

2025 年度 修士論文

TA 地表検出器アレイを用いた機械学習による  
最高エネルギー宇宙線の原子核種推定

(Mass identification of the highest-energy cosmic ray  
using machine learning  
with the Telescope Array Surface Detector array)

大阪公立大学大学院理学研究科  
物理学専攻  
宇宙線物理学研究室

博士前期課程 2 年  
学籍番号 BHB24071

遠藤 康平

## 概要

$10^{18}$  eV を超えるエネルギーを持つ宇宙線の観測を目的とするテレスコプアレイ実験 (Telescope Array : TA) は、2021 年 5 月 21 日に 244 EeV という極めて高いエネルギーを持つ宇宙線を観測した。この宇宙線は「アマテラス粒子」と名付けられ、最高エネルギー宇宙線の起源探索に非常に重要なイベントとして注目されている。宇宙線は荷電粒子であり、銀河系内外の磁場の影響を受け進行方向が曲げられるため、到来方向からその起源を特定することは困難である。しかし、アマテラス粒子はその極めて高いエネルギーゆえに磁場から受ける影響が小さく、ほぼ直進すると期待される。また宇宙空間の光子との相互作用によってエネルギーを失うため、その起源は約 50 Mpc 以内の近傍に制限される。そのためアマテラス粒子の起源特定は可能かと思われたが、観測によって得られた到来方向は宇宙空間のボイドに位置しており、近傍には起源となりうる天体が存在しないことが分かっている。ここで注目されるのがアマテラス粒子の原子核種である。仮にアマテラス粒子が陽子のような軽い原子核であれば磁場の影響をほぼ受けず直進するため、既存の候補天体では説明できない起源の可能性が生まれる。反対に鉄のような重い原子核であれば磁場の影響を強く受け、起源天体の候補領域が拡大する。

本研究では、TA 地表検出器アレイを用いた機械学習によるアマテラス粒子の原子核種推定を行う。シミュレーションを用いてアマテラス粒子の観測結果を再現する空気シャワーイベントを生成し、シミュレーションデータから抽出した特徴量を学習データとする機械学習モデルを構築する。機械学習手法は CNN とランダムフォレストを採用した。学習モデルの正解率は約 85% となり、アマテラス粒子に対するモデルの出力値「Ironness (陽子を 0、鉄を 1 とした場合の鉄らしさ)」は約 0.7 となり鉄原子核に近い判定となった。またランダムフォレストによる特徴量重要度の評価から、地表検出器データのうち空気シャワー中のミューオン数が影響するピーク波形の情報や、空気シャワーフロントプレーンが作る波形全体の形状が原子核種推定に効果的であることが判明した。

# 目次

第 1 章	最高エネルギー宇宙線	1
1.1	最高エネルギー宇宙線観測	1
1.2	最高エネルギー宇宙線の起源候補	2
1.3	空気シャワー	3
1.4	アマテラス粒子	5
1.5	原子核種の重要性	7
第 2 章	テレスコープアレイ実験	10
2.1	地表検出器アレイ	10
2.2	標準的な解析手法と本研究のアプローチ	13
第 3 章	機械学習用空気シャワーイベントの生成	15
3.1	アマテラス粒子の再構成	15
3.2	空気シャワー生成シミュレーション	17
3.3	SD 応答シミュレーション	18
3.4	シミュレーションパラメータの調整	19
第 4 章	CNN を用いた原子核種推定	25
4.1	畳み込みニューラルネットワーク	25
4.2	データセットの作成	26
4.3	学習モデル	26
4.4	結果	29
第 5 章	ランダムフォレストを用いた原子核種推定	33
5.1	ランダムフォレスト	33
5.2	データセットの作成	33
5.3	学習モデル	35
5.4	結果	37
第 6 章	議論	44
6.1	モデルの推定結果	44

---

6.2	特徴量重要度 . . . . .	44
6.3	課題と展望 . . . . .	45
第 7 章	結論	46
付録		47
A	QGSJET-II-04 によるシミュレーションデータ . . . . .	47
B	QGSJET-II-04 による学習モデルの特徴量重要度 . . . . .	47
謝辞		56

# 目次

1.1	原子核種を区別しない宇宙線のエネルギースペクトル	1
1.2	Hillas plot	2
1.3	空気シャワーの模式図	4
1.4	陽子と鉄の $X_{\max}$ の比較	5
1.5	アマテラス粒子が観測された際のイメージ図 ([7] より掲載)。	6
1.6	アマテラス粒子の到来方向を表した図	7
1.7	アマテラス粒子の原子核種ごとの起源予測	9
2.1	TASD の配置	11
2.2	SD の模式図	12
2.3	波形データの一例	13
3.1	アマテラス粒子の観測結果	16
3.2	シニングとデシニング	18
3.3	MC イベントの再構成結果 (調整前)	21
3.4	MC イベントの再構成結果 (調整前)	22
3.5	MC イベントの再構成結果 (調整後)	23
3.6	MC イベントの再構成結果 (調整後)	24
4.1	CNN による学習の模式図。2D アレイに対して畳み込みによる特徴抽出を行い、NN の入力層とする。本研究では最終的に 1 つの出力を持つ。	26
4.2	TASD, 埋めあり	27
4.3	CNN による学習の学習曲線	30
4.4	EPOS-LHC のデータセットを使ったモデルの予測結果	31
4.5	QGSJETII-04 のデータセットを使ったモデルの予測結果	32
5.1	TASD, 埋めなし	35
5.2	EPOS-LHC のデータセットを使ったモデルの予測結果	38
5.3	QGSJET-II-04 のデータセットを使ったモデルの予測結果	39
5.4	重要度上位 20 の特徴量	40

---

5.5	重要度上位 20 の特徴量 (Centroid) . . . . .	41
5.6	重要度上位 20 の特徴量 (MFCC-3) . . . . .	42
5.7	重要度上位 20 の特徴量 (Spectrogram mean coefficient 800kHz) . . . . .	43
A.1	MC イベントの再構成結果 (調整前) . . . . .	48
A.2	MC イベントの再構成結果 (調整前) . . . . .	49
A.3	MC イベントの再構成結果 (調整後) . . . . .	50
A.4	MC イベントの再構成結果 (調整後) . . . . .	51
B.5	重要度上位 20 の特徴量 . . . . .	52
B.6	重要度上位 20 の特徴量 (Centroid) . . . . .	53
B.7	重要度上位 20 の特徴量 (MFCC-3) . . . . .	54
B.8	重要度上位 20 の特徴量 (Spectrogram mean coefficient 800kHz) . . . . .	55

# 表目次

3.1	アマテラス粒子の再構成データ . . . . .	15
3.2	シミュレーション設定 . . . . .	19
3.3	シミュレーション設定 (調整後) . . . . .	20
4.1	パラメータ探索範囲の設定 . . . . .	28
4.2	機械学習モデル設定 . . . . .	29
4.3	学習モデルの正解率 . . . . .	30
5.1	パラメータ探索範囲 . . . . .	36
5.2	機械学習モデル設定 . . . . .	36
5.3	学習モデルの正解率 (全イベント) . . . . .	37
5.4	学習モデルの正解率 (鉄イベントのみ) . . . . .	37

## 第 1 章

# 最高エネルギー宇宙線

### 1.1 最高エネルギー宇宙線観測

宇宙線とは、宇宙空間を飛び交う高エネルギーの原子核（陽子やヘリウム、鉄など）である。その中でも  $10^{18}$  eV を超えるものを超高エネルギー宇宙線（Ultra-High-Energy Cosmic Ray: UHECR）と呼ぶ。UHECR の持つエネルギーは、地上の加速器実験（Large Hadron Collider, LHC）などによって到達可能なエネルギーと比較しても桁違いに高く、 $10^{20}$  eV を超える極めて高いものもある。本研究ではこのような宇宙線を最高エネルギー宇宙線とし、その解析を行う。

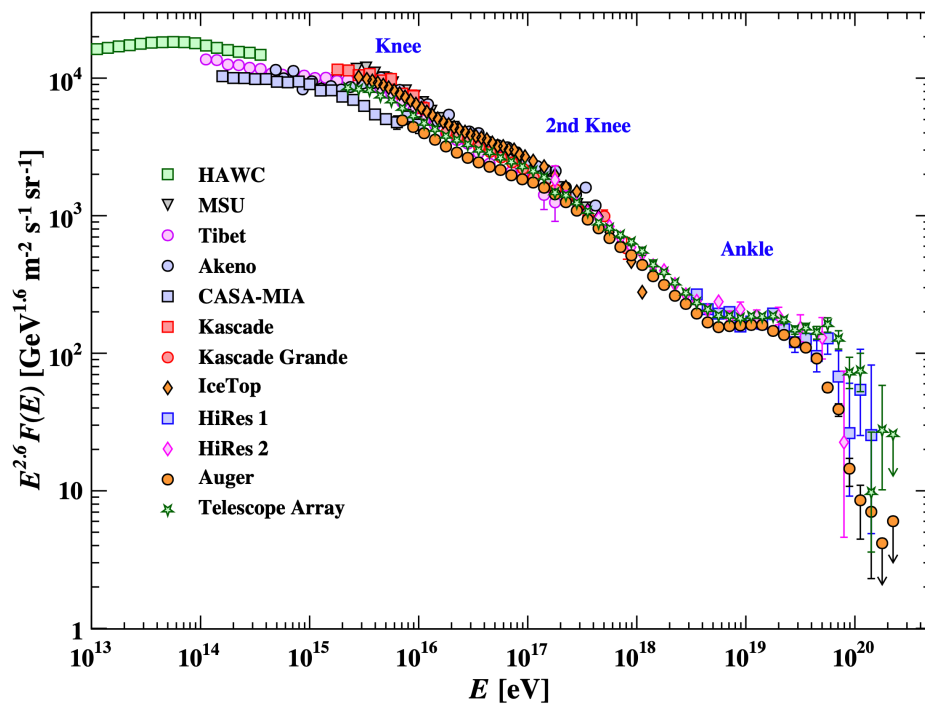


図 1.1 原子核種を区別しない宇宙線のエネルギースペクトル。([1] から掲載)

### 1.1.1 宇宙線エネルギースペクトル

宇宙線は、その到来頻度（フラックス）がエネルギーが高くなるにつれ急激に低下する。図 1.1 に原子核種を区別しない宇宙線のエネルギースペクトルを示す [1]。宇宙線のエネルギースペクトルはべき乗則に従い、おおよそ  $E^{-2.7}$  で減少する。特に  $10^{20}$  eV を超えるような最高エネルギー宇宙線の到来頻度は、 $1 \text{ km}^2$  あたりに 100 年に 1 イベント程度と極めて稀である。そのためその観測には数百  $\text{km}^2$  以上の広大な有効面積を持つ検出器が必要である。

## 1.2 最高エネルギー宇宙線の起源候補

宇宙線の加速は、磁場の存在する領域に閉じ込められ、時間をかけて行われる。ある領域で加速可能なエネルギー上限は、宇宙線を閉じ込めるための条件として、加速領域の磁場  $B$  と大きさ  $L$  を用いて次のように表される [2]。

$$E_{\text{max}} \approx \beta ZeBL \tag{1.1}$$

ここで  $\beta$  は加速領域における衝撃波の速度、 $Z$  は粒子の電荷である。図 1.2 は縦軸に磁場の強さ、横軸に天体の大きさを取り、陽子を仮定した際に  $10^{20}$  eV の宇宙線を閉じ込められる天体の条件を示したものであり、Hillas plot と呼ばれる。

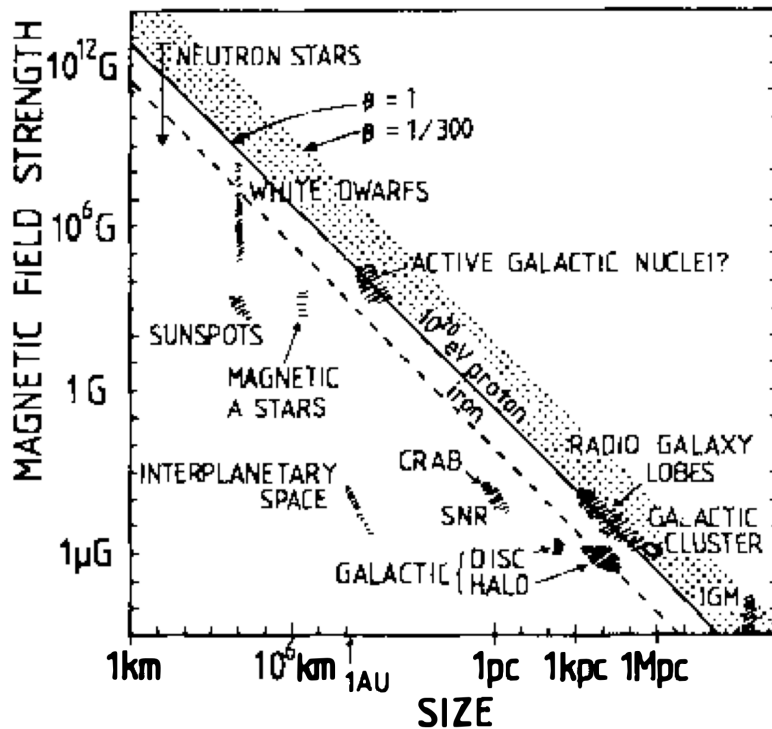


図 1.2 Hillas plot ([2] より掲載)

宇宙線をアマテラス粒子のような極めて高いエネルギーにまで加速できる起源候補の例として、以下の天体现象が挙げられる。

### 活動銀河核 (AGN)

活動銀河核では、中心からのジェット内の衝撃波統計加速（フェルミ加速）により、粒子が極限まで加速されると考えられている。近傍の AGN としては、ケンタウルス A などが知られており、到来方向の異方性との相関も議論されている。

### スターバースト銀河 (SBG)

活発な星形成が行われているスターバースト銀河では、複数の超新星爆発によって銀河規模の「スーパーウィンド」が形成される。この強大な銀河風に伴う衝撃波が、UHECR の加速を担っている可能性がある。実際、M82 や NGC253 といったスターバースト銀河の方向と、観測された UHECR の過剰領域（ホットスポット）との相関が報告されている [3]。

### ガンマ線バースト (GRB)

ガンマ線バーストは、巨大恒星の爆発や、中性子星合体によって引き起こされる宇宙最大規模の爆発現象である。ガンマ線バーストで放出される超音速の衝撃波によって、宇宙線が最高エネルギーまで加速される可能性がある。

### 新物理起源

最高エネルギー宇宙線はその極めて高いエネルギーから、標準模型を超えた新物理による加速機構の考察が存在し、非常に重いダークマターの崩壊などが挙げられる。

## 1.3 空気シャワー

前節で述べたように、最高エネルギー宇宙線の到来頻度は非常に低く、広大な有効面積を確保するためには地表でその観測を行う必要がある。地表での宇宙線観測では、以下に述べる空気シャワー現象を用いる。

高エネルギーの宇宙線が大気圏に突入すると、大気中の窒素や酸素原子核と相互作用を起こし、パイ中間子などの二次粒子を生成する。

$$A + N_{\text{air}} \rightarrow p(n) + \pi^{\pm} + \pi^0 + K^{\pm} + \dots \quad (1.2)$$

その二次粒子がさらに以下のような相互作用や粒子生成を繰り返す。

中性パイ中間子の崩壊によるガンマ線生成、ガンマ線の電子対生成、電子陽電子の制動放射によるガンマ線生成

$$\begin{aligned} \pi^0 &\rightarrow \gamma + \gamma \\ \gamma &\rightarrow e^+ + e^- \\ e^{\pm} &\rightarrow e^{\pm} + \gamma \end{aligned} \quad (1.3)$$

荷電パイオンの崩壊によるミューオン生成

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu \end{aligned} \tag{1.4}$$

このようにしてカスケード的に大量の二次粒子群（電子、陽電子、ガンマ線、ミューオンなど）が生成される現象を空気シャワーと呼ぶ。空気シャワーの模式図を図 1.3 に示す。

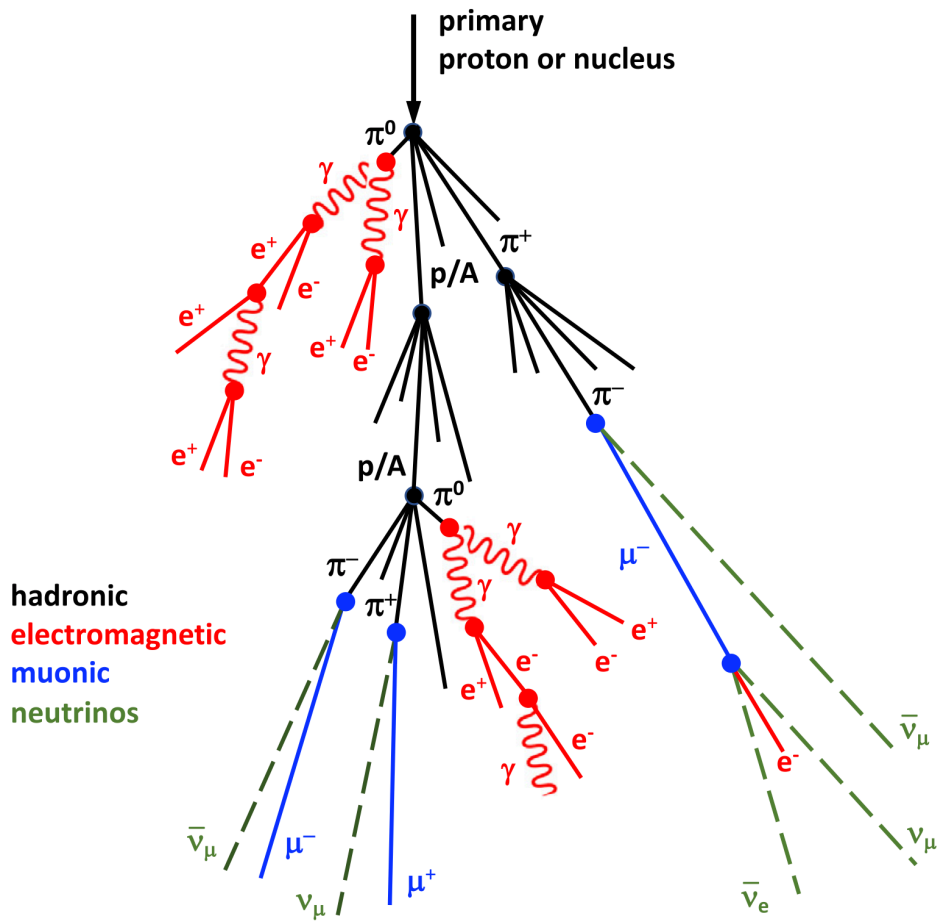


図 1.3 空気シャワー発達の様子を描いた図。([4] から掲載)

### 1.3.1 空気シャワーの発達

生成された二次粒子群は大気中を進むにつれてその数を増やすが、同時に個々の持つエネルギーは減少していく。粒子の持つ平均エネルギーが電磁相互作用の臨界エネルギーに達すると、粒子数は減少に転じる。粒子数が最大になるまでに進んだ大気の深さを  $X_{\max}$  という。

このような空気シャワーの縦方向（進行方向）の発達と、一次宇宙線の原子核種には密接な関係がある。エネルギー  $E_0$  を持つ質量数  $A$  の原子核は、エネルギー  $E_0/A$  を持つ核子  $A$  個の集合体として振る舞うと近似できる（重ね合わせの原理）。そのため鉄のような重い原子核は、陽子のような軽い原子

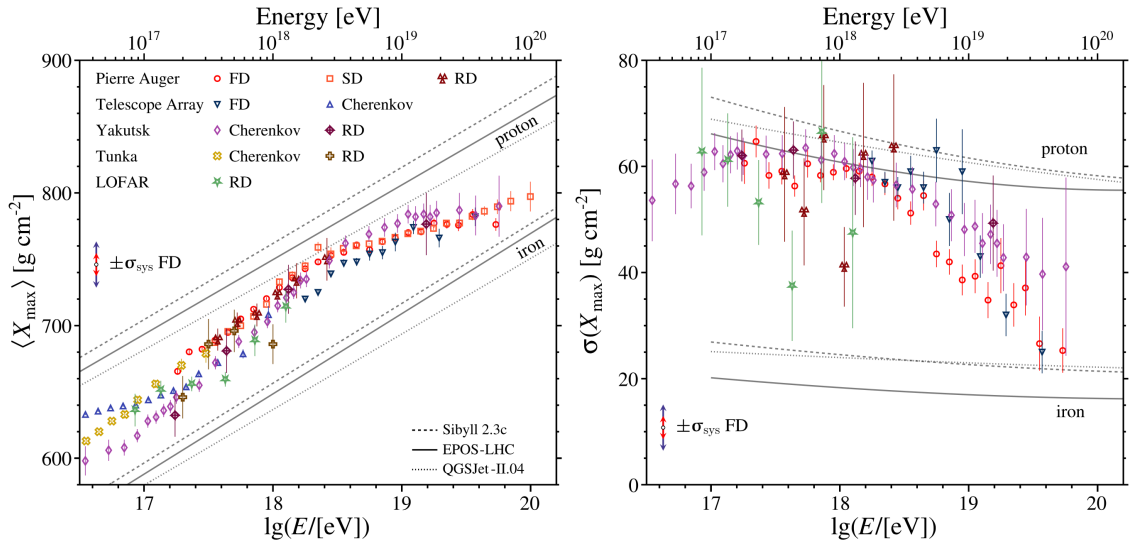


図 1.4  $X_{\max}$  についてのシミュレーションによる陽子と鉄および観測値との比較 ([5] から掲載)。左図は  $X_{\max}$  を、右図はその標準偏差を示す。

核と比較して核子当たりのエネルギーが低く、かつ大気中の原子核との相互作用断面積が大きいため、大気の上層で相互作用を開始し、より早い段階でシャワーを発達させる性質を持つ。その結果、図 1.4 の左図に示すように、鉄のような重い原子核ほど  $X_{\max}$  の値は小さく、陽子のような軽い原子核ほど  $X_{\max}$  の値は大きくなる傾向がある [6]。また、図 1.4 の右図に示すように、鉄の方が  $X_{\max}$  の幅が小さくなる。核種の違いは空気シャワーの横方向（進行方向に垂直な方向）分布にも現れる。ミューオンなどのハドロン成分は電磁成分に比べてエネルギー損失や散乱が少なく、シャワー軸から遠くまで到達する傾向がある。重い原子核ほど空気シャワー中のハドロン成分が多くなり、空気シャワーの横方向の広がりが大きくなる。

### 1.3.2 空気シャワー観測

宇宙線の持つ高いエネルギーは二次粒子群に分配され、円盤状になって地表へ到達する。この時の空気シャワーの横方向の広がり、 $10^{20}$  eV の宇宙線で約 10 km に達する。地表での宇宙線観測実験では、地表に数 km 間隔で検出器を並べることでこの空気シャワー粒子を捉え、その密度分布や到来時刻の情報から、宇宙線のエネルギーや到来方向を再構成している。

## 1.4 アマテラス粒子

宇宙線のエネルギースペクトルの正確な測定や、UHECR の起源探査のために、米国ユタ州で観測を行っているテレスコプアレイ (Telescope Array : TA) 実験で、宇宙線観測史上 2 番目となる極めて高いエネルギーを持つ宇宙線が観測された [7]。この宇宙線のエネルギーは  $244 \pm 29$  (stat.)  $^{+51}_{-76}$  (syst.) EeV ( $2.44 \times 10^{20}$  eV) と推定され、1991 年に同じく米国ユタ州のフライズアイ実験で観測された「オーマイゴッド粒子 ( $320 \pm 90$  EeV) [8]」に匹敵する。この宇宙線は観測時刻

が現地の明け方であったことから、日本神話の太陽神にちなんで「アマテラス粒子」と名付けられた (図 1.5)。

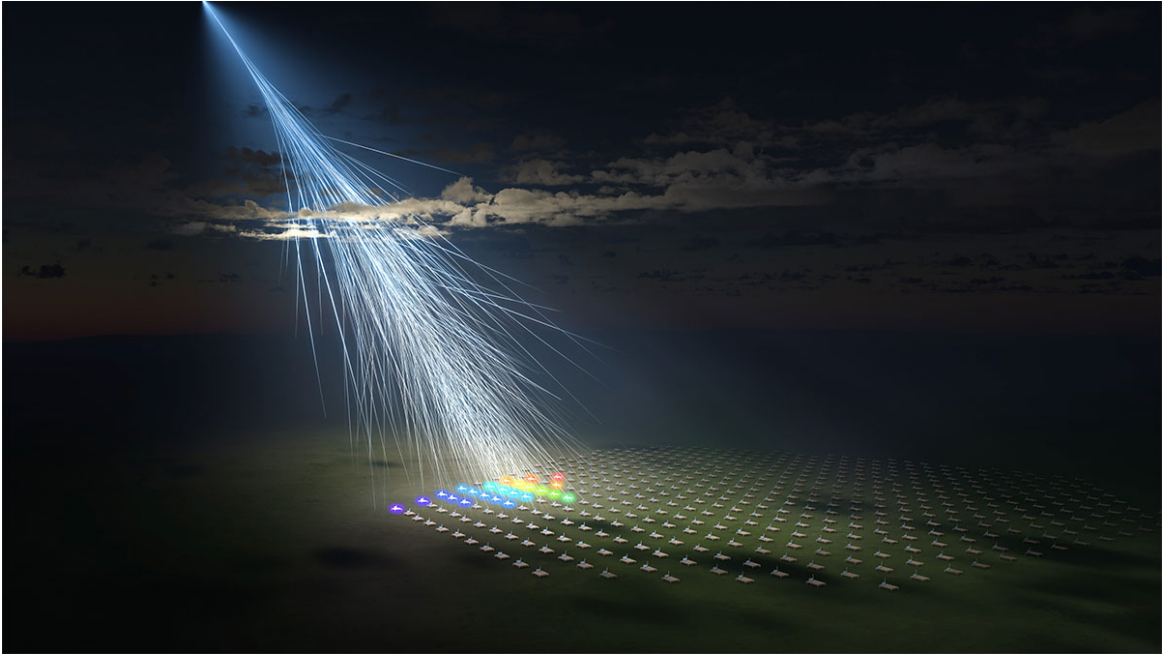


図 1.5 アマテラス粒子が観測された際のイメージ図 ([7] より掲載)。

#### 1.4.1 アマテラス粒子の到来方向

アマテラス粒子について、その極めて高いエネルギーだけでなく、到来方向も宇宙線研究において非常に重要である。解析の結果、アマテラス粒子の到来方向は、天の川銀河の円盤面に近い領域を指していることが判明した。しかし、図 1.6 に示すように、その到来方向の近傍には、宇宙線をこのようなエネルギーまで加速できるような候補天体（活動銀河核やスターバースト銀河など）が既知の範囲では存在しない。それどころか、銀河がほとんど存在しない「ローカル・ボイド (Local Void)」と呼ばれる空洞領域と一致している。また、アマテラス粒子のような極めて高いエネルギーを持つ宇宙線は、宇宙空間の光子と相互作用しパイ中間子を生成することでエネルギーを急速に失う [9, 10]。

$$\begin{aligned} p + \gamma_{\text{CMB}} &\rightarrow p + \pi^0 \\ p + \gamma_{\text{CMB}} &\rightarrow n + \pi^+ \end{aligned} \quad (1.5)$$

そのため起源は比較的近傍 (<~ 50 Mpc) にあるはずである。また、これほど高いエネルギーであれば、宇宙空間の磁場による進行方向への影響もそれほど受けず、到来方向はそのまま起源を指し示していると考えられる。しかし、観測結果は起源が存在すべき方向に候補天体が存在しないという矛盾を示している。

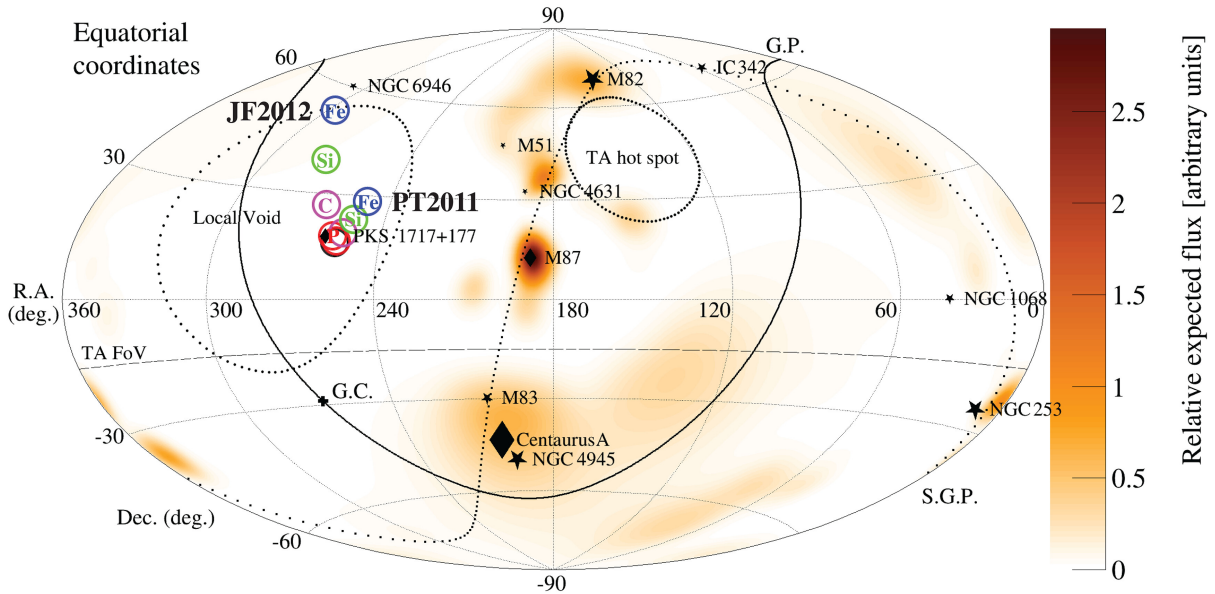


図 1.6 アマテラス粒子の到来方向を赤道座標で表した図。黒い円はアマテラス粒子の到来方向、その他の円は異なる 2 つの磁場モデルを用いた核種ごとの起源候補領域である。主要な起源天体も合わせて描いている。[7] から掲載。

## 1.5 原子核種の重要性

この矛盾を解決する手掛かりとなりうるのが、宇宙線の原子核種である。

### 1.5.1 磁場による偏向

宇宙線は荷電粒子であるため、加速源から地球に到達するまでに銀河系内外の磁場によるローレンツ力を受け、その進行方向が曲げられる。この時の偏向角  $\delta$  は、宇宙線のエネルギー  $E$  と電荷  $Z$  に依存し、近似的に以下の関係で表される。

$$\delta \propto \frac{Z}{E} \int B_{\perp} dl \quad (1.6)$$

ここで  $B_{\perp}$  は経路上の磁場の垂直成分、 $l$  は経路長である。この式は、宇宙線のエネルギー  $E$  が高いほど直進性が増す一方で、その電荷  $Z$  が大きいほど（重い原子核であるほど）磁場の影響を強く受け、進行方向を大きく曲げられることを示している。

本研究ではアマテラス粒子の原子核について、次の 2 つのシナリオを考える。

#### 陽子である場合 ( $Z = 1$ )

磁場による偏向は数度程度と小さく、到来方向はほぼ発生源を指しているとみなせる。しかし、その方向はボイドであるため、最高エネルギー宇宙線の起源として、未知の天体现象や、現在の物理学では説明できない標準模型を超えた物理（ダークマターの崩壊など）を考える必要が生じる [11]。

### 鉄である場合 ( $Z = 26$ )

陽子に比べて電荷が 26 倍大きいため、同じエネルギーであっても磁場によって大きく進路を曲げられる。この場合、見かけの到来方向がボイドであったとしても、実際の発生源はそこから数十度離れた位置にある既知の天体 (NGC 6946 など) である可能性が浮上する [12]。原子核種ごとの起源候補領域を図 1.7 に示す。

このように、アマテラス粒子の原子核種を特定することは、その起源探査の候補領域を決定し、未知の物理現象の可能性を検証する上で決定的に重要である。

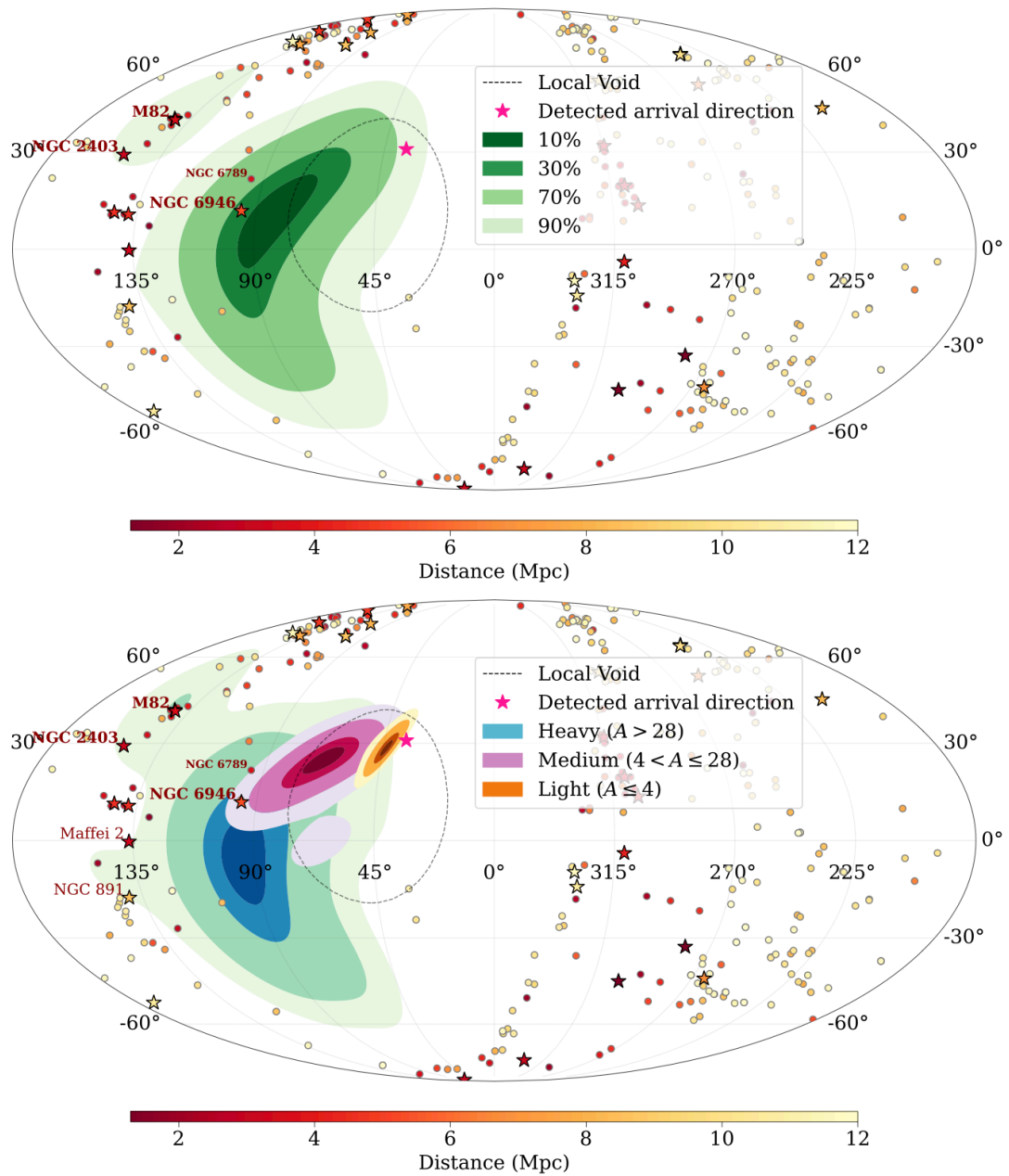


図 1.7 原子核種ごとのアマテラス粒子の起源候補領域を銀河座標で示した図。オレンジはヘリウムまでの軽い原子核、紫は窒素までの中間質量原子核、青はそれ以上の重い原子核を示す。(図は [Bourriche2024\_Amaterasu] より掲載)

## 第 2 章

# テレスコープアレイ実験

テレスコープアレイ (Telescope Array : TA) 実験は、北半球最大の宇宙線観測実験であり、日本、アメリカなど 9 つの国、地域の参加する国際共同実験である [13]。観測サイトはアメリカユタ州の砂漠地帯に位置し、約  $700 \text{ km}^2$  という広大な有効面積を有している。標高は約  $1400 \text{ m}$  (大気深さ約  $875 \text{ g/cm}^2$ ) であり、超高エネルギー宇宙線が作る空気シャワーが最大発達を迎える大気深さに近い地点での観測が可能である。

TA 実験では、地表検出器 (Surface Detector : SD) アレイと大気蛍光望遠鏡 (Fluorescence Detector : FD) という異なる 2 種類の検出器による観測を行っている。SD アレイは昼夜天候を問わない定常観測が可能で、地表に届いた空気シャワー中の荷電粒子を観測している [14]。一方で、FD は空気シャワー粒子による大気蛍光を観測する。空気シャワーを直接観測するため、エネルギー決定精度が高く、 $X_{\text{max}}$  も観測可能である [15, 16]。そのため宇宙線の原子核種の推定も可能だが、大気蛍光は微弱な光のため、よく晴れた月明かりのない夜にしか観測できないというデメリットがある。SD 観測の稼働率がほぼ 100 % であるのに対し、FD 観測は 10 % 程度である。アマテラス粒子の観測時は FD 観測が行えない明け方だったため、本研究では SD データのみを扱う。以下に SD について詳述する。

### 2.1 地表検出器アレイ

TA 実験の SD アレイは、507 台の SD を  $1.2 \text{ km}$  間隔で正方格子状に配置したものである。SD アレイの配置図を図 2.1 に示す。SD アレイの中心には大気透明度の測定や FD の較正を行うための Central Laser Facility (CLF) が設置されている。TA 実験のデータ解析における座標系はこの CLF の位置を原点として定義されており、本研究における SD の位置や空気シャワーのコア位置も、この座標系に従う。個々の SD は空気シャワー中の荷電粒子 (電子、陽電子、ミューオンなど) が通過した際に生じるシンチレーション光を検出する。本節では後述する機械学習における入力データとなる SD の波形データの特徴について述べる。

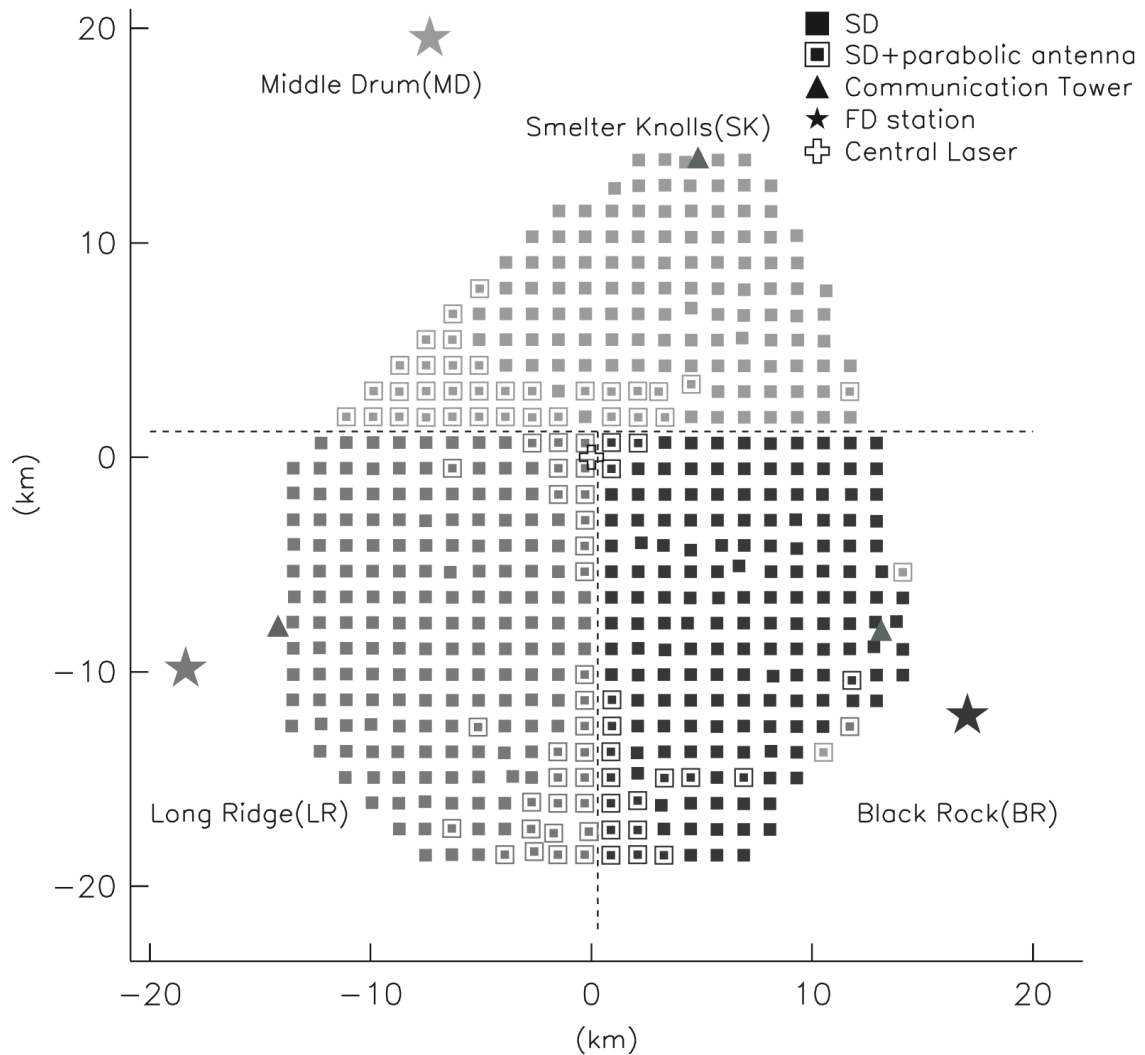


図 2.1 TA の検出器全体を表した図 ([14] より掲載)。中央の十字は CLF を、黒い四角は SD を、3つの星は FD ステーションを表す。

### 2.1.1 検出器の構造

SD の模式図を図 2.2 に示す。各 SD は面積  $3\text{ m}^2$ 、厚さ  $1.2\text{ cm}$  のプラスチックシンチレータを 2 層 (上層・下層) に重ねた構造を持つ。それぞれのシンチレータはタイベックシートで覆われ、層の間には厚さ  $1\text{ mm}$  のステンレスの板が挟まれており、上層と下層は光学的に分離されている。荷電粒子によるシンチレーション光は、シンチレータ表面の溝に埋め込まれた波長変換ファイバーによって集光され、光電子増倍管に導かれる。PMT は上層用と下層用で独立しており、SD ごとに 2 つの信号を取得する仕組みとなっている。2 層のコインシデンスを取ることで、PMT の熱ノイズなどのバックグラウンドを除去し、高エネルギー粒子の信号のみを記録することができる。

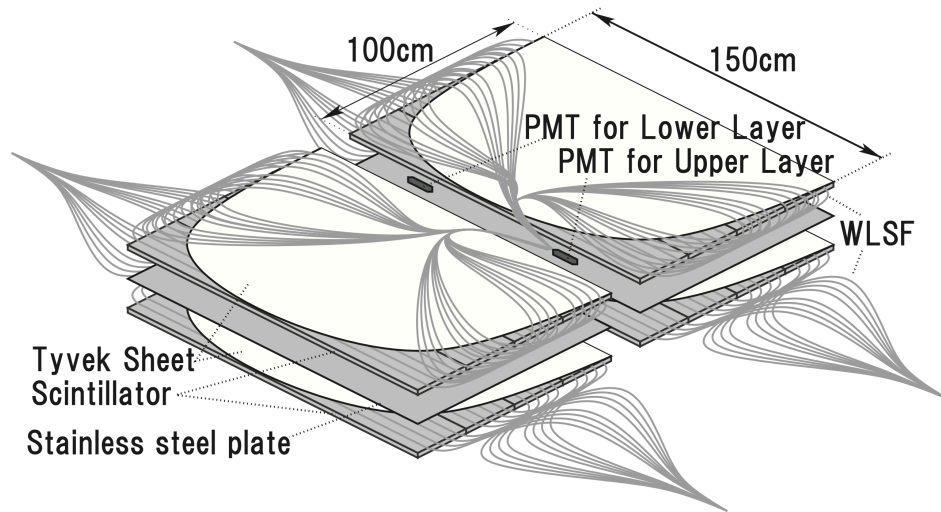


図 2.2 SD の構造を表した図 ([14] より掲載)。それぞれ光学的に分離されたシンチレータに、波長変換ファイバーが埋め込まれ PMT へと繋がれる。

### 2.1.2 PMT 信号と波形データ

PMT からのアナログ信号（シンチレータの発光量）は、エレクトロニクス回路において FADC (Flash Analog-to-Digital Converter) によって波形データへと変換される。サンプリングレートは 50 MHz (時間分解能 20 ns) であり、トリガー条件を満たすと、128 ビン (時間幅 2.56  $\mu$ s) の波形データが上層と下層独立に保存される。波形データは他数の粒子の重ね合わせであり、粒子ごとの特徴が波形に現れる。

電磁成分（電子、陽電子、ガンマ線）は数十 MeV と比較的エネルギーが小さく、空気シャワー中の 90 % 以上を占めるほど数が多いため、大多数はまとまって図 2.3 の左側にあるような波形を形成する。

ミュオンなどは、数は全体の 10 % 程度と少ないが、一つ一つのエネルギーが数 GeV 以上と高く、シンチレータをほぼ直線的に貫通する。その結果短い一定の発光量となり、図 2.3 の右側にあるような鋭く尖った波形となる。

波形データの例を図 2.3 に示す。鉄が作る空気シャワーでは陽子が作る空気シャワーに比べミュオン成分が多くなるため、SD に記録される波形データは鋭いピークが多いものになると予想される。

取得された信号 (FADC count) を物理的なエネルギー基準に換算するために、TA 実験では、1 個のミュオンがシンチレータを垂直に貫通した際に落とすエネルギー損失量を「1 MIP (Minimum Ionizing Particle) 相当」と定義し、これをエネルギー較正の基準単位としている。各 SD では約 10 分ごとにバックグラウンドのミュオン粒子を用いた較正が行われており、温度変化などによる PMT ゲインの変動を補正している。

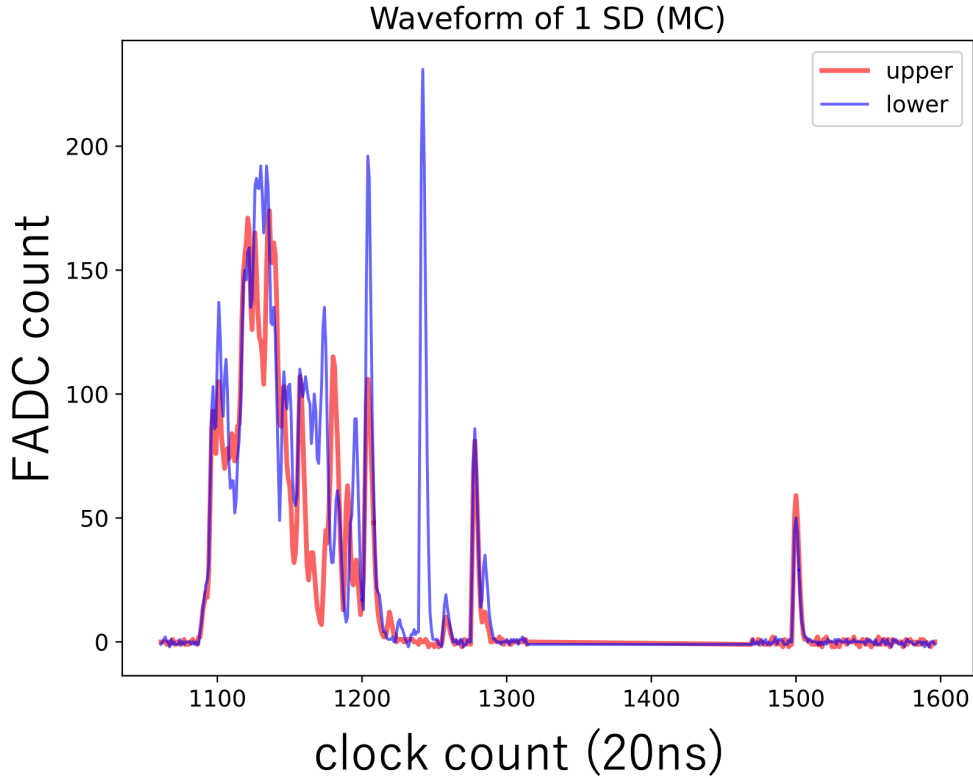


図 2.3 SD に記録される波形データの例。

## 2.2 標準的な解析手法と本研究のアプローチ

TA 実験における SD アレイデータの標準的な解析と、本研究での扱いについて述べる [17]。

### 2.2.1 幾何学的再構成

まず、複数の SD の信号立ち上がり時刻（タイミング情報）を用いて、空気シャワーの到来方向（天頂角・方位角）を決定する。空気シャワーフロントプレーンは光速に近い速度で移動する薄い円盤として近似でき、各 SD への到達時間の差からシャワーの到来方向を再構成できる。

### 2.2.2 LDF フィッティングとエネルギー再構成

各 SD が取得した波形データ（FADC count）を時間積分し、総電荷量を粒子数密度（VEM 単位）に変換する。空気シャワーのコアからの距離  $r$  に対する粒子数密度の分布を横方向分布関数（Lateral Distribution Function: LDF）を用いて計算する。TA 実験では、AGASA 実験で用いられた LDF をベースにした関数 [18] を用いてフィッティングを行い、シャワー軸から 800 m 地点における粒子数密度  $S_{800}$  を求める。モンテカルロシミュレーションによって作成されたエネルギー推定テーブルを用いて、 $S_{800}$  と天頂角から宇宙線のエネルギーを決定する [19]。

この解析では、各 SD の持つ波形データは上層、下層を合わせた積分値が計算され、さらに  $S_{800}$  という 1 つの値に集約される。しかし本研究では、波形データに、空気シャワーの粒子群が SD アレイに到達した時間分布（空気シャワーフロントプレーンの厚みや構造）という物理的に重要な情報が含まれていると考えた。従来解析では主に積分値のみが使われてきたが、本研究ではこの時間発展形状そのものも機械学習用の特徴量として利用する。

### 2.2.3 本研究での SD データの扱い

TA では SD を用いた宇宙線の質量推定がすでに行われており、結果が報告されている [20, 21]。先行研究では T ASD 全体を用いた解析が行われたが、本研究ではアマテラス粒子単体に特化し、個々の SD の波形データを用いた機械学習を以下の 2 段階に分けて行う。

#### 積分値を用いた CNN による原子核種推定

4 章では  $S_{800}$  を計算する際に使われる SD ごとの波形データの積分値と到来時刻と SD アレイの格子状の配置を 1 枚の画像に見立てて、CNN を用いた特徴抽出を行う。本研究では、個々の SD の信号値の特徴を最大限に生かすため信号値は FADC count のまま使用し、上下層の積分値の平均のみを使用し、空気シャワーフロントプレーンの持つ横方向の空間的な特徴を用いた原子核種推定を行う。

#### 多次元特徴量を用いたランダムフォレストによる原子核種推定

5 章では、4 章の結果を受け、より詳細に波形データを解析するために、特徴量抽出ライブラリを用いて取得した多数の特徴量を用いたランダムフォレストによる原子核種推定を行う。上下層それぞれの波形を独立に使用し、空気シャワーフロントプレーンによる SD 波形データ個別の詳細な情報を活かした解析を行う

## 第 3 章

# 機械学習用空気シャワーイベントの生成

本研究の機械学習は、正解となる原子核種が判明している空気シャワーシミュレーションイベントを用いた教師あり学習である。モデルはシミュレーション空気シャワーの観測結果から原子核種固有の特徴を学習し、構築した学習モデルでアマテラス粒子の観測結果から原子核種を推定する。

### 3.1 アマテラス粒子の再構成

本研究では、機械学習用にアマテラス粒子の観測結果を再現するイベントを使用する。そのため空気シャワーシミュレーションの際のパラメータ設定に、TA 実験の標準解析 [19] によるアマテラス粒子の再構成結果を使用する。アマテラス粒子の図を 3.1、アマテラス粒子の主要な再構成パラメータは表 3.1 に示す。以下に各パラメータについて述べる。

表 3.1 アマテラス粒子の再構成データ

パラメータ名	値
エネルギー	$244 \pm 29 \text{ EeV}$
天頂角	$38.6 \pm 0.4^\circ$
方位角	$206.8 \pm 0.6^\circ$
空気シャワー中心 $x$ 座標 (CLF を中心とする)	$-9471 \pm 23 \text{ m}$
空気シャワー中心 $y$ 座標 (CLF を中心とする)	$1970 \pm 31 \text{ m}$
$S_{800}$	$530 \pm 57$

#### エネルギー

SD アレイによる再構成のエネルギー決定は、シミュレーションを用いて作成したエネルギー決定テーブルを用いて行われる。TA 実験では、エネルギー決定精度の高い大気蛍光望遠鏡 (FD) を用いた再構成手法によるエネルギーと合わせるために、SD アレイによる再構成で得られたエネルギーに 1/1.27 をかけた値を宇宙線のエネルギーとしている [19]。そのためシミュレーションに与える値は 1.27 倍された値となる。

(A) Surface detector array of TA

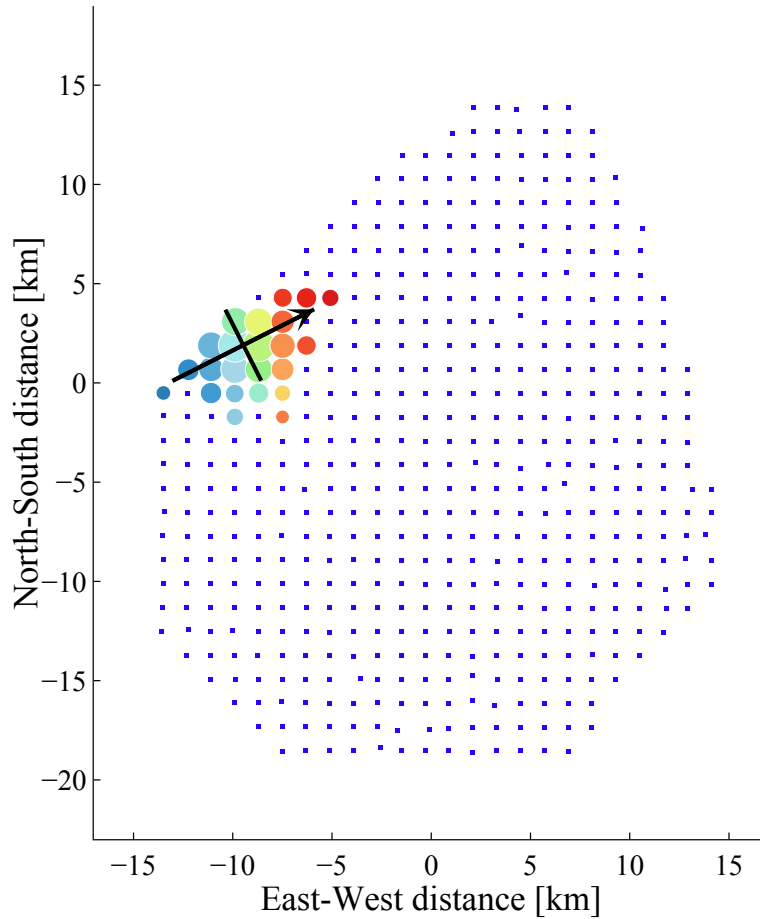


図 3.1 アマテラス粒子の観測結果を可視化したもの。円の大きさは粒子数、色は到来時刻差、矢印は到来方向の方位角を表す。

### 天頂角、方位角

宇宙線の到来方向は天頂角および東を 0 度とした反時計回りの方位角を用いて表す。

### 空気シャワー中心座標

空気シャワー中心の座標は、2.1 で述べたように SD アレイの中心である CLF を 0 としている。

### $S_{800}$

$S_{800}$  とは、空気シャワー軸から 800 m 地点での空気シャワーの粒子数密度であり、望遠鏡アレイ実験において宇宙線のエネルギーを決定する際に使用するパラメータである。空気シャワーシミュレーションの設定値ではないが、宇宙線のエネルギーを設定する際の指標として用いる。

空気シャワーイベントの生成は、宇宙線と大気中の原子核との相互作用による空気シャワー生成シ

ミュレーションと、SD アレイでの応答シミュレーションの2段階に分かれている。空気シャワーイベントの生成について、本研究では、望遠鏡アレイ実験で使用しているシミュレーションプログラム [22] を使用し、アマテラス粒子の検出結果を再現するイベントを生成するためにパラメータの調整のみを行う。

本研究ではこれに加えて、得られたシミュレーションデータについて、アマテラス粒子に用いたものと同じ再構成プログラムによる再構成を行い、各パラメータの再構成結果について、シミュレーションデータと観測データとの比較を行うことで、機械学習用のデータとして、アマテラス粒子の検出結果を再現できているかの追加検証を行う。

## 3.2 空気シャワー生成シミュレーション

### 3.2.1 COsmic Ray Simulations for KASCADE (CORSIKA)

宇宙線による空気シャワー生成シミュレーションには CORSIKA [23] (CORSIKA-73500) を用いる。CORSIKA は高エネルギー宇宙線が大気に入射した際に作る空気シャワーの全粒子を追跡し、詳細にシミュレートするプログラムである。空気シャワー粒子と大気との相互作用を計算する際に使用するハドロン相互作用モデルについては、望遠鏡アレイ実験では現在高エネルギー帯 (80 GeV 以上) で QGSJET-II-03、低エネルギー帯 (80 GeV 以下) では FLUKA を使用しているが、本研究では高エネルギー帯でより新しいハドロン相互作用モデルである EPOS-LHC [24]、QGSJET-II-04 [25] を用いて空気シャワーシミュレーションを行う。

### 3.2.2 空気シャワーパラメータの設定

本研究の空気シャワーシミュレーションでは、宇宙線のパラメータのうち一次宇宙線のエネルギーと到来方向の天頂角のみを設定する。本研究ではアマテラス粒子の再構成結果から得られた値を使用する。しかし、エネルギーを設定する際に2つの問題がある。

まず、アマテラス粒子の観測データから得られたエネルギーである 244 EeV は、現在の再構成プログラムで得られたエネルギーの 309 EeV に 3.1 による補正をかけたものである。しかし、現在の再構成プログラムは QGSJET-II-03 のシミュレーションデータを元に作成されており、本研究の空気シャワーシミュレーションでは、EPOS-LHC、QGSJET-II-04 を用いるため、これに加えて相互作用モデルごとのバイアスも考慮する必要がある。

次に原子核種について、1.3.1 で述べたように、鉄と陽子では核子の数の違いから空気シャワーが異なる広がり方をするため、地表での検出結果に違いが生まれる。

これらを考慮するために、本研究ではシミュレーションでのエネルギーの設定について、相互作用モデル、核種ごとに、SD データと深い関わりを持つパラメータである  $S_{800}$  がアマテラス粒子の再構成結果と一致することを指標とする。地表での空気シャワー粒子の分布がアマテラス粒子の観測結果を再現するようにシミュレーション時に設定するエネルギーを調整する。

天頂角成分についてはアマテラス粒子の再構成パラメータによる固定値を用いる。

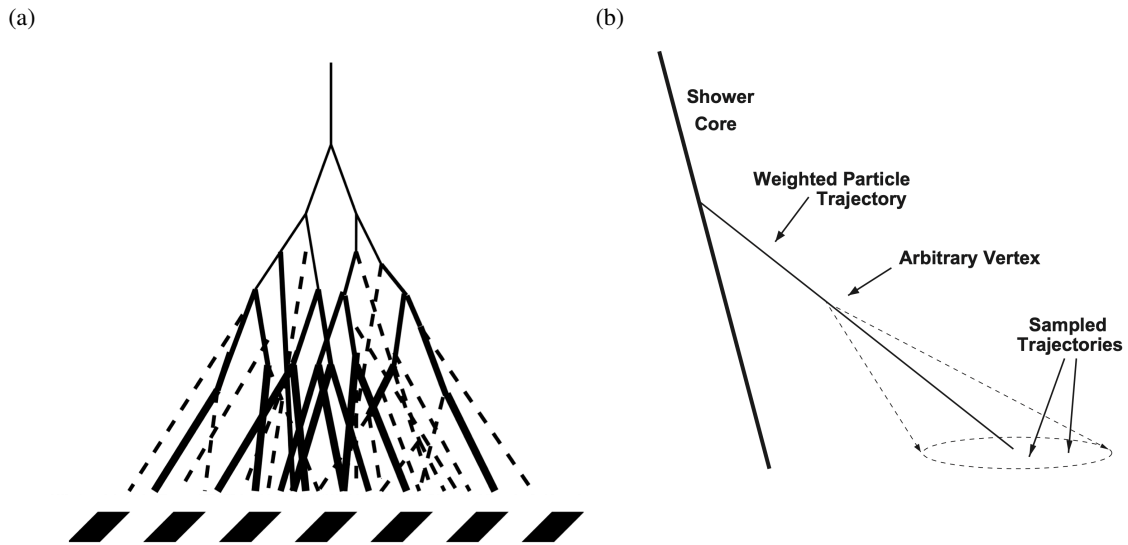


図 3.2 左図はシニングを示す [27]。実線が代表粒子、点線が計算が行われない粒子を表す。右図がデシニングを示す [28]。代表粒子について、ガウス分布に従い、付近の粒子が復元される。

### シニングとデシニング

$10^{18}$  eV を超えるような極めて高いエネルギーの宇宙線による空気シャワーでは膨大な数の粒子が発生し、すべての粒子について計算を行うと計算時間が非常に大きくなってしまふ。そのためシャワー生成中では代表粒子を決め、その他の粒子の計算を行わないシニングという処理と、シャワー生成後に代表粒子付近に粒子を増やすデシニングという処理を行う [26]。シニングとデシニングについて図 3.2 に示す。

到来方向の方位角成分についても CORSIKA で設定できるが、TA 実験ではまず西を  $0^\circ$  とした反時計回り  $45^\circ$  に固定し空気シャワーを生成し、後述する検出器シミュレーションの際に調整しており、本研究でも同様の手法をとる。

アマテラス粒子のような極めて高いエネルギーを持つ空気シャワーシミュレーションを行う際は、シニングやデシニングを行ったとしても、1 シャワーにつき 1 GB ほどとなり、大量のシャワー生成を行うことは困難である。そのため、作成する空気シャワーの数を減らし、後述する検出器シミュレーションを行う際にシャワーの使い回しを行うことでイベント数を確保する。

本研究では、相互作用モデル、核種ごとに 100 シャワーずつ、合わせて 400 シャワーを生成した。

### 3.3 SD 応答シミュレーション

SD 応答シミュレーションに使用するための、空気シャワー粒子が検出器に入射した際の応答テーブルを、シミュレーションプログラムの GEANT4 [29] で作成しておく。これを用いて、作成した空気シャワーに対して、宇宙線の到来方向の方位角、空気シャワー中心位置を設定する。空気シャワー生

成と同様に、本研究では機械学習用にアマテラス粒子の検出結果を再現するためのパラメータの設定のみを行う。

### 3.3.1 パラメータの設定

作成した空気シャワーデータはテレスコープアレイ実験が行われる標高における  $6 \times 6 \text{ m}^2$  の範囲ごとの、検出器内での空気シャワー粒子のエネルギー損失量として得られる。これに対して、本研究では検出器アレイ側を移動、回転させることで、宇宙線到来方向の方位角成分、空気シャワー中心の  $xy$  座標をアマテラス粒子を再現するように設定する。

到来方向の方位角成分はアマテラス粒子の再構成パラメータによる固定値を使用する。

空気シャワー中心の  $xy$  座標については、アマテラス粒子の再構成パラメータの不定性である半径約  $30 \text{ m}$  の円内で一様にランダムな値を使用する。ここで、イベントごとの結果のランダム性を考慮しつつ機械学習用のデータ量を確保するため、空気シャワーごとに 100 回の使い回しを行う。同じシャワーを中心位置を変えながら繰り返し使用することで、生成する空気シャワーの数が少なくても、機械学習に必要なアマテラス粒子の観測結果と再現するイベントを確保できる。

シミュレーションで設定するパラメータは表 3.2 に示す。

表 3.2 シミュレーション設定

ハドロン相互作用モデル	EPOS-LHC		QGSJET-II-04	
宇宙線核種	陽子	鉄	陽子	鉄
エネルギー	261 EeV	240 EeV	300 EeV	272 EeV
天頂角	38.6°			
方位角	206.8°			
空気シャワー中心 $x$ 座標	−9471 m			
空気シャワー中心 $y$ 座標	1904 m			

## 3.4 シミュレーションパラメータの調整

アマテラス粒子の核種推定を行うためには、アマテラス粒子を精巧に再現したシミュレーションデータが学習に使われる必要がある。そのため機械学習データセットを作成する前に、シミュレーションデータがアマテラス粒子の観測データを再現できているか確認を行う。シミュレーションデータは実際の実験と同様に、検出器に入射した荷電粒子によるシンチレータの発光量を電気信号に変換したものである。

### 3.4.1 再構成パラメータの比較

作成したシミュレーションデータに対して、アマテラス粒子と同様の空気シャワー再構成を行い、得られたパラメータの比較を行った。

図 3.3 と図 3.4 に、EPOS-LHC で作成した陽子と鉄のシミュレーションデータにおける各パラメータの再構成結果を示す。各パラメータごとにアマテラス粒子とシミュレーションデータとで再構成結果に違いが見られる。これは現在の再構成プログラムで再構成を行う際のバイアスによるものと考えられる。このバイアスの原因は現時点では特定できていないが、機械学習用のデータセットにはアマテラス粒子と似たイベントを使用することが好ましいため、このバイアスを考慮し、再構成後のパラメータがアマテラス粒子の再構成結果と一致するように、シミュレーションに与えるパラメータを調整する。具体的には、それぞれのパラメータについて差分を取り、表 3.3 に示すパラメータ設定で前述のシミュレーションを再度行う。 $S_{800}$  のズレについては、シミュレーションとアマテラス粒子との再構成結果の比の分、設定するエネルギーの値を変えることで調整を行う。

図 3.5 と図 3.6 に示す調整後のパラメータを見ると、ほとんどのパラメータで再構成結果での一致を確認できるが、到来方向の方位角成分ではわずかなずれが残っている。これはシャワー中心の位置を変えたことによるものと思われるが、誤差の範囲で一致しているとして、本研究ではこのシミュレーションデータを用いることにする。

本研究では、アマテラス粒子の核種推定と並行してハドロン相互作用モデルごとの結果の違いの評価も行う。そのため学習モデルはハドロン相互作用モデルごとに独立したものを作成する。それぞれの学習モデルで約 20000 イベント（陽子：鉄 = 1：1）が使用され、ハドロン相互作用モデル間の相互評価も行う。

得られた信号を用いた機械学習用データセットの作成については、手法ごとに各章で述べる。

表 3.3 シミュレーション設定（調整後）

ハドロン相互作用モデル 宇宙線核種	EPOS-LHC		QGSJET-II-04	
	陽子	鉄	陽子	鉄
エネルギー	259 EeV	231 EeV	300 EeV	269 EeV
天頂角	38°			
方位角	207.4°			
空気シャワー中心 $x$ 座標	-9267 m			
空気シャワー中心 $y$ 座標	1970 m			

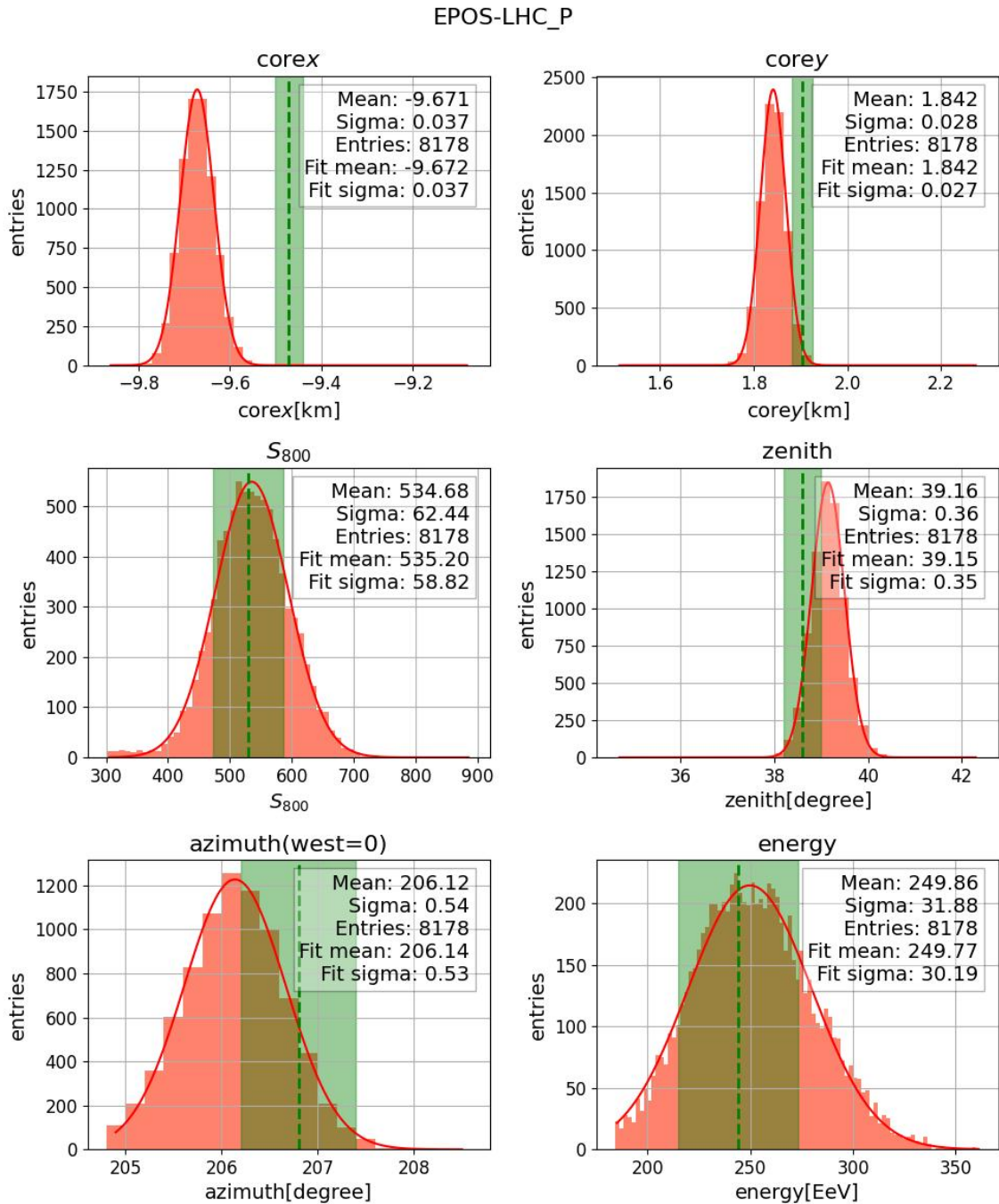


図 3.3 EPOS-LHC を使った陽子のシミュレーションイベントの再構成結果（調整前）。上段左から空気シャワーの中心の  $x$  座標、 $y$  座標、中段に  $S_{800}$  と天頂角、下段に方位角と一次宇宙線のエネルギーを示す。ヒストグラムはそれぞれガウシアンでフィットしている。緑色の点線とバンドはアマテラス粒子の再構成結果とその不定性である。

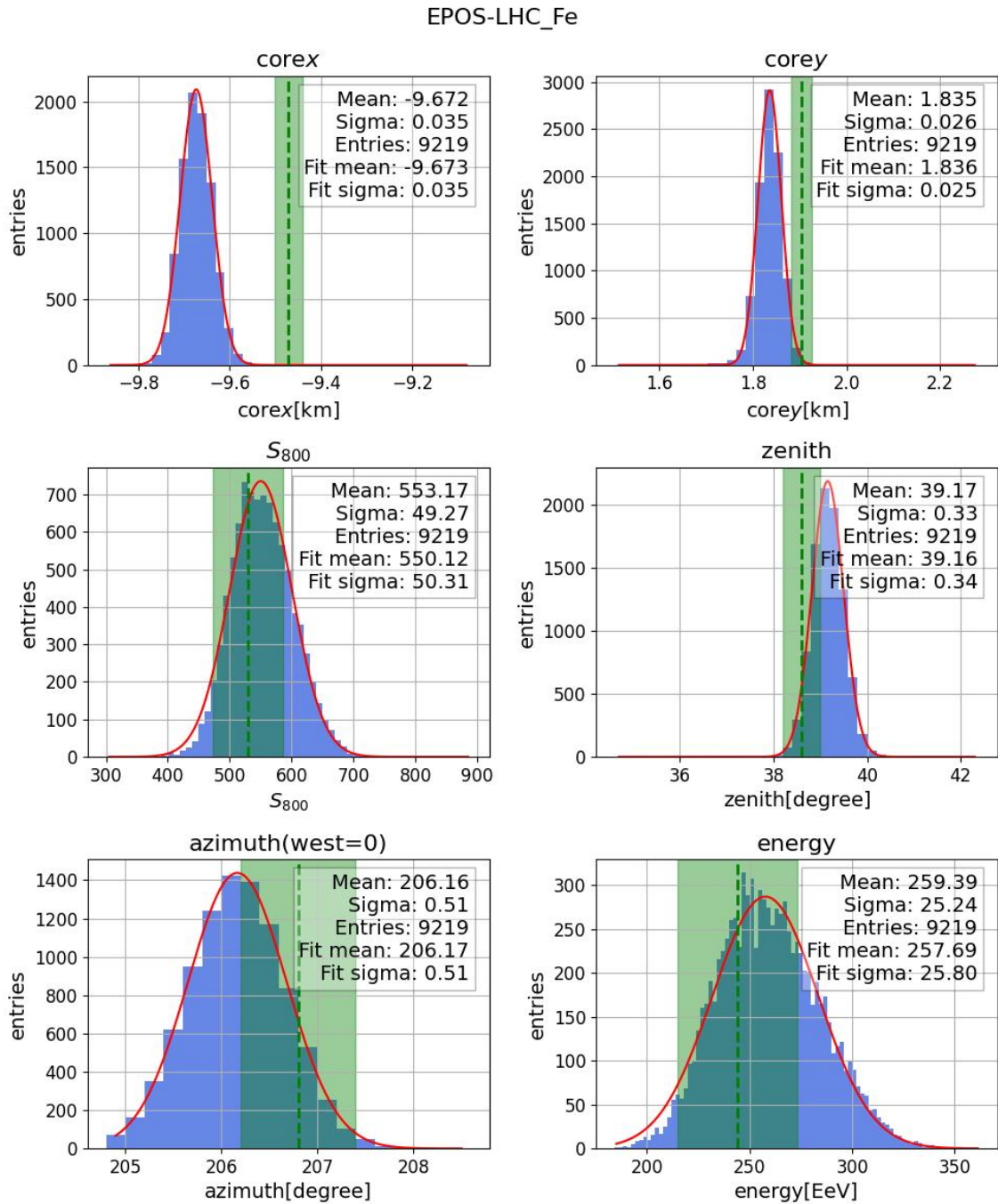


図 3.4 EPOS-LHC を使った鉄のシミュレーションイベントの再構成結果（調整前）。

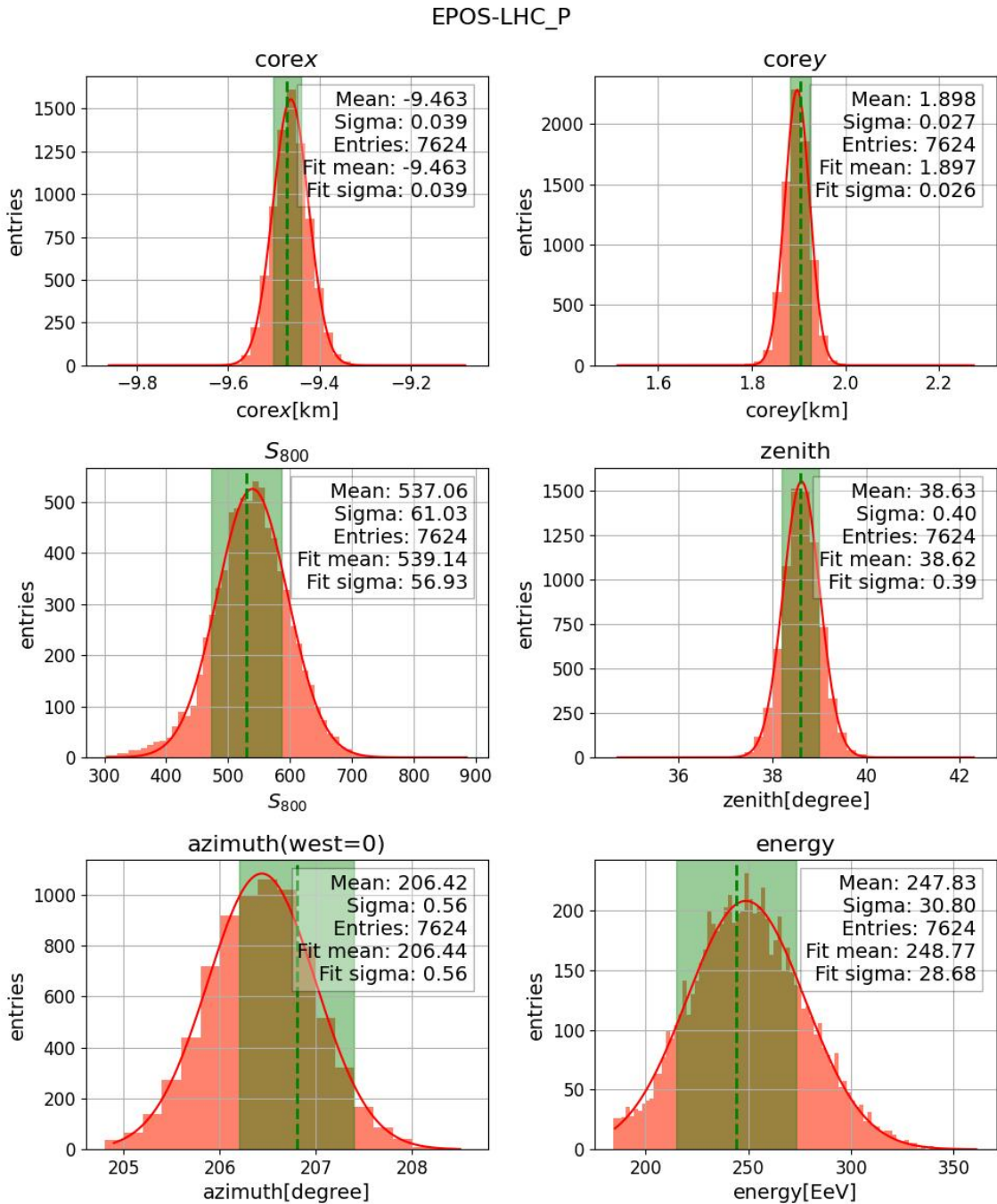


図 3.5 EPOS-LHC を使った陽子のシミュレーションイベントの再構成結果（調整後）。上段左から空気シャワーの中心の  $x$  座標、 $y$  座標、中段に  $S_{800}$  と天頂角、下段に方位角と一次宇宙線のエネルギーを示す。ヒストグラムはそれぞれガウシアンでフィットしている。緑色の点線とバンドはアマテラス粒子の再構成結果とその不定性である。

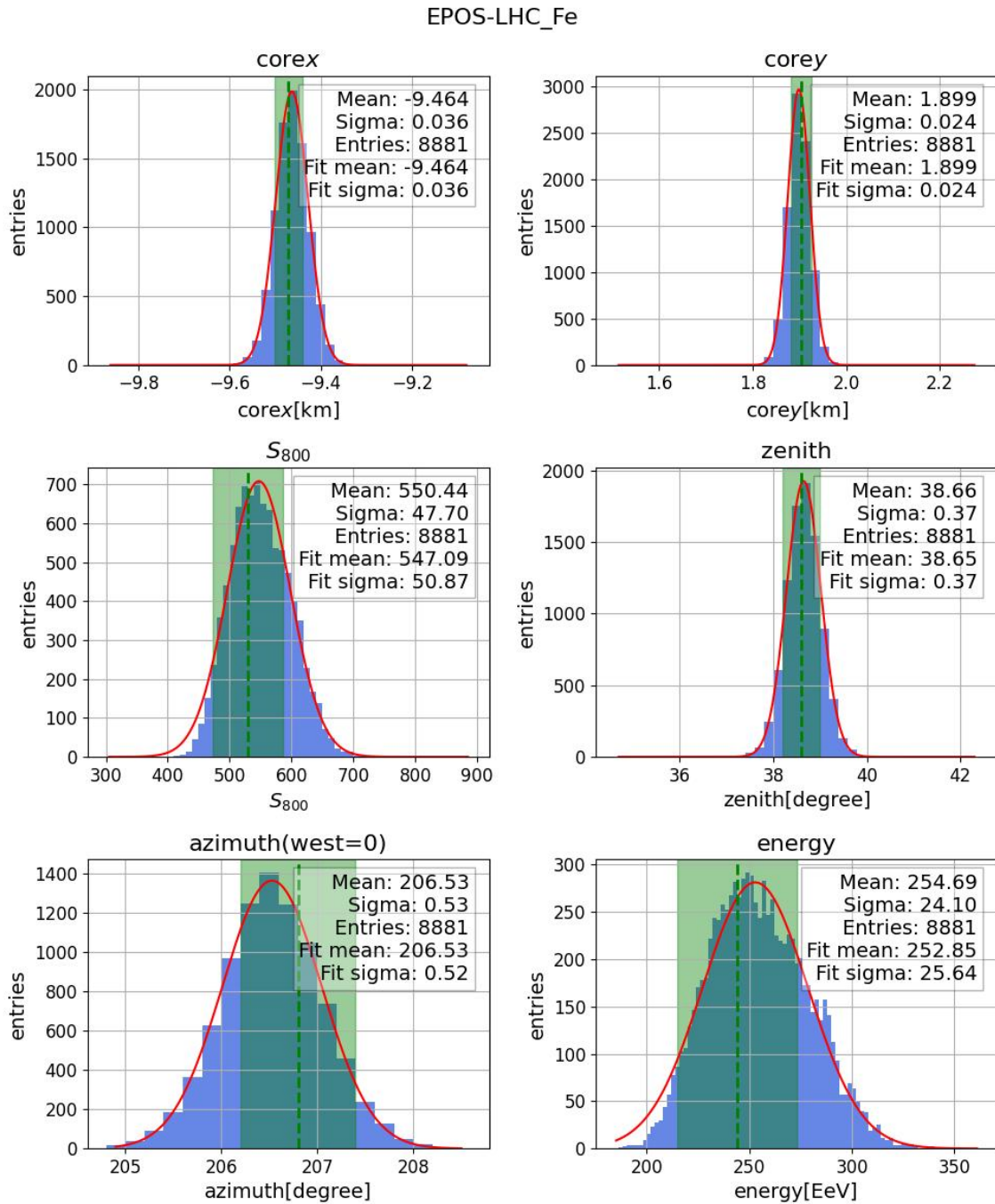


図 3.6 EPOS-LHC を使った鉄のシミュレーションイベントの再構成結果（調整後）。

## 第 4 章

# CNN を用いた原子核種推定

### 4.1 畳み込みニューラルネットワーク

TA 実験の SD アレイは格子状に並べられているため、それぞれの SD が持つ特徴量をまとめたデータは画像のように扱えると判断した。本章では、画像認識分野で広く使われる畳み込みニューラルネットワーク (Convolutional Neural Network : CNN) を導入することで、空気シャワーの空間的な特徴を抽出する。研究の初期段階として個々の SD では「到来時刻」と「信号の積分値」の 2 つの特徴量に絞ってデータセットとして使用する。

以下に NN および CNN について述べる。

#### 4.1.1 NN

ニューラルネットワーク (Neural Network : NN) は人工知能の一種であり、人間の脳の神経回路 (ニューロン) を模したシステムを用いて入力データの特徴を学習し、分類や出力の精度を向上させる機械学習アルゴリズムである [30]。NN は入力層、隠れ層、出力層に分かれており、入力層は隠れ層と、隠れ層はさらに次の隠れ層もしくは出力層と繋がれており、それぞれの層を構成するノード間の重みづけを何度も繰り返し更新することで、学習モデルの精度を向上させる。

#### 4.1.2 CNN

CNN (畳み込みニューラルネットワーク) とは NN を用いた深層学習 (DNN) の一種である [31]。アレイ状のデータに対して複数のカーネルを用いた畳み込みによる特徴抽出を行い、抽象化されたデータを 1 次元に変換したのち、NN の入力層として使用する。CNN の模式図を図 4.1 に示す。

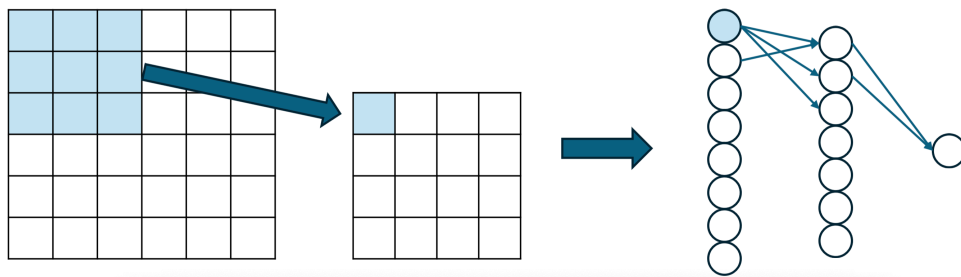


図 4.1 CNN による学習の模式図。2D アレイに対して畳み込みによる特徴抽出を行い、NN の入力層とする。本研究では最終的に 1 つの出力を持つ。

## 4.2 データセットの作成

### 4.2.1 アレイデータ

本章では、アマテラス粒子が観測された点付近の  $8 \times 8$  のアレイを用いる。しかし、アマテラス粒子が到来したのは SD アレイの端であり、 $8 \times 8$  を考える際に SD が存在しない場所がある。これについては図 4.2 のように仮想の SD を配置し、信号の存在しない SD と同じ扱いとする。

### 4.2.2 各 SD のもつデータ

入力データの各ピクセル (SD) はそれぞれ波形データの開始時刻 (到来時刻) と波形の積分値 (上層と下層の平均値) の 2 つのデータを持つ。データセットの前処理として 2 つの特徴量についてそれぞれ正規化を行い、0 から 1 の範囲の数値データに変換する。この処理は、モデルの学習において、特徴量の持つ数値の大きさによるバイアスが生じてしまうのを防ぐために行われる。

### 4.2.3 データセットまとめ

1 つのイベントは SDID 64 行に到来時刻と波形データの積分値の 2 列のデータを持った csv ファイルにまとめられ、ハドロン相互作用モデル、核種ごとに約 10000 ファイルずつのデータセットを使用する。

## 4.3 学習モデル

学習モデルはハドロン相互作用モデル間の比較を行うためにそれぞれ独立したものを構築する。そのためそれぞれの学習モデルに使用されるデータセットは約 20000 イベントずつとなる。1 割をテストデータとして使用し、残りの 9 割のデータをトレーニングデータとして学習を行う。学習およびテストに使われるデータの特異性がモデルの性能に及ぼす影響を緩和させるために、全データを 10 分割し、それぞれが 1 割のテストデータの役割を果たすように 10 個の Fold を作成し、すべてのデータ

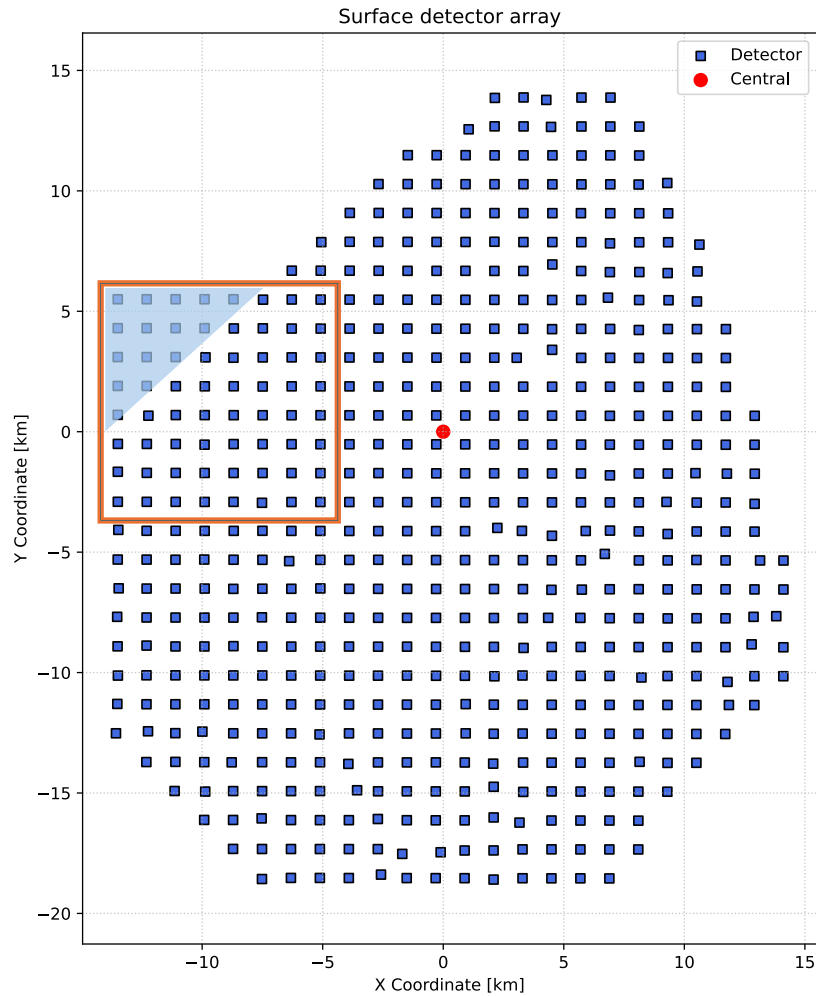


図 4.2 T ASD の配置と学習に用いる範囲。CNN での学習のために仮定の SD を配置している。

がテストデータとして評価されるよう 10 回の独立した学習を行う「10-fold cross-validation」を採用する。

#### 4.3.1 畳み込みによる特徴量抽出

Fold ごとに、まずデータセットからの特徴量抽出を行う。8 × 8 のアレイに対して 3 × 3 のカーネルを用いた畳み込み（アレイサイズが変わらないように外側に 0 埋めを行う）と、抽象化のための MaxPooling（2 × 2 のうち一番大きい値 1 つのみを使う）を行うことで、最終的な入力データは 4 × 4 × 特徴量 2 つ × カーネル数、これを 1 次元データに変換し、NN の入力層とする。

#### 4.3.2 モデルの学習

本研究ではモデルの出力を「Ironness（鉄らしさ）」と定義し、0 を陽子、1 を鉄とする。

NN の隠れ層は 1 層とし、出力層は Ironness を出力する。入力時には陽子のイベントは Ironness が

表 4.1 パラメータ探索範囲の設定

ハイパーパラメータ	探索範囲
畳み込みカーネル数	16, 32, 64
隠れ層ノード数	16, 32, 64
ドロップアウト率	0.2 - 0.5
学習率	$10^{-2}$ - $10^{-5}$
Epoch 数	0 - 50

0、鉄のイベントは Ironness が 1 であるとしてモデルに学習させる。学習モデルはまずトレーニングデータの中から一部のイベントを検証用に分ける（検証用データ）。残ったイベントを用いて 1 Epoch ごとに「バッチサイズの大きさだけイベントを取り出し、イベント内の特徴量使って予測を行い正解との差を見る。結果から 1 イベントが持つ特徴量ごとの重みづけを更新する」という動作を、トレーニングデータがなくなるまで続ける。各 Epoch の終わりには分けておいた検証用データを未知のイベントとした検証を行い、その時点でのモデルの性能評価を行う。

### 4.3.3 Optuna を用いたパラメータ探索

モデルの推定精度を高めるために、機械学習のハイパーパラメータを適切な値に設定する必要がある。本研究では表 4.1 に示すパラメータ探索のために特徴量探索ライブラリ「Optuna」を使用した。

#### Optuna

Optuna とは設定した範囲内のパラメータから最も精度が向上するものを自動的に探索する Python パッケージである [32]。

パラメータ探索は正解率（陽子の鉄らしさを 0.5 未満、鉄の鉄らしさを 0.5 以上と出力した場合を正解とする）を最大にするように全イベントの半分を使って行われる。見つかった最良のパラメータを用いて全データを使った学習を行い、10 Fold 分のテストデータでの予測結果、アマテラス粒子への適用結果を得る

### 4.3.4 学習モデルまとめ

コードの作成は Python を用いて行う。

学習モデルは NN などの高度な処理に適した Tensorflow.Keras[33] を用いて構築し、最適化関数は NN での学習で広く使われる Adam[34]、モデルの予測正解のずれの評価を行う損失関数には交差エントロピー誤差 [30] を使用し、陽子または鉄を判別する二値分類による評価を行う。Optuna のパラメータ探索をふまえた機械学習モデルの設定は表 4.2 のとおりである。

表 4.2 機械学習モデル設定

フレームワーク	Tensorflow.keras	
ハドロン相互作用モデル	EPOS-LHC	QGSJETII-04
最適化関数	Adam	
損失関数	交差エントロピー誤差	
畳み込みカーネル数	32	16
隠れ層ノード数	16	64
ドロップアウト率	0.319	0.498
学習率	$1.68 \times 10^{-3}$	$1.24 \times 10^{-3}$
Epoch 数	19	27

## 4.4 結果

以下に CNN を用いた機械学習の結果について述べる。

### 4.4.1 機械学習モデルの評価

#### 学習曲線

学習曲線とは CNN の学習の進行を可視化したものであり、横軸に Epoch、縦軸に Accuracy や Loss をとったグラフで描かれる。本研究では 10 回の独立した学習を行うため、学習曲線は重ね書きしている。

図 4.3 の学習曲線において、トレーニングデータでの Accuracy の向上、Loss の減少が見られるのに対してテストデータでは不安定な推移が見られる。これには 2 つの原因が考えられ、一つは学習不足であり、Accuracy、Loss とともに一定の Epoch 以降改善が見られない点からモデルの性能が頭打ちになっていることがわかる。もう一つとして、モデルがデータセットの持つ情報をほぼすべて網羅してしまい、トレーニングデータにしかないようなノイズなどを過度に学習してしまうことで、結果未知のデータに対して汎化性能を持っていない「過学習」状態にあるとも考えられる。結果は、モデルが正確な推定を行うための情報をデータセットから得られていないことを示している。

### 4.4.2 モデルの予測結果

各 Fold において、トレーニング前にあらかじめ除いておいた 1 割の未知のテストデータセットと、もう片方の相互作用モデルのデータセット全てに対して予測を行うことでモデルの最終的な性能を評価する。結果は図 4.4 および図 4.5 に示す。赤線は陽子イベント、青色は鉄イベントに対する予測で、緑はアマテラス粒子に対する予測である。実線は 10 Folds の平均値、バンドは標準偏差を表す。

モデルのテストデータセットでの正解率は表 4.3 にまとめている。

アマテラス粒子への適用結果は EPOS-LHC のデータセットを学習したモデルでは Ironness =

$0.8330 \pm 0.0862$ 、QGSJETII-04 のデータセットを学習したモデルでは Ironness =  $0.8889 \pm 0.0688$  となった (図 4.4、図 4.5 参照)。



図 4.3 CNN による学習の学習曲線 (上が EPOS-LHC、下が QGSJETII-04 のデータセットでの学習)。それぞれ 10 Fold 分の結果を重ね描きしている。

表 4.3 学習モデルの正解率

	EPOS-LHC テストデータセット	QGSJETII-04 テストデータセット
EPOS-LHC 学習モデル	$0.6515 \pm 0.0257$	$0.6720 \pm 0.0057$
QGSJETII-04 学習モデル	$0.6598 \pm 0.0069$	$0.6978 \pm 0.0422$

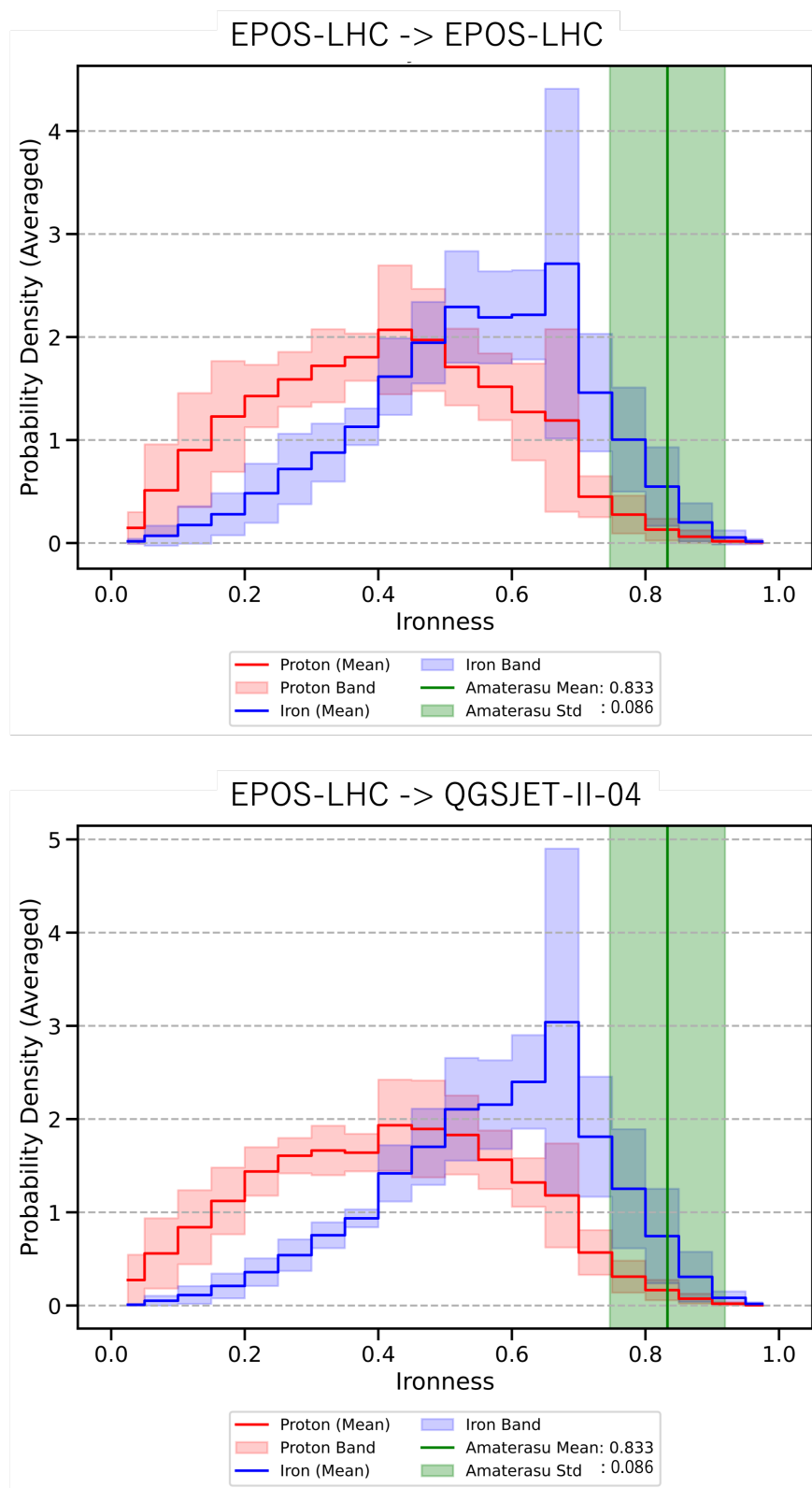


図4.4 EPOS-LHC のデータセットを使ったモデルの予測結果(上が EPOS-LHC、下が QGSJETII-04 のデータセットへの適用結果)。ビンごとに、10 Fold 分の結果の平均値および標準偏差を表している。赤が陽子のイベント、青が鉄のイベントに対する適用結果である。緑の線とバンドはアマテラス粒子への適用結果の 10 Fold 分の平均値と標準偏差である。

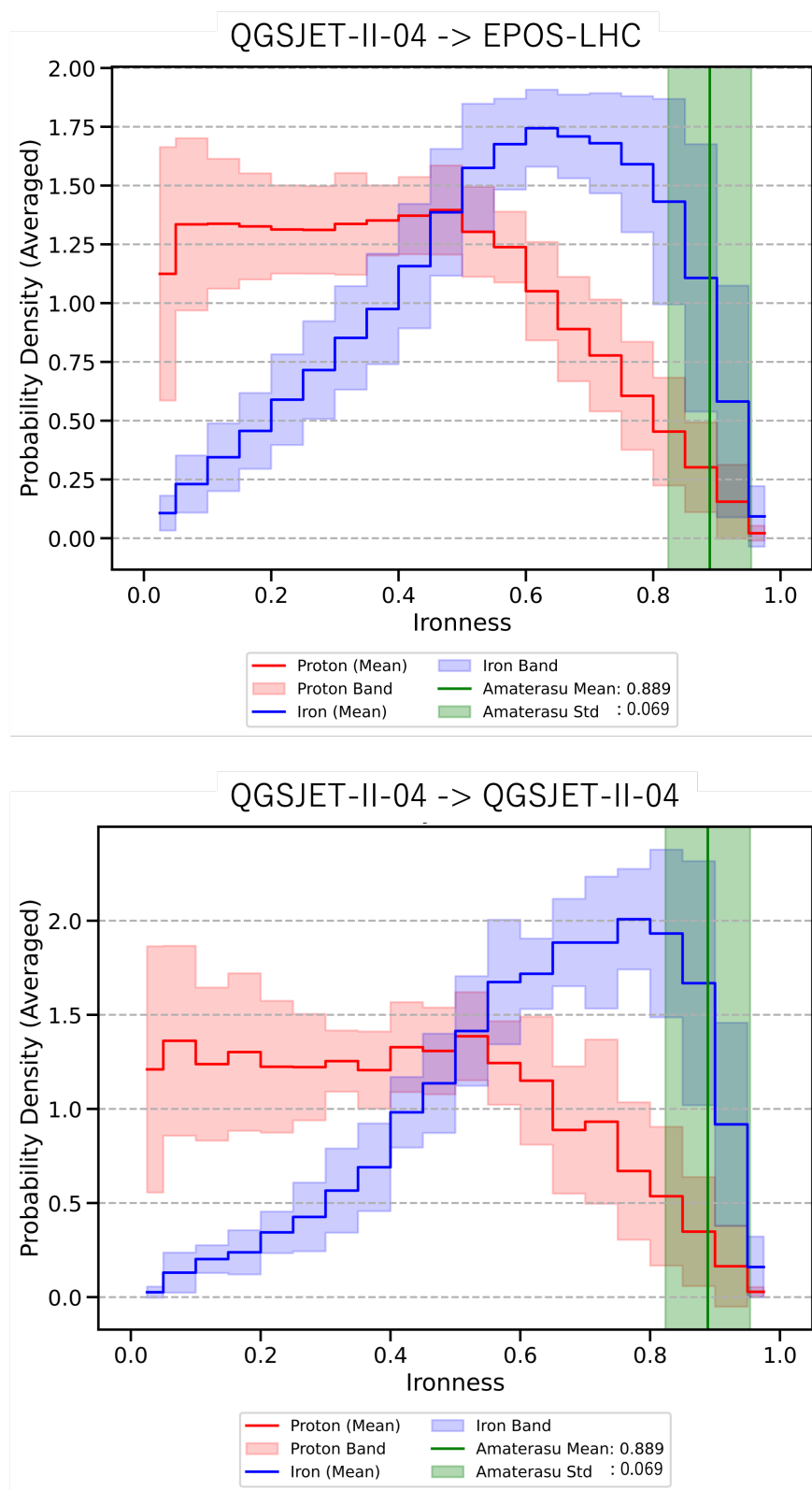


図 4.5 QGSJETII-04 のデータセットを使ったモデルの予測結果（上が EPOS-LHC、下が QGSJETII-04 のデータセットへの適用結果）。

## 第 5 章

# ランダムフォレストを用いた原子核種推定

4 章で用いた学習モデルの問題点を解決するために、より多くの特徴量を用いた新たなモデルを構築する。学習手法について、膨大な特徴量を従来の CNN のような全結合型の手法で扱う場合、計算時間が非常に長くなってしまったため、実現可能な計算時間で特徴量を適切に扱えるアンサンブル型の手法であるランダムフォレストを採用する [35]。

### 5.1 ランダムフォレスト

ランダムフォレストとは決定木に基づくアンサンブル型の学習アルゴリズムである。それぞれの決定木は独立に陽子か鉄かの判定を行い、複数の決定木の結果による推定を行う。決定木ごとに独立したイベント選択、特徴量選択を用いた推定を行うことで、膨大な特徴量のデータセットに対しても過学習を抑えつつ汎化性能を高めることができる。

#### 5.1.1 特徴量重要度

ランダムフォレストの強みとして、特徴量重要度を可視化できるという点がある。モデルが推定を行う際に不要だった特徴量を除くことで、学習効率を高めることができる。これにより性能を保ったまま計算時間を減らすことができれば、データセットの数を増やしたり、NN のような複雑な手法を用いたより高度な学習を現実的な計算時間で行える。

### 5.2 データセットの作成

CNN による学習 (4 章) では、波形データから得る特徴量は到来時刻と積分値のみであった。本章の学習では、各イベントの情報をより細かく抜き出すために、TSFEL (Time Series Feature Extraction Library) [36] を用いた波形データからの特徴量抽出を行う。

### 5.2.1 波形データ

機械学習用にまとめた地表検出器が持つ波形データは 1 列目に時刻のカウント (1 カウント 20 ns)、2 列目に上層の信号、3 列目に下層の信号をもつ csv ファイルとなっている。波形データはトリガーがかかるたびに 128 ビン (2.56  $\mu$ s) のデータを保存するため、検出器ごとに波形の長さが異なる。128 ビンの一番短い波形でも TSFEL での特徴量抽出は可能なため、本研究では制限をかけずに全ての波形を使用する。この波形データに対して TSFEL を用いた特徴量抽出を行い、機械学習用のデータセットを作成する。

### 5.2.2 TSFEL

TSFEL は時系列データから特徴量を網羅的に抽出するために開発された Python ライブラリであり、主な特徴として統計的、時間的、スペクトル的な 100 種類以上の特徴量を自動的に抽出することができる。本研究では、特徴量不足によると考えられる機械学習モデルの性能の頭打ちを解消するべく、TSFEL を用いて検出器の波形データからの機械学習に使用する特徴量を抽出する。

### 5.2.3 TSFEL を用いた特徴抽出

抽出する特徴量は統計的、時間的、スペクトル的な特徴量として TSFEL で抽出可能な 156 のすべてのものである。これに到来時刻を加えた 157 の特徴量を各検出器の特徴量として使用する。

#### 使用する地表検出器

本研究では、イベントごとに同じ範囲の検出器群を用いるため本質的に検出器の座標といった空間的な特徴を必要としないため、CNN において必要だった正方形の検出器配置は本来不要である。不要な特徴量はモデルの学習の際ノイズとなるため、CNN での学習の際に導入した仮想の検出器は除き、49 台の实在の検出器のみを用いる (図 5.1)。

それぞれの検出器について、上層と下層について独立した特徴量抽出が行われ、最終的に機械学習プログラムに与えられる総特徴量は 1 イベントにつき  $49 \times 2 \times 157$  の 15386 となる。4 章での 128 特徴量と比較した大幅な特徴量の増加によりモデルが空気シャワーの情報を細かく学習できるため、推定精度の向上が期待できる。

### 5.2.4 データセットまとめ

モンテカルロシミュレーションによる空気シャワーイベントは 4 章で使用したものと同じものを用いる。TSFEL を用いて抽出された特徴量は、上層と下層別に地表検出器が 49 行、特徴量が 157 列の csv ファイルにまとめられる。それぞれの特徴量は scaler を用いて正規化され、特徴量の数値の大きさでのバイアスが発生しないようにする。

ファイルは相互作用モデル、核種ごとに 10000 イベントずつの計 40000 イベントあり、CNN の学

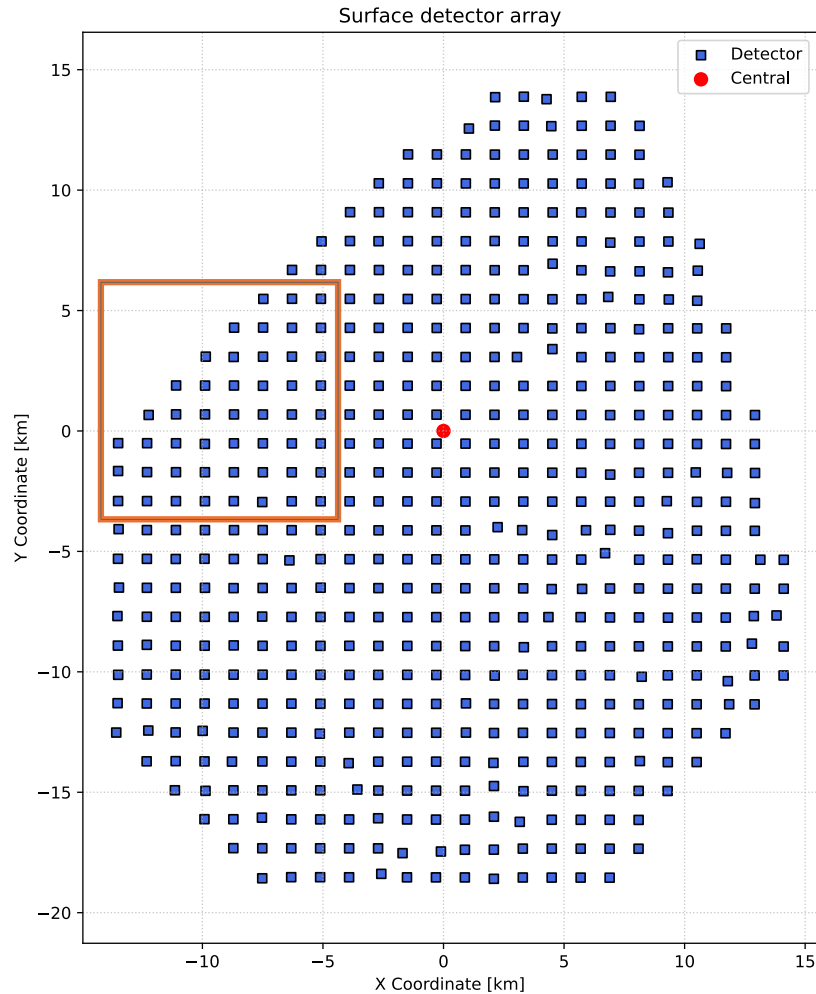


図 5.1 TA 地表検出器の配置と学習に用いる範囲。実在の検出器のみを用いる。

習と同様に相互作用モデルごとに独立な学習モデルを構築するためデータセットも 20000 ファイルずつ独立に使用する。

## 5.3 学習モデル

モデルは CNN での学習と同様にハドロン相互作用モデルごとに独立したものを作成し、データセットも独立したものを使用する。モデルの出力について、本章でも「Ironness」を採用する。

### 5.3.1 パラメータ探索

Optuna によるハイパーパラメータ探索はランダムフォレストでの学習においても効果的であるため採用する。設定するパラメータは以下の 4 つ。決定木の深さは決定木ごとに枝分かれの条件で決まるため、余裕を持たせ 40 を上限とする。Optuna は CNN の学習と同様に正解率を最大にするように探索を行う。探索を行うパラメータおよび探索範囲は表 5.1 に示す。

表 5.1 パラメータ探索範囲

パラメータ名	探索範囲
決定木の本数	350 - 450
決定木の深さの上限	40
枝分かれ時に必要なイベント数	36 - 44
枝分かれ後に必要なイベント数	18 - 22

表 5.2 機械学習モデル設定

フレームワーク	scikit-learn	
	EPOS-LHC	QGSJET-II-04
決定木の本数	435	385
決定木の深さの上限		40
枝分かれ時に必要なイベント数	40	38
枝分かれ後に必要なイベント数	22	19

決定木の本数はモデルの汎化性能を高める。決定木の深さおよび枝分かれ前後で要求されるイベント数はモデルがイベントをどれだけ細かく分別するかを表している。

Optuna が探索するパラメータの範囲を表 5.1 に示す。計算時間を抑えるため、またモデルがトレーニングデータのみに含まれる微小なノイズなどを学習することによって引き起こされる過学習を防ぐために、本研究では枝分かれ前後の要求イベント数を粗めに設定している。同時にそれぞれの決定木が間違える可能性も上昇するため決定木の本数を多めに設定してモデルの汎化性能を高めるよう設定する。

### 5.3.2 特徴量選択

Optuna によるパラメータ探索の際に特徴量の評価も同時に行い、後の本学習で使用する特徴量を選別する。特徴量は重要度の高い順に選ばれ、合計 95 % を超えるまでに選ばれたものを記録する。

### 5.3.3 学習モデルまとめ

ランダムフォレストの学習ではデータセットの前処理や性能評価の可能なライブラリである scikit-learn を採用する [37]。Optuna の探索によって得られたパラメータを含めた学習モデルの設定は表 5.2 にまとめる。

パラメータ探索、特徴量重要度評価を行った後に、本学習を行う。ランダムフォレストの学習でも、CNN の学習と同様に全データを 10 分割し、それぞれのブロックがテストに使われるように 10 回の学習を行う 10-fold cross-validation を採用する。

表 5.3 学習モデルの正解率 (全イベント)

	EPOS-LHC テストデータセット	QGSJET-II-04 テストデータセット
EPOS-LHC 学習モデル	0.8483 ± 0.0077	0.8745 ± 0.0011
QGSJET-II-04 学習モデル	0.8217 ± 0.0009	0.9072 ± 0.0074

表 5.4 学習モデルの正解率 (鉄イベントのみ)

	EPOS-LHC テストデータセット	QGSJET-II-04 テストデータセット
EPOS-LHC 学習モデル	0.9240 ± 0.0101	0.9406 ± 0.0015
QGSJET-II-04 学習モデル	0.8937 ± 0.0016	0.9631 ± 0.0069

## 5.4 結果

以下にランダムフォレストによる学習の結果を述べる。

### 5.4.1 モデルの適用結果

CNN での学習と同様に、トレーニング前にあらかじめ除いておいた 1 割の未知のテストデータセットと、もう片方の相互作用モデルのデータセット全てに対して予測を行うことでモデルの最終的な性能を評価する。結果は図 5.2、図 5.3 に示す。赤線は陽子イベント、青色は鉄イベントに対する予測で、緑はアマテラス粒子に対する予測である。実線が 10Fold の平均値、バンドは標準偏差を表す。

モデルのテストデータセットにおける正解率は表 5.3 にまとめている。正解は陽子イベントで Ironness が 0.5 未満、鉄イベントで Ironness が 0.5 以上の出力値となった場合を意味する。

アマテラス粒子への適用結果は EPOS-LHC のデータセットを学習したモデルでは Ironness =  $0.7078 \pm 0.0122$ 、QGSJET-II-04 のデータセットを学習したモデルでは Ironness =  $0.6404 \pm 0.0155$  となった (図 5.2、図 5.3 を参照)。

アマテラス粒子への適用結果が 0.6 付近であったことを踏まえて、条件付きの正解率を計算した。結果を図 5.4 に示す。ここでは扱うイベントを Ironness が 0.6 以上の出力値になったもののみとし、鉄イベントのみでの正解率を計算した。

### 5.4.2 特徴量重要度の評価

10Fold での特徴量重要度の評価を図 5.4 に示す。重要度は全て足し上げた際に 1 になる。突出して重要な特徴量はなく、最も高いもので 1 % 程度の寄与である。頻出する特徴量についての重要度を図 5.5、図 5.6、図 5.7 に示す。

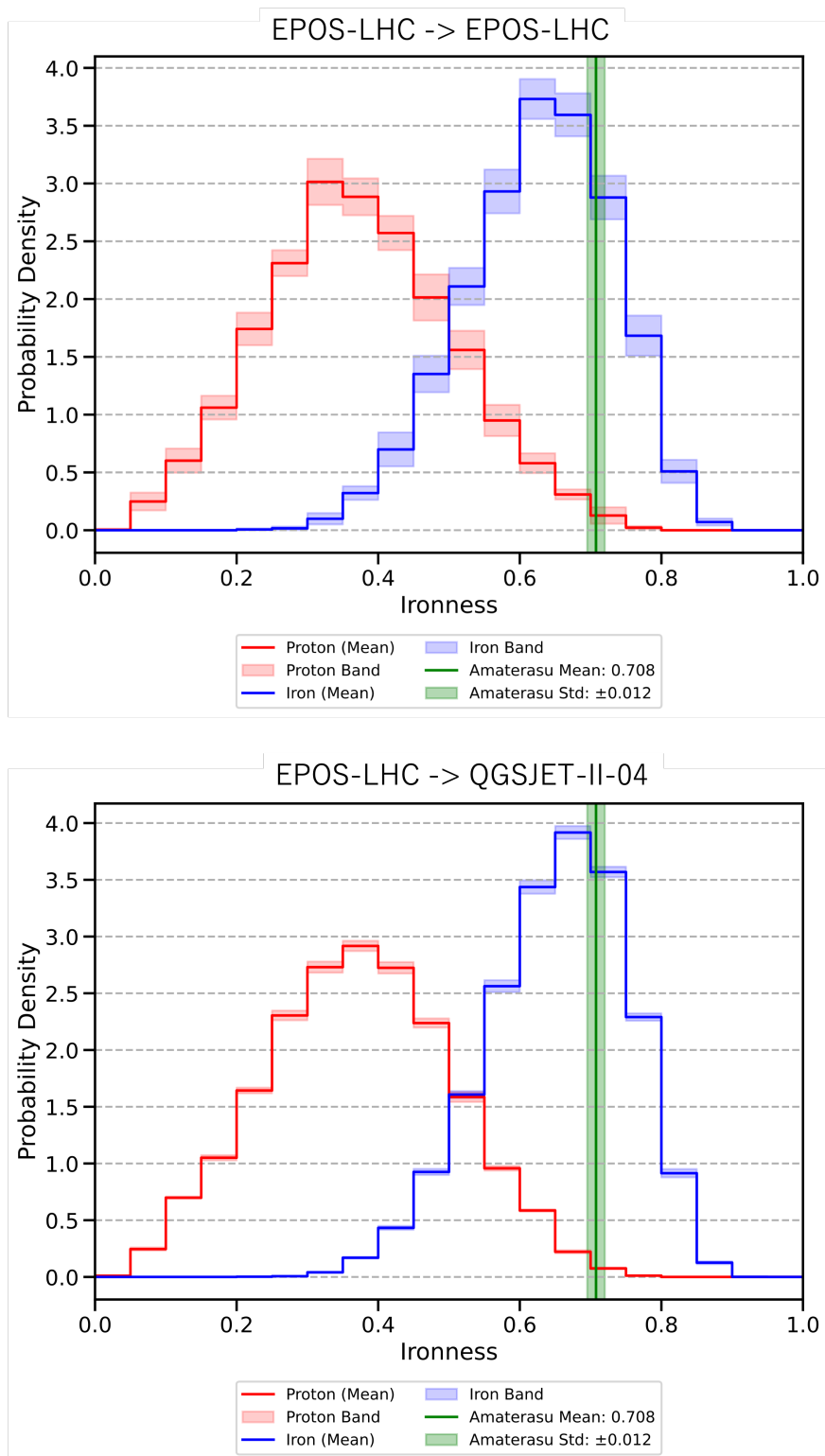


図 5.2 EPOS-LHC のデータセットを使ったモデルの予測結果（上が EPOS-LHC、下が QGSJET-II-04 のデータセットへの適用結果）。ピンごとに、10 Fold 分の結果の平均値および標準偏差を表している。赤が陽子のイベント、青が鉄のイベントに対する適用結果である。緑の線とバンドはアマテラス粒子への適用結果の 10 Fold 分の平均値と標準偏差である。

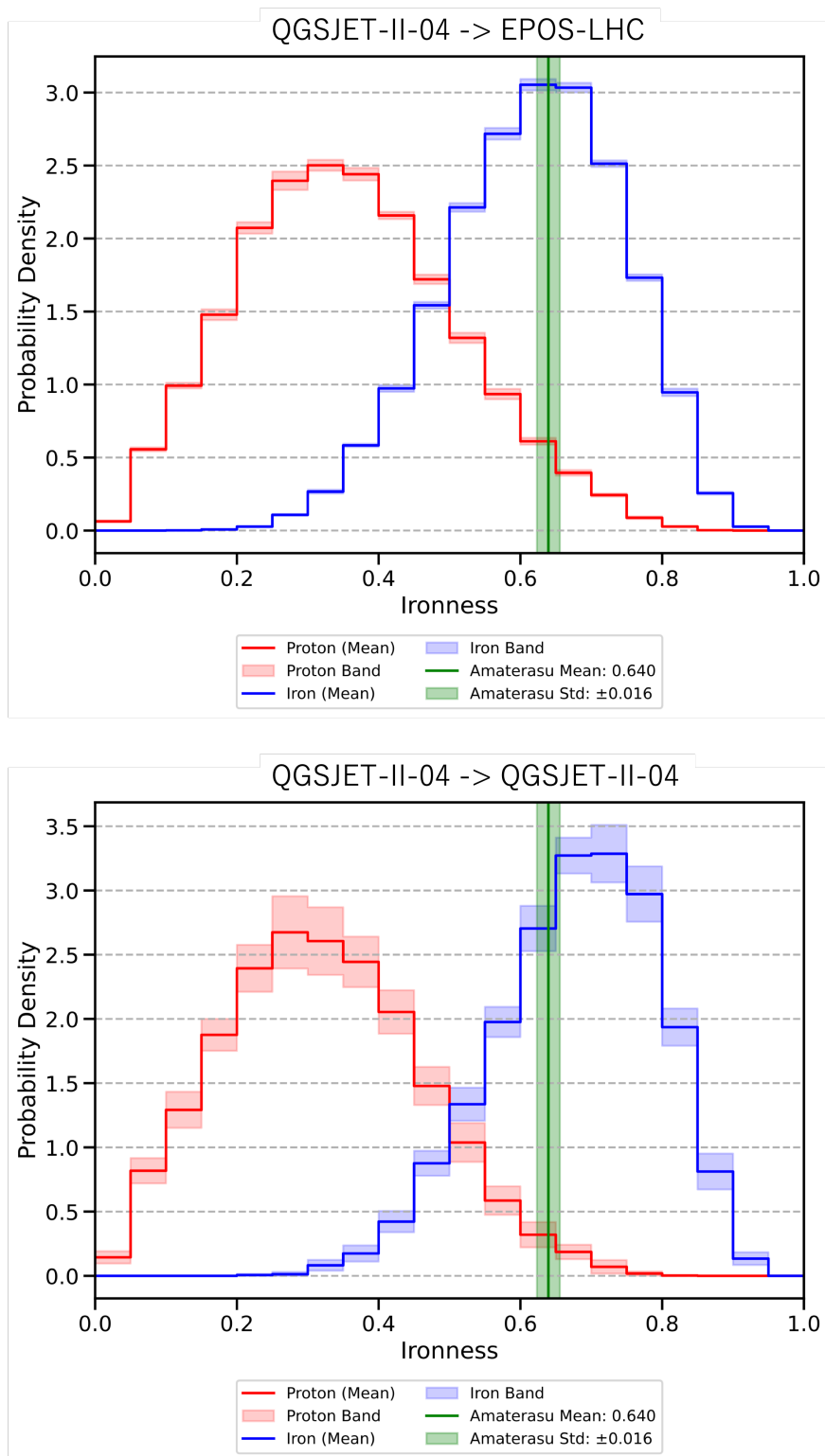


図 5.3 QGSJET-II-04 のデータセットを使ったモデルの予測結果（上が EPOS-LHC、下が QGSJET-II-04 のデータセットへの適用結果）。

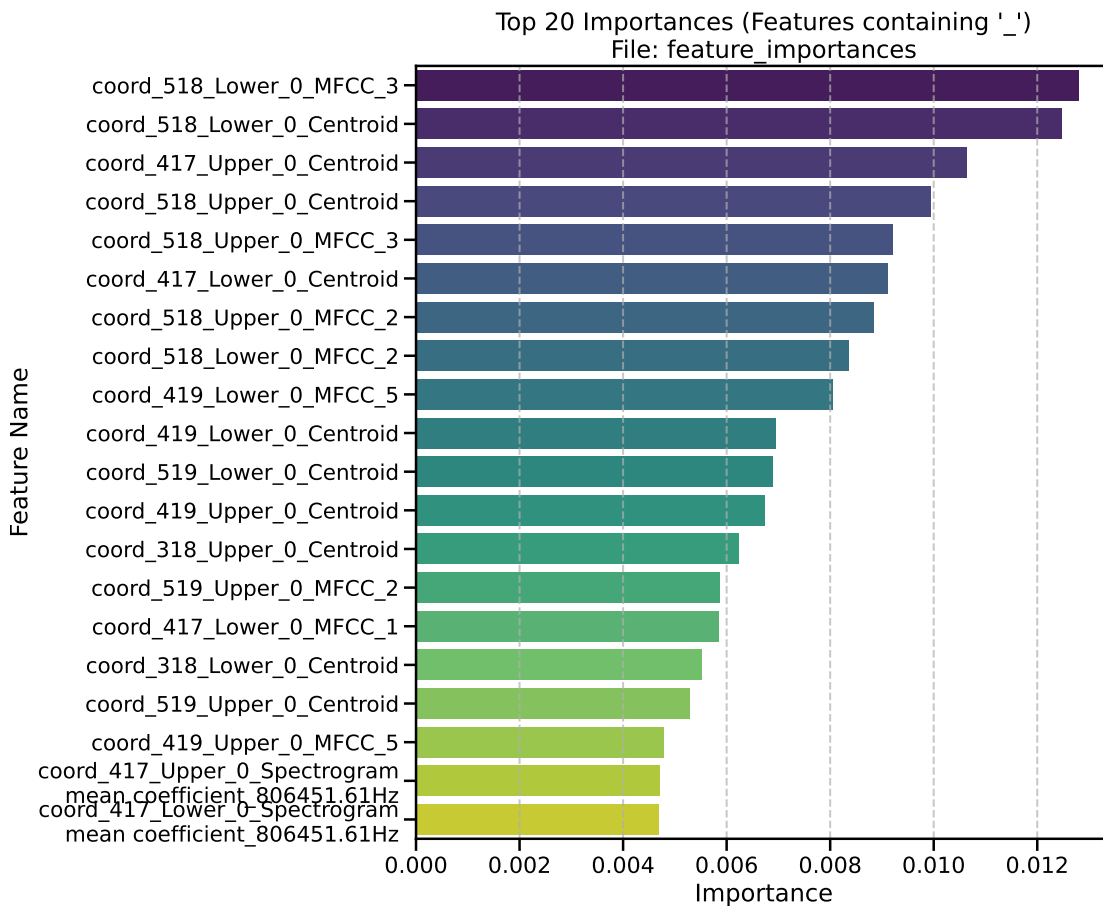


図 5.4 重要度上位 20 の特徴量。EPOS-LHC のデータセットを使った学習モデルによるもの。

### Centroid

信号の重心を表す特徴量。空気シャワー粒子の大部分によって作られる波形に対して、ミューオンのようなピーク波形の存在によってこの値が変化すると考えられる。

### MFCC

メル周波数ケプストラム係数。本来は人の耳が聞き取りやすい周波数帯の音に焦点を当てて解析を行う際に用いられる係数であり、高周波数 (個々の粒子が作る幅の狭い信号) よりも、低周波数 (粒子群による幅の広い信号) に焦点を当てて計算する特徴量を表す。

### Spectrogram mean coefficient

波形の中に特定の周波数の信号がどれだけ含まれているかを表す。50MHz のサンプリングレート (20 ns) に対して、重要とされる 800kHz 付近の周波数帯 (1  $\mu$ s) は、MFCC と同様に粒子群による幅の広い信号に由来する特徴量を表す。

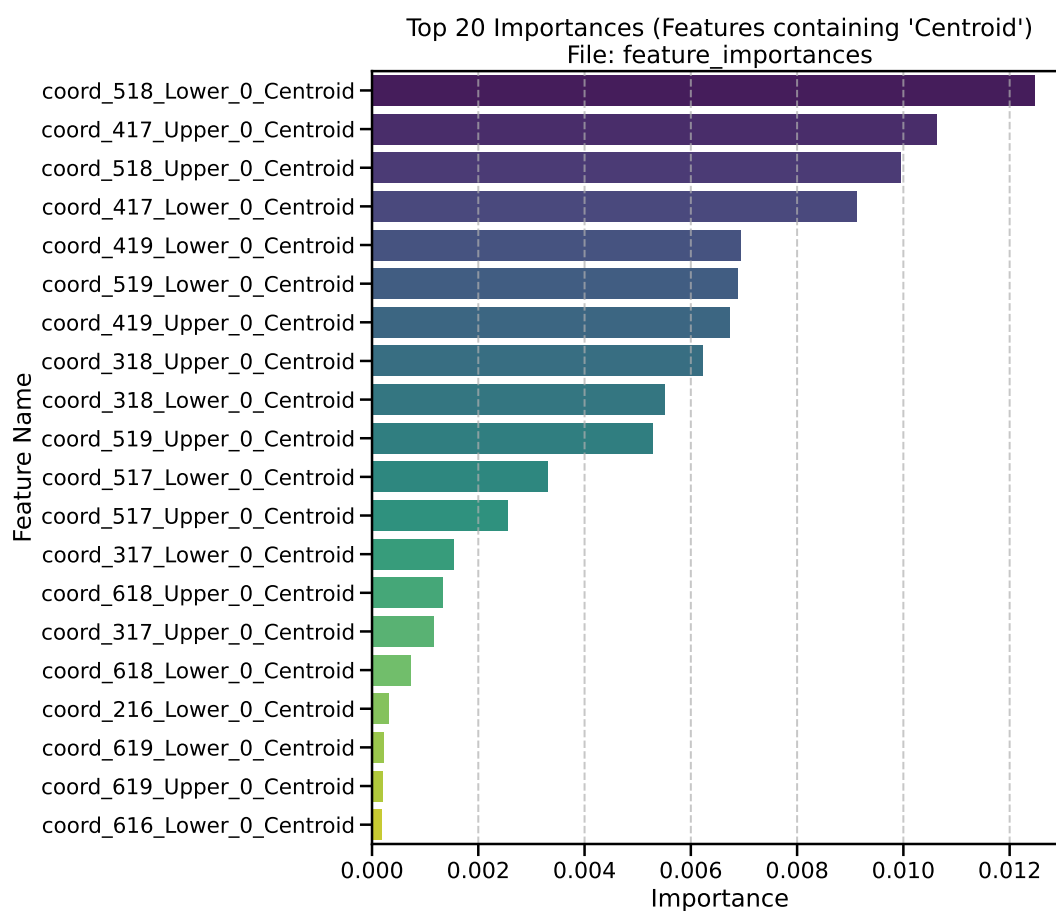


図 5.5 重要度上位 20 の特徴量 (Centroid)。EPOS-LHC のデータセットを使った学習モデルによるもの。

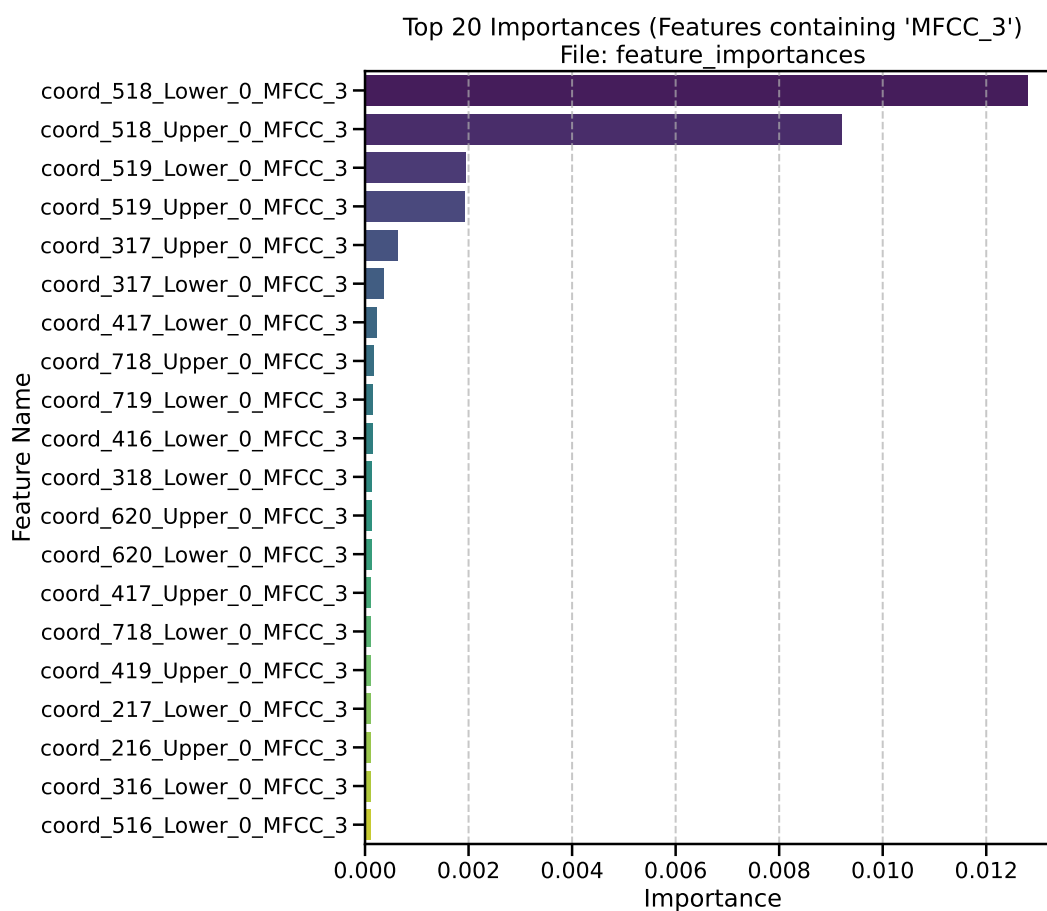


図 5.6 重要度上位 20 の特徴量 (MFCC-3)。EPOS-LHC のデータセットを使った学習モデルによるもの。

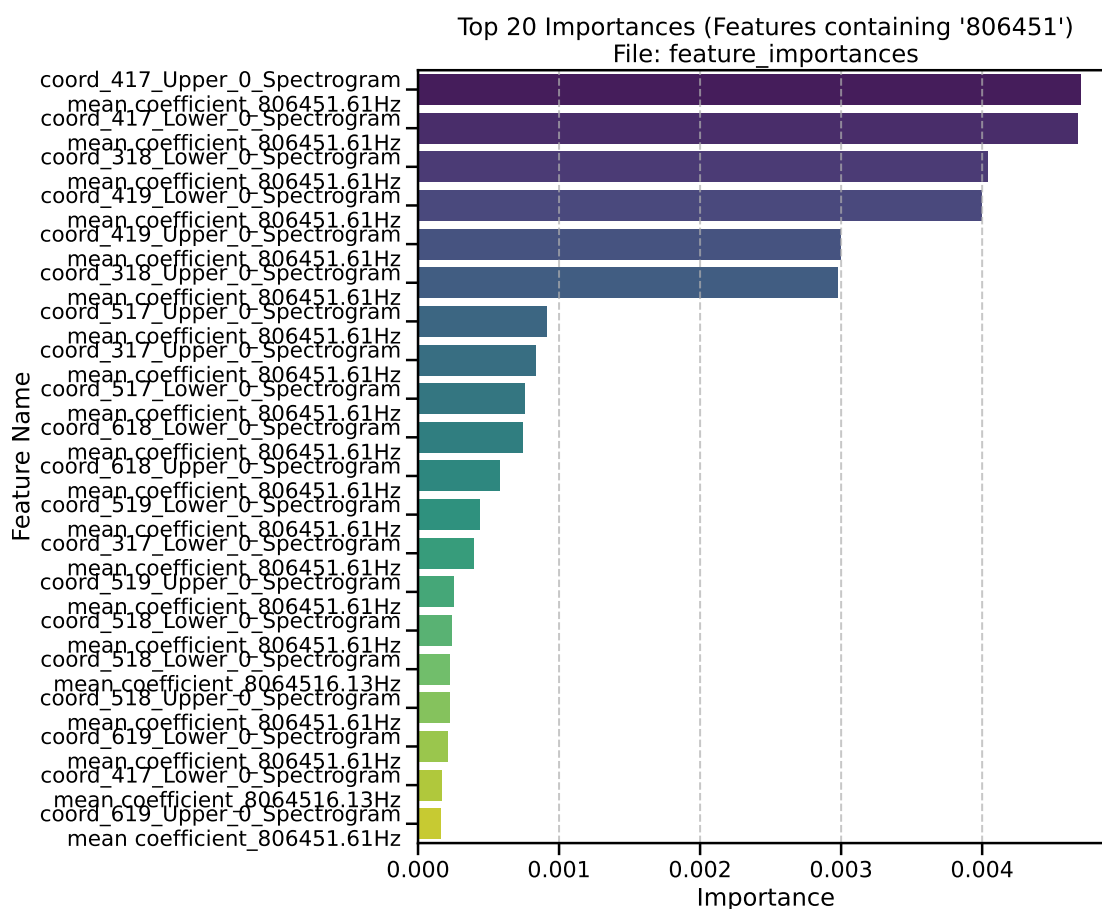


図 5.7 重要度上位 20 の特徴量 (Spectrogram mean coefficient 800kHz)。EPOS-LHC のデータセットを使った学習モデルによるもの。

## 第 6 章

# 議論

### 6.1 モデルの推定結果

CNN を用いた原子核種推定では学習モデルの過学習の傾向と、推定性能の限界が見られた。これらの問題が起きる原因として、データセットの持つ特徴量の少なさが考えられる。ここで使われている特徴量は各 SD への空気シャワー粒子の到来時刻と波形データの積分値のみであり、モデルは特徴量から判別に必要な情報を得られておらず、結果 CNN のような高度な手法を用いるにはアレイサイズも特徴量も小さすぎたことで、過学習や性能の限界が見られたと考えられる。

また、CNN の結果ではテストデータの推定値、アマテラス粒子の推定値ともに幅がある。これについて、特徴量の少なさが原因と考える。CNN でのアマテラス粒子の推定値は鉄側に見えるが、鉄イベントの大部分が位置する場所よりもさらに右側に位置しており、鉄らしいと言い切ることはできない。

この結果を踏まえて、使用する特徴量を増やすことで問題の解決を狙ったランダムフォレストでの推定結果では、学習モデルの正解率が向上し、推定値の幅は小さくなった。アマテラス粒子の推定値は鉄イベントの大部分と近い出力となった。

### 6.2 特徴量重要度

ランダムフォレストによる特徴量重要度の評価では、Centroid や MFCC といった特徴量が重要であると分かった。これらの性質を見ると、学習モデルは電磁成分とミュオンなどの違いが現れる個々の粒子の作る細かな波形と、粒子群の重ね合わせによる大きな波形の形状をどちらも重要視していることがわかる。波形全体の形状は空気シャワープレーンの厚みを表しているため、本研究によって、SD の波形データにも原子核種の推定に効果的な情報が含まれていることを示した。

また、特徴量重要度のヒストグラムからは突出して重要な特徴量はないように見えるが、実際は重要度の大部分 (70% 以上) はシャワー中心付近の検出器の特徴量が占めている。シャワー中心付近は粒子数が多いため波形からの特徴量抽出が効率よく行われ、結果として分類に強く寄与していると考えられる。今後は範囲を絞って空気シャワーコア付近の SD に範囲を絞ってより細かく解析を行うことで、波形の特徴を効果的に活用できると考えられる。

### 6.3 課題と展望

本研究では全体を通して、シミュレーション結果を信頼して使用していることに注意する必要がある。一般に、現在広く使われているハドロン相互作用モデルには、シミュレーション時に実際よりミューオン数を少なく見積もってしまうという問題がある。超高エネルギー宇宙線でもこの特徴が現れると考えると、鉄を想定しているシミュレーションはより軽い原子核の振る舞いをしている可能性がある。本研究の推定結果を踏まえて、アマテラス粒子が鉄よりも軽い原子核（中間質量である炭素、窒素、酸素など）である可能性を考えなければならない。

また、近年では鉄よりも重い原子核の可能性があるという主張も存在する [38, 39]。本研究では陽子と鉄の 2 つのみを考えたが、推定精度を向上させることができれば、複数の原子核種を想定した学習が可能になり、より細かく核種推定を行うことができると考えられる。

機械学習モデルについて、本研究はアマテラス粒子という 1 つのイベントに特化した学習であるため、学習モデルの信頼度を評価する必要がある。TA 実験では FD と SD を組み合わせたハイブリッド観測も行われており、FD 観測によって核種推定が行われている宇宙線イベントの SD 情報を用いて同様の解析を行うことでの学習モデルの評価を検討している。

また、本研究で用いたイベント特化型の機械学習手法を 2、3 番目... と順に他の最高エネルギー宇宙線に対して適用し、推定値を得ることで、最高エネルギー宇宙線の核種について統計的な解析を行うことができる。このデータは TA 実験における異方性解析にも利用できる可能性がある。

## 第 7 章

# 結論

本研究では、テレスコープアレイ (Telescope Array : TA) 実験において観測された実験史上最高エネルギー (244 EeV) を持つ宇宙線「アマテラス粒子」の起源特定のため、機械学習を用いた原子核種推定を行った。

まず、ハドロン相互作用モデル (EPOS-LHC、QGSJET-II-04) を用いた空気シャワーシミュレーションにより、アマテラス粒子の観測パラメータを再現する学習データセットを生成した。推定手法として CNN およびランダムフォレストを採用した結果、波形データの積分値と到来時刻のみを用いた CNN では正解率が約 65% に留まり、学習不足および過学習の傾向が見られた。一方、TSFEL を用いて波形データから多次元的な特徴量を抽出したランダムフォレストでは、テストデータに対して約 85% の高い正解率を達成した。

シミュレーションデータをランダムフォレストを用いて学習させ、構築した 2 つのモデルをアマテラス粒子の観測データに適用した結果、推定値 (Ironness) はそれぞれ  $\text{Ironness} = 0.7078 \pm 0.0122$ 、 $\text{Ironness} = 0.6404 \pm 0.0155$  となり、陽子に比べて鉄原子核に近いという結果となった。特徴量重要度の解析からは、波形のピークや、全体の形状を反映する波形の「Centroid (重心)」や「MFCC (低周波数成分)」が判別に大きく寄与していることが判明し、SD データのみでも核種推定が可能であることを示した。

以上の結果から、本研究はアマテラス粒子は重い原子核である可能性が高く、銀河系内外の磁場によって大きく偏向を受けた結果、見かけ上の到来方向であるボイド領域とは異なる近傍天体 (NGC 6946 等) を起源としているシナリオを支持する結果となった。今後は、イベント数を増やした学習モデルの性能向上、中間質量の原子核を含めた分類モデルの構築や、TA ハイブリッド観測データを用いたイベント特化型学習の評価などを行い、最高エネルギー宇宙線の核種推定精度の向上を目指す。

# 付録

## A QGSJET-II-04 によるシミュレーションデータ

3 章でのシミュレーションデータの再構成結果について、QGSJET-II-04 によるものを以下に示す。

## B QGSJET-II-04 による学習モデルの特徴量重要度

5 章での特徴量重要度の評価について、QGSJET-II-04 を用いた学習モデルのものを以下に示す。

QGSJET-II-04\_P

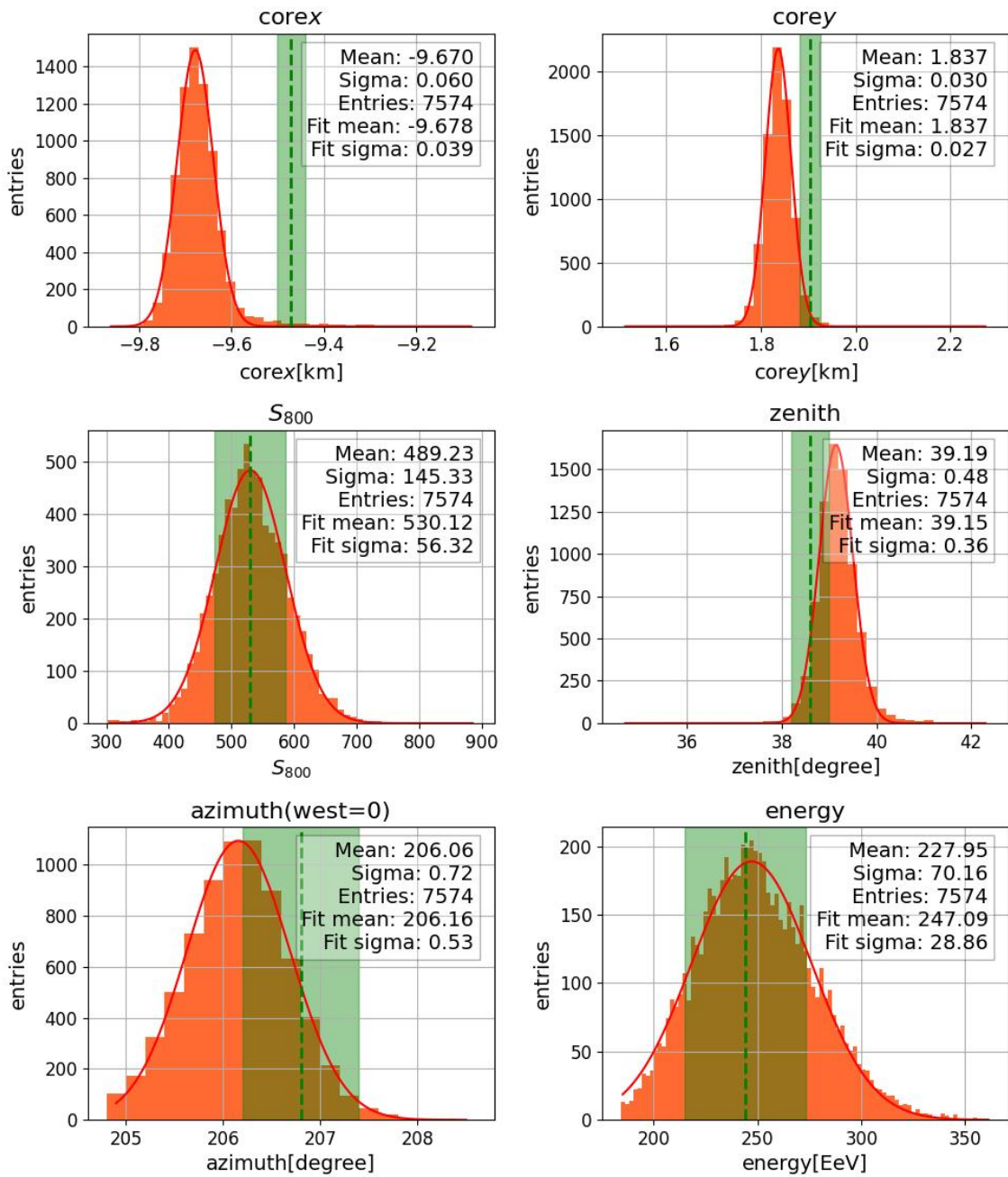


図 A.1 QGSJET-II-04 を使った陽子のシミュレーションイベントの再構成結果（調整前）。

QGSJET-II-04\_Fe

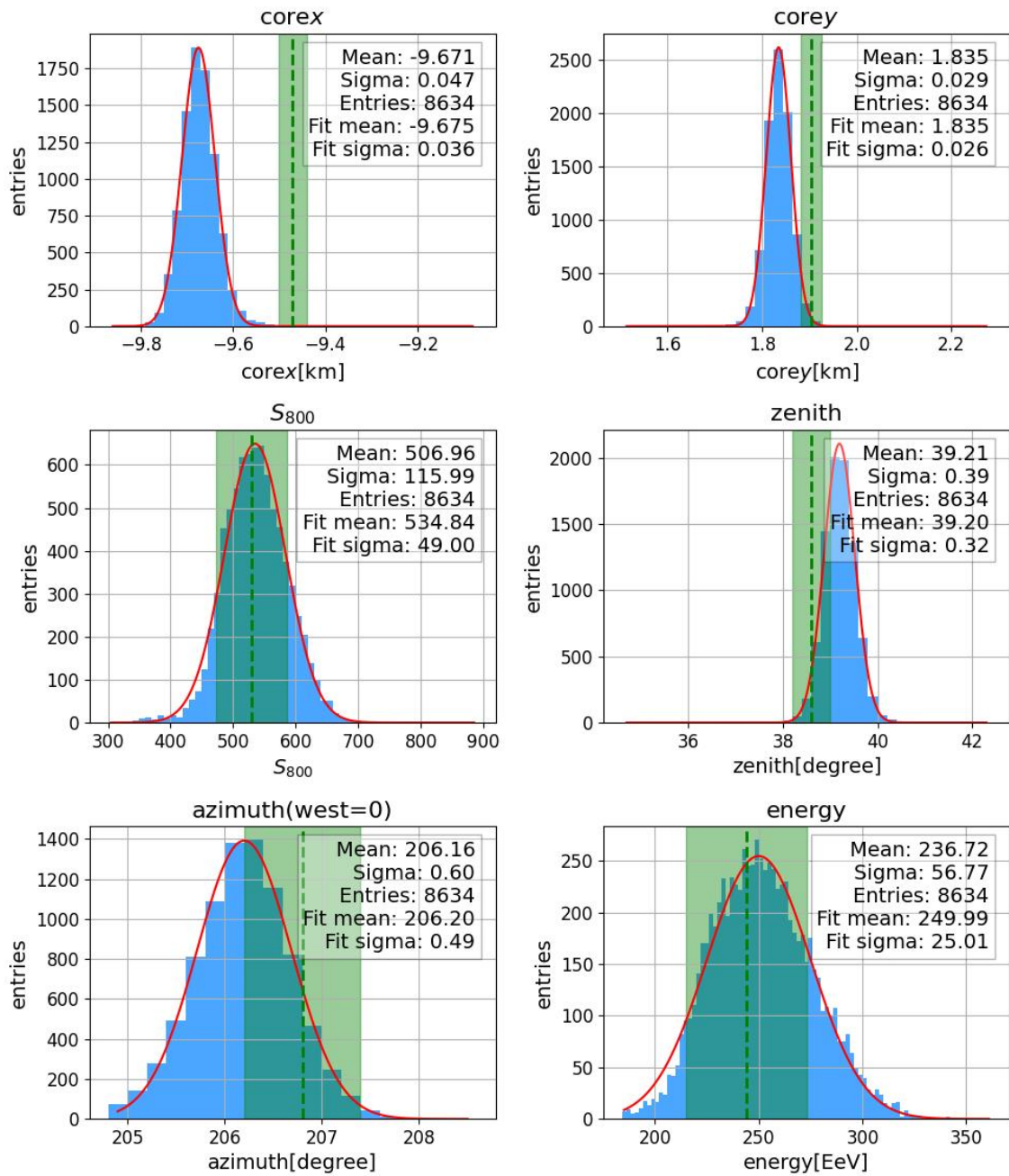


図 A.2 QGSJET-II-04 を使った鉄のシミュレーションイベントの再構成結果 (調整前)。

QGSJET-II-04\_P

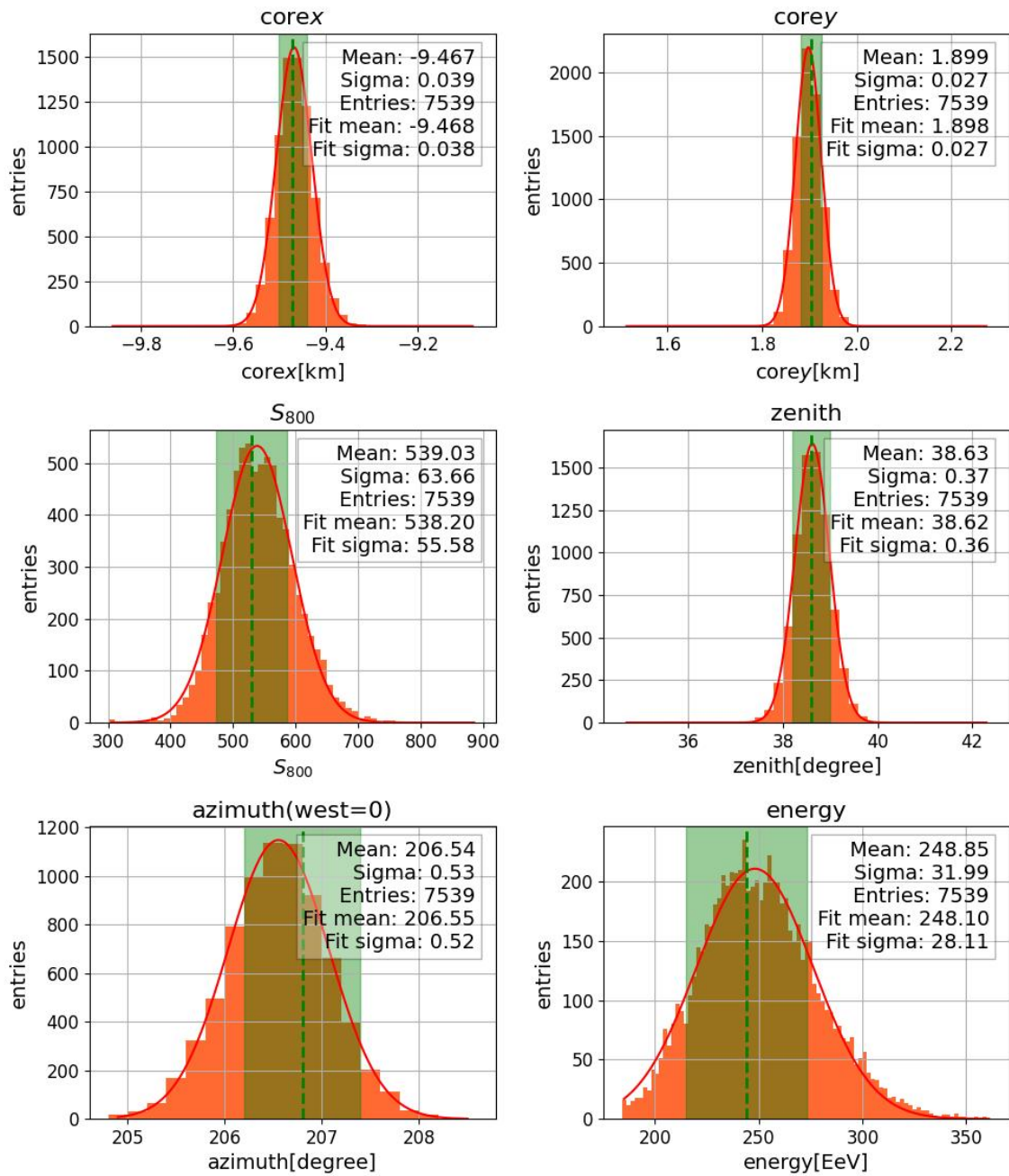


図 A.3 QGSJET-II-04 を使った陽子のシミュレーションイベントの再構成結果（調整後）。

QGSJET-II-04\_Fe

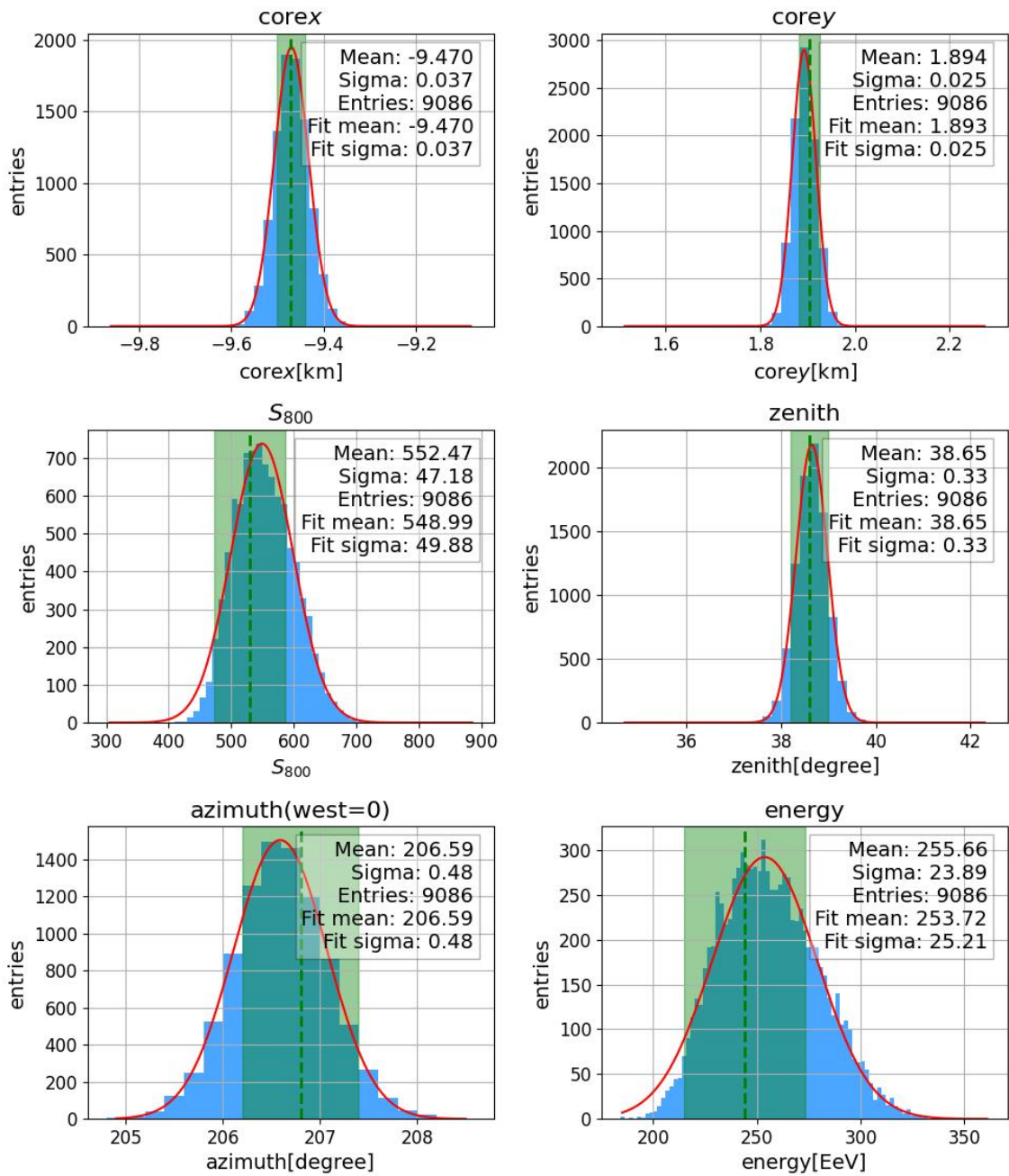


図 A.4 QGSJET-II-04 を使った鉄のシミュレーションイベントの再構成結果 (調整後)。

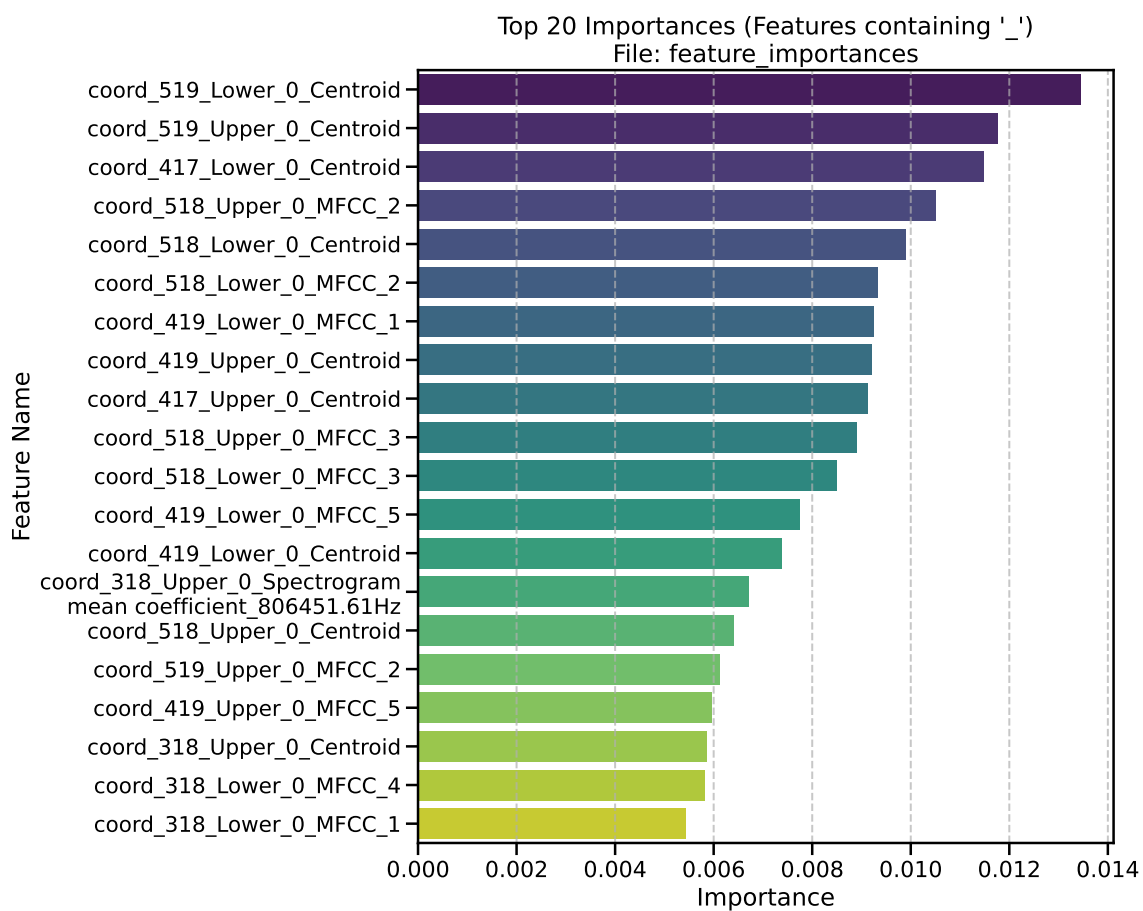


図 B.5 重要度上位 20 の特徴量。QGSJET-II-04 のデータセットを使った学習モデルによるもの。

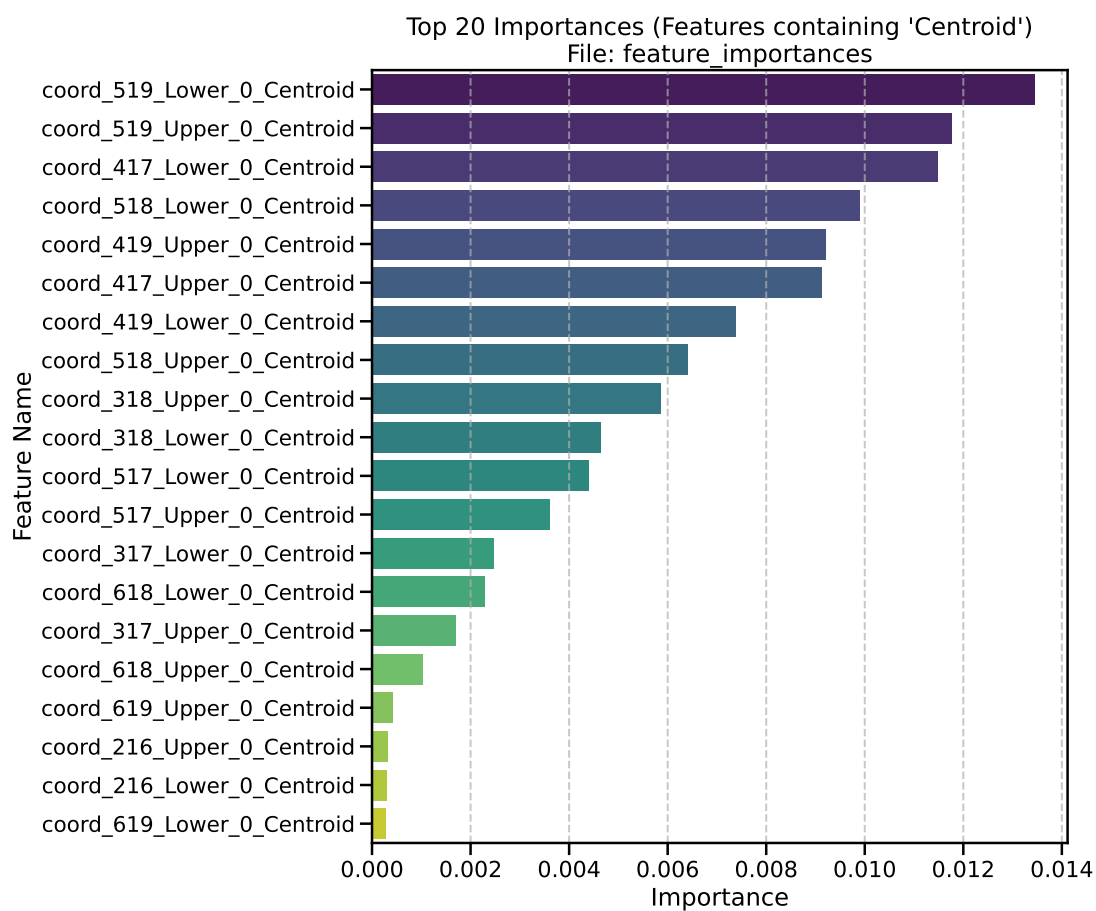


図 B.6 重要度上位 20 の特徴量 (Centroid)。QGSJET-II-04 のデータセットを使った学習モデルによるもの。

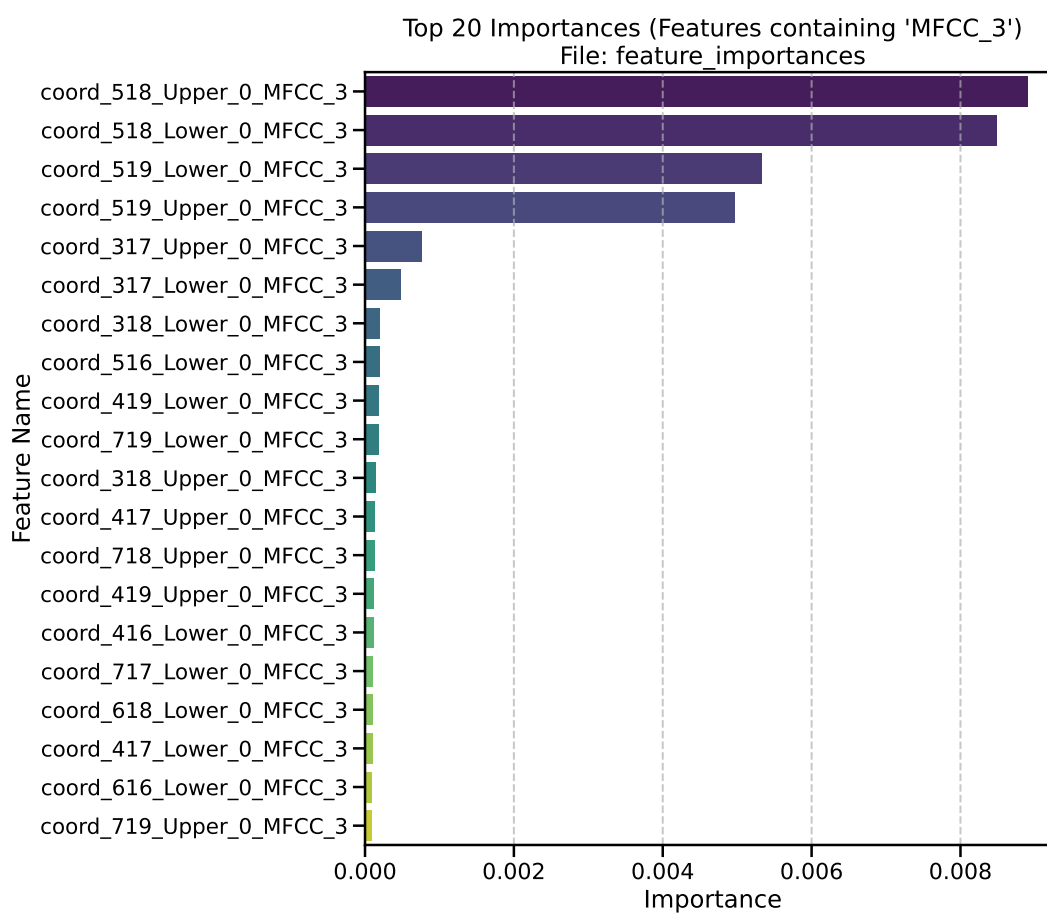


図 B.7 重要度上位 20 の特徴量 (MFCC-3)。QGSJET-II-04 のデータセットを使った学習モデルによるもの。

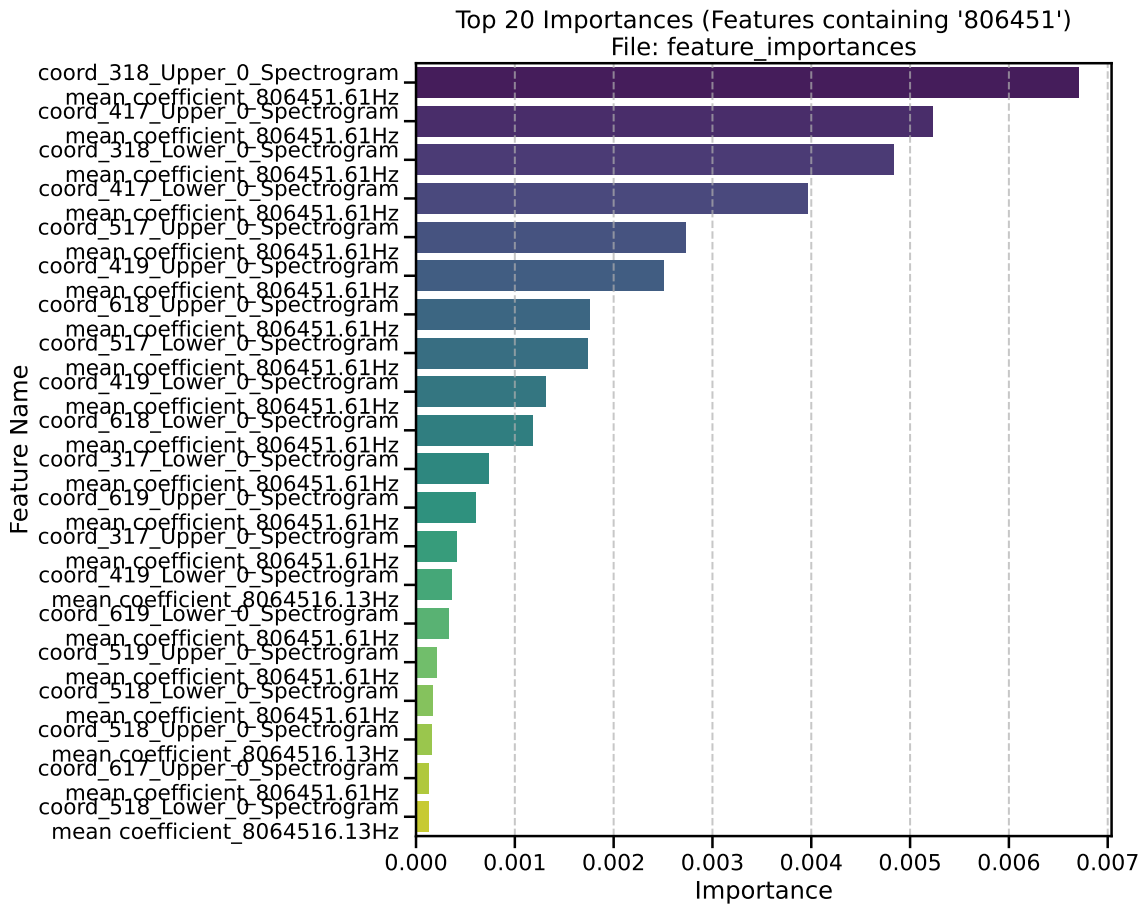


図 B.8 重要度上位 20 の特徴量 (Spectrogram mean coefficient 800kHz)。QGSJET-II-04 のデータセットを使った学習モデルによるもの。

## 謝辞

本研究を進めるにあたり、多くの方にご助言、ご支援をいただきました。皆様に心より感謝申し上げます。

指導教員でもある藤井俊博准教授には本研究のテーマであるアマテラス粒子の解析を提案していただきました。宇宙線分野において非常に重要な研究に携わることができ、とても充実した研究生活を送ることができました。研究について何をやっていくか悩んでいた際にも発破をかけていただき、さまざまなアプローチを提案していただけたおかげで、実のある結果を出すことができました。

常定芳基教授にはゼミや進捗報告を通して、宇宙物理学の研究者としての基礎知識、また日本語の基礎についてご指導をいただきました。初めての読み合わせの際に、準備不足だった私を追い返してくださったことで、このままではいけないと生まれ変わることができました。この3年間で私も少しはまともな言葉を使えるようになったのではないかと考えています。

金智賢准教授には、TA 実験のプロフェッショナルとして、解析について多くのアドバイスをいただきました。韓国語や英語について、また文化について、私の知らないことを多く学べたことがとても嬉しかったです。

申興秀特任助教にはデータ解析の面で大変お世話になりました。申さんがいらっしゃらなければ、私は卒業研究の段階で詰んでいました。困っていることを伝えるといつでもサポートしていただけたおかげで、本研究の土台を作ることができました。麻雀セット、ありがとうございました。

櫻井駿介特任助教には、日頃の研究や機械学習関係で大変お世話になりました。機械学習を使った研究者として右も左も知らなかった私がこの研究を進められたのは櫻井さんから多くのアドバイスをいただいたからです。ご飯もたくさんご馳走になりました。今度はご馳走させていただきます。

Telescope Array 実験の共同研究者の皆様には、進捗報告会の場などでご指導をいただき、大変お世話になりました。貴重な経験を積めたことで人として成長できたと感じています。Academia Sinica の藤末紘三特別研究員には、本研究を進める際に、個別でミーティングの時間を確保していただき、シミュレーションについてたくさんのお話を教えていただきました。初めて訪れる台湾の研究会でも大変お世話になりました。

研究室の方々、また研究室をご卒業された先輩方には、研究だけではない日々の生活、毎週の進捗報告会などで大変お世話になりました。Fraser William Bradfield 氏には機械学習の先輩として、また初めての海外の友人として大変お世話になりました。またテニスを教えてください。古前壱朗氏、河内祐輔氏、小林拓郎氏、古坊龍一氏、中原美紅氏には、研究室の先輩として大変お世話になりました。この論文を書き上げられたのは1つ上の先輩方のおかげです。後輩の橘春花氏、中川心媛氏、中本寛

輝氏、松下治斗氏、板谷臯生氏、田中花歩氏、湯淺圭太氏には研究のモチベーションを上げていただきました。熱心に研究に励むみんなを見て、自分も頑張ろうと思いました。同期の石井達希氏、太田新一郎氏は研究グループは違えど、共に戦う仲間としてとても心強かったです。また遊びましょう。能勢幸弘氏は学部時代の1年間でしたが、とても優秀な姿を見せられ、励みになりました。敷田淳氏とは真の同期として一緒に研究ができて嬉しかったです。お互いにこれからも頑張りましょう。

最後に、研究生生活を支えていただいた家族、友人に深く感謝申し上げます。特に母には大変迷惑をかけました。これからもお世話になります。3年間ありがとうございました。

## 引用文献

- [1] R. L. Workman et al., “Review of Particle Physics”, *Progress of Theoretical and Experimental Physics* 2022.8 (2022), 083C01 (2022), DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097), URL: <https://doi.org/10.1093/ptep/ptac097>.
- [2] A. M. Hillas, “The Origin of Ultra-High-Energy Cosmic Rays”, *Annual Review of Astronomy and Astrophysics* 22 (1984), 425–444 (1984).
- [3] R. U. Abbasi et al., “Indications of Intermediate-Scale Anisotropy of Cosmic Rays with Energy Greater Than 57 EeV in the Northern Sky Measured with the Telescope Array Experiment”, *The Astrophysical Journal Letters* 790 (2014), L21 (2014).
- [4] A. Addazi et al., “Quantum gravity phenomenology at the dawn of the multi-messenger era—A review”, *Progress in Particle and Nuclear Physics* 125 (2022), 103948 (2022), DOI: [10.1016/j.pnpnp.2022.103948](https://www.sciencedirect.com/science/article/pii/S0146641022000096), URL: <https://www.sciencedirect.com/science/article/pii/S0146641022000096>.
- [5] B. T. Zhang et al., “Ultraheavy Ultrahigh-Energy Cosmic Rays”, 2024, arXiv: [2405.17409](https://arxiv.org/abs/2405.17409) [astro-ph.HE], URL: <https://arxiv.org/abs/2405.17409>.
- [6] J. Matthews, “A Heitler model of extensive air showers”, *Astroparticle Physics* 22.5-6 (2005), 387–397 (2005), DOI: [10.1016/j.astropartphys.2004.09.003](https://doi.org/10.1016/j.astropartphys.2004.09.003).
- [7] R. U. Abbasi et al., “An extremely energetic cosmic ray observed by a surface detector array”, *Science* 382.6673 (2023), 903–907 (2023), DOI: [10.1126/science.abo5095](https://doi.org/10.1126/science.abo5095).
- [8] D. J. Bird et al., “The Cosmic-Ray Energy Spectrum Observed by the Fly’s Eye”, *The Astrophysical Journal* 441 (1995), 144 (1995), DOI: [10.1086/175344](https://doi.org/10.1086/175344).
- [9] K. Greisen, “End to the Cosmic-Ray Spectrum?”, *Phys. Rev. Lett.* 16 (1966), 748–750 (1966), DOI: [10.1103/PhysRevLett.16.748](https://doi.org/10.1103/PhysRevLett.16.748).
- [10] G. T. Zatsepin and V. A. Kuzmin, “Upper Limit of the Spectrum of Cosmic Rays”, *JETP Lett.* 4 (1966), 78 (1966).
- [11] P. Sarmah et al., “Constraining the superheavy dark matter origin of ultrahigh-energy cosmic rays with the Amaterasu event”, *Physical Review D* 111.8 (2025), 083048 (2025), DOI: [10.1103/PhysRevD.111.083048](https://journals.aps.org/prd/abstract/10.1103/PhysRevD.111.083048), URL: <https://journals.aps.org/prd/abstract/10.1103/PhysRevD.111.083048>.

- [12] N. Bourriche and F. Capel, “Beyond the Local Void: A Data-driven Search for the Origins of the Amaterasu Particle”, *The Astrophysical Journal* 997.2 (2026), 264 (2026), DOI: [10.3847/1538-4357/ae2c89](https://doi.org/10.3847/1538-4357/ae2c89), URL: <https://iopscience.iop.org/article/10.3847/1538-4357/ae2c89>.
- [13] H. Kawai et al., “Telescope Array Experiment”, *Nuclear Physics B - Proceedings Supplements* 175-176 (2008), 221–226 (2008), DOI: [10.1016/j.nuclphysbps.2007.11.037](https://doi.org/10.1016/j.nuclphysbps.2007.11.037).
- [14] T. Abu-Zayyad et al., “The surface detector array of the Telescope Array experiment”, *Nuclear Instruments and Methods in Physics Research Section A* 689 (2012), 87–97 (2012), DOI: [10.1016/j.nima.2012.05.079](https://doi.org/10.1016/j.nima.2012.05.079).
- [15] H. Tokuno et al., “New air fluorescence detectors employed in the Telescope Array experiment”, *Nuclear Instruments and Methods in Physics Research Section A* 676 (2012), 54–65 (2012), DOI: [10.1016/j.nima.2012.02.044](https://doi.org/10.1016/j.nima.2012.02.044).
- [16] R. U. Abbasi et al., “Depth of Ultra High Energy Cosmic Ray Induced Air Shower Maxima Measured by the Telescope Array Fluorescence Detectors”, *The Astrophysical Journal* 858.2 (2018), 76 (2018), DOI: [10.3847/1538-4357/aabad7](https://doi.org/10.3847/1538-4357/aabad7).
- [17] T. Abu-Zayyad et al., “The surface detector array of the Telescope Array experiment”, *Nuclear Instruments and Methods in Physics Research Section A* 689 (2013), 87–97 (2013), DOI: [10.1016/j.nima.2012.05.079](https://doi.org/10.1016/j.nima.2012.05.079).
- [18] M. Takeda et al., “Energy determination in the Akeno Giant Air Shower Array experiment”, *Astroparticle Physics* 19.4 (2003), 447–462 (2003), DOI: [10.1016/S0927-6505\(02\)00243-8](https://doi.org/10.1016/S0927-6505(02)00243-8).
- [19] T. Abu-Zayyad et al., “The Cosmic-Ray Energy Spectrum Observed with the Surface Detector of the Telescope Array Experiment”, *The Astrophysical Journal Letters* 768.1 (2013), L1 (2013), DOI: [10.1088/2041-8205/768/1/L1](https://doi.org/10.1088/2041-8205/768/1/L1).
- [20] R. U. Abbasi et al., “Mass composition of ultrahigh-energy cosmic rays with the Telescope Array Surface Detector data”, *Physical Review D* 99 (2 2019), 022002 (2019), DOI: [10.1103/PhysRevD.99.022002](https://doi.org/10.1103/PhysRevD.99.022002).
- [21] J. Albright et al., “Isotropy of Cosmic Rays beyond  $10^{20}$  eV Favors Their Heavy Mass Composition”, *Physical Review Letters* 133 (4 2024), 041001 (2024), DOI: [10.1103/PhysRevLett.133.041001](https://doi.org/10.1103/PhysRevLett.133.041001).
- [22] T. Abu-Zayyad et al., “CORSIKA Simulation of the Telescope Array Surface Detector”, 2014, arXiv: [1403.0644](https://arxiv.org/abs/1403.0644) [astro-ph.IM], URL: <https://arxiv.org/abs/1403.0644>.
- [23] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, tech. rep. FZKA 6019, Forschungszentrum Karlsruhe, 1998.
- [24] T. Pierog et al., “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”, *Physical Review C* 92 (2015), 034906 (2015), DOI: [10.1103/PhysRevC.92.034906](https://doi.org/10.1103/PhysRevC.92.034906).

- [25] S. Ostapchenko, “Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model”, *Physical Review D* 83 (2011), 014018 (2011), DOI: [10.1103/PhysRevD.83.014018](https://doi.org/10.1103/PhysRevD.83.014018).
- [26] P. Billoir, “Sampling procedure for the regeneration of particles in extensive air showers (EM showers)”, *Astroparticle Physics* 30.5 (2008), 270–285 (2008), DOI: [10.1016/j.astropartphys.2008.10.002](https://doi.org/10.1016/j.astropartphys.2008.10.002).
- [27] D. Ivanov, “Energy spectrum measured by the Telescope Array surface detector”, PhD thesis, University of Utah, 2012, URL: [http://www.physics.utah.edu/~thomson/talks/Ivanov\\_Thesis.pdf](http://www.physics.utah.edu/~thomson/talks/Ivanov_Thesis.pdf).
- [28] B. T. Stokes et al., “Dethinning extensive air shower simulations”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 678 (2012), 121–128 (2012), DOI: [10.1016/j.nima.2012.02.039](https://doi.org/10.1016/j.nima.2012.02.039).
- [29] S. Agostinelli et al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), 250–303 (2003), DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>, MIT Press, 2016.
- [31] Y. LeCun et al., “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE* 86.11 (1998), 2278–2324 (1998).
- [32] T. Akiba et al., “Optuna: A Next-generation Hyperparameter Optimization Framework”, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, 2623–2631 (2019).
- [33] M. Abadi et al., “TensorFlow: A System for Large-Scale Machine Learning”, *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, 265–283 (2016).
- [34] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *International Conference on Learning Representations (ICLR)*, 2015.
- [35] L. Breiman, “Random Forests”, *Machine Learning* 45.1 (2001), 5–32 (2001), DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [36] M. Barandas et al., “TSFEL: Time Series Feature Extraction Library”, *SoftwareX* 11 (2020), 100456 (2020), DOI: [10.1016/j.softx.2020.100456](https://doi.org/10.1016/j.softx.2020.100456).
- [37] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research* 12 (2011), 2825–2830 (2011).
- [38] G. R. Farrar and M. S. Pshirkov, “Ultra-high-energy Cosmic-Ray Production in Binary Neutron Star Mergers”, *The Astrophysical Journal Letters* 954.1 (2023), L17 (2023), DOI: [10.3847/2041-8213/aceef0](https://doi.org/10.3847/2041-8213/aceef0).
- [39] B. T. Zhang et al., “Ultraheavy Ultrahigh-Energy Cosmic Rays”, 2024, arXiv: [2405.17409](https://arxiv.org/abs/2405.17409) [astro-ph.HE].