Depth of shower maximum and mass composition of cosmic rays from 50 PeV to 2 EeV measured with the LOFAR radio telescope

A. Corstanje,[1,2,*] S. Buitink,[1,2] H. Falcke,[2,3,4] B. M. Hare,[5] J. R. Hörandel,[2,4,1] T. Huege,[6,1] G. K. Krampah,[1]
P. Mitra,[1] K. Mulrey,[1] A. Nelles [ID],[7,9] H. Pandya,[1] J. P. Rachen [ID],[1] O. Scholten [ID],[8] S. ter Veen,[3]
S. Thoudam,[10] G. Trinh,[11] and T. Winchen [ID][12]

[1]*Astrophysical Institute, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*
[2]*Department of Astrophysics/IMAPP, Radboud University Nijmegen,*
P.O. Box 9010, 6500 GL Nijmegen, Netherlands
[3]*Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, Netherlands*
[4]*Nikhef, Science Park Amsterdam, 1098 XG Amsterdam, Netherlands*
[5]*University of Groningen, Kapteyn Astronomical Institute, Groningen 9747 AD, Netherlands*
[6]*Institut für Astroteilchenphysik, Karlsruhe Institute of Technology (KIT),*
P.O. Box 3640, 76021 Karlsruhe, Germany
[7]*DESY, Platanenallee 6, 15738 Zeuthen, Germany*
[8]*Interuniversity Institute for High-Energy, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*
[9]*ECAP, Friedrich-Alexander-University Erlangen-Nürnberg, 91058 Erlangen, Germany*
[10]*Department of Physics, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates*
[11]*Department of Physics, School of Education, Can Tho University Campus II,*
3/2 Street, Ninh Kieu District, Can Tho City, Vietnam
[12]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany*

We present an updated cosmic-ray mass composition analysis in the energy range $10^{16.8}$ to $10^{18.3}$ eV from 334 air showers measured with the LOFAR radio telescope and selected for minimal bias. In this energy range, the origin of cosmic rays is expected to shift from galactic to extragalactic sources. The analysis is based on an improved method to infer the depth of the maximum X_{max} of extensive air showers from radio measurements and air shower simulations. We show results of the average and standard deviation of X_{max} versus primary energy and analyze the X_{max} dataset at the distribution level to estimate the cosmic ray mass composition. Our approach uses an unbinned maximum likelihood analysis, making use of existing parametrizations of the X_{max} distributions per element. The analysis has been repeated for three main models of hadronic interactions. Results are consistent with a significant light-mass fraction, at best fit 23% to 39% protons plus helium, depending on the choice of hadronic interaction model. The fraction of the intermediate-mass nuclei dominates. This confirms earlier results from LOFAR, with systematic uncertainties on X_{max} now lowered to 7 to 9 g/cm^2. We find agreement in mass composition compared to results from Pierre Auger Observatory, within statistical and systematic uncertainties. However, in line with earlier LOFAR results, we find a slightly lower average X_{max}. The values are in tension with those found at Pierre Auger Observatory but agree with results from other cosmic ray observatories based in the northern hemisphere.

I. INTRODUCTION

Cosmic rays arrive at the Earth's atmosphere in an energy range from below 10^9 to above 10^{20} eV. Upon interacting in the atmosphere, they produce a cascade of secondary particles called *extensive air shower*, which is measurable in ground-based detector arrays for energies above about 10^{14} eV. At the high end of the energy spectrum, these particles have the highest energy of the known particles in the Universe. Therefore, the questions about their origin and their mass composition have raised considerable interest, and cosmic-ray air showers are measured in observatories around the world. The largest is the Pierre Auger Observatory in Argentina [1,2], spanning an area of 3000 km^2.

In this analysis, we study cosmic rays with a primary energy between $10^{16.8}$ and $10^{18.3}$ eV, the energy range where a transition is expected from particles originating from within the Galaxy, to an extragalactic origin. Heavy

---
[*]Corresponding author.
A.Corstanje@astro.ru.nl

nuclei from the Galaxy are expected to reach higher energies than protons, as they are more easily magnetically contained due to their higher charge (i.e., the Hillas criterion [3]). Therefore, composition measurements in this energy region are interesting for comparison with models of cosmic-ray sources and propagation. For instance, in [4], it is argued that a secondary Galactic component may (still) dominate around 10^{17} eV. In one scenario where supernovas of Wolf-Rayet stars are the main sources, one expects a rather low proton fraction together with a higher helium and C/N/O fraction, before proton-dominated extragalactic cosmic rays take over around 10^{18} eV.

Along the track of an air shower, the number of secondary particles reaches a maximum, at a depth expressed in g/cm^2 of traversed matter, referred to as X_{max}. This maximum is reached for almost all showers in our energy range, typically at altitudes of 2 to 7 km. At a given primary energy, X_{max} depends on the mass of the primary particle. It is different for protons compared to heavy nuclei, both on average and in distribution. The shift in average X_{max} with respect to protons is approximately proportional to $\ln A$, for particles with an atomic mass number A, where protons have the deepest shower maximum on average. Thus, measuring X_{max} for a collection of air showers gives information about their composition and is the basis for the present analysis.

There are three main techniques for measuring X_{max}: (i) measuring fluorescence light along the trail of the air shower, (ii) measuring Cherenkov light, and (iii) measuring the radio signal using antennas on the ground [5,6]. Measuring secondary particles on the ground, especially the electron/muon ratio, yields composition information without (explicitly) measuring X_{max}. The radio detection technique has shown substantial development in recent years, leading to a method to determine X_{max} with a resolution about 20 g/cm^2 [7]. The method has been demonstrated using the LOFAR radio telescope, showing that the cosmic rays around 10^{17} eV have a considerable light-mass component [8]. Here, we present a method that has been improved on several points, thus lowering the systematic uncertainties, and an extended dataset.

The method relies on air shower simulations tracking individual particles, and summing up their contributions to the radio signal measured on the ground. For this, the CORSIKA [9] simulation program has been used, with its plugin CoREAS [10] for computing the radio signal. For an ensemble of simulated air showers, their lateral intensity distribution or "radio footprint" is fitted to the measurements, from which X_{max} and the energy of the measured shower are reconstructed.

The X_{max} -distributions for the different elements have substantial overlap. Achieving low systematic uncertainties on X_{max} is therefore a crucial point for composition measurements, besides a good X_{max} resolution per shower. This is done by a fiducial sample selection based on the

CoREAS simulations per shower, and lowering known contributions to systematic uncertainties where possible. For example, in [11], it was shown that accurately representing local atmospheric conditions (refractive index) at the time of the air shower, removes a systematic error of 4 to 11 g/cm^2.

Other improvements to the analysis include a radio-only reconstruction of both X_{max} and energy, the latter using a new calibration based on Galactic emission [12], which halves the systematic energy uncertainty compared to the earlier particle-based treatment. Using a fast precomputation of shower simulations with CONEX [13] streamlines the reconstruction, as showers can be preselected for their X_{max}. The selection criteria to obtain a bias-free X_{max} sample have been improved, and a refined statistical analysis has been done. All these increase the accuracy of the composition analysis, by lowering systematic and/or statistical uncertainties.

The Low Frequency Array (LOFAR) [14] is a radio telescope consisting of many separate antennas. The core region in the north of the Netherlands has a high density of antennas. The antennas are grouped in stations, each of which in the Netherlands contains 96 low-band antennas (LBA), working in the 10 to 90 MHz range, and 48 high-band antennas (HBA) operating at 110–240 MHz. The center of LOFAR is a circular area of 320 m diameter, with six of those stations. In a core region of about 6 km^2, there are 18 more stations. LOFAR uses ring buffers to store up to 5 s of the raw measured signals at each antenna, which are used to measure the radio signals of air showers. For air shower measurements, we use signals from the low-band antennas, filtered to 30 to 80 MHz.

To trigger a buffer readout when an air shower arrives, a particle detector array called LORA (LOFAR Radboud Air shower Array) [15] is located inside the innermost ring of LOFAR. With 20 scintillator detectors monitored in real time, a trigger is sent to LOFAR when a threshold of 13 coincident detections is reached, a level that is optimal for our purposes.

The paper is organized as follows: in Sec. II, we present the method of fitting air shower simulations to measured data to infer X_{max}. Furthermore, we discuss the selection criteria used to obtain a bias-free sample of showers. In Sec. III, the statistical analysis to infer particle composition from the X_{max} values is explained. The results are split into two sections, Sec. IV for the X_{max} distribution from our dataset and Sect. V for the composition results. A summary is given in Sec. VI.

II. METHOD

The discussion of the methods is split into five sections. After an introduction to the use of CORSIKA and CoREAS simulations, we give a brief review of the procedure to infer X_{max} for individual measured air showers. A more detailed explanation is found in [7]; the details that have changed in

this version are given below. The method to estimate the primary energy, and its uncertainties, are discussed in Sec. II C. We show how including the local atmospheric conditions into the simulations leads to improved accuracy. Finally, we explain our method to select showers in order to create an unbiased sample.

A. CORSIKA/CoREAS simulations

For the reconstruction of LOFAR-measured air showers, we use CORSIKA (version 7.7100) to simulate air showers, with its plugin CoREAS, which calculates the radio emission from the particle showers. The simulation uses a "microscopic" approach: it simulates individual particles and their contribution to radio emission, as they are produced along the evolution of the shower. Air showers are simulated from a Monte Carlo approach to particle interactions and applying classical electrodynamics (Maxwell's equations) to obtain the radio signal at the antennas. For the particle part, three main models of hadronic interactions have been considered: QGSJetII-04 [16], EPOS-LHC [17], and Sibyll-2.3d [18,19]. Their differences represent the intrinsic (systematic) uncertainty on hadronic interactions at the high energy levels of cosmic rays in our energy range and beyond.

The radio signals are calculated from first principles, i.e., without free parameters other than those from discretization approaches, which are set to values fine enough to reach convergence in results. This is important for accuracy in reconstructing X_{max}. Calculations are based on the "end point formalism" presented in [20]. In particular, there is no distinction between separate emission mechanisms such as geomagnetic and charge-excess contributions (see, e.g., [21]), as these are naturally included. The radio signals at ground level are a (coherent) superposition of contributions from particles along the shower track, propagated geometrically to the antennas. Therefore, one reconstructs essentially the (geometric) distance to X_{max} from the radio signals.

When fitting simulated air showers to LOFAR measurements, close agreement is found, for pulse energy [8] as well as for detailed measurements such as circular polarization [22]. Comparisons of results from Corsika/CoREAS with another simulation program based on the same principles, ZHAireS [23], show close agreement [6,24,25], and remaining differences are ascribed to details in the simulation of particle interactions. The given detailed and, where possible, parameter-free approach, together with agreement between different extensively developed simulation codes, gives a solid basis for accurate air shower reconstructions.

B. Using CoREAS simulations to estimate X_{max} of measured air showers

A starting point is a set of air showers measured with LOFAR. When an event is triggered by the particle detector array LORA, its radio dataset is passed through our analysis pipeline [26]. Its primary output parameter for this analysis is pulse "energy," defined as the square of the measured voltage in an antenna, integrated over a time window of 55 ns (11 samples) around the pulse maximum. When reconstruction quality criteria are passed, a dataset consists of pulse energy, including its uncertainty, per antenna in at least three LOFAR stations, an accurate measurement of the incoming direction, and an initial estimate of X_{max} and primary energy from fitting a (parametrized) lateral distribution function [27]. The initial estimates are used as a starting point for the simulations. Simulations are iterated if the initial parameters are found to be inaccurate. Therefore, the final estimates do not depend on them.

For each shower measured with LOFAR, we produce an ensemble of CoREAS showers, spanning the natural range of X_{max} for protons and for iron nuclei. We use the QGSJetII-04 hadronic interaction model [16] to produce the particle showers with CORSIKA. As simulation energy, we use an estimate from fitting an analytic description of the radio footprint [27] or an estimate from the particle detectors when the fit failed to converge.

As a precomputation stage, we produce 600 showers with the fast simulation method CONEX, version 4.3700 [13], of which 150 have an iron primary while the others start from a proton. This is suitable to select those random number seeds to span the natural range of X_{max} roughly uniformly with about 15 showers. The same random number seeds are used in the full CORSIKA (version 7.7100) simulations. The number of CONEX showers is high enough to sample into the tails of the X_{max} distributions at a level corresponding to the size of our final dataset ($N = 334$). The aim is twofold, to have simulated showers covering the entire range of X_{max}, which is important for the selection criteria for a bias-free sample (see Sec. II E), and to have a region around the best-fitting X_{max} with extra dense coverage, to improve precision.

Therefore, ten additional showers are simulated in a region of ± 20 g/cm^2 around the first X_{max} estimate, aiming to have a high density of simulations close to the reconstructed X_{max}. The total number of simulated showers is around 30 per measured shower. If the reconstructed X_{max} deviates from the initial fit, extra showers are simulated to match the dense region with the reconstructed X_{max}. An example is shown in the middle panel of Fig. 1; existing showers with $X_{max} > 700$ fall outside the plotted vertical range.

The radio signal of each simulated shower is passed through our antenna model for the LOFAR LBA antennas [26], and through the bandpass filter used in the data analysis, to be able to compare with LOFAR data. The signal energy for each simulated shower is then matched per antenna to the LOFAR measurements. In this fit, the core position and an overall scaling factor are free parameters. This gives a chi-squared value for each shower,
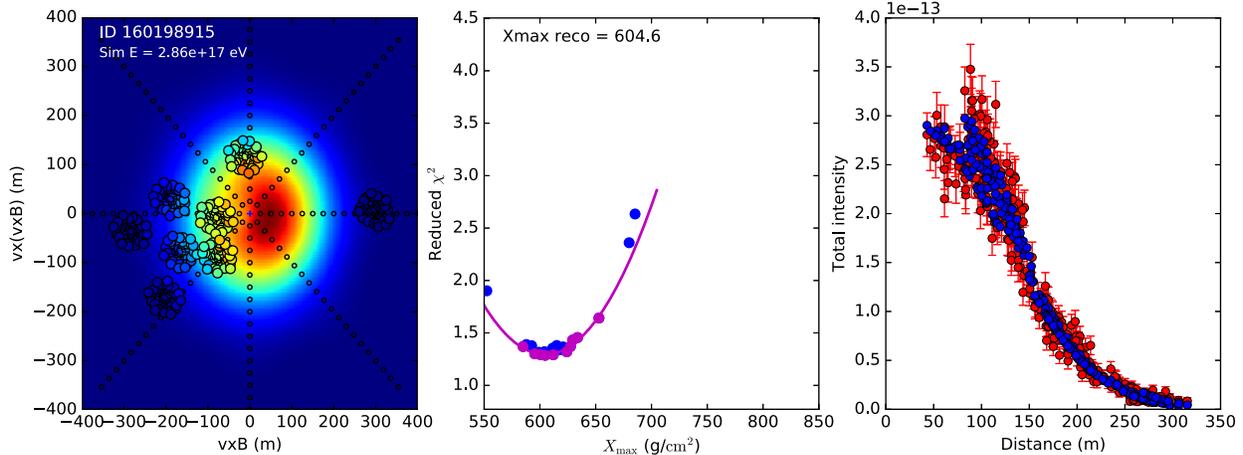
FIG. 1.   An example of fitted CoREAS showers to a LOFAR-measured shower. The left panel shows simulated signal energy along with the measurements, in the shower plane, for the best-fitting shower. The lateral intensity (and pulse energy) distribution is not rotationally symmetric. The middle panel shows the reduced χ^2 as a function of $X_{\max}$, with a parabolic fit through the lower envelope denoted by the magenta line. The right panel shows a 1D lateral distribution function, where red points with uncertainties denote the measurements, and blue points denote the simulated intensities.

$$\chi^2_{\rm radio} = \sum_{\rm antennas} \left(\frac{P_{\rm ant} - f_r^2 P_{\rm sim}(x_{\rm ant} - x_0, y_{\rm ant} - y_0)}{\sigma_{\rm ant}} \right)^2, \qquad (1)$$

where $P_{\rm ant}$ and $\sigma_{\rm ant}$ denote the measured signal energy and its uncertainty, and $P_{\rm sim}$ is the simulated pulse energy. The overall scaling factor is f_r^2, and (x_0, y_0) is the fitted shower core position. In contrast to the method in [7], we perform the fit based on the radio signals only, making the radio reconstruction self-sufficient. In the previous analysis, the fit included both radio and particle detector data. As a consequence, showers for which the reconstruction cannot be done accurately without the particle detector signals are now (automatically) discarded.

The result of the fitting procedure for one of our measured showers is shown in Fig. 1. In the left panel, the best-fitting simulated shower is shown (background color) together with the measurements (colored circles). The colored circles blend in well with the background color, indicating a good fit. This is confirmed by the middle plot, showing a reduced χ^2 of 1.3 for the best fit, and a clear minimum as a function of $X_{\max}$. The right panel shows a one-dimensional representation of the simulated and measured intensities per antenna.

We employ a Monte Carlo procedure, using the simulated ensemble of showers to infer the uncertainties on $X_{\max}$, the energy and the shower core position. For each simulated shower in our ensemble, we create three mock datasets as they would have been measured, i.e., adding the noise level found in each LOFAR antenna. They represent three different realizations of the random noise, at a fixed shower core position. This simulated shower is then reconstructed through the above procedure, using the ensemble of all other simulations. Statistically comparing

the reconstructions with the real $X_{\max}$, core position, and radio scale factor, which are known in simulations, yields their uncertainties. The uncertainties thus found are calculated from the entire simulated ensemble and are applicable to the measured shower as well as to each simulated shower; this is important in the bias-free sample selection procedure explained in Sec. II E. This procedure, relying on the reconstruction method described above, now also uses only the radio signals.

C. Estimate of the primary particle energy

We estimate the energy of the primary particle by comparing the pulse energy of the measured radio signal with the predicted radio signal from CoREAS, which was produced at a given simulation energy obtained from the initial fit. The intensity, and measured pulse energy, of the radio signal scale quadratically with the primary energy [28–30]. Fitting CoREAS radio signal energy to LOFAR data produces an overall scale factor. The square root of this is taken as a correction to the simulation energy, giving an estimate of the primary particle energy. Moreover, following [24], we apply a correction factor of 11% to the simulated pulse intensities. This accounts for a finite step size (electron multiple scattering length) in tracking the particle cascade in the simulation.

This procedure of matching simulated and measured pulse energy relies on an accurate absolute calibration of the radio antennas at LOFAR. The calibration has been improved with respect to the previous analysis [12]. It uses the emission from the Galaxy, which enters the measured time traces of the antenna signals as "noise." The galactic emission model LFMap [31] is used to obtain the contribution of each point on the sky at a given (sidereal) time.

This is integrated over the visible sky, using the antenna response as known from [26]. Apart from this, the contributions of electronic noise in multiple stages of the signal chain have been fitted, by comparing the measured variations with sidereal time to the curve from LFMap.

Uncertainties in the calibration, and in the directional dependence in the antenna model, translate into systematic uncertainties on the energy estimates. The X_{max} estimate is, at least to lowest order, not affected, as all signals arrive from the same direction (to within a degree) for a given air shower. Systematic uncertainties on the calibration amount to 13% in total, where the main contributor (by 11%) is the uncertainty on the sky temperature in our frequency range. The other, minor contributions are uncertainties in the electronic noise levels and from the antenna model.

In [25], a cross-check is described between the energy scales determined from radio and from the particle detectors, respectively. For the particle-based reconstruction, one compares the particle footprint from the best-fitting Corsika shower to the signals at the LORA detectors. The conversion from particles reaching the ground to LORA signals is simulated using the GEANT4 simulation package [32], a procedure described further in Sec. II E 1. Agreement within 10% was found between the resulting energy scales from radio and from particles.

Additional contributions to the systematic uncertainty on primary energy have been tested and were found to be small. These arise from the choice of simulation code and from the choice of hadronic interaction model within the simulation code. Switching from CoREAS to ZHAireS [23] yields a difference below 3%. Thus, importantly, two independent simulation codes aimed at detailed simulations produce very close radio energy levels. A cross analysis of the QGSJetII-04 interaction model versus Sibyll-2.3c gives another contribution of 3%.

The resulting systematic uncertainty is 14%, from adding the contributions in quadrature. This is a considerable improvement from the 27% in the previous analysis based on the particle detectors.

The statistical uncertainty on the energy estimate follows from our Monte Carlo uncertainty analysis per shower (see Sec. II B). The average values for the uncertainties on energy and log-energy, i.e., σ_E and $\sigma_{\log E}$, correspond to 9%. Again, this is a notable improvement over the 32% uncertainty in [8], arising from the large number of radio antennas compared to the 20 particle detectors used earlier.

D. Including local atmospheric parameters

To improve the accuracy of the simulations and the X_{max} reconstructions, we have codeveloped an updated version of CORSIKA and CoREAS, which allows us to include local atmospheric altitude profiles of density and refractive index into the simulation runs (v7.7100 includes the update).

The atmospheric parameters at the time of each air shower are taken from the Global Data Assimilation System (GDAS) [33], which gives pressure, temperature, and humidity in 24 layers in the atmosphere. These are data used, e.g., in weather models.

The important quantities for us are the altitude profiles of density and refractive index. The density profile determines the amount of matter traversed by the particles along the shower evolution, and therefore, the shower geometry depends on this. Due to natural variations in air pressure and temperature, the geometric distance to X_{max} may be under- or overestimated, leading to a systematic error per shower on the order of 15 to 20 g/cm^2. In the earlier analysis of [7], the GDAS density profile was used to correct to first order the X_{max} estimate from simulations using the US Standard Atmosphere.

The refractive index n and its variations are important for the radio emission processes. The refractive index is a function of both the density and the humidity. Natural variations in n make the Cherenkov angle wider or narrower, thus affecting the intensity footprint on the ground [11]. Typical variations of $(n-1)$ are on the order of 4%, and introduce a systematic error on the inferred X_{max}. From simulations, this error was found to be about 4 to 11 g/cm^2, depending on the zenith angle.

Residual uncertainties in $(n-1)$ as taken from GDAS temperature, pressure, and humidity are about 0.5%. From this, uncertainties on X_{max} are on the order of 1 to 2 g/cm^2 and will vary between positive and negative from one shower to another, adding to the statistical uncertainty per shower. We have thus removed a systematic uncertainty that is important for precision measurements.

In CORSIKA, five layers are used to parametrize the atmospheric density profile as a function of altitude. In each layer (except for the top layer), the density is set to fall off exponentially with altitude, with a scale height as a free parameter. We have used least-squares curve fitting to determine the optimal parameters to match the five-layer model atmosphere to the GDAS representation [34]. The error on X_{max} induced by the five-layer approximation was found to be about 4 g/cm^2 and adds to the statistical uncertainty per shower. It introduces a systematic uncertainty of 1 to 2 g/cm^2, depending on altitude, hence taken as 2 g/cm^2.

E. Bias-free sample selection

In this section, we show how to apply fiducial cuts, i.e., to reject showers that would introduce a composition bias to the sample. Cuts are made only based on the simulated ensemble of showers, not on the measured data (for instance, through fit quality). For the composition measurement, we aim to obtain a sample which is unbiased in X_{max}. We do not expect, however, to obtain a sample reflecting the natural cosmic-ray energy spectrum, as the

effective exposure area, both on the ground and on the sky, depends strongly on energy.

A bias may arise from the particle detector trigger, which is reached more easily for showers penetrating deeper into the atmosphere (high X_{max}). Another, opposite source of bias arises from the radio detection threshold. We require at least three LOFAR stations to detect significant pulses for a given shower. Showers with low X_{max} have a larger radio footprint and hence, are more likely to trigger three LOFAR stations.

We analyze each measured shower given its energy, reconstructed shower core position, and incoming direction. The central requirement is that this shower would have produced a trigger in both the particle detectors and in the radio data, if it had any other value of X_{max} in the natural range. Moreover, it must meet the core reconstruction quality criterion explained below. As noted in Sec. II B, our simulated ensemble for each measured shower is based on a preselection from 600 random showers simulated with CONEX. This sufficiently represents the natural range of X_{max}, as our (final) dataset is smaller than this.

A dataset comprising all measured showers that meet this requirement is then unbiased in X_{max}, so this requirement is a sufficient condition. Due to the irregular array layout and moderate event count, a per-shower inclusion criterion is more efficient than attempting to construct a fiducial volume in parameter space (which would also be rather irregular).

1. Removing selection bias arising from the particle trigger

For each measured shower, we use the set of all simulated showers, including their particle content, to see if each simulated shower would have triggered LORA.

For this, we use the GEANT4 simulation tool [32], which simulates the particles traversing the detectors and their deposited energy. The simulation of the LORA detectors was also used in the measurement of the cosmic-ray energy spectrum presented in [35]. Only if all showers in the ensemble are able to trigger, the measured shower is included in the sample.

From CORSIKA, we obtain a list of particles reaching the ground, with their respective positions and momenta. In the GEANT4 simulation, this is converted to an energy deposit at the detector locations. We divide the energy deposit by an average value of 6.2 MeV per particle. The value of 6.2 MeV arises from the most probable energy deposit of single, high-energy muons from an all-sky distribution [35]. Although the muons vary in energy, their deposit is nearly constant with energy.

At the time of each measured shower, we note the trigger threshold of each detector, which was derived during operation from the baseline and standard deviation of its signal time trace. This can be expressed in equivalent

muons. When particles hit a detector and produce enough energy deposit, it will trigger. This is subject to Poisson statistics. We evaluate the probability of having $\geq n$ particles giving an energy deposit of 6.2 MeV each, where n is the first integer above the ratio trigger threshold /1 equivalent muon.

In our trigger setup, a number k out of 20 LORA detectors must trigger in coincidence for the radio data of the air shower to be recorded. The threshold k has been variable over the years of measurements, where $k = 13$ was the most common value. Changes have been made mainly when one or more detectors were down. For each measurement, we use the trigger setting at that time. Hence, also from all simulated showers, we require that with a probability of 99%, at least k detectors would trigger (due to statistical fluctuations, the probability cannot reach exactly 100%).

This test has a tendency to remove showers from the sample which have large reconstructed X_{max} values, i.e., at relatively low altitude in the atmosphere, and/or high inclination. In this case, the given measured shower has produced a trigger, but had its X_{max} been lower, the number of particles would have been too small. Similarly, showers with low energy and/or a core position far from the LORA detectors are more likely to be rejected.

2. Removing bias arising from the radio detection threshold

We perform a test against bias from the finite radio detection threshold. The criterion is, similar to the particle detection bias test, that the radio signal for each simulated shower in the ensemble would have been detected above the noise in at least three LOFAR stations.

To this end, we take the core position of the shower that fits best to the LOFAR-measured shower, and position also all other simulated showers here with respect to LOFAR. From the best-fitting shower, we have a fitted scale factor relating simulated to measured pulse energy. Using this scale factor, we obtain the pulse intensities for each simulated shower and for each antenna. The noise intensities from the LOFAR-measured showers are taken as reference, and a threshold criterion is set as an energy signal-to-noise ratio of 6 in each antenna. In the data processing pipeline, we have a (somewhat arbitrary) threshold requiring half of the antennas per station to trigger to have a "good" detection. Although that detection is amplitude-based, an energy signal-to-noise ratio of 6 was found to be slightly conservative and otherwise in good agreement with the amplitude threshold detection.

This test typically rejects showers with a small reconstructed X_{max} value, i.e., relatively high in the atmosphere, and/or zenith angle; a shower with the same parameters would then have a much smaller radio footprint at high X_{max}, which may not be able to trigger three LOFAR stations.
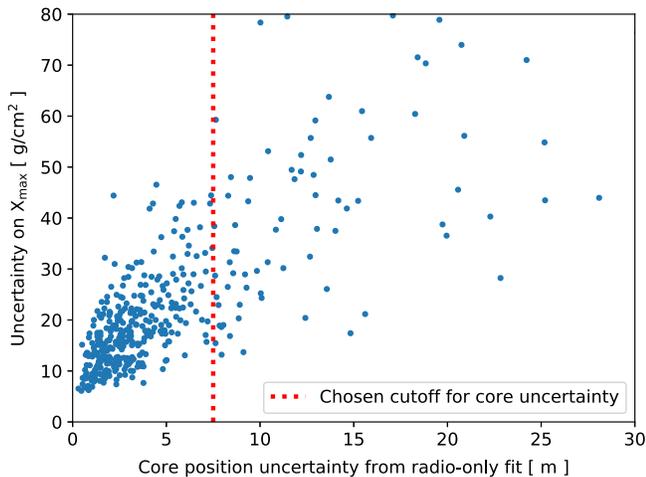
FIG. 2.   The uncertainty on X_{max} versus the uncertainty on the core position, per measured shower, after cuts on energy and fiducial selection criteria. The dotted line indicates the chosen cutoff.

F. Reconstruction quality cuts

The procedure described in Sec. II B to infer the uncertainty on X_{max} is also useful as a test of the reconstruction quality of the radio signal. Apart from the X_{max} uncertainty, it also gives an uncertainty on the fitted shower core position and on the energy. These uncertainties are calculated from the entire simulated ensemble, and hence, they are the same for each simulated shower being tested by the two above procedures.

From the three uncertainties, the precision of the core position reconstruction is arguably the most relevant indicator of overall shower reconstruction quality. When this precision is low, one cannot expect either X_{max} or energy to be reconstructed accurately. Shown in Fig. 2 is the uncertainty on X_{max} versus the core position uncertainty. They are clearly correlated, and a cut on the reconstruction uncertainty at 7.5 m was found to be sufficient to reject the majority of poorly reconstructed showers, while retaining showers with low X_{max} uncertainty.

The appearance of poorly reconstructed showers, despite meeting the other criteria, comes mainly from the position of some showers with respect to the LOFAR array geometry. Most notably, when the core position is outside the array and/or only three stations have been triggered at low signal-to-noise ratio, the reconstruction precision becomes well below average. This criterion catches these cases automatically.

G. Systematic uncertainties

Our method to determine X_{max} is affected by the following systematic uncertainties, which are summarized in Table I. The choice of the hadronic interaction model used in CORSIKA, in this case QGSJetII-04, introduces a systematic uncertainty of 5 g/cm^2 [8] in the X_{max}

TABLE I.   Systematic uncertainties in the X_{max} reconstruction.

	Syst. uncertainty	Added stat. unc.
Choice of hadronic interaction model	5 g/cm^2	
Remaining atmospheric uncertainty	~1 g/cm^2	~2 g/cm^2
Five-layer atmosphere CORSIKA	2 g/cm^2	4 g/cm^2
Possible residual bias	3.3 g/cm^2	
Curve fit for χ^2 optimum	≤1 g/cm^2	
Total, added in quadrature	7 g/cm^2	

measurements, due to minor differences in radio footprints when changing the model, for example, to EPOS-LHC [17]. The choice of the hadronic interaction model also causes another, larger uncertainty in the composition analysis, as the average X_{max} for a given element varies by up to about 15 g/cm^2 between models. This is treated separately by repeating the composition analysis with different models.

Residual systematic uncertainties due to variations in the atmosphere, local weather etc. are about 2 g/cm^2 from the five-layer approximation of CORSIKA. This approximation also produces an additional statistical uncertainty of 4 g/cm^2, which is added in quadrature to the statistical uncertainty on X_{max} per shower. A systematic uncertainty, or bias, in averages of X_{max} may arise from possible residual bias after applying the above selection criteria. We test this in Sec. IV B, obtaining a value of 3.3 g/cm^2 to be added as a systematic uncertainty on X_{max}. Hence, a total systematic uncertainty on X_{max} of 7 g/cm^2 follows. This is comparable to the systematic uncertainty on X_{max} in the measurements of [36], who find a value between 7 and 10 g/cm^2 for primary energies above $10^{17.8}$ eV.

When performing the parabolic fit to the χ^2 values per simulation, as in the middle panel of Fig. 1, a systematic error of up to 5 g/cm^2 may arise if the fit optimum is not contained in the dense region of simulations. This is removed by simulating extra showers around the optimum when needed. A Monte Carlo simulation shows no residual systematic error (≤ 1 g/cm^2) if the dense region is positioned asymmetrically around the optimum but does contain it.

The systematic uncertainty in the energy estimate from the radio antennas was found to be 14%, or 0.057 in lg(E) [12]; by convention, we write lg $E \equiv \log_{10} E$.

III. MASS COMPOSITION ANALYSIS

Having established the set of showers for the mass composition analysis, we perform statistical analysis on the measured data, being $(X_{max}, \sigma_{X_{max}}, \lg E, \sigma_{\lg E})$ for each shower. We make use of the probability density functions of

X_{max} as a function of energy and atomic mass number A, as parametrized by [37] and updated by [38]. The parametrizations follow a generalized Gumbel distribution, which is a function with three parameters, yielding a variable mean, spread, and tail-end asymmetry, respectively. The function has been fitted to a large sample of CONEX showers and has a precision within 2 g/cm^2 for both average and standard deviation of X_{max}, as well as a close fit to the distribution itself; the high-end tail, which drops exponentially, was shown to be well represented. It should be noted that CONEX is a faster but less complete shower simulation method than CORSIKA. Average X_{max} values were found to deviate by 4 to 5 g/cm^2. This is therefore treated as an additional systematic uncertainty on X_{max}, which for the composition analysis then amounts to 8 g/cm^2.

Example curves are shown in Fig. 3 for $E = 10^{17}$ eV for protons, helium, nitrogen, and iron nuclei. The functions overlap substantially, limiting the extent to which the individual elements can be distinguished. This is the statistical challenge in performing a composition analysis on X_{max} data. The mean X_{max} shifts approximately proportionally to $\ln A$. Therefore, for a four-component model of astrophysically relevant elements, a reasonable choice is to take p, He, C/N/O, and Fe. These are roughly equally spaced in $\ln A$; as C, N, and O cannot be readily distinguished, either of them can be chosen as a proxy for all three. We choose nitrogen, as this is in between carbon and oxygen, and the best choice in terms of equal spacing in $\ln A$.

A. Statistical analysis

We use an unbinned maximum likelihood method to determine the best-fitting parameters for the four-component composition model. This has the advantage
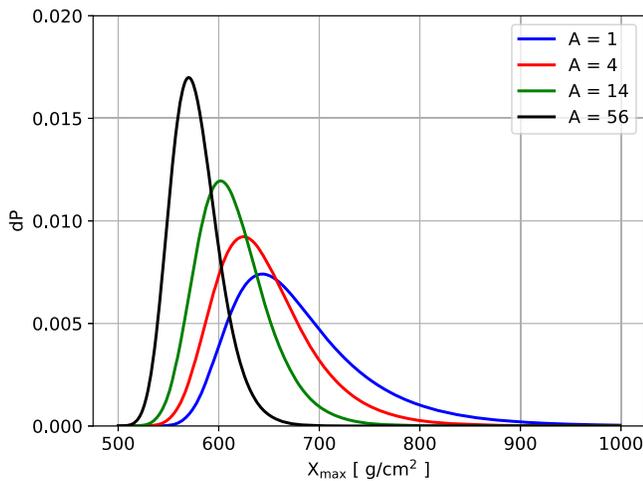
FIG. 3. The probability density functions for the depth of shower maximum X_{max}, for the elements H, He, N, and Fe, at energy 10^{17} eV and hadronic interaction model QGSJetII-04.

of treating each shower separately, instead of relying on X_{max} histograms and/or binning in energy. This is especially suitable when the dataset is relatively small and a narrow binning in shower energies is inappropriate.

Given a measured shower with parameters (X_{max}, $\sigma_{X_{max}}$, $\lg E$, $\sigma_{\lg E}$), its likelihood function for a given element is described by the curves in Fig. 3, convolved with a Gaussian for the uncertainties $\sigma_{X_{max}}$ and $\sigma_{\lg E}$. For a mixed composition, the likelihood function is a weighted average of these; the mix fractions maximizing the likelihood is taken as the best-fitting composition.

With this method, a complementary goodness-of-fit test is needed, for which we use a Kolmogorov-Smirnov test, enhanced with Monte Carlo simulation. This is a simple, well-known method, comparing the cumulative distribution function (cdf) of the best-fit model to the empirical cumulative distribution of the data. The Kolmogorov-Smirnov test statistic is defined as the maximum difference between the model's cdf $F(X)$ and the empirical distribution $E(X)$,

$$K = \sup_X |F(X) - E(X)|. \qquad (2)$$

For our case, where the best-fitting distribution has been estimated from data, standard critical values of the test statistic K do not apply. Instead, to determine the p value corresponding to K, we use parametric bootstrap sampling from $F(X)$, counting how often the K value is larger than the one for the dataset. This tests the null hypothesis that the dataset is a random drawing from $F(X)$.

The best-fit model X_{max} distribution $F(X)$ is taken as the cumulative integral of the linear combination of X_{max} distributions $f(X_{max}, E)$ for the best-fitting composition,

$$f_{sum}(X_{max}) = \frac{1}{N} \sum_{k=1}^{N} \sum_i \alpha_i f_i(X_{max}, E_k)$$
$$* \mathcal{N}_X(X_{max}, \sigma^2_{X_{max},k}) * \mathcal{N}_{\lg E}(\lg E_k, \sigma^2_{\lg E,k}), \quad (3)$$

summing over all showers (index k) and over the elements in the composition model (index i). Here, $*$ denotes convolution, in this case, with Gaussians corresponding to uncertainties in X_{max} and log energy.

For the uncertainty analysis, we use a likelihood ratio test. Denoting the likelihood of the best-fitting composition as $L(\{\hat{\alpha}\})$, we fix one of the element fractions, say the proton fraction, scanning over the range from 0 to 1. We then find the maximum likelihood composition given the fixed proton fraction, $L(\alpha_p, \{\hat{\alpha}_i\})$, again optimizing over the free parameters indexed by i. This gives the test statistic D,

$$D = 2\ln\left(\frac{L(\{\hat{\alpha}\})}{L(\alpha_p, \{\hat{\alpha}_i\})}\right). \qquad (4)$$

This is nonnegative by definition, and in the large-N limit, it follows a chi-squared distribution with 1 degree of freedom, when fixing 1 parameter. The confidence intervals at significance level $1 − p$ then follow directly from the critical values of the chi-squared distribution. For confidence levels of 68%, 95%, and 99%, these are 1.00, 3.84, and 6.64, respectively.

Confidence intervals for two elements simultaneously, such as used in the contour plot Fig. 11 in Sec. V are computed analogously, fixing two parameters instead of one, and noting that the test statistic D then follows a $\chi^2(2)$ distribution.

When splitting the sample into two equal-sized bins, such as done in Sec. V D, one can use another likelihood ratio test to assess the significance of the difference between results in the two bins. For instance, in a model with three independent parameters, such as used in this analysis, splitting into two bins adds three parameters to the (model) description of the data. As a result, the combined (log)likelihood, and test statistic D, will be higher than in a single-bin analysis. The difference follows a $\chi^2(3)$ distribution under the null hypothesis that the data in both bins are a drawing from the same model distributions. This yields a p value for the difference between the bin results.

IV. RESULTS: THE MEASURED $X_{\max}$ DISTRIBUTION

In the following sections, we present the results regarding statistics on $X_{\max}$, such as the estimate of the mean and standard deviation of the $X_{\max}$ distribution. After this, the implications for the cosmic-ray composition in our energy range are given, based on the statistical analysis presented in Sec. III.

The results are based on a dataset of $N = 334$ cosmic rays with energies between $10^{16.8}$ and $10^{18.3}$ eV, which pass all selection criteria for a bias-free sample with accurately reconstructable showers, as explained in Sec. II E. It is a subset of 720 showers measured in at least three LOFAR stations, of which 469 have a core reconstruction precision better than 7.5 m. Another 135 showers did not meet the sample selection criteria; their inclusion would lead to a dataset biased in $X_{\max}$. The uncertainty on the $X_{\max}$ measurement per shower is on average 19 g/cm². The average fit quality of the best-fitting simulation to the measured LOFAR data is $\chi^2/\text{d.o.f.} = 1.19$, indicating a good fit.

A. Mean and standard deviation of $X_{\max}$ as a function of primary energy

We have divided the dataset into energy bins of width 0.25 in $\lg(E/\text{eV})$ and computed the mean and standard deviation in each bin.

The sample averages are shown in Fig. 4. The given uncertainty is the uncertainty on the mean of the $X_{\max}$
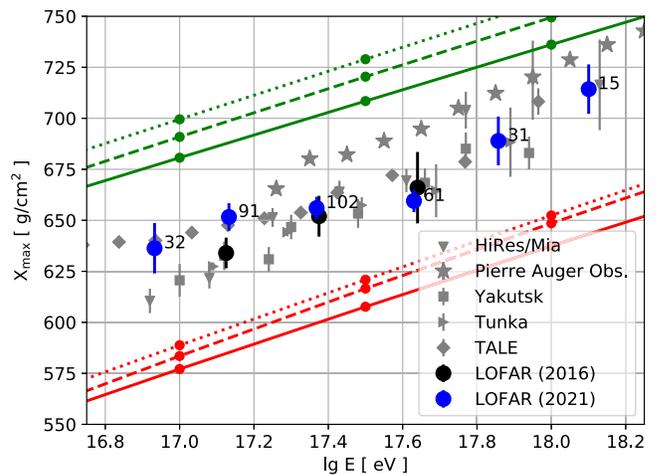
FIG. 4. The average depth of shower maximum $X_{\max}$, as a function of primary particle energy. The annotated numbers indicate the number of showers in each bin, and the error margins indicate the uncertainty on the mean of the $X_{\max}$ distribution. The upper lines indicate the mean values expected for protons, from simulations with QGSJetII-04 (solid), EPOS-LHC (dashed) and Sibyll-2.3d (dotted). The lower lines show the mean predicted values for iron nuclei. For comparison, results from Pierre Auger [39], Yakutsk [40], Tunka [41], HiRes/Mia [42], and TALE [43] are included.

distribution, i.e., $\sigma/\sqrt{N_{\text{bin}}}$, with a sample standard deviation σ. For positioning the points, we have used the average log-energy inside each bin. Two showers above $\lg(E/\text{eV}) = 18.25$ were discarded, as no meaningful average can be taken from them.

For comparison, results are included from the Pierre Auger Observatory [39], HiRes [42], Tunka [41], and Yakutsk [40]. We also include recent results from TALE [43], noting that their method to infer a bias-corrected $\langle X_{\max} \rangle$ is different and assumes the EPOS-LHC hadronic interaction model.

The differences with respect to the earlier LOFAR results [8] can be explained through statistical fluctuations, and from the revised treatment of systematic effects including the atmosphere and the radio-derived energy scale. The lowest-energy data point stands out somewhat, with a difference of 17.6 g/cm², at statistical uncertainties of 7.5 and 6.8 g/cm² for the older and newer result, respectively. Such a difference in one of three overlapping data points is not unreasonable just from statistics; it is also possible that improvements in fiducial selection criteria make some difference here, as differences are expected to appear especially at lower energies where signals are closer to trigger thresholds.

The average $X_{\max}$ agrees reasonably well with the other experiments such as Tunka, Yakutsk, HiRes/Mia, and TALE, especially for $\lg E > 17.2$. However, the results from the Pierre Auger Observatory, which is the largest experiment, are somewhat higher starting at the bin around
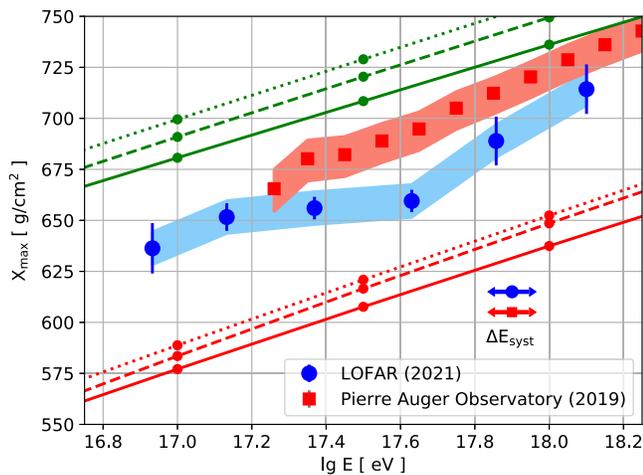
FIG. 5. Average X_{max} versus primary energy, for LOFAR and Pierre Auger Observatory, with colored bands indicating their systematic uncertainty on X_{max}. The uncertainty margins per data point are statistical uncertainties only. The systematic uncertainty on energy is the same for both experiments and is indicated by the arrows to the lower right.

FIG. 6. The standard deviation of X_{max} as a function of primary particle energy. The margins indicate the uncertainty on the standard deviation of the X_{max} distribution. Results from the Pierre Auger Observatory are shown, together with the values from simulations of protons and iron nuclei (high and low lines, respectively).

$lg\,E = 17.325$. Their statistical uncertainty is smaller than the plotted symbols, arising from a high number of showers (1000 to 2600) per individual bin. Systematic uncertainties on X_{max} in this energy range are about 11 g/cm^2 for Auger [44], and about 7 g/cm^2 for LOFAR. Additionally, there is a systematic uncertainty in energy, which for LOFAR as well as Auger [45], is about 0.057 in $lg\,E$.

To better compare the results of LOFAR and Pierre Auger Observatory and their systematic uncertainties, we have plotted these separately in Fig. 5. The band plots are seen to have little to no overlap, although systematic uncertainties in energy could shift either result horizontally in this plot, according to the margins indicated by arrows.

Thus, there is tension between these results. There is a notable difference in methodology to measure X_{max}, being direct fluorescence detection versus radio detection with Corsika/CoREAS simulations. In both cases, considerable attention was given to estimating systematic uncertainties from different contributions. In the range where LOFAR and Auger overlap, our results are in agreement with the other shown experiments on the Northern hemisphere, including the recent results from TALE. Also, our earlier results from 2016 are consistent with the present results, while for the latter, systematic uncertainties on energy and atmospheric effects have been lowered considerably. Hence, the apparent difference is not fully explained at present.

In Fig. 6, we show the standard deviation in each bin, along with its uncertainty. To calculate these, as an estimator $\hat{\sigma}$ of the underlying X_{max} -distribution's standard deviation, we subtract the variance caused by the X_{max} uncertainty per measured shower,

$$\hat{\sigma} = \sqrt{\sigma^2 - \frac{1}{N_j}\sum_{i=1}^{N_j} u_i^2}, \qquad (5)$$

with σ the sample standard deviation, u_i the X_{max} uncertainty on each shower, and N_j is the number of showers in energy bin j. The uncertainty on the standard deviation of the distribution is estimated using a parametric bootstrap, taking the best-fit mass composition from Sec. VA. In this method, one needs to assume a particular model distribution, but it is suitable for estimating uncertainties in small samples. Switching the hadronic interaction model was found to make little difference here.

The results are consistent with those from the Pierre Auger Observatory, except for one bin around $lg\,E = 17.625$. However, a caveat is the relatively low number of showers per bin, and the exponential tail of the X_{max} distributions. The showers at the high end of X_{max}, roughly $X_{max} > 800$ g/cm^2, appear in our dataset only at a low-number statistics level, while their presence shifts the standard deviation considerably upward. In the energy bin from $lg\,E = 17.5$ to 17.75, there happen to be none thus lowering the sample standard deviation.

The results are summarized in Fig. 7, showing the histograms in each energy bin.

B. Tests for residual bias

In our energy range, the average X_{max} of the measured showers is expected to be independent of shower parameters such as the zenith angle. This follows from the fact that the cosmic-ray composition in our energy range is independent of time and incoming direction, as far as is
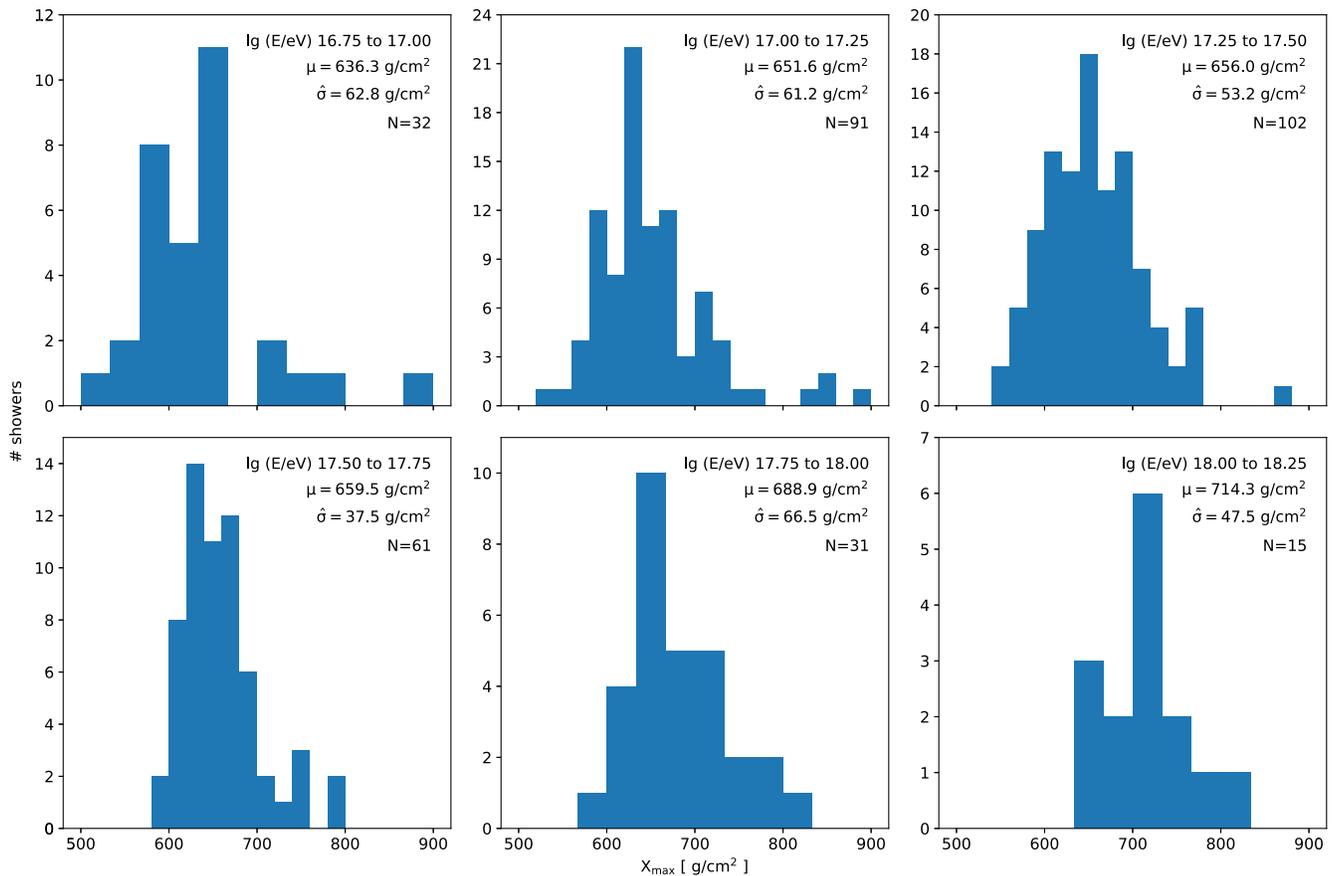
FIG. 7.   Histograms of X_{max} for each energy bin. The average (μ) and standard deviation estimates ($\hat{\sigma}$) per bin are as shown in Figs. 4 and 6, respectively.

known from experiments. This allows us to perform a test of our sample for a residual bias in X_{max} due to variations in these parameters. Our sample of 334 showers has an average X_{max} of 659 ± 3.3 g/cm^2.

As discussed in Sec. II E, a biased sample would readily show a dependence of the average X_{max} on the zenith angle. However, the average X_{max} also depends on the energy; its expected value is to a good approximation linear in lg E over our energy range. From the parametrization using Gumbel distributions, as discussed in Sec. III, and for the QGSJetII-04 hadronic interaction model, we find, for a factor of 10 increase in energy, an average rise in X_{max} (elongation rate) of 55.4, 57.8 and 60.3 g/cm^2, for protons, nitrogen, and iron nuclei, respectively. This is in good agreement with the elongation rate of 58 g/cm^2 predicted, for example, by the Heitler-Matthews model [46].

The average (log) energy of our measured showers tends to rise with the zenith angle. This is no problem, as long as there is no bias in X_{max}. A possible residual bias in X_{max}, corrected for the influence of varying energy, is evaluated by introducing a parameter Y for each shower, as

$$Y = X_{max} + 57(\lg(E/\text{eV}) - 17.4) \text{ g/cm}^2, \quad (6)$$

where 17.4 is approximately the average value of log energy in our sample.

The results are shown in the right panel of Fig. 8, together with a linear fit. The uncertainty margins are once again given by the standard error of the mean. A constant fit of $Y = 660$ as well as a linear fit are shown.

The linear fit has a slope parameter of -0.10 ± 0.30. Hence, the slope is compatible with zero, and no residual bias is evident. The high value near 700 for the rightmost bin appears suggestive, but as it contains only eight showers and has a correspondingly large uncertainty, it is not significant. The constant fit has an uncertainty of 3.3 g/cm^2. A bias at this level cannot be ruled out; hence, this is added as a contribution to the systematic uncertainty on X_{max}.

We also show a complete scatter plot of X_{max} versus the zenith angle, in Fig. 9. This plot shows the effect of the bias tests for the radio and particle detectors and the corresponding fiducial cuts (Secs. II E 1 and II E 2). As expected, the particle bias test flags most events at high inclination and high X_{max}, especially above 45 deg. The radio bias test flags mostly the opposite region, low X_{max}, and low inclination.

Consequently, we see only few showers passing the tests at $\theta < 10°$, and there are only eight in the highest zenith
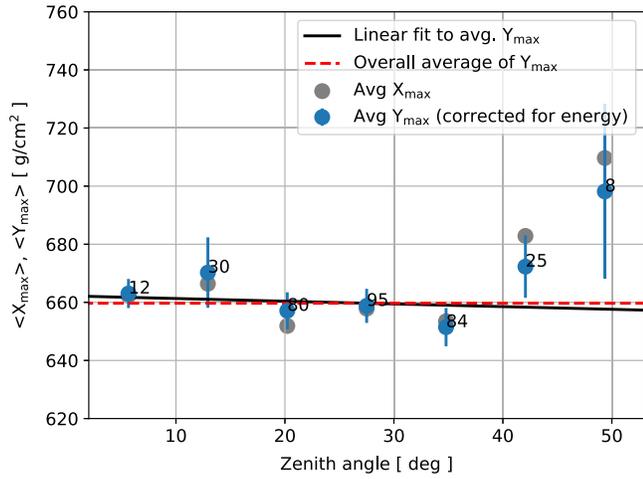
FIG. 8.   Average Y [from Eq. (6)] versus zenith angle, together with a constant and linear fit.
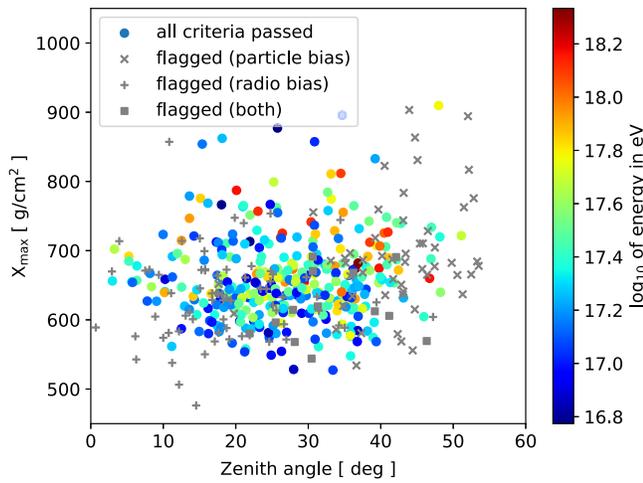
FIG. 9.   Scatter plot of X_{max} versus zenith angle, for all 469 showers with core reconstruction precision better than 7.5 m. Colored circles represent the 334 showers passing all criteria, with the color denoting their energy. The showers flagged by the particle and radio bias tests are also shown.

V. MASS COMPOSITION RESULTS

We have applied the statistical analysis in Sec. III to the set of 334 showers passing all selection criteria. Results are shown in detail for the QGSJetII-04 model and summarized for the other two considered, which are EPOS-LHC [17] and Sibyll-2.3d [18]. We show the results for the full range, followed by a division in two equally sized energy bins. Owing to the strong overlap in X_{max} distributions (see Fig. 3), an $N = 334$ dataset is modest-sized for mass

composition analysis based on a multielement model; hence, further binning is not appropriate. The statistical method presented in Sec. III A is well suited for analysis of a wide energy range without loss of measurement information.

A. Statistics for the QGSJetII-04 hadronic interaction model

The maximum likelihood estimate was found to be 28% protons, 11% helium, 60% nitrogen, and 1% iron. A histogram of X_{max} is shown in Fig. 10 (top), for the full energy range. The coverage of this energy range can be summarized by a mean log-energy of $\lg(E/eV) = 17.39$
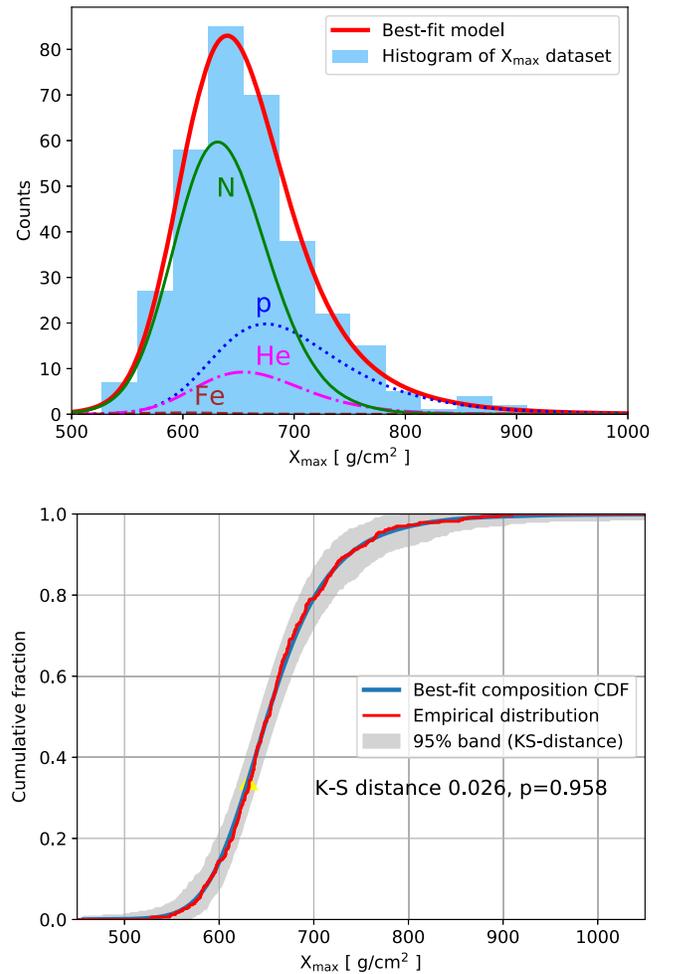
FIG. 10.   Top: A histogram of X_{max}, together with the best-fitting distribution from Eq. (3). Bottom: The cumulative distribution corresponding to the best-fit composition model, together with the empirical distribution from our dataset. The yellow arrow indicates where the distance between CDF and empirical distribution is maximal. The grey band is the envelope of all simulated empirical distributions that have a K-S distance to the CDF below its 95-percentile level. Both graphs indicate a good fit.

angle bin above 45°. The plot makes clear that the fiducial cuts from Sec. II E are necessary, as there would have been a strong zenith angle dependence, and thus a biased X_{max} sample, had it been omitted.

and a standard deviation of 0.32. Hence, the "center of mass" of the dataset lies between 17.07 and 17.71.

The red (solid) curve is the best-fitting distribution, found using the maximum likelihood method and Eq. (3). The distributions for the elements that make up the best-fitting distribution are also shown, scaled by their respective mix fractions.

We have tested the goodness-of-fit of the best-fitting model to our dataset, using the Kolmogorov-Smirnov distance between the cumulative and empirical distributions (Sec. III A) as compared to random drawings from the best-fit distribution. As shown in Fig. 10 (bottom), the model is a good fit to the data ($p = 0.96$). Switching the hadronic interaction model to EPOS-LHC or Sibyll-2.3d produces about equally good fits to the data, at $p = 0.93$ and $p = 0.90$ respectively. Hence, all three models fit the data well, at their respective best-fitting composition. It is of course possible that this would change with a larger dataset and/or smaller energy bins.

We observe that protons and helium are to a significant degree interchangeable in the statistical analysis of our model, given our dataset. This is readily seen in the contour plot in Fig. 11, showing the D statistic for the likelihood ratio test, versus proton and helium fractions. The contours show the allowed regions with confidence levels one-sigma (68%), 95%, and 99%, respectively. Within the one-sigma region, one can exchange helium for protons in a ratio of about 3 to 1. The contour plot underlines, for example, that the dataset only allows a very low proton fraction if the helium fraction is rather large instead.

B. Accounting for systematic uncertainties

The systematic uncertainty in X_{max} amounts to ± 8.1 g/cm^2 (before roundoff), including the contribution
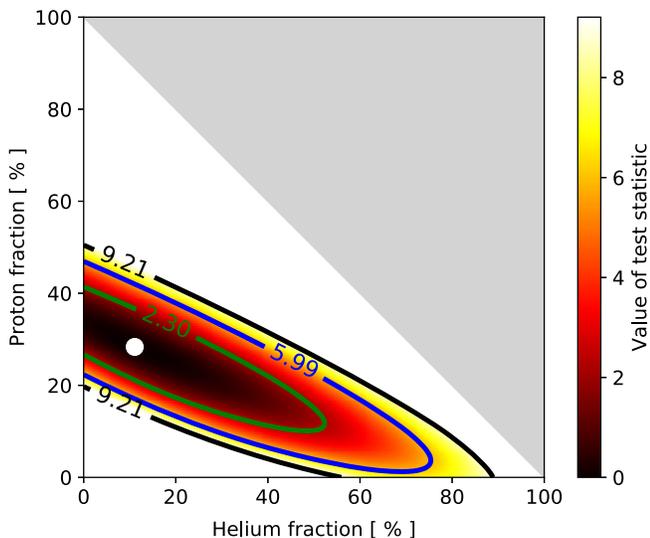
from the CONEX-based parametrizations. The energy uncertainty of 14%, or 0.057 in $\lg E$ has, to first order, the effect of an overall shift of X_{max} in the X_{max} distributions [see Eq. (6)], of 3.1 g/cm^2. By adding both uncertainties in quadrature, we obtain a systematic uncertainty of 9 g/cm^2.

Evaluating the composition results for $X_{max} \pm 9$ g/cm^2 for all showers, we obtain limits for the best fit, as well as the (expanded) confidence intervals that arise for a systematic shift in this range. Noteworthy is for example that when the average X_{max} is shifted downward, the helium fraction is fitted much higher at the expense of the nitrogen fraction. Helium is then favored over nitrogen in the fit, due to the lower expected X_{max} at lower energies, and the longer "tail" of the X_{max} distribution for helium.

C. Results for three hadronic interaction models

The results for the hadronic interaction models QGSJetII-04, EPOS-LHC, and Sibyll-2.3d are plotted in Fig. 12.

For EPOS-LHC, it is seen that the fit favors a more heavy composition, with more iron instead of nitrogen, and with a significant upper bound on the helium fraction. Sibyll-2.3d shows a still somewhat heavier composition, with a lower proton fraction, and a higher nitrogen fraction at best fit. Otherwise, the results are very similar across these three hadronic interaction models, especially considering the uncertainty margins. The intermediate-mass elements in the C/N/O range (possibly stretching to somewhat higher mass numbers as well) are dominant, and there is a significant fraction of light elements, i.e., protons and helium, at best fit ranging from 23% to 39%, depending on the interaction model. Apart from a crossover from C/N/O to iron, the choice of hadronic interaction model has

FIG. 11. Contour plot of the proton and helium fraction, giving the regions consistent with a one-sigma, 95% and 99% confidence level, respectively.
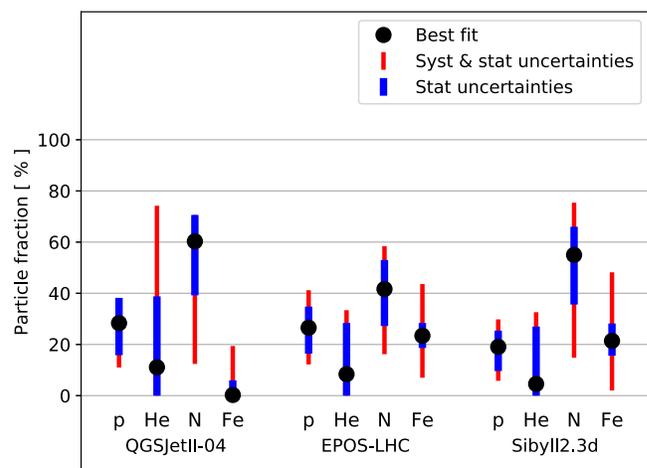
FIG. 12. Composition results from our dataset, assuming each of the three hadronic interaction models shown at the bottom. The best fit is shown along with statistical and systematic uncertainties.

only a limited effect on the best-fit results. From the large intermediate-mass contribution, it is clear that the composition cannot be described as a two-component mixture of protons and iron. This is confirmed by a likelihood ratio test with respect to the four-component model, yielding a p value $p < 10^{-10}$. However, a two-component mixture of protons and nitrogen would work for QGSJetII-04 ($p = 0.9$).

For this dataset, helium and nitrogen are not fully resolved in the composition model. The fitted values for the helium and nitrogen fraction are highly anticorrelated, which follows from the requirement that all mix fractions sum up to 1. Helium and nitrogen are a factor 3.5 apart in atomic mass number, whereas the other consecutive elements are a factor of 4 apart; constant factors here correspond to a constant shift in $\ln A$, and the mean of the X_{max}-distributions varies by an amount proportional to $\ln A$. Moreover, helium and nitrogen have two "neighboring" elements in the composition model, unlike hydrogen and iron. This increases sensitivity to a systematic shift (up or down) in X_{max}.

Importantly, when comparing the current results to the earlier LOFAR results published in [8], the results are found to be consistent, after various improvements to the analysis setup, the systematic uncertainties, and having a larger dataset.

The previous, smaller dataset allowed for a near-100% helium fraction and essentially, no protons; this scenario is now very unlikely. As shown in Fig. 11, a similar but somewhat more constrained interchange between protons and helium is still possible. There are astrophysical scenarios where one would see very few protons but a large helium and carbon fraction. For example, a transition from a Galactic component dominated by Wolf-Rayet supernovae to a particularly strong extragalactic component at these energies could produce a helium fraction near 60% plus a small proton fraction of about 5% at our central energy around 2×10^{17} eV [4]. This is still allowed within 95% confidence limits.

Comparing the results to those from Pierre Auger Observatory [47], starting at $\lg(E/\text{eV}) = 17.25$, agreement is found within statistical and systematic margins. At best fit, their proton fractions are higher, in line with the higher average X_{max}. The difficulty in resolving helium and nitrogen remains, at their higher level of statistics.

Generally, the statistical margins indicate that the analysis would improve with more data. This is no surprise, at a modest number of showers. However, systematic uncertainties are also important at any level of statistics, as they enlarge the margins of statistical plus systematic uncertainties together. This is especially evident in the fitted iron fraction, which is well bounded by statistics, but has substantially expanded margins when systematic uncertainties are included. Also, looking once more at the substantially overlapping X_{max} distributions in Fig. 3, it

is clear that achieving lower systematic offsets in X_{max} is still important, to improve the resolution of the composition analysis as well as the separability of the element fractions.

D. Analysis in two energy bins

We have divided the dataset into two bins with equal number of showers, being those below versus above the median of $\lg(E/\text{eV}) = 17.34$. This is a conservative choice with respect to statistical significance. The results for the three interaction models are shown in Fig. 13. The coverage of the energy bins, summarized as an average and standard deviation, is mainly at $\lg(E/\text{eV}) = 17.14 \pm 0.13$, and 17.65 ± 0.23, respectively.

Notable is the best-fit proton fraction, which is lower in the high-energy bin for all three interaction models. However, from a likelihood ratio test (see Sec. III A, bottom), the difference between the two bins was found not to be statistically significant. That is, a null hypothesis that the observed difference between the two bins would arise from a mass composition constant with energy, as in Fig. 12, is not rejected ($p = 0.58$).

A similar, simpler test is to use Eq. (6) to take out the first-order energy-dependence of X_{max} and split the dataset in two equal-sized bins, for lower and higher energy, respectively. This is a nonparametric test that does not depend on element-based composition models using X_{max} distributions. Also, the elongation rate does not have a strong dependence on the interaction model.

The difference in average Y_{max} between the low and the high-energy bin is not significant ($p = 0.25$). Thus, it is clear that a larger dataset is needed to draw conclusions on possible variations of the element fractions with energy. Would such a trend towards lower proton fractions be confirmed, it would challenge the hypothesis of a transition from a helium and C/N/O-dominated Galactic to a
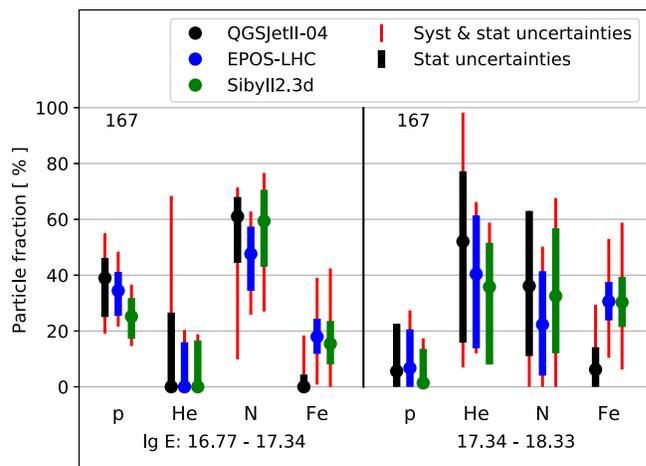
FIG. 13. Mass composition results in two energy bins. The energy ranges per bin are shown; their main coverage is at 17.14 ± 0.13, and 17.65 ± 0.23, respectively.

proton-dominated extragalactic component at energies below 10^{18} eV.

VI. SUMMARY

We have presented an updated cosmic-ray mass composition analysis from LOFAR radio data, built on an improved method for measuring the depth of shower maximum X_{max} of air showers. For the reconstruction of shower parameters, CORSIKA / CoREAS simulations have been used.

We have incorporated several refinements to our analysis, such as including local atmospheric parameters, the Galaxy-based radio calibration and energy measurement, and improved detector characterization in the fiducial sampling procedure. This leads to an energy resolution of 9% and an X_{max} resolution of 19 g/cm^2 on average, and a systematic uncertainty of 14% on energy and 7 g/cm^2 on X_{max}, or 9 g/cm^2 in the mass composition analysis.

To obtain an unbiased dataset suitable for composition studies, three selection criteria were applied to a set of 720 reconstructed showers, based on the ensemble of simulations per measured shower. Requiring a shower core precision better than 7.5 m gives a sufficient general cut on reconstruction quality. We further require that each simulated shower in the CORSIKA / CoREAS ensemble must be able to trigger the LORA particle detector array and also pass the detection and quality criteria of the LOFAR radio analysis pipeline. This procedure leaves a sample of 334 showers for analysis of X_{max} -statistics and (mixed) element composition.

The inferred average and standard deviation of X_{max} have been presented in a (log-)energy range of $16.75 < \lg(E/\mathrm{eV}) < 18.25$. The average X_{max} was found to be in line with results from northern hemisphere-based observatories such as Tunka, Yakutsk, HiRes/Mia, and TALE. However, the values are somewhat lower than those from the Pierre Auger Observatory, where their energy range overlaps. The origin of this tension remains unclear and requires additional research.

Apart from the first two moments, we have also analyzed the X_{max} data at (complete) distribution level, using a four-component model of elements, about equally spaced in $\ln A$. An unbinned maximum likelihood method was found suitable to obtain the best-fitting mass composition in our energy range, together with a separate goodness-of-fit test. This relies on high-precision parametrizations of the X_{max} distributions of the elements, as produced by [37] and updated by [38].

From this analysis, the best-fitting mass composition for our dataset is 28% protons, 11% helium, 60% nitrogen, and 1% iron, assuming the QGSJetII-04 hadronic interaction model. This is averaged over our energy range, with coverage mainly at a log energy of $\lg(E/\mathrm{eV}) = 17.39 \pm 0.32$. The EPOS-LHC and Sibyll-2.3d models tend towards a heavier composition, with an iron fraction just above 20% as a best fit. The light-mass elements together, protons plus helium, form a fraction of 23% to 39% at best fit. Overall, the differences in composition results between these three important hadronic interaction models are minor, apart from a shift between nitrogen and iron. For these differences to become significant, a larger dataset would be needed. A division of the dataset into two equal-sized bins, for lower and higher energy, respectively, yielded no significant difference, both in the model-based composition analysis and in the difference in elongation-rate-corrected X_{max} [Eq. (6)].

The present results are consistent with the earlier LOFAR results [8], thus confirming an appreciable light-mass component in this energy range, at a lower level of systematic uncertainties and from an extended dataset. The element-based mass composition results are in agreement with those from the Pierre Auger Observatory, within systematic and statistical uncertainties. Hence, the tension in the average X_{max} results does not translate to element-based results outside their uncertainty margins. To conclude, we have shown that our X_{max} analysis per air shower achieves an accuracy in line with the current state of the art and demonstrates the value of the radio detection method for measuring air showers.

[1] J. Abraham *et al.*, Trigger and aperture of the surface detector array of the Pierre Auger Observatory, Nucl. Instrum. Methods Phys. Res., Sect. A **613**, 29 (2010).

[2] J. Abraham *et al.*, Properties and performance of the prototype instrument for the Pierre Auger Observatory, Nucl. Instrum. Methods Phys. Res., Sect. A **523**, 50 (2004).

[3] A. Hillas, The origin of ultra-high energy cosmic rays, Annu. Rev. Astron. Astrophys. **22**, 425 (1984).

[4] S. Thoudam, J. P. Rachen, A. van Vliet, A. Achterberg, S. Buitink, H. Falcke, and J. R. Hörandel, Cosmic-ray energy spectrum and composition up to the ankle: The case for a second Galactic component, Astron. Astrophys. **595**, A33 (2016).

[5] K.-H. Kampert and M. Unger, Measurements of the cosmic ray composition with air shower experiments, Astropart. Phys. **35**, 660 (2012).

[6] T. Huege, Radio detection of cosmic ray air showers in the digital era, Phys. Rep. **620**, 1 (2016).

[7] S. Buitink *et al.*, Method for high precision reconstruction of air shower X_{max} using two-dimensional radio intensity profiles, Phys. Rev. D **90**, 082003 (2014).

[8] S. Buitink *et al.*, A large light-mass component of cosmic rays at $10^{17} - 10^{17.5}$ electronvolts from radio observations, Nature (London) **531**, 70 (2016).

[9] D. Heck *et al.*, CORSIKA: A Monte Carlo code to simulate extensive air showers (1998).

[10] T. Huege, M. Ludwig, and C. W. James, Simulating radio emission from air showers with CoREAS, ARENA 2012, AIP Conf. Proc. **1535**, 128 (2013).

[11] A. Corstanje *et al.*, The effect of the atmospheric refractive index on the radio signal of extensive air showers, Astropart. Phys. **89**, 23 (2017).

[12] K. Mulrey *et al.*, Calibration of the LOFAR low-band antennas using the Galaxy and a model of the signal chain, Astropart. Phys. **111**, 1 (2019).

[13] T. Bergmann, R. Engel, D. Heck, N. Kalmykov, S. Ostapchenko, T. Pierog, T. Thouw, and K. Werner, One-dimensional hybrid approach to extensive air shower simulation, Astropart. Phys. **26**, 420 (2007).

[14] M. P. van Haarlem *et al.*, LOFAR: The Low Frequency Array, Astron. Astrophys. **556**, A2 (2013).

[15] S. Thoudam *et al.*, LORA: A scintillator array for LOFAR to measure extensive air showers, Nucl. Instrum. Methods Phys. Res., Sect. A **767**, 339 (2014).

[16] S. Ostapchenko, QGSJET-II: Physics, recent improvements, and results for air showers, EPJ Web Conf. **52**, 02001 (2013).

[17] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider, Phys. Rev. C **92**, 034906 (2015).

[18] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, Hadronic interaction model Sibyll 2.3d and extensive air showers, Phys. Rev. D **102**, 063002 (2020).

[19] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, Cosmic ray interaction event generator SIBYLL 2.1, Phys. Rev. D **80**, 094003 (2009).

[20] C. W. James, H. Falcke, T. Huege, and M. Ludwig, General description of electromagnetic radiation processes based on instantaneous charge acceleration in endpoints, Phys. Rev. E **84**, 056602 (2011).

[21] O. Scholten, K. Werner, and F. Rusydi, A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers, Astropart. Phys. **29**, 94 (2008).

[22] O. Scholten, T. N. G. Trinh *et al.*, Measurement of the circular polarization in radio emission from extensive air showers confirms emission mechanisms, Phys. Rev. D **94**, 103010 (2016).

[23] J. Alvarez-Muñiz, W. R. Carvalho, and E. Zas, Monte Carlo simulations of radio pulses in atmospheric showers using ZHAireS, Astropart. Phys. **35**, 325 (2012).

[24] M. Gottowik, C. Glaser, T. Huege, and J. Rautenberg, Determination of the absolute energy scale of extensive air showers via radio emission: Systematic uncertainty of underlying first-principle calculations, Astropart. Phys. **103**, 87 (2018).

[25] K. Mulrey *et al.*, On the cosmic-ray energy scale of the LOFAR radio telescope, J. Cosmol. Astropart. Phys. 11 (2020) 017.

[26] P. Schellart, A. Nelles *et al.*, Detecting cosmic rays with the LOFAR radio telescope, Astron. Astrophys. **560**, A98 (2013).

[27] A. Nelles *et al.*, The radio emission pattern of air showers as measured with LOFAR: A tool for the reconstruction of the energy and the shower maximum, J. Cosmol. Astropart. Phys. 05 (2015) 018.

[28] Pierre Auger Collaboration, Measurement of the Radiation Energy in the Radio Signal of Extensive Air Showers as a Universal Estimator of Cosmic-Ray Energy, Phys. Rev. Lett. **116**, 241101 (2016).

[29] A. Nelles, S. Buitink, H. Falcke, J. R. Hörandel, T. Huege, and P. Schellart, A parameterization for the radio emission of air showers as predicted by CoREAS simulations and applied to LOFAR measurements, Astropart. Phys. **60**, 13 (2015).

[30] C. Glaser, M. Erdmann, Jörg R. Hörandel, T. Huege, and J. Schulz, Simulation of radiation energy release in air showers, J. Cosmol. Astropart. Phys. 09 (2016) 024.

[31] E. Polisensky, LFmap: A low frequency sky map generating program, http://www.faculty.ece.vt.edu/swe/lwa/memo/lwa0111.pdf.

[32] S. Agostinelli *et al.*, GEANT4: A simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

[33] National Oceanic and Atmospheric Administration, Global Data Assimilation System (GDAS), www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas. https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas

[34] P. Mitra *et al.*, Reconstructing air shower parameters with LOFAR using event specific GDAS atmosphere, Astropart. Phys. **123**, 102470 (2020).

[35] S. Thoudam *et al.*, Measurement of the cosmic-ray energy spectrum above 10^{16} eV with the LOFAR Radboud air shower array, Astropart. Phys. **73**, 34 (2016).

[36] Pierre Auger Collaboration, Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV, Phys. Rev. D **90**, 122005 (2014).

[37] M. De Domenico, M. Settimo, S. Riggi, and E. Bertin, Reinterpreting the development of extensive air showers initiated by nuclei and photons, J. Cosmol. Astropart. Phys. 07 (2013) 050.

[38] S. Petrera (private communication).

[39] A. Yushkov *et al.*, Mass composition of cosmic rays with energies above $10^{17.2}$ eV from the hybrid data of the Pierre Auger Observatory, in *Proceedings of the 36th International Cosmic Ray Conference, Madison, USA* (2019), p. 482.

[40] S. Knurenko *et al.*, Mass composition of cosmic rays in the energy region 10^{16}—10^{18} eV by data the small Cherenkov array at Yakutsk. Comparison with other arrays, in *Proceedings of the 34th International Cosmic Ray Conference, The Hague, The Netherlands* (2015), p. 254.

[41] V. Prosin *et al.*, Primary CR energy spectrum and mass composition by the data of Tunka-133 array, EPJ Web Conf. **99,** 04002 (2015).

[42] P. Sokolsky, Final results from the High solution Fly's Eye (HiRes) Experiment, Nucl. Phys. B, Proc. Suppl. **212–213,** 74 (2011).

[43] Telescope Array Collaboration, The cosmic-ray composition between 2 PeV and 2 EeV observed with the TALE detector in monocular mode, Astrophys. J. **909,** 178 (2021).

[44] J. Bellido *et al.*, Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and composition implications, in *Proceedings of the 35th International Cosmic Ray Conference, Busan, Korea* (2017), p. 506.

[45] B. Dawson *et al.*, The energy scale of the Pierre Auger Observatory, in *Proceedings of the 36th International Cosmic Ray Conference, Madison, USA, Proc. Sci.* (2019), p. 231.

[46] J. Matthews, A Heitler model of extensive air showers, Astropart. Phys. **22,** 387 (2005).

[47] S. Petrera, Recent results from the Pierre Auger Observatory, EPJ Web Conf. **208,** 08001 (2019).