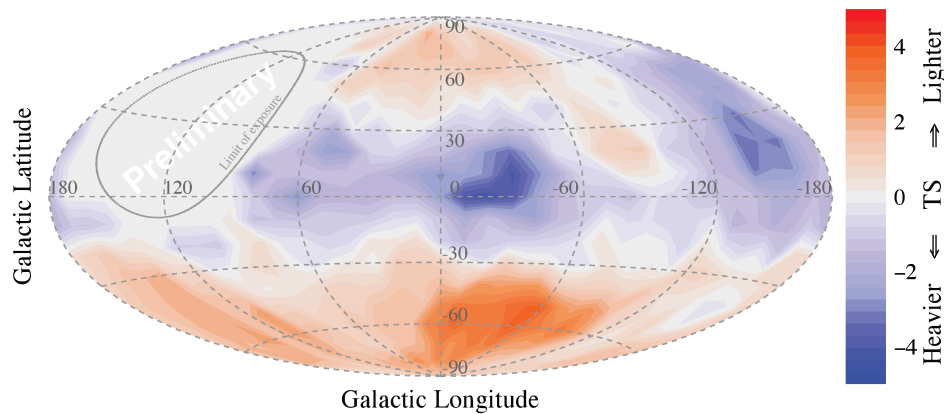


Figure 13: Sky map of cosmic ray composition for $E \geq 10^{18.7}$ eV

within 30° of the galactic plane is visible. This can not be due to detector systematics as they would be declination dependent and appear as radial patterns centered on $\ell = -57^\circ, b = -27^\circ$ due to the geographic location of the Observatory.

7 Conclusions and Outlook

There is an apparent difference in the mean mass of primaries with energies greater than $10^{18.7}$ eV that arrive from within 30° of the galactic plane. This has been observed at least at the 3.3σ level in the standard hybrid data set used for X_{\max} -based composition analyses. It has now been independently confirmed with an additional 2.1σ significance in a second hybrid data set formed from high quality events cut by a selection aimed at reducing the bias caused by the X_{\max} -dependent event acceptance. The combined significance of these two results has not yet been evaluated. Further tests of the on/off-plane difference are being planned using analyses of data from the SD and will be reported elsewhere.

Currently, this result should be considered to primarily provide a new verification of a mixed composition above the ankle as it is clear no such difference could be observed in a flux with a single mass component. Though the analysis provides a possible indication that the galactic magnetic field may have an observable impact on mass-dependent anisotropies, the result found in this analysis does not necessarily support a causal relationship with galactic structures. The differing horizons of different nuclear species at a given energy could also result in composition-dependent anisotropic patterns [26]. It is important, however, to note there is significant tension with models [27]. Alternative scenarios are being explored along these lines of thought.

References

- [1] A. Aab et al. (Pierre Auger), Nucl. Instrum. Meth. A **798**, 172 (2015), 1502.01323
- [2] I. Allekotte et al. (Pierre Auger), Nucl. Instrum. Meth. A **586**, 409 (2008), 0712.2832
- [3] J. Abraham et al. (Pierre Auger), Nucl. Instrum. Meth. A **620**, 227 (2010), 0907.4282
- [4] C. Bonifazi (Pierre Auger), Nucl. Phys. B Proc. Suppl. **190**, 20 (2009), 0901.3138
- [5] M. Mostafa, Nucl. Phys. B Proc. Suppl. **165**, 50 (2007), astro-ph/0608670
- [6] A. Aab et al. (Pierre Auger), PRD **90**, 122005 (2014)
- [7] E. Mayotte et al. (Pierre Auger), PoS ICRC2021, 321 (2021)
- [8] A. Aab et al. (Pierre Auger), Phys. Rev. Lett. **125**, 121106 (2020), 2008.06488
- [9] J. Linsley (1963), ICRC 8 77
- [10] A. Aab et al. (Pierre Auger), Science **357**, 1266 (2017)
- [11] A.A. Watson, Nucl. Phys. B Proc. Suppl. **22**, 116 (1991)
- [12] J. Szabelski, J. Wdowczyk, A.W. Wolfendale, J. Phys. G **12**, 1433 (1986)
- [13] Pierre Auger Collaboration, PRD **90**, 122006 (2014)
- [14] P. Abreu et al. (Pierre Auger), PoS ICRC2021, 311 (2021)
- [15] A. Yushkov (Pierre Auger), PoS ICRC2019, 482 (2020)
- [16] A. Aab et al. (Pierre Auger), JCAP **04**, 038 (2017)
- [17] T. Bergmann et al., Astropart. Phys. **26**, 420 (2007)
- [18] F. Riehn et al., PoS ICRC2015, 558 (2016)
- [19] P. Abreu et al. (Pierre Auger), Astropart. Phys. **34**, 368 (2011)
- [20] J. Bellido (Pierre Auger), PoS ICRC2017, 506 (2018)
- [21] T.W. Anderson, D.A. Darling, Ann. Math. Statist. **23**, 193 (1952)
- [22] F.W. Scholz, M.A. Stephens, JASA **82**, 918 (1987)
- [23] S. Engmann, D. Cousineau, JAQM **6** (2011)
- [24] T. Pierog et al., PRC **92**, 034906 (2015)
- [25] B.L. Welch, Biometrika **29**, 350 (1938)
- [26] C. Ding, N. Globus, G.R. Farrar, Astrophys. J. Lett. **913**, L13 (2021), 2101.04564
- [27] D. Allard, J. Aublin, B. Baret, E. Parizot, Astron. Astrophys. **664**, A120 (2022), 2110.10761