Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV

Hans Peter Dembinski[*]
Max Planck Institute for Nuclear Physics, Heidelberg, Germany
E-mail: hdembins@mpi-hd.mpg.de

Ralph Engel
Karlsruhe Institute of Technology, Karlsruhe, Germany
E-mail: ralph.engel@kit.edu

Anatoli Fedynitch
DESY Zeuthen, Zeuthen, Germany
E-mail: anatoli.fedynitch@desy.de

Thomas Gaisser
Bartol Institute, University of Delaware, Newark DE, USA
E-mail: gaisser@bartol.udel.edu

Felix Riehn
LIP, Lisbon, Portugal
E-mail: friehn@lip.pt

Todor Stanev
Bartol Institute, University of Delaware, Newark DE, USA
E-mail: stanev@bartol.udel.edu

We present a new parametrization of the cosmic-ray flux and its mass composition over an energy range from 10 GeV to 10^{11} GeV. Our approach is data-driven and relies on theoretical assumptions as little as possible. We combine measurements of the flux of individual elements from high-precision satellites and balloon experiments with indirect measurements of mass groups from the leading air shower experiments. To our knowledge, we provide the first fit of this kind that consistently takes both statistical and systematic uncertainties into account. The uncertainty on the energy scales of individual experiments is handled explicitly in our mathematical approach. Part of our results is a common energy scale and adjustment factors for the energy scales of the participating experiments. Our fit has a reduced χ^2-value of 0.5, showing that experimental data are in good agreement, if systematic uncertainties are considered. Our model may serve as a world-average of the measured fluxes for individual elements from proton to iron from 10 GeV to 10^{11} GeV. It is useful as an input for simulations or theoretical computations based on cosmic rays. The experimental uncertainties of the input data are captured in a covariance matrix, which can be propagated into derived quantities.

[*]Speaker.

1. Introduction and motivation

The cosmic-ray flux spans more than 11 decades in energy. Individual experiments cover only a part of that range, which means that a common picture needs to be pieced together from many data sets. The *Database of Charged Cosmic Rays* (CRDB) [1] is a commendable effort in providing central machine-readable access to cosmic-ray data from satellite and balloon experiments. These experiments are able to separate elements by their charge Z and report the flux per element.

Above a few 100 TeV, direct observation is not feasible and ground-based experiments take over. These measure a cosmic ray indirectly through secondary particles produced in a cosmic-ray induced air shower. Air-shower experiments achieve large apertures, but loose the ability to discriminate the charge Z of each cosmic ray. The distinguishing feature is now the mass A, which has a measurable effect on several air-shower observables, but stochastic fluctuations in the shower development overshadow these signatures. Ground-based experiments report the all-particle flux, and potentially the flux fractions contributed by elements which fall into a given mass range.

The inference of the cosmic-ray energy E and mass A from air-shower observables depends on theoretical models. The air-shower development is driven by soft QCD interactions, which are computed from phenomenological models tuned to collider data. The theoretical uncertainties in these models cause larger systematic uncertainties on the reported results compared to direct measurements. In particular, the energy scale of each experiment has a non-negligible uncertainty of about $(10-20)\%$, which produces apparent discrepancies between independent experiments.

We present a new data-driven model, called the *Global Spline Fit* (GSF). It parametrizes the latest and most detailed measurements of the cosmic-ray flux and its composition from 10 GeV to 10^{11} GeV, combining direct and air-shower observations. Smooth curves of the fluxes of all elements from proton to nickel are provided. Energy-scales of included experiments are cross-calibrated based on the successive overlap of data sets in energy, and a common energy scale is established which is fixed by direct measurements. The GSF yields a covariance matrix which represents the experimental uncertainty of the input data. The covariance matrix can be used to compute standard deviations for quantities derived from the cosmic ray flux and composition.

In contrast to other recent models [2, 3, 4], the flux in this work is not constructed from components with power-law shape and rigidity-dependent cut-offs. The GSF only assumes that the flux is smoothly varying and otherwise relies only on the available data. It is not a model in the usual sense since it does not try to explain the data, it just describes an average. This approach has become feasible recently, since air-shower experiments now cover the energy range above 1 PeV with detailed data about the cosmic-ray mass composition.

The GSF model, as presented here, is used in a state-of-the-art calculation of atmospheric lepton fluxes also presented at this conference [5]. In the following, we will describe the GSF approach, and discuss our preliminary results. The final data tables and code for the GSF model will be published later this year, allowing everyone to reproduce these results.

2. Global Spline Fit

In the GSF model, the cosmic-ray flux is divided into four mass groups, which cover roughly equal ranges in logarithmic mass ln$A$, as shown on the left-hand side in Fig. 1. We split in ln$A$, since
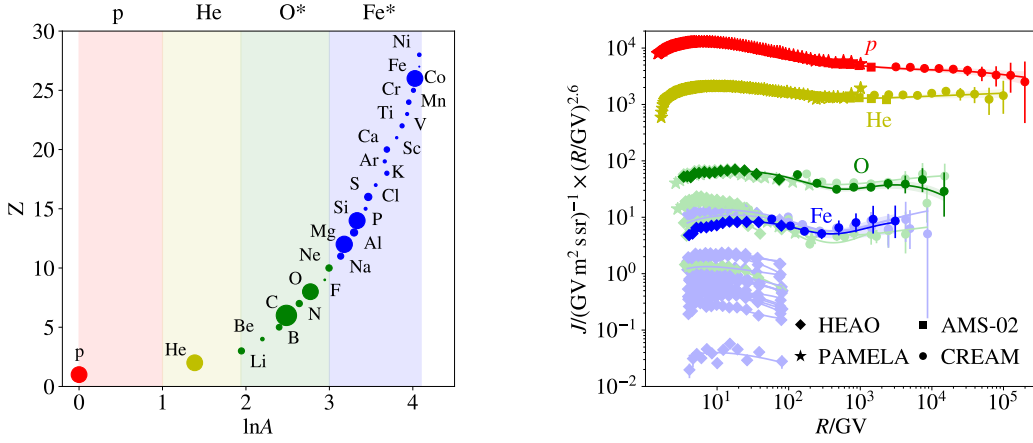
Figure 1: *Left:* Elements from proton to nickel are sorted into four mass groups, covering roughly equal intervals in logarithmic mass ln$A$. The groups are named after the leading element, which provides most to the differential flux $\propto \mathrm{d}N/\mathrm{d}E$ of the group. The size of the marker indicates the flux ratio f_{iL} relative to the leading element L of the group, as described in the text. The values are obtained from fits to HEAO data [6]. *Right:* Fluxes of individual elements measured by satellites and balloons [6, 7, 8, 9, 10, 11, 12], together with fitted spline curves. Highlighted points and curves show the four leading elements, fainter points and curves sub-leading elements from the oxygen and iron groups.

air-shower measurements are sensitive to changes in ln$A$ rather than A. Each group has a leading element L that contributes most of the flux per energy interval. We note that if two elements have the same abundance in flux per rigidity interval $J(R) \propto \mathrm{d}N/\mathrm{d}R$, the element with the higher charge contributes more to the flux per energy interval $\mathrm{d}N/\mathrm{d}E$. The leading elements are thus the heaviest abundant elements in each group; namely proton, helium, oxygen, and iron. The oxygen and iron groups contain many sub-leading elements. In the oxygen group, carbon contributes nearly as much as oxygen. In the GSF model, the flux $J_i(R)$ of a sub-leading element i is kept in a constant ratio f_{iL} to the leading element $L \in \{p, \mathrm{He}, \mathrm{O}, \mathrm{Fe}\}$ of its group, $J_i(R) = f_{iL} \times J_L(R)$.

This treatment is motivated empirically by low energy data, shown on the right-hand-side in Fig. 1. If the differential flux is plotted as a function of rigidity, the fluxes of neighboring elements roughly have the same shape. In other words, the flux ratio of elements within a group is a very slowly varying function of ln$R$. We approximate by keeping this ratio constant.

This approximation is rough, but only applied to sub-leading elements. Leading elements are fitted to experimental data and pull sub-leading elements with them. For many quantities computed from the GSF model, the exact ratios of sub-leading elements are not important; for example, for the nucleon flux and mean-logarithmic mass. Sub-leading elements however cannot be neglected. The fluxes of sub-leading elements contribute roughly a factor of two enhancement over the elemental fluxes of oxygen and iron for the respective mass groups. Air-shower measurements can only distinguish mass groups, so this additional flux has to be included. Neglecting this leads to artifacts in the description of the transition from direct to air-shower measurements.

The differential flux $J_L(R)$ of each leading element L is parametrized by a smooth curve. We use a modified spline curve, build from a linear combination of B-splines [13] shaped by a power-

law term,

$$J_L(R) = \sum_k a_{Lk} \times b_k\big(\ln(R/\mathrm{GV})\big) \times (R/\mathrm{GV})^{-3}, \tag{2.1}$$

where $b_k\big(\ln(R/\mathrm{GV})\big)$ are standard cubic B-splines defined over a common vector of knots, a_{Lk} are the coefficients adjusted in a global fit to all data sets, and $(R/\mathrm{GV})^{-3}$ is the shaping factor. The latter stabilizes numerical results. Without the shaping factor, the coefficients a_{Lk} would vary over many orders of magnitude. This would lead to precision loss in numerical computations.

The knot locations form a regular grid in $\ln(R/\mathrm{GV})$, with a typical step size of 0.5. We double (half) the step size in regions where the flux varies more (less). We tested that the exact placement of knots has negligible impact on our results by varying all knot positions by ± 0.1.

Our model parametrizes the differential flux of nuclei per rigidity interval $J(R)$. Air-shower measurements of the cosmic-ray flux are reported as the differential flux of nuclei per energy interval $J(E)$. The latter is computed from the former as $J(E) = J(R)\,\mathrm{d}R/\mathrm{d}E$. The relationship between total energy E and rigidity R depends on the number of nucleons A and protons Z and is computed individually for each element. Air-shower measurements describe the flux of mass groups. We sum the flux of all elements in each group when comparing the model to such measurements.

We consider the effect of a potential energy-scale offset $z_E = (\tilde{E} - E)/E$ for each experiment, where $\tilde{E}$ is the energy reported by an experiment for the differential flux $\tilde{J}(\tilde{E})$ reported by an experiment. If $J(E)$ is the true flux at the true energy E, the apparent flux is

$$\tilde{J}(\tilde{E}) = J(E)\frac{\mathrm{d}E}{\mathrm{d}\tilde{E}} = J\left(\frac{\tilde{E}}{1+z_E}\right)\frac{1}{1+z_E}. \tag{2.2}$$

For a power-law spectrum $J(E) \propto E^{-\alpha}$, the apparent flux is changed by a factor of $(1 + z_E)^{\alpha-1}$ when plotted on the same graph. For example, a small energy-scale offset of 10 % results in an apparent increase of 18 % for a spectral index $\alpha = 2.7$. This effect is well known [2, 3], and has been corrected by hand before. To our knowledge, the GSF approach is the first that derives the energy-scale offsets from a global fit, so that the corrections are a result instead of an input.

With these ingredients, we can compute residuals between measurements and the GSF model, and find spline amplitudes which minimize these residuals. We use a global least-squares fit to all data sets at once. We numerically find the spline amplitudes a_{Lk} and energy-scale offsets $z_{E,l}$ by minimizing the function

$$F' = \sum_i F_i + \sum_i \left(\frac{z_{E,i}}{\sigma_{E,i}}\right)^2, \tag{2.3}$$

where the term F_i represents the sum of squared residuals between data points and the GSF model for an experiment i, which depend on the spline amplitudes and the energy-scale offset $z_{E,i}$ for that experiment. To compute residuals, we take both statistical and systematic uncertainties into account. We also include correlations of systematic uncertainties, because systematic effects typically affect several data points similarly. A detailed treatment of systematic uncertainties is very important in a fit of cosmic-ray data, since they are often the largest uncertainties.

Each offset $z_{E,i}$ is handled as a squared residual over the relative energy-scale uncertainty $\sigma_{E,i}$ of experiment i and added to F'. If the offsets $z_{E,i}$ are zero (no energy-scale adjustment), the second sum is minimal, but the first sum may increase. The global fit finds a balance, where energy-scale offsets minimize the deviations of experiments from the model (and therefore from each other), but

do not become too large compared to the energy-scale uncertainties. This approach is a consistent extension of the standard least-squares method [14].

3. Data sets selected as input

Many experiments provide cosmic-ray data [15]. For the GSF, we use only the most recent measurements which divide the flux into separate contributions from mass groups. In case of direct measurements, we favor satellite over balloon experiments, since satellites do not have to correct for interactions in the residual atmosphere above the balloon. We make exceptions from these rules only to fill gaps in our targeted energy range.

We use direct measurements from the HEAO satellite [6], PAMELA [9, 10], AMS-02 [7, 8], CREAM-I and II [11, 12]. Solar modulation is corrected based on the force-field approximation [16], using modulation parameters from the CRDB [1, 17]. We use air-shower measurements from ARGO-YBJ [18], TUNKA [19, 20], IceCube [21], KASCADE-Grande (KG) [22, 23], Telescope Array (TA) [24], and the Pierre Auger Observatory (Auger) [25, 26]. At the time of this writing, Telescope Array has not published a split of their all-particle flux into individual mass groups. It does not pass our criteria, but is included for its special significance of providing an all-particle flux that covers six decades in energy. For ARGO-YBJ, we use the reported proton+helium flux only below 1 PeV; above 1 PeV the flux drops sharply, in contradiction to three other data sets.

For a global fit, it is important that experiments report systematic uncertainties. The composition results from TUNKA, IceCube, and KASCADE-Grande do not yet have published systematic uncertainties. Fitting data sets with and without systematic uncertainties leads to inconsistencies. It puts the former at a disadvantage; data sets with systematic uncertainties have less pull on the model, although an evaluation of systematic uncertainties increases the reliability of results.

To prevent such inconsistencies, we assign 10 % uncertainty to composition measurements reported by TUNKA, and 15 % to those of IceCube. These estimates are based on the observed deviations of these results from each other. We assign 20 % to the KASCADE-Grande composition, a preliminary conservative estimate suggested by the authors. This is not a satisfactory solution, of course, but one we suggest until final results with systematic uncertainties are published.

Composition results from air-shower experiments are sensitive to the hadronic interaction model used to interpret the raw data, and models vary in our data sets. A single-model interpretation over the whole energy range is currently not possible with the available data. The Pierre Auger Observatory provides three composition results for three different hadronic interaction models [26]. We average these here, and treat the model variation as a systematic uncertainty.

4. Results and discussion

The fitted GSF model is shown in Fig. 2 and Fig. 3. The global fit together with the adjusted energy-scales reveals detailed structure in the all-particle flux. We observe an unnamed dip around 5 TeV, the *knee* around 3 PeV, a *second knee* around 100 PeV, the *ankle* around 8 EeV, a slight hardening around 20 PeV, and finally the *toe* around 60 EeV. The energy-scale offsets found by the GSF are compatible with the reported systematic uncertainties.
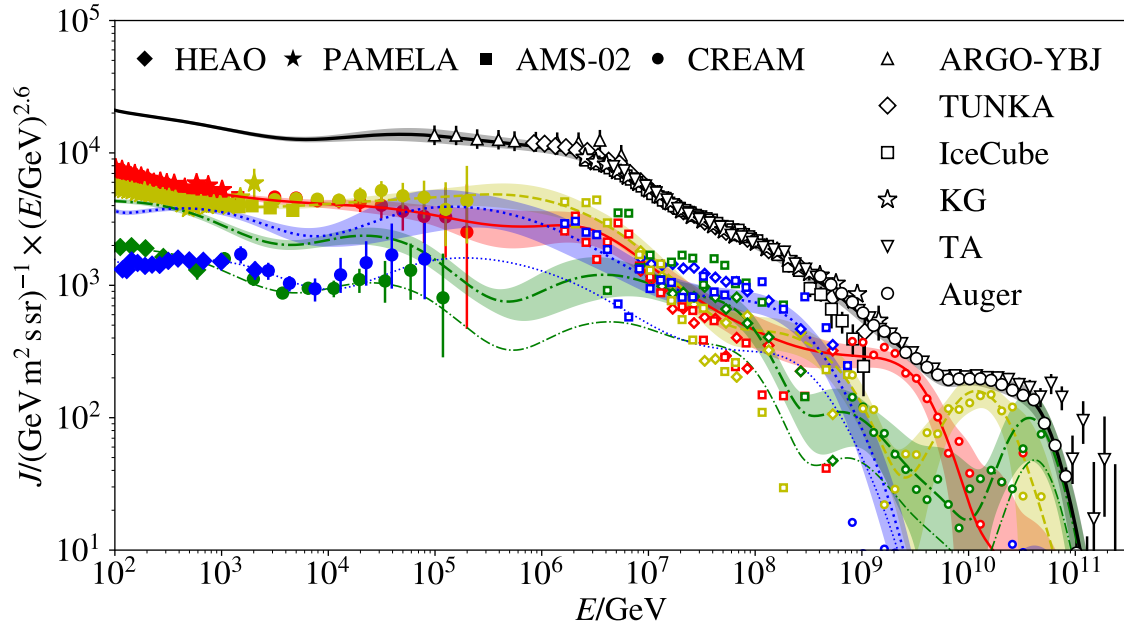
4

Figure 2: All-particle flux (black thick solid line), the flux contributed by protons (red line solid line), helium (yellow dashed line), the oxygen group (gray dash-dotted line), and the iron group (blue dotted line). Bands around the model lines show a variation of one standard deviation. Data points show measurements which were energy-scale adjusted as described in the text. Error bars represent combined statistical and systematic uncertainties. Data points of composition measurements from air-showers are not shown without error bars for clarity. In case of oxygen and iron, both the elemental flux and the group flux are shown; the smaller flux without error band is the elemental flux in each case. TA stands for Telescope Array, KG for KASCADE-Grande.
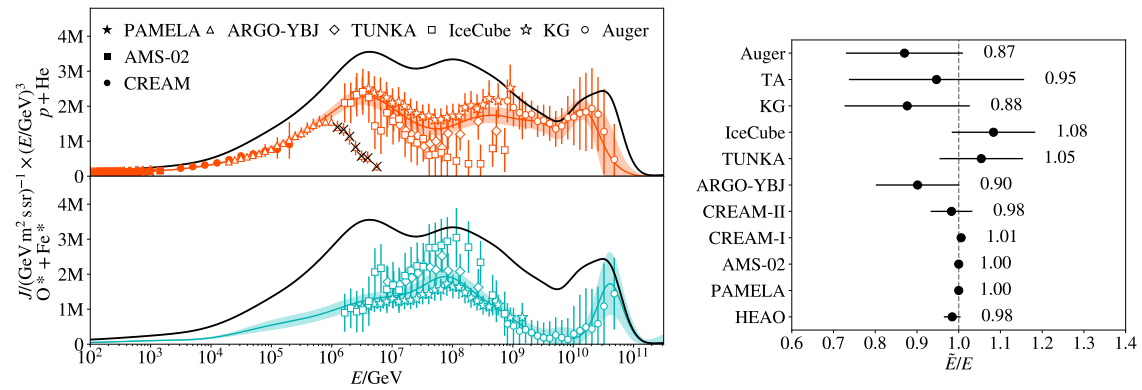

Figure 3: *Left:* Particle flux in linear scale, split into a light (proton and helium) and heavy (others elements) component. The black solid line represents the all-particle flux. KASCADE-Grande and ARGO-YBJ reported their composition measurements in this split. For other experiments, synthetic points are generated for visual comparison from the more detailed composition data. The ARGO-YBJ points marked with crosses are not used in the fit. *Right:* Ratios of energy scales used by an experiment relative to the cross-calibrated energy scale established by the GSF model. Error bars represent the reported systematic uncertainties of the energy scales. TA stands for Telescope Array, KG for KASCADE-Grande.

We quantify the agreement of the GSF with the input data by computing the sum of squared residuals, taking both statistical and systematic uncertainties into account. We obtain a χ^2 of 385.2 for 724 degrees of freedom, which indicates a good fit and implies that the data sets are overall consistent with each other, if systematic uncertainties are taken into account. We remind, however, that this is achieved by rejecting part of the ARGO-YBJ data, and assigning $(10-20)\%$ systematic uncertainty to results where none were reported.

Telescope Array and the Pierre Auger Observatory show remarkable agreement, except for a notable discrepancy in the all-particle flux at the toe. It has been noted that these experiments observe different hemispheres and so the flux could be different, but so far there is no conclusive evidence for this hypothesis [25]. The GSF model is build on the assumption that the flux is isotropic and therefore pulled towards the Auger data points, which have much smaller uncertainties.

Based on the GSF view on the flux and its mass composition, we can make a few observations. The first knee is created by a maximum in the proton and helium components, the second knee mostly by a maximum in the iron mass group. The mass composition data from the Pierre Auger Observatory at the high end of the cosmic-ray spectrum matches a Peters cycle [27] with a rigidity-dependent cut-off. The proton component drops before the ankle, so that the toe is dominated by helium and the oxygen group. The flux has a unique proton-maximum around 2 EeV, before the ankle. We note that the all-particle flux between the second knee and the ankle looks like a simple power-law, although the mass composition changes drastically in the same energy range.

The GSF model is a framework to compute a world-average of the cosmic-ray flux and elemental composition, inspired by what the Particle Data Group does for various other measurements. We believe that a comprehensive model based on experimental data is a valuable input for many other analyses that compute derived quantities, like the atmospheric lepton flux. The GSF model summarizes the data provided by leading experiments, providing a combined estimate and a covariance matrix to compute the standard deviation of derived quantities. It establishes an common energy scale by cross-calibrating overlapping data sets, which is fixed by satellite and balloon experiments.

5. Acknowledgments

References

[1] D. Maurin, F. Melot and R. Taillet, *A database of charged cosmic rays*, *Astron. Astrophys.* **569** (2014) A32, [`1302.5525`].

[2] J. R. Hoerandel, *On the knee in the energy spectrum of cosmic rays*, *Astropart. Phys.* **19** (2003) 193–220, [`astro-ph/0210453`].

[3] T. K. Gaisser, T. Stanev and S. Tilav, *Cosmic Ray Energy Spectrum from Measurements of Air Showers*, *Front. Phys.(Beijing)* **8** (2013) 748–758, [`1303.3565`].

[4] T. K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, *Astroparticle Physics* **35** (July, 2012) 801–806.

[5] A. Fedynitch, H. Dembinski, R. Engel, T. Gaisser and T. Stanev, *State-of-the-art calculation of atmospheric lepton fluxes*, in *proceedings of ICRC 2017, Busan, Korea*, `PoS(ICRC2017)1019`, 2017.

[6] J. J. Engelmann, P. Ferrando, A. Soutoul, P. Goret and E. Juliusson, *Charge composition and energy spectra of cosmic-ray for elements from Be to NI - Results from HEAO-3-C2*, *Astron. Astrophys.* **233** (1990) 96–111.

[7] AMS collaboration, M. Aguilar et al., *Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **114** (2015) 171103.

[8] AMS collaboration, M. Aguilar et al., *Precision Measurement of the Helium Flux in Primary Cosmic Rays of Rigidities 1.9 GV to 3 TV with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **115** (2015) 211101.

[9] PAMELA collaboration, O. Adriani et al., *PAMELA Measurements of Cosmic-ray Proton and Helium Spectra*, *Science* **332** (2011) 69–72, [`1103.4055`].

[10] PAMELA collaboration, O. Adriani et al., *Time dependence of the proton flux measured by PAMELA during the July 2006 - December 2009 solar minimum*, *Astrophys. J.* **765** (2013) 91, [`1301.4108`].

[11] CREAM collaboration, H. S. Ahn et al., *Energy spectra of cosmic-ray nuclei at high energies*, *Astrophys. J.* **707** (2009) 593–603, [`0911.1889`].

[12] CREAM collaboration, Y. S. Yoon et al., *Cosmic-Ray Proton and Helium Spectra from the First CREAM Flight*, *Astrophys. J.* **728** (2011) 122, [`1102.2575`].

[13] C. de Boor, *A Practical Guide to Splines*. Springer Verlag, New York, Heidelberg, Berlin, 1978.

[14] R. J. Barlow, *Combining Experiments with Systematic Errors*, `1701.03701`.

[15] K.-H. Kampert and M. Unger, *Measurements of the Cosmic Ray Composition with Air Shower Experiments*, *Astropart. Phys.* **35** (2012) 660–678, [`1201.0018`].

[16] L. J. Gleeson and W. I. Axford, *Solar Modulation of Galactic Cosmic Rays*, *Astrophys. J.* **154** (1968) 1011.

[17] I. G. Usoskin, G. A. Bazilevskaya and G. A. Kovaltsov, *Solar modulation parameter for cosmic rays since 1936 reconstructed from ground-based neutron monitors and ionization chambers*, *Journal of Geophysical Research (Space Physics)* **116** (Feb., 2011) A02104.

[18] ARGO-YBJ collaboration, P. Montini and S. M. Mari, *The bending of the proton plus helium flux in primary cosmic rays measured by the ARGO-YBJ experiment in the energy range from 20 TeV to 5 PeV*, `1608.01389`.

[19] TUNKA collaboration, V. V. Prosin et al., *Tunka-133: Results of 3 year operation*, *Nucl. Instrum. Meth.* **A756** (2014) 94–101.

[20] E. E. Korosteleva, V. V. Prosin, L. A. Kuzmichev and G. Navarra, *Measurement of cosmic ray primary energy with the atmospheric Cherenkov light technique in extensive air showers*, *Nucl. Phys. Proc. Suppl.* **165** (2007) 74–80.

[21] ICECUBE collaboration, K. Rawlins and T. Feusels, *Latest Results on Cosmic Ray Spectrum and Composition from Three Years of IceTop and IceCube*, in *Proceedings of ICRC 2015, The Hague, Netherlands*, `Pos(ICRC2015)334`, 2016.

[22] KASCADE-GRANDE collaboration, W. D. Apel et al., *The spectrum of high-energy cosmic rays measured with KASCADE-Grande*, `1206.3834`.

[23] KASCADE-GRANDE collaboration, S. Schoo, *The energy spectrum of cosmic rays in the range from 10^{14} to 10^{18}eV*, in *Proceedings of ICRC 2015, The Hague, Netherlands*, `PoS(ICRC2015)263`, 2016.

[24] TELESCOPE ARRAY collaboration, D. Ivanov, *TA Spectrum Summary*, in *Pproceedings of ICRC 2015, The Hague, Netherlands*, `Pos(ICRC2015)349`, 2016.

[25] PIERRE AUGER collaboration, I. Valino, *The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory*, in *Proceedings of ICRC 2015, The Hague, Netherlands*, `PoS(ICRC2015)271`, 2016.

[26] PIERRE AUGER collaboration, A. Aab et al., *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications*, *Phys. Rev.* **D90** (2014) 122006, [`1409.5083`].

[27] B. Peters, *Primary Cosmic Radiation and Extensive Air Showers*, *Il Nuovo Cim. XXII* (1961) 800–819.