

Vrije Universiteit Brussel



Faculty of Science

Measuring the length of air showers with LOFAR and SKA

Bachelor's thesis Physics and Astronomy

Louis Van Langendonck

Promoted by Professor Stijn Buitink

Academic year: 2020 – 2021

Contents

1	Introduction	2
2	Cosmic rays	3
2.1	Galactic cosmic rays	4
2.2	Extra-galactic cosmic rays	5
2.3	From galactic to extra-galactic cosmic rays	6
3	Air showers	7
3.1	Hadronic Component	7
3.2	Electromagnetic component	8
3.2.1	Radio emission mechanism	8
3.2.2	Observations with LOFAR and SKA	10
3.3	Shower parameters of interest	11
3.3.1	Cosmic ray energy E_0	12
3.3.2	Atmospheric depth of shower maximum X_{max}	12
3.3.3	Length parameter L and asymmetry parameter R	13
3.4	Simulation techniques	13
3.4.1	CoREAS	14
3.4.2	Template Synthesis	16
4	Depth of shower maximum X_{max} reconstruction	18
4.1	Fit procedure	18
4.2	Performance analysis	19
4.3	Using template synthesis	20
5	Implementing the length L of the shower	23
5.1	The case for L as a useful reconstruction parameter	23
5.2	Reconstruction technique	26
5.3	Performance analysis	29
6	Conclusions	31

Abstract

Cosmic rays from the second knee up to the ankle in the all particle energy spectrum are subject to several uncertainties about its source as well as about the hadronic interactions taking place in the air shower it creates. This thesis explores the possibility of expanding on the current parameter reconstruction routine, performed with radio observations made by the Low Frequency Array (LOFAR), by implementing a new parameter L , associated with the length of the shower. Moreover, a new simulation technique called ‘template synthesis’ is used in this implementation. For LOFAR observations, the desired L reconstruction resolution is most likely difficult to obtain. However, this thesis puts forth a method such that with the Square Kilometre Array (SKA), a radio telescope which will be operational soon and will allow for more accurate observations, the desired reconstruction resolution of $\sim 3 \text{ g/cm}^2$ or better should be achievable. By adding the reconstruction of L to the current shower maximum X_{max} reconstruction routine, the combination of these shower parameters can improve differentiation between several proposed source- and interaction models, hence help answer the open questions surrounding cosmic rays in the transition region.

1 Introduction

The Earth is continuously bombarded by charged nuclei, known as cosmic rays, from both galactic and extra-galactic origin. However, various question about their sources remain unanswered, one of which is identifying the nature and particle energy level of the transition from galactic to extra-galactic cosmic rays. A key to solving this puzzle concerns understanding and observing the mass composition of these nuclei. A primary way of doing this is by measuring the radio emission of so-called air showers, cascades of secondary particles as a result of cosmic rays colliding with atmosphere molecules, and then investigating shower parameters sensitive to the composition of the primary particle. The general strategy for achieving this, is by performing a fit of the measured radio emission to various simulated air showers. However, using the current method, it may still be too difficult to distinguish the differences in the particle composition and hence the different possible sources. This Bachelor’s thesis will explore a potential addition to the method by implementing the length of the air shower in the fitting routine. Moreover, a new and potentially more efficient simulation method called ‘template synthesis’ is used. Currently, the radio observations are performed by the Low-Frequency Array (LOFAR), but with the imminent

opening of a new radio telescope called the Square Kilometre Array (SKA), more accurate measurements are expected. With both improved observation quality and the potential expenditure of air shower property reconstruction, progress could be made in answering the open questions about the nature of cosmic rays.

This thesis is structured as follows: First, an overview of cosmic-ray properties and its primary research subjects is presented. Next, in Sec. 3, the focus is shifted to extensive air showers. The section opens with an account of the particle and radiation physics in the air shower, after which a description is given of the air shower parameters of interest, such that observing these might help answering the open questions about cosmic-rays. Eventually, parameter L , associated with the length of air showers, is introduced. The final part of Sec. 3 considers the air shower simulation techniques needed for researching cosmic rays, one of which is the template synthesis method. In the following section, Sec. 4, the reconstruction technique of shower parameter X_{max} is described and analysed. Afterwards, in Sec. 5, a parameter reconstruction routine for the L -parameter is proposed. The possibility of using this in air shower observations, performed at LOFAR and eventually SKA, is investigated. Finally, a conclusion of the results is presented in Sec. 6.

2 Cosmic rays

Charged nuclei from outside the solar system, known as cosmic rays, are continuously observed entering the Earth's atmosphere. The flux of these particles, which range from protons to heavier nuclei, follow a power law $Cte \times E^{-\gamma}$, with E the particle's energy and γ the spectral index [17]. This relation is known as the all particle cosmic-ray energy spectrum and is shown in Figure 1. The spectrum has four characterizing features: a steepening of the flux from $\gamma \approx 2.7$ to 3.1 at $E \approx 3 \times 10^{15}$ eV, called the 'first knee', another such steepening to $\gamma \approx 3.3$ at $E \approx 10^{17}$ eV, called the 'second knee', a flattening of the flux at $E \approx 4 \times 10^{18}$ eV, called the 'ankle' and finally, a suppression of the flux at the far end of the spectrum [17]. From these distinctions, information about the particles origin is retrieved, but several questions remain under active research [3, 12, 16, 18]. Two of the most prominent ones relate to explaining the behaviour of the energy spectrum and the particle mass composition after the first knee and to locate the transition from galactic to extra-galactic cosmic rays [16].

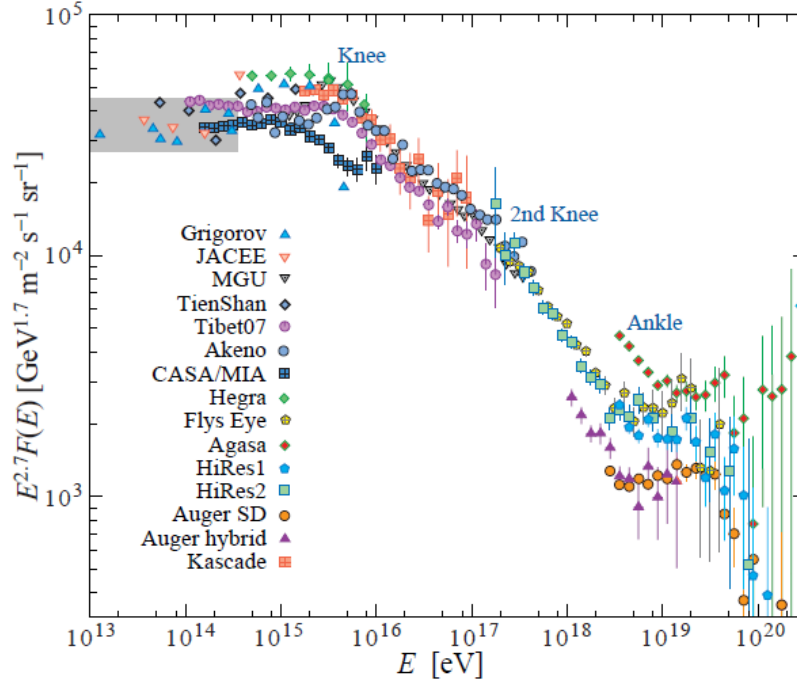


Figure 1: All particle energy spectrum. Figure taken from reference [18].

2.1 Galactic cosmic rays

Up to almost the second knee, cosmic rays are considered to originate from galactic supernova remnants [8, 12, 16, 17]. The shock waves associated with these allow for a diffusive shock acceleration process, which accelerates the particles to high enough energies. As particles consequently escape the remnant, they undergo diffusive propagation through the galaxy, under influence of the galactic magnetic field [8, 16]. Up to almost the second knee, both radio observations at earth and direct observations of particles in supernova remnants are in agreement with this model, but extrapolating it to energies above 2×10^{16} eV, it fails to explain the observed spectrum, with the second knee as its most distinct feature [16]. Moreover, the oddly varying distribution of the composition of the particles in this region, roughly from the second knee up to the ankle, raise some questions [12, 16]. Here, it is important to note that it is expected that if galactic particles gain enough energy, the galactic magnetic field is too weak to prevent the particles from escaping. This is called the rigidity cutoff and is described by $E_{max}(^Z_A X) = Z \times E_{max}(p)$ [12], thus predicting a gradual increase in average particle mass near the end of the contributions of galactic cosmic-rays to the spectrum. However, it is observed that instead of the particles steadily becoming heavier after the second knee,

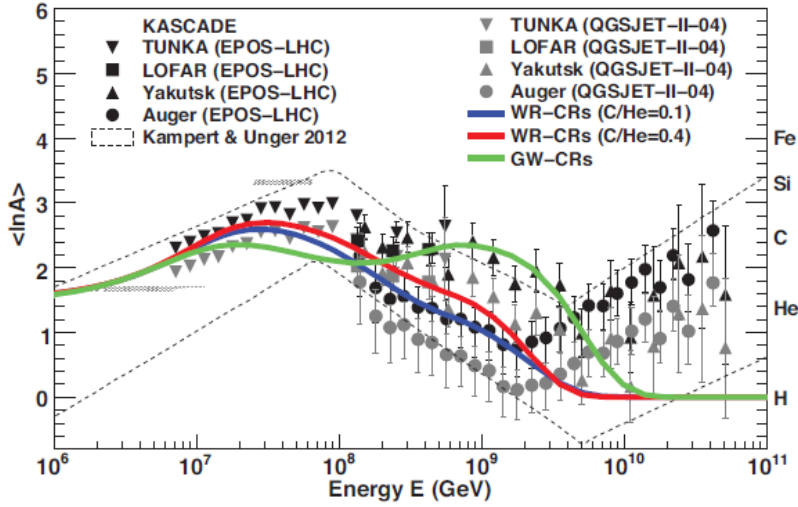


Figure 2: Logarithmic mass distribution in function of the cosmic-ray energy measured by various experiments. The colored lines represent simulated predictions of different models. Figure taken from reference [8]

the mean mass reaches a first maximum at about $E = 6 \times 10^{17}$ eV, decreases up to the ankle and from there on gradually increases again [12, 16, 17]. Observations of the mean logarithmic mass in function of the energy, together with the predictions of three different models, are shown on Figure 2.

Both the observed spectrum and mass distribution suggest a second galactic component in which the two respective knees are a direct result of a different galactic component [8, 16]. As already established, particles directly originating from galactic supernova remnants would thus dominate up to almost the second knee. For explaining the spectrum around and above the second knee, actively researched theories for a second galactic component include incorporating re-acceleration due to the shock waves from older remnants in the interstellar medium and the potential influence of distinct nearby sources like supernova explosions of Wolf-Rayet stars [8, 16]. However, additional research is needed to confirm that these adaptations can simultaneously explain both the observed spectrum and the mass composition evolution [8, 12, 16].

2.2 Extra-galactic cosmic rays

From about the ankle and, in part, even the second knee, cosmic rays are considered to originate from extra-galactic sources, with proposed origins ranging

from clusters of galaxies and AGN jets to gamma-ray bursts [12, 16, 18]. That these particles are extra-galactic, mainly follows from the observed isotropy of the incoming cosmic-rays, because at these high energies, the galactic magnetic field has little effect on the particles path, such that a strong preference for directions coming from the galactic center is expected if the particles are of galactic origin instead [8, 12]. However, several questions remain unanswered [12, 16, 18]. For instance, it is not yet clear at which exact energy these extra-galactic particles begin to contribute. As a result, the mechanism responsible for the second knee and the ankle is up for debate [3, 12, 16, 18].

Finally, the details of the mechanism that causes the suppression of particles at the highest energies, around 10^{20} eV, have yet to be uncovered too [3]. A leading explanation for this suppression considers particle interactions with the cosmic microwave background radiation photons, known as the Greisen, Zatsepin and Kuzmin (GZK) cutoff [18], but inconsistencies with experimental data suggest the possibility of other phenomena playing a role [3].

2.3 From galactic to extra-galactic cosmic rays

As discussed above, it is well established that cosmic-rays up to the about the second knee are from galactic origin and towards the end of the spectrum from extra-galactic origin, but the nature and location of the transition between both cases is uncertain [3, 12, 16]. Several identifications of galactic and extra-galactic sources and the corresponding particle-acceleration and -propagation mechanisms have been brought forth and are often adapted and combined to try and fit the spectrum and observed mass distribution [3, 7, 12, 16, 18].

Each of these theories predict distinct energy-dependent mass distributions, such that it is clear that more particle elemental composition measurements in this transition region are needed to address the open questions regarding the nature and location of the transition [13]. A primary approach for cosmic-ray property reconstruction in this energy region is by studying the radio emission of cosmic-ray air showers, which will be further discussed in the subsequent paragraphs [8, 10, 13].

3 Air showers

Cosmic rays entering the Earth's atmosphere with energies higher than some 10^{14} eV are known to interact with nuclei in the air and consequently produce an 'extensive air shower', cascades of secondary particles and electromagnetic radiation [7, 13]. Because it is impossible to directly observe cosmic rays at these energies, these air shower observations are used to reconstruct information about the primary particle [7, 13]. The general strategy to derive the energy and composition of the primary particle is by reconstructing distinct energy- and mass sensitive shower parameters. In air showers, two components are typically differentiated: a hadronic component, considering the hadrons in the cascade, and an electromagnetic component, considering mostly positrons and electrons. The methods used in this paper for reconstructing the energy- and mass sensitive parameters rely heavily on shower simulations and in order to be able to perform accurate simulations, insight in both shower components is required.

3.1 Hadronic Component

The hadronic component of the shower considers the cascade of different interactions and decays between hadrons as a result of the collision of the primary particle.

Consider as example a proton striking an air molecule. This collision produces a number of both neutral and charged pions [14]. Neutral pions decay to photons almost immediately. This produces electron-positron pairs in its interaction with nuclei, which in turn initiates the electromagnetic subshowers further discussed in Sec. 3.2 [14]. The charged pions on the other hand, travel some distance and interact, producing another generation of pions [14]. This multiplication continues until the individual pion energies drop below a certain critical energy. For charged pions below this energy it is more likely that they will decay rather than interact [14]. Therefore, from this point on, the charged pions are assumed to decay to muons, which in turn are observed at the ground [14].

The exact cascade of different hadronic interactions and decays taking place are too complex to derive analytically and thus require computationally heavy simulation techniques [7]. Moreover, at higher energies, uncertainties in the hadronic interaction models increase as the energies of these cosmic rays exceed those from experiments conducted in the LHC at CERN, which is currently limited to about 10^{17} eV [13, 15]. The estimation of mass-

composition of the incoming particle relies heavily on shower simulations, which will be further discussed in Sec. 3.4. These simulations are in turn dependant on the underlying hadronic interaction models. Hence, the hadronic component of the air shower, together with the mass composition of the cosmic ray, are important sources of systematic uncertainty in the interpretation of cosmic ray data [15]. In order to constrain both uncertainties, very precise observations and reconstructions of the longitudinal evolution of air showers are needed. One possibility, explored in this thesis, is reconstructing the length L of the shower from observation with radio arrays like LOFAR or SKA. Both the systematic uncertainty and the potential of favouring certain interaction models above the other, will be further discussed in Secs. 4-5 about the parameter reconstruction methods. Currently, some of the prominent high-energy interaction models are called the DPMJET model, the EPOS model, the QGSJETII model and the SIBYLL model [15]. A detailed account of these models can be found in reference [15].

3.2 Electromagnetic component

3.2.1 Radio emission mechanism

In contrast to the hadronic component, the mechanisms behind the electromagnetic component are considered to be accurately understood [7]. Although electromagnetic radiation from air shower at ground-level isn't limited to only the radio spectrum, the coherence of the signal at frequencies below about 100 Mhz, which will further discussed in this section, results in a strong and useful radio component. This radio emission from extensive air showers is the result of the superposition of two components: geomagnetic emission and charge excess emission [9].

Geomagnetic emission considers the secondary electrons and positrons, which in turn, are accelerated in the Earth's magnetic field [9]. For these individual particles, the Lorentz force

$$\vec{F}_L = q (\vec{v} \times \vec{B}) \quad (1)$$

suggests circular motion and hence synchrotron radiation, with q the particle charge, $\vec{v}$ the particles velocity vector and $\vec{B}$ the present magnetic field. However, the particles continuously collide with air molecules such that only a net drift, i.e. electrical current, is observed, instead of pure synchrotron radiation [9]. Furthermore, the evolution of electron and positron production has to be taken into account as well. This evolution of charged particles in the shower, called the longitudinal evolution of the electromagnetic air shower,

is presented in Figure 3. As it is known that the number of these particles

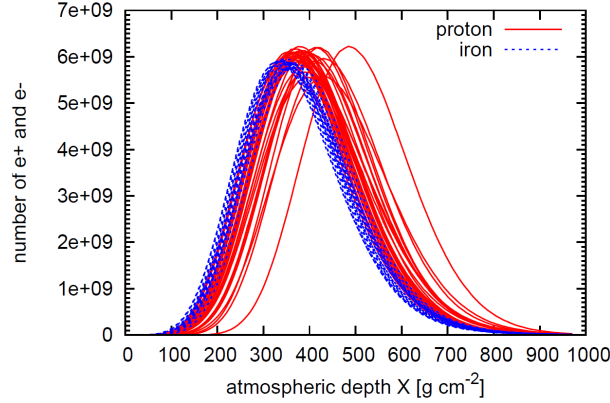


Figure 3: Simulation of longitudinal evolution profiles of electrons and protons of air showers initiated by proton and iron primary particles. Figure taken from reference [9].

increase up to a maximum N_{max} after which it again declines and eventually dies out, the corresponding current undergoes the same evolution. It is the time-variation of these currents that produce the geomagnetic radio emission [9]. A microscopical interpretation states that these time-varying currents have to correspond with individually accelerating charged particles, which in turn result in observable radiation. The particles' propagation direction $\hat{v}$ can be approximated by the shower axis, which is the axis determined by the primary particles' arrival direction [9]. Because the geomagnetic radiation is the result of the Lorentz force (1), it is polarised along the $(\vec{v} \times \vec{B})$ -direction [9].

The second and smaller contribution to the radio emission is the result of the negative charge excess, present in the shower. This negative charge excess of 10 – 20% is the result of ionisation electrons, which are swept along together with other shower particles, while the heavier ionized nuclei stay behind [9]. Once again, this process evolves together with the particle number evolution of the shower, such that it is the time-variation of the charge excess responsible for the radiation. This radiation component is linearly polarised as well, but orientated radially, perpendicular to the shower axis. Figure 4 shows a schematic representation of the two radiation components in air showers.

Because the radiating particles along the shower axis move at relativistic speeds, the relativistic larmor formula dictates that radiation is forward beamed along the shower axis, thus creating a ellipse-shaped illuminated

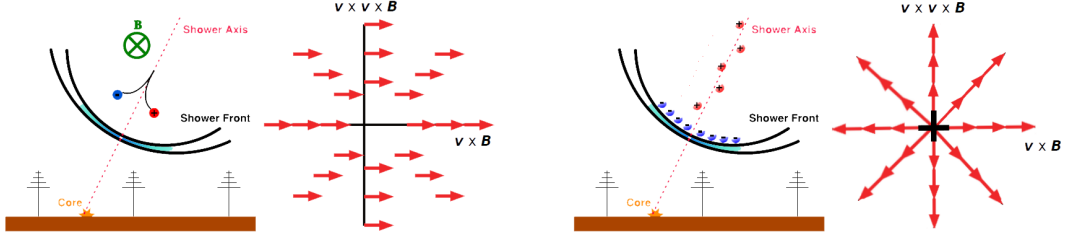


Figure 4: left: Schematic representation of geomagnetic radiation and its polarization in radio air showers. right: Schematic representation of charge excess radiation and its polarization in radio air showers. Figure adapted from reference [9].

area at the ground [9]. This power distribution is typically referred to as the lateral distribution of the radio emission. Moreover, each shower appears to have a distinct solid angle, the Cherenkov angle, for which increased radiation is observed at ground, due to Cherenkov-like effects [4, 9]. This effect is the result of the non-unitary refractive index compressing the pulse in time [9].

Because the two radiation components are polarised in different directions, the total emission pattern is the vector addition of the two [4]. The resulting power pattern on the ground, holding into account all phenomena described above, is a rotationally asymmetric, bean-shaped distribution with increased constructive interference along the $(\vec{v} \times \vec{B})$ -axis [4]. Finally, it is important to note that the observed radiation is coherent, as below a frequency of about ~ 100 MHz, the wavelength exceeds the shower front, which is the disc in which the radiating particles are travelling [4, 9]. At the Cherenkov angle, because of the increased compactness of the particles, the signal can be coherent for up to GHz frequencies [4]. Coherency implies that the power measured at the ground scales quadratically with the number of radiating particles instead of the linear dependency of incoherent signals [9]. This property is useful because it yields a strong measurable signal and allows for estimations of the energy of the primary particle [9].

3.2.2 Observations with LOFAR and SKA

To observe the radio emission described in Sec. 3.2.1, thousands of dipole radio antennas at ground level are needed. The Low Frequency Array (LOFAR) is a radio telescope in the North of the Netherlands used for these measurements, with a low-band frequency range of 30 – 80 MHz [4]. It is

organized in stations of 144 antennas. Six closely packed stations make up the center, around which other stations are placed at increasing distance. The core of 24 stations with a diameter of ~ 2 km is ultimately used for air shower detection [4]. As will be further explained in Sec. 3.3, the atmospheric depth of the shower maximum X_{max} , is one of the primary shower parameters reconstructed using these radio observations. For LOFAR, the average X_{max} resolution using the techniques described in Sec. 4 is about ~ 17 g/cm².

In 2023, however, the low frequency part of a new radio telescope, called the Square Kilometre Array (SKA), will go online in Australia [10]. With over 60000 pieces, SKA's core of radio antennas will be very dense and homogeneous in contrast to LOFAR. Moreover, the frequency coverage will increase from 30 – 80 MHz in LOFAR to 50 – 350 Mhz in SKA [10]. With these changes, X_{max} -reconstruction resolutions are expected improve to about ~ 6 g/cm² [10]. To illustrate the difference in coverage, a simulation of radio sampling for both LOFAR and SKA is presented in Figure 5.

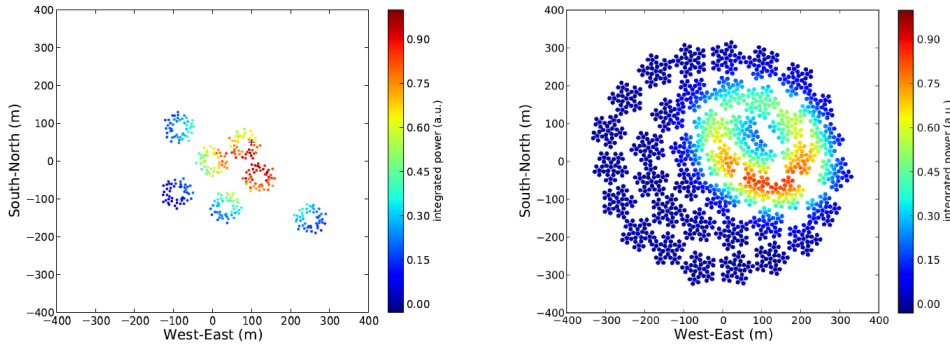


Figure 5: simulation of radio sampling for both LOFAR (left) and SKA (right). Figure taken from reference [10].

3.3 Shower parameters of interest

In order to answer the research questions about cosmic rays described in Sec. 2, observations are focused on energy- and mass-sensitive shower parameters that can differentiate between the distinct source models. In radio observations, these shower parameters of interest are typically the cosmic-ray energy E_0 and the atmospheric depth of the shower maximum X_{max} [9, 13]. However, other observables are under consideration as well. Among these are the

length L of the air shower and an asymmetry parameter R . The potential utility of parameter L is explored in Sec. 5.

3.3.1 Cosmic ray energy E_0

The all particle cosmic ray spectrum, presented in Sec. 2, is reliant on the accurate determination of the absolute energy scale of cosmic rays. Although not the only technique to reconstruct E_0 , radio measurements have been shown to efficiently yield accurate results for a few reasons [1, 9]. Firstly, as the radio signal is coherent, the measured power scales quadratically with the number of electrons and positrons in the shower, which in turn scales linearly with the energy of the primary particle [1, 9]. Secondly, at the ground, the electromagnetic component holds from 80 % up to 100% of the energy in the shower because of the efficient decay of charged hadrons, such that it strongly represents the energy of the primary component [9]. Finally, the radio signal travels essentially uninterrupted through the atmosphere [9]. This all results in radio observations at the ground essentially yielding a calorimetric energy of the shower [1, 9]. The power measurements at a characteristic lateral distance from the shower axis are now used to estimate the energy of the primary particle [1, 9]. Moreover, to cross-calibrate these estimations, the results are compared to universal and very accurate energy determinations in the 30 – 80 Mhz-range, accomplished in the Pierre Auger observatory [1, 9].

3.3.2 Atmospheric depth of shower maximum X_{max}

The atmospheric depth (in g/cm²) indicates how far a shower permeates the atmosphere and is defined by

$$X(h) = \int_h^\infty \rho(y) dy, \quad (2)$$

where h is the altitude measured from the ground, $\rho(h)$ is the matter density and the integral is evaluated along a straight vertical line [6]. The shower maximum is the point in longitudinal development, at which the air shower reaches its maximal radiating particle number N_{max} [9]. Examples of this longitudinal development are presented in Figures 3 and 6. The mean depth of the shower maximum at a given energy is now used to estimate the composition of the cosmic ray, using the relation [13]

$$\langle X_{max} \rangle \approx \sum_i f_i \left(c + D_p \ln\left(\frac{E}{A_i}\right) \right), \quad (3)$$

with f_i the fractions of nuclei with mass A_i , E the total electromagnetic shower energy and parameters c and D_p dependent on characteristics of hadronic interactions. For this estimation, the semi-superposition theorem is assumed. It states that a primary nucleus of mass A and energy E can be treated as a superposition of A nucleons of energy $E' = E/A$ together with some modifications in the treatment of its interaction length [13]. Moreover, showers initiated by heavy nuclei have a higher cross-section and multiplicity such that they will reach the maximum faster with less event-to-event fluctuations than those expected for proton initiated showers. Because of the direct correlation of X_{max} with the composition of the particle, this parameter can help differentiate between different source- and interaction models. Up until now, numerous approaches to measuring X_{max} have been developed. The primary approach for doing this in radio astronomy is described in further detail in Sec. 4.

3.3.3 Length parameter L and asymmetry parameter R

Although the maximum number of particles N_{max} together with the corresponding shower depth X_{max} , strongly characterise the longitudinal evolution, these aren't its sole determining features [2]. Describing the full shape of the profile may yield more information on the primary particle and first hadronic interactions [2]. To further characterise the shower, parameters L , associated with the length of the longitudinal distribution, and R , associated with the distribution's level of asymmetry, are introduced.

It has been shown that the longitudinal profiles of air showers have a universal shape after translation $X' = X - X_{max}$ and normalization $N' = N/N_{max}$, as shown on Figure 6. This profile can now be parametrised using equation [2]

$$N' = \left(1 + \frac{RX'}{L}\right)^{R-2} \exp\left(-\frac{X'}{LR}\right). \quad (4)$$

The length parameter L is associated with the width of the profile and is defined as $L = \sqrt{|X'_0\lambda|}$, with X'_0 the translated atmospheric depth of first interaction and λ an effective interaction length. The asymmetry parameter R is associated with the shape of the profile and is defined as $R = \sqrt{\lambda/|X'_0|}$.

3.4 Simulation techniques

The approach for air shower property reconstruction, further described in Secs. 4 and 5, is contingent on comparisons between measured and simulated two-dimensional radio distribution signals at the ground. Evidently,

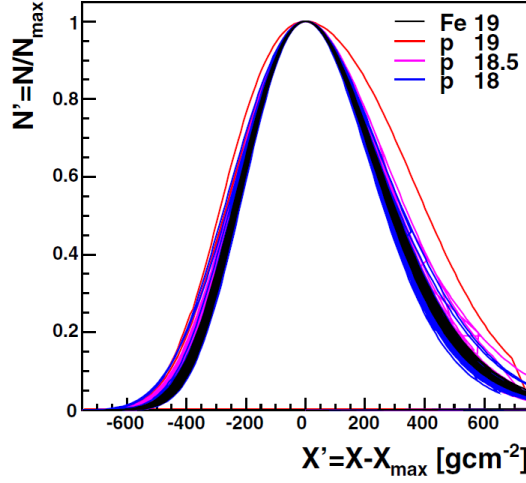


Figure 6: Normalized and translated longitudinal profiles for 100 showers of different energies and primary particles exhibiting the universal shape. Figure taken from reference [2].

this method requires an accurate, and preferably efficient, simulation routine [4]. A primary example of this is CoREAS, whose simulated profiles have been shown to consistently match observed showers at all geometries and energies and is hence consistently used in LOFAR observations [4, 11]. However, because this method is computationally heavy, more efficient techniques are explored. Given some generalisation, a promising model called template synthesis, reliant on synthesising templates of simulated air showers, could soon enhance CoREAS in shower property reconstruction performed by LOFAR and eventually SKA [5].

3.4.1 CoREAS

CoREAS uses a microscopic approach in simulating the radio signal of air showers, which means that each single electron and positron in the shower is considered separately [9]. The distribution of the charged particles is delivered by the particle simulation code CORSIKA [4]. From this distribution, the ‘endpoint formalism’ is used to calculate the resulting radio emission. This formalism calculates the signal of the moving particles by the radiation from the instantaneous accelerations of charges at the beginnings and ends of straight track segments [4, 9, 11]. A Monte Carlo simulation of the electromagnetic cascade using this formalism can now yield the electric field vector at a given location at the ground [4, 11]. For a full two-dimensional mapping of the power at the ground, the simulation is typically run for about

160 ground positions in a star-shaped pattern in the shower plane with eight arms, of which two are aligned with the $(\vec{v} \times \vec{B})$ -axis and an other pair with the $(\vec{v} \times \vec{v} \times \vec{B})$ -axis. Because the signal is known to be continuous and smooth for small differences in distance, linear interpolation can now yield the full two-dimensional power profile. An example of such a simulated power footprint is presented in Figure 7.

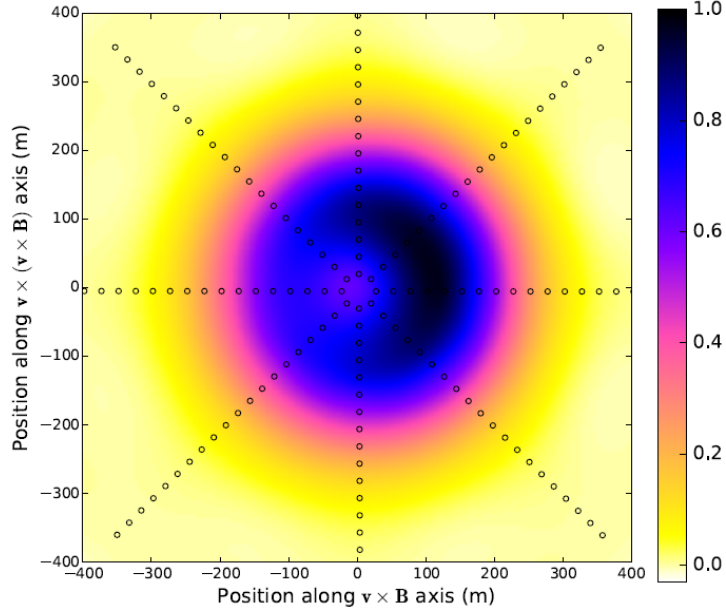


Figure 7: Simulated relative received power profile by CoREAS for proton shower with $X_{max} = 794 \text{ g/cm}^2$. The circles are the directly simulated antenna positions used for linear interpolation. Figure taken from reference [4].

Because CoREAS calculates the radiation directly from the positron-electron distribution given by CORSIKA, the simulation is parameter free such that there is no ambiguity in differentiating between different mechanisms. Therefore, the uncertainties from this model arise strictly from the uncertainties in the hadronic interaction models used in CORSIKA [9]. Finally, it is important to note that because the method requires a full microphysics Monte Carlo simulation, it is computationally demanding. Therefore, given that running a lot of simulations is important for the quality of parameter reconstruction, more efficient models would benefit shower observation research [5].

3.4.2 Template Synthesis

The longitudinal development of the shower is the starting point for using simulated radio pulses as templates. Evidently, this distribution is sensitive to the choice of interaction- and source model. However, from this point on, calculating the resulting radiation is strictly reliant on electrodynamics, hence independent from the underlying models for the longitudinal distribution. The template synthesis method takes advantage of this fact by reshaping a single longitudinal distribution as follows: A template is constructed by cutting the longitudinal particle distribution in slices such that the radiation of every particle, calculated by CoREAS, is assigned to exactly one slice as if it were one macroscopic source [5]. At a given position at the ground, the direct sum of the total electric field from every slice then yields the observable electric field for the whole shower. Using this template, a first attempt at simulating a target pulse can be achieved as follows: Naively assuming that the electric field in every slice is solely proportional to the number of particles N , the signal for each slice of the template can be rescaled to a target shower by varying N [5]. The direct sum of slices now yields the synthesized electric field $\vec{E}^{Synth}$ of the target shower, such that [5]

$$\vec{E}^{Synth}(\vec{r}, t) = \sum_X \frac{N^{Target}(X)}{N^{Temp}(X)} \vec{E}_{Slice}^{Target}(X, \vec{r}, t), \quad (5)$$

where X is the shower depth of a slice.

Although this yields promising results, the accuracy consistently falls short of expectations, especially for larger differences in the templates' and targets' X_{max} . To account for differences in the shower other than the particle number, the amplitude spectrum for every slice is incorporated. This amplitude spectrum of a slice for a given position on the ground $A(X, \vec{r}, X_{max}, f)$ is obtained by an analytical fit and is dependent on both the showers' X_{max} and the signals' frequency f . Because of the frequency dependence, this additional scaling factor undergoes an inverse Fourier transform back to the time domain. The final configuration of the method now becomes [5]

$$\vec{E}^{Synth}(\vec{r}, t) = \sum_X \frac{N^{Target}(X)}{N^{Temp}(X)} \mathcal{F}^{-1} \left[\frac{A(X, \vec{r}, X_{max}^{Target}, f)}{A(X, \vec{r}, X_{max}^{Temp}, f)} \vec{E}_{Slice}^{Target}(X, \vec{r}, t) \right]. \quad (6)$$

This refined template synthesis model has been shown to yield the desired precision and speed in simulating radio signals for arbitrary X_{max} [5]. In order to see if this technique generalises as well outside of the controlled

test conditions, template synthesis has to be applied in situations where CoREAS is typically used, like reconstructing the depth of shower maximum X_{max} . Moreover, the method is currently limited to simulating non-inclining showers, which means that each incoming particle is considered to arrive perpendicular to the atmosphere, with a core arriving position in the center of the plane. As long as these parameters can't be varied, these assumptions are too restrictive to apply the method to actual observations.

Because this method is contingent on the rescaling of the longitudinal development of the shower, implementing parameters that describe this distribution, other than X_{max} , can be easily implemented. Because of this flexibility and the favorable simulation speed, the template synthesis simulation method is the simulation tool of choice for most of the thesis. First, it will be used to test the X_{max} -reconstruction routine, described in Sec. 4. Moreover, in Sec. 5, the implementation of parameter L relies entirely on the methods simulation files too.

The concrete version of the simulation tool, used in this paper, functions as follows. A list of longitudinal shower distributions are provided by CORSIKA, each for non-inclined showers with a different maximum depth X_{max} and with a core arriving position right above the center of the ground plane. Each distribution can now be used as rescalable templates using the template synthesis method. Finally, the electric field for each of the rescaled distributions is calculated at six position at the ground, with radial distances of 1, 40, 75, 110, 150 and 375 m from the shower core, not taking into account the typical rotational asymmetry of the radio footprint. However, because the antenna at a radial distance of 1 m lies right under the arrival point of the shower, the simulated signal here is physically less reliable and subject to heavy fluctuations. Moreover, the simulated signal for the farthest antenna, at a radial distance of 375 m, is typically multiple orders of magnitude smaller than the at the other antenna positions. Therefore, both these antennas are discarded in calculations of the power distribution of an air shower that will be of further use in the following sections about parameter reconstruction. In Figure 8, a plot is shown of several power simulation attempts from templates with different $X_{max,temp}$, each attempting to construct the pulse of a target shower with shower maximum $X_{max,target} = 659.32 \text{ g/cm}^2$, for an antenna at radial distance of 75 m. After studying multiple of these pulse creations, it seems that this version of the template synthesis method provides quite accurate results. However, for increasing differences between $X_{max,temp}$ and $X_{max,target}$, the quality of reconstruction tends to deteriorate.

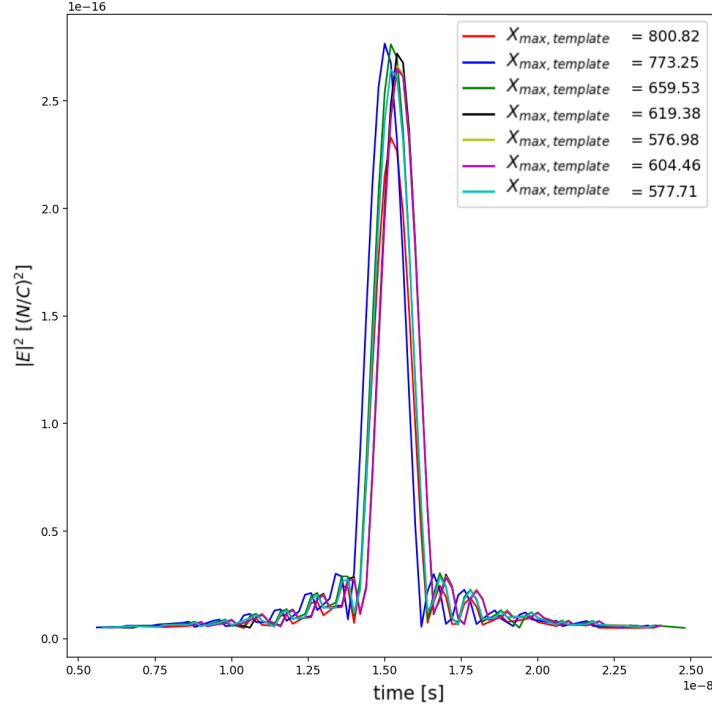


Figure 8: The simulated power provided by the template synthesis method, for different template showers, for an antenna at radial distance 75 m, for a target shower with shower maximum $X_{max, target} = 659.32$ g/cm².

4 Depth of shower maximum X_{max} reconstruction

Because the atmospheric depth of the shower maximum X_{max} gives direct insight into the composition of the primary particle, it is an important shower parameter to estimate. In radio observations, this is achieved by simulating the two-dimensional power profile for up to about forty showers per every real radio observation, each with different values for X_{max} , and then identifying the best-fitting simulation [4, 9].

4.1 Fit procedure

A radio observation of an air shower consists of measuring the electric field strength in function of the time at different antenna positions, each with coordinates (x_{ant}, y_{ant}) , out of which the power at those positions $P_{ant}(x_{ant}, y_{ant})$ is then easily calculated. Additionally, a noise level σ_{ant} on these measurements is derived from the observation. For each observation, forty two full

mappings of the two-dimensional power profiles of an air shower, with either a protons or iron primary particle and each with different X_{max} , are simulated such that at any position in the observatory plane, P_{sim} can be retrieved. To calculate how well a shower fits the data, the minimum- χ^2 fitting method is used. Hence, the quantity that has to be minimised is [4]

$$\chi^2 = \sum_{antennas} \left[\frac{P_{ant}(x_{ant}, y_{ant}) - f_r^2 P_{sim}(x_{ant} - x_0, y_{ant} - y_0)}{\sigma_{ant}} \right]^2, \quad (7)$$

with three free fit parameters: x_0 and y_0 , which represent the location of the shower axis, and f_r , a scaling parameter for calibration [4]. An example of such a fit for a real LOFAR shower observation, where the simulation that yielded the lowest χ^2 is used, is plotted in Figure 9. Note that the observation points are organised in superterp station, as described in Sec. 3.2.2.

Now, the minimal χ^2 -value of every simulation is plotted in function of its maximum atmospheric depth X_{max} . This plot is presented in Figure 9, where the blue circles represent proton shower simulations and the magenta squares represent iron shower simulations. Because of the approximately quadratic relation, the plot clearly shows that the value for X_{max} significantly influences the fit quality. However, in general, the data doesn't lie on a smooth curve, which might imply that other parameters influence the fit quality as well. Finally, in order to retrieve the estimation for X_{max} of the real shower, a parabolic fit around a selection of the lowest χ^2 -values is performed and its minimum calculated [4]. This minimum is now the estimation for X_{max} of the air shower observed by LOFAR. Now, a resolution of the reconstruction has to be approximated. But because the data doesn't lie on a smooth curve, using the width of the fitted parabola isn't an accurate estimation for the uncertainty on this X_{max} [4]. Hence, another method is used and is described in the subsequent paragraph.

4.2 Performance analysis

To determine the resolution of this X_{max} -reconstruction method, a specific method using only simulation data is proposed. It consists of applying the following routine to every simulated antenna. First the simulated power is evaluated on the positions of the LOFAR (or eventually SKA) antennas and then, Gaussian noise, equivalent to the one observed in the real shower, is added [4]. The resulting power now emulates real observation data. The remaining simulations are then used to apply the X_{max} fitting routine on, described in Sec. 4.1, to yield a value $X_{max, reco}^{sim}$. This can be compared to the actual X_{max}

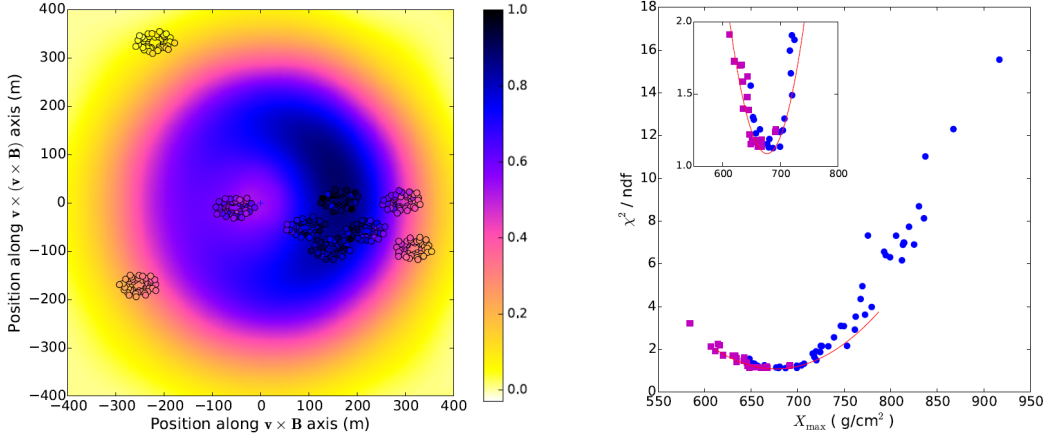


Figure 9: left: Two-dimensional fit of the LOFAR observations (circles) to the simulated power distribution that yields the lowest χ^2 [4]. right: Plot and parabolic fit of each simulations χ^2 -value in function of the atmospheric depth of the shower maximum X_{max} [4].

of the simulation, here named $X_{max,real}^{sim}$. After executing this entire procedure for every simulated shower, the distribution $|X_{max,reco}^{sim} - X_{max,real}^{sim}|$ can be visualised using a histogram. Finally, from such a histogram, of which an example is shown on Figure 10, the 1σ uncertainty is derived by selecting 68 % of the histogram distribution mass [4]. Applying this method to LOFAR observations, the typical resolution obtained for X_{max} -reconstruction is about $\sim 17 \text{ g/cm}^2$ [4]. However, with the increased sampling- and frequency coverage of SKA, which should go online soon, resolutions are expected to improve to about $\sim 6 \text{ g/cm}^2$.

However, because the χ^2 -values are known to consistently deviate from a perfect parabola, it is clear that additional parameters might influence the quality of fit. The case for parameter L , associated with the length of the shower, which will be further discussed in Sec. 5.

4.3 Using template synthesis

Finally, in order to test the quality of the template synthesis simulations, the method for determining the reconstruction resolution, described in Sec. 4.2, is applied to the simulation data described in Sec. 3.4.2. Forty six shower simulations with different X_{max} -values are generated using the template synthesis method. Because of the small amount of antenna positions, the power fitting routine considers the two polarisation direction separately. Moreover,

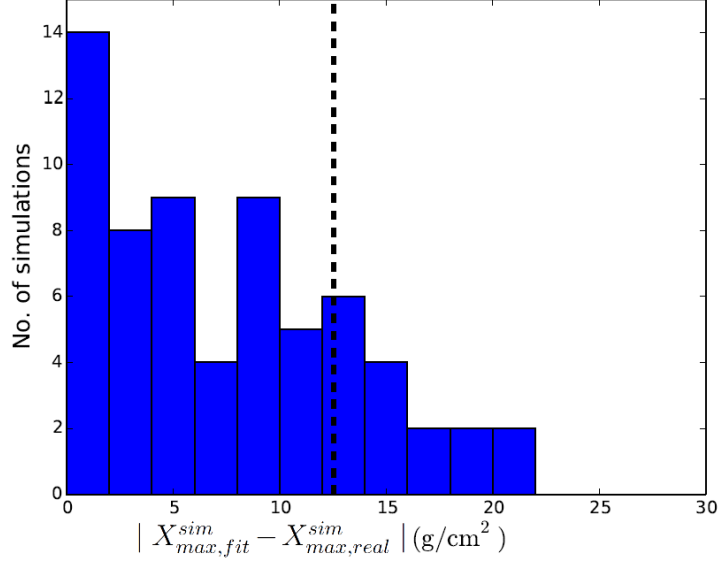


Figure 10: Distribution of $|X_{max,fit}^{sim} - X_{max,real}^{sim}|$ for a set of forty simulations from which the 1σ uncertainty is derived. Figure taken from reference [4].

because template synthesis considers only non-inclined showers with a core arriving position above the center of the ground plane, the shower axis coordinates (x_0, y_0) can be excluded from the fit, given by equation (7). Finally, in order to emulate the signal of actual measurements on which the other simulated power values are fitted, Gaussian noise with $\sigma_{Gauss,sim} = 0.032 \cdot P_{max}$, with P_{max} the maximum simulated power at a certain position, is added. This corresponds to a signal-to-noise ratio of 3.2%, which is substantially lower than the noise in real observations which is considered to be at maximum about 20%, a rather conservative approximation [4]. The reasons for the choice of this seemingly low signal-to-noise ratio on the simulation data can be motivated as follows: For real observations, around at maximum $N \sim 300$ data points per LOFAR observation and at maximum $N \sim 5000$ data points for SKA-observations, are expected, instead of the $N_{sim} = 8$ data points used here. Hence, it is safe to assume that the Gaussian error chosen to emulate observation data can be interpreted as larger, when extrapolating it to real measurements. The additional data points can be interpreted as corresponding to a decrease in signal-to-noise ratio approximated by the following formula:

$$\sigma_{Gauss,real} \approx \sqrt{\frac{N_{observation}}{N_{sim}}} \cdot \sigma_{Gauss,sim}, \quad (8)$$

with $N_{observation}$ the number of antennas used in by the telescope. This equation generally holds for increasing measurements without any extra information. Therefore, as the additional datapoints also probe more radial distances from the shower core, using this equation can be considered as a conservative approximation. Now, $\sigma_{Gauss,sim} = 0.032 \cdot P_{max}$ implies the use of about 300 antennas, which according to (8), corresponds to the conservative but more realistic signal-to-noise ratio of 20%. This number of antennas implies that the subsequent X_{max} -reconstruction resolution results should be obtainable in both LOFAR- and SKA observations.

Putting this all together, the resulting minimum χ^2 -fit consists of a two-dimensional fit with the scaling factor f_r as only fit parameter. In Figure 11, an example of a X_{max} reconstruction fit with template synthesized data is presented, together with a histogram of the resolution for all forty six simulation files with the corresponding uncertainty determination.

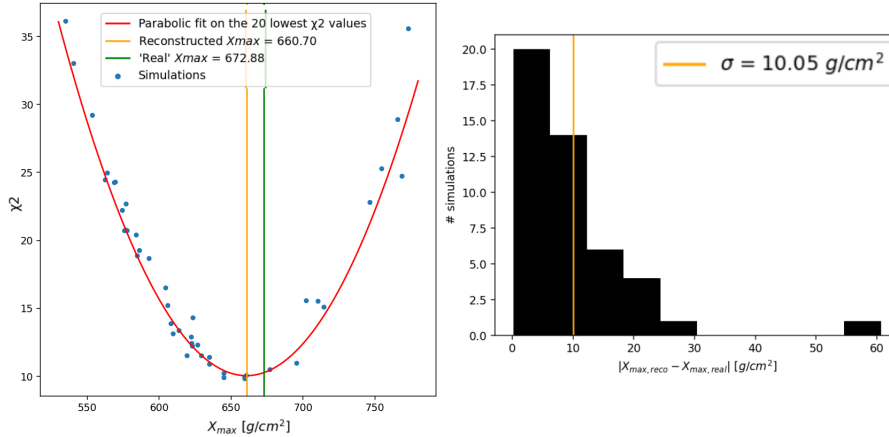


Figure 11: left: Example of X_{max} reconstruction fit for template synthesis generated data. right: Distribution of $|X_{max,fit}^{sim} - X_{max,real}^{sim}|$ for a set of forty six simulations provided by the template synthesis method, from which the 1σ uncertainty is derived.

The results indicate that the fitting routine works as it should for data generated by the template synthesis method. The differences in X_{max} show the same semi-parabolic behaviour as for the data simulated by CoREAS. Applying the routine on all of the forty six showers and calculating the standard deviation of the individual reconstruction resolutions, an uncertainty of around $\sim 10 \text{ g/cm}^2$ is typically found, which is an excellent resolution for X_{max} -reconstruction. It can therefore be concluded that it is highly likely that

the template synthesis model is suitable for the existing X_{max} -reconstruction fitting routine.

5 Implementing the length L of the shower

In order to improve primary particle composition measurements and further distinguish the effects of different source- and hadronic interaction models, implementing the reconstruction of shower parameters other than X_{max} , is looked into. Therefore, the case for L as such a parameter is explored, after which a reconstruction routine is proposed. This method, just like for X_{max} -reconstruction, relies on fitting simulated radiation footprints to actual measurements. Ideally, a practical implementation would mean reconstructing the L -value after X_{max} -reconstruction, both using the template synthesis method for the needed simulations.

5.1 The case for L as a useful reconstruction parameter

In order for L to be a useful shower parameter to reconstruct, it has to meet two criteria. First, it should be sensitive to the primary mass and hadronic physics in the shower, such that it can help differentiate between different source- and interaction models. Secondly, it has to significantly influence the quality-of-fit, such that a fitting routine similar to X_{max} -reconstruction is possible.

The possible use of L in differentiating between different source- and interaction models can be illustrated as follows: A lot of simulations are run for combinations of different source- and hadronic interaction models at 10^{18} eV and the X_{max} - and L values calculated. Each dot represents a possible outcome of X_{max} - and L after observing 1000 showers including statistical uncertainties. The resulting plot is presented in Figure 12. In order to improve differentiation, the different source model are strategically chosen. For instance, the different assumptions for the extragalactic component, labeled MIN, PCS and UFA, represent different extremes in contribution in the transition region. The minimal model (MIN) minimises the extragalactic contribution under the ankle. PCS, on the other hand, proposes a very high contribution. The third model, UFA, lies somewhere in the middle of these. A more detailed account of these models can be found in [16]. For the galactic component a prediction for Wolf-Rayet supernovas is used. For the primary compositions, two extremes (C/He=0.1 and C/He=0.4) were used in the calculation. Intermediate values are indicated with the colored bars

connecting these two extremes 12.

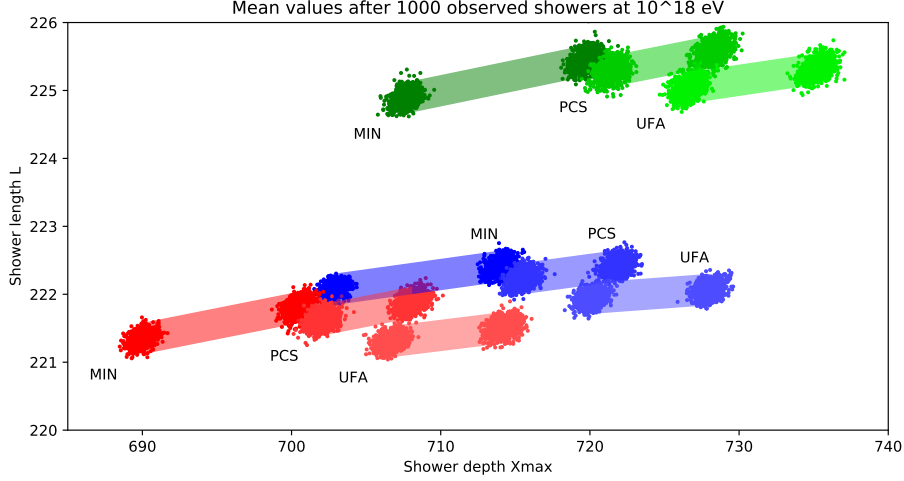


Figure 12: Simulation of average X_{max} and L observations with SKA at 10^{18} eV. The different colors indicate hadronic interaction models (green = Sibyll, blue = EPOS, red = QGSJET). Each label indicates a different assumption for the extragalactic component (MIN, PCS, UFA). For the galactic component a prediction for Wolf-Rayet supernovas is used. The different models occupy different regions in L - X_{max} space, but are highly degenerate in traditional measurements of only X_{max} [16].

From this plot it is clear that L is sensitive to the choice of source- and interaction models and that, in combination with X_{max} -measurements, it can be used to untangle the different hypotheses. If L can now be accurately reconstructed, actual observations of air showers can be used to see where they appear on this plot. If it lands in the designated area of a certain combination, it can be assumed that the respective models are more realistic descriptions of reality than the others. This could provide a lot of insight in the physics of cosmic rays at energies exceeding those of experiments conducted at CERN.

As discussed in Sec. 4, the almost-parabolic behaviour of the χ^2 -values for fits on simulation showers with different X_{max} , show that the quality of fit is dependent on X_{max} . In order for L to be a suitable parameter to reconstruct, it should behave similarly for simulation showers with different L -values. To check this, the template synthesis method is used to create signals at the ground as described in Sec. 3.4.2, for seven L -values ranging from 200 to

230 g/cm^2 and a constant $X_{max,target} = 555 \text{ g}/\text{cm}^2$. However, it is important to note that these are created out of templates of showers with different $X_{max,temp}$ -values. For this first small sample of data, a χ^2 minimisation method is used, analogous to the one for X_{max} described in Sec. 4.2. It considers imitating ‘real’ data points by adding Gaussian noise relative to the maximum simulated power of the respective simulation and then fitting these on the other simulated power patterns. An example of a typical distribution of these χ^2 -values is visible on Figure 13. A few things can be derived

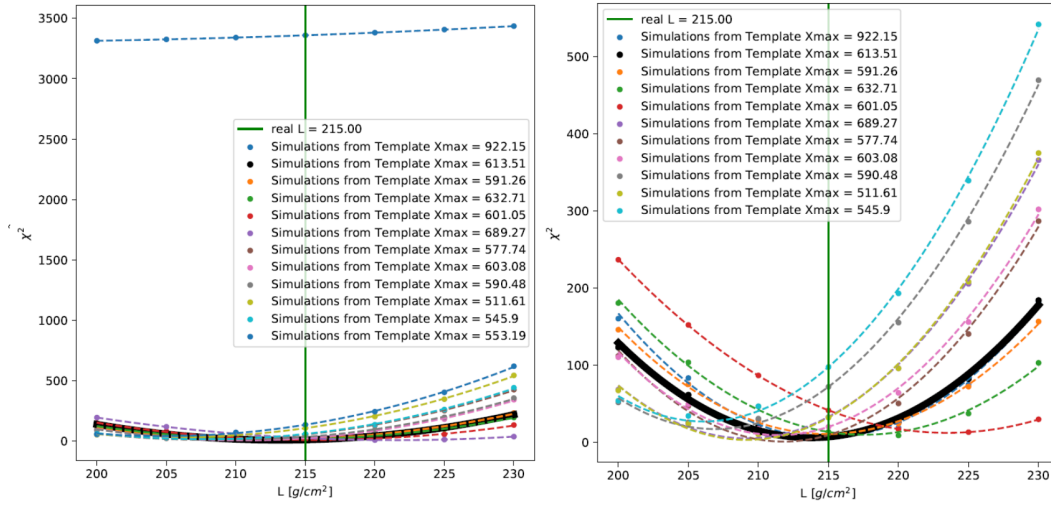


Figure 13: χ^2 -values distribution resulting from fitting to simulation files with different L -values and constant $X_{max} = 555 \text{ g}/\text{cm}^2$. Showers are created out of templates with different $X_{max,temp}$ and such that each parabola is plotted through the data points originating from the same template. The parabola through the template data points which also contains the observation emulating data, is indicated by the bold black parabola. On the left, χ^2 -values from template with $X_{max,temp} = 922.15 \text{ g}/\text{cm}^2$ are included while on the right these are omitted.

from this distribution. First, it is clear that most of the data points show the desired parabolic behaviour around the L -value on which the data is created, indicated by the green line. This shows that the quality-of-fit is sensitive to the simulated L -value, just like it is with X_{max} . Secondly, although all simulation files have the same target $X_{max,target} = 555 \text{ g}/\text{cm}^2$, the distribution shows a dependency on the atmospheric depth $X_{max,temp}$ of the template out of which simulation files are created. It is clear that for data from templates with $X_{max,temp}$ further away from the one out of which the data is created $X_{max,temp,data}$, the χ^2 -values rise substantially and the parabolic behaviour

decreases. An example of this is the $X_{max,temp} = 922.15 \text{ g/cm}^2$ which results in simulations that perform significantly worse than all the others, which are simulated by templates with $X_{max,temp}$ closer to the $X_{max,temp,data} = 613.51 \text{ g/cm}^2$. This implies that the template synthesis method shows measurable inaccuracies dependent on the choice of template and hence should always be chosen mindfully. The most logical way of doing this, is by first fitting a X_{max} -value for a given observation and then simulating pulses for showers with this atmospheric depth as target $X_{max,target}$ out of templates with $X_{max,temp}$ -values close to it. The next step is to find a reconstruction routine similar to the parabolic fit through a few of the lowest χ^2 points in X_{max} reconstruction.

5.2 Reconstruction technique

To further develop the reconstruction technique, a large batch of 336 simulation files is generated. All have a target shower maximum $X_{max,target} = 655 \text{ g/cm}^2$, with $X_{max,temp}$ -values ranging from 627 to 673 g/cm^2 such that the differences between the two are significantly smaller than in the first batch. For the different L -values, smaller intervals are chosen compared to the first batch, with values ranging from 200 to 230 g/cm^2 , in steps of 2 g/cm^2 . The typical Gaussian noise addition and χ^2 -fitting routine is applied to one of the files. The resulting distribution of χ^2 is plotted on the top left of Figure 14 and further confirms the behaviour described in Sec. 5.1. Now, a logical choice for reconstructing a L -value, is by using the same method as in the X_{max} -reconstruction. This means taking a few of the lowest χ^2 -points, fitting a parabola through them and finding the L -value for the minimum of the parabola. However, for L -reconstruction, the differences in $X_{max,temp}$ result in parabolas closely stacked on top of each other. Hence, selecting a few of the best points results in a elliptical cluster of data points, through which fitting a parabola is difficult and yields bad results. To evade this problem, new techniques are explored.

Because of the stacked parabola structure, a parabola can be fitted through each of these. Note, that in order to do this, it should be tracked from which template shower each χ^2 point originates. It is through these points that each parabola is fitted. A first possible reconstruction method considers calculating the mean of the L -values of the minima of all these parabola and use this as L_{reco} . This method is shown on the top right of Figure 14. However, this first method generally performs poorly. This is because the parabolas with generally higher χ^2 -values, tend to show some inaccuracies, thus yielding minima increasingly divergent from the actual L_{real} . These inaccuracies

increase considerably for increasing Gaussian noise for the data creation, a phenomena which will be further explored in Sec. 5.3.

To avoid the imprecision of the top parabolas, it can be chosen to simply select the parabola with the lowest minimum and use its L -value as L_{reco} . An example of this technique is shown on the bottom left of Figure 14. Although this second method often yields accurate results, it is sensitive to significant outliers. First of all, because it only selects one parabola, it only uses a small amount of data points for the reconstruction. This means that small fluctuations in these points can lead to significant inaccuracies, certainly when increasing the Gaussian noise when creating data. Secondly, the worst fitting parabola can wrongly be selected as the parabola to use. This is because these points can deviate enough from a parabolic shape that they form a line, in which case the fitted parabola's minimum can be very low but far away from the actual data points through which it is fitted. Naturally, the resulting L_{reco} then yields inaccurate and physically impossible results. Thirdly, this method is very sensitive to the selection of test showers. Hence, removing or adding one might result in very different results. Because of these three reasons this method is considered too unstable.

A third technique attempts to strike a balance between the two former ones. First, in order to remove the effect of a bad parabola with a low minimum, like described in the second method, the highest χ^2 -data point is found and its corresponding parabola discarded. Now for the remaining χ^2 -values, in order to increase the amount of data points used compared to the second technique without using all parabolas like in the first method, a few of the lowest points are selected, which tend to result in elliptical clusters of points. All parabolas with at least one of its points in this cluster are used, the rest is discarded. For these remaining 'good' parabolas, the minima are calculated. The mean of these minima is now considered the value for L_{reco} . An example of this routine is found on the bottom right of Figure 14. In general, this third method performs the best out of the three. However, applying this particular reconstruction technique on other simulation sets, the method perform differently. This suspected poor generalisation is the result of a particular weakness in the technique: discarding only the single parabola containing the highest χ^2 -point, may be inadequate if multiple parabolas deviate significantly from the rest. Although revisiting this problem in future applications is recommended, the results in this thesis should still be indicative of obtainable resolutions. In order to further analyse the performance of all three the methods, the performances of the different methods are quantified in the subsequent paragraph.

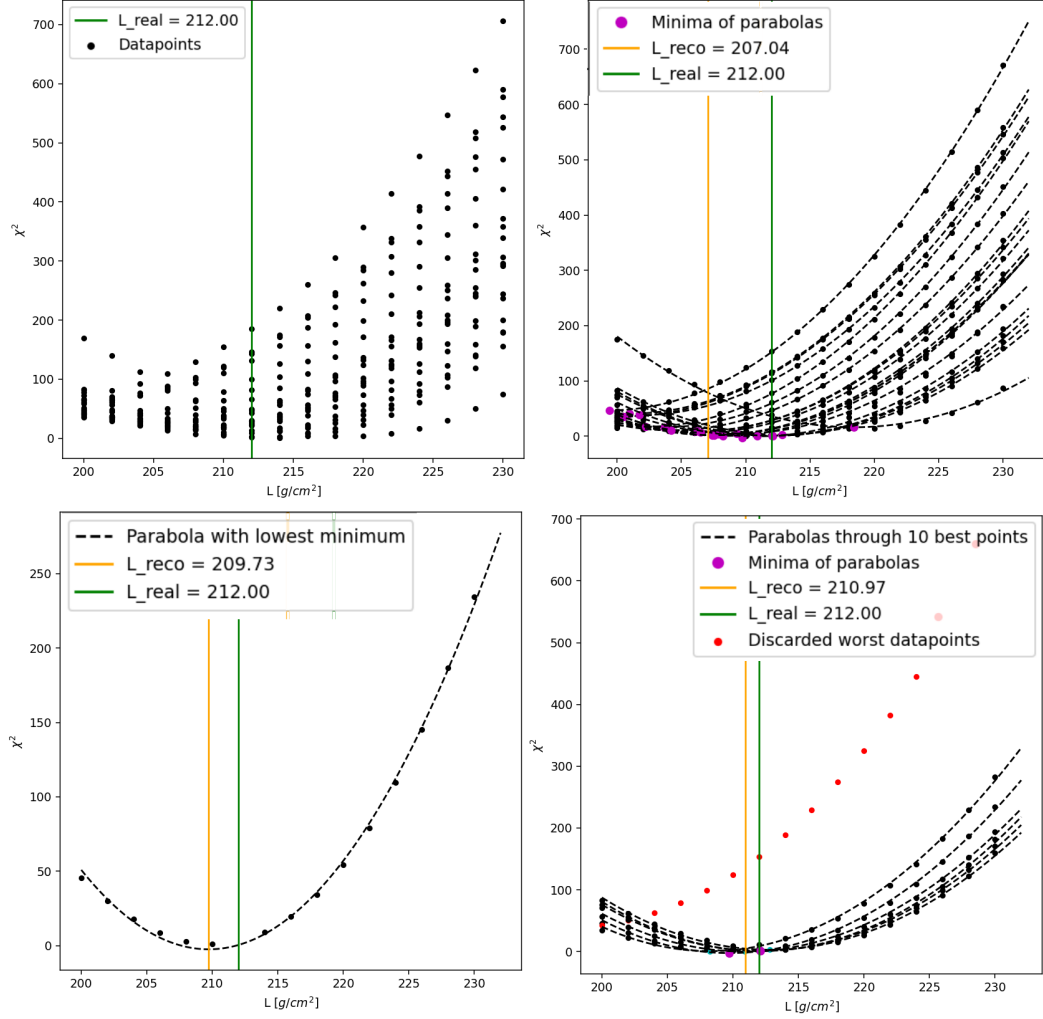


Figure 14: top left: χ^2 -values distribution for fitting 335 simulation files to one simulation file with Gaussian noise of $\sigma_{\text{Gauss}, \text{sim}} = 0.003 \cdot P_{\text{max}}$. top right: Finding L_{reco} by taking mean of minima of all parabolas. bottom left: Finding L_{reco} by taking minimum of lowest parabola. bottom right: Finding L_{reco} by taking mean of minima of a few ‘good’ parabolas by selecting the ten lowest χ^2 -data points.

5.3 Performance analysis

In order to predict if these L -reconstruction techniques can be applied to measurements performed by LOFAR and eventually SKA, accurate predictions of the obtainable resolution are needed. Because L -values typically only range from about 200 to 230 g/cm², the desired reconstruction resolution is smaller than the one for X_{max} . For this research, resolutions of 3 g/cm² or better are considered accurate enough. An estimation for the general reconstruction resolution for each fitting method can be easily retrieved by first performing the fitting routine for every simulation file, then plotting the resolution of each respective reconstruction in a histogram and finally calculating a definitive resolution-value such that it contains about 68% of the histogram mass, just like in X_{max} -reconstruction. However, these results are very sensitive to the choice of Gaussian noise $\sigma_{Gauss,sim}$ added to a simulation file to emulate real observation data. A few resolution calculations using the histogram method for some different $\sigma_{Gauss,sim}$ and fitting techniques are shown in Figure 15.

The noise on real observations can be approximated with a relative Gaussian error of $\sigma_{Gauss,real} = 0.2 \cdot P_{max}$. This 20% signal-to-noise ratio is a safe choice, because for observations this will often be lower. Moreover, because the relative error is chosen relative to the maximum power, the antennas observing a weaker signal, will have over-estimated noise added to them. Hence, the resolutions obtained for simulations with a 20% signal-to-noise ratio on the maximum power, can be considered as a lower limit on the actual reconstruction. However, as can be seen in Figure 15, using $\sigma_{Gauss,sim} = 0.2 \cdot P_{max}$ on the template synthesis data, the desired resolutions are not obtained. Nevertheless, the $\sigma_{Gauss,sim}$ may be chosen significantly smaller than $\sigma_{Gauss,real}$ for the same reasons described in Sec. 4.3: The template synthesis model uses only $N_{sim} = 8$ data points in the fitting routine, which is very little compared to observations made by LOFAR and eventually SKA. For LOFAR this corresponds to a maximum of about $N_{LOFAR} \sim 300$ and for SKA this will correspond to a maximum of about $N_{SKA} \sim 5000$. The corresponding improvements in signal-to-noise ratio are again approximated by equation (8). Aiming for an observation noise of $\sigma_{Gauss,real} = 0.2 \cdot P_{max}$, the obtained resolution using the third fitting routine described in Sec. 5.2, is calculated in function of the number of antennas $N_{observations}$ to be used in an observation. The results are plotted and shown in Figure 16.

Figure 16 clearly shows that for observations performed with at least about 5000 antennas, it is expected that L can be accurately reconstructed. Know-

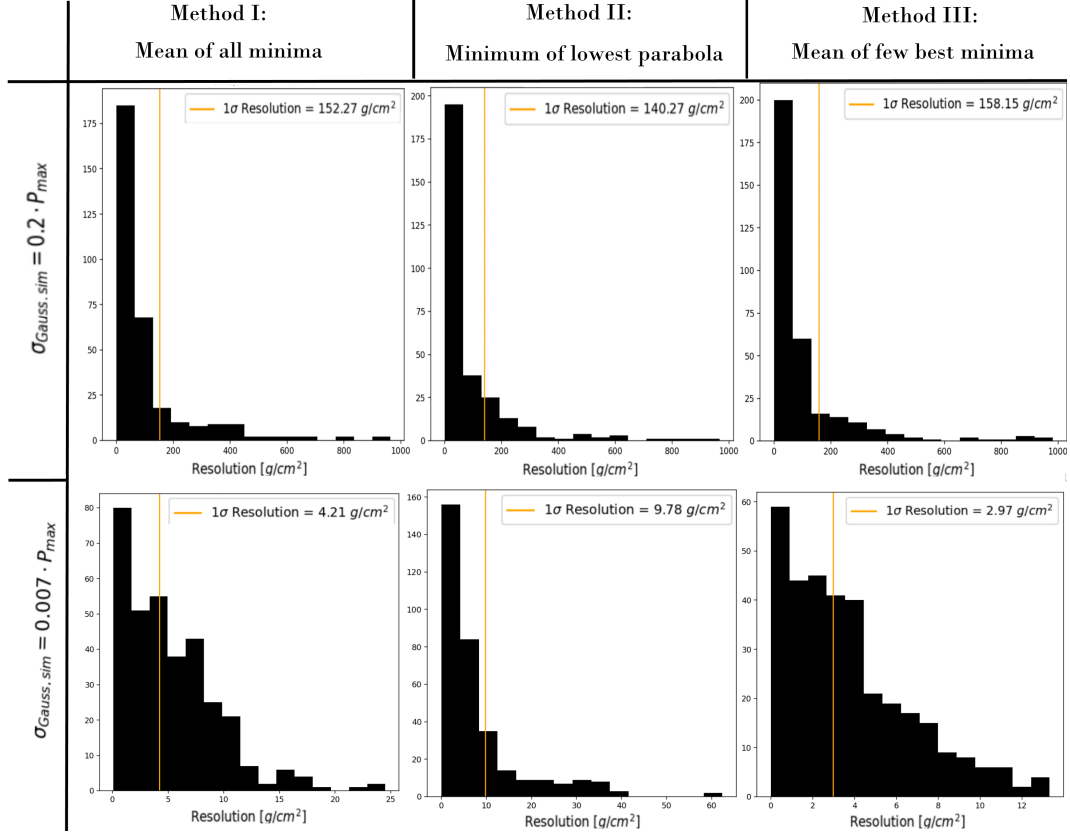


Figure 15: Resolution determination histograms of 336 simulations for the three techniques described in Sec. 5.2, for added noise of either $\sigma_{Gauss, sim} = 0.2 \cdot P_{max}$ or $\sigma_{Gauss, sim} = 0.007 \cdot P_{max}$ to emulate data. In the third technique, the ten lowest χ^2 points are used to select the ‘best’ parabolas.

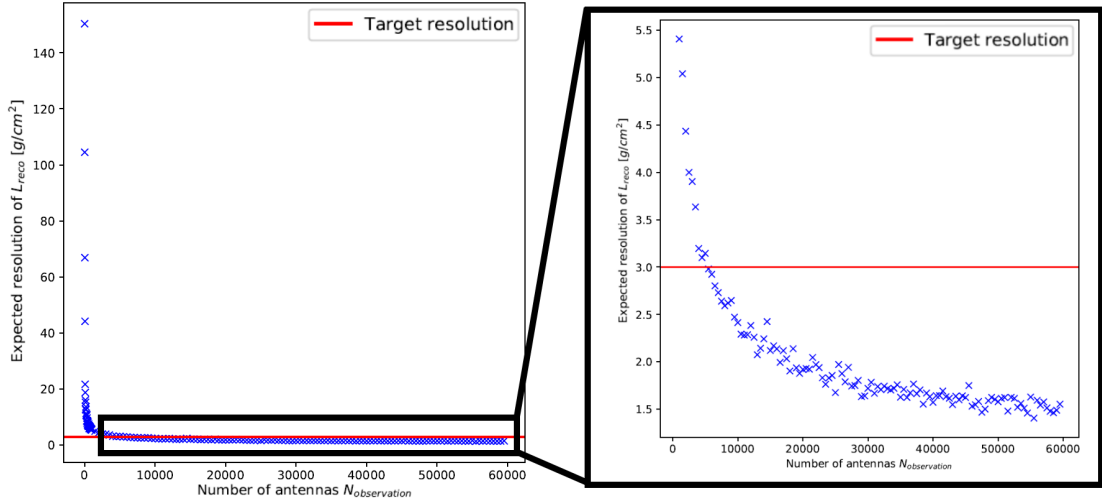


Figure 16: The expected resolutions, calculated via the method described in Sec. 5.3, for the third fitting routine, described in Sec. 5.2, in function of the number of antennas $N_{\text{observations}}$ to be used in an observation.

ing that LOFAR has about ~ 300 active antennas in a radio footprint, it will be difficult to reach the desired resolution. However, it has to be noted that the approximation was rather conservative for two reasons. First of all, the choice of a realistic noise of 20% of the maximum power is generally an overestimation. Secondly, the additional datapoints compared to the 8 used in this routine, contain more information than just an increase in measurements because they probe much radial distances from the shower core. Therefore, reconstructing L for LOFAR observations might still yield insightful results, although a resolution below $3 \text{ g}/\text{cm}^2$ is unlikely. In contrast, with the prospect of about 5000 antennas per observation in SKA, it is highly likely that for these future observations, L can be reconstructed with excellent accuracy.

6 Conclusions

This thesis explored adding the reconstruction of air shower parameter L to the current routine used in LOFAR and eventually SKA. Implementing this parameter might lead to more information about the sources and hadronic interactions taking place for cosmic rays in the transition region of the energy spectrum, i.e. from the second knee up to the ankle. This information might consequently contribute in answering the open questions about the nature and location of the transition from galactic to extra-galactic cosmic rays.

Additionally, a promising new simulation tool called ‘template synthesis’ is used and its applicability for future research analysed.

First, the template synthesis method was applied to the existing reconstruction routine of the shower maximum X_{max} . Despite some conservative assumptions, the routine has shown to yield desirable results when applied on the template synthesis data. The obtained expected reconstruction resolution for LOFAR observations of about $X_{max} \sim 10 \text{ g/cm}^2$ is very promising. Moreover, as SKA will use more antennas in its measurements, it is reasonable to expect even higher resolutions.

Secondly, reconstructing shower parameter L , associated with the length of the longitudinal development of the shower, is looked into. A first simulation batch showed that L is indeed a suitable parameter to reconstruct, as it is sensitive to the primary mass and hadronic physics in the shower and has significant influence on the quality-of-fit. A second simulation batch was used for exploring a fitting routine and approximating an obtainable corresponding resolution. One proposed fitting routine, which calculates the mean of the minima of a select number of paraboles of χ^2 -values, yielded desirable reconstruction resolutions for this particular simulation set. Using some conservative assumptions, this method is expected to yield a L -reconstruction of at minimum $\sim 3 \text{ g/cm}^2$ for observations with at least about 5000 antennas. Although these are substantially more antennas than in a typical LOFAR observation, SKA should have around this number of antennas available per observation. It is therefore expected that observations by SKA will be able to accurately and efficiently reconstruct parameter L , using template synthesised simulation data.

To further substantiate this positive result, some remarks have to be made. First of all, several approximations are used in extrapolating the results of the data from small scale simulations to what is expected in actual observations. One of which is the choice of a target signal-to-noise ratio of 20% relative to the maximum power, which can actually be considered an overestimation, hence favoring the positive final results in L -reconstruction. Moreover, equation (8), central in the extrapolation, does not incorporate the additional information actual observations will give by probing more of the antennas’ radial distances from the shower core. Again, this provides an optimistic view on the obtained resolution. A second note concerns the generalisation of the used fitting routine. Although the proposed method of removing the worst fitting parabola yields the desired results on the used simulation files, it might be sensitive to failure when applying it on actual measurements and differ-

ent simulation files. This is because multiple parabolas might significantly deviate instead of only one. It is therefore recommended to reinvestigate this fitting routine and to generalise it before applying it to actual observations. However, holding into account the relatively stable behaviour of the χ^2 datapoints on which to apply the fitting routine, it is safe to consider that a more stable fitting routine is achievable and will yield similar results to the ones obtained in this thesis. Finally, the current template synthesis method is limited to only non-inclined showers at core position (0,0). To apply the simulation routine on actual observations, different inclinations and positions have to be incorporated in the simulation tool. Moreover, the current restrictions on the template synthesis method result in omitting the core fit in the χ^2 calculations (7). For SKA observations, implementing a core fit probably won't significantly affect the reconstruction resolution, as the ground plane has enough antennas to perform equally accurate measurements for air showers with arbitrary core position. In LOFAR observations however, the dipole stations are positioned around the core, such that a shower with a different core position might severely influence the quality of observation and hence the obtained resolution.

In conclusion, it is safe to assume that, using SKA, implementing reconstruction of L with a resolution of ~ 3 g/cm² or better is achievable. Moreover, because L has shown to help differentiate between difference source- and interaction models, implementing it in future research will most likely help answer the open questions regarding the sources and hadronic interactions at energies ranging from the second knee up to the ankle. Assuming the template synthesis method will be further developed and with the imminent opening of SKA in 2023, actual implementation of the proposed reconstruction routine could be close at hand.

References

- [1] AAB, A., ET AL. Measurement of the radiation energy in the radio signal of extensive air showers as a universal estimator of cosmic-ray energy. *Physical Review Letters* 116, 24 (2016).
- [2] ANDRINGA, S., CONCEIÇÃO, R., AND PIMENTA, M. Mass Composition and Cross-section from the Shape of Cosmic Ray Shower Longitudinal Profiles. *Astroparticle Physics* 34, 6 (2010), 360.
- [3] BATISTA, A., ET AL. Open questions in cosmic-ray research at ultrahigh energies. *Frontiers in Astronomy and Space Sciences* 6 (2019), 23.
- [4] BUITINK, S., CORSTANJE, A., ENRIQUEZ, J., FALCKE, H., HORANDEL, J., HUEGE, T., NELLES, A., RACHEN, J., SCHELLART, P., SCHOLTEN, O., TER VEEN, S., THOUDAM, S., AND TRINH, T. Method for high precision reconstruction of air shower x-max using two-dimensional radio intensity profiles. *Physical Review D* 90 (08 2014).
- [5] BUTLER, D., HUEGE, T., ENGEL, R., AND SCHOLTEN, O. Universality and template synthesis of cosmic ray air shower radio emission (phd thesis), 2019.
- [6] CRUZ MORENO, J., AND SCIUTTO, S. Characterization of the atmospheric depth profile using the ground-level temperature: The case of Malargüe, Argentina. *European Physical Journal Plus* 128 (2013), 104.
- [7] ENGEL, R., HECK, D., AND PIEROG, T. Extensive air showers and hadronic interactions at high energy. *Annual Review of Nuclear and Particle Science* 61, 1 (2011), 467–489.
- [8] HÖRANDEL, J. R., ET AL. High-energy cosmic rays: galactic or extragalactic? *Frascati Phys. Ser.* 64 (2017), 95–102.
- [9] HUEGE, T. Radio detection of cosmic ray air showers in the digital era. *Physics Reports* 620 (Mar 2016), 1–52.
- [10] HUEGE, T., BRAY, J. D., BUITINK, S., BUTLER, D., DALLIER, R., ECKERS, R. D., ENSSLIN, T., FALCKE, H., HAUNGS, A., JAMES, C. W., ET AL. Ultimate precision in cosmic-ray radio detection — the ska. *EPJ Web of Conferences* 135 (2017), 02003.
- [11] HUEGE, T., LUDWIG, M., AND JAMES, C. W. Simulating radio emission from air showers with CoREAS. *AIP Conf. Proc.* 1535, 1 (2013), 128.

- [12] KACHELRIESS, M. Transition from galactic to extragalactic cosmic rays. *EPJ Web Conf.* 210 (2019), 04003.
- [13] KAMPERT, K.-H., AND UNGER, M. Measurements of the cosmic ray composition with air shower experiments. *Astroparticle Physics* 35, 10 (2012), 660–678.
- [14] MATTHEWS, J. A Heitler model of extensive air showers. *Astropart. Phys.* 22 (2005), 387–397.
- [15] PIEROG, T. Hadronic interactions and air showers: Where do we stand? *EPJ Web Conf.* 208 (2019), 02002.
- [16] THOUDAM, S., ET AL. Cosmic-ray energy spectrum and composition up to the ankle: the case for a second galactic component. *A&A* 595 (2016), A33.
- [17] TODOR, S. Cosmic rays and extensive air showers. *Cern Proceedings* 2010 (2010).
- [18] UNGER, M. Cosmic rays above the knee. *arXiv: Astrophysics* (2008).