Measurement of the Depth of Maximum of Air-Shower Profiles with energies between $10^{18.5}$ and 10^{20} eV using the Surface Detector of the Pierre Auger Observatory and Deep Learning

A. Abdul Halim[13], P. Abreu[71], M. Aglietta[53,51], I. Allekotte[1], K. Almeida Cheminant[79,78,69], A. Almela[7,12], R. Aloisio[44,45], J. Alvarez-Muñiz[77], J. Ammerman Yebra[77], G.A. Anastasi[57,46], L. Anchordoqui[84], B. Andrada[7], L. Andrade Dourado[44,45], S. Andringa[71], L. Apollonio[58,48], C. Aramo[49], P.R. Araújo Ferreira[41], E. Arnone[62,51], J.C. Arteaga Velázquez[66], P. Assis[71], G. Avila[11], E. Avocone[56,45], A. Bakalova[31], F. Barbato[44,45], A. Bartz Mocellin[83], C. Berat[35], M.E. Bertaina[62,51], G. Bhatta[69], M. Bianciotto[62,51], P.L. Biermann[a], V. Binet[5], K. Bismark[38,7], T. Bister[78,79], J. Biteau[36,k], J. Blazek[31], C. Bleve[35], J. Blümer[40], M. Boháčová[31], D. Boncioli[56,45], C. Bonifazi[8], L. Bonneau Arbeletche[22], N. Borodai[69], J. Brack[f], P.G. Brichetto Orchera[7], F.L. Briechle[41], A. Bueno[76], S. Buitink[15], M. Buscemi[46,57], M. Büsken[38,7], A. Bwembya[78,79], K.S. Caballero-Mora[65], S. Cabana-Freire[77], L. Caccianiga[58,48], F. Campuzano[6], R. Caruso[57,46], A. Castellina[53,51], F. Catalani[19], G. Cataldi[47], L. Cazon[77], M. Cerda[10], B. Čermáková[40], A. Cermenati[44,45], J.A. Chinellato[22], J. Chudoba[31], L. Chytka[32], R.W. Clay[13], A.C. Cobos Cerutti[6], R. Colalillo[59,49], M.R. Coluccia[47], R. Conceição[71], A. Condorelli[36], G. Consolati[48,54], M. Conte[55,47], F. Convenga[56,45], D. Correia dos Santos[27], P.J. Costa[71], C.E. Covault[82], M. Cristinziani[43], C.S. Cruz Sanchez[3], S. Dasso[4,2], K. Daumiller[40], B.R. Dawson[13], R.M. de Almeida[27], B. de Errico[27], J. de Jesús[7,40], S.J. de Jong[78,79], J.R.T. de Mello Neto[27], I. De Mitri[44,45], J. de Oliveira[18], D. de Oliveira Franco[47], F. de Palma[55,47], V. de Souza[20], E. De Vito[55,47], A. Del Popolo[57,46], O. Deligny[33], N. Denner[31], L. Deval[40,7], A. di Matteo[51], J.A. do[13,68], M. Dobre[72], C. Dobrigkeit[22], J.C. D'Olivo[67], L.M. Domingues Mendes[16,71], Q. Dorosti[43], J.C. dos Anjos[16], R.C. dos Anjos[26], J. Ebr[31], F. Ellwanger[40], M. Emam[78,79], R. Engel[38,40], I. Epicoco[55,47], M. Erdmann[41], A. Etchegoyen[7,12], C. Evoli[44,45], H. Falcke[78,80,79], G. Farrar[86], A.C. Fauth[22], T. Fehler[43], F. Feldbusch[39], F. Fenu[40,h], A. Fernandes[71], B. Fick[85], J.M. Figueira[7], P. Filip[38,7], A. Filipčič[75,74], T. Fitoussi[40], B. Flaggs[88], T. Fodran[78], T. Fujii[87,j], A. Fuster[7,12], C. Galea[78], B. García[6], C. Gaudu[37], A. Gherghel-Lascu[72], P.L. Ghia[33], U. Giaccari[47], J. Glombitza[41,i], F. Gobbi[10], F. Gollan[7], G. Golup[1], M. Gómez Berisso[1], P.F. Gómez Vitale[11], J.P. Gongora[11], J.M. González[1], N. González[7], D. Góra[69], A. Gorgi[53,51], M. Gottowik[40], F. Guarino[59,49], G.P. Guedes[23], E. Guido[43], L. Gülzow[40], S. Hahn[38], P. Hamal[31], M.R. Hampel[7], P. Hansen[3], D. Harari[1], V.M. Harvey[13], A. Haungs[40], T. Hebbeker[41], C. Hojvat[d], J.R. Hörandel[78,79], P. Horvath[32], M. Hrabovský[32], T. Huege[40,15], A. Insolia[57,46], P.G. Isar[73], P. Janecek[31], V. Jilek[31], J.A. Johnsen[83], J. Jurysek[31], K.-H. Kampert[37], B. Keilhauer[40], A. Khakurdikar[78], V.V. Kizakke Covilakam[7,40], H.O. Klages[40], M. Kleifges[39], F. Knapp[38], J. Köhler[40], F. Krieger[41], N. Kunka[39], B.L. Lago[17], N. Langner[41], M.A. Leigui de Oliveira[25], Y. Lema-Capeans[77], A. Letessier-Selvon[34], I. Lhenry-Yvon[33], L. Lopes[71], L. Lu[89], Q. Luce[38], J.P. Lundquist[74], A. Machado Payeras[22], M. Majercakova[31], D. Mandat[31], B.C. Manning[13], P. Mantsch[d], F.M. Mariani[58,48], A.G. Mariazzi[3], I.C. Mariş[14], G. Marsella[60,46], D. Martello[55,47], S. Martinelli[40,7], O. Martínez Bravo[63], M.A. Martins[77], H.-J. Mathes[40], J. Matthews[g], G. Matthiae[61,50], E. Mayotte[83], S. Mayotte[83], P.O. Mazur[d], G. Medina-Tanco[67], J. Meinert[37], D. Melo[7], A. Menshikov[39], C. Merx[40], S. Michal[31], M.I. Micheletti[5], L. Miramonti[58,48], S. Mollerach[1], F. Montanet[35], L. Morejon[37], K. Mulrey[78,79], R. Mussa[51], W.M. Namasaka[37], S. Negi[31], L. Nellen[67], K. Nguyen[85], G. Nicora[9], M. Niechciol[43], D. Nitz[85], D. Nosek[30], V. Novotny[30], L. Nožka[32], A. Nucita[55,47], L.A. Núñez[29], C. Oliveira[20], M. Palatka[31], J. Pallotta[9], S. Panja[31], G. Parente[77], T. Paulsen[37], J. Pawlowsky[37], M. Pech[31], J. Pękala[69], R. Pelayo[64], V. Pelgrims[14], L.A.S. Pereira[24], E.E. Pereira Martins[38,7], C. Pérez Bertolli[7,40], L. Perrone[55,47], S. Petrera[44,45], C. Petrucci[56], T. Pierog[40], M. Pimenta[71], M. Platino[7], B. Pont[78], M. Pothast[79,78], M. Pourmohammad Shahvar[60,46], P. Privitera[87], M. Prouza[31], S. Querchfeld[37], J. Rautenberg[37], D. Ravignani[7], J.V. Reginatto Akim[22], M. Reininghaus[38], A. Reuzki[41], J. Ridky[31], F. Riehn[77], M. Risse[43], V. Rizi[56,45], W. Rodrigues de Carvalho[78], E. Rodriguez[7,40], J. Rodriguez Rojo[11], M.J. Roncoroni[7], S. Rossoni[42], M. Roth[40], E. Roulet[1], A.C. Rovero[4], A. Saftoiu[72], M. Saharan[78], F. Salamida[56,45], H. Salazar[63], G. Salina[50], J.D. Sanabria Gomez[29], F. Sánchez[7], E.M. Santos[21], E. Santos[31], F. Sarazin[83], R. Sarmento[71], R. Sato[11], P. Savina[89], C.M. Schäfer[38], V. Scherini[55,47], H. Schieler[40], M. Schimassek[33], M. Schimp[37], D. Schmidt[40], O. Scholten[15,b], H. Schoorlemmer[78,79], P. Schovánek[31], F.G. Schröder[88,40], J. Schulte[41], T. Schulz[40], S.J. Sciutto[3], M. Scornavacche[7,40], A. Sedoski[7], A. Segreto[52,46], S. Sehgal[37], S.U. Shivashankara[74], G. Sigl[42], K. Simkova[15,14], F. Simon[39], R. Smau[72], R. Šmída[87], P. Sommers[e], R. Squartini[10], M. Stadelmaier[48,58,40], S. Stanič[74], J. Stasielak[69], P. Stassi[35], S. Strähnz[38], M. Straub[41], T. Suomijärvi[36], A.D. Supanitsky[7], Z. Svozilikova[31], Z. Szadkowski[70], F. Tairli[13], A. Tapia[28], C. Taricco[62,51], C. Timmermans[79,78], O. Tkachenko[31], P. Tobiska[31], C.J. Todero Peixoto[19], B. Tomé[71], Z. Torrès[35], A. Travaini[10], P. Travnicek[31], M. Tueros[3], M. Unger[40], R. Uzeiroska[37], L. Vaclavek[32], M. Vacula[32], J.F. Valdés Galicia[67], L. Valore[59,49], E. Varela[63], V. Vašíčková[37], A. Vásquez-Ramírez[29], D. Veberič[40], I.D. Vergara Quispe[3], V. Verzi[50], J. Vicha[31], J. Vink[81], S. Vorobiov[74], C. Watanabe[27], A.A. Watson[c], A. Weindl[40], L. Wiencke[83], H. Wilczyński[69], D. Wittkowski[37], B. Wundheiler[7], B. Yue[37], A. Yushkov[31], O. Zapparrata[14], E. Zas[77], D. Zavrtanik[74,75], M. Zavrtanik[75,74]

(Pierre Auger Collaboration)

[1] *Centro Atómico Bariloche and Instituto Balseiro (CNEA-UNCuyo-CONICET), San Carlos de Bariloche, Argentina*
[2] *Departamento de Física and Departamento de Ciencias de la Atmósfera y los Océanos, FCEyN, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina*
[3] *IFLP, Universidad Nacional de La Plata and CONICET, La Plata, Argentina*
[4] *Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), Buenos Aires, Argentina*
[5] *Instituto de Física de Rosario (IFIR) – CONICET/U.N.R. and Facultad de Ciencias Bioquímicas y Farmacéuticas U.N.R., Rosario, Argentina*
[6] *Instituto de Tecnologías en Detección y Astropartículas (CNEA, CONICET, UNSAM), and Universidad Tecnológica Nacional – Facultad Regional Mendoza (CONICET/CNEA), Mendoza, Argentina*
[7] *Instituto de Tecnologías en Detección y Astropartículas (CNEA, CONICET, UNSAM), Buenos Aires, Argentina*
[8] *International Center of Advanced Studies and Instituto de Ciencias Físicas, ECyT-UNSAM and CONICET, Campus Miguelete – San Martín, Buenos Aires, Argentina*
[9] *Laboratorio Atmósfera – Departamento de Investigaciones en Láseres y sus Aplicaciones – UNIDEF (CITEDEF-CONICET), Argentina*
[10] *Observatorio Pierre Auger, Malargüe, Argentina*
[11] *Observatorio Pierre Auger and Comisión Nacional de Energía Atómica, Malargüe, Argentina*
[12] *Universidad Tecnológica Nacional – Facultad Regional Buenos Aires, Buenos Aires, Argentina*
[13] *University of Adelaide, Adelaide, S.A., Australia*
[14] *Université Libre de Bruxelles (ULB), Brussels, Belgium*
[15] *Vrije Universiteit Brussels, Brussels, Belgium*
[16] *Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, RJ, Brazil*
[17] *Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Petropolis, Brazil*
[18] *Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ), Brazil*
[19] *Universidade de São Paulo, Escola de Engenharia de Lorena, Lorena, SP, Brazil*
[20] *Universidade de São Paulo, Instituto de Física de São Carlos, São Carlos, SP, Brazil*
[21] *Universidade de São Paulo, Instituto de Física, São Paulo, SP, Brazil*
[22] *Universidade Estadual de Campinas (UNICAMP), IFGW, Campinas, SP, Brazil*
[23] *Universidade Estadual de Feira de Santana, Feira de Santana, Brazil*
[24] *Universidade Federal de Campina Grande, Centro de Ciencias e Tecnologia, Campina Grande, Brazil*
[25] *Universidade Federal do ABC, Santo André, SP, Brazil*
[26] *Universidade Federal do Paraná, Setor Palotina, Palotina, Brazil*
[27] *Universidade Federal do Rio de Janeiro, Instituto de Física, Rio de Janeiro, RJ, Brazil*
[28] *Universidad de Medellín, Medellín, Colombia*
[29] *Universidad Industrial de Santander, Bucaramanga, Colombia*
[30] *Charles University, Faculty of Mathematics and Physics, Institute of Particle and Nuclear Physics, Prague, Czech Republic*
[31] *Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic*
[32] *Palacky University, Olomouc, Czech Republic*
[33] *CNRS/IN2P3, IJCLab, Université Paris-Saclay, Orsay, France*
[34] *Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), Sorbonne Université, Université de Paris, CNRS-IN2P3, Paris, France*
[35] *Univ. Grenoble Alpes, CNRS, Grenoble Institute of Engineering Univ. Grenoble Alpes, LPSC-IN2P3, 38000 Grenoble, France*
[36] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay, France*
[37] *Bergische Universität Wuppertal, Department of Physics, Wuppertal, Germany*
[38] *Karlsruhe Institute of Technology (KIT), Institute for Experimental Particle Physics, Karlsruhe, Germany*
[39] *Karlsruhe Institute of Technology (KIT), Institut für Prozessdatenverarbeitung und Elektronik, Karlsruhe, Germany*
[40] *Karlsruhe Institute of Technology (KIT), Institute for Astroparticle Physics, Karlsruhe, Germany*
[41] *RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany*
[42] *Universität Hamburg, II. Institut für Theoretische Physik, Hamburg, Germany*
[43] *Universität Siegen, Department Physik – Experimentelle Teilchenphysik, Siegen, Germany*
[44] *Gran Sasso Science Institute, L'Aquila, Italy*
[45] *INFN Laboratori Nazionali del Gran Sasso, Assergi (L'Aquila), Italy*
[46] *INFN, Sezione di Catania, Catania, Italy*
[47] *INFN, Sezione di Lecce, Lecce, Italy*
[48] *INFN, Sezione di Milano, Milano, Italy*
[49] *INFN, Sezione di Napoli, Napoli, Italy*
[50] *INFN, Sezione di Roma "Tor Vergata", Roma, Italy*
[51] *INFN, Sezione di Torino, Torino, Italy*
[52] *Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo (INAF), Palermo, Italy*
[53] *Osservatorio Astrofisico di Torino (INAF), Torino, Italy*
[54] *Politecnico di Milano, Dipartimento di Scienze e Tecnologie Aerospaziali , Milano, Italy*
[55] *Università del Salento, Dipartimento di Matematica e Fisica "E. De Giorgi", Lecce, Italy*
[56] *Università dell'Aquila, Dipartimento di Scienze Fisiche e Chimiche, L'Aquila, Italy*
[57] *Università di Catania, Dipartimento di Fisica e Astronomia "Ettore Majorana", Catania, Italy*
[58] *Università di Milano, Dipartimento di Fisica, Milano, Italy*
[59] *Università di Napoli "Federico II", Dipartimento di Fisica "Ettore Pancini", Napoli, Italy*

[60] *Università di Palermo, Dipartimento di Fisica e Chimica "E. Segrè", Palermo, Italy*
[61] *Università di Roma "Tor Vergata", Dipartimento di Fisica, Roma, Italy*
[62] *Università Torino, Dipartimento di Fisica, Torino, Italy*
[63] *Benemérita Universidad Autónoma de Puebla, Puebla, México*
[64] *Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional (UPIITA-IPN), México, D.F., México*
[65] *Universidad Autónoma de Chiapas, Tuxtla Gutiérrez, Chiapas, México*
[66] *Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México*
[67] *Universidad Nacional Autónoma de México, México, D.F., México*
[68] *Universidad Nacional de San Agustin de Arequipa, Facultad de Ciencias Naturales y Formales, Arequipa, Peru*
[69] *Institute of Nuclear Physics PAN, Krakow, Poland*
[70] *University of Łódź, Faculty of High-Energy Astrophysics,Łódź, Poland*
[71] *Laboratório de Instrumentação e Física Experimental de Partículas – LIP and Instituto Superior Técnico – IST, Universidade de Lisboa – UL, Lisboa, Portugal*
[72] *"Horia Hulubei" National Institute for Physics and Nuclear Engineering, Bucharest-Magurele, Romania*
[73] *Institute of Space Science, Bucharest-Magurele, Romania*
[74] *Center for Astrophysics and Cosmology (CAC), University of Nova Gorica, Nova Gorica, Slovenia*
[75] *Experimental Particle Physics Department, J. Stefan Institute, Ljubljana, Slovenia*
[76] *Universidad de Granada and C.A.F.P.E., Granada, Spain*
[77] *Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain*
[78] *IMAPP, Radboud University Nijmegen, Nijmegen, The Netherlands*
[79] *Nationaal Instituut voor Kernfysica en Hoge Energie Fysica (NIKHEF), Science Park, Amsterdam, The Netherlands*
[80] *Stichting Astronomisch Onderzoek in Nederland (ASTRON), Dwingeloo, The Netherlands*
[81] *Universiteit van Amsterdam, Faculty of Science, Amsterdam, The Netherlands*
[82] *Case Western Reserve University, Cleveland, OH, USA*
[83] *Colorado School of Mines, Golden, CO, USA*
[84] *Department of Physics and Astronomy, Lehman College, City University of New York, Bronx, NY, USA*
[85] *Michigan Technological University, Houghton, MI, USA*
[86] *New York University, New York, NY, USA*
[87] *University of Chicago, Enrico Fermi Institute, Chicago, IL, USA*
[88] *University of Delaware, Department of Physics and Astronomy, Bartol Research Institute, Newark, DE, USA*
[89] *University of Wisconsin-Madison, Department of Physics and WIPAC, Madison, WI, USA*

—

[a] *Max-Planck-Institut für Radioastronomie, Bonn, Germany*
[b] *also at Kapteyn Institute, University of Groningen, Groningen, The Netherlands*
[c] *School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom*
[d] *Fermi National Accelerator Laboratory, Fermilab, Batavia, IL, USA*
[e] *Pennsylvania State University, University Park, PA, USA*
[f] *Colorado State University, Fort Collins, CO, USA*
[g] *Louisiana State University, Baton Rouge, LA, USA*
[h] *now at Agenzia Spaziale Italiana (ASI). Via del Politecnico 00133, Roma, Italy*
[i] *now at ECAP, Erlangen, Germany*
[j] *now at Graduate School of Science, Osaka Metropolitan University, Osaka, Japan*
[k] *Institut universitaire de France (IUF), France*

We report an investigation of the mass composition of cosmic rays with energies from 3 to 100 EeV (1 EeV=10^{18} eV) using the distributions of the depth of shower maximum $X_{\rm max}$. The analysis relies on $\sim$50,000 events recorded by the Surface Detector of the Pierre Auger Observatory and a deep-learning-based reconstruction algorithm. Above energies of 5 EeV, the data set offers a 10-fold increase in statistics with respect to fluorescence measurements at the Observatory. After cross-calibration using the Fluorescence Detector, this enables the first measurement of the evolution of the mean and the standard deviation of the $X_{\rm max}$ distributions up to 100 EeV. Our findings are threefold:

(i) The evolution of the mean logarithmic mass towards a heavier composition with increasing energy can be confirmed and is extended to 100 EeV.

(ii) The evolution of the fluctuations of $X_{\rm max}$ towards a heavier and purer composition with increasing energy can be confirmed with high statistics. We report a rather heavy composition and small fluctuations in $X_{\rm max}$ at the highest energies.

(iii) We find indications for a characteristic structure beyond a constant change in the mean logarithmic mass, featuring three breaks that are observed in proximity to the ankle, instep, and suppression features in the energy spectrum.

I. INTRODUCTION

To understand the physics of ultra-high-energy cosmic rays (UHECRs), including their origin, the measurement of their mass composition is of fundamental importance. On the one hand, an event-by-event determination of mass enables estimations of the particle charges, which are valuable when performing arrival-direction analyses in the presence of magnetic fields. On the other hand, it provides insights into whether the observed flux suppression at the end of the cosmic-ray spectrum [1–4] is a signature of the interaction of the particles with the cosmic microwave background [5, 6], a consequence of a limit of the maximum injection energy of the cosmic accelerators [7, 8], or a combination of both [9, 10]. Whereas for the former, due to photodisintegration during the propagation, a change in the composition is expected that scales with the energy per nucleon (E/A), for the latter, the so-called Peters cycle, a change in composition scaling with rigidity (E/Z) is expected.

Due to the rapid decrease in particle flux at ultra-high energies, modern cosmic-ray observatories perform indirect measurements of the rare particles by detecting generated air showers instead. The influence of the primary mass on the shower development can be characterized mainly by the number of muons and the atmospheric depth of the shower maximum X_{max} at which the shower reaches its maximum size. With increasing primary mass, the number of induced sub-showers increases, and the energy per nucleon reduces, leading to an X_{max} higher up in the atmosphere and decreasing shower fluctuations. The increase in the number of sub-showers additionally causes an increase in the number of produced muons. Since the current generation of hadronic interaction models cannot describe the muon component in full detail [11–14], currently the most precise composition studies at ultra-high energies rely on measurements of X_{max}.

Studying the energy evolution of $\langle X_{\mathrm{max}} \rangle$, the mean of the distribution, enables us to directly examine the evolution of $\langle \ln \mathrm{A} \rangle$, the mean logarithmic mass. The evolution of the fluctuations $\sigma(X_{\mathrm{max}})$, i.e., the standard deviation of the distribution, provides additional insights into the composition and its mixing [15, 16].

In the past two decades, significant progress in our understanding of UHECRs has been made, largely attributed to the establishment of the Pierre Auger Observatory [17] and the Telescope Array Project [18]. The Pierre Auger Observatory is the world's largest cosmic ray detector and is composed of a Surface Detector (SD) and a Fluorescence Detector (FD). By observing the longitudinal shower profile of extensive air showers, the FD telescopes of the Observatory not only enable the precise determination of the shower energy but also provide an accurate determination of X_{max} [19, 20]. Currently, the most precise mass composition studies rely on these fluorescence observations. However, the operation of fluorescence telescopes is confined to dark and moonless nights, resulting in a duty cycle of around 15%. Additionally, for an unbiased X_{max} data set, further cuts have to be applied. In contrast, the duty cycle of the SD is close to 100%, enabling composition studies with high statistics.

Recently, several methods have been developed to infer mass-sensitive information using the SD. Using the risetime of signals in the water-Cherenkov detectors, the evolution of the average mass composition as a function of energy can be studied with good precision [21]. Furthermore, the phenomenological approach of shower universality [22], based on a decomposition of the measured detector signals into the different shower components, has shown first promising results in the reconstruction of $\langle \ln \mathrm{A} \rangle$. To determine the fluctuations in X_{max}, i.e., to measure $\sigma(X_{\mathrm{max}})$, yet more precise, event-by-event measurements are needed. Measuring $\sigma(X_{\mathrm{max}})$ is particularly important as its interpretation does not depend strongly on hadronic interaction models, as a considerable part of the fluctuations depends on the mean free path of the first interaction and, thus, the cross-section at the highest energies.

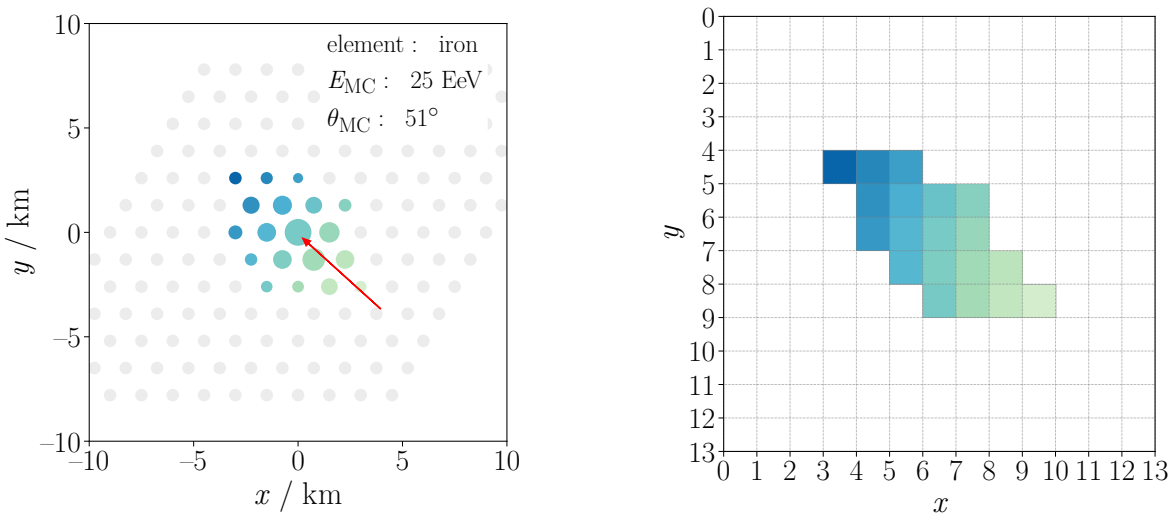
With the advent of deep learning, new possibilities have emerged for designing learning algorithms, i.e., deep neural networks (DNNs), to analyze high-dimensional and complex data in computational sciences [23, 24] as well as in physics [25]. Trained on large simulation libraries, these algorithms are capable of recognizing small patterns to which conventional methods were previously not sensitive. This recent progress provided improved reconstruction algorithms in astroparticle physics, e.g., imaging air Cherenkov telescopes [26], gravitational wave detection [27], neutrino [28–30] and cosmic-ray observatories [31], including the reconstruction of X_{max} [32, 33] and other air shower properties [34, 35]. So far, the potential of deep-learning-based methods for improved reconstruction in astroparticle physics has been demonstrated, but the application to measured data, including a comprehensive study of systematic uncertainties and associated new insights, is limited. This work aims to close this gap and shows a successful application to measured data starting at the raw detector signals.

In this article and the accompanying Letter [36], we report on the first investigation of the UHECR mass composition based on the first and second moment of the X_{max} distributions from 3 to 100 EeV using the SD. The data set, reconstructed using a novel deep-learning-based reconstruction method, offers an increase in statistics with respect to analyses based on fluorescence observations, which amounts to a factor of ten above 5 EeV.

By cross-calibrating the developed algorithm using hybrid events — events that feature SD and FD reconstruction — we find an excellent agreement with previous analyses. The new measurement of $\langle X_{\mathrm{max}} \rangle$ and $\sigma(X_{\mathrm{max}})$ up to the highest energies is subject to minor systematic uncertainties and avoids the large statistical uncertainties present in previous work. As a result, it offers new insights into the composition at ultra-high energies.

II. THE PIERRE AUGER OBSERVATORY

The Pierre Auger Observatory, fully commissioned in 2008, is located in the Pampa Amarilla in Argentina at an altitude of $\sim$1500 m, which corresponds to about 875 $\mathrm{g\,cm^{-2}}$ of atmospheric overburden. The SD of the Observatory [17] com-



(a) (b)

FIG. 1: Simulated footprint of a cosmic-ray event measured by the SD. (a) Cutout of 13×13 stations of an SD event containing the induced signal pattern on the triangular grid. The marker size indicates the logarithm of the total measured signal, the color denotes the arrival time (green for early, blue for late), the arrow marks the projection of the shower axis on the ground, and its tip denotes the shower core. (b) Representation of the event on a Cartesian grid after pre-processing as a cutout with dimensions 13×13 as used for the DNN after axial indexing. The color indicates the arrival time of the shower front at each station.

prises an array of 1660 water-Cherenkov detectors (WCDs) placed on a triangular grid with a spacing of 1500 m and covering an area of about 3000 km^2. Each WCD is composed of a sealed liner with a diameter of 3.6 m and a height of 1.2 m filled with 12,000 liters of ultra-pure water. Three 9-inch photomultiplier tubes (PMTs) look downward through transparent windows into the water volume to record the Cherenkov light of relativistic charged particles penetrating the walls. The signal measured by each PMT is digitized by a 40 MHz Flash Analog-to-Digital Converter, corresponding to a bin width of 25 ns in time. Due to the limited available bandwidth, the time-dependent signals, i.e., *signal traces*, are only collected if a signal was measured in at least three WCD stations in temporal and spatial coincidence. In addition, the current parameters for calibrating the signals into units of VEM [37] (vertical equivalent muon) — defined as the average signal of a single muon induced when passing the detector vertically through the center of the tank — are sent. These are updated every minute and provide reliable signal sizes even during strongly varying operation conditions. This in situ calibration, together with solar-powered electronics and a battery, offers a duty cycle of the SD close to 100%.

The SD array is overlooked by 27 telescopes located at four different sites at the borders of the Observatory. Three sites host six, and one hosts nine Schmidt telescopes, each composed of a 13 m^2 mirror and a 440-pixel camera to observe the longitudinal shower development using the isotropically-emitted fluorescence light. At the Coihueco site, three High Elevation Auger Telescopes are used to detect low-energy (down to 10^{17} eV) showers. To ensure the most precise observations, the atmospheric conditions are monitored using probing beams of two laser facilities placed close to the center of the array. For more details on the design and operation of the Observatory we refer to Ref. [17].

A. Surface Detector data

A typical air shower with a zenith angle below 60° and $E > 10$ EeV induces a footprint with the size of tens of square kilometers at the Earth's surface, on average triggering around ten stations. See Fig. 1a, for a simulated example event. For each triggered station, three signal traces are recorded, one measured by each PMT. The trace has a length of 768 time steps of 25 ns, resulting in a total length of ∼20 µs. These traces are further processed with a peak finder to search for the signal window. In this work, the signal window has a width of 120 time steps (3 µs), which includes more than 99% of the signals. Simulated example traces of the event shown in Fig. 1 are depicted in Fig. 2 for stations located at three different distances to the shower core. Note that, in contrast to the standard reconstruction, which integrates over the signal window to estimate the shower energy, in this work, we make use of the full signal trace for the X_{max} reconstruction.

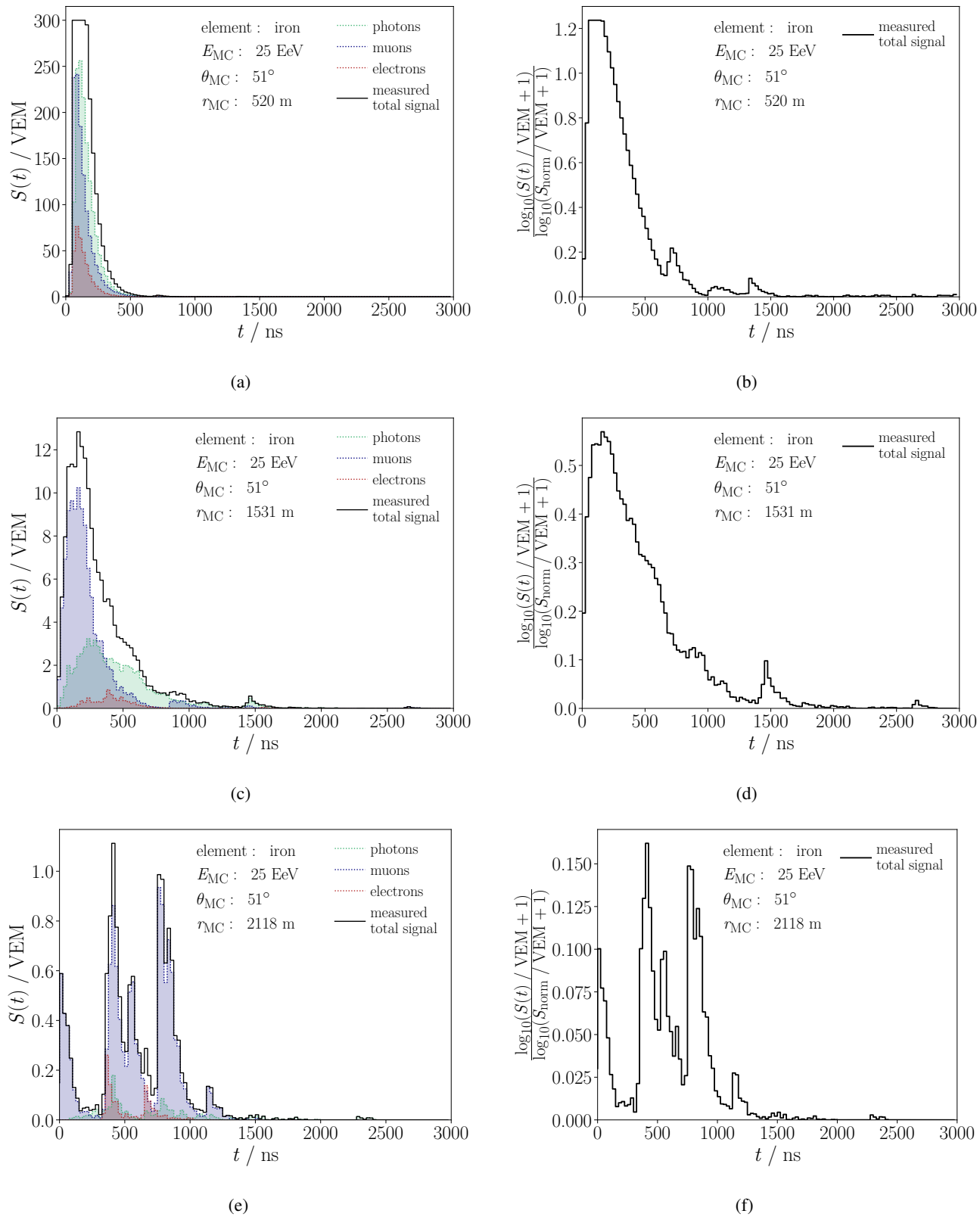


FIG. 2: Simulated signal traces of a cosmic-ray event in the SD stations before (left) and after the pre-processing (right). (a) Simulated signal trace of a station close to the shower core (top), at a distance of around 1000 m (middle), and at a distance of around 3500 m to the shower core (bottom). Different colors indicate signals from different shower components. The black line denotes the total measured signal, including the saturation effects of the electronics that are only simulated for the sum of all shower components.

TABLE I: Basic and analysis-specific selections (separated by a line) for the SD data set.

Cut	Events	ε (%)
reconstructed vertical event ($\theta < 60°$)	5,994,712	—
is 6T5	4,858,291	81.1
$\log_{10}(E/\mathrm{eV}) > 18.5$	133,167	2.7
hardware status	129,403	97.2
station start slot	128,308	99.2
$2.75 < A/P < 3.45$	126,033	98.2
$\bar{S}_{\mathrm{tot}} > 5$ VEM in surrounding hexagon	125,828	99.8
350 m < core distance < 1000 m	101,392	80.6
fiducial SD cut	48,824	48.1

TABLE II: Selections for hybrid data.

Cut	Events	ε (%)
number of events	25,076	—
telescope cuts	19,733	78.7
hardware status	16,916	85.7
aerosols/clouds	9,822	58.1
hybrid geometry	9,157	93.2
fiducial FoV cut	3,497	38.2
profile cuts	3,331	95.3
passed SD selection	3,086	92.6
analysis-specific cuts	1,642	53.2

In addition to the three traces, the *arrival time* of the shower front at each station is estimated based on the starting time of the signal window and the trigger time of each WCD station. These arrival times, combined with the station positions, encode information on the arrival direction of the primary particle. They are used in the standard reconstruction to determine the shower axis by fitting a model of an inflating sphere [38]. The algorithm for the reconstruction of X_{max} relies on the position of the triggered stations and on both the arrival times and the signal traces measured at each station.

B. Data selection

The data set for the measurement of the cosmic-ray mass composition consists of air shower events recorded with the SD. Additionally, for the calibration and the validation of the reconstruction, hybrid measurements — events detected by both the SD and the FD — are utilized.

1. SD data set

The data selection for mass composition studies mostly follows the criteria used for determining the energy spectrum [39] and is summarized in Table I. As pre-selection criteria, we require a successful energy reconstruction, a zenith angle $< 60°$ to consider vertical showers only, and exclude lightning-induced events. We further require that the stations with the largest measured signals are surrounded by six working stations (a so-called 6T5 trigger) to ensure that the footprint is sufficiently sampled by the SD and that the events with shower cores outside the array are rejected. In this analysis, we only consider events with $\log_{10}(E/\mathrm{eV}) > 18.5$, where the SD is fully efficient in the selected zenith angle range, and keep only events when the SD is properly operational.

In the analysis-specific post-selection, we remove a minor fraction of events where the starting bin of any single signal trace seems to be mis-reconstructed by the peak finder. We only accept events with an average integrated signal $\bar{S}_{\mathrm{tot}} >$ 5 VEM, for all triggered stations surrounding the station with the largest signal to ensure an adequate measurement of the signal trace. Additionally, we reject events with very large and very low area-over-peak (A/P) ratios that cannot be linearly calibrated during the aging calibration, which is discussed in Section III D 1. We additionally remove events with small (350 m) and large (1000 m) distances of the reconstructed shower core to the station with the largest integrated signal to remove events with saturated stations and station multiplicities challenging to reconstruct, respectively. Since the multiplicity of triggered stations is small at low energies, and the sampling fluctuations of the WCDs are — due to the smaller particle density — high, the X_{max} reconstruction bias depends on the zenith angle and the energy. To obtain an unbiased dataset, we therefore only accept events where the composition bias is small. This fiducial selection was derived using simulation and is discussed in Section III B. It is strict at energies below 10 EeV, causing, due to the steeply falling energy spectrum, a low overall selection efficiency. However, at high energies, it hardly causes any statistical disadvantage for the energy-dependent study of the UHECR mass composition. In total, the SD data set contains, after selection, 48,824 events recorded between 1 January 2004 and 31 August 2018.

2. Hybrid measurements

For the calibration of the DNN to the X_{max} scale of the FD, hybrid events with a high-quality reconstruction of the SD and FD data are used. Thus, besides the selection of the SD events, FD cuts are applied to this data set that follow the selection used in previous composition analyses [20]. The selection is summarized in Table II. The pre-selection ensures good data-taking conditions by accepting only events with a stable gain calibration of the FD PMTs and adequate observation conditions, i.e., featuring a clear sky and a measurement of the vertical aerosol optical depth within the last hour that guarantees precise measurements. To ensure an adequate air-shower reconstruction, we require a good fit of the Gaisser–Hillas profile, a minimum observed track length of 200 $\mathrm{g\,cm^{-2}}$, and the X_{max} to be reconstructed in the field of view of the telescope with an X_{max} uncertainty smaller than 40 $\mathrm{g\,cm^{-2}}$. Since the condition on X_{max} to fall in the field of view constrains X_{max} and, thus, results in an acceptance that depends on the mass of the primary particle, a fiducial field-of-view cut is applied. The cut is derived in a data-driven fashion and ensures that only shower geometries that provide unbiased views of the bulk of the X_{max} distribution are selected. This strict crite-

rion removing a significant fraction of events guarantees an unbiased, i.e., composition-independent data sample and has an efficiency of slightly less than 40%. See Ref. [20] for more details. In addition, we remove events with holes in the profile that exceed 20% of the observed track length, and those events with an uncertainty on the energy reconstruction above 12%. In total, 3,331 events remain after the FD selection. After applying the same SD selection as described above, the hybrid data set comprises 1,642 events measured between 1 January 2004 and 31 December 2017.

III. RECONSTRUCTION OF THE SHOWER MAXIMUM USING THE SURFACE DETECTOR AND DEEP LEARNING

The previous reconstruction of X_{max} on an event-by-event basis was confined to fluorescence telescope data. To obtain high-statistic measurements of the UHECR mass composition at the highest energies, the reconstruction of X_{max} using the SD is a promising solution. The reconstruction is challenging as, in contrast to the FD, the SD does not directly measure the longitudinal shower development — enabling straightforward observations of X_{max}— but subsamples the particle density of the particle cascade at the ground. To infer information on the shower development, the temporal structure of the particle footprint has to be exploited. On the one hand, different particles induce different signals in the WCDs [34, 40], e.g., a single muon typically induces a clear spike as it crosses the station in a straight line (cf. Fig. 2e). Thus, the SD signals contain information on the absolute density of the respective content. Additionally, the temporal structure of the signals encodes information on the shower development. For example, γ, e^+, and e^- that form the electromagnetic component undergo multiple scattering when penetrating the atmosphere, leading to delay and broadening of the signals (cf. Fig. 2a and cf. Fig. 2c), which scale with the distance to X_{max}. The temporal structure of the measured signal in a single station, however, is more complex as it further depends on additional kinematics like the energy and mass of the primary cosmic ray, the zenith angle, and the distance of the station to the shower core. Previous approaches rely on measuring the signal risetimes [21] and thus provide insights into the muon content. However, this data-driven approach does not consider all available information on the shower development. The complex temporal and spatial information in the SD signals are intractable to analyze using analytical models. Therefore, complicated parameterizations are needed that rely on simulation libraries. The phenomenological approach of air-shower universality [22, 41] utilizes simplifications in order to parameterize and decompose the expected signals, limiting the performance of the algorithm, especially when exploiting the temporal structures of signals with strong fluctuations (signal spikes) beyond the average. Thus, in this work, we use an alternative approach based on deep neural networks (DNNs).

A. Deep-learning-based reconstruction

The DNN trained for the reconstruction of X_{max} is based on the signal traces measured using the WCDs of the triangular grid of the SD array and the arrival times. To process the temporal and spatial structure of the particle footprint, the DNN uses the following architecture methodology, separating the analysis in space and time. Since the SD grid is triangular with a regular spacing of 1500 m, we use the axial representation for re-indexing into a Cartesian grid [31]. For a memory-efficient re-indexing, we use a cutout of 13×13 stations, where the station with the largest signal defines the center of this grid. The dimensions of 13×13 stations guarantee that, on average, more than 99.9% of the triggered stations per event are contained within this sub-array. See Fig. 1a that visualizes this process using an example SD event.

The time traces $S(t)$ at each time step t are re-scaled using a logarithmic transformation

$$\tilde{S}_i(t) = \frac{\log_{10}\left(S_i(t)/\mathrm{VEM}+1\right)}{\log_{10}\left(S_{\mathrm{norm}}/\mathrm{VEM}+1\right)} \tag{1}$$

that maps stations with a large signal of $S_{\mathrm{norm}} = 100$ VEM to 1 and maintains the physical property that non-triggered stations keep zero signals. This normalization stabilizes the training process of the DNN. In a similar way the shower arrival time $t_{0,i}$ at each WCD is normalized with respect to the arrival time τ_{center} measured at the station with the largest signal, i.e., the center of the cutout, and the standard deviation $\sigma_{t,\mathrm{data}} = 48.97$ ns of the arrival times estimated over the whole training data set,

$$\tilde{t}_{0,i} = \frac{t_{0,i} - \tau_{\mathrm{center}}}{\sigma_{t,\mathrm{data}}}. \tag{2}$$

To analyze the temporal structure of the signal traces, recurrent long short-term memory (LSTM) layers [42] are utilized in the first part of the network. The identical network subpart with the same adaptive parameters is applied to each signal trace, i.e., we apply *weight sharing* along all stations as similar particles induce similar responses in the detector. The output of this network can be interpreted as an image of 13×13 pixels (stations) with ten channels instead of three in a natural RGB image, as the recurrent network part characterizes the traces of each station into ten features. These features are the input for the convolutional part and are concatenated with the channel of arrival times and an additional channel characterizing the detector states of the surrounding WCDs (working/not working). The next stage is based on an advanced type of convolutional neural network (CNN) [23] to exploit the spatial structure of the event. We make use of the so-called HexaConv layers [43] and residual connections [44, 45], which extends the principle of a filter sliding along an image by a rotation. This is a meaningful extension as the induced signal patterns are to first order independent of the azimuth angle. Finally, after a ResNet-like architecture, a single node for the prediction of X_{max} forms the output of the DNN. For a detailed description of the DNN architecture, we refer to Ref. [32].

The network was trained using a library [46, 47] of 400,000 events with equal fractions of proton-, helium-, oxygen-, and

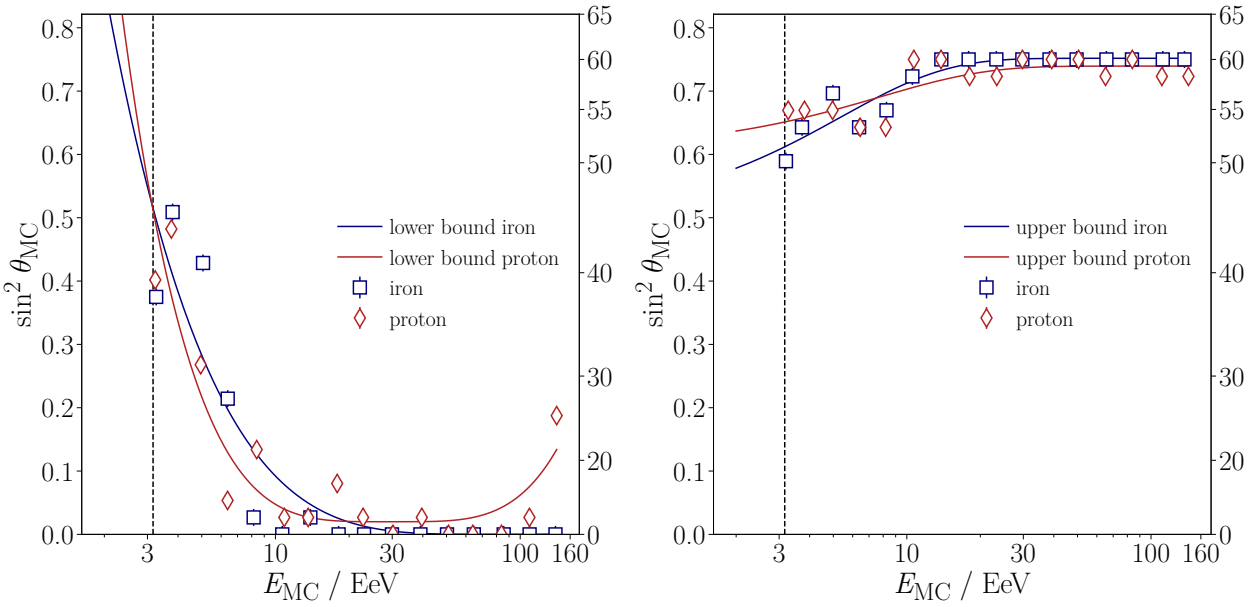


FIG. 3: Determination of the fiducial selection with events simulated using the Sibyll2.3c hadronic interaction model. We show the lower (left) and upper (right) bounds of the selection as derived using a pure proton (red) and pure iron (blue) composition. The markers at a given energy indicate the minimum (lower bound) and maximum (upper bound) zenith angle bin where the reconstruction bias is less than 10 g cm^{-2}. Derived parameterizations are shown as continuous curves. In the ongoing analysis, only events between the upper and lower bounds are accepted.

iron-induced showers in an energy range of 1 to 160 EeV with a spectral index of $\gamma = -1$ simulated using CORSIKA [48] with the hadronic interaction model EPOS-LHC [49] and the FLUKA model [50]. We use only events with zenith angles $\theta < 60^\circ$ and the full azimuth range ($0^\circ - 360^\circ$). During the training, we perform on-the-fly augmentation of the data using varying detector states[1] to increase the diversity of our data and mimic real operational conditions. Technical details of the training and the model can be found in Ref. [32].

B. Fiducial event selection

The mass-composition analysis in this work relies on the first and second moments of the measured $X_{\rm max}$ distributions and their energy evolution. An unbiased selection of the reconstructed events has to be ensured for a precise determination of the moments. In contrast to the FD, $X_{\rm max}$ cannot directly be observed using the SD but needs to be inferred from the time-resolved particle density at the ground. Due to the attenuation of the particle density for increased distances between the shower maximum and the detector plane, which further scales with the zenith angle, the amount of information encoded in the sampled signals depends on the shower geometry and the energy. For example, at very low energies, there are fewer particles in the shower, and $X_{\rm max}$ is farther away from the detector. This will lead to a smaller particle density at the ground, i.e., fewer triggered stations, and fewer particles arriving per station to be analyzed by the DNN, making the already challenging measurement of $X_{\rm max}$ intractable.

To avoid selections depending on $X_{\rm max}$, and thus the composition itself, we derive upper and lower zenith angle bounds for the selection of air-shower events as a function of energy. We scan the reconstruction bias for proton and iron-induced showers[2] as a function of energy and estimate the minimum (maximum) zenith angle at which the absolute reconstruction bias is below $|\Delta X_{\rm max}| < 10$ g cm^{-2} to derive a lower (upper) bound on the zenith angle. This is visualized in Fig. 3.

At low energies, for almost vertical showers, the number of triggered stations is small (around 6), and for events with large zenith angles, the signals decrease by up to 50% due to the increased atmospheric attenuation[3] between shower maximum and ground level [17]. Whereas the energy and arrival direction can be accurately reconstructed using the SD [38], this leads to a reconstruction bias in $X_{\rm max}$. At very high energies, events can be reconstructed for zenith angles up to 60°, but for smaller angles, proton showers can develop the shower maximum below the ground, causing biased $X_{\rm max}$ reconstructions. Therefore, we accept only events if they have zenith

[1] This includes malfunctioning stations, faulty PMTs, and varying saturation thresholds of the WCD electronics.

[2] As the reconstruction of proton and iron showers is subject to the largest reconstruction biases. Also see Fig. 4

[3] At 10 EeV, $S(1000)$ decreases from around 55 VEM at 0° to 25 VEM at 60°.

angles above the lower and below the upper iron *and* proton bounds at a given energy. We find the very same dependence across the investigated hadronic interaction models EPOS-LHC, QGSJetII-04, and Sibyll2.3c [51] and use the selection derived using Sibyll2.3c for further analyses, as it results in the most conservative cut. The selection removes more than 50% of events below 5 EeV and is very relaxed above 10 EeV. Note that the cut is independent of the primary particle mass since only a cut on the zenith angle is performed. However, due to the steep cosmic ray spectrum, strict cuts at lower energies still enable statistically powerful measurements. Using this fiducial cut, high quality X_{max} measurements are ensured, and a merit factor[4] of separating proton and iron close to 1.5 can be reached [52].

C. Reconstruction of the X_{max} moments

The determination of the first two moments $\langle X_{\mathrm{max}}\rangle$ and $\sigma(X_{\mathrm{max}})$ of the X_{max} distribution and its evolution with energy relies on the X_{max} reconstruction of the DNN and the energy estimator S_{38}[5] from the standard reconstruction of SD data [38]. To examine the quality of the reconstructed X_{max} moments, both resolution and bias must be considered. Therefore, we study hereafter the reconstruction of the X_{max} moments using a forward-folding approach. The bias and resolution of the X_{max} and energy reconstruction depend on the composition and energy. The finite resolution of the energy estimator and its composition bias can cause a spillover of events into neighboring energy bins, depending on the underlying spectrum and the composition. To handle this effect, we utilize the latest measurement of the UHECR spectrum [53] and consider the trigger efficiency of the SD at low energies. We investigate this forward-folding approach for the energy evolution of $\langle X_{\mathrm{max}}\rangle$ and $\sigma(X_{\mathrm{max}})$ for three different composition scenarios following the Auger spectrum [53].

Since proton and iron showers feature the largest reconstruction biases, we study a pure proton and a pure iron composition. Note that this is a conservative approach since previous analyses strongly disfavor significant iron fractions at low and significant proton fractions at high energies [20]. As the most realistic scenario, we also use the *Auger mix*, the composition fractions derived by fitting a template of simulations to the X_{max} distributions measured using the FD [54]. Since the measurement of the FD ends at about 50 EeV, we assume the composition remains unchanged from there onward.

To finally estimate the reconstruction performance, we compare the reconstructed $\langle X_{\mathrm{max}}\rangle$ and $\sigma(X_{\mathrm{max}})$ after the forward-folding process with the injected moments from Monte-Carlo simulations. To study the composition bias, we use bootstrapping in each bin to estimate $\langle X_{\mathrm{max}}\rangle$ and its statistical uncertainty. Since a composition bias in the X_{max} reconstruction translates into an X_{max} dependence of the reconstruction bias, the variance of the reconstructed distribution can be expressed as

$$\sigma^2(X_{\mathrm{max,DNN}}) = \underbrace{\sigma^2(X_{\mathrm{max}})}_{\text{phys. fluct.}} + \underbrace{\sigma^2_{\mathrm{res}}(X_{\mathrm{max,DNN}})}_{\text{resolution}} + \underbrace{2\,\mathrm{Cov}(X_{\mathrm{max}}, X_{\mathrm{max,DNN}} - X_{\mathrm{max}})}_{X_{\mathrm{max}}\text{ dependence of bias}}.$$

Thus, to reconstruct $\sigma(X_{\mathrm{max}})$, not only the resolution but due to the composition dependence, i.e., X_{max} dependence also, the covariance term has to be considered. In the absence of sufficient information, estimators like DNNs trained with the mean-squared-error objective function tend to predict samples close to the mean since the reconstruction is ambiguous. Therefore, at low energies in particular, the DNN is likely to reconstruct, on average, iron with positive and proton with negative X_{max} bias. In turn, the covariance term is negative for the DNN, i.e., it will in part cancel the resolution term. We studied the dependence of the sum of both terms, the covariance and the resolution for various hadronic models and compositions as a function of energy. In the case of our trained DNN, we found that, to a good approximation, both terms cancel or are small in comparison to the physical fluctuations in X_{max}. Therefore, the standard deviation of the distribution formed by the DNN predictions is used as the estimate for $\sigma(X_{\mathrm{max}})$, and the statistical uncertainty on $\sigma(X_{\mathrm{max}})$ is obtained using bootstrapping. Deviations, i.e., scenarios with non-canceling contributions of the resolution and covariance, will translate into a composition bias of the second moment in this forward-folding approach and propagate into the systematic uncertainties of the $\sigma(X_{\mathrm{max}})$ measurement.

In Fig. 4, the performance in reconstructing the evolution of the moments $\langle X_{\mathrm{max}}\rangle$ (top) and $\sigma(X_{\mathrm{max}})$ (bottom) using the SD is depicted for three different scenarios. A pure proton composition is shown in red, pure iron in blue, and the Auger mix in yellow for the three hadronic interaction models EPOS-LHC (filled diamonds), QGSJetII-04 (grey-filled squares), and Sibyll2.3c (open circles). As a reference, the injected (true) moments are shown as white boxes where their vertical sizes indicate the statistical uncertainty prior to the X_{max} reconstruction by the DNN and the energy reconstruction [38]. Note that only EPOS-LHC was used as a hadronic interaction model for training the DNN. Since a large fraction of the EPOS-LHC simulations was used for the DNN training, the statistical uncertainty is larger for EPOS-LHC events than for QGSJetII-04 and Sibyll2.3c, where all simulations could be used for testing. We find that the performance in the determination of $\langle X_{\mathrm{max}}\rangle$ depends on energy and the hadronic interaction models. Above 10 EeV the performance differences across the models and mass composition scenarios are small [6].

[4] The merit factor is defined as

$$f_{\mathrm{MF}} = \frac{|\langle X_{\mathrm{max,P}}\rangle - \langle X_{\mathrm{max,Fe}}\rangle|}{\sqrt{\sigma^2(X_{\mathrm{max,P}}) + \sigma^2(X_{\mathrm{max,Fe}})}}, \tag{3}$$

[5] Defined as the signal a station measures at a distance of 1000 m to the shower core if the shower would have arrived at a zenith angle of 38°.

[6] A residual plot for the three investigated models can be found in Fig. 17

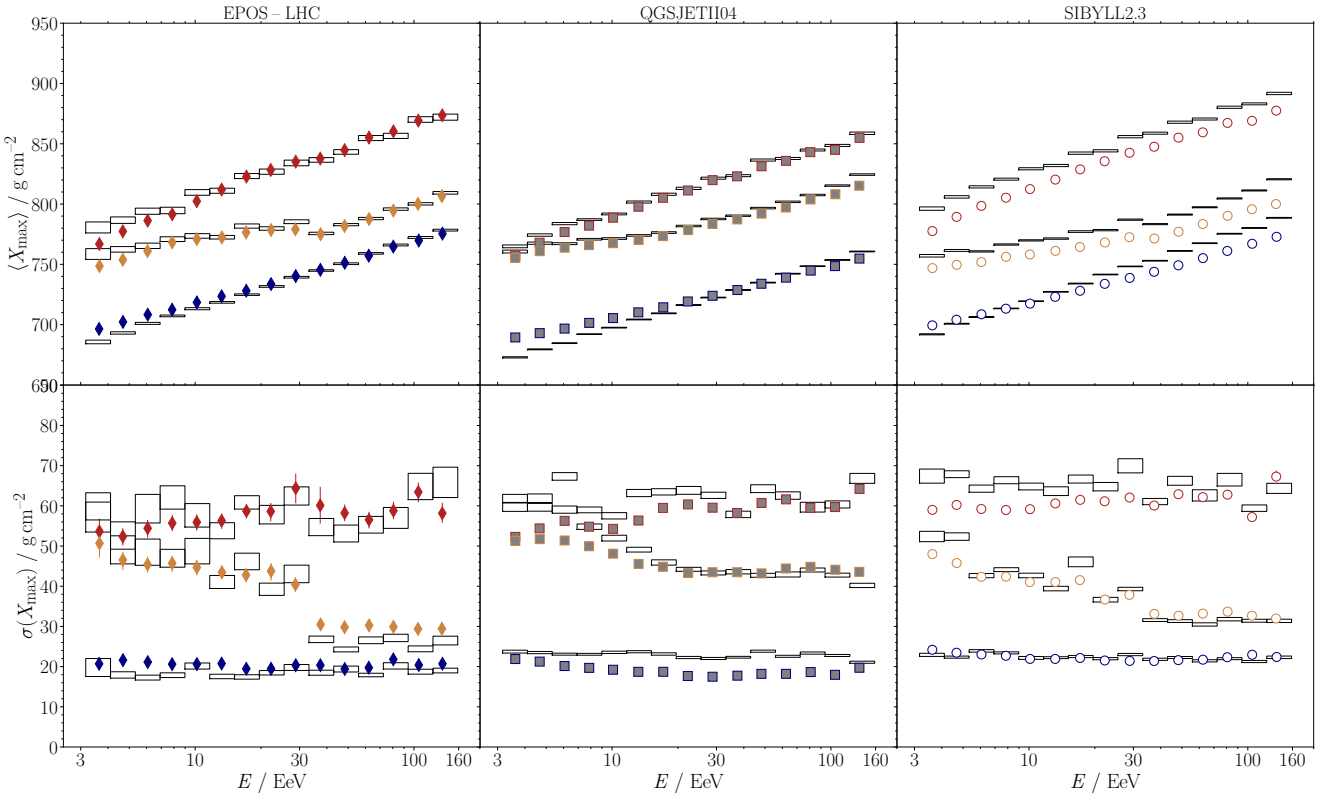


FIG. 4: Reconstructed (top) first moments and (bottom) second moments of the $X_{\max}$ distributions as a function of energy using the SD for the scenarios of a pure proton (red), Auger mix (yellow), and pure iron (blue) composition for showers simulated using EPOS-LHC (left, filled diamonds), QGSJetII-04 (center, grey-filled squares), and Sibyll2.3c (right, open circles). The injected (true) $X_{\max}$ moments, prior to the energy and $X_{\max}$ reconstruction, are shown as white boxes. The reconstruction of the DNN and systematic effects such as composition-dependent resolution and bias of the SD-based energy reconstruction are considered using forward folding. The fiducial SD selection is applied. Note that EPOS-LHC was used as a hadronic interaction model for training the DNN.

The reconstruction bias shows a dependence on the composition at low energies. At 3 EeV, the pure iron composition is subject to a bias of up to 10 to 15 $\mathrm{g\,cm^{-2}}$ that reduces with energy. For a pure proton composition, a similar dependence is visible, of up to $-10\ \mathrm{g\,cm^{-2}}$ decreasing with energy.

For the Auger mix, the composition bias is independent of energy for QGSJetII-04 and Sibyll2.3c. Only for EPOS-LHC, a small composition bias of up to $-5\ \mathrm{g\,cm^{-2}}$ can be seen below 6 EeV. Otherwise, we find an interaction-model bias of $-5\ \mathrm{g\,cm^{-2}}$ for QGSJetII-04 that applies to all compositions. For Sibyll2.3c, this bias amounts to $-12\ \mathrm{g\,cm^{-2}}$. Because EPOS-LHC was used for training, no such bias is visible for this model. Both the composition bias and the hadronic-interaction model bias would propagate into the systematic uncertainty of the $\langle X_{\max}\rangle$ measurement. However, since the Observatory features a hybrid detector design, both biases, and their energy dependence can be removed (as investigated in a simulation study) by re-calibrating the DNN measurement using fluorescence observations (as discussed in Section III D 2).

The composition bias of the $\sigma(X_{\max})$ reconstruction as seen in Fig. 4 (bottom) depends on energy and is below 5 $\mathrm{g\,cm^{-2}}$ above 10 EeV for proton and iron. For the Auger mix it is even lower at these high energies and not significant. Overall, the biases observed for reconstructing $\sigma(X_{\max})$ are small over a large range of energies. For this reason, no calibration using the FD will be performed. Furthermore, the biases found here will be transferred to the systematic uncertainty of our composition measurements. Different from the measurement of $\langle X_{\max}\rangle$, the estimation of $\sigma(X_{\max})$ is not subject to a strong dependence on the hadronic interaction model. This can be explained by the fact that a large part of the shower fluctuations depends on the fluctuations of X_1, i.e., the traversed depth prior to the first interaction, owing to the given nuclear cross sections with air molecules. Depending on the primary, the expected fluctuations in X_1 are in the order of 10 $\mathrm{g\,cm^{-2}}$ (50 $\mathrm{g\,cm^{-2}}$) for iron (proton) nuclei.

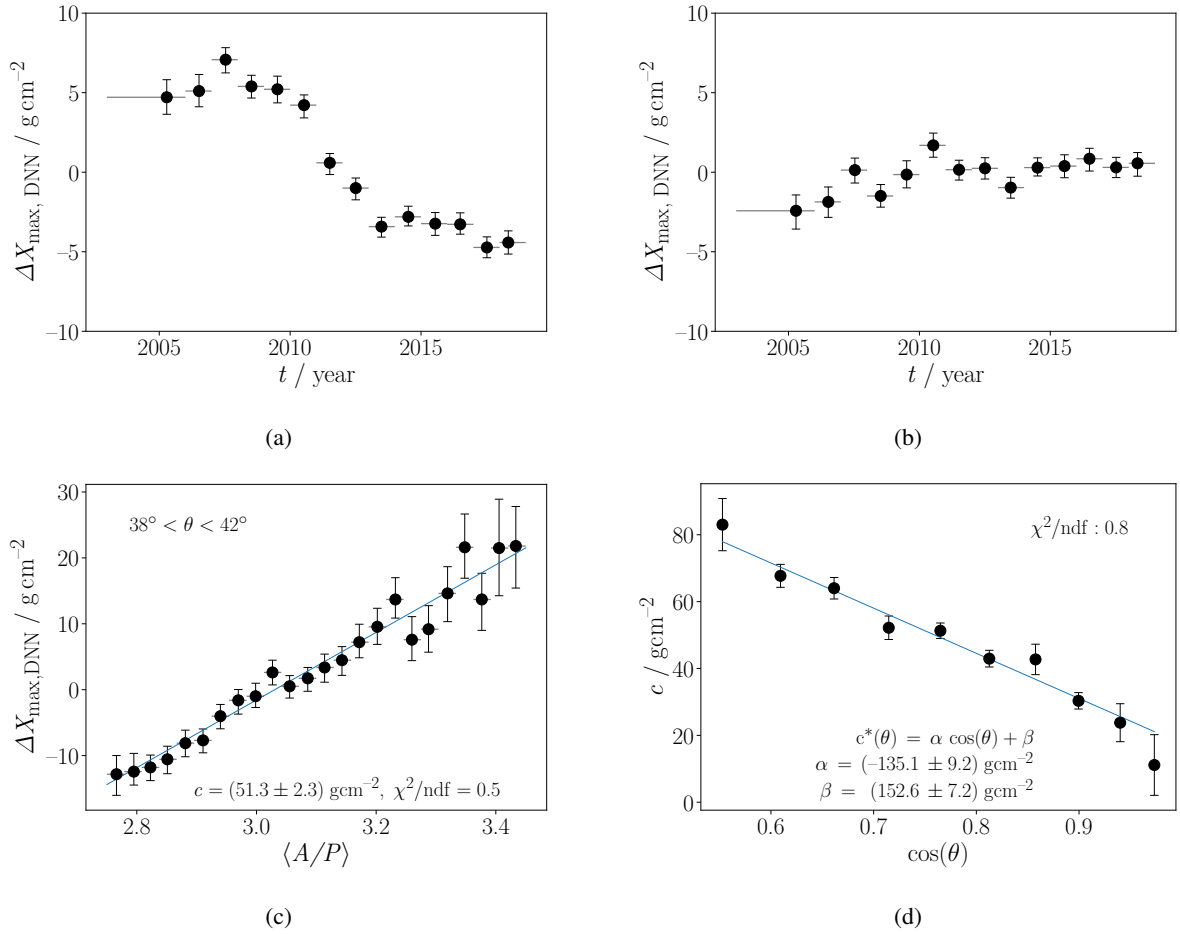


FIG. 5: Correction of the DNN X_{max} predictions for detector aging effects of the SD. Decay of the predicted X_{max} during the lifetime of the Observatory before (a) and after (b) the calibration. (c) Dependence of the X_{max} predictions on the $\langle A/P \rangle$ for an example zenith angle bin. The fitted calibration function is shown in blue. (d) Obtained slope c as a function of the zenith angle. The blue line denotes the fitted parameterization $c^*(\theta)$. The obtained parameter for plot (c) corresponds to $\cos(40°) \approx 0.77$.

D. Calibration

After training the DNN using simulations, the algorithm is applied to the measured data. Even though the simulation is continuously developed and improved, important differences exist between the measured data and the simulations. To remove such differences arising from inaccurate modeling, we perform calibrations using the SD data set by studying the X_{max} reconstruction as a function of physics and monitoring observables. We examine the reconstruction bias $\Delta X_{\mathrm{max,DNN}} = X_{\mathrm{max,DNN}} - \langle X_{\mathrm{max,DNN}} \rangle$, estimated with respect to the average X_{max} prediction. For each variable y we intend to correct with, we perform an event-by-event correction with

$$X'_{\mathrm{max}} = X_{\mathrm{max}} - f_{\Delta X_{\mathrm{max}}}(y), \tag{4}$$

where $f_{\Delta X_{\mathrm{max}}}(y)$ denotes the dependence of the X_{max} prediction on the variable y. This approach performed for each event separately ensures meaningful corrections of the predictions beyond the first moment of X_{max}.

Finally, using hybrid events, we calibrate the X_{max} predictions to the scale of the FD and remove the dependence on the hadronic interaction model used during the algorithm training, the composition, and any remaining differences in the detector simulation.

1. *Corrections using Surface Detector data*

The WCD stations of the SD are exposed to a harsh environment, with changes in temperature covering several tens of °C. During the many years of operation, the PMTs, read-out electronics, water, and reflective liner are subject to aging effects. By utilizing muons that constantly cross the detector stations, the average shape of the single-muon signal is monitored using the area over peak ratio A/P, which relates the deposited charge in the detector (integrated pulse) to its height and is a rough measure of the signal duration. Since A/P is a monitoring observable summarizing the characteristics of the

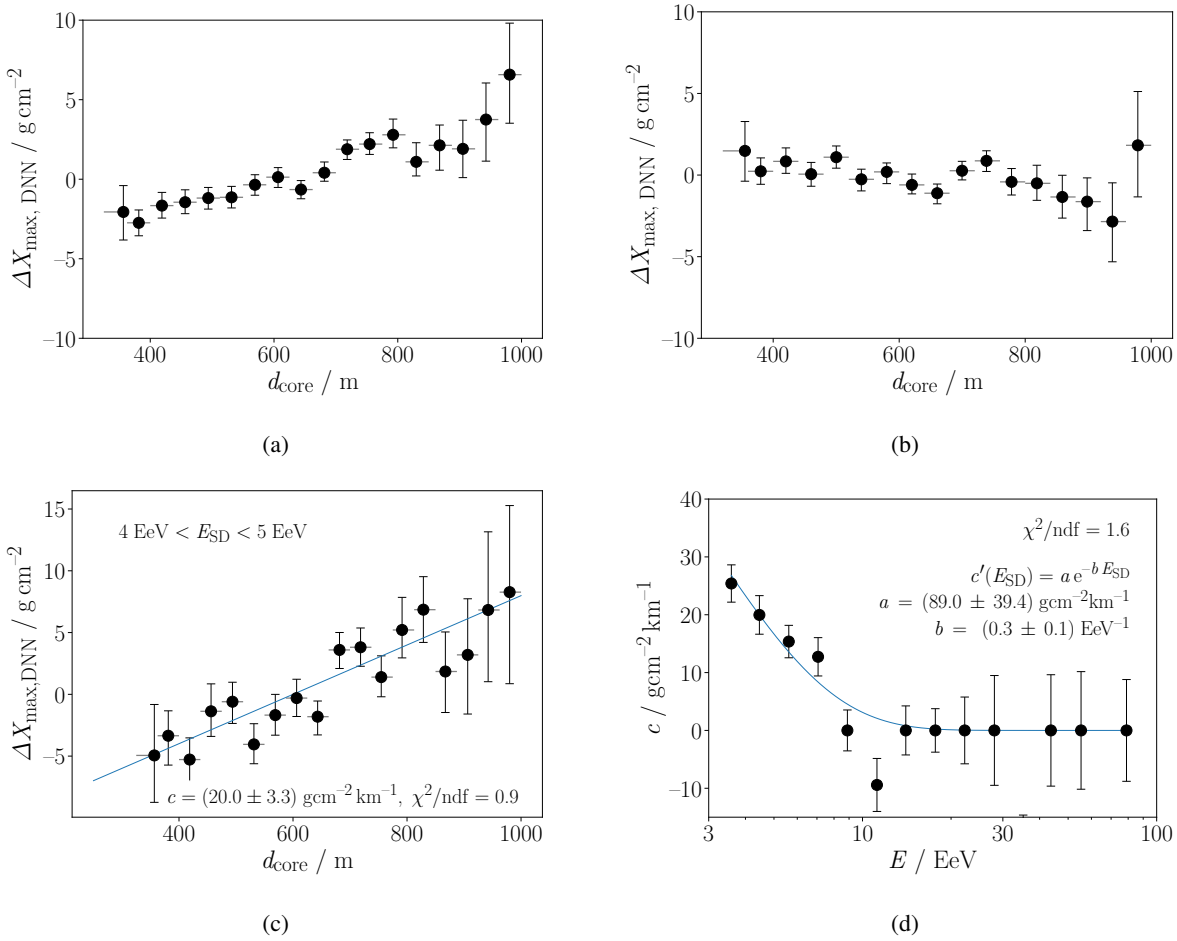


FIG. 6: Correction of the DNN X_{max} predictions as a function of the core distance d_{core} to the station with the largest signal. (a) Reconstruction bias before the calibration. (b) Bias after performing the core calibration. (c) Linear calibration of the X_{max} reconstruction bias as a function of the core distance for an example energy bin from 4 to 5 EeV. (d) Obtained slopes c of the linear fit as a function of energy E. The determined parameterization of the slope is shown as a blue line.

individual PMT responses, water quality, and liner reflectivity, it is specific to every station and changes as a function of time.

a. Aging calibration. When monitoring the distribution of A/P values of all stations, over the years, a decrease in its average $\langle A/P\rangle$ from 3.20 to 2.95 can be observed [55, 56]. We assume this to be mainly caused by the decrease in the liner reflectivity or water transparency, leading to a drift of the X_{max} predictions as a function of time (see Fig. 5a). Since our simulation library is currently limited to simulated stations with $A/P = 3.2$, the predictions of the DNN have to be calibrated as a function of the $\langle A/P\rangle$ — the average A/P of all triggered SD stations in a given event — to remove possible time dependencies of the predictions. As depicted in Fig. 5c, the dependence of the X_{max} predictions on $\langle A/P\rangle$ can be modeled linearly and does not depend on energy. Additionally, we find an increase of this dependence as a function of the zenith angle, likely caused by the fact that the average distance a particle travels through the detector rises with the zenith angle. By parameterizing (blue line in Fig. 5d) this dependence as $c^*(\theta) = \alpha\cos(\theta) + \beta$, with $\alpha = (-135.1 \pm 9.2)$ g cm^{-2}, and $\beta = (152.6 \pm 7.2)$ g cm^{-2}, we calibrate the predictions using:

$$f_{\Delta X_{\mathrm{max}}}(\theta, A/P) = \underbrace{c^*(\theta)\left(\langle A/P\rangle - \overline{\langle A/P\rangle}\right)}_{\text{zenith \& } A/P \text{ dependence}} + \underbrace{\Delta X}_{\text{scale}}, \qquad (5)$$

where $\overline{\langle A/P\rangle} = 3.03$ denotes the average of the distribution calculated over the full SD data set. Additionally, we introduce an absolute shift $\Delta X = \langle X_{\mathrm{max}}\rangle - \langle X_{\mathrm{max},A/P=3.2}\rangle = -9.5\,\mathrm{g/cm^2}$to adjust the X_{max} scale by considering the different averages in the A/P distributions in simulations $\overline{\langle A/P\rangle}_{\mathrm{MC}} = 3.2$ and data $\overline{\langle A/P\rangle}_{\mathrm{data}} = 3.03$.

The reconstruction bias of X_{max} shows a dependence on the distance between the station with the largest signal and the shower core reconstructed using fitting of a lateral distribution function [38]. At low and high energies, the biases are relatively large since at small distances, the stations with the largest signal are often saturated, and at low energies, the

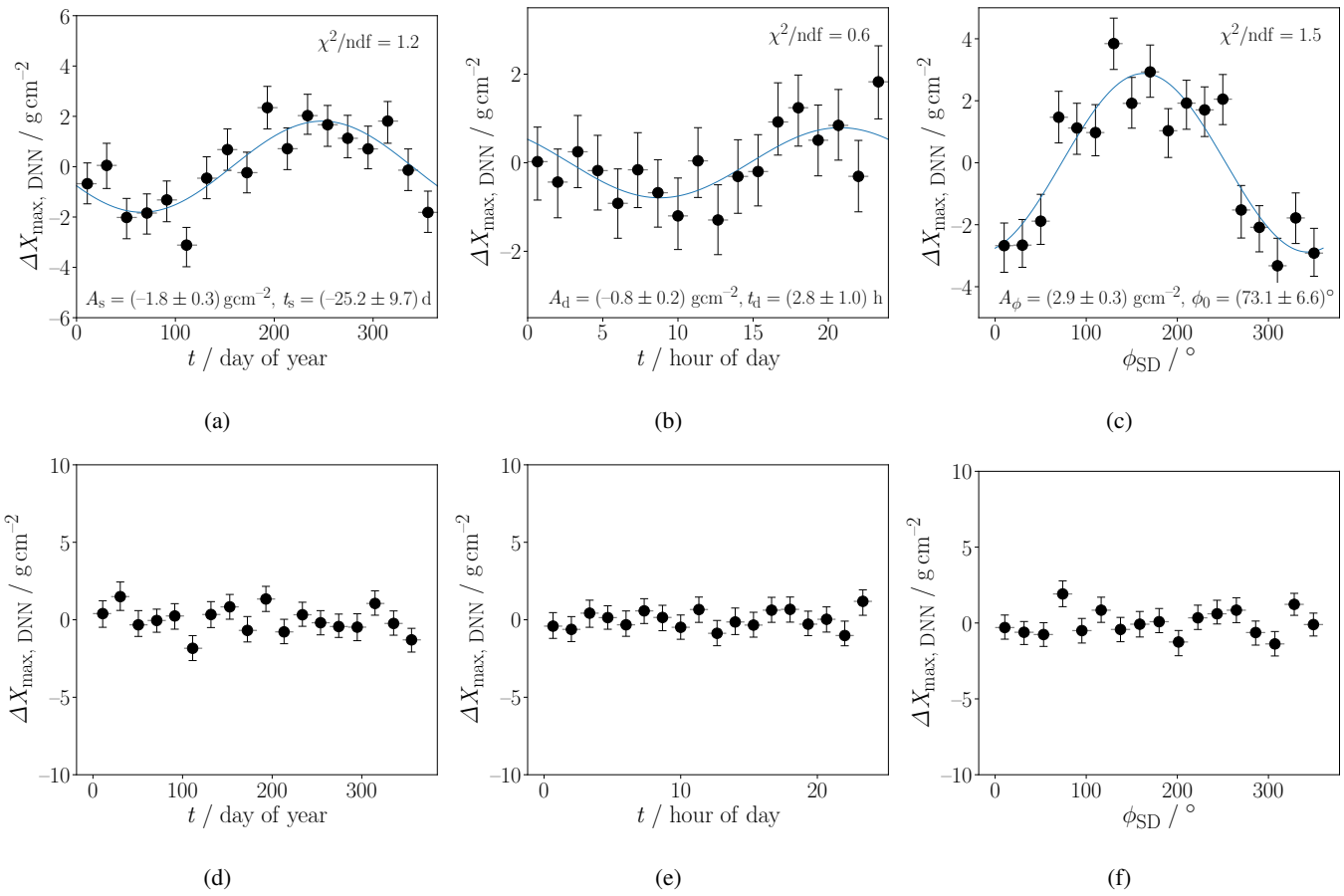


FIG. 7: Data-based correction of the DNN X_{max} prediction. Top: dependence of the X_{max} reconstruction as a function of (a) season (time of year), (b) time of day (UTC), and (c) azimuth angle of the shower. The fitted calibration functions are shown in blue. Bottom: dependence of the reconstruction as a function of (d) season, (e) time of day (UTC), and (f) azimuth angle after the calibration.

station multiplicity is small, making a detailed reconstruction more challenging. Furthermore, the number of triggered stations is on average, lower for cores located close to one station since in this case the distances to the next stations of the grid all become close to 1500 m. A similar effect applies to events with reconstructed shower cores far away from the station with the highest signal. Simulation studies show that at a distance of roughly 600 m, the bias is smallest. In general, events with core distances larger than 350 m and smaller than 1000 m feature a small bias and a dependence of the reconstruction bias on the core distance, as visible in Fig. 6a. Events outside this regime that exhibit larger biases were rejected during the data selection.

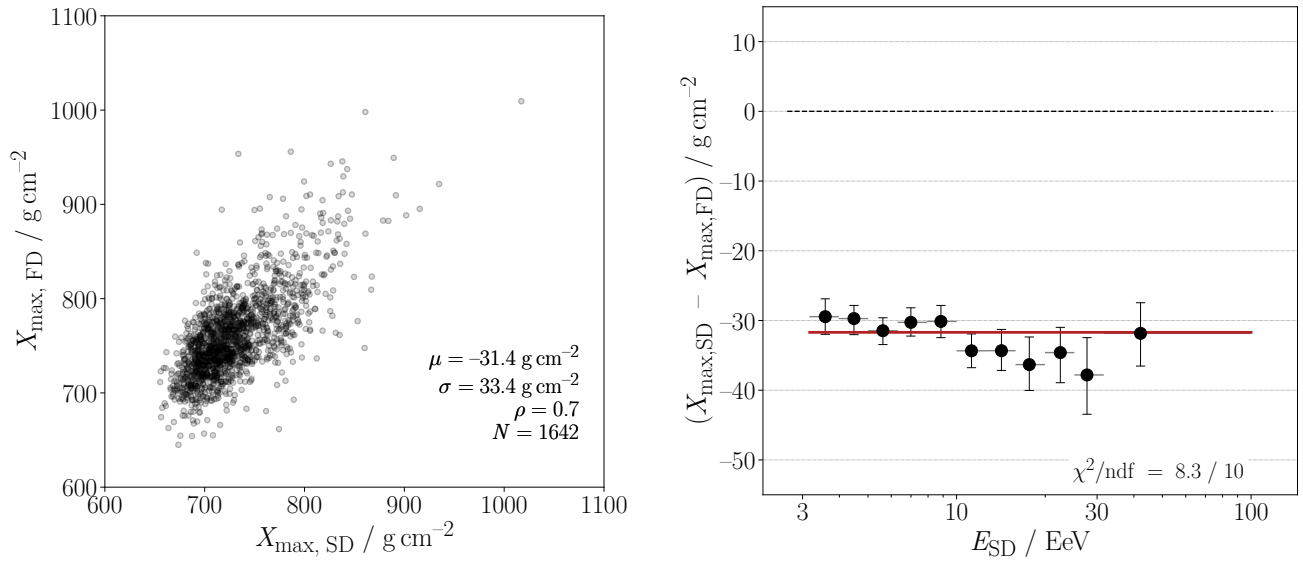
The described effect mainly concerns events with a low multiplicity of triggered stations, i.e., showers produced by primaries having energies below ∼10 EeV. Therefore, we apply an energy-dependent calibration to remove the X_{max} bias at small energies. In bins of energy, we perform a linear fit $f_{\Delta X_{\mathrm{max}}}(d) = c\,(d - 0.6\ \mathrm{km})$ of the reconstruction bias, as shown in Fig. 6c for the example energy bin from 4 to 5 EeV. Above 10 EeV, a constant fit usually shows a better χ^2 and is preferred over a linear fit. The dependence of the slope with energy is shown in Fig. 6d, and parameterized by fitting the function $c'(E_{\mathrm{SD}}) = a\,e^{-b\,E_{\mathrm{SD}}}$. The obtained values are $a = (89.0 \pm 39.4)\,\mathrm{g\,cm^{-2}/km}$ and $b = (0.3 \pm 0.1)\,\mathrm{EeV^{-1}}$. The final calibration of the DNN is performed using

$$f_{\Delta X_{\mathrm{max}}} = c'(E_{\mathrm{SD}})\,(d - 0.6\ \mathrm{km}) \tag{6}$$

and using Eq. (4) on an event-by-event basis.

b. Temporal variations. The change in pressure and temperature due to diurnal and seasonal variations causes small influences on the detector response and the conversion of distance to X_{max}, hence affecting its reconstruction. To remove seasonal and diurnal variations, we investigated the reconstruction as a function of time on a yearly and daily basis[7], as shown in Fig. 7a and Fig. 7b. We find small variations of the size of 2 $\mathrm{g\,cm^{-2}}$ or 1 $\mathrm{g\,cm^{-2}}$, respectively. We first calibrate the X_{max} predictions to remove the seasonal variation by fitting a sine function to the data and correcting it by using the

[7] The time is given in UTC.



(a) (b)

FIG. 8: Application of the DNN X_{max} estimation to hybrid data. (a) Correlation between FD observations and DNN predictions using SD data. (b) The energy-dependent bias of the SD-based reconstruction of the DNN when compared to the reconstruction of the fluorescence detector.

time of year on an event-by-event basis. These predictions are also used to remove the diurnal variations by again fitting a sine wave and correcting events using Eq. (4). The dependencies after correction are depicted in Fig. 7d and Fig. 7e.

c. Angular dependence. In contrast to our simulation study, we find a dependence of the X_{max} reconstruction on the azimuth angle (see Fig. 7c). The dependence is small, and its fluctuations are around 3 g cm^{-2} and possibly caused by a slight slope of the SD array tilted away from the Andes mountains. We remove the dependence by fitting a cosine and calibrating the predictions using the azimuth angle on an event-by-event level. The reconstruction after calibration is shown in Fig. 7f. We also tested the reconstruction for a possible dependence on the zenith angle. Therefore, we studied X_{max} as a function of the zenith angle for different energy intervals to account for the fiducial cut but could not find any indications for a dependence.

2. *Calibration using hybrid events*

The hybrid design of the Pierre Auger Observatory enables a cross-calibration of the SD X_{max} measurements with FD X_{max} of hybrid events. The dependence of the X_{max} scale of the DNN on the hadronic interaction model can thus be eliminated by calibrating the DNN predictions with the FD X_{max} scale that can be accurately determined [20].

The event-by-event correlation between the FD and the SD is shown in Fig. 8a. We find a Pearson correlation of $\rho = 0.70 \pm 0.03$, which is in good agreement with the expectations from idealized simulations ($\rho_{\mathrm{MC}} = 0.73$). The absolute bias, however, amounts to -31.4 ± 0.8 g cm^{-2}. This bias is larger than expected from simulation studies (up to -15 g cm^{-2} assuming the Auger mix) with interaction models different from those used in the algorithm training (cf. Fig. 4 and Ref. [32, 52]). The observation of negative bias, i.e., a heavier composition in data (smaller X_{max} values), is in line with findings in previous analyses, where the average signal footprint measured using surface detector arrays seems to favor a composition heavier than expected from simulations [11, 14, 21]. In particular, recent works indicate that the current generation of hadronic interaction models may not model the muonic component in full detail [11, 12, 57]. Additionally, adjustment of the longitudinal shower profile might be needed [58]. In contrast, the relative fluctuations in the muon component seem to be reasonably modeled [14]. Using the exotic hadronic interaction model Sibyll* [59] that features ad-hoc modifications of the shower content, a significant increase of the muon number can be accomplished. A test using Sibyll* that predicts an increase of the muon number by 40% for protons with respect to Sibyll2.3d shows that a bias of -40 g cm^{-2} could be reproduced, indicating that the observed scale of the bias could be explained by a mis-modeling of the muonic component of current interaction models. However, it is unclear if such ad-hoc adjustments or data-based refinements [60] offer a realistic solution. In addition, note that a non-perfect detector simulation could cause deviations and that the systematic uncertainty on the FD X_{max} scale amounts to roughly 10 g cm^{-2} [61].

We find no significant energy dependence when studying the bias as a function of energy (see Fig. 8b). This is consistent with our simulation study since no strong energy de-

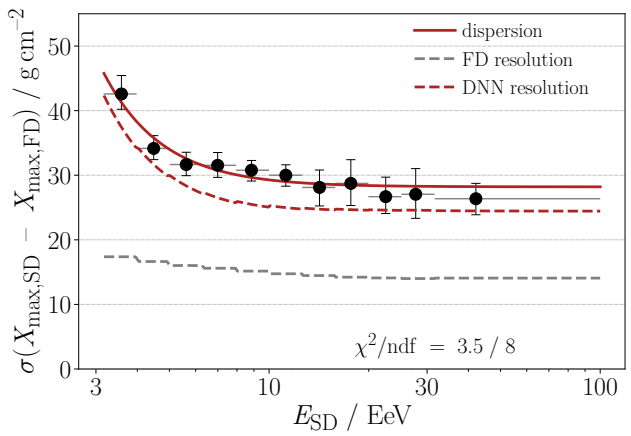


FIG. 9: Energy-dependent resolution of the DNN (dashed red line) obtained by subtracting the FD resolution (dashed grey line) from the standard deviation of the FD and the SD X_{max} reconstruction (red line) in the hybrid data set.

pendence was found when studying the reconstruction bias of the most likely composition scenario, i.e., the Auger mix, for Sibyll2.3c and QGSJetII-04 (see Fig. 4). Since the X_{max} scale can be precisely defined using the FD, we re-calibrate the predictions of the DNN for the SD events with a constant offset of $(-31.7 \pm 0.7)\,\mathrm{g\,cm^{-2}}$ obtained by the fit depicted as a red line in Fig. 8b. Due to the calibration, we adopt the uncertainty of the FD X_{max} scale to the systematic uncertainties on the $\langle X_{\mathrm{max}} \rangle$ measurement. This enables us to remove the composition and interaction-model-dependent contributions to the systematic uncertainty of the $\langle X_{\mathrm{max}} \rangle$ measurement with the DNN and the SD. Since with the low statistics in the hybrid data set at high energies, we cannot exclude a small energy dependence of the DNN reconstruction, deviations from a constant calibration offset are examined using energy-dependent calibrations. We consider this in an energy-dependent systematic uncertainty on $\langle X_{\mathrm{max}} \rangle$ measurements. Note that we are using in this work the same data production as Ref. [62], which covers the same data-taking period. Ongoing work on the FD reconstruction has led to refinements in the X_{max} scale [63, 64] that have not been considered, but remain below 5 $\mathrm{g\,cm^{-2}}$ in $\langle X_{\mathrm{max}} \rangle$ [65].

In Fig. 9, we show the event-by-event resolution of reconstructing X_{max} using the DNN (dashed red line) after subtracting the FD resolution [61] (dashed grey line) in quadrature from the standard deviation (continuous red line), found using the hybrid data. The resolution improves from $40\,\mathrm{g\,cm^{-2}}$ at low energies to $25\,\mathrm{g\,cm^{-2}}$, which is in good agreement with simulations studies [32].

a. Crosscheck of SD-based calibrations. We additionally checked the event-by-event correlation between the FD and the SD reconstruction before and after each calibration described in Section III D 1 to ensure its validity. We found an increase in correlation with the FD X_{max} measurement after performing each SD-based calibration. Furthermore, the Pearson correlation coefficient increased from 0.62 to 0.7 by applying all the SD calibrations and the analysis-specific cuts, thus confirming the validity of the calibrations and the selection.

E. Systematic uncertainties

The systematic uncertainties of the $\langle X_{\mathrm{max}} \rangle$ measurement using the SD are shown in Fig. 10a. The X_{max}-scale uncertainty of the FD, as inherited by the DNN during the calibration using hybrid measurements to remove the dependence on hadronic interaction models, is depicted as a dash-dotted line. It contains uncertainties regarding the reconstruction, the atmosphere, and the calibration of the FD. Whereas the latter is independent of energy, the energy dependence is caused by the former two contributions. At low energies, reconstruction uncertainties of the FD dominate. These are surpassed by atmospheric uncertainties with increasing energy since more distant showers can be detected with correspondingly larger corrections for the light transmission between the shower and the detector. For more details on the FD uncertainty, we refer to Ref. [20]. Uncertainties from the SD are denoted as dashed region and comprise the remaining uncertainties of the detector aging ($< 0.5\,\mathrm{g\,cm^{-2}}$), diurnal variations ($1\,\mathrm{g\,cm^{-2}}$) — since the FD calibration is performed at night — and the uncertainty on the calibration using hybrid events. The calibration uncertainty has two parts, the uncertainty of the definition of the absolute X_{max} scale, which is estimated to be $\pm 2\,\mathrm{g\,cm^{-2}}$, and the energy dependence of the calibration. To estimate the energy-dependent uncertainty of the calibration, we compare the assumed constant calibration to a calibration function linear in $\log_{10}(E_{\mathrm{SD}}/\mathrm{eV})$ and use the observed differences as the upper and lower uncertainty on our calibration (compare Fig. 18a in the appendix).

The resulting total uncertainties of the SD-based $\langle X_{\mathrm{max}} \rangle$ measurement are of the order of $\pm 10\,\mathrm{g\,cm^{-2}}$ and shown as a continuous red line. In general, the obtained uncertainty is very similar to the FD uncertainty. Only at high energies, due to the limited statistics of hybrid events, is the uncertainty on the calibration rising slightly. Nevertheless, at high energies, substantial deviations from the applied calibration are not to be expected since the simulation study (see Section III C) indicated only a very small reconstruction bias above 30 EeV.

In contrast with the measurement of the first moment, no strong dependence of $\sigma(X_{\mathrm{max}})$ on hadronic interaction models was found. Therefore, no calibration is performed using the FD. Hence, the measurement is independent of the FD and the systematic uncertainties contain only SD contributions. Fig. 10b displays the different contributions as a function of energy, where effects are only shown that contribute more than $1\,\mathrm{g\,cm^{-2}}$. The largest source of uncertainty at low energies is the composition bias that was found to be independent of the interaction models. It was derived from the simulation studies reported in Section III C by assuming for each energy bin the largest reconstruction bias found in studies of a pure proton, a pure iron, and the Auger mix composition. The assumed parameterizations are a conservative estimate as, in nature, a pure proton, a pure iron, and an Auger mix composition can-

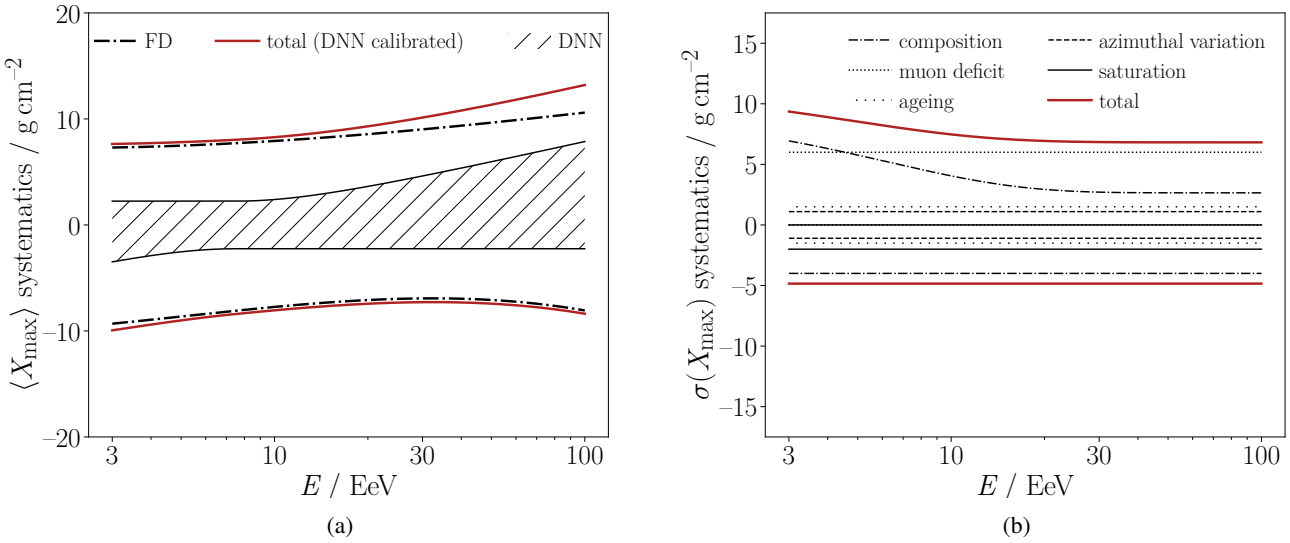


FIG. 10: Energy-dependent systematic uncertainties for the measurement of $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$. (a) Total uncertainty of the calibrated DNN (continuous red lines) resulting from adding the uncertainty on the FD $X_{\rm max}$ scale (dash-dotted line) and the DNN uncertainties (hatched region) after the hybrid calibration in quadrature. (b) Systematic uncertainties for the DNN reconstruction of $\sigma(X_{\rm max})$ as a function of energy. The total systematic uncertainty is denoted as the continuous red lines. Only contributions > 1 g cm^{-2} are presented.

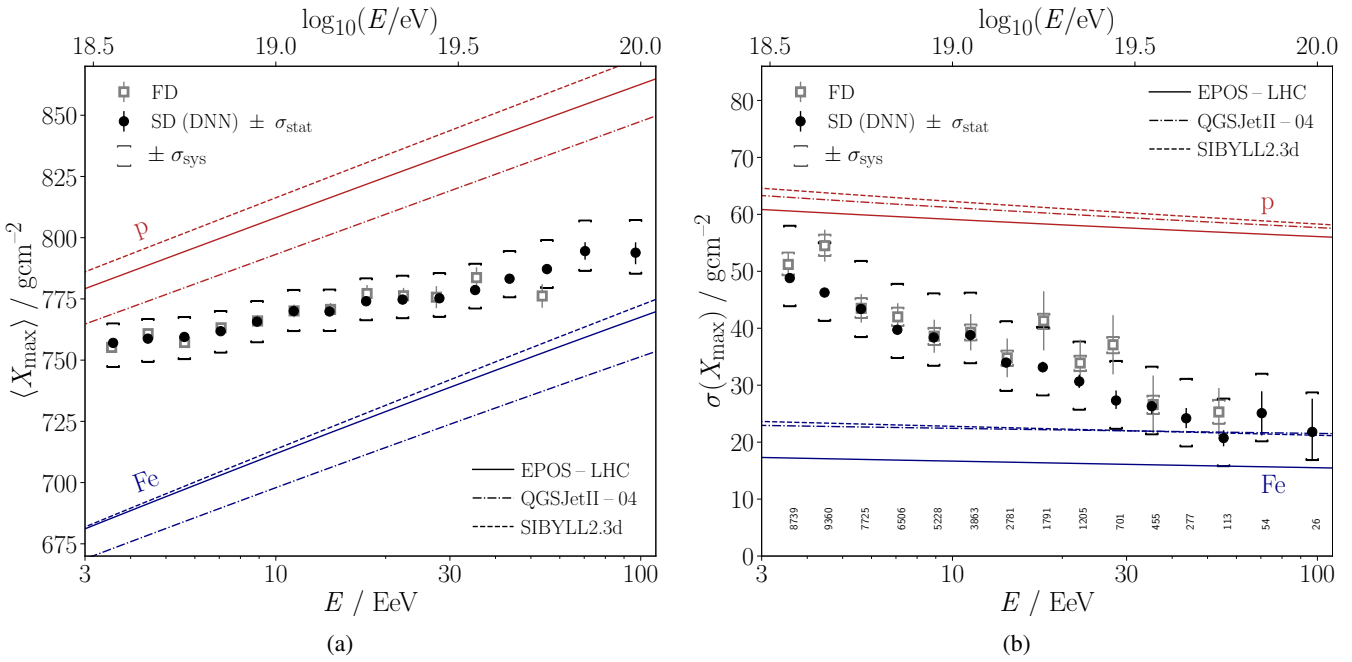
not exist at the same time. Nonetheless, this provides an estimate for all potential scenarios, even though a substantial proton or iron fraction at high or low energies, respectively, are extremely unlikely. In this work, we use the bias parameterization[8] obtained for the EPOS-LHC interaction model. To examine potential shortcomings of the modeling of the muon component as intensively discussed in the literature [11–14], we studied Sibyll* with a significantly increased muon number and found a slight underestimation of the fluctuations of the order of 5 g cm^{-2}, constant for all compositions and independent of energy. In particular, at high energies, we account for a potential muon deficit, contributing more than all other factors to the uncertainty of $\sigma(X_{\rm max})$. Other contributions come from saturation effects (-2 g cm^{-2}) and detector aging (± 1.5 g cm^{-2}).

IV. INVESTIGATION OF THE UHECR MASS COMPOSITION USING THE SURFACE DETECTOR

In the following section, we present inferences on the UHECR mass composition based on the first two moments — $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ — of the reconstructed $X_{\rm max}$ distributions. The composition analysis is based on 48,824 events recorded using the SD and its evolution studied in bins of $\Delta\log_{10}(E/{\rm EeV}) = 0.1$ with an integral bin above $10^{19.9}$ eV. For comparison, we use FD data of Ref. [62] covering the same data-taking period. Since the full FD data set features — in comparison to the hybrid data set used for calibrating the DNN, which requires a full SD and FD reconstruction as well as a specific geometry due to the different efficiencies of the two detectors as a function of zenith angle — an increase in statistics of a factor of almost five, we allow for a constant shift of the SD $X_{\rm max}$ scale when comparing the DNN and FD measurements. The adjustment on top of the hybrid events study amounts to -1.7 g cm^{-2} and is within our statistical uncertainty (2 g cm^{-2}) of the calibration.

In Fig. 11, we present the energy evolution of $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ as reconstructed using the DNN based on SD data (black circles) and as obtained using the standard FD reconstruction (open grey squares). The statistical uncertainties are estimated using bootstrapping and are shown as vertical lines, whereas the systematic uncertainties, discussed in Section III E, are depicted as brackets. We do not show the systematic uncertainty for the measurement of $\langle X_{\rm max}\rangle$ using the FD as it is part of the SD uncertainty due to the cross-calibration we conducted. Since for $\sigma(X_{\rm max})$ no calibration is performed, we show the systematics for both measurements. The measured data are compared to predictions [66] for protons (red) and iron (blue) of the three hadronic interaction models EPOS-LHC, Sibyll2.3d, and QGSJetII-04, denoted by different line styles. In the right plot, we further show the number of events in each bin of the SD data, which is the same in both plots. The evolution of $\langle X_{\rm max}\rangle$ in Fig. 11a as a function of energy shows an excellent agreement between the SD and the FD measurements with only very small deviations that can be explained purely by statistics. This extends the $X_{\rm max}$ measurements to 100 EeV and confirms the transition from a lighter to a heavier composition with increasing

[8] No significant differences can be observed between different interaction models. The obtained parameterization can be found in Fig. 17.



(a) (b)

FIG. 11: Energy evolution of (a) the average depth of shower maximum $\langle X_{\rm max}\rangle$ and (b) the fluctuations of the shower maximum $\sigma(X_{\rm max})$ as determined using the FD reconstruction [62] (grey open squares) and the DNN $X_{\rm max}$ predictions (black circles). Red (blue) lines indicate expectations for a pure proton (iron) composition for various hadronic models.

energy, also reported in previous SD-based studies using the risetime of signals in the WCDs [21].

The *elongation rate* D_{10} is defined by the change of $\langle X_{\rm max}\rangle$ per decade of energy

$$D_{10} = \frac{\mathrm{d}\langle X_{\rm max}\rangle}{\mathrm{d}\log_{10}(E)} = \hat{D}_{10}\left(1 - \frac{\mathrm{d}\langle \ln A\rangle}{\mathrm{d}\ln(E)}\right),$$

where A denotes the primary particle mass. When measuring D_{10}, a deviation from the elongation rate $\hat{D}_{10}$, which is in a very good approximation, universal across all hadronic interaction models and primary nuclei, can be traced back to a change in the primary mass composition. The elongation rate obtained using the SD over the whole energy range amounts to $D_{10} = (24.1 \pm 1.2)\ \mathrm{g\,cm^{-2}\,decade^{-1}}$ in good agreement with the FD result $\left((26 \pm 2)\ \mathrm{g\,cm^{-2}}\right)$ [62]. However, the reduced $\chi^2/\mathrm{ndf} = 46.7/13$ obtained for the SD data indicates that another substructure exists, as will be comprehensively discussed in the next Section IV A.

The evolution in $\sigma(X_{\rm max})$, sensitive to the composition mixing, is shown in Fig. 11b. We find a decrease of $\sigma(X_{\rm max})$ as a function of energy and a very good agreement between the measurements of the SD and the FD. This confirms for the first time the transition from a lighter and mixed composition into a heavier and purer composition with large statistics. At the highest, previously inaccessible energies (> 50 EeV), the fluctuations appear to stabilize and remain small. However, more statistics are needed to examine the composition evolution at these energies in more detail. Given the limited differences in the interaction model predictions of $\sigma(X_{\rm max})$, the small fluctuations in $X_{\rm max}$ beyond 30 EeV clearly exclude a scenario with a substantial fraction of protons and light nuclei in the UHECR composition. Additionally, at around 10 EeV, the fluctuations appear to stay constant.

A. Discussion of breaks in the elongation rate

The observation of an elongation rate similar to the FD but obtained using the comprehensive SD data set that features $\chi^2/\mathrm{ndf} \approx 3.6$, indicates that a simple linear model is not describing the data well (see Fig. 12a), suggesting the existence of a substructure to be analyzed. The measurement of $\sigma(X_{\rm max})$ also shows a non-continuous decrease of fluctuations with energy.

In Fig. 12, we study the evolution in the UHECR mass composition using different models. We analyze the evolution using broken-line fits with a different number of breaks. The simplest model beyond a constant elongation rate is a broken-line fit with one fitted break point shown in Fig. 12b that also cannot describe our data reasonably ($\chi^2/\mathrm{ndf} \approx 3.4$). Considering Wilks' theorem, we compared the χ^2 values of two nested models, in which the model of a constant elongation rate is used as the null hypothesis and test if it can be rejected with more complex models. A model with two breaks in the elongation rate can reject the constant elongation rate hypothesis at a significance of 3.4σ (see Fig. 12c). In Fig. 12d, we show a model with three breaks in the elongation rate, where the slopes and the break position were determined by a fit. This model can reject the hypothesis of a constant elongation rate at a level of 4.6σ and a single-break model at a level of

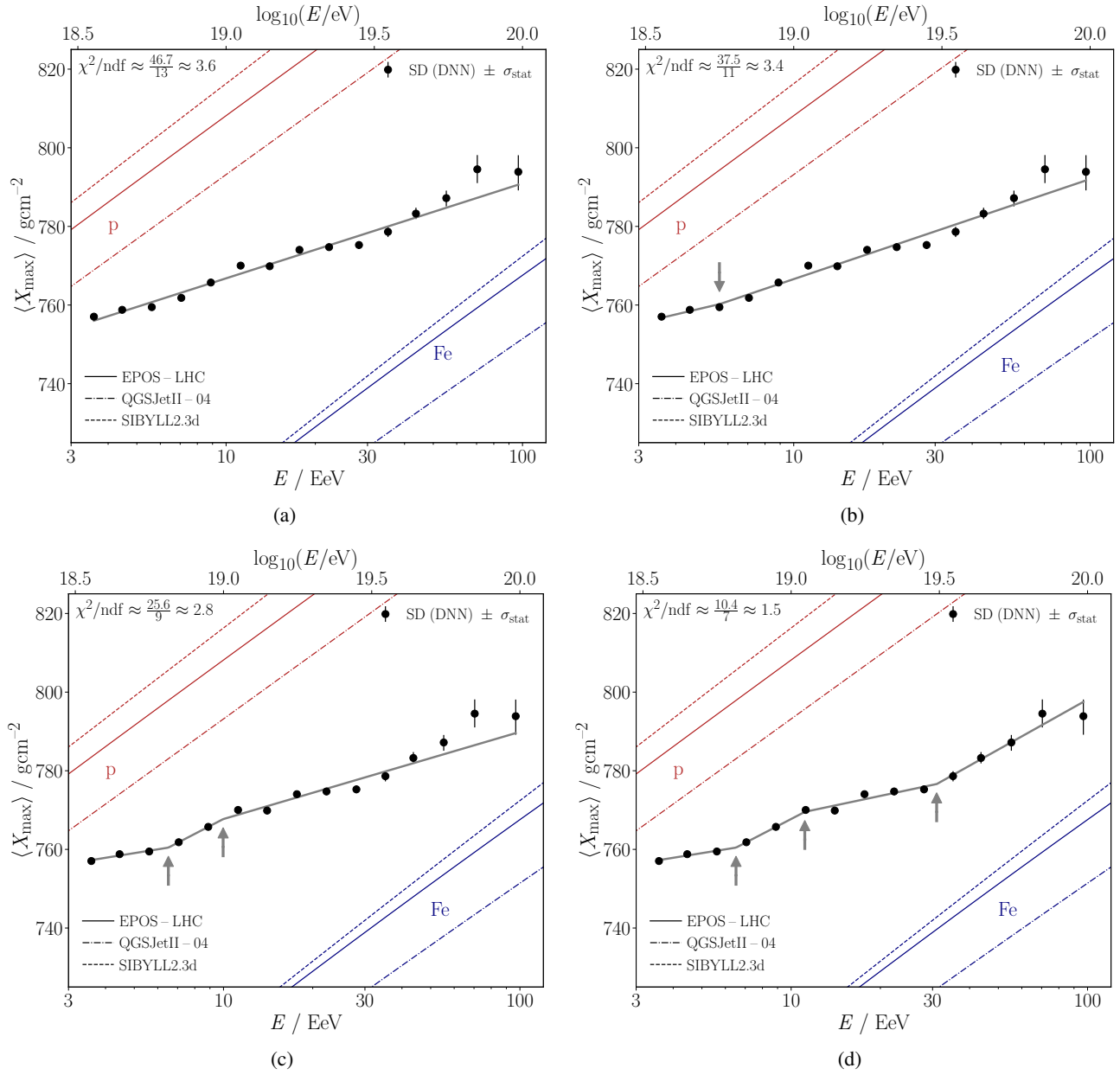


FIG. 12: Investigated models (grey lines) describing the evolution of $\langle X_{\rm max}\rangle$ as a function of energy E. The studied models are piecewise-linear in $\log_{10}(E/{\rm eV})$. (a) Fit of a constant elongation rate, as suggested by the FD data analyses above 3 EeV. More complex models describing a scenario beyond a constant evolution: piecewise-linear models with (b) one break, (c) two breaks, and (d) three breaks. The locations of the breaks are indicated by grey arrows.

4.4σ, which, on a statistical basis, indicates a substructure in the evolution of the UHECR composition. The significance of rejecting the hypothesis of a two-break model using the three-break model amounts to 3.3σ.

The investigated models and their parameters are summarized in Table III, including statistical and systematic uncertainties, and compared to the positions of the energy spectrum features identified at ultra-high energies. Systematic uncertainties were estimated by shifting the measurement by the upper and lower energy-dependent uncertainties discussed in Section III E and re-fitting the data and furthermore incorporating the uncertainty on the hybrid calibration by an energy-dependent calibration (see Fig. 18) affecting the size of the elongation rate. The breaks in the evolution of $\langle X_{\rm max}\rangle$ in all models are observed to be at similar energies as the features of the UHECR energy spectrum [53], i.e., the ankle at $(4.9 \pm 0.1\,({\rm stat}) \pm 0.8\,({\rm sys}))$ EeV, instep at $(14 \pm 1\,({\rm stat}) \pm 2\,({\rm sys}))$ EeV and suppression at $(47 \pm 3\,({\rm stat}) \pm 6\,({\rm sys}))$ EeV. Note that, even for a joint astrophysical interpretation, features in the energy spectrum and

TABLE III: Best-fit parameters with statistical and systematic uncertainties for the studied elongation models that feature up to three changes at energies (E_1, E_2, E_3) in the elongation rate (D_0, D_1, D_2, D_3) and an offset b ($\langle X_{\rm max}\rangle$ at 1 EeV), without including a systematic uncertainty of 14% on the energy scale. Also given are the positions of the energy spectrum features measured at the Pierre Auger Observatory [53] in the same data-taking period.

Parameter	Const. elong.	1-break model	2-break model	3-break model	Energy spectrum
	Val $\pm\,\sigma_{\rm stat}\pm\sigma_{\rm sys}$	Val $\pm\,\sigma_{\rm stat}\pm\sigma_{\rm sys}$	Val $\pm\,\sigma_{\rm stat}\pm\sigma_{\rm sys}$	Val $\pm\,\sigma_{\rm stat}\pm\sigma_{\rm sys}$	Val $\pm\,\sigma_{\rm stat}\pm\sigma_{\rm sys}$
b / g cm^{-2}	$743\pm5\pm13$	$743\pm5\pm13$	$750.5\pm4\pm13$	$750.5\pm3\pm13$	
D_0 / g cm^{-2} decade^{-1}	$24\pm1\pm4$	$23\pm2\pm12$	$12\pm6\pm5$	$12\pm5\pm6$	
E_1 / EeV		$35\pm12\pm16$	$6.5\pm0.9\pm1$	$6.5\pm0.6\pm1$	$4.09\pm0.1\pm0.8$
D_1 / g cm^{-2} decade^{-1}		$39\pm14\pm12$	$39\pm12\pm10$	$39\pm5\pm14$	
E_2 / EeV			$10\pm2\pm3$	$11\pm2\pm1$	$14\pm1\pm2$
D_2 / g cm^{-2} decade^{-1}			$22\pm3\pm8$	$16\pm3\pm6$	
E_3 / EeV				$31\pm5\pm3$	$47\pm3\pm6$
D_3 / g cm^{-2} decade^{-1}				$42\pm9\pm12$	

the evolution of $\langle X_{\rm max}\rangle$ do not have to coincide in energy, as, for example, the break in the elongation rate observed around 2 EeV [61] is physically interpreted in association with the ankle [9, 10], located at 5 EeV.

We analyzed the $\sigma(X_{\rm max})$ measurement for characteristics similar to the ones found in the evolution of $\langle X_{\rm max}\rangle$. Between $E_0 = 6.5$ EeV and $E_1 = 11$ EeV, where the observed elongation rate is within uncertainties compatible with a constant composition, also $\sigma(X_{\rm max})$ appears to stay constant. Furthermore, beyond $E \approx 30$ EeV (at E_2), the decrease in the fluctuations appears to stop, which would be consistent with the elongation rate that was found to be close to that of a constant composition at the highest energies. Due to the increasing statistical uncertainties, more data are needed for a definite statement. A quantitative test of a structure in $\sigma(X_{\rm max})$ with breaks at positions that agree with the ones found in the elongation rate study, however, is not significant. The null hypothesis of a linear decrease of $\sigma(X_{\rm max})$ can be rejected at only a 2.2σ significance level, using a model with three break positions fixed to the ones found in the elongation rate study, which nonetheless seems to be compatible with the data ($\chi/{\rm ndf} = 10.3/10$). Reduced uncertainties and more data are required to analyze the structure in the evolution of $\sigma(X_{\rm max})$ in detail.

Note that a one-to-one agreement of breaks and structures generally, in the measurements of $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$, is not to be expected since a change in the mean logarithmic mass does not need to coincide with a similar change in the measurement of $\sigma(X_{\rm max})$, i.e., the composition mixing [16]. It would rather reveal a characteristic structure of the composition. Interestingly, breaks at similar positions in the energy evolution of $X_{\rm max}$ and $\sigma(X_{\rm max})$ can be obtained when fitting a simplified astrophysical model using the FD $X_{\rm max}$ data and the Auger spectrum as measured by the SD (see Fig. 3 and Fig. 6 in Ref. [10]). A dedicated analysis focusing on the astrophysical interpretation and investigating the non-trivial interplay between the spectrum, $\langle X_{\rm max}\rangle$, and $\sigma(X_{\rm max})$, is ongoing and will be discussed in a future publication.

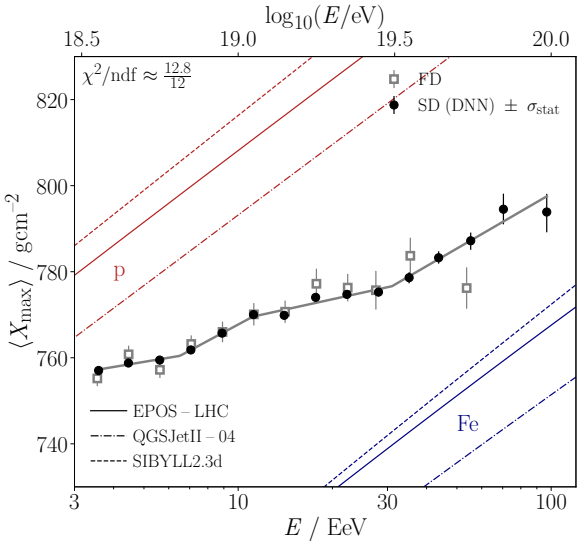


FIG. 13: The found elongation rate model with three breaks obtained using SD data (continuous grey line) compared to the evolution of $\langle X_{\rm max}\rangle$ as measured using the FD (open grey boxes) and the SD (black markers). The χ^2 shown refers to the FD data.

1. *Crosscheck and comparison with the FD*

The obtained model exhibiting a characteristic structure beyond a constant change in the mean logarithmic mass has to be consistent with the FD measurements. The comparison of the elongation model with the FD and SD is presented in Fig. 13. The model describes the FD data adequately with $\chi^2/{\rm ndf} = 1.1$, demonstrating the consistency of the model with FD data.

Additionally, we tested the obtained model by investigating the fit with a different binning in energy and with different selections of core distance, zenith angle, and azimuth angle. We further studied the influence of seasonal, diurnal (day-night), and aging effects. No significant effects could be found for any of the tests. Next, we investigated the existence of the fea-

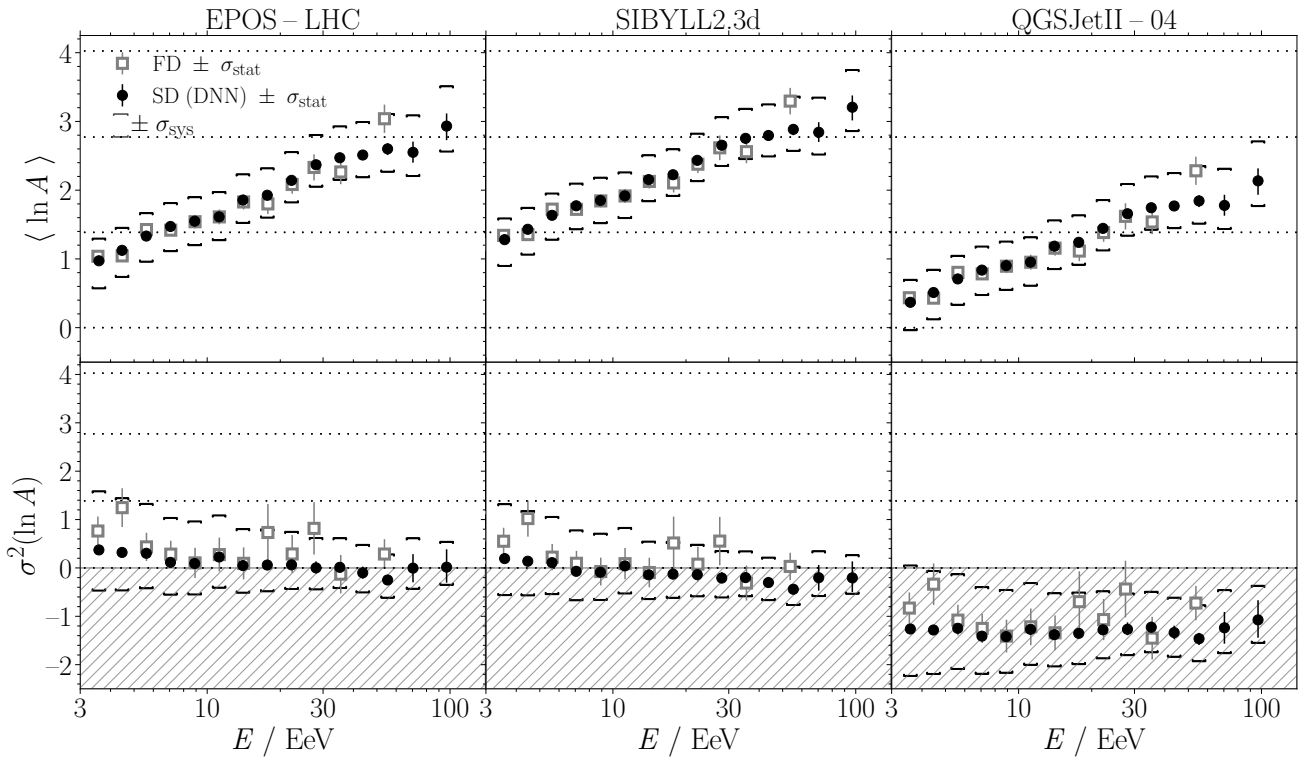


FIG. 14: Evolution of the first two moments of the ln$A$ distribution as a function of energy E for the hadronic interaction models EPOS-LHC, Sibyll2.3d, and QGSJetII-04, determined using the FD [62] (grey open squares) and the DNN (black circles).

tures before applying calibrations and without applying quality cuts and found a statistical significance level larger than 5σ using the model featuring three breaks. Since our selection removes approximately half of the events, we estimate that the expected median significance for identifying the breaks with a data set of similar statistics, and same quality as the full data set without any cuts, is 3.4σ. The finding of a 4.6σ significance using our data selection confirms the expectation that the significance should increase with improving data quality.

We additionally tested energy-dependent calibrations of the DNN with the hybrid data by employing various broken-line model fits. The tested calibration functions are summarized in Fig. 18. None of the tested calibration functions reduced the significance of rejecting a constant elongation rate but showed, due to the energy dependence of the calibration, a stronger rejection of a constant elongation. In addition, for each studied hybrid calibration, we examined the energy-dependent FD $X_{\rm max}$ scale uncertainty. The significance of rejecting the constant elongation rate with the three-break model remains of the same order, with a minimum of 4.4σ observed for the cases where the total lower and upper uncertainty is applied to the measurement. The two-break model can be rejected at a significance level of around 3σ in most cases. Only for more complex functions (compare Fig. 18e and Fig. 18f), which cannot be strongly constrained due to the low statistics in the hybrid sample, the significance level drops to around 2σ. The rejection of a single-break model consistently remains above the 3σ level and is at the 4σ level in most cases.

Rejecting a constant elongation rate using the two-break model is very stable and above a significance of 3.4σ for all scenarios. Applying instead of the FD calibration a correction of the SD $X_{\rm max}$ reconstruction based on the expected composition bias of the Auger mix using simulations (compare Fig. 4), a constant elongation rate can be rejected by more than 5σ assuming EPOS-LHC, as the first break is strongly pronounced. Therefore, we find a robust indication at a 4.4σ level for structures beyond a constant elongation rate. However, more statistics and/or a reduction in energy-dependent uncertainties are needed to confidently reject the 2-break model, i.e., to investigate the existence and nature of the third break.

B. Interpretation using hadronic interaction models

By interpreting the reconstructed moments $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ using hadronic interaction models, the measurement can be converted into the first two moments of the distributions of the logarithmic mass [67, 68], its mean $\langle \ln {\rm A}\rangle$ and variance $\sigma^2(\ln A)$. In Fig. 14, the derived moments are shown using air-shower simulations based on the interaction models EPOS-LHC, Sibyll2.3d, and QGSJetII-04. The evolution of the mean logarithmic mass with energy shows a trend from a light composition towards a heavier composition, including the same characteristic breaks at three energies. Likewise, at

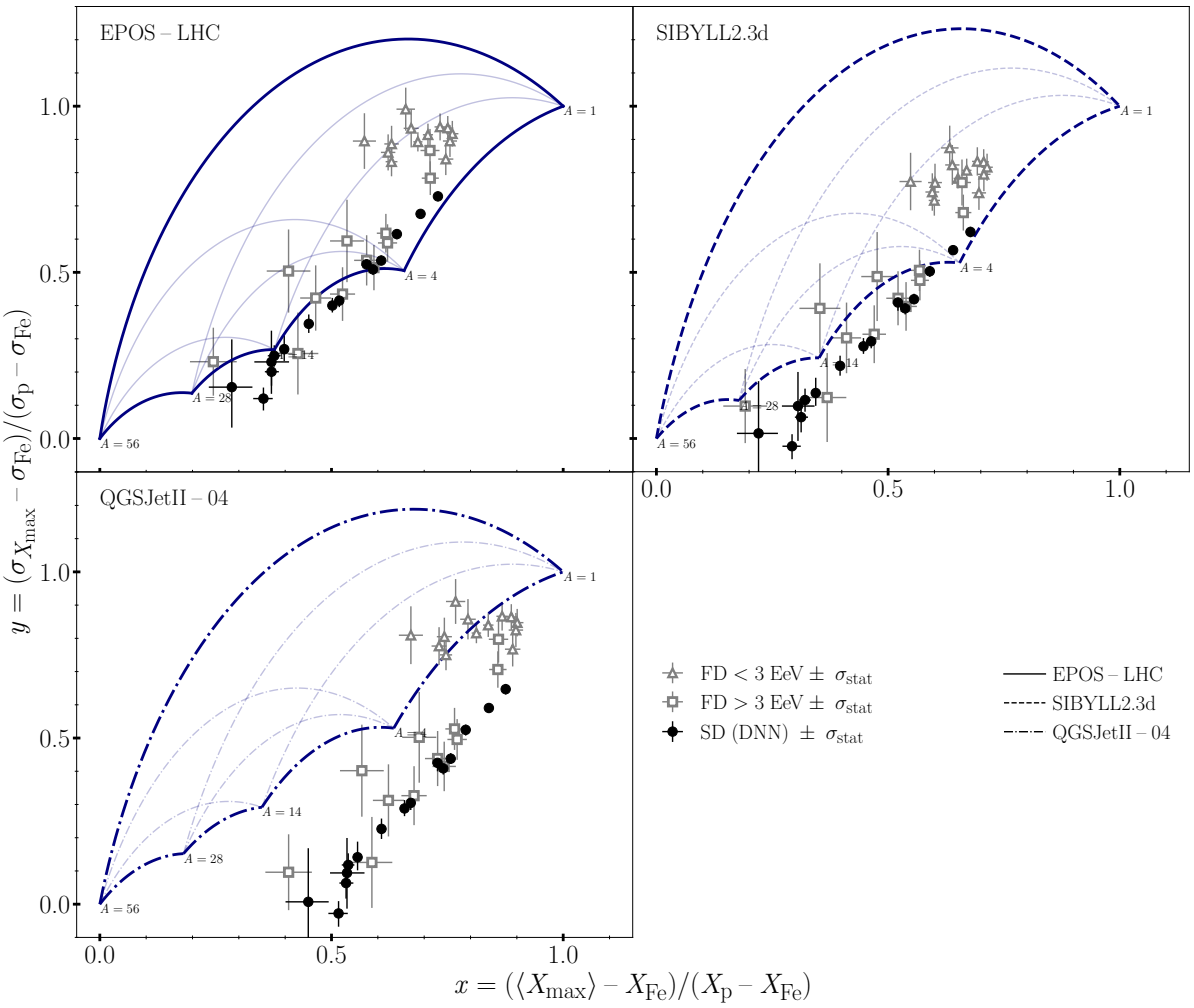


FIG. 15: Evolution of the measurements determined using the FD [62] (grey markers) and the DNN (black circles) in the re-scaled $\langle X_{\rm max}\rangle$ vs. $\sigma(X_{\rm max})$ representation compared to specific composition predictions for the hadronic interaction models EPOS-LHC, Sibyll2.3d, and QGSJetII-04 at 10 EeV. The arc-like curves denoted by blue lines indicate transitions between pairs of pure compositions labeled with their mass numbers (resulting in a so-called "umbrella" plot).

around 10 EeV and 30 EeV, the $\langle \ln {\rm A}\rangle$ shows indications of an almost constant composition. For all interaction models, the fluctuations $\sigma^2(\ln A)$ in $\ln A$ are small, indicating a composition dominated by a single type of nucleus. This observation exhibits a distinct characteristic that is quite compatible with the expectations for the Peters cycle. However, for quantitative results on the fluctuations of $\ln A$, the systematic uncertainties in the measurements, as well as the uncertainties in the interaction models, will need to be reduced.

Nonphysical negative fluctuations are found for QGSJetII-04 across the whole energy range, strongly disfavoring the model, in line with previous studies [21, 61, 62, 69]. Negative fluctuations for Sibyll2.3d and EPOS-LHC are also visible but are compatible with zero within uncertainties. Note that this result does not state that the fluctuations are not correctly modeled in simulations but rather that the fluctuations expected from a composition derived from the $\langle X_{\rm max}\rangle$ measurement are in tension with the model predictions. In fact, the uncertainties from the interaction-model description of the fluctuations are rather small, and parts of the mismatch found could likely originate from differences in the $X_{\rm max}$ scale in measured data and simulations. Indications for such a tension in the $\langle X_{\rm max}\rangle$ scale in simulation and data were previously reported in other studies [12, 58].

Another way of comparing the measured data to model predictions is the illustration of the data in a re-scaled $\sigma(X_{\rm max})$ vs. $X_{\rm max}$ plane [15, 70, 71]. First, in this representation, the measurements of $\langle X_{\rm max}\rangle$ are transformed into the scale of the respective model. Thus, $x = 0$ translates to a pure iron composition, and $x = 1$ corresponds to a pure proton composition. A similar transformation is applied to $\sigma(X_{\rm max})$ and denoted with y. Note that extremely mixed compositions would fea-

ture values larger than $y = 1$. Since the elongation rate for pure beams is, to a good approximation, universal across all interaction models and the energy-dependence of $\sigma(X_{\rm max})$ is small, the representation allows for a concise interpretation in which transitions between two pure compositions follow arc-like curves in these "umbrella" plots.

In Fig. 15, we show the measurement of the SD (black dots) and the FD (open grey squares), including only statistical uncertainties, and compare them to the predictions of the hadronic models EPOS-LHC, Sibyll2.3d, and QGSJetII-04. The blue lines indicate transitions between pairs of pure compositions at an energy of 10 EeV. FD measurements below 3 EeV are depicted as open grey triangles. The statistical uncertainties are estimated using bootstrapping and shown as vertical lines. In the matching energy range, the SD and FD data agree well and show the same evolution, demonstrating a consistent measurement of the two $X_{\rm max}$ moments.

Our measurements are consistent with a relatively heavy and pure composition for EPOS-LHC and Sibyll2.3d, within systematic uncertainties. Again, QGSJetII-04 shows a significant tension and is disfavored by our measurements. Furthermore, it can be seen that the features we found in the energy evolution of $\langle X_{\rm max}\rangle$ yield a consistent picture when including the $\sigma(X_{\rm max})$ measurement. Energy regions with smaller changes in $\langle \ln {\rm A}\rangle$ would appear as clusters of points. Two such regions are suggested by the elongation rate studies (cf. Fig. 13) and are also visible in Fig. 15. For example, for EPOS-LHC, the two regions at around 10 EeV and 50 EeV are close to the mass groups of $A \cong 4$ and $A \cong 14$.

V. SUMMARY

In this work, we have presented a study of the UHECR mass composition based on the first two moments of the distribution of depth of maximum, $X_{\rm max}$, of air shower profiles using surface detector data of the Pierre Auger Observatory recorded between 2004 and 2018. With the use of deep learning, a novel reconstruction technique was developed, enabling for the first time a precise reconstruction of $X_{\rm max}$ using the recorded time-dependent SD signals on an event-by-event level. Our approach included cross-calibration with the complementary FD to remove mismatches between simulations and measured data and investigate systematic uncertainties, highlighting the importance of an independent data set for calibrating and validating machine learning algorithms. After cross-calibrating the method using fluorescence observations, we have studied the energy evolution of $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ from 3 EeV up to 100 EeV. Due to the superior duty cycle of the SD in comparison to the FD, the statistics for composition studies using $X_{\rm max}$ are increased by a factor of ten for energies above 5 EeV, enabling for the first time a measurement of $\sigma(X_{\rm max})$, sensitive to the composition mixing, beyond 50 EeV. We have found excellent agreement of the $\langle X_{\rm max}\rangle$ measurement with previous studies using the FD and confirm the transition of $\langle \ln {\rm A}\rangle$ from a lighter to a heavier composition. Furthermore, our $\sigma(X_{\rm max})$ measurement, which is independent of the FD calibration, agrees very well with previous studies using fluorescence telescopes. The finding of a decrease in the fluctuations with energy is confirmed, indicating a transition to a heavier and purer composition. The observation of very small fluctuations appears to exclude a large fraction of light nuclei at the highest energies.

With the increase in statistics, we have found evidence at a level of 4.4σ for a characteristic structure in the evolution of the mass composition beyond a constant elongation rate, considering both statistical and systematic uncertainties. The model describing our data best features three breaks in the energy evolution of the composition and is compatible with the FD measurements. The locations of the identified breaks are found at energies similar to the ankle, instep, and suppression features identified in the UHECR energy spectrum. An interesting structure, while not statistically significant, is visible in $\sigma(X_{\rm max})$, which could suggest breaks at similar energies. However, more statistics and reduced systematic uncertainties are needed to study the nature of the identified breaks and, in particular, investigate the existence of the third break.

The study presented here is one of the first that uses deep learning to analyze measured detector data in astroparticle physics, including a comprehensive study of systematic uncertainties. The demonstrated performance, superior to previous approaches for mass composition studies using SD data, shows promising potential for machine-learning-based methods in astroparticle physics. The ongoing AugerPrime upgrade, including the upgrade of the water-Cherenkov detectors [72], as well as further improvements in machine-learning-based analysis strategies, opens up new and far-reaching prospects for understanding cosmic rays, their mass composition at ultra-high energies, and ultimately constraining astrophysical models of their origin.

ACKNOWLEDGMENTS

The successful installation, commissioning, and operation of the Pierre Auger Observatory would not have been possible without the strong commitment and effort from the technical and administrative staff in Malargüe. We are very grateful to the following agencies and organizations for financial support:

Argentina – Comisión Nacional de Energía Atómica; Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT); Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Gobierno de la Provincia de Mendoza; Municipalidad de Malargüe; NDM Holdings and Valle Las Leñas; in gratitude for their continuing cooperation over land access; Australia – the Australian Research Council; Belgium – Fonds de la Recherche Scientifique (FNRS); Research Foundation Flanders (FWO), Marie Curie Action of the European Union Grant No. 101107047; Brazil – Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Financiadora de Estudos e Projetos (FINEP); Fundação de Amparo à Pesquisa do Estado de Rio de Janeiro (FAPERJ); São Paulo Research Foundation (FAPESP) Grants No. 2019/10151-2, No. 2010/07359-6 and No. 1999/05404-3; Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC);

Czech Republic – GACR 24-13049S, CAS LQ100102401, MEYS LM2023032, CZ.02.1.01/0.0/0.0/16_013/0001402, CZ.02.1.01/0.0/0.0/18_046/0016010 and CZ.02.1.01/0.0/0.0/17_049/0008422 and CZ.02.01.01/00/22_008/0004632; France – Centre de Calcul IN2P3/CNRS; Centre National de la Recherche Scientifique (CNRS); Conseil Régional Ile-de-France; Département Physique Nucléaire et Corpusculaire (PNC-IN2P3/CNRS); Département Sciences de l'Univers (SDU-INSU/CNRS); Institut Lagrange de Paris (ILP) Grant No. LABEX ANR-10-LABX-63 within the Investissements d'Avenir Programme Grant No. ANR-11-IDEX-0004-02; Germany – Bundesministerium für Bildung und Forschung (BMBF); Deutsche Forschungsgemeinschaft (DFG); Finanzministerium Baden-Württemberg; Helmholtz Alliance for Astroparticle Physics (HAP); Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF); Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen; Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg; Italy – Istituto Nazionale di Fisica Nucleare (INFN); Istituto Nazionale di Astrofisica (INAF); Ministero dell'Università e della Ricerca (MUR); CETEMPS Center of Excellence; Ministero degli Affari Esteri (MAE), ICSC Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union NextGenerationEU, reference code CN_00000013; México – Consejo Nacional de Ciencia y Tecnología (CONACYT) No. 167733; Universidad Nacional Autónoma de México (UNAM); PAPIIT DGAPA-UNAM; The Netherlands – Ministry of Education, Culture and Science; Netherlands Organisation for Scientific Research (NWO); Dutch national e-infrastructure with the support of SURF Cooperative; Poland – Ministry of Education and Science, grants No. DIR/WK/2018/11 and 2022/WK/12; National Science Centre, grants No. 2016/22/M/ST9/00198, 2016/23/B/ST9/01635, 2020/39/B/ST9/01398, and 2022/45/B/ST9/02163; Portugal – Portuguese national funds and FEDER funds within Programa Operacional Factores de Competitividade through Fundação para a Ciência e a Tecnologia (COMPETE); Romania – Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, contract no. 30N/2023 under Romanian National Core Program LAPLAS VII, grant no. PN 23 21 01 02 and project number PN-III-P1-1.1-TE-2021-0924/TE57/2022, within PNCDI III; Slovenia – Slovenian Research Agency, grants P1-0031, P1-0385, I0-0033, N1-0111; Spain – Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación (PID2019-105544GB-I00, PID2022-140510NB-I00 and RYC2019-027017-I), Xunta de Galicia (CIGUS Network of Research Centers, Consolidación 2021 GRC GI-2033, ED431C-2021/22 and ED431F-2022/15), Junta de Andalucía (SOMM17/6104/UGR and P18-FR-4314), and the European Union (Marie Sklodowska-Curie 101065027 and ERDF); USA – Department of Energy, Contracts No. DE-AC02-07CH11359, No. DE-FR02-04ER41300, No. DE-FG02-99ER41107 and No. DE-SC0011689; National Science Foundation, Grant No. 0450696; The Grainger Foundation; Marie Curie-IRSES/EPLANET; European Particle Physics Latin American Network; and UNESCO.

[1] R. U. Abbasi (High Resolution Fly's Eye Collaboration) *et al.*, Phys. Rev. Lett. **100**, 101101 (2008).
[2] J. Abraham (Pierre Auger Collaboration) *et al.*., Phys. Rev. Lett. **101**, 061101 (2008).
[3] T. Abu-Zayyad (Telescope Array Project) *et al.*, ApJL **768**, L1 (2013).
[4] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. Lett. **125**, 121106 (2020).
[5] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
[6] G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4**, 78 (1966).
[7] B. Peters, Il Nuovo Cimento (1955-1965) **22**, 800 (1961).
[8] D. Allard, N. G. Busca, G. Decerprit, A. V. Olinto, and E. Parizot, JCAP **2008** (10), 033.
[9] A. Aab (Pierre Auger Collaboration) *et al.*, JCAP **2017** (04), 038.
[10] A. Abdul Halim (Pierre Auger Collaboration) *et al.*, JCAP **2023** (05), 024.
[11] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **91** (2015).
[12] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. Lett. **117**, 192001 (2016).
[13] A. Aab (Pierre Auger Collaboration) *et al.*, Eur. Phys. J. C **80**, 751 (2020).
[14] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. Lett. **126**, 152002 (2021).
[15] K.-H. Kampert, M. Unger, Astropart. Phys. **35**, 660 (2012).
[16] P. Abreu (Pierre Auger Collaboration) *et al.*, JCAP **2013** (02), 026.
[17] A. Aab (The Pierre Auger Collaboration) *et al.*, Nucl. Instrum. Methods A **798**, 172 (2015).
[18] H. Kawai *et al.*, Nucl. Phys. B. - Proceedings Supplements **175-176**, 221 (2008), proceedings of the XIV International Symposium on Very High Energy Cosmic Ray Interactions.
[19] J. Abraham (Pierre Auger Collaboration) *et al.*, Phys. Rev. Lett. **104**, 091101 (2010).
[20] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **90**, 122006 (2014).
[21] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **96**, 122003 (2017).
[22] A. Bridgeman on behalf of the Pierre Auger Collaboration, PoS **ICRC2017**, 323 (2017).
[23] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[25] M. Erdmann, J. Glombitza, G. Kasieczka, and U. Klemradt, *Deep Learning for Physics Research* (World Scientific, 2021).
[26] I. Shilon *et al.*, Astropart. Phys. **105**, 44 (2019).
[27] E. Cuoco *et al.*, Mach. learn.: sci. technol. **2**, 011002 (2020).
[28] R. Abbasi (IceCube Collaboration) *et al.*, JINST **16** (07), P07041.
[29] S. Aiello (KM3NeT Collaboration) *et al.*, JINST **15** (10), P10005.
[30] R. Abbasi (IceCube Collaboration) *et al.*, Science **380**, 1338 (2023).

[31] M. Erdmann, J. Glombitza, and D. Walz, Astropart. Phys. **97**, 46 (2018).
[32] A. Aab (Pierre Auger Collaboration) *et al.*, JINST **16** (07), P07019.
[33] J. Glombitza on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 278 (2023).
[34] A. Aab (Pierre Auger Collaboration) *et al.*, JINST **16** (07), P07016.
[35] J. Glombitza on behalf of the Pierre Auger Collaboration, PoS **ICRC2019**, 270 (2019).
[36] A. Abdul Halim (Pierre Auger Collaboration) *et al.*, companion Letter, submitted to Phys. Rev. Lett. (2024).
[37] X. Bertou (Pierre Auger Collaboration) *et al.*, Nucl. Instrum. Methods A **568**, 839–846 (2006).
[38] A. Aab (Pierre Auger Collaboration) *et al.*, JINST **15** (10), P10021.
[39] A. Aab (Pierre Auger Collaboration) *et al.*., Phys. Rev. D **102**, 062005 (2020).
[40] M. Ave, M. Roth, and A. Schulz, Astropart. Phys. **88**, 46 (2017).
[41] M. Stadelmaier, M. Roth, D. Schmidt and D. Veberic, PoS **ICRC2021**, 432 (2021).
[42] S. Hochreiter and J. Schmidhuber, Neural Comput. **9**, 1735 (1997).
[43] E. Hoogeboom, J. W. T. Peters, T. S. Cohen, and M. Welling, (2018), arXiv:1803.02108 [cs, stat].
[44] K. He, X. Zhang, S. Ren, and J. Sun, (2015), arXiv:1512.03385 [cs].
[45] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, (2018), arXiv:1608.06993 [cs].
[46] E. Santos on behalf of the Pierre Auger Observatory, PoS **ICRC2023**, 248 (2023).
[47] E. Santos on behalf of the Pierre Auger Observatory, PoS **ICRC2021**, 232 (2021).
[48] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, FZKA **6019** (1998).
[49] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, Phys. Rev. C **92**, 034906 (2015).
[50] A. Fasso *et al.*, (2003), arXiv:hep-ph/0306267 [hep-ph].
[51] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, Phys. Rev. D **102**, 063002 (2020).
[52] J. Glombitza on behalf of the Pierre Auger Collaboration, PoS **ICRC2021**, 359 (2021).
[53] P. Abreu (Pierre Auger Collaboration) *et al.*, EPJ C **81** (2021).
[54] J. Bellido on behalf of the Pierre Auger Collaboration, PoS **ICRC2017**, 506 (2017).
[55] R. Sato on behalf of the Pierre Auger Collaboration, in *32nd ICRC* (2011) arXiv:1107.4806.
[56] O. Zapparrata on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 266 (2023).
[57] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **90**, 012012 (2014).
[58] A. Abdul Halim (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **109**, 102001 (2024).
[59] F. Riehn, A. Fedynitch, and R. Engel, Astropart. Phys. **160**, 102964 (2024).
[60] M. Erdmann, L. Geiger, J. Glombitza, and D. Schmidt, Comput Softw Big Sci **2**, 4 (2018).
[61] A. Aab (Pierre Auger Collaboration) *et al.*, Phys. Rev. D **90**, 122005 (2014).
[62] A. Yushkov on behalf of the Pierre Auger Collaboration, PoS **ICRC2019**, 482 (2019).
[63] J. Bellido on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 211 (2023).
[64] V. M. Harvey on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 300 (2023).
[65] T. Fitoussi on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 319 (2023).
[66] M. de Domenico, M. Settimo, S. Riggi, and E. Bertin, JCAP **2013** (07), 050.
[67] P. Abreu (Pierre Auger Collaboration) *et al.*, JCAP **2013** (02), 026.
[68] A. Aab (Pierre Auger Collaboration) *et al.*, (2013), arXiv:1307.5059 [astro-ph.HE].
[69] T. Fitoussi on behalf of the Pierre Auger Collaboration, PoS **ICRC2023**, 319 (2023).
[70] J. Linsley, Proc. 19th ICRC 6 , 1 (1985).
[71] P. Lipari, Phys. Rev. D **103** (2021).
[72] A. Castellina on behalf of the Pierre Auger Collaboration, EPJ Web Conf. **210**, 06002 (2019).

APPENDIX

Derivation of the formula for the reconstruction of $\sigma(X_{\rm max})$. Here, $X_{\rm max}$ denotes the true depth of the shower maximum and $X_{\rm max,DNN}$ the reconstruction of the DNN.

$$\begin{aligned}\sigma^2(X_{\rm max,DNN}) &= \sigma^2(X_{\rm max} + X_{\rm max,DNN} - X_{\rm max}) \\ &= \sigma^2(X_{\rm max}) + \sigma^2(X_{\rm max,DNN} - X_{\rm max}) \\ &\quad + 2\,{\rm Cov}(X_{\rm max}, X_{\rm max,DNN} - X_{\rm max})\end{aligned} \tag{7}$$

TABLE IV: First two moments of the $X_{\rm max}$ distributions. Energies are given in $\log_{10}E/{\rm eV}$ and $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ are given in $\mathrm{g\,cm^{-2}}$ followed by their statistical and systematic uncertainties.

$\log_{10}E/{\rm eV}$ bin	$\langle\log_{10}E/{\rm eV}\rangle$	$\langle X_{\rm max}\rangle$ / $\mathrm{g\,cm^{-2}}$	$\sigma(X_{\rm max})$ / $\mathrm{g\,cm^{-2}}$
18.5–18.6	18.55	$757.0 \pm 0.5^{+7.7}_{-9.6}$	$48.8 \pm 0.5\ ^{+9.1}_{-4.8}$
18.6–18.7	18.65	$758.8 \pm 0.5^{+7.8}_{-9.2}$	$46.3 \pm 0.5\ ^{+8.7}_{-4.8}$
18.7–18.8	18.75	$759.5 \pm 0.5^{+7.9}_{-8.8}$	$43.4 \pm 0.5\ ^{+8.3}_{-4.8}$
18.8–18.9	18.85	$761.8 \pm 0.5^{+8.0}_{-8.5}$	$39.7 \pm 0.5\ ^{+8.0}_{-4.8}$
18.9–19.0	18.95	$765.7 \pm 0.5^{+8.2}_{-8.2}$	$38.4 \pm 0.5\ ^{+7.6}_{-4.8}$
19.0–19.1	19.05	$770.0 \pm 0.6^{+8.4}_{-7.9}$	$38.8 \pm 0.6\ ^{+7.3}_{-4.8}$
19.1–19.2	19.15	$769.9 \pm 0.6^{+8.7}_{-7.7}$	$34.0 \pm 0.6\ ^{+7.1}_{-4.8}$
19.2–19.3	19.25	$774.0 \pm 0.8^{+9.1}_{-7.5}$	$33.2 \pm 0.8\ ^{+7.0}_{-4.8}$
19.3–19.4	19.35	$774.7 \pm 0.9^{+9.5}_{-7.4}$	$30.7 \pm 0.9\ ^{+6.9}_{-4.8}$
19.4–19.5	19.45	$775.3 \pm 1.0^{+10.0}_{-7.3}$	$27.3 \pm 1.0\ ^{+6.8}_{-4.8}$
19.5–19.6	19.55	$778.6 \pm 1.3^{+10.5}_{-7.3}$	$26.3 \pm 1.3\ ^{+6.8}_{-4.8}$
19.6–19.7	19.64	$783.2 \pm 1.4^{+11.0}_{-7.3}$	$24.2 \pm 1.4\ ^{+6.8}_{-4.8}$
19.7–19.8	19.74	$787.2 \pm 2.0^{+11.6}_{-7.5}$	$20.7 \pm 2.0\ ^{+6.8}_{-4.8}$
19.8–19.9	19.85	$794.5 \pm 3.6^{+12.2}_{-7.8}$	$25.1 \pm 3.6\ ^{+6.8}_{-4.8}$
> 19.9	20.00	$793.9 \pm 4.5^{+13.1}_{-8.3}$	$21.8 \pm 4.5\ ^{+6.8}_{-4.8}$

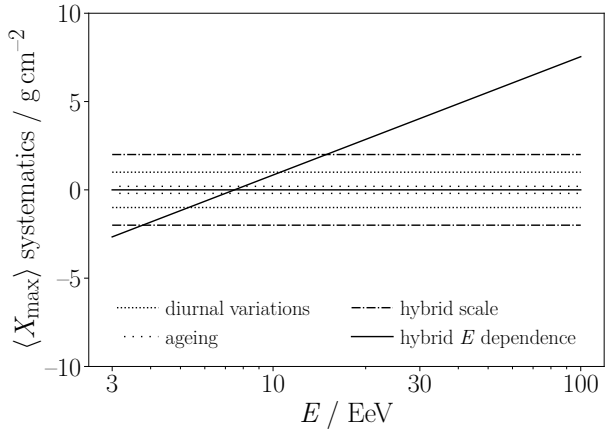
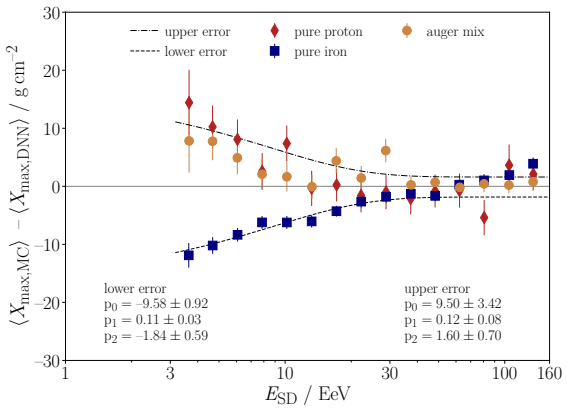


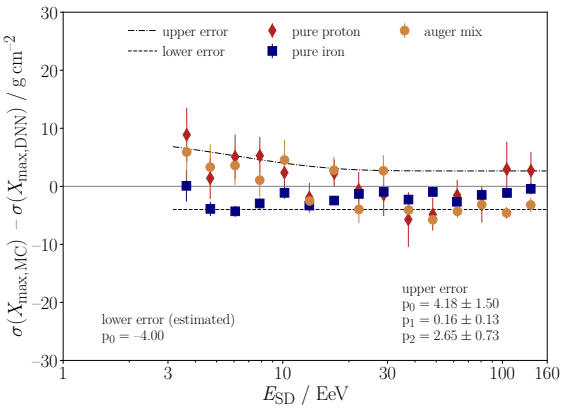
FIG. 16: Energy-dependent contributions of the SD systematic uncertainties for the measurement of $\langle X_{\rm max}\rangle$ after calibrating DNN to FD observations.

TABLE V: Available statistics for determining the UHECR composition using the FD and the DNN. FD data are taken from Ref. [62].

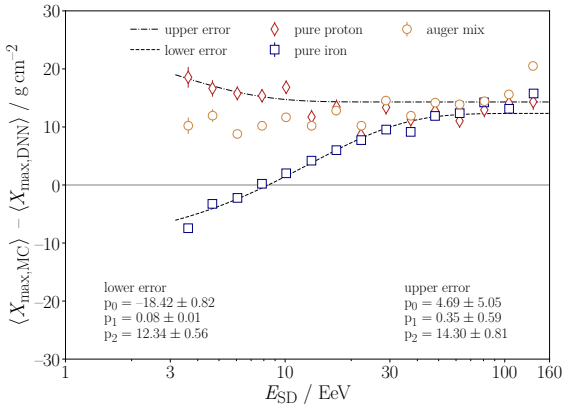
$\log_{10}E/{\rm eV}$ bin	$X_{\rm max,FD}$	$X_{\rm max,DNN}$
18.5–18.6	1,347	8,739
18.6–18.7	1,007	9,360
18.7–18.8	707	7,725
18.8–18.9	560	6,506
18.9–19.0	417	5,228
19.0–19.1	312	3,863
19.1–19.2	253	2,781
19.2–19.3	159	1,791
19.3–19.4	122	1,205
19.4–19.5	80	701
19.5–19.6	50	455
19.6–19.7	35	277
19.7–19.8		113
19.8–19.9		54
> 19.9		26



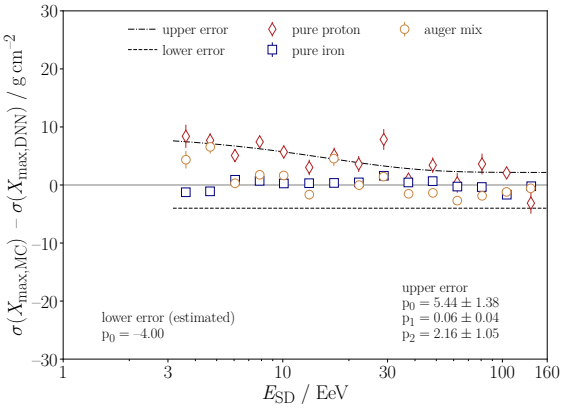
(a) EPOS-LHC



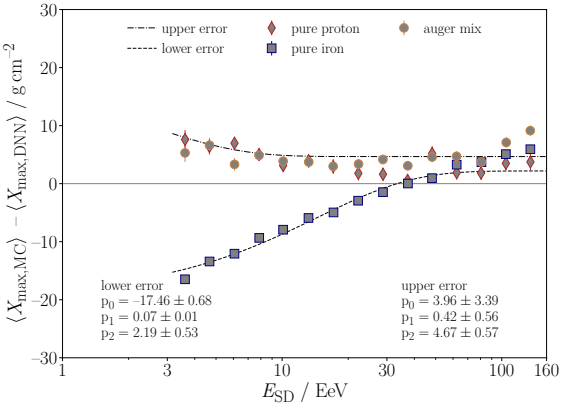
(b) EPOS-LHC



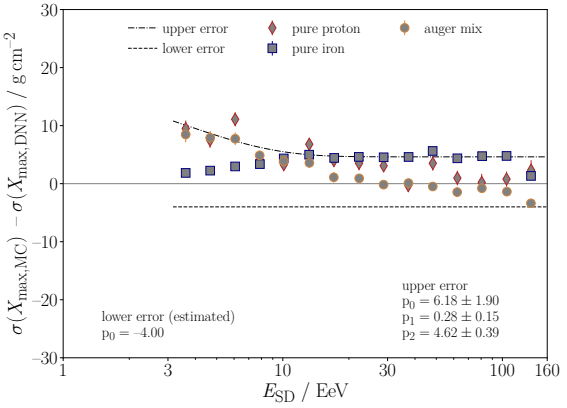
(c) Sibyll2.3c



(d) Sibyll2.3c



(e) QGSJetII-04



(f) QGSJetII-04

FIG. 17: Expected composition bias for measuring the first moment $\langle X_{max} \rangle$ (left) and the second moment $\sigma(X_{max})$ (right) of X_{max} distributions as a function of energy for EPOS-LHC, Sibyll2.3c, and QGSJetII-04 (from top to bottom) after forward-folding of all systematic effects on the measurement. The different compositions are denoted by different colors. The dashed line indicates a parameterization for the composition bias. Note that only the parameterization for $\sigma(X_{max})$ propagates into the uncertainty of the measurement, as for $\langle X_{max} \rangle$ the method is cross-calibrated using the FD. Note that EPOS-LHC was used as the hadronic interaction model in the training of the network.

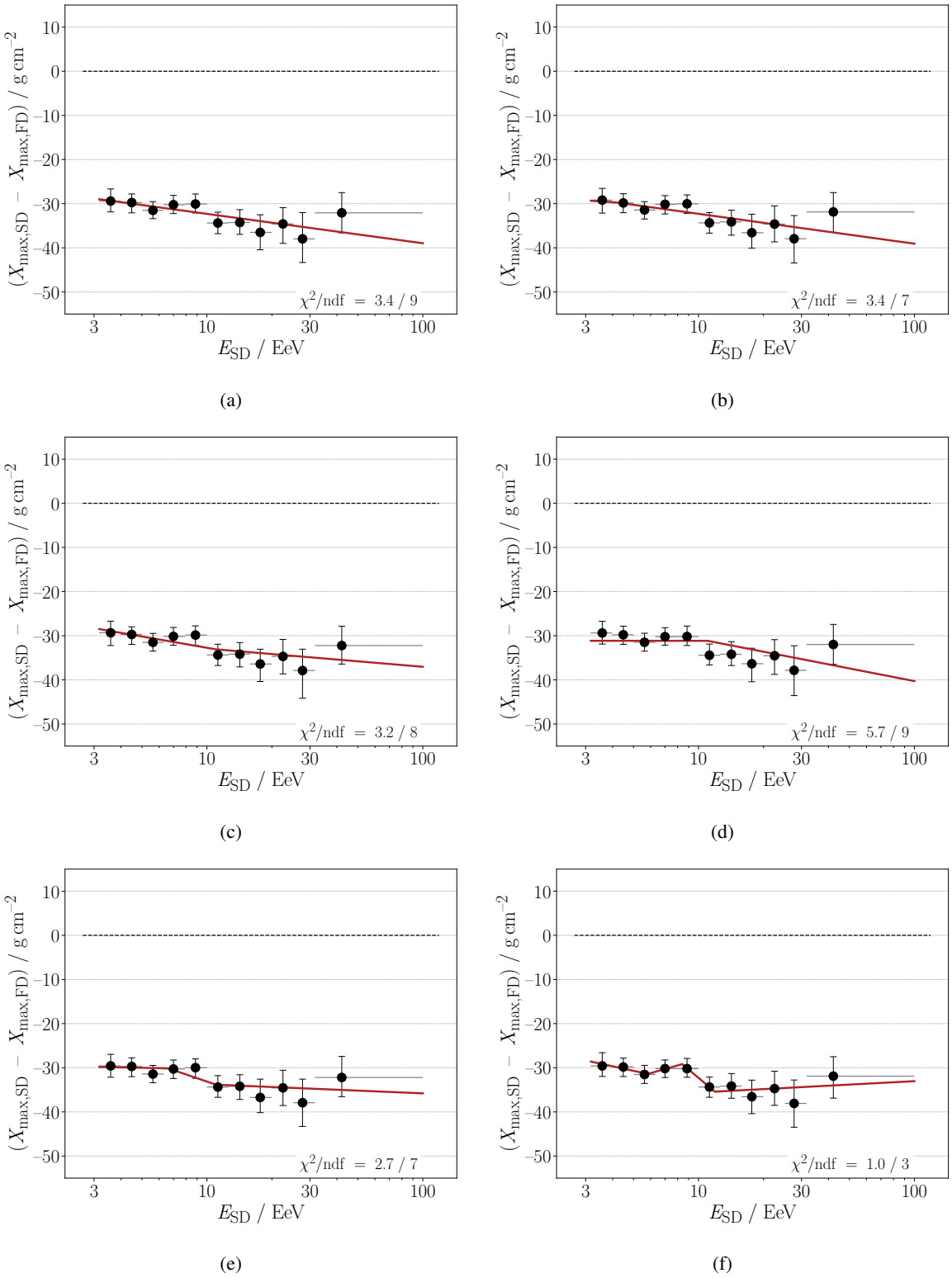


FIG. 18: Models used for studying the effect of energy-dependent calibrations on the measurement of $\langle X_{\mathrm{max}} \rangle$ and the significance of the identified features. (a) Linear function considered for a global energy dependence of the calibration (used for estimating the systematic uncertainty of $\langle X_{\mathrm{max}} \rangle$). (b) Piecewise-linear function. (c) Piecewise-linear function with the break fixed to the position of the fitted second break. A similar dependence could also be motivated by the composition bias of EPOS-LHC (used for training) using the Auger mix (cf. Fig. 17a). (d) Piecewise-linear fit with the first slope fixed to 0 g cm^{-2} and the break fixed to the position of the second break (e) 3-fold piecewise-linear fit with the first and second break fixed to the position of the first and second break. (f) Piecewise-linear function with three adaptive breaks. None of the calibrations lowers the significance considerably.