

## Combined analysis of accelerator and ultra-high energy cosmic ray data

---

**Sebastian Baur, Colin Baus\*, Anatoli Fedynitch, Igor Katkov, Tanguy Pierog, Ralf Ulrich, Hauke Wöhrmann**

*Karlsruhe Institute of Technology*

*E-mail:* [colin.baus@kit.edu](mailto:colin.baus@kit.edu)

The complexity of the development of extensive air showers makes it extremely difficult to study the nature and the sources of cosmic rays at ultra-high energies. The largest uncertainties are related to the modelling of hadronic interactions in the air shower cascade. The sensitivity to the theoretical description is maximised when measurements in the forward phase-space at accelerators are combined with measurements based on cosmic ray data. So far, this method has never been applied to constrain interaction models. Here, we outline a strategy to find an optimal global description of the underlying physics, ultimately leading to a consistent understanding of astrophysics as well as hadronic interaction physics at ultra-high energies.

*The 34th International Cosmic Ray Conference,  
30 July- 6 August, 2015  
The Hague, The Netherlands*

---

\*Speaker.

## 1. Introduction

Hadronic interaction models introduce large uncertainties on any cosmic ray measurement. In the era of the LHC at CERN a vast but incomplete set of measurements is available from multiple experiments in different phase-space regions. At the LHC proton-proton collisions are studied along with heavy ion collisions, i.e. proton-proton, proton-lead, and lead-lead, up to  $\sqrt{s_{\text{NN}}} = 13 \text{ TeV}$  or equivalently  $E_{\text{lab}} = 9 \times 10^{16} \text{ eV}$ .

Hadronic interaction models used to simulate extensive air showers (EAS) comprise tens of adjustable parameters. The models include them in the phenomenological description of the important soft interactions inaccessible to perturbative calculations that use the theory of strong interactions, quantum chromodynamics. One important aspect is that the models are foremost adjusted to describe collider data at central pseudorapidities,  $|\eta| < 2.5$ , where typically the best tracking and calorimetric information is available [1]. Forward measurements offer far more constraining data for the important parameters [2, 3]. It is evident that a systematic study of the agreement of the model predictions with forward measurements is needed.

Hadronic interaction models must also be able to describe EAS data and accurately predict hadronic interactions at laboratory energies three orders of magnitude higher than the currently available collider energies. Many observables derived from EAS are, in fact, sensitive to the underlying hadronic physics [4]. Even without knowing the exact mass composition of primary cosmic rays, the measured observables can be exploited to constrain model parameters at ultra-high energies. At the same time the relevant phase-space region is probed.

The parameter sets of models are optimised by studying the theory-data agreement. The combination of the RIVET and PROFESSOR toolkits [5, 6] are, for example, used to optimise multi-purpose generators at the LHC. The former program allows to compare to collider-based analyses based on information provided by the generated particles and the latter takes model predictions derived from a scan of the parameters as input and allows to run a  $\chi^2$ -test and minimisation.

Here, the performance is explored when EAS measurements are included, which is not possible with the above tools. Their limitation arises from the missing interface to simulate EAS with air shower simulation packages and from the inflexible data-theory comparison of the  $\chi^2$  method. To simultaneously study collider and air shower observables, the following programs are employed:

- CONEX [7]: A hybrid Monte Carlo (MC) program to simulate air showers in one dimension. It uses cascade equations for particles below a certain energy threshold. The simulations give access to longitudinal shower information, such as  $X_{\text{max}}$  and  $X_{\text{max}}^{\mu}$ .
- CRMC [8]: A MC program for collider-type events that can be simulated with any hadronic interaction models used in air showers.
- A new framework to adjust model parameters and simulate air showers and collider-type collisions for any given parameter set and compare to available data from both, LHC experiments and the Pierre Auger Observatory.

## 2. A combined analysis

The combined analysis presented here is based on a likelihood approach. In this way, it is

flexible enough to implement asymmetric and even piecewise-defined probability distributions for any measurement. Firstly, a set of three model parameters for an event generator is chosen. The focus lies on model parameters bearing influence on fragmentation, cross section calculations, and diffractive processes. Subsequently, a set of measurements that are sensitive to these parameters from collider experiments is selected. For each parameter phase-space point, the likelihood to describe each data set is calculated. For most measurements this is derived from Gaussian statistics since the exact likelihood function is unknown. Given the known standard deviations  $\sigma_+$  and  $\sigma_-$  of a measurement, an asymmetric piecewise Gaussian distribution is used to obtain the probability for a model prediction,  $y$ , to agree with the measured value,  $y_{\text{data}}$ :

$$P(y|y_{\text{data}}, \sigma_+, \sigma_-) = \begin{cases} c_1 \text{Gaus}(y_{\text{data}} - y, \sigma_+) & \text{for } y \geq y_{\text{data}} \\ c_2 \text{Gaus}(y_{\text{data}} - y, \sigma_-) & \text{for } y < y_{\text{data}}, \end{cases} \quad (2.1)$$

where the constants  $c_1$  and  $c_2$  are chosen such that  $\int_{-\infty}^{\infty} dy P(y|y_{\text{data}}, \sigma_+, \sigma_-) = 1$  and  $c_1 \text{Gaus}(0, \sigma_+) = c_2 \text{Gaus}(0, \sigma_-)$ . Obviously, for  $\sigma = \sigma_+ = \sigma_-$  the method is equivalent to the  $\chi^2$ -distribution. The only technical difference is that the probability,  $P_i$ , for each data point  $y_{\text{data},i}$  is  $P_i = 1/2$  in case the model deviates by  $1\sigma$  from the measurement. The logarithm of the likelihood can be represented by a sum and is here scaled to match the number of degrees of freedom (ndf):

$$\log(L) = \sum_i \log(P_i) / \text{ndf}. \quad (2.2)$$

To project the 3-dimensional parameter space into one dimension, the average likelihood is defined, which is the average over the two remaining dimensions that are not selected to study.

Assuming  $\mathcal{P}_{\text{sel}}$  to be the selected parameter of interest and the remaining parameters to be  $\mathcal{P}_{1,i}$  and  $\mathcal{P}_{2,j}$  with  $i$  and  $j$  representing the simulated points in the parameter space, the average likelihood is defined as

$$\langle -\log L(\mathcal{P}_{\text{sel}}) \rangle = \frac{1}{NM} \sum_i^N \sum_j^M -\log(L(\mathcal{P}_{\text{sel}}, \mathcal{P}_{1,i}, \mathcal{P}_{2,j})) \quad (2.3)$$

and the relative likelihood as

$$\text{relative likelihood}(\mathcal{P}_{\text{sel}}) = \langle -\log L(\mathcal{P}_{\text{sel}}) \rangle - \langle -\log L(\mathcal{P}_{\text{sel,default}}) \rangle, \quad (2.4)$$

where  $\mathcal{P}_{\text{sel,default}}$  is the default value of the fixed parameter.

## 2.1 Modification to the hadronic interaction model

Measurements by collider-based experiments constrain the parameters of the hadronic interaction model. This method will, in the following, be demonstrated with EPOS as the hadronic interaction model. It is one of the most widely used models to simulate high-energy collisions in air showers. The latest version, EPOS-LHC [9], is optimised to describe the first LHC data. It uses Gribov-Regge theory to calculate multiple scatterings of nucleons and nuclei with correct energy sharing between the collisions. Additionally, an effective hydrodynamic evolution of the quark gluon plasma is included and can, thus, realistically describe collisions of ions up to lead.

Three parameters that were modified compared to the previous version and that are directly linked to the cross-section evolution at high energies and diffractive processes are studied in this report. The selected parameters are listed with the default values highlighted in bold font:

- Name: “wdiff(2)”:  $w_{\text{diff}} = \{0.865, 0.870, \mathbf{0.875}, 0.880, 0.885\}$   
This parameter is related to the probability of simulating diffractive hadronic collisions. The “(2)” indicates that the behaviour is changed for proton interactions.
- Name “epscrx”:  $\epsilon_{\text{crx}} = \{0.25, 0.3, 0.35, \mathbf{0.4}, 0.45, 0.5\}$   
This parameter affects mainly the cross section extrapolation and multiplicity of particle production in an interaction. Modification of this parameter changes the saturation scale of the hard component of the scattering amplitude. Increasing the saturation scale dampens a too fast rise of the cross section.
- Name “alpdI”:  $\alpha_{\text{diff}} = \{0.2, 0.4, 0.725, \mathbf{1.05}, 1.35, 1.65\}$   
This parameter determines the shape of the remnant mass distribution for diffractive processes. The distribution follows  $1/M^{2\alpha_{\text{diff}}}$ . Therefore, for larger values of  $\alpha_{\text{diff}}$  more low-mass states are present, which influences in particular the forward multiplicity.

All three parameters bear large influence on the shower development since not only the height of the first interactions but also the momentum transfer between the interacting nucleons is changed.

## 2.2 Collider-based measurements

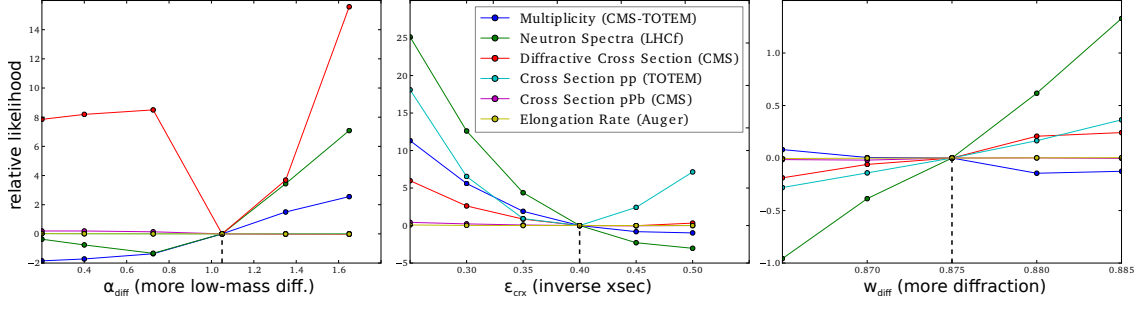
Multiple measurements mostly by the LHC experiments that are sensitive to the forward phase-space are selected. The predictions of the model for a specific parameter set is tested to fit the inelastic cross sections (“Cross Section pp” [10, 11] and “Cross Section pPb” [12]), the mass distribution of single and double diffractive cross sections (“Diffractive Cross Section” [13]), match spectra of forward neutrons (“Neutron Spectra” [14]), and describe the central and forward multiplicity for three different event selections (“Multiplicity” [15]). The centre-of-mass energies per nucleon pair extends up to 8 TeV.

The additionally shown analysis “Elongation Rate” is derived from EAS observables and its calculation of the relative likelihood is treated specially due to the nature of the mass composition of cosmic rays. The data [16] are extended by the low-energy measurement with the HEAT telescopes.

The relative likelihood is presented for each analysis individually in Fig. 1. A value below 0 indicates an improvement compared to the default parameter values. In many cases the default values are at the minimum of the relative likelihood function. Some tension exists between various analyses, e.g., for  $\epsilon_{\text{crx}}$  larger than the default value. The parameter  $\epsilon_{\text{crx}}$  is at the same time the most constrained one by the measurements. The default values of EPOS-LHC are already optimised to match the proton-proton cross section measurement at  $\sqrt{s} = 7\text{ TeV}$ . Concerning the two diffractive parameters,  $\alpha_{\text{diff}}$  is by far more important than  $w_{\text{diff}}$  in the sampled parameters range. The “Diffractive Cross Section” measurement favours the new default value for  $\alpha_{\text{diff}} = 1.05$  over the value used in the previous version (EPOS1.99),  $\alpha_{\text{diff}} = 0.45$ , for which the relative likelihood is high with a value of about 8.

## 2.3 Technical requirements for including observables of extensive air showers

Simulating EAS requires the generation of a large number of individual collisions of primary and secondary particles with air nuclei in the atmosphere. With 1D-showers simulated with CONEX, the time consumption can be reduced by about a factor of 100 compared to 3D-calculations even



**Figure 1:** Sensitivity of the likelihood for the EPOS-LHC model to changes in one parameter ( $\mathcal{P}_{\text{sel}}$ ) while averaging over the parameter grid in the other dimensions. The y-axis shows the relative likelihood as defined in (2.4). The default values are marked with a dashed vertical line.

when “thinning” is switched on. With the settings used in this report a shower induced by an iron primary with energy  $E_0 = 10^{19}$  eV takes about 30 CPU minutes to compute on a modern Intel Xeon CPU at 2.6 GHz. The spatial information for particles arriving at ground is lost but would be needed to compare to many measurements of the surface detector of ultra-high energy cosmic ray experiments. Without time-intensive 3D-simulations, comparison to, e.g., the muon production depth (MPD) measured for muons that arrive at a certain distance from the shower core is inaccessible.

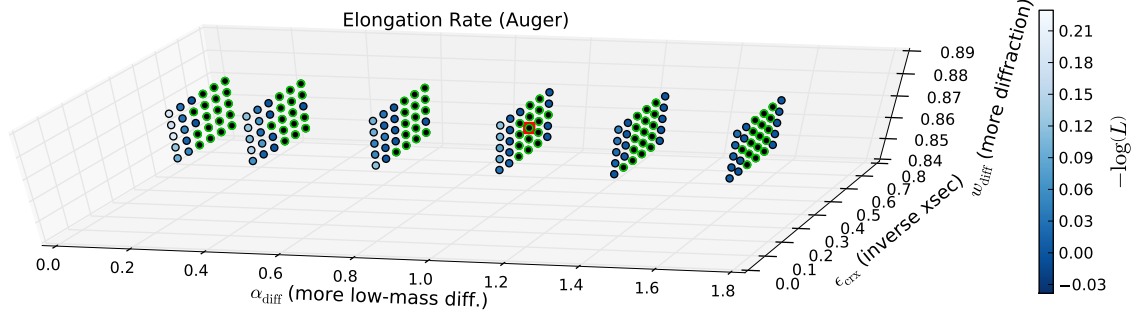
For each parameter set (in this report 180 phase-space points are used) the tables required by the interaction model are regenerated. The tables contain, e.g., cross section values at different energies. Here, the tables in `epos.inirj` and `epos.inics` take about 1 CPU hour to produce. Furthermore in the air shower simulation, the energy deposit for the particle types of  $p$ ,  $n$ ,  $\pi$ ,  $K^\pm$ ,  $K_L^0$  for 24 logarithmic energy bins from  $10^{10.5}$  eV up to  $10^{16}$  eV with 20,000 events in each bin is calculated to use as input for the cascade equations. The threshold to switch to cascade equations is uniformly for all particle type set to  $0.001 E_0$ , enabling simulations up to  $E_0 = 10^{19}$  eV.

For LHC-type collisions, simulation of sufficient statistics does not require a long simulation time but large storage space. Contrary to simulation of EAS, which are saved in the well compressed ROOT file format, `zlib`-compressed `HepMC` event records [17] are saved. Altogether large amounts of both storage and CPU time are required.

## 2.4 Constraints from astroparticle physics on physics of hadronic interactions

Until now constraints from astroparticle physics have not been successfully applied to a systematic study of model parameters. The technical reasons that arise when relying on air shower measurements are expressed above and can nowadays be overcome with large computer clusters. On top, conceptual difficulties to extract information about hadronic physics exist. In this report solutions to overcome these difficulties are presented and put to a first test by studying the shower profile of electromagnetic particles and of muons.

For ultra-high energy cosmic rays the mass composition of the particles arriving at Earth is not well known. Observables like the mean depth of the shower maximum,  $\langle X_{\text{max}} \rangle$ , and the muon production depth,  $\langle X_{\text{max}}^\mu \rangle$ , are dependent on the primary mass and become smaller with increasing



**Figure 2:** Comparison of the EPOS-LHC predictions of all 180 probed parameter sets to data from the  $X_{\max}$  measurement of the Pierre Auger Observatory. The colour code of each marker represents the likelihood of data-MC agreement. Darker blue denotes a better agreement whereas the light blue shows stronger disagreement. The parameters with the smallest  $-\log(L)$  values are marked with green borders and the default parameter set is drawn with a red border.

mass of the nucleus. Two ways are explored to derive constraints. For the shower maximum the data is compared to the two extreme cases, proton and iron primaries.

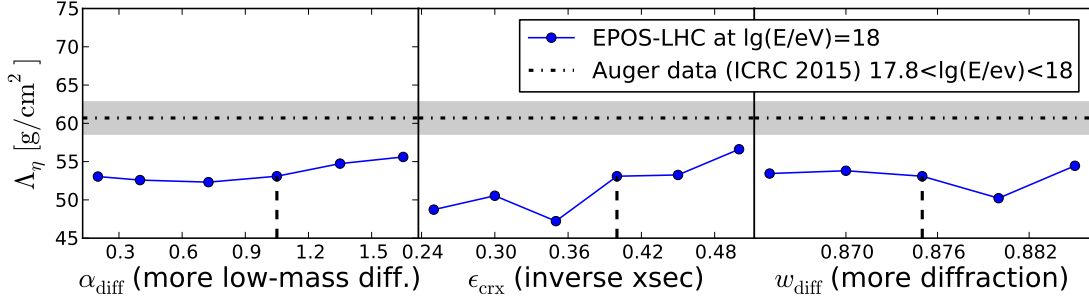
A straight-forward implementation uses a flat likelihood function when the data is bracketed by the model predictions made for proton,  $X_{\max}^p$ , and iron primaries,  $X_{\max}^{\text{Fe}}$ . However, when the data are outside this range, the likelihood drops with Gaussian tails that are determined by the uncertainties of the measurement. For the measured value  $X_{\max} \pm \sigma$  with the corresponding prediction derived from iron and proton simulations  $X_{\max}^{\text{Fe/p}}$ , the probability is defined as

$$P(X_{\max}^{\text{Fe/p}} | X_{\max}, \sigma) = \begin{cases} c_1 \text{Gaus}(X_{\max} + \sigma - X_{\max}^p, \sigma) & \text{for } X_{\max} \geq X_{\max}^p - \sigma \\ c_2 & \text{for } X_{\max}^{\text{Fe}} + \sigma < X_{\max} < X_{\max}^p - \sigma \\ c_1 \text{Gaus}(X_{\max} - \sigma - X_{\max}^{\text{Fe}}, \sigma) & \text{for } X_{\max} \leq X_{\max}^{\text{Fe}} + \sigma. \end{cases}$$

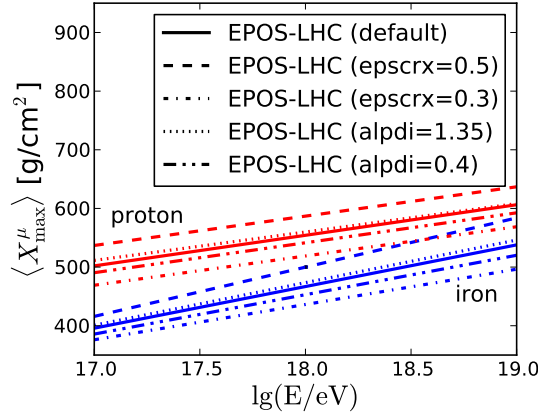
The normalisation constants  $c_1$  and  $c_2$  are chosen such that the distribution is continuous and its integral is unity. Furthermore, a prior distribution must be multiplied that compensates the dependence of  $c_2$  on the size of the difference  $X_{\max}^p - X_{\max}^{\text{Fe}}$ . This is achieved by dividing by  $c_2$  as a last step.

The result of the procedure is shown in Figs. 1 and 2. Especially the latter figure demonstrates well that all three parameters can be constrained by this method. However, only the values at the edge of the chosen parameter range have a decreased likelihood.

The second way is to compare the  $X_{\max}$  values of the 20% most deeply penetrating showers as is done for deriving the p-Air cross section for the data of the Pierre Auger Observatory [18]. The tail of the  $X_{\max}$  distribution is dominated by protons with only a slight contamination by helium nuclei. Since the contamination is estimated by comparison to model predictions, the helium fraction must either be treated as a free parameter or—as in the approach used here—the slope of the exponential tail,  $\Lambda_\eta$ , is derived from combined  $X_{\max}$  values of proton and helium primaries. The two distributions are weighted by the fitted proton ( $f_p$ ) and the helium ( $1 - f_p$ ) fractions, respectively. The method has a slight bias towards higher likelihood values but the proton fraction is large



**Figure 3:** The slope of the exponentially dropping distribution of the 20% deepest penetrating showers,  $\Lambda_\eta$ , derived for modified values of  $\alpha_{\text{diff}}$  (left),  $\epsilon_{\text{crx}}$  (middle), and  $w_{\text{diff}}$  (right) while keeping the other parameters at their respective default value. The default values are marked with dashed vertical lines.



**Figure 4:**  $\langle X_{\text{max}}^\mu \rangle$  over the logarithm of the energy of the primary cosmic ray. The predictions for protons (red) and iron (blue) are shown for modified parameters,  $\alpha_{\text{diff}}$  and  $\epsilon_{\text{crx}}$ , while the default parameter set is shown as a solid line.

for all parameter values ( $f_p \gtrsim 90\%$ ). The detector resolution is neglected, which is also expected to render only a small ( $< 0.5 \text{ g cm}^{-2}$ ) effect on  $\Lambda_\eta$ . Both effects are smaller than the uncertainty of about 4% due to the low MC statistics in the energy bin at  $10^{18}$  eV. The derived values for  $\Lambda_\eta$  are shown in Fig. 3 and the strongest dependence comes from the saturation scale  $\epsilon_{\text{crx}}$ . A larger saturation scale entails a smaller cross section and, thus, a larger value of  $\Lambda_\eta$ . The dependence on the other parameters is less clear and needs larger statistics to be resolved.

Another interesting observable to be studied is  $\langle X_{\text{max}}^\mu \rangle$ . The data reported in [19] reveal tension with the simulations done with the EPOS-LHC model. With the 1D simulations no direct comparison is possible but the data prefer a smaller increase with energy. The parameters tested in this report clearly change the shape of  $\langle X_{\text{max}}^\mu \rangle$  (see Fig. 4) but 3D-air shower simulations are required to conclude on the matter.



### 3. Summary

For the first time a combined analysis was presented that tests hadronic interaction models using air shower and collider data simultaneously. A framework has been set up to manage the demands on the computer cluster, the generation of collider-type and EAS-type events, and the comparison of theory predictions with data. It could be shown that priors from astroparticle physics can be used to constrain parameters of hadronic interaction models. The sensitivity of the measurements to three parameters has been shown.

The combined analysis was tested with three parameters of the hadronic interaction model EPOS-LHC, which are in general well tuned. Some measurements prefer a slight deviation from the default settings. However, the optimal value of one measurement is sometimes in tension with other measurements.

The framework has a huge potential to help with optimisation of model parameters. The tool greatly reduces manual labour in the parameter optimisation process but the relevance of the physics processes connected to each parameter must not be neglected when using it. In the future, more analyses need to be added. A Markov Chain-based search for an optimal parameter set is currently under scrutiny.

#### Acknowledgment

This work is supported in part by the Helmholtz Young Investigator Group HHNG VH-NG-733 and the Helmholtz Alliance for Astroparticle Physics HAP, which are both funded by the Initiative and Networking Fund of the Helmholtz Association. We also would like to thank Ralph Engel, Michael Unger, Nobuyuki Sakurai, and the rest of the LHCf group at SteLab, Nagoya, for fruitful discussions.

#### References

- [1] D. d'Enterria, R. Engel, T. Pierog, S. Ostapchenko, and K. Werner, *Constraints From the First Lhc Data on Hadronic Event Generators for Ultra-High Energy Cosmic-Ray Physics*, *Astropart. Phys.* **35** (2011) 98, [[1101.5596](#)].
- [2] **Pierre Auger and Yakutsk** Collaboration, J. Allen et al., *Air Shower Simulation and Hadronic Interactions*, *EPJ Web Conf.* **53** (2013) 01007, [[1306.6090](#)].
- [3] C. Baus, R. Engel, and R. Ulrich, *Relation between Hadronic Interactions and Air Showers*, *Proceedings of the 2014 ISVHECRI (to be published)* (2014).
- [4] T. Pierog, *LHC Data and Extensive Air Showers*, *EPJ Web Conf.* **52** (2013) 03001.
- [5] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, et al., *Rivet User Manual*, *Comput. Phys. Commun.* **184** (2013) 2803, [[1003.0694](#)].
- [6] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, *Systematic Event Generator Tuning for the LHC*, *Eur. Phys. J. C* **65** (2010) 331, [[0907.2973](#)].
- [7] T. Bergmann, R. Engel, D. Heck, N. Kalmykov, S. Ostapchenko, et al., *One-Dimensional Hybrid Approach to Extensive Air Shower Simulation*, *Astropart. Phys.* **26** (2007) 420, [[astro-ph/0606564](#)].
- [8] <https://web.ikp.kit.edu/rulrich/crmc.html>.



- [9] T. Pierog, I. Karpenko, J. Katzy, E. Yatsenko, and K. Werner, *EPOS LHC: Test of Collective Hadronization with LHC Data*, e-print (2013) [[1306.0121](#)].
- [10] **TOTEM** Collaboration, G. Antchev et al., *Measurement of Proton-Proton Inelastic Scattering Cross-Section at  $\sqrt{s} = 7$  TeV*, *Europhys. Lett.* **101** (2013) 21003.
- [11] **TOTEM** Collaboration, G. Antchev et al., *Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at  $\sqrt{s} = 8$  TeV*, *Phys. Rev. Lett.* **111** (2013), no. 1 012001.
- [12] **CMS** Collaboration, *Measurement of the Inelastic pPb Cross Section at  $\sqrt{s_{NN}} = 5.02$  TeV*, CMS Physics Analysis Summary CMS-PAS-FSQ-13-006, CERN, Geneva, 2013.
- [13] **CMS** Collaboration, V. Khachatryan et al., *Measurement of Diffraction Dissociation Cross Sections in pp Collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **92** (2015), no. 1 012003, [[1503.08689](#)].
- [14] K. Kawade, *Measurement of Neutron Production in the Very Forward Rapidity at LHC  $\sqrt{s} = 7$  TeV pp Collision*. PhD thesis, Nagoya University, Nagoya, 2014.
- [15] **CMS+TOTEM** Collaboration, S. Chatrchyan et al., *Measurement of Pseudorapidity Distributions of Charged Particles in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV by the CMS and TOTEM Experiments*, *Eur. Phys. J. C* **74** (2014), no. 10 3053, [[1405.0722](#)].
- [16] **Pierre Auger** Collaboration, A. Aab et al., *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above  $10^{17.8}$  eV*, *Phys. Rev. D* (2014) [[1409.4809](#)].
- [17] M. Dobbs and J. B. Hansen, *The HepMC C++ Monte Carlo Event Record for High Energy Physics*, *Comput. Phys. Commun.* **134** (2001) 41.
- [18] **Pierre Auger** Collaboration, A. Aab et al., *Extension of the Measurement of the Proton-Air Cross Section with the Pierre Auger Observatory*, *This conference* (2015).
- [19] **Pierre Auger** Collaboration, A. Aab et al., *Muons in Air Showers at the Pierre Auger Observatory: Measurement of Atmospheric Production Depth*, *Phys. Rev. D* **90** (2014) 012012, [[1407.5919](#)].