

# A Bayesian method for air-shower reconstruction using Information Field Theory

Karen Terveer <sup>a</sup>, Sjoerd Bouma <sup>a</sup>, Stijn Buitink <sup>b,c</sup>, Arthur Corstanje <sup>b,c</sup>,  
 Vital De Henau <sup>b</sup>, Vincent Eberle <sup>d,e</sup>, Torsten A. Enßlin <sup>d,e,f</sup>, Philipp Frank <sup>g,d</sup>,  
 Tim Huege <sup>h,b</sup>, Philipp Laub <sup>a</sup>, Katharine Mulrey <sup>c,i</sup>, Anna Nelles <sup>a,j</sup>, Simon Strähnz <sup>h</sup>,  
 Satyendra Thoudam <sup>k</sup>, Keito Watanabe <sup>h</sup>

<sup>a</sup> Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

<sup>b</sup> Inter-University Institute For High Energies (IIHE), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium

<sup>c</sup> Department of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

<sup>d</sup> Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

<sup>e</sup> Ludwig-Maximilians-Universität München (LMU), Geschwister-Scholl-Platz 1, 80539 München, Germany

<sup>f</sup> Deutsches Zentrum für Astrophysik, Postplatz 1, 02826 Görlitz, Germany

<sup>g</sup> Kavli Institute for Particle Astrophysics & Cosmology (KIPAC), Stanford University, CA 94305, Stanford, USA

<sup>h</sup> Institut für Astroteilchenphysik, Karlsruhe Institute of Technology (KIT), P.O. Box 3640, 76021 Karlsruhe, Germany

<sup>i</sup> Nikhef, Science Park Amsterdam, 1098 XG Amsterdam, The Netherlands

<sup>j</sup> Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

<sup>k</sup> Department of Physics, Khalifa University of Science and Technology, P.O. Box 127788, Abu Dhabi, United Arab Emirates

## ARTICLE INFO

### Keywords:

Air shower reconstruction  
 Radio detection  
 Bayesian inference  
 Information Field Theory  
 LOFAR

## ABSTRACT

The radio detection of extensive air showers provides a powerful method for studying the origin of high-energy cosmic rays. The Low-Frequency Array (LOFAR) offers unprecedentedly detailed measurements of the radio emission footprint. However, fully exploiting this information requires advanced reconstruction techniques. In this paper, we introduce a novel framework for air shower reconstruction based on Bayesian inference and Information Field Theory (IFT). Our method is built on a fully differentiable forward model of the radio signal, which incorporates a physical emission parameterization and a precise wavefront model. Additionally, we augment this physical model with Gaussian processes to account for systematic uncertainties in both the signal fluence and arrival timing. By leveraging gradient information, our approach enables efficient (three orders of magnitude acceleration w.r.t. the legacy method) and robust inference of the underlying physical shower parameters, such as primary energy and the depth of shower maximum,  $X_{\max}$ . This work provides not only point estimates but also a rigorous quantification of uncertainties. We achieve a resolution in  $X_{\max}$  of 25 g cm<sup>-2</sup> and a radiation energy resolution of 12% on simulations for LOFAR.

## 1. Introduction

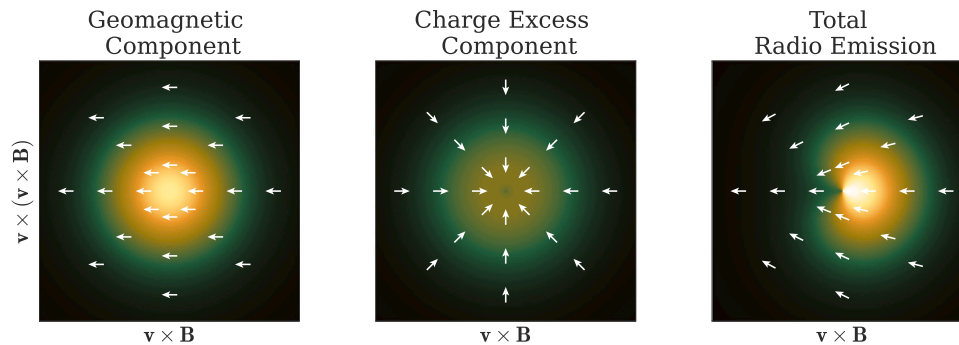
The origin of high-energy cosmic rays remains one of the most significant open questions in astroparticle physics. As charged particles, their trajectories are deflected by Galactic and intergalactic magnetic fields, obscuring the path to their astronomical sources. With their flux decreasing according to a power-law with energy, a direct detection at highest energies is no longer feasible. Our primary means of studying them is through the Extensive Air Showers (EAS) they initiate upon entering Earth's atmosphere. The properties of an EAS, particularly the energy of the primary particle ( $E$ ) and the atmospheric depth of the shower maximum ( $X_{\max}$ ), from which statistically the mass

composition can be inferred, are critical observables for constraining acceleration models and studying the mass composition of the cosmic-ray flux [1].

For decades, EAS have been observed using ground arrays of particle detectors and optical telescopes that detect atmospheric fluorescence or Cherenkov light. Although successful, these techniques have limitations; for example, fluorescence detection is limited to clear nights without moon, resulting in a duty cycle of 10%–15% [2]. In the last two decades, the detection of EAS radio emission has emerged as a complementary and standalone technique [3]. This emission provides a

\* Corresponding author.

E-mail address: [karen.terveer@fau.de](mailto:karen.terveer@fau.de) (K. Terveer).



**Fig. 1.** The radio pattern of the two emission mechanisms in the 30-80MHz band, the Geomagnetic effect (left) and the Charge Excess effect (middle) along with the direction of their E-Field indicated by arrows. The total emission (right) shows a characteristic asymmetric pattern due to the interference of the two effects. Patterns produced using parameterization from [19].

calorimetric measurement of the shower's electromagnetic component, is not significantly absorbed in the atmosphere, and can be detected with a near-100% duty cycle.

Previous work has shown excellent agreement between radio and other methods like Fluorescence detection [4], Cherenkov-light detection [5], and particle detectors [6], delivering superior per-event precision [7]. It has also been shown that the obtainable energy-scale has comparably small uncertainties and is consistent between experiments [8–10]. However, a consistent reconstruction that uses all information available in the electric field (i.e. amplitude, polarization, phase) without relying on very costly per-event simulations has been lacking.

We present a new approach to analyzing the datasets provided by modern radio arrays. We introduce a novel, end-to-end framework for air shower reconstruction that leverages the principles of Bayesian inference. We treat the reconstruction as a probabilistic inference problem, which allows for the consistent propagation of uncertainties and the incorporation of prior physical knowledge. To address the high dimensionality inherent in analyzing the full radio signal field, we employ the formalism of Information Field Theory (IFT) [11].

IFT has already been applied to various astrophysical problems, including reconstructions of large-scale cosmological structures [12], mapping of the Galactic interstellar medium [13], and reconstructing electric fields of radio signals from cosmic rays and neutrinos [14, 15]. Recent works have demonstrated the potential of IFT in air-shower reconstruction using parameterized forward modeling of the radio signal [16,17] and template-based synthesis of the full electric-field signal [18].

We demonstrate that this method can infer key shower parameters - arrival direction, core position, energy, and  $X_{\max}$  - along with their full posterior probability distributions, all while treating the air shower as a single phenomenon rather than a collection of separate observables, thereby maximizing the scientific potential of radio-detection arrays. This method simultaneously utilizes timing and fluence data with shared parameters, distinguishing it from previous approaches that relied on fluence only. Additionally, we will show that this method can reconstruct a larger selection of events than previous methods, allowing a future analysis of the complete LOFAR1.0 data set for a mass composition study. The treatment of the air shower as a random field additionally allows us to recover not just point estimates of the shower parameters, but their full posterior distributions and thus uncertainties on all parameters consistent with the underlying prior model and the data.

We present the method, tested on simulations and applied to a first selection of EAS measured with LOFAR.

### 1.1. Radio emission in extensive air showers

The charged particles in an EAS propagate in a thin disk called the *shower front*, which propagates along the direction of the primary cosmic ray, the shower axis  $\mathbf{v}$ . There are two primary mechanisms producing radio emission in an air shower. The dominant one, also called the geomagnetic effect, arises due to the deflection of the electrons and positrons into opposite directions by the earth's magnetic field  $\mathbf{B}$  in combination with their deceleration through interactions with the atoms of the atmosphere. The Lorentz force is given by  $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$ . The resulting charge separation creates a transverse current  $\mathbf{J}_{\perp}$  flowing perpendicular to the shower axis. With the longitudinal development of the shower, this transverse current first rises and then dies out with the number of charged particles, causing a time-dependent change in the transverse current  $\frac{\partial \mathbf{J}_{\perp}}{\partial t}$ , which radiates electromagnetic waves [20,21]. The strength of this emission scales with  $\sin(\alpha)$ , where  $\alpha$  is the angle between the geomagnetic field and the shower axis, meaning that the geomagnetic effect is nearly zero if the shower axis is parallel to the geomagnetic field and the emission is strongest for showers perpendicular to it [22]. The emission from the geomagnetic effect is linearly polarized in the direction of the Lorentz force,  $(\mathbf{v} \times \mathbf{B})$ .

The second, generally sub-dominant effect is the charge excess effect, also referred to as Askaryan Effect [23]. As the shower develops, low-energy electrons are continuously stripped from atmospheric molecules via Compton scattering, and positrons from the shower annihilate with electrons. This results in the shower front accumulating a net negative charge excess, typically on the order of 10%–20% of the total number of particles [24]. As this net charge also varies in time while the shower develops, it radiates a sub-dominant radio pulse. The electric field from this effect is linearly polarized with the electric field vectors pointing radially inwards towards the shower axis in the shower plane.

As radio waves experience negligible attenuation in the atmosphere, the total emission from both effects reaches the ground. The two components superpose as  $\mathbf{E} = \mathbf{E}_{\text{geo}} + \mathbf{E}_{\text{ce}}$ , under the approximation that the components arrive in phase, yielding a purely linearly polarized signal. As both components have different polarization angles, constructive interference between the two on one side of the shower footprint generates a stronger signal, while the polarization vectors are opposite on the other side, weakening the electric field through destructive interference. Fig. 1 showcases this phenomenon and the resulting asymmetric radio footprint that reaches the ground.

As the emission is coherent at frequencies of tens of MHz, the electric fields from the individual particles add up constructively, resulting in the total radiated energy scaling with the number of particles in the shower ( $E_{\text{rad}} \propto N^2$ ) and thus also scaling quadratically with the energy of the primary particle  $E_{\text{CR}}$ . The radio signal therefore

provides a calorimetric measure of the electromagnetic component of the shower [6].

The emission is beamed forward because the shower front moves relativistically, and for precise calculations, it is important to account for the index of refraction of the atmosphere, which causes signal compression at the Cherenkov angle. Consequently, the shower pattern is sensitive to the geometrical distance to the shower maximum and thus also the slant distance  $D_{X_{\max}}$  in  $\text{g cm}^{-2}$ . To first order, this impacts the size of the footprint on the ground, as showers that developed at larger altitudes will produce much larger footprints than ones that develop closer to the ground [25]. The  $D_{X_{\max}}$  controls the shape of the radio footprint by determining how mature the air shower is when observed: showers close to the ground give a narrow, axis-centered peak, intermediate ones a broad Gaussian-like pattern, and far showers a distinct Cherenkov ring with a conical peak [19].

Sampling the radio footprint on the ground with an antenna array therefore provides information on the primary shower parameters. There are various radio arrays that observe data from EAS, like AERA [26], and the Low-Frequency-Array (LOFAR) [27], which also serves as a pathfinder experiment for the next-generation radio array SKAO [28].

### 1.2. Radio detection of air showers with LOFAR

The Low-Frequency Array (LOFAR) is a multi-purpose radio telescope with stations distributed across the Netherlands and several other European countries [29]. For cosmic-ray science, the key component is the dense core located in Exloo, the Netherlands. The core consists of a dense center region, the *Superterp*, with six antenna fields within a 320 m radius. With increasing distance there are 18 more core stations around the superterp. Each station contains 96 Low-Band Antennas (LBAs), operating in the 30 to 80 MHz frequency range, and 48 high-band antennas (HBA, 110 to 240 MHz). In the first operational stage of LOFAR, from 2012–2024, only the LBA were used for cosmic-ray measurements as the HBA data was beamformed and because LBA and HBA antennas could not observe simultaneously due to hardware limitations [25,27]. With the upcoming LOFAR2.0 upgrade, the HBA data will be available together with the LBA data, enabling a sampling of the higher frequency footprint [30]. This current work, however, focuses on the 30–80 MHz band exclusively, but could in the future be extended to a broader frequency range.

For air shower detection, LOFAR operates in a triggered mode. An independent, co-located array of scintillator particle detectors, the LOFAR Radboud Air shower Array (LORA) [31], provides an external trigger when an EAS is detected. Upon receiving a trigger, the time-series of raw Analog-To-Digital-Converter (ADC) counts from the LBA are saved for a short time window ( $\sim 2$  ms) around the event time. This captures the brief, nanosecond-scale radio pulse of the air shower.

To convert ADC counts into physically meaningful units, an absolute calibration of the LOFAR antennas using a model of the Galaxy is used [32], which yields a systematic uncertainty of 13%.

The standard method for reconstructing air shower properties with LOFAR [33] (here referred to as *legacy method*) is based on comparing fluence data to CoREAS [34] simulations. The fluence is a measure of energy deposit per area and can be calculated from the electric field via

$$F(\mathbf{r}_{\text{ant}}) = \epsilon_0 c \sum_{\text{pol}} \int_T |E_{\text{pol}}(t, \mathbf{r}_{\text{ant}})|^2 dt, \quad (1)$$

where  $E_{\text{pol}}$  is the electric field trace in each E-field component, which varies as a function of time and antenna position.

In the legacy method, an analogous quantity to the fluence is calculated directly from voltage data of the instrumented measurement channels to prevent the passing of noise through the antenna model. This fluence-like quantity is then compared to simulated voltage data

after passing the simulated electric fields through the forward response of the antenna.

For each measured shower, an ensemble of simulations is produced and their voltage-level fluences interpolated to LOFAR positions. These are then fitted to the measured pulse energies by minimizing a chi-squared function with free parameters for shower core position and an overall fluence scaling factor, and the resulting chi-squared values are plotted against  $X_{\max}$ , with a parabolic fit to the lower envelope yielding the reconstructed  $X_{\max}$ .

Uncertainties on  $X_{\max}$  are determined by adding realistic noise to the simulations and reconstructing them against the ensemble. The primary energy is estimated by applying the square root of the radio signal scaling factor to the simulation energy, given that the initial simulation energy is already close to the reconstructed energy.

The mean uncertainty on  $X_{\max}$  of the method is  $19 \text{ g cm}^{-2}$ , which is the current state-of-the-art [33]. The drawback of this method is its computational demand, as already one of the 60 CoREAS simulations that need to be produced per measured LOFAR event has a runtime of 1 CPU day. In the past year, an approach to decrease the computational load via template synthesis has become available [35], but has not yet been used for a reconstruction study.

Additionally, the legacy method relies solely on fluence data, discarding the information that other parts of the radio data contains. For example, the pulse arrival times also provide information on the  $X_{\max}$  [36,37]. No established reconstruction approach has combined both timing and fluence so far.

Our new approach relies on Bayesian inference, more specifically the Information Field Theory framework that is ideal for reconstructing high-dimensional, noisy data. It not only encompasses more information, but is also faster in reconstructing shower properties than the conventional analysis by relying on a semi-parametric air shower model instead of simulations, while retaining a higher accuracy than simple parametric approaches.

### 1.3. Information field theory

Information Field Theory (IFT) provides a probabilistic framework for reconstructing physical fields by treating them as continuous random variables defined through probability distributions [38]. This Bayesian foundation allows the framework to address high-dimensional inference problems.

The objective of IFT is to infer an unknown signal field  $\phi$  from a set of measurements  $d$ . Bayes' theorem expresses the posterior distribution as

$$P(\phi|d) \propto P(d|\phi) P(\phi),$$

where the likelihood  $P(d|\phi)$  encodes the measurement process — including instrumental response and noise — while the prior  $P(\phi)$  incorporates physical knowledge about the signal's statistical properties, such as its correlation structure.

In many cases, Gaussian statistics provide an adequate description:

$$P(d|\phi) \propto \exp\left(-\frac{1}{2}(d - R(\phi))^{\dagger} N^{-1}(d - R(\phi))\right), \quad (2)$$

$$P(\phi) \propto \exp\left(-\frac{1}{2}\phi^{\dagger} S^{-1}\phi\right), \quad (3)$$

with  $R(\phi)$  denoting the (possibly non-linear) response operator,  $N$  the noise covariance matrix, and  $S$  the signal covariance matrix. In its simplest incarnation, the inference problem reduces to minimizing the information Hamiltonian (which is equivalent to maximum a posteriori (MAP)), defined as the negative log-posterior:

$$\mathcal{H}(\phi|d) = -\ln P(\phi|d).$$

In variational inference (VI) approaches within IFT, the objective is to minimize the variational information distance between a variational distribution  $Q(\phi)$  and the posterior  $P(\phi|d)$ , given by

$$D_{\text{KL}}(Q(\phi) \| P(\phi)) = -\mathbb{E}_Q \left[ \ln \frac{P(\phi|d)}{Q(\phi)} \right] \quad (4)$$

where  $D_{\text{KL}}$  is the Kullback–Leibler divergence and  $\mathbb{E}_Q$  denotes the expectation with respect to  $Q$ . In its VI version (with  $Q$  being the first argument), it measures the amount of spurious information introduced by approximating  $P(\phi|d)$  by  $Q(\phi)$ . This should be as small as possible, hence the minimization [39,40].

The VI approach of IFT enables the quantification of uncertainty by characterizing the posterior distribution rather than providing only point estimates, as given for example by MAP based methods. In non-linear settings, the posterior is typically non-Gaussian, preventing exact computation. To address this, approximation techniques based on VI have been developed within the IFT framework, such as Metric Gaussian Variational Inference (MGVI) [39]. MGVI approximates the posterior with a Gaussian distribution that depends only on a posterior mean value for the quantities that are to be inferred, while the uncertainty covariance is expressed in terms of this mean with the help of the Fisher information metric. This improves computational efficiency compared to standard sampling methods. This efficiency is necessary for processing large datasets from radio observatories such as LOFAR and the SKA.

To address the limitations of the Gaussian approximation used by MGVI, Geometric Variational Inference (geoVI) [40] employs normalizing flows to learn a non-linear mapping from a simple base distribution to the target distribution. This enables the variational family to approximately represent non-Gaussian posteriors and represent intricate degeneracies, providing more detailed uncertainty quantification at the cost of increased computational demand.

In this work, we present a full air shower reconstruction method implemented using the Numerical Information Field Theory (NIFTy) framework in Python [41–44]. NIFTy utilizes Google’s JAX library [45] for automatic differentiation and just-in-time compilation, enabling GPU acceleration for efficient data processing.

## 2. The differentiable forward model

Our method is centered around a Bayesian inference engine that connects a detailed physical forward model of the air shower signal to the observed data. A key innovation of our work is the development of a *fully differentiable* forward model, which allows us to leverage powerful gradient-based inference techniques.

The forward model,  $F(\vec{\lambda})$ , maps a set of physical air shower parameters,  $\vec{\lambda}$ , to the predicted data observables. It is constructed as a chain of modular, differentiable operations implemented in JAX.

### 2.1. Coordinate systems and transformations

The physics of the emission is most naturally described in the **shower-plane coordinate system**. This is a right-handed system with its origin at the shower core (i.e., impact) position on the ground,  $(X_{\text{core}}, Y_{\text{core}})$ , and its  $\hat{z}'$ -axis pointing upwards along the shower propagation direction  $\mathbf{v}$ . The transverse axes  $(\hat{x}', \hat{y}')$  are aligned with the polarization vector of the geomagnetic  $(\mathbf{v} \times \mathbf{B})$  emission and the direction perpendicular to it  $(\mathbf{v} \times (\mathbf{v} \times \mathbf{B}))$ . The parameterizations of our model are all described in this coordinate frame, however the forward model itself contains coordinate transformations such that its output is always in the ground plane  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$  (cartesian coordinates, with  $x$  pointing east) for a given arrival direction  $(\theta, \phi)$  and core position. This transformation enables using ground plane data for the reconstruction, and embeds the shower-geometry-dependent coordinate transformation into the optimization, enhancing core and arrival direction precision and minimizing the geometry assumptions (i.e. known arrival direction) that need to be made pre-reconstruction.

### 2.2. Fluence model: The ‘geoceLDF’ parametrization

We base our fluence prediction on the semi-analytical ‘geoceLDF’ model [19], which provides a physically motivated Lateral Distribution Function (LDF) by combining the two primary emission mechanisms (Section 1.1) in superposition. The fluence quantity we use is calculated from the electric fields in  $\text{eV}/\text{m}^2$ , not the voltage-level fluence previous analyses LOFAR used.

The parameterization is most accurate for showers with  $\theta < 60^\circ$ , which is ideal for the dense LOFAR core. It models both the geomagnetic and the charge excess effect contributions in the 30 to 80 MHz band separately, along with their dependence on the shower parameters  $E_{\text{rad}}$ , the slant distance  $D_{X_{\text{max}}}$  and the shower direction  $\theta, \phi$ . Site-specific dependencies such as the magnetic field vector or the atmospheric model and observational height can be defined as well, making the model instrument-independent.

The total fluence at any position  $\mathbf{r} = \begin{pmatrix} r \\ \varphi_{\text{obs}} \end{pmatrix}$  in the shower plane is a superposition that depends on the observer’s polar angle relative to the shower core  $\varphi_{\text{obs}}$ :

$$F(\mathbf{r}) = F_{\text{geo}}(r) + F_{\text{ce}}(r) + 2\sqrt{F_{\text{geo}}(r)F_{\text{ce}}(r)} \cos(\varphi_{\text{obs}}) \quad (5)$$

Through this dependence on observer angle, the interference caused by the different polarization patterns of the emission mechanisms can be caught.

We have implemented two critical enhancements to this model. First, we have extended it to directly accept the absolute atmospheric depth of the shower maximum,  $X_{\text{max}}$  (in  $\text{g}/\text{cm}^2$ ), as a primary input parameter. The  $X_{\text{max}}$  and  $D_{X_{\text{max}}}$  can, for less inclined showers ( $\theta < 60^\circ$ ) be related via

$$D_{X_{\text{max}}} = \frac{X_{\text{atm}}}{\cos(\theta)} - X_{\text{max}}, \quad (6)$$

where  $X_{\text{atm}}$  is the vertical atmospheric depth at ground level, which depends on the atmospheric model that can be chosen based on the site’s atmospheric conditions. For real events, a unique atmospheric profile of the event time can be imported. This is a crucial modification, as  $X_{\text{max}}$  is the key physical observable for mass composition studies, and an inaccurate atmosphere model can easily introduce biases of  $\mathcal{O}(10\%)$  [46].

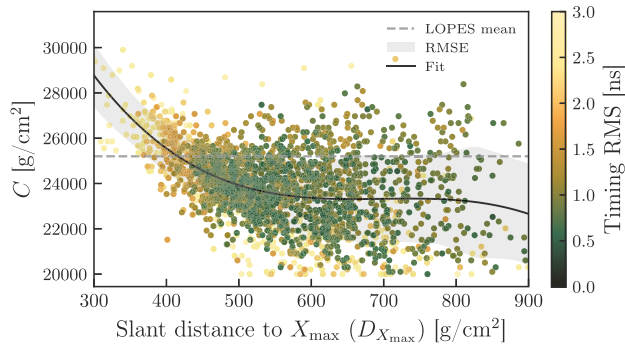
Secondly, the entire model was re-implemented from the ground up to be end-to-end differentiable. This required a complete rewrite of the source code, but enables IFT’s powerful gradient-based inference techniques.

### 2.3. Arrival time model: Hyperbolic wavefront

The temporal structure of the radio signal’s arrival is modeled primarily using a hyperbolic wavefront, as measured by LOFAR [36], which accounts for the refractive index of the atmosphere and the relativistic movement of the shower. The geometric arrival time, i.e. the time of the maximum of the Hilbert envelope of the total electric field,  $\tau_{\text{geo},i}$  at antenna  $i$  is given by the parameterization derived for LOPES [37,47]:

$$\tau_{\text{geo},i} = \frac{1}{c} \left( \sqrt{(d_i \sin \rho)^2 + (c \cdot b_{\text{offset}})^2} + z'_i \cos \rho \right) + t_0. \quad (7)$$

Here,  $d_i = \sqrt{x_i'^2 + y_i'^2}$  is the lateral distance from the shower axis, and  $t_0$  is a global time offset. The parameter  $\rho$  represents the cone opening angle, and  $b_{\text{offset}}$  is fixed at 3 ns. The  $X_{\text{max}}$  dependence of the cone is described via



**Fig. 2.** Dependence of the optimal timing parameter  $C$  on the slant distance to the shower maximum,  $D_{X_{\max}}$ . Each point represents an individual CoREAS simulation, where the signal times were determined via the Hilbert envelope maximum of the total electric field trace, colored by the root-mean-square (RMS) of the timing residuals for the best fit. The solid black line shows the fit using a 3rd-degree polynomial (Table 1), with the shaded region indicating the root-mean-square (RMS) deviations from the fit. The dashed gray line indicates the fixed value used by LOPES,  $C = 25200$  g/cm<sup>2</sup>.

**Table 1**

Coefficients of the third-degree polynomial fit  $C(x) = p_3x^3 + p_2x^2 + p_1x + p_0$ , where  $x = D_{X_{\max}}$  in g/cm<sup>2</sup>.

Polynomial Term	Coefficient Value
$p_3 \cdot x^3$	$-8.955 \times 10^{-5} (\text{g/cm}^2)^{-2}$
$p_2 \cdot x^2$	$1.873 \times 10^{-1} (\text{g/cm}^2)^{-1}$
$p_1 \cdot x^1$	$-1.302 \times 10^2$
$p_0$	$5.338 \times 10^4 \text{ g cm}^{-2}$

$$X_{\max} = C \cdot \rho \cdot \cos^{-\gamma} \theta, \quad (8)$$

where  $\gamma, C$  were fixed constants. We reparameterized these two constants on LOFAR simulations and found  $\gamma = 1.465 \pm 0.292$ . For the constant  $C$  we found a correlation with the slant distance to  $X_{\max}$ , shown in Fig. 2, to which we fit a third degree polynomial (see Table 1).

We note that the RMS in Fig. 2 indicates the goodness of fit decreases for smaller slant distances. This is likely due to the shower being closer to the ground.

To account for potential tensions between the shower geometry preferred by the signal fluence and the wavefront timing we introduce a decoupling parameter,  $\Delta X_{\max}^{\text{time}}$ . While the fluence model (Section 2.2) depends directly on the primary  $X_{\max}$ , the wavefront parametrization utilizes an effective depth  $X_{\max}^{\text{time}} = X_{\max} + \Delta X_{\max}^{\text{time}}$ . This nuisance parameter is constrained by a Gaussian prior with  $\mu = 0$  and  $\sigma = 15$  g/cm<sup>2</sup>, allowing the inference to resolve minor inconsistencies between amplitude and timing data. This stabilizes the fit but we do not observe a significant bias introduced by it.

#### 2.4. Particle model: NKG function

For completeness, our framework also includes a differentiable model for the ground-level particle density, based on the well-established Nishimura–Kamata–Greisen (NKG) function [48,49]. This model predicts the lateral distribution of secondary particles as a function of shower parameters.

We note that the particle signal is significantly more complex to describe than the radio model due to stronger event-by-event fluctuations. In the current implementation, we find that including the particle data does not improve the reconstruction precision for key observables like  $X_{\max}$  or core position compared to a radio-only analysis, and causes tension. Consequently, the particle component is **not used in the primary reconstruction** presented in this work.

However, the model is retained for two important reasons: (1) it allows verification that the LORA particle detector array would have triggered on any given event, serving as an independent cross-check; and (2) it provides a foundation for future combined radio-particle analyses.

The full particle model is described in Appendix C.

#### 2.5. Modeling systematics with correlated fields

Parametric models as described in Sections 2.2 and 2.3 are idealized representations of the complex statistical processes involved in an air shower. Based on simulations, they describe the idealized or at best average case of an air shower. In reality, the particle physics processes in an air shower undergo fluctuations, local atmospheric conditions [46] or clouds [50] impact the propagation of radio signals, and a perfect detector performance is never reached.

Standard least-squares fitting treats all deviations as uncorrelated white noise. However, many systematic effects are spatially correlated. By modeling correlation explicitly, we prevent the solver of the inference problem from treating systematic physical effects as random instrument noise.

Lastly, a purely parametric model is often too rigid (under-fitting), forcing physical parameters such as the shower maximum or core position to bias themselves to compensate for data that does not fit the ideal shape, leading to large  $\chi^2$ -values and inaccurate posterior distributions. Adding correlated deviations to the model adds thousands of potential degrees of freedom (one for every point in space), but tames their behavior via a prior that forces them to be a smooth function. This allows the model to bend where necessary to fit the data, without breaking the global parametric model.

IFT offers the modeling of correlations via Correlated Fields [51]. Correlated Fields do not treat the systematic effects as fixed parameters, but as continuous random fields with a Gaussian Process prior. In harmonic space, the power spectrum of the field we use is defined using a power law  $P(k) \propto k^{-\alpha}$ . A higher spectral index suppresses high-frequency modes, causing the field to be smoother and more correlated on large scales. A spectral index of 0 would correspond to uncorrelated white noise.

To maintain differentiability in JAX, the field is not sampled directly, but modeled via a generative process.

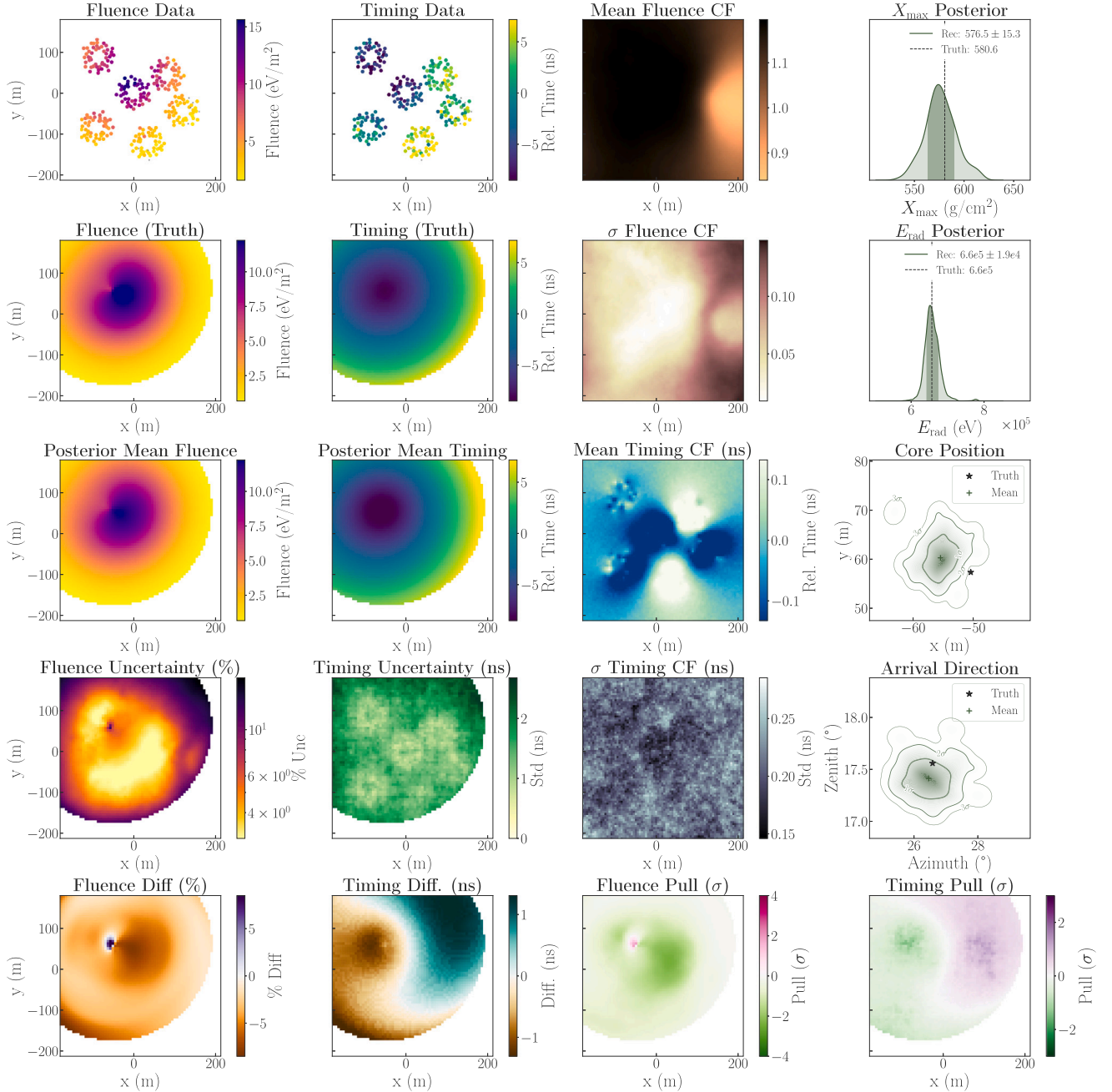
By including these fields in the Bayesian inference, we effectively marginalize over the systematic uncertainties. The posterior distribution of the physical parameters (core, energy) widens, leading to more truthful and robust error estimates.

We introduce two Correlated Fields to our model, one acting on the fluence and one acting on the timing. As fluence variations scale with the signal strength, we apply the fluence Correlated Field multiplicatively, i.e. allowing it to alter each fluence point by a few percent. This is best achieved via an exponential factor  $s_{\text{sys}} = \exp(s'_{\text{sys}})$ , where the fluctuations of  $s'_{\text{sys}}$  are modeled as normally distributed around zero. That way, if  $s'_{\text{sys}} = 0$  the Correlated Field becomes = 1 and does not affect the fluence. Furthermore, we assign a spectral index of the field of  $\alpha = -4.5 \pm 0.5$ , favoring larger scale correlations.

The timing Correlated Field  $\delta\tau_{\text{sys}}$  acts additively on the wavefront model. We define zero-centered, normally distributed fluctuations with  $\sigma = 1.5$  ns, and a spectral index of  $\alpha = -1 \pm 0.5$ , allowing smaller-scale variations.

Examples of posterior Correlated Fields are shown in Fig. 3 (third column from the left), where they are plotted exactly as they act on the parameterized models.

We apply nonlinearities (clipping) to the fields. This is critical when mixing physical and non-parametric models. It ensures the “flexible” field can correct local shapes but is mathematically forbidden from drifting so far that it replaces the parametric relations (e.g., the field cannot simply “create” the air shower signal from scratch, it can only modulate the physical fluence prediction by  $\pm 20\%$  and the timing prediction by  $\pm 15$  ns).



**Fig. 3.** Reconstruction summary for a simulated air shower event, an iron primary with an energy of  $2.44 \times 10^{17}$  eV. The first two columns display the ground distributions of fluence and arrival timing (with the plane-wave component subtracted to better emphasize the hyperbolic shower front), showing (top to bottom) input data at LOFAR positions, CoREAS truth, reconstructed Posterior means, relative uncertainties, and residuals. The third column maps the inferred spatially Correlated Fields (CF). The fourth column presents the marginal posterior distributions for the depth of shower maximum  $X_{\max}$ , radiation energy ( $E_{\text{rad}}$ ), core position ( $X_0, Y_0$ ), and arrival direction ( $\theta, \phi$ ), with ground truth values indicated. The last row compares obtained differences with the corresponding pull.

## 2.6. The full model

Combining the physical parameterizations with the non-parametric systematic fields for both fluence and timing, the full forward model for the observables at antenna  $i$  is:

$$\mathcal{F}_i(\vec{\lambda}, s_{\text{sys}}) = \left[ \mathcal{F}_i(\vec{\lambda}_{\text{shower}}) \cdot \exp(s'_{\text{sys},i}) \right], \quad (9)$$

$$\tau_i(\vec{\lambda}, \delta\tau_{\text{sys}}) = \tau_{\text{geo},i}(\vec{\lambda}_{\text{time}}) + \delta\tau_{\text{sys},i}. \quad (10)$$

The parameter vector that is to be inferred now includes the physical shower parameters  $\vec{\lambda}$  and the latent variables defining both the

fluence correlated field  $s_{\text{sys}}$  and the timing correlated field  $\delta\tau_{\text{sys}}$ . Correlated fields can be turned “off” and “on” separately, also enabling a fully parametric forward model on demand.

**Fig. 3** is an example reconstruction using this model. This visualizes the components involved in the forward model. For example, it shows the access to the parameter distributions ( $X_{\max}$ ,  $E_{\text{rad}}$ , arrival direction, core position) along with the spatial fluence and timing distributions. In principle, one can access the approximation of the full fluence or timing posterior distribution of each point in space, as well as the Correlated Fields at each point in space. This is difficult to visualize, so instead we

show the mean and standard deviation of the posterior samples at each point.

The model is now semi-parametric. The fluence and timing parameterizations are deterministic, meaning with the same set of shower parameters the model will always return identical values. The Correlated Fields are what adds randomness to the model, allowing for small deviations given the same parameter set.

The implementation of this entire chain — from the primary shower parameters, the parametric and non parametric components to the clipping operations — is done in the differentiable JAX framework. This allows us to calculate the analytical gradient of the posterior probability,  $\nabla_{\vec{\theta}} \ln P(\vec{\theta}|d)$ , unlocking efficient inference algorithms that can navigate this high-dimensional space without becoming trapped in local minima.

### 3. Bayesian inference and parameter estimation

Having established the generative forward model in Section 2.6, we now turn to the inverse problem: reconstructing the set of physical and systematic parameters,  $\vec{\theta}$ , given the observed radio measurements. We adopt a Bayesian framework to quantify the posterior probability distribution,

$$P(\vec{\theta} | d) \propto \mathcal{L}(\vec{\theta}) \pi(\vec{\theta}), \quad (11)$$

where  $\mathcal{L}(\vec{\theta}) := P(d|\vec{\theta})$  is the likelihood of the data and  $\pi(\vec{\theta}) := P(\vec{\theta})$  represents the prior knowledge encoded in the model structure.

This section details the three main components of our inference strategy. First, we define the likelihood function, which accounts for the specific noise properties of the radio antennas and the masking of unreliable timing data. Second, we describe the numerical optimization scheme used to approximate the high-dimensional posterior via VI. Finally, we report on the computational cost and efficiency of this reconstruction pipeline compared to legacy simulation-matching approaches.

#### 3.1. The likelihood and noise model

The observed data vector for a single event consists of two distinct measurement types collected at  $N_{\text{ant}}$  antennas:

- Fluence:  $\mathcal{F}_i^{\text{obs}}$  (in  $\text{eV m}^{-2}$ ),  $i = 1, \dots, N_{\text{ant}}$
- Arrival-time:  $\tau_i^{\text{obs}}$  (in seconds),  $i = 1, \dots, N_{\text{ant}}$

Only a subset of antennas provide reliable timing information (after outlier rejection, see Section 4.2). We therefore introduce a binary mask vector  $\mathbf{m} \in \{0, 1\}^{N_{\text{ant}}}$  such that  $m_i = 1$  if antenna  $i$  contributes a timing measurement and  $m_i = 0$  otherwise.

The forward model presented in Section 2.6 predicts, for any parameter configuration  $\vec{\theta}$ , the corresponding fluence and arrival time at every antenna position:

$$\mathcal{F}^{\text{model}}(\vec{\theta}) = \mathcal{F}_{\text{phys}}(\vec{\lambda}_{\text{shower}}) \odot \exp(\mathbf{s}'_{\text{sys}}), \quad (12)$$

$$\boldsymbol{\tau}^{\text{model}}(\vec{\theta}) = \boldsymbol{\tau}_{\text{geo}}(\vec{\lambda}_{\text{time}}) + \boldsymbol{\delta}\boldsymbol{\tau}_{\text{sys}}, \quad (13)$$

where  $\odot$  denotes element-wise multiplication,  $\mathbf{s}'_{\text{sys}}$  is the interpolated and clipped fluence systematic field, and  $\boldsymbol{\delta}\boldsymbol{\tau}_{\text{sys}}$  is the interpolated and clipped timing systematic field.

We model the measurement errors as independent Gaussian noises with *known* (data-driven) variances. The full likelihood is therefore

$$\mathcal{L}(\vec{\theta}) \equiv P(\mathcal{F}^{\text{obs}}, \boldsymbol{\tau}^{\text{obs}} | \vec{\theta}) = \mathcal{L}_{\mathcal{F}}(\vec{\theta}) \mathcal{L}_{\tau}(\vec{\theta}), \quad (14)$$

with the two decoupled contributions

$$\mathcal{L}_{\mathcal{F}}(\vec{\theta}) = \prod_{i=1}^{N_{\text{ant}}} \mathcal{N}(\mathcal{F}_i^{\text{obs}} | \mathcal{F}_i^{\text{model}}(\vec{\theta}), \sigma_{\mathcal{F},i}^2) \quad (15)$$

$$\mathcal{L}_{\tau}(\vec{\theta}) = \prod_{i=1}^{N_{\text{ant}}} \left[ \mathcal{N}(\tau_i^{\text{obs}} | \tau_i^{\text{model}}(\vec{\theta}), \sigma_{\tau,i}^2)^{m_i} \times \mathcal{N}(\tau_i^{\text{obs}} | 0, 10^{24} \text{ s}^2)^{1-m_i} \right]. \quad (16)$$

The second factor in Eq. (16) is a technical addition: for antennas that do *not* contribute timing information ( $m_i = 0$ ), we impose an extremely weak constraint (variance  $10^{24} \text{ s}^2$ ) centered on zero. This is numerically equivalent to ignoring those data points while still keeping the likelihood well-defined over the full  $2N_{\text{ant}}$ -dimensional data space.

In matrix form the negative log-likelihood (up to an irrelevant constant) used by the IFT optimizer is

$$-\ln \mathcal{L}(\vec{\theta}) = \frac{1}{2} (\mathbf{d} - \mathbf{h}(\vec{\theta}))^{\top} \mathbf{N}^{-1} (\mathbf{d} - \mathbf{h}(\vec{\theta})), \quad (17)$$

where the concatenated data and model vectors are

$$\mathbf{d} = \begin{pmatrix} \mathcal{F}^{\text{obs}} \\ \boldsymbol{\tau}^{\text{obs}} \end{pmatrix} \in \mathbb{R}^{2N_{\text{ant}}}, \quad \mathbf{h}(\vec{\theta}) = \begin{pmatrix} \mathcal{F}^{\text{model}}(\vec{\theta}) \\ \boldsymbol{\tau}^{\text{model}}(\vec{\theta}) \end{pmatrix}. \quad (18)$$

The noise covariance matrix  $\mathbf{N}$  is approximated as diagonal, which we find sufficient for our case as the off-diagonal elements are generally small [52]:

$$\mathbf{N} = \text{diag}(\sigma_{\mathcal{F},1}^2, \dots, \sigma_{\mathcal{F},N_{\text{ant}}}^2, \sigma_{\tau,1}^2, \dots, \sigma_{\tau,N_{\text{ant}}}^2), \quad (19)$$

so that its inverse, the precision matrix, is simply

$$\mathbf{N}^{-1} = \text{diag}(1/\sigma_{\mathcal{F},1}^2, \dots, 1/\sigma_{\mathcal{F},N_{\text{ant}}}^2, 1/\sigma_{\tau,1}^2, \dots, 1/\sigma_{\tau,N_{\text{ant}}}^2). \quad (20)$$

Because the noise covariance is diagonal and the forward model is implemented with JAX, which is equipped with autodifferentiation, the full gradient  $\nabla_{\vec{\theta}}(-\ln \mathcal{L})$  is analytically exact and extremely cheap to evaluate, which is crucial for efficient sampling in the high-dimensional space spanned by the two correlated fields (approximately 180 000 latent amplitudes in total for the  $300 \times 300$  grid).

#### 3.2. Gradient-based variational inference

The posterior distribution  $P(\vec{\theta}|d)$  is a high-dimensional function that cannot be computed analytically. We therefore use VI to find a tractable, approximate posterior distribution  $Q(\vec{\theta})$ . This is achieved by minimizing the Kullback–Leibler (KL) divergence between  $Q$  and  $P$ .

We perform this optimization using the NIFTy library [41–44]. We utilize NIFTy's `optimize_kl` routine, which implements advanced VI methods such as `geoVI` [40]. To ensure robust convergence, our inference strategy employs a multi-stage approach. We begin with a resampling-based optimization (`nonlinear_resample`) to explore the parameter space globally and avoid local minima, before switching to a gradient-based update mode (`nonlinear_sample`) to refine the solution. The result is a multi-dimensional approximate posterior,  $Q(\vec{\theta})$ , whose mean provides the maximum a posteriori point estimates and whose covariance matrix provides a full quantification of parameter uncertainties and correlations.

#### 3.3. Computational demand

The reconstruction framework presented here offers a substantial improvement in computational efficiency compared to standard simulation-matching techniques.

A typical reconstruction requires an average wall-clock time of  $48 \pm 13$  minutes. The peak memory usage remains stable across events at approximately 6.2 GiB. Both values depend on the number of data points, and on whether correlated fields are activated or not. This efficiency is a direct consequence of the differentiability of the forward model, which allows the optimizer to navigate the high-dimensional parameter space using exact gradients rather than brute-force sampling.

In contrast, the legacy reconstruction method relies on  $\sim 60$  individual simulations per measured event. With the generation of a

single detailed simulation taking roughly 1 CPU day, the computational demand of the legacy reconstruction, based on generating the simulations alone, amounts to approximately 1440 CPU hours per event, not including the demand of the consequent fitting to the data.

Therefore, the method proposed in this work reduces the computational cost by three orders of magnitude—from weeks of CPU time to approximately 1 CPU hour per event. This renders the detailed reconstruction of large datasets feasible on standard computing clusters without requiring massive high-performance computing allocations, while retaining accuracy superior to simple parametric reconstructions.

#### 4. Performance on simulations

The first test of this framework is performed on CoREAS simulations. This is the best way to compare the reconstructed shower properties to the truth and fully understand the methods biases and limitations.

##### 4.1. Generation of realistic data

The simulations used are a pre-existing set for LOFAR conditions, with primary energies between  $10^{16.5}$  eV and  $10^{18.5}$  eV and  $X_{\max}$  values between 500 and 1000  $\text{g cm}^{-2}$ . The set contains both iron and proton primaries and used the hadronic interaction model QGSJetII-04 [53]. We are limiting our dataset to near-vertical simulations with a zenith angle  $< 45^\circ$ , consistent with the geometry of typical LOFAR events. Showers at higher inclinations produce larger radio footprints where the energy is distributed over a larger area, causing signals to drop below detection threshold. Additionally, only a small amount of particles reach sea-level at high zenith angles, resulting in LORA not triggering for these events, resulting in LOFAR datasets being dominated by lower zenith angles.

For convenience we use the NuRadioReco software [54,55] to read in CoREAS simulations and interpolate their electric field traces to LOFAR positions using Fourier-based interpolation [56], assigning a random core from  $\mathcal{N}(0|100)$  m, where 0 m is the center of the LOFAR superterp. This core distribution was chosen as it matches the distribution of cores in the LOFAR mass composition analysis [57].

We then converted the interpolated electric fields to voltages using the LOFAR antenna response. In order to more realistically simulate the behavior of real data with simulations, we added measured noise from LOFAR data at the voltage level to the traces. Previous analyses used the voltages directly to calculate a fluence-like quantity to avoid passing noise through the antenna response. However, as our forward model expects the fluence in units of  $\text{eV m}^{-2}$ , we need the electric-field fluence. We thus add measured voltage noise at this stage to simulate the passing of noise through the antenna response.

For all subsequent steps of the reconstruction, the true arrival direction and core position of the simulation is randomized slightly within the known LORA precision [31] ( $\sigma_{\theta,\phi} = 0.7^\circ$  and  $\sigma_{X_C,Y_C} = 6$  m for cores within 150 m of the LORA center, and a more generous  $\sigma_{\theta,\phi} = 2^\circ$  and  $\sigma_{X_C,Y_C} = 20$  m for cores further out), to simulate that in real data we would not know the true geometry either.

We want to focus only on events that would have realistically been detected by LOFAR. For previous analyses, only simulations where more than 50% of antennas in at least 3 stations had an energy signal-to-noise ratio  $\left(\frac{f_{\text{signal}}^{55 \text{ ns}}}{f_{\text{noise}}^{55 \text{ ns}}}\right)$  of  $> 6$  were included [57]. This was a conservative filter to prevent composition bias, which we will loosen to require only 1 station to fulfill this check. We thus allow both smaller and fainter footprints in our analysis, widening the prospective real dataset. We note that nevertheless, for a future mass composition analysis using this method, further tests against composition bias will need to be implemented.

##### 4.2. Pulse finding

If an event is considered *triggered* (i.e. at least one station has more than 50% of antennas with an energy  $\text{SNR} > 6$ ), it enters the pulse finding stage. For this, the voltage traces are first converted into electric fields via

$$\begin{pmatrix} V_1(f) \\ V_2(f) \end{pmatrix} = \begin{pmatrix} \mathcal{H}_1^\theta(f) & \mathcal{H}_1^\phi(f) \\ \mathcal{H}_2^\theta(f) & \mathcal{H}_2^\phi(f) \end{pmatrix} \begin{pmatrix} E^\theta(f) \\ E^\phi(f) \end{pmatrix}, \quad (21)$$

where  $\mathcal{H}$  is the antenna response,  $V$  are the voltages from the two antenna polarizations, using the arrival direction guess and solving for the electric field components  $E$ .

Finding the exact pulse location is essential for a successful reconstruction, however especially at low SNR, pulse finding proves challenging.

Before analyzing individual antennas, we perform a robust “incoherent beamforming” sum of all antennas per station to find a global timestamp estimate ( $t_{\text{BF}}$ ), which will then define the region of interest (ROI). This is not performed globally as it assumes a plane wavefront, which suffices as approximation only for smaller subsections of the footprint.

For arrival direction  $(\theta, \phi)$ , the plane-wave arrival time delay  $\tau_i$  for antenna  $i$  at position  $(x_i, y_i)$  relative to the station center  $(x_0, y_0)$  is

$$\tau_i = \frac{1}{c} \left( \sin \theta \cos \phi \cdot (x_i - x_0) + \sin \theta \sin \phi \cdot (y_i - y_0) \right). \quad (22)$$

We calculate the magnitude of the sum of electric-field components of each trace and apply a Gaussian smoothing ( $\sigma \approx 10$  ns) to create a broad pulse. This is to ensure that even if the arrival direction guess is slightly off, the pulses from different antennas still overlap when summed.

The smoothed envelopes are then shifted by their geometric delays and averaged:

$$S_{\text{beam}}(t) = \frac{1}{N_{\text{ant}}} \sum_{i=1}^{N_{\text{ant}}} E_{i,\text{smooth}}(t - \tau_i) \quad (23)$$

The time of the resulting maximum is then used as the center of the ROI for the pulse search in individual antennas. In this region of interest, the maximum of the Hilbert envelope of the trace is calculated (on the non-smoothed trace). This is the simplest way to find pulses, and is successful for most signals.

Next, we perform a cleaning via a local outlier rejection. Each antenna is compared to its 10 nearest neighbors by calculating the median and the standard deviation of all their timing values. If the antenna timing deviates by  $2.5\sigma$  from the median value, the timing value is given no weight in the reconstruction, as described in Section 3. This pre-processing of the timing data is necessary as any strong deviations from the true timing values will massively impact the reconstruction performance, which can happen in low-SNR signals.

This is not the case for fluence data. For very low SNR, where the timing guess might be hundreds of nanoseconds off, the calculated fluence value will not differ much from its true counterpart, as noise dominates here. The low-signal antennas are important to include in the data, as they provide a valuable upper limit for the footprint extent.

Thus, for all antennas in a triggered station, the fluence is calculated via Eq. (1) in a time window of 300 ns around the determined pulse maximum. This ensures that the entire signal is well within the fluence calculation window, even if the determined pulse time is slightly off. A window of 300 ns is larger than usually used in analyses and results in a larger noise fraction in the fluence, however we find the advantages of having more of the signal in the window outweigh the disadvantage of a larger noise fraction. A smaller signal window causes the fluence to be underestimated which propagates into an energy underestimation in the reconstruction.

#### 4.2.1. Modeling of fluence noise $\sigma_{F,i}$

We do not perform any noise subtraction but pass the raw, noisy fluence as data into the optimization. Accordingly, we add a log-normal noise scale prior  $\sigma_{\text{noise}}$  to the fluence forward model, which is calculated as the mean fluence noise from multiple non-overlapping noise windows (see Table 2).

Furthermore, for every antenna  $i$  the fluence uncertainty is taken to be the quadratic sum of two independent contributions:

$$\sigma_{F,i}^2 = \sigma_{F,i}^{\text{stat}2} + \sigma_{F,i}^{\text{sys}2} = \sigma_{F,i}^{\text{stat}2} + (0.1 \cdot \mathcal{F}_i^{\text{obs}})^2. \quad (24)$$

The statistical term  $\sigma_{F,i}^{\text{stat}}$  is estimated from the standard deviation of multiple non-overlapping noise windows. The 10% systematic uncertainty is added in quadrature and dominates at high signal-to-noise ratios. The systematic uncertainty arises from both the absolute calibration of the antennas, and also the finite sampling in the time domain.

#### 4.2.2. Modeling of timing noise $\sigma_{\tau,i}$

The timing noise is more complicated, as it cannot directly be calculated from data and depends on the signal strength of the event. For high SNR events, the time of the maximum signal will usually be within 5 ns (sampling time of LOFAR) of the true signal time, but for lower SNR events, due to the noise, the maximum might jump back or forth a few samples. Using the same timing noise estimate for all data points is not an option, so we need to estimate the timing noise on a per-antenna basis using local clusters.

A global plane-wave fit is insufficient because its residuals combine systematic error from the wavefront curvature with the true random noise, which would incorrectly inflate the uncertainty estimate and reduce sensitivity to  $X_{\text{max}}$ . To address this, we perform a polynomial fit per antenna that includes a radial curvature term, capturing both the plane-wave delay and the wavefront curvature. For each antenna  $i$ , we define a local cluster  $C_i$  consisting of its  $k = 20$  nearest neighbors. Positions within this cluster are centered and normalized for numerical stability:

$$\tilde{x}_j = \frac{x_j - \bar{x}_{C_i}}{\sigma_r}, \quad \tilde{y}_j = \frac{y_j - \bar{y}_{C_i}}{\sigma_r}, \quad \tilde{r}_j^2 = \tilde{x}_j^2 + \tilde{y}_j^2, \quad (25)$$

where  $\sigma_r$  is the standard deviation of radial distances within the cluster. A least-squares fit is performed using the design matrix  $\mathbf{A} = [1, \tilde{x}, \tilde{y}, \tilde{r}^2]$ , solving for coefficients  $\vec{c}$  such that  $t_j \approx \mathbf{A}_j \cdot \vec{c}$ . The residuals  $\epsilon_j = t_j - \mathbf{A}_j \cdot \vec{c}$  represent pure measurement noise, free from systematic curvature contributions. The per-antenna timing uncertainty is then:

$$\sigma_{\tau,i} = \max\left(\text{std}(\{\epsilon_j\}_{j \in C_i}), \sigma_{\tau}^{\min}\right), \quad (26)$$

where  $\sigma_{\tau}^{\min} = 1.5$  ns is a floor value to prevent unrealistically small uncertainties. If a cluster has fewer than 8 antennas, a fallback uncertainty of 5 ns is used.

For antennas that pass all quality cuts ( $m_i = 1$ ), this per-antenna uncertainty is used. For antennas that are excluded from the timing fit ( $m_i = 0$ ), the effective uncertainty is set to  $\sigma_{\tau,i} = 10^{12}$  s (see Eq. (16)).

#### 4.2.3. Estimation of the time offset $t_0$

The global time offset parameter  $t_0$  (Eq. (7)), essentially the time at the shower core, requires a specific prior derivation. While the geometric parameters are centered on the LORA reconstruction,  $t_0$  is derived from the radio timing data itself. We utilize the spatial polynomial fit described in Section 4.2.2 and extrapolate it to the estimated core position. The value of this fit at the core defines the mean of the  $t_0$  prior ( $\mu_{t_0}$ ), while the prior width ( $\sigma_{t_0}$ ) is derived from the covariance of the fit combined with a distance-dependent uncertainty term to account for extrapolation errors.

This ensures that the  $t_0$  prior does not prevent the optimizer from being able to fit the timing data by being too narrow, which is especially important for those events with cores outside of the LOFAR superterp, where the  $t_0$  estimate is usually less accurate.

**Table 2**

Prior distributions for the scalar shower parameters.  $\mathcal{N}(\mu, \sigma)$  denotes a Normal distribution,  $\mathcal{LN}(\mu, \sigma)$  denotes a Log-Normal distribution, and  $\mathcal{U}(a, b)$  denotes a Uniform distribution. The geometric priors are dynamically centered on the initial guess ( $\phi_{\text{init}}, \theta_{\text{init}}$ ) and particle core ( $x_{c,\text{init}}$ ).

Parameter	Symbol	Prior Distribution
Radiation energy	$E_{\text{rad}}$	$\mathcal{LN}(\ln(10^7), 3.0)$ [eV]
Shower Azimuth	$\phi$	$\mathcal{U}(\phi_{\text{init}} - 3^\circ, \phi_{\text{init}} + 3^\circ)$
Shower Zenith	$\theta$	$\mathcal{U}(\theta_{\text{init}} - 3^\circ, \theta_{\text{init}} + 3^\circ)$
Core Position (X)	$X_c$	$\mathcal{U}(X_{c,\text{init}} - w, X_{c,\text{init}} + w)^a$
Core Position (Y)	$Y_c$	$\mathcal{U}(Y_{c,\text{init}} - w, Y_{c,\text{init}} + w)^a$
Depth of Max.	$X_{\text{max}}$	$\mathcal{U}(450, 1020)$ [g/cm <sup>2</sup> ]
Time Offset	$t_0$	$\mathcal{N}(\mu_{t_0}, \sigma_{t_0})$
Noise Scale	$\sigma_{\text{noise}}$	$\mathcal{LN}(\mu_{\text{data}}, 0.1 \cdot \mu_{\text{data}})$

<sup>a</sup> The prior width for the core position  $w$  is based on the core location  $d_c$ . Within 150 m of the LORA center, we can assume the resolution of the core guess is very precise [31], while outside this radius the resolution is worse. Hence, we use  $w_{d_c \leq 150\text{m}} = 30$  m and  $w_{d_c > 150\text{m}} = 60$  m.

#### 4.2.4. Prior distributions

We apply constraints on the hyperparameter prior distributions as displayed in Table 2. The radiation energy  $E_{\text{rad}}$  is sampled in logarithmic space, while geometric parameters are constrained by a uniform prior centered around the initial reconstruction provided by the LORA particle detector array.

On simulations, we found the best reconstruction results when deactivating the fluence Correlated Field (Section 2.5), as spatial deviations on the fluence can cause biases in  $X_{\text{max}}$  by distorting the footprint shape. It might be beneficial to re-activate it for reconstructing data, and making the field more noise-like by lowering the fields spectral index. This could help with per-antenna calibration or hardware differences. As the impact on reconstruction bias is complex, this would need further investigation before using it, thus we only show the fluence Correlated Fields for example events, but do not use them for the simulation or data sets we reconstruct in this work.

We use a correlated field on timing data, with a spectral index of  $\alpha = -1 \pm 0.5$ , allowing spatial correlations over smaller distances, but capping the field at  $\pm 10$  ns. We notice (see Fig. 3) that the correlated field appears to vary by antenna/station positions, indicating that it ‘‘irons’’ out station and antenna-wise timing offsets.

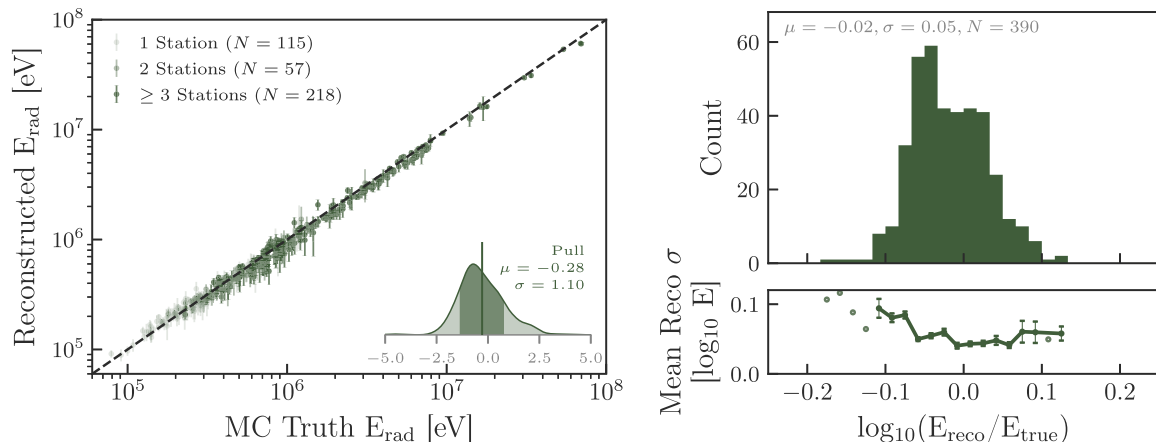
#### 4.3. Reconstruction performance

Post-optimization, the forward model can be evaluated for the generated samples to obtain Posterior distributions of both the shower parameters and the full forward model. This allows studying the results beyond just point estimates. An example reconstruction result is shown in Fig. 3.

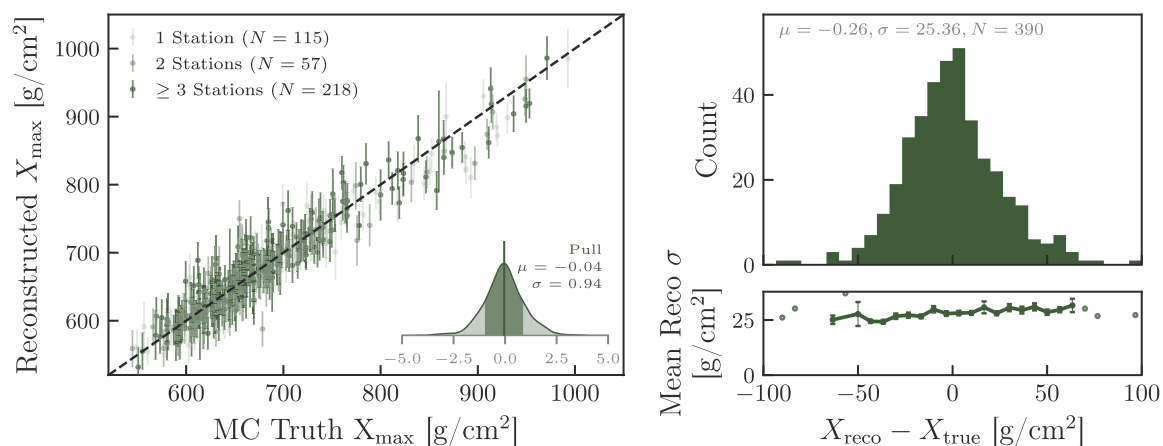
For the purpose of this example plot, we activated the fluence Correlated Field to demonstrate the model’s capacity to infer spatial systematics. The plot shows a good agreement between truth and reconstructed signal, with absolute fluence residuals of less than 10% and timing residuals within 2 ns.

Of all 500 simulations in the set, 390 (78%) passed the post-reconstruction quality cuts, demanding the reduced  $\chi^2$  to be  $< 1.25$  and the core reconstruction uncertainty to be below 10 m. These cuts ensure a high reconstruction quality without removing too many events. These cuts will be applied to real data as well.

We find that especially for higher energy events, the posteriors resulting from the reconstruction are too narrow. For lower energies, noise dominates the data, which is reflected in broader posterior distributions. For higher energy events, systematic data-model differences dominate, resulting lower uncertainties. Activating the fluence correlated field could be a solution to that in the future. For now, we perform a calibration of uncertainties based on the reconstructed radiation energy. The full calibration procedure is detailed in Appendix B. This



**Fig. 4.** Performance of the radiation energy  $E_{\text{rad}}$  reconstruction on CoREAS simulations. The left side displays both the reconstructed values compared to the CoREAS truth for all 390 successful reconstructions, as well as the pull from the reconstruction. The color-coding of the scatter points is based on how many stations were considered “triggered” for this specific event. The right side of the plot shows a histogram of the reconstruction errors at the top and how the reconstruction uncertainties correlate with reconstruction error.



**Fig. 5.** Performance of the  $X_{\text{max}}$  reconstruction on CoREAS simulations. (Left) Reconstructed vs. true  $X_{\text{max}}$  and pull distribution. The color-coding of the scatter points is based on how many stations were considered “triggered” for this specific event. (Right) Error distribution and uncertainty correlation.

calibration has been applied to all of the results that are discussed in the following sections.

We investigated, whether this underestimation of uncertainties is inherently due to the usage of VI, which in some cases can lead to too narrow posteriors. Running the reconstruction again using a No U-Turn Sampler (NUTS) from the NIFTY framework revealed no significant change in posterior width, indicating that the under-coverage stems from the model and not the inference engine.

#### 4.3.1. $E_{\text{rad}}$ reconstruction performance

The performance of the reconstruction of the radiation energy  $E_{\text{rad}}$  for the whole simulation set is shown in Fig. 4.

We achieve an overall precision on  $E_{\text{rad}}$  of 11.5% with a bias of 4.5%, leading to an overall resolution of 12.4%. The Figure also displays a Pull, which is an excellent measure of how well the reconstructed uncertainties are estimated. Post-calibration, the  $\sigma$  of the Pull is 1.1. A pull  $\sigma$  of 1 would correspond to ideal overall coverage. Pre-calibration the uncertainties were underestimated by an average factor of 4.45.

#### 4.3.2. $X_{\text{max}}$ reconstruction performance

Analogously, the performance on the  $X_{\text{max}}$  is shown in Fig. 5. We achieve an overall precision on the  $X_{\text{max}}$  of 25.4 g/cm<sup>2</sup> with a bias of  $-0.3$  g/cm<sup>2</sup>. The Pull  $\sigma$  post-calibration of uncertainties is 0.94. Before calibration, the Pull  $\sigma$  was 1.52.

We observe that especially for larger shower maxima, where the footprints tend to extend only over  $< 100$  m, the reconstruction predicts lower shower maxima. Since both the fluence and the timing parameterization are less accurate for large  $X_{\text{max}}$ , this behavior is somewhat expected.

Additionally to the combined timing-fluence model, we also tested an only fluence-based reconstruction. This was to test how much the timing information contributes to the fit. For this, we prepared the data exactly as for the timing-fluence case, but discarded timing data and the time component of the forward model. With that approach, we achieved a resolution in  $X_{\text{max}}$  of 36 g cm<sup>-2</sup>, meaning the timing information improved the resolution substantially, as further discussed in Appendix A.

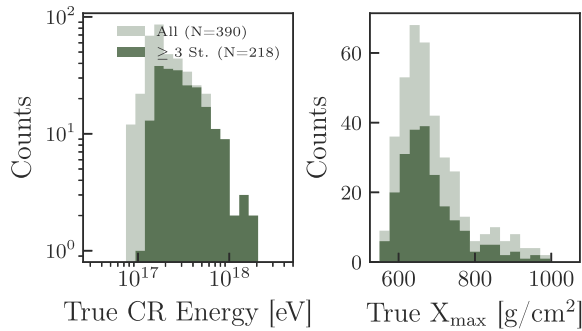


Fig. 6. Distribution of primary energies and depths of shower maximum, showcasing the effect of loosening the trigger criterion.

#### 4.3.3. Shower geometry performance

The shower geometry being part of the forward model means the zenith angle, azimuth angle and core position are optimized during the reconstruction and we can access their posteriors. We find a precision on the zenith angle  $\theta$  of  $0.5^\circ$  and an azimuth angle ( $\phi$ ) precision of  $1^\circ$ , in line with the size of the LOFAR core and observation wavelength. The overall core accuracy achieved was 8.4 m.

The fluence-based fit, in which timing data was discarded, achieved a similar resolution in azimuth as the combined fit, but only a  $1.5^\circ$  resolution on the zenith angle. This further shows how the timing data is beneficial to the overall reconstruction (see Appendix A).

#### 4.4. Impact of the trigger criterion

Prior reconstructions placed a rather strict constraint on events that would be considered for a mass composition. In this work we have also considered events where just one station was triggered. This will increase the number of reconstructable events, allowing a mass composition study with improved statistics and extending to lower energies. We have shown the events from our whole simulation set in Fig. 6, and marked the events that contained at least 3 triggered stations, and thus would have been considered in past analyses. As expected, more lower energy events made the less strict condition.

In a mass composition analysis these events would nevertheless have to be checked against a composition bias event by event, however [57] indicated that the previous trigger criterion was rather conservative.

#### 4.5. Resolution and error distribution analysis

The resolution in  $X_{\max}$  (Fig. 7) as compared to the legacy method is evaluated by analyzing the distribution of signed errors, defined as the residual  $\Delta X_{\max} = X_{\text{rec}} - X_{\text{true}}$ . This contrasts the observed residuals from the IFT pipeline against the expected performance of the legacy method.

The IFT distribution (labeled ‘IFT’) is derived directly from the calculated residuals of the filtered event set. The reference distribution (labeled ‘2021’) is generated synthetically to model the theoretical resolution of the legacy pipeline, assuming unbiased reconstruction errors. This is achieved via a bootstrap resampling procedure:

1. A pool of event uncertainties ( $\sigma_{2021}$ ) is extracted from the legacy dataset.
2. For each event in the comparison set, a  $\sigma_i$  is randomly sampled with replacement from this pool.
3. A synthetic error is generated by drawing from a normal distribution  $\mathcal{N}(\mu = 0, \sigma = \sigma_i)$ .

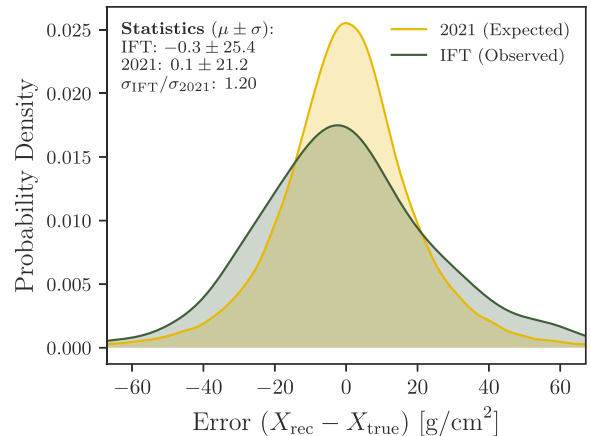


Fig. 7. Comparison of the  $X_{\max}$  error distribution between the IFT method and the legacy 2021 method. The expected error from the legacy method is based on its distribution of reconstruction uncertainties, while the IFT method represents direct reconstruction errors. The ratio of the two is displayed in the Statistics legend.

Both distributions are visualized as continuous probability density functions using Gaussian Kernel Density Estimation (KDE). The arithmetic mean ( $\mu$ ) and standard deviation ( $\sigma$ ) for both distributions are calculated to quantify the bias and resolution, respectively, with the ratio  $\sigma_{\text{IFT}}/\sigma_{2021}$  serving as a metric for relative precision.

While the IFT method appears a factor 1.2 less precise in this comparison, the resulting  $\sim 25 \text{ g/cm}^2$  resolution remains better than the  $30\text{--}40 \text{ g/cm}^2$  typical of sparse arrays in this energy range [5,58] and approaches the benchmark of the legacy LOFAR analysis [57]. Furthermore, the achieved energy resolution of 12.4% competes well with the 17% precision reported by AERA [6], albeit this includes further detector effects.

Combined with a computation time decrease of a factor of 1000 w.r.t. the legacy method and less strict triggering conditions allowing more ‘‘challenging’’ events (i.e. lower energies, smaller shower footprints) in our simulation set, this method is ideal for a mass composition study with a full LOFAR1.0 data set.

## 5. Application to LOFAR data

To verify the method further, we select a subset of LOFAR events previously reconstructed in Ref. [57]. The data are read in and converted into electric field traces using NuRadioReco [54,55].

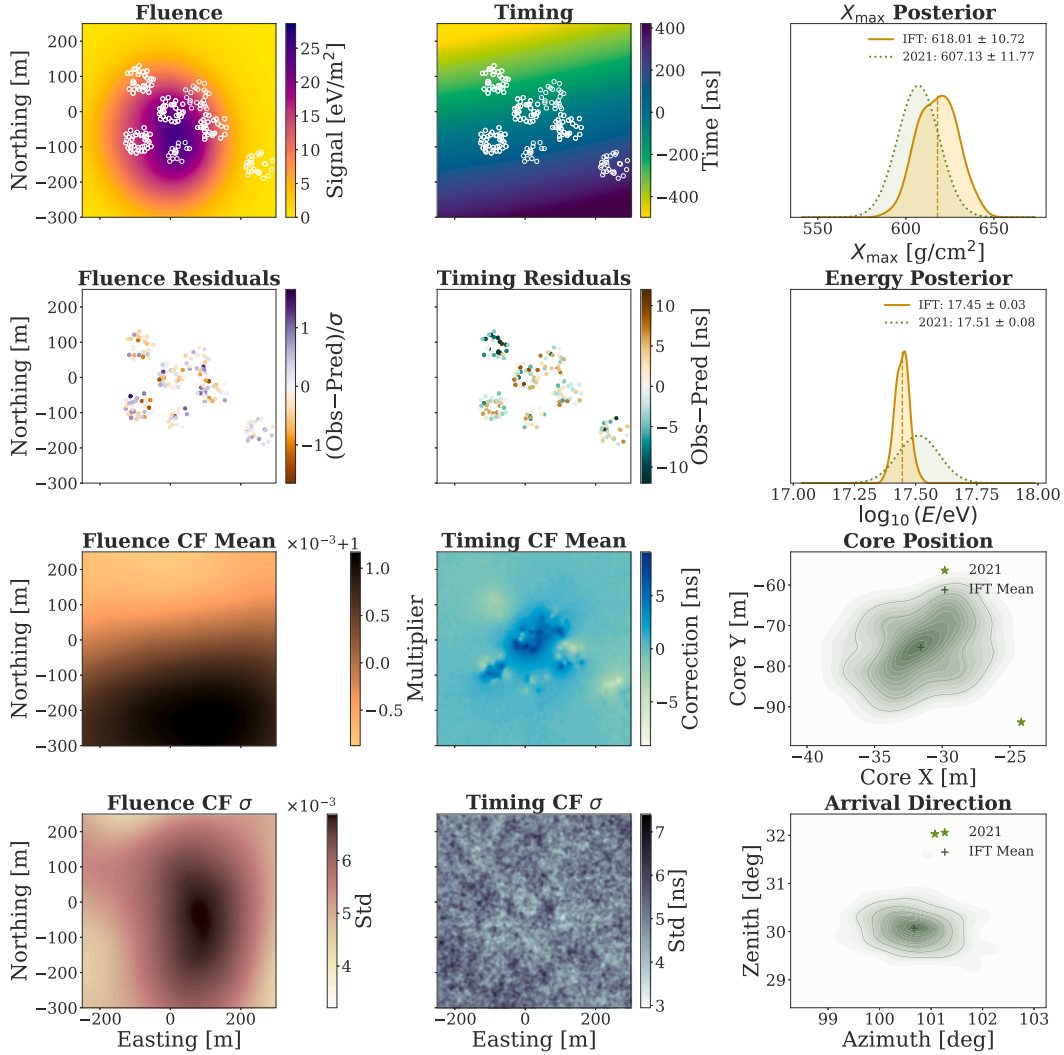
We use the stored initial estimate of the LORA particle data fit on the core position and shower direction as our prior mean for the shower geometry, but maintain our own prior widths as used in Section 4.2.4. The resulting prior distribution is usually wider than the expected LORA accuracy, but a wider prior can aid the optimizer in navigating through the posterior landscape and avoid getting stuck in local minima. This also makes us more independent of the goodness of fit of the particle data.

The pulse finding and data processing are handled as described in Section 4, without any further adjustment needed from what has been done for simulations. We use an individual atmosphere model for each event, matching the conditions of the event time.

Post-reconstruction we apply the simulation-based calibration described in Appendix B to the reconstruction uncertainties to avoid underestimating them.

To fully compare the method with the legacy reconstruction, we need to calculate the primary energy  $E_{\text{CR}}$  from the reconstructed radiation energy  $E_{\text{rad}}^{30\text{--}80\text{MHz}}$ . This can be done via the relation [6]

$$E_{\text{rad}}^{30\text{--}80\text{MHz}} / \sin^2(\alpha) = A \times 10^7 \text{ eV} (E_{\text{CR}}/10^{18} \text{ eV})^B. \quad (27)$$



**Fig. 8.** One of the LOFAR event reconstructions. Displayed are fluence in the left column, timing in the middle and posterior distributions compared with the legacy 2021 reconstruction on the right, where the arrival direction was determined via a planewave fit. For the fluence and timing columns, the top row shows the data along with the posterior mean in the background. The second row shows the residuals and the third and fourth row the inferred Correlated Field mean and standard deviation.

A more accurate relation would be only correcting the geomagnetic component of the radiation energy by  $\sin^2(\alpha)$  [59], however for the purpose of validating the method we use the simpler relation from Eq. (27). We calibrate the parameters  $A$  and  $B$  on the reconstructed radiation energies of our simulation set (a mixed proton/iron composition), yielding  $A = 9.57 \pm 0.90 \text{ MeV}$  and  $B = 2.010 \pm 0.011$ . We note that this is smaller than the value previously obtained for LOFAR data [9], but consistent within systematic uncertainties. It still warrants further investigation when analyzing the whole LOFAR1.0 data-set, whether this is due to method choices like the time-integration window, the choice of hadronic interaction model, or a previously overlooked effect. For validating the method we have found this to be sufficiently self-consistent.

### 5.1. Reconstruction quality

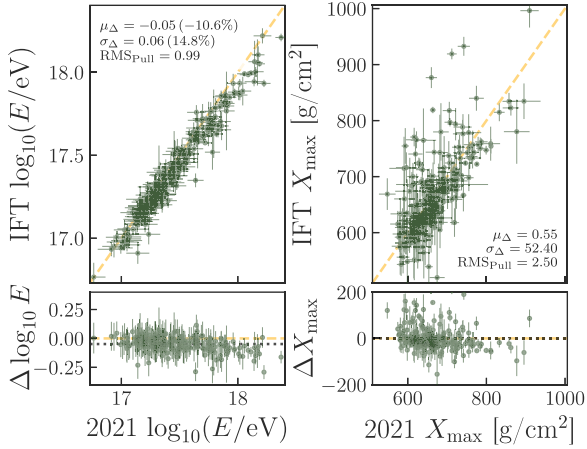
Of the 289 LOFAR events, 231 reconstructions were successful, which is with about 80% similar to the simulations. For the unsuccessful reconstructions, the optimizer failed to achieve a reduced  $\chi^2 < 1.25$ , or

returned a core uncertainty of  $> 10 \text{ m}$ . The test on simulations revealed that for those cases, the reconstruction accuracy was very low.

Post-reconstruction, we check for each of the samples generated during the optimization whether it would have triggered the required 3 stations. If not, the sample is discarded, ensuring that the posterior distribution only contains possible realizations of the model. In reality, only few ( $< 1\%$ ) samples were discarded over the full event set, indicating that the optimizer rarely found unrealistic realizations.

An example reconstruction and its parameter comparison to the 2021 reconstruction is shown in Fig. 8. The residuals show little structure, indicating they are mostly noise and not signal. The agreement between the 2021 reconstruction and this work for  $X_{\text{max}}$  and energy is good and in line with the behavior discussed in Section 4.3.1. The legacy method only uses a planewave to describe the arrival direction, which explains the obtained offset in zenith angle.

Furthermore, we show the event-wise comparison in reconstructed  $X_{\text{max}}$  and  $E_{\text{CR}}$  in Fig. 9. The  $X_{\text{max}}$  distribution on the right shows no bias and a standard deviation of  $52.4 \text{ g cm}^{-2}$ . Assuming the statistical uncertainties of the two methods are uncorrelated, the expected standard deviation of the difference distribution based on the respective



**Fig. 9.** (Top) Comparison of reconstructed shower parameters  $E_{CR}$  (left) and  $X_{max}$  (right) by the legacy method (2021) and this work (IFT). The statistics in the top panel mark the mean, standard deviation and the standard deviation of the pull. The dashed black line in the bottom panel marks the mean bias.

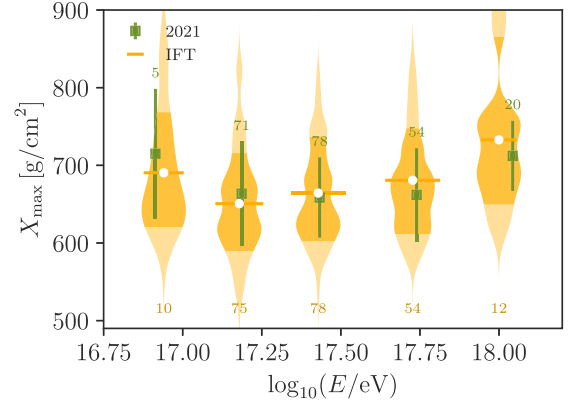
resolutions would be  $\sigma_{\Delta} = \sqrt{\sigma_{IFT}^2 + \sigma_{2021}^2} \approx 33 \text{ g cm}^{-2}$ . Upon further investigation of the differences in  $X_{max}$  between the two methods, we notice that this large scatter is primarily outlier-driven, as the difference distribution is slightly asymmetric with a non-Gaussian tail, indicating a few strong outliers where the IFT method predicts much larger  $X_{max}$  values than the legacy method. To account for this, we recalculated the scatter using a robust standard deviation estimator: the 68% interquartile range, defined as  $\sigma_{rob} = (P_{84} - P_{16})/2$ . This metric isolates the true  $1\sigma$  width of the dense core of the data while ignoring the extreme outlier tails. Using this metric, we retrieve a scatter of only  $35 \text{ g cm}^{-2}$ , however we also find that the median difference between the two methods is  $-9.7 \text{ g cm}^{-2}$  using this metric. Nevertheless, for the absolute majority of reconstructions the methods agree within their nominal precisions for  $X_{max}$ . The strong outliers warrant further investigation for a future composition study.

The reconstructed energies are systematically lower than in the legacy method, with the average reconstructed energy being 10% lower and with a standard deviation of 15%. As the energy bias seems to increase with energy, it cannot easily be corrected, however this might be due to the simplified conversion formula (Eq. (27)) that we used and will be investigated in the future before a full analysis of the LOFAR1.0 data set can be performed.

## 5.2. Mass composition performance

Fig. 10 shows the  $X_{max}$  distributions of the reconstructed event sample. This was not cross-checked against composition biases and purely serves as a comparison between what the two methods predict for the same LOFAR event set. To facilitate a direct comparison between the reconstruction pipelines, the dataset is strictly filtered to include only those events successfully reconstructed by both the legacy method and the new Information Field Theory approach. The data are grouped into energy bins of width  $\Delta \log_{10}(E/eV) = 0.3$  over the range  $16.5 \leq \log_{10}(E/eV) \leq 18.5$ . The number of events contributing to each bin ( $N$ ) is annotated explicitly in the figure.

For the legacy reconstruction, the binned data are represented by their arithmetic means and standard deviations. The square markers indicate the mean energy and mean depth of shower maximum ( $X_{max}$ ) of the events within the bin. The associated vertical and horizontal error bars represent the standard deviations of the  $X_{max}$  and  $\log_{10} E$  distributions, respectively.



**Fig. 10.**  $X_{max}$  distributions as determined by the method (labeled “IFT”). The violin plot shows summed posterior distributions of the reconstructed  $X_{max}$  for each energy bin, with the  $1\sigma$  band highlighted. The horizontal error bar represents the statistical and systematic energy uncertainty. The legacy method (labeled “2021”) is marked for comparison purposes. Error bars show both the statistical per-bin uncertainties and the event-by-event uncertainties. The numbers around each energy bin represent the number of events in the respective energy bin.

For the IFT reconstruction, the visualization leverages the full posterior probability distributions provided by the algorithm. Rather than reducing each event to a single scalar value, the posterior  $X_{max}$  samples from all events falling within a specific energy bin are summed. The resulting distribution is visualized using a Gaussian Kernel Density Estimation (KDE) to illustrate the total probability density of the reconstructed mass composition within that bin. The white circular markers indicate the global mean of these stacked samples, while the horizontal error bars denote the standard deviation of the mean reconstructed energies of the constituent events.

Within uncertainties, we observe good agreement between both methods.

## 6. Summary and outlook

Our novel Bayesian approach using Information Field Theory has shown promising results with an accuracy comparable to the method previously used for LOFAR data. The main improvements are the factor of 1000 decrease in runtime w.r.t. the direct simulation comparison and the wider selection of events we were able to reconstruct in tests on simulations. The added timing information as well as the non-parametric additions to the model make the reconstruction significantly more accurate than a solely fluence-based fit, and more flexible in the application on real data. On realistic simulations, we achieved an  $X_{max}$  resolution of  $25.4 \text{ g cm}^{-2}$  and a resolution in  $E_{rad}$  of 12.4%. The application to a small set of LOFAR data revealed no significant bias in  $X_{max}$  and slight bias of 10% in primary energy.

In the future, further investigation of the impact of the fluence Correlated Field will be performed, both to find out what impact it has on shower parameters and also whether it is beneficial to be added to data. Overall, a deeper insight in the behavior of the Correlated Fields on real data will be helpful, as they can model systematic behavior and spatial correlations that a simple uncorrelated noise model cannot.

While tested on LOFAR data, we note that the method can in principle be used site-independently for near-vertical air shower reconstruction in the 30 – 80 MHz band. The implementation of the method along with the full forward model and nifty8 optimizer is available for use [60].

Future improvements of the method will entail reconstructing based on the full voltage traces, as they also contain information on key

shower parameters. While parameterizations of the electric field spectra already exist [61], these were developed for inclined showers and found to not hold up for near-vertical showers. A development of a parameterization for these spectra is thus necessary. Additionally, the LOFAR antennas specifically are highly resonant, which makes the reconstruction of a spectrum across the full bandwidth challenging.

Future radio telescopes such as the SKA, which will have a more uniform antenna response over the full bandwidth, might be more suited for this approach. We are nevertheless working on the full waveform reconstruction, along with extensions of this model to broader frequency bands for LOFAR2.0 and the SKA.

### CRediT authorship contribution statement

**Karen Terveer:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sjoerd Bouma:** Writing – review & editing. **Stijn Buitink:** Writing – review & editing, Data curation. **Arthur Corstanje:** Writing – review & editing, Data curation. **Vital De Henau:** Writing – review & editing. **Vincent Eberle:** Writing – review & editing. **Torsten A. Enßlin:** Writing – review & editing, Funding acquisition. **Philipp Frank:** Writing – review & editing. **Tim Huege:** Writing – review & editing, Funding acquisition. **Philipp Laub:** Writing – review & editing. **Katharine Mulrey:** Writing – review & editing, Data curation. **Anna Nelles:** Writing – review & editing, Supervision, Funding acquisition, Data curation. **Simon Strähnz:** Writing – review & editing. **Satyendra Thoudam:** Writing – review & editing, Data curation. **Keito Watanabe:** Writing – review & editing.

### Acknowledgments

This work was supported by the German Federal Ministry for Research, Technology, and Space as part of the projects ErUM-IFT and D-LOFAR-ERIC, as well as the German Research Foundation through project 531213488.

We also acknowledge support by the Dutch Research Council (NWO) grant number OCENW.XS25.1.237.

### Declaration of competing interest

Given his role as Editor, Prof. Tim Huege had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor.

### Appendix A. Impact of added timing information

To determine the benefit of adding timing to the forward model, we performed a separate test on simulation and data using only the fluence implementation of the forward model. We kept all other steps identical to the data processing described in Section 4, including prior widths, pulse finding and noise modeling. For the simulation set we can compare the resolution with respect to the MC truth, which is displayed in Fig. 11 (top panel) for the radiation energy, the  $X_{\max}$  and the zenith angle.

The figure not only shows a clear improvement of resolution in  $X_{\max}$  and zenith angle when using the full forward model with timing for simulations, but it also shows that the estimated uncertainties are lower both for data and also for simulations, to a similar extent, as compared to a solely fluence-based reconstruction (bottom panel).

### Appendix B. Calibration of uncertainties

To account for the underestimation of uncertainties, we calibrate based on the Pull  $\sigma$  of our simulation set. Fig. 12 shows the Pull  $\sigma$  as a function of shower parameters for the radiation energy.

The reconstruction appears overly confident for higher energy events, likely because for lower energies the data are more dominated by noise than by systematic differences from the forward model. We model the Pull  $\sigma$  as a function of reconstructed radiation energy via a third degree polynomial fit. The posterior distribution of the radiation energy is then scaled by this function. Fig. 12 also shows how a global correction by the overall Pull  $\sigma$  would perform. Overall we find the  $E_{\text{rad}}^{\text{reco}}$ -based correction yields a better coverage.

The  $X_{\max}$  reconstruction in Fig. 5 also indicates undercoverage. As for the radiation energy, we find that the Pull increases as a function of energy, so we calibrate based on reconstructed radiation energy. Fig. 13 shows the Pull dependence for the  $X_{\max}$  as function of shower parameters, the cubic fit used for the calibration and the comparison with a simple global correction.

We note that there seems to be an additional dependence on zenith angle, as the bottom right subfigure of Fig. 13 indicates. We attempted also a two-dimensional calibration based on both radiation energy and zenith angle, however we found that the resulting overall coverage in  $X_{\max}$  was worse than for a one-dimensional calibration.

We suspect that, especially for the radiation energy reconstruction, the usage of correlated fields in the fluence forward model would provide a better coverage due to the added degrees of freedom in the model. In the future, the effect of the fluence correlated field on  $X_{\max}$  could be further investigated, possibly resulting in a similar reconstruction precision and better coverage without the need for a simulation-based calibration.

### Appendix C. Particle model

The NKG function describes the particle density  $\rho(r)$  at lateral distance  $r$  from the shower axis as:

$$\rho(r) = N_e \cdot C(s) \cdot \left(\frac{r}{R_M}\right)^{s-2} \left(1 + \frac{r}{R_M}\right)^{s-4.5}, \quad (28)$$

where  $N_e$  is the total number of particles at ground level,  $s$  is the shower age parameter, and  $R_M$  is the Molière radius. The normalization constant  $C(s)$  ensures proper integration over the lateral plane:

$$C(s) = \frac{\Gamma(4.5 - s)}{2\pi R_M^2 \Gamma(s) \Gamma(4.5 - 2s)}. \quad (29)$$

The shower age  $s$  characterizes the evolutionary stage of the cascade and is typically  $s \approx 1$  at shower maximum. For showers observed at ground level,  $s > 1$  indicates post-maximum development. The total particle number  $N_e$  is related to the primary cosmic ray energy  $E_{\text{CR}}$  via an empirical relation:

$$\log_{10}(N_e^{\text{max}}) = \frac{\log_{10}(E_{\text{CR}}/\text{eV}) - A}{B}, \quad (30)$$

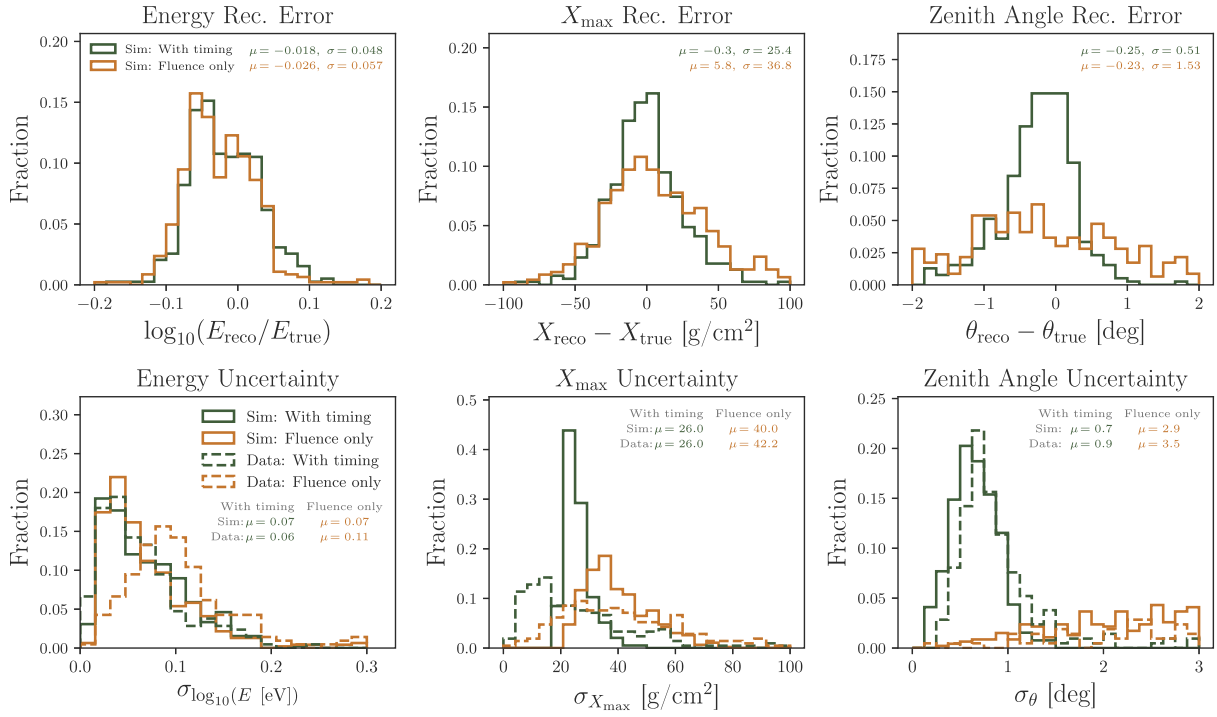
where  $A$  and  $B$  are constants calibrated to simulations.

An attenuation factor accounts for the decrease in particle density beyond shower maximum. This is modeled as:

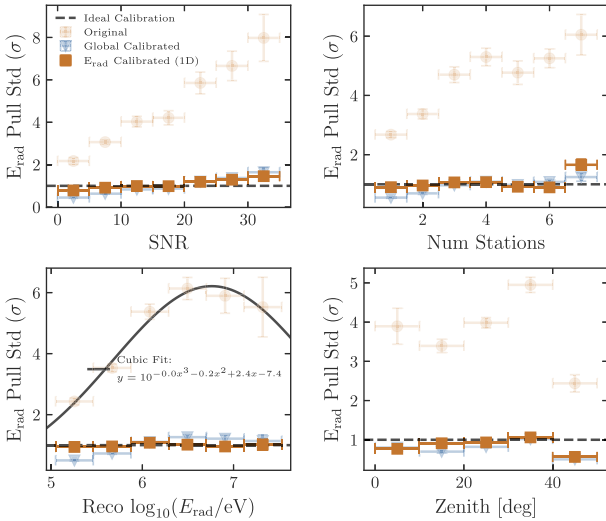
$$\ln(N_e^{\text{ground}}) = \ln(N_e^{\text{max}}) - \frac{D_{X_{\max}}}{A_{\text{att}}}, \quad (31)$$

where  $A_{\text{att}}$  is the attenuation length for electromagnetic particles in air.

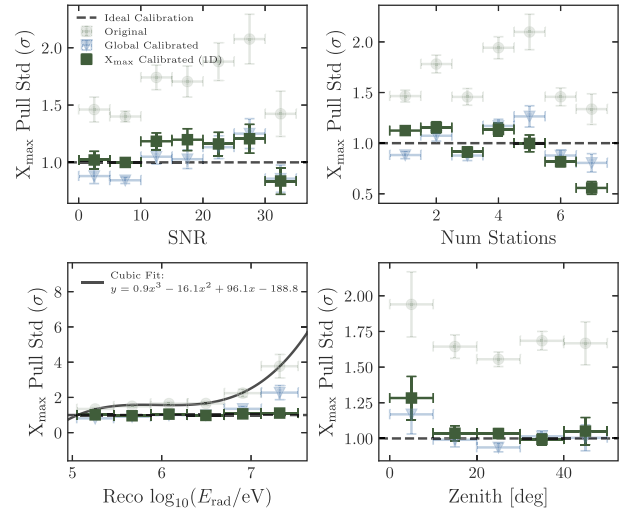
The LORA detectors do not count particles directly but instead measure the total deposited energy. This quantity depends not only on the total number of particles hitting the detector but also on the track length through the scintillator, which is approximately  $L/\cos(\theta)$  for fully contained tracks, with  $L$  representing the thickness of the scintillator. In practice, the effective length is somewhat shorter because some particles leave or enter through the sides; nevertheless, this geometric effect largely compensates for the reduction in the projected detector



**Fig. 11.** Resolution comparison of the pure fluence forward model compared to the model with timing information for three shower parameters (radiation energy,  $X_{\max}$  and zenith angle). The top panel shows the error distributions of reconstructed value compared to the simulated MC truth, with the statistics of each in the top right corner of each respective subfigure. The bottom panel shows uncertainty distributions both for simulations (solid lines) and data (dashed lines) derived from the IFT reconstruction, comparing fluence-only (yellow) and the full model (green). The mean uncertainty values are displayed in a table in the top right corner of each subfigure.



**Fig. 12.** Pull standard deviation of the  $E_{\text{rad}}$  for different shower parameters: SNR (maximum fluence SNR across the event), number of triggered stations, reconstructed radiation energy and zenith angle. The original Pull  $\sigma$  are compared to a simple global correction and a correction based on a cubic fit to the Pull  $\sigma$  as function of reconstructed radiation energy.



**Fig. 13.** Pull standard deviation of the  $X_{\max}$  for different shower parameters: SNR (maximum fluence SNR across the event), number of triggered stations, reconstructed radiation energy and zenith angle. The original Pull  $\sigma$  are compared to a simple global correction and a correction based on a cubic fit to the Pull  $\sigma$  as function of reconstructed radiation energy.

area. Consequently, the predicted count in detector  $i$  with effective area  $A_{\text{eff}}$  is:

$$\lambda_i = \rho(r_i) \cdot A_{\text{eff}} + n_{\text{bg}}, \quad (32)$$

where  $n_{\text{bg}}$  represents a small background term.

As with the radio model, the particle prediction includes a non-parametric correlated field to model systematic uncertainties. This field

is clipped to prevent unphysical corrections. The full implementation is differentiable, sharing the core position and arrival direction with the radio model while maintaining separate particle-specific parameters ( $s$ ,  $R_M$ , overall scale). In the future one could even link the particle-specific parameters to the radio parameters, as the shower age depends on  $X_{\max}$  and the Molière radius can be parameterized from energy,  $X_{\max}$ , arrival direction and the atmospheric model. This however goes

beyond the scope of this work as for now we do not actively use the particle model for reconstruction.

## Data availability

The forward model code has been made public (Reference [60]).

## References

- [1] K. Kampert, M. Unger, Measurements of the cosmic ray composition with air shower experiments, *Astropart. Phys.* 35 (10) (2012) 660–678.
- [2] A. Aab, et al., Pierre Auger Collaboration Collaboration, The pierre auger cosmic ray observatory, *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrometers Detect. Assoc. Equip.* 798 (2015) 172–213.
- [3] T. Huege, Radio detection of cosmic ray air showers in the digital era, *Phys. Rep.* 620 (2016) 1–52.
- [4] A. Abdul Halim, et al., Pierre Auger Collaboration, Demonstrating agreement between radio and fluorescence measurements of the depth of maximum of extensive air showers at the pierre auger observatory, *Phys. Rev. Lett.* 132 (2) (2024) 021001.
- [5] P.A. Bezyazeev, et al., Tunka-Rex Collaboration, Radio measurements of the energy and the depth of the shower maximum of cosmic-ray air showers by Tunka-Rex, *J. Cosmol. Astropart. Phys.* 01 (2016) 052.
- [6] A. Aab, et al., 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 (24) (2016) 241101.
- [7] S. Buitink, et al., LOFAR Collaboration, A large light-mass component of cosmic rays at  $10^{17} - 10^{17.5}$  ev from radio observations, *Nature* 531 (2016) 70.
- [8] A. Abdul Halim, et al., Pierre Auger Collaboration, Determination of the energy scale of cosmic ray measurements using the auger engineering radio array, *PoS ICRC2025* (2025) 292.
- [9] K. Mulrey, et al., LOFAR Collaboration, On the cosmic-ray energy scale of the LOFAR radio telescope, *J. Cosmol. Astropart. Phys.* 11 (2020) 017.
- [10] W.D. Apel, et al., Tunka-Rex, LOPES Collaboration, A comparison of the cosmic-ray energy scales of tunka-133 and KASCADE-grande via their radio extensions tunka-rx and LOPES, *Phys. Lett. B* 763 (2016) 179–185.
- [11] T.A. Enßlin, M. Frommert, F.S. Kitaura, Information field theory for cosmological perturbation reconstruction and non-linear signal analysis, *Phys. Rev. D* 80 (2009) 105005.
- [12] J. Jasche, et al., Bayesian power-spectrum inference for large-scale structure data, *Mon. Not. R. Astron. Soc.* 406 (2010) 60.
- [13] G. Edenhofer, et al., A parsec-scale galactic 3D dust map out to 1.25 kpc from the sun, *Astron. Astrophys.* 685 (2024) A82.
- [14] C. Welling, et al., Reconstructing non-repeating radio pulses with information field theory, *J. Cosmol. Astropart. Phys.* 2021 (04) (2021) 071.
- [15] S. Strähnz, et al., Electric field reconstruction with information field theory, 2024, <http://dx.doi.org/10.22323/1.470.0056>, *PoS(ARENA2024)056*.
- [16] K. Terveer, et al., Reconstructing air showers using probabilistic forward modelling for LOFAR, in: 39th International Cosmic Ray Conference, 2025, p. 413, <http://dx.doi.org/10.22323/1.501.0413>.
- [17] S. Strähnz, T. Huege, T. Enßlin, K. Terveer, A. Nelles, Information field theory based event reconstruction for cosmic ray radio detectors, *PoS ICRC2025* (2025) 402.
- [18] K. Watanabe, et al., A novel approach for air shower profile reconstruction with dense radio antenna arrays using information field theory, in: 39th International Cosmic Ray Conference, 2025, p. 436, <http://dx.doi.org/10.22323/1.501.0436>.
- [19] C. Glaser, et al., An analytic description of the radio emission of air showers based on its emission mechanisms, *Astropart. Phys.* 104 (2019) 64–77.
- [20] H.R. Allan, Low frequency radio emission from extensive air showers, *Nature* 237 (1972) 384–385.
- [21] K.D. de Vries, A.M. van den Berg, O. Scholten, K. Werner, Coherent cherenkov radiation from cosmic-ray-induced air showers, *Phys. Rev. Lett.* 107 (2011) 061101.
- [22] C.W. James, H. Falcke, T. Huege, M. Ludwig, General description of electromagnetic radiation processes based on instantaneous charge acceleration in ‘endpoints’, *Phys. Rev. E* 84 (2011) 056602.
- [23] G.A. Askaryan, Excess negative charge of an electron-photon shower and its coherent radio emission, *Sov. Phys. JETP* 14 (1962) 441.
- [24] P. Schellart, et al., LOFAR Collaboration, Polarized radio emission from extensive air showers measured with LOFAR, *J. Cosmol. Astropart. Phys.* 10 (2014) 014.
- [25] A. Nelles, et al., LOFAR Collaboration, Measuring a cherenkov ring in the radio emission from air showers at 110–190 MHz with LOFAR, *Astropart. Phys.* 65 (2015) 11–21.
- [26] A.A. Halim, et al., Pierre Auger Collaboration, Long-term calibration and validation of stability of the auger engineering radio array using the diffuse galactic radio emission, *J. Instrum.* 20 (12) (2025) P12017.
- [27] P. Schellart, et al., LOFAR Collaboration, Detecting cosmic rays with the LOFAR radio telescope, *Astron. Astrophys.* 560 (2013) A98.
- [28] A. Corstanje, et al., LOFAR-style reconstruction of cosmic-ray air showers with SKA-low, *Phys. Rev. D* 112 (2) (2025) 023017.
- [29] M.P. van Haarlem, et al., LOFAR Collaboration, LOFAR: The LOW-frequency array, *Astron. Astrophys.* 556 (2013) A2.
- [30] K. Mulrey, et al., Cosmic-ray detection with LOFAR 2.0, *PoS ICRC2025* (2025) 340.
- [31] S. Thoudam, et al., LORA: A scintillator array for LOFAR to measure extensive air showers, *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrometers Detect. Assoc. Equip.* 767 (2014) 339–346.
- [32] K. Mulrey, et al., LOFAR Collaboration, Calibration of the LOFAR low-band antennas using the galaxy and a model of the signal chain, *Astropart. Phys.* 111 (2019) 1–11.
- [33] S. Buitink, et al., LOFAR Collaboration, Method for high precision reconstruction of air shower  $X_{max}$  using two-dimensional radio intensity profiles, *Phys. Rev. D* 90 (8) (2014) 082003.
- [34] T. Huege, M. Ludwig, C.W. James, Simulating radio emission from air showers with CoREAS, *AIP Conf. Proc.* 1535 (1) (2013) 128–132.
- [35] M. Desmet, S. Buitink, T. Huege, K. Watanabe, SMIET: Fast and accurate synthesis of radio pulses from extensive air shower using simulated templates, *Astropart. Phys.* 175 (2026) 103182.
- [36] A. Corstanje, et al., LOFAR Collaboration, The shape of the radio wavefront of extensive air showers as measured with LOFAR, *Astropart. Phys.* 61 (2015) 22–31.
- [37] W. Apel, et al., LOPES Collaboration, The wavefront of the radio signal emitted by cosmic ray air showers, *J. Cosmol. Astropart. Phys.* 2014 (09) (2014) 025.
- [38] T.A. Enßlin, Information field theory, *Ann. Phys., Lpz.* 531 (2019) 1800127.
- [39] J. Knollmüller, T.A. Enßlin, Metric Gaussian variational inference, 2020.
- [40] P. Frank, R. Leike, T.A. Enßlin, Geometric variational inference, *Entropy* 23 (7) (2021) 853.
- [41] G. Edenhofer, et al., Re-envisioning numerical information field theory (NIFTY.re): A library for Gaussian processes and variational inference, *J. Open Source Softw.* 9 (98) (2024) 6593.
- [42] P. Arras, et al., NIFTY 5: Numerical information field theory, 2019, *Astrophysics Source Code Library*.arXiv:asc1:1903.008.
- [43] T. Steininger, et al., NIFTY 3 - numerical information field theory: A python framework for Bayesian signal inference on HPC clusters, *Ann. Phys., Lpz.* 531 (2019) 1800290.
- [44] M. Selig, et al., NIFTY: Numerical information field theory. a versatile Python library for signal inference, *Astron. Astrophys.* 554 (2013) A26.
- [45] J. Bradbury, et al., JAX: composable transformations of Python+NumPy programs, 2018, <http://github.com/google/jax>.
- [46] P. Mitra, et al., Reconstructing air shower parameters with LOFAR using event specific GDAS atmosphere, *Astropart. Phys.* 123 (2020) 102470.
- [47] T. Huege, et al., LOPES Collaboration, Radio detection of cosmic ray air showers with the LOPES experiment, *J. Phys.: Conf. Ser.* 110 (6) (2008) 062012.
- [48] K. Greisen, Cosmic ray showers, *Ann. Rev. Nucl. Sci.* 10 (1960) 63–108.
- [49] K. Kamata, J. Nishimura, The lateral and the angular structure functions of electron showers, *Prog. Theor. Phys. Suppl.* 6 (1958) 93–155.
- [50] G. Cantarini, Modeling the Reflection of Radio Signals from Atmospheric Clouds as a Near-Horizon Background for Astroparticle Observatories (Masterarbeit), University of Milano Bicocca, 2025, p. 92, URL <https://bib-pubdb1.desy.de/record/644602>.
- [51] P. Arras, et al., Variable structures in M87\* from space, time and frequency resolved interferometry, *Nat. Astron.* 6 (2022) 259–269.
- [52] M. Ravn, C. Glaser, T. Glüsenkamp, A. Öcelikkale, A. Coleman, Likelihood reconstruction for radio detectors of neutrinos and cosmic rays, 2025, URL <https://arxiv.org/abs/2510.21925>, arXiv:2510.21925.
- [53] S. Ostapchenko, QGSJET-II: physics, recent improvements, and results for air showers, *EPJ Web Conf.* 52 (2013) 02001, *ISVHECRI 2012 – XVII International Symposium on Very High Energy Cosmic Ray Interactions*.
- [54] C. Glaser, et al., NuRadioReco: a reconstruction framework for radio neutrino detectors, *Eur. Phys. J. C* 79 (6) (2019).
- [55] S. Bouma, M. Desmet, C. Glaser, P. Laub, A. Nelles, L. Pyras, F. Schlüter, K. Terveer, Integrating radio detectors of cosmic-ray air showers into the open-source NuRadio framework, *PoS ICRC2025* (2025) 345.
- [56] A. Corstanje, et al., A high-precision interpolation method for pulsed radio signals from cosmic-ray air showers, *J. Instrum.* 18 (09) (2023) P09005.
- [57] A. Corstanje, et al., LOFAR Collaboration, Depth of shower maximum and mass composition of cosmic rays from 50 PeV to 2 EeV measured with the LOFAR radio telescope, *Phys. Rev. D* 103 (10) (2021).
- [58] A. Abdul Halim, et al., Pierre Auger Collaboration Collaboration, Radio measurements of the depth of air-shower maximum at the pierre auger observatory, *Phys. Rev. D* 109 (2024) 022002.
- [59] C. Glaser, M. Erdmann, J.R. Hörandel, T. Huege, J. Schulz, Simulation of radiation energy release in air showers, *J. Cosmol. Astropart. Phys.* 2016 (09) (2016) 024.
- [60] K. Terveer, Air shower reconstruction code, 2026, <https://github.com/karenterveer/RIA>.
- [61] S. Martinelli, F. Schlüter, T. Huege, Parameterization of the frequency spectrum of radio emission in the 30-80 MHz band from inclined air showers, *PoS ARENA2022* (2023) 036.