Study of Monte Carlo event generators for proton-proton collisions at LHC energies in the forward region*

Alexandru Cătălin ENE[1,2;1)] Alexandru JIPA[2] Lavinia-Elena GIUBEGA[1]

[1]Department of Elementary Particles Physics, Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering (IFIN-HH), Reactorului 30, RO- 077125, P.O.B. MG-6, Măgurele-Bucharest, Romania
[2]Faculty of Physics of the University of Bucharest, Atomiştilor 405, RO-077125, P.O.B. MG-11, Măgurele-Bucharest, Romania

Abstract: In this paper we present a comparative study between PYTHIA, EPOS, QGSJET, and SIBYLL generators. The global event observables considered are the charged energy flow, charged particle distributions, charged hadron production ratios and V^0 ratios. The study is performed in the LHCb and TOTEM fiducial phase spaces on minimum bias simulated data samples for pp collisions at $\sqrt{s}$ = 7 TeV , using the reference measurements from these experiments. In the majority of cases, the measurements are within a band defined by the most extreme predictions. The observed differences between the predictions and the measurements seem to be, in most part, caused by extrapolation from the central pseudorapidity region ($|\eta|\leqslant 2.5$), in which the generators were mainly tuned.

Keywords: phenomenological models, event generators, LHCb

PACS: 12.38.Bx, 12.39.-x, 12.40.N **DOI:** 10.1088/1674-1137/43/8/083001

1 Introduction

One of the most important sources of information concerning elementary particle physics is the study of high energy cosmic rays. Up until the advent of powerful particle accelerators in the 1950s, the only source of high energy particles were the cosmic rays. The cosmic ray spectrum reaches energies of the order of 10^{20} eV [1], whilst the most powerful collider to date, the Large Hadron Collider, reaches energies of 13 TeV in the center of mass frame or about 10^{17} eV fixed target equivalent. So, there are two independent sources of information for pp collisions at the same energy scale. Combining the two helps create a better picture of the phenomena that take place in such collisions. Although the cross-section of hard interactions is considerable at these energy scales, the soft interaction part is still large. As soft processes imply non-perturbative QCD, we rely on phenomenological models and effective theories for predictions. Hadronic interaction generators have been developed for the description of the physics at the aforementioned energy scales, with an emphasis on either cosmic rays or collider physics. In recent years, cosmic rays generators have been extensively tuned to collider physics measurements, especially in the context of the newly available data from the LHC. In this paper we compare the predictions obtained with EPOS LHC [2], QGSJETII-04 [3] and SIBYLL 2.3 [4] generators, included in the CRMC package [5] , and the widely used event generator for the LHC physics PYTHIA (versions 8.186 [6] and 8.219 [7]), for pp interactions at $\sqrt{s}$ = 7 TeV , with the measurements from the LHCb and TOTEM experiments. The generators studied are all tuned using various observables measured by the LHC experiments. Predictions obtained with PYTHIA 8.186 using the non-LHC tune 2M are also shown for reference. Throughout this paper we refer to the measurements/tunes performed in the "central" and "forward" regions, defined with respect to the pseudorapidity of the particles. The central pseudorapidity region is defined as $|\eta| \leqslant 2.5$, corresponding to the ATLAS, ALICE and CMS acceptances [8-10], and the forward pseudorapidity region as $\eta \geqslant 2.5$, corresponding to the LHCb ($2 \leqslant \eta \leqslant 5$) and TOTEM ($3.1 \leqslant |\eta| \leqslant 6.5$) acceptances [11, 12].

Received 27 January 2019, Revised 28 March 2019, Published online 26 June 2019

* Supported by the following projects: Proiect IDEI (56/07.10.2011) and the IFA project România-CERN (7/16.03.2016) (LHCb-Ro)

1) E-mail: alexandru.catalin.ene@cern.ch

2 The Monte Carlo event generators

2.1 General description

The generators used for this study are PYTHIA, a collider physics generator, and EPOS, QGSJET and SIBYLL, which are cosmic ray collision generators. They can be split in three categories according to the models on which they are based. PYTHIA is a parton based generator and it simulates parton interactions and parton showers, where the hadronization is treated using the Lund string fragmentation model [13, 14]. Another category are the generators based on the Regge theory, such as QGSJET and SIBYLL. These models treat soft and semi-hard interactions as Pomeron exchanges ("soft" and "semi-hard" Pomerons), but also mix perturbative methods in the treatment of hard interactions [14, 15]. EPOS is in a distinct category in which the parton based description is mixed with aspects from the Regge theory [14]. The focus of this study is on minimum bias physics measurements and the generators used, especially the cosmic ray ones, are developed for the description of such observables. The selection of these particular generators is justified by their varied usage and basic assumptions, while at the same time sharing similarities, as well as having been tuned to the LHC data, as will be discussed below.

PYTHIA is one of the most widely used Monte Carlo event generators for collider physics with an emphasis on *pp* interactions. It is mainly based on the Leading Order (LO) QCD, having implemented LO matrix elements and usually using LO PDF sets (NLO PDF sets are also available) [7, 16, 17]. The main event in a *pp* collision (internally called "hard process") can be represented by a plethora of processes like elastic and diffractive (described using Pomerons) [7, 13, 18], soft and hard QCD processes, electroweak processes, top quark production etc. The generator also implements parton showers (Initial State Radiation, ISR, and Final State Radiation, FSR) in Leading Log (LL) approximation with matching and merging methods between them and the hard processes [7, 16]. Given that the colliding hadrons have a complex partonic structure, other partonic interactions aside from the main event are expected. These are called multi-parton interactions (MPI) and are usually soft in nature, but the momentum transfer can also reach the hard interaction energy scale. PYTHIA implements a description of both types and also of the beam remnants which form after the extraction of MPI initiator partons [7]. The hadronization mechanism is based on the Lund string fragmentation model [7].

The Parton-Based Gribov-Regge Theory is an effective field theory using concepts from QCD in which the elementary interactions between the constituent partons of nucleons/nuclei proceed via exchanges of parametrized objects called Pomerons, which have the quantum numbers of vacuum [19, 20]. In this theory the elementary collisions are treated as a sum of soft, semi-hard and hard contributions. If one considers a cutoff value of the momentum transfer squared of $Q_0^2 \sim 1$ GeV^2, below which perturbative QCD calculations can no longer be done, then the soft contribution (non-perturbative) is represented by processes with $Q^2 < Q_0^2$ and the hard contribution (perturbative) by processes with $Q^2 > Q_0^2$. The processes in which sea partons with $x \ll 1$ (Björken x) are involved are called semi-hard and are represented by a parton ladder with soft Pomeron ends [19].

The generator EPOS is based on the effective theory described above [2]. EPOS is an acronym for **E**nergy conserving quantum mechanical approach, based on **P**artons, parton ladders, strings, **O**ff-shell remnants, and **S**plitting of parton ladders [21]. In EPOS, the interaction of the two beam particles is described by means of Pomeron exchanges. As discussed above, these Pomerons can be soft, semi-hard or hard. A soft Pomeron can be viewed from a phenomenological standpoint as two parton ladders (or cut Pomeron) connected to the remnants by two color singlets (legs) from the parton sea [22]. A cut Pomeron can be viewed as two strings which fragment to create hadrons. The flavors of the string ends need to be compensated within the remnants. Thus, particle production in EPOS comes from two sources, namely cut Pomerons and the decay of remnants [22]. Through a recent development (from EPOS 1.99 onwards), EPOS is now a core-corona model. The core represents a region with a high density of string segments that is larger than some critical density for which the hadronization is treated collectively, and the corona is the region with a lower density of string segments for which the hadronization is treated non-collectively. The strings from the core region form clusters which expand collectively. This expansion has two components, namely radial and longitudinal flow. Through this core-corona approach, EPOS takes into account effects not accounted for in other HEP models [2]. In EPOS, in the case of multiple scatterings (multi-Pomeron exchanges), the energy scales of the individual scatterings are taken into account when calculating the respective cross-sections, while in the other models based on the Gribov-Regge theory this is not the case. This leads to a consistent treatment of both exclusive particle production and cross-section calculation, taking energy conservation into account in both cases [19, 22]. The multiplicity and inelastic cross-section predictions of the model are directly influenced by energy-momentum sharing and beam remnant treatment [22].

The elementary scatterings in QGSJET are also treated as Pomeron exchanges [15]. QGSJET is based on the Quark-Gluon string model, which is in turn based on

the Gribov-Regge model [23]. In this model the Pomeron exchange can be viewed as an exchange of a non-perturbative gluon pair. Each of the colliding protons can be considered as a system of a quark and a diquark with opposite transverse momenta. The quark from the first proton exchanges a non-perturbative gluon with the diquark from the second proton and viceversa, thus creating two quark-gluon strings which will decay according to the fragmentation functions to create hadrons [24]. In a similar manner to EPOS, the soft (non-perturbative) and hard (perturbative) contributions are separated by a cutoff value of Q_0^2. In QGSJET, a Pomeron is actually a sum of two contributions: a "soft" Pomeron and a "semi-hard" Pomeron contribution. The soft part represents a purely non-perturbative parton cascade, while the "semi-hard" Pomeron can be viewed as two "soft" Pomerons connected by a parton ladder [25]. At very high energies as those at the LHC and/or small impact parameters, the semi-hard contribution dominates and so it is crucial to take it into account [15, 23]. In these high energy collisions a large number of parton-parton interactions occur, the resulting cascades interacting with one another (Pomeron-Pomeron interactions), and thus their evolution is no longer independent but correlated. QGSJET-II takes into account these non-linear effects, which are computed with enhanced Pomeron diagrams [15, 23].

SIBYLL is based on the dual parton model (DPM), using the mini-jet model for hard interactions and the Lund string fragmentation model for hadronization [26, 27]. Similarly to both EPOS and QGJSET, soft and hard interactions are separated by a transverse momentum scale cutoff value. The soft interactions are treated using the dual parton model (DPM) in which the nucleon is treated as consisting of a quark and a diquark, and similar to the Quark-Gluon string model described above, a quark (diquark) from the projectile combines with the diquark (quark) from the target to form two strings which are fragmented separately using the Lund string fragmentation model. In SIBYLL 1.7, the cutoff value was set to $p_T^{\min} = \sqrt{5}$ GeV, but from version 2.1 onwards it was changed to a function of the collision energy, which for $\sqrt{s} = 7$ TeV returns $p_T^{\min} \approx 3.87$ GeV [26].

2.2 Versions used in the study

The default tune for PYTHIA 8.186 is Tune 4C with the CTEQ6L1 LO PDF set as the default one [7, 28]. Tune 4C (default from version 8.150 onwards [29]) is obtained starting from Tune 2C for which the Tevatron data have been used, by varying MPI and color reconnection parameters to fit the measurements for minimum bias (MB) and underlying event (UE) observables from the ALICE and ATLAS experiments at various collision energies (0.90, 2.36 and 7 TeV). The observables used are for example: charged multiplicity and rapidity distributions, transverse momentum distributions, mean transverse momentum distributions as a function of charged multiplicity, transverse momentum sum densities etc. Tune 2M is obtained in a similar manner to 2C, using the measurements from the CDF experiment at Tevatron, but uses the modified PDF set MRST LO** instead of the CTEQ6L1 LO PDF set [30]. From here on, PYTHIA 8.186 with Tune 2M will be refered to as PYTHIA 8.1 2M.

PYTHIA 8.219 has the Monash 2013 tune as its default (with the NNPDF3.3 QCD+QED LO PDF set) [7, 29]. The Monash 2013 tune has been created for a better description of minimum bias and underlying event observables. Similar observables as for the previous tune have been used, with the measurements from the ATLAS and CMS experiments, and the charged pseudorapidity distribution from TOTEM in the forward region. The flavor-selection parameters of the string fragmentation model have been re-tuned using a combination of data from PDG and from the LEP experiments, resulting in an overall increase of about 10% in strangeness production, and a similar decrease of the production of vector mesons. The kaon yields have clearly improved with respect to the CMS measurements, and the yields of hyperons are also slightly improved. The minimum bias charged multiplicity has also increased by about 10% in the forward region [31].

EPOS LHC fundamental parameters are tuned to the cross-section measurements from the TOTEM experiment at $\sqrt{s} = 7$ TeV, leading to a highly improved description of charged multiplicity (compared to EPOS 1.99). In EPOS LHC, the radial flow calculations are corrected. This correction affects the high multiplicity region, again leading to a highly improved description of this observable in this particular region. In EPOS 1.99, the baryon-antibaryon pair and strangeness production were largely overestimated in high energy collisions. This issue was corrected in EPOS LHC, and by using the same string fragmentation parameters as for e^+e^- collisions, kaon/pion and proton/pion ratio measurements from CMS at $\sqrt{s} = 7$ TeV are reasonably well described [2]. The statistical particle production mechanism from the core affects strangeness production by removing its suppression. This leads to a good description of the strange baryon yield measurements from CMS at $\sqrt{s} = 7$ TeV, as shown in Figure 10 from [2]. The radial flow parameters are tuned using charged particle transverse momentum distributions (for minimum bias pp collisions) obtained by the ATLAS experiment at $\sqrt{s} = 0.9$ and 7 TeV. This leads to a very good agreement with the experimental transverse momentum distributions of identified particles [2].

QGSJETII-04 is distinguished from the previous version, QGSJETII-03, by taking into account all significant enhanced Pomeron diagram contributions, including

Pomeron loops, and the tuning to the new LHC data [32]. As QGSJET is used for high energy cosmic rays studies, the current version of the generator has been tuned to the LHC measurements for observables to which the extensive air shower (EAS) muon content is sensitive. Examples of such observables are: charged particle multiplicities and densities, anti-proton and strange particle yields etc. QGSJETII-03 predicts a steeper increase in multiplicity in pseudorapidity plots from $\sqrt{s}$ = 0.9 to 7 TeV than what is observed in the ATLAS measurements for these collision energies. As a consequence, the Q_0^2 separation scale between soft and hard interactions has been increased from 2.5 GeV^2 to 3.0 GeV^2. For a better description of the ALICE measurements of the antiproton transverse momentum spectrum at $\sqrt{s}$ = 0.9 TeV, the anti-nucleon yield was slightly reduced and the hadronization parameters have been modified so as to enlarge the average transverse momentum of the anti-nucleons. The strangeness production has been enhanced to better describe K_S^0 and Λ rapidity distributions measured by CMS for $\sqrt{s}$ = 0.9 TeV and 7 TeV pp collisions. Another major tuning was done using the inelastic cross-section measurements at $\sqrt{s}$ = 7 TeV from the TOTEM experiment [33].

SIBYLL is a relatively simpler model and emphasis is put on describing observables on which the evolution of extensive air showers depends, like energy flow and particle production in the forward region [34]. In SIBYLL 2.3 soft gluons can also be exchanged between sea quarks or sea and valence quarks. A new feature in version 2.3 is the beam remnant treatment which is similar to that in QGSJET. This new treatment allows particle production in the forward region to be tuned without modifying the string fragmentation parameters. A major tuning procedure has been done for the description of leading particle measurements from the NA22 and NA49 experiments [4]. SIBYLL 2.3 has also been tuned using measurements from $\sqrt{s}$ = 7 TeV pp collisions from LHC experiments, namely the inelastic cross-section from TOTEM, and the average antiproton multiplicities and charged particle differential cross-sections as a function of transverse momentum obtained by CMS. The version SIBYLL 2.1 was tuned using the Tevatron data, and it describes, for example, charged pseudorapidity density measurements reasonably well, even from CMS at $\sqrt{s}$ = 7 TeV, as can be seen in Figure 4 from [35]. At the same time, SIBYLL 2.1 overestimates the inelastic cross-section measurements at high collision energies (beyond 1 TeV), leading to the tuning of version 2.3 with the $\sigma_{pp}^{\rm inel}$ measurements at $\sqrt{s}$ = 7 TeV from TOTEM. The antiproton multiplicities measured in fixed target experiments at low collision energies seem to be reasonably well described by version 2.1, but the measurements obtained by the CMS experiment for various collision energies are largely underestimated. To correct this effect in SIBYLL 2.3, a different value of the quark/diquark production probability, $P_{q/qq}$, has been assigned for the fragmentation of mini-jets than for all the other fragmentation processes. The value of $P_{q/qq}$ in SIBYLL 2.1 was fixed to 0.04 for all processes. SIBYLL 2.3 uses the same effective parton density function as the previous version, but the quark and gluon contributions are obtained from the same parametrizations used to calculate the mini-jet cross-section. This leads to a steeper parton distribution function for low Björken x , which combined with the correction of the definition of $p_T^{\rm min}$, leads in turn to a better description of the measurements for charged particle cross-sections as a function of transverse momentum obtained by CMS in the range $2 \leqslant p_T \leqslant 5$ GeV/c . Also, a charm hadron production model was implemented in version 2.3 [35].

3 Data generation and analysis strategy

Samples of 10^6 inelastic minimum bias pp events at $\sqrt{s}$ = 7 TeV were generated for each generator. For all generators, a stable particle definition of $c\tau \geqslant 3$ m was used, where τ is the mean proper lifetime of the particle species.

This study treats five distinct aspects: charged energy flow, charged particle distributions, charged hadron production ratios and V^0 ratios.

Charged energy flow is computed as the total energy of stable charged particles (p, $\bar{p}$, $K^{\pm}$, $\pi^{\pm}$, $\mu^{\pm}$ and $e^{\pm}$) in the interval $1.9 \leqslant \eta \leqslant 4.9$ (10 bins of $\Delta\eta = 0.3$), divided by the width of the pseudorapidity bin and normalized to the number of visible inelastic pp interactions $N_{\rm int}$ or:

$$\frac{1}{N_{\rm int}}\frac{{\rm d}E_{\rm total}}{{\rm d}\eta} = \frac{1}{\Delta\eta}\left(\frac{1}{N_{\rm int}}\sum_{i=1}^{N_{{\rm part},\eta}} E_{i,\eta}\right), \quad (1)$$

where $N_{{\rm part},\eta}$ is the number of stable charged particles (as defined above) in a $\Delta\eta = 0.3$ bin and $E_{i,\eta}$ is the energy of the particles from the respective bin (see [36]).

There are four event classes considered for the charged energy flow: inclusive minimum bias events, hard scattering events, diffractive enriched events and non-diffractive enriched events. The inclusive minimum bias events are required to have at least one charged particle in the range: $1.9 \leqslant \eta \leqslant 4.9$. The hard scattering events require at least one charged particle with $p_T \geqslant 3$ GeV/c in the aforementioned range. Diffractive enriched events require that no particles are generated in the pseudorapidity range of $-3.5 < \eta < -1.5$ and non-diffractive enriched events require at least one particle in this range. These event class definitions are compatible with the ones from [36], from which the LHCb reference measurements were taken.

The purity of the diffractive enriched and non-diffractive enriched events samples have been studied for both versions of PYTHIA (as the generator has readily accessible event type information) and are about 94% and 92%, respectively. In Fig. 1, the transverse momentum scale distributions of the hardest parton collisions from hard and soft (non-hard and non-diffractive) events, obtained with PYTHIA 8.186, are shown. As can be seen, the peaks are reasonably well separated with $\mu \approx 8.7$ GeV/c, $\sigma \approx 4.5$ GeV/c, for hard events and $\mu \approx 4.2$ GeV/c, $\sigma \approx 3.2$ GeV/, for soft events. The fraction of events that pass both the hard and diffractive enriched event class conditions are negligible.

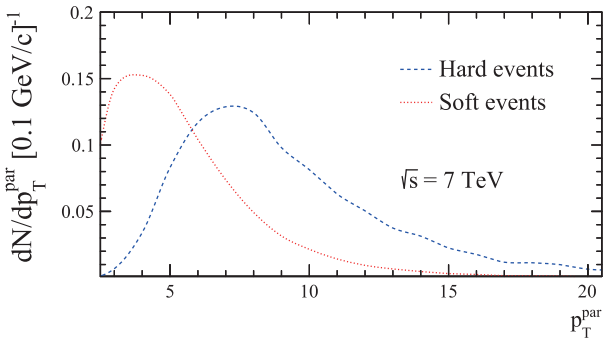

Fig. 1. (color online) Transverse momentum scale of the hardest subprocess obtained with PYTHIA 8.186 for hard and soft events. The distributions were normalized to the number of visible events for each event class.

The number of visible events for the different event classes are given in Table 1.

Table 1. Number of visible events for different event classes. N_{MB}, the number of visible minimum bias events, is expressed as a percentage of the total number of generated inelastic events $N_{\rm gen} = 10^6$. $N_{\rm hard}$ and $N_{\rm dif}$, the numbers of visible hard and diffractive events, respectively, are expressed as percentage of $N_{\rm MB}$.

Generator	$N_{\rm MB}$	$N_{\rm hard}$	$N_{\rm dif}$
PYTHIA 8.186	88.20%	5.63%	7.04%
PYTHIA 8.219	88.11%	5.05%	7.10%
EPOS LHC	84.92%	4.87%	6.26%
QGSJETII-04	86.72%	7.94%	5.52%
SIBYLL 2.3	89.55%	6.43%	6.47%
PYTHIA 8.1 2M	86.89%	5.08%	7.97%

The transverse momentum, pseudorapidity and multiplicity distributions of charged stable particles (p, π, K, e, μ) are presented in Figs. 3-6. The distributions were scaled with the number of visible events from the sample. The visible events are required to contain a minimum of one charged particle satisfying the criteria listed below:

- Figure 3: $2 < \eta < 4.8$, $p \geqslant 2$ GeV/c and $p_T > 0.2$ GeV/c [37].
- Figure 4: $2 < \eta < 4.5$ [38].
- Figure 5: $2.5 < \eta < 4.5$ and $p_T > 1$ GeV/c. These events are called "hard" [38].
- Figure 6: $5.3 < \eta < 6.5$ and $p_T > 40$ MeV/c [39].

The number of minimum bias and hard events with a minimum of one charged particle in the range $2 < \eta < 4.5$ are given in Table 2.

Table 2. Number of events with a minimum of $n_{ch} \geqslant 1$ in $2 < \eta < 4.5$, expressed as a percentage of the total number of generated inelastic events $N_{\rm gen} = 10^6$. Hard events require a minimum of one charged particle with $p_T \geqslant 1$ GeV/c in $2.5 < \eta < 4.5$.

Generator	minimum bias	hard events [% of minbias]
PYTHIA 8.186	87.28%	43.90%
PYTHIA 8.219	87.17%	42.83%
EPOS LHC	83.81%	44.86%
QGSJETII-04	85.57%	54.01%
SIBYLL 2.3	88.19%	46.68%
PYTHIA 8.1 2M	85.87%	37.37%

For all distributions mentioned above, pull plots of $(x_{\rm gen} - x_{\rm exp})/\sigma_{\rm exp}$ have been drawn.

A particle is defined as prompt if the sum of mean proper lifetimes of its ancestors is less than 10 ps, as in [37-39].

The prompt charged hadron production ratios $\bar{p}/p$, π^-/π^+, K^-/K^+, $(K^+ + K^-)/(\pi^+ + \pi^-)$, $(p + \bar{p})/(K^+ + K^-)$ and $(p + \bar{p})/(\pi^+ + \pi^-)$ are shown in Figs. 9-11 as a function of pseudorapidity. These ratios are computed in the phase space defined by $2.5 \leqslant \eta \leqslant 4.5$ and $p \geqslant 5$ GeV/c, and in three transverse momentum intervals, namely $p_T < 0.8$ GeV/c, $0.8 \leqslant p_T < 1.2$ GeV/c and $p_T \geqslant 1.2$ GeV/c [40].

The prompt V^0 particle ratios $\bar{\Lambda}/\Lambda$ and $\bar{\Lambda}/K^0_S$ as a function of rapidity are shown in Fig. 12. The ratios are computed in the phase space defined by $2 \leqslant y \leqslant 4.5$ and three p_T intervals: $0.15 < p_T < 0.65$ GeV/c, $0.65 < p_T < 1.00$ GeV/c and $1.00 < p_T < 2.50$ GeV/c. Figs. 13-14 show the prompt V^0 particle ratios as a function of rapidity and as a function of transverse momentum in the $2 \leqslant y \leqslant 4.5$ rapidity interval and the full p_T interval $0.15 < p_T < 2.50$ GeV/c [41].

The statistical uncertainties of the MC predictions are negligible, reaching a maximum of about 3% in the least populated bins at the edges of the considered phase space regions, while for the rest of the bins the uncertainties are of the order of 0.1%.

The sources of the reference measurements used in the plots are given at the end of the captions.

4 Results and discussion

The charged energy flow for different event classes is

presented in Fig. 2. In Figs. 1 and 2 from [36] , one can find the predictions for the older pre-LHC tuned versions of the generators used in this study.

The predictions of PYTHIA 6 versions [36] seem to be reasonably good in the central region (with the exception of diffractive events), but largely underestimate the measured values in the forward region in all cases. PYTHIA 8.135 predictions have a good description for the inclusive minimum bias, diffractive enriched and non-diffractive enriched event classes, but overestimate the measured values for the hard events.

PYTHIA 8.1 2M exhibits a slight decrease in overall values relative to version 8.135 (which uses the older Tune 1 [29]) for the minimum bias, non-diffractive enriched and hard event classes. The description for the hard event class is improved, while for the other two event classes a tendency to underestimate is observed. There is no major difference between the two versions for the diffractive event class.

With the exception of SIBYLL, a generator tuned to reproduce energy flow measurements, PYTHIA 8.186 seems to give the best description overall of the LHC-tuned generators. Its predictions for the diffractive enriched class are very similar to version 8.135, but for the other event classes the predictions are further away from the measurements, exhibiting a constant tendency to overestimate.

PYTHIA 8.219 gives a good description of the charged energy flow for the diffractive enriched class, and is similar to version 8.186. One can see that the predictions tend to increasingly overestimate in the forward region, but are similar to version 8.186 in the central region. The differences can be explained by the 10% increase in charged particle densities in the forward region implemented with the Monash 2013 tune [31].

EPOS 1.99 [36] describes reasonably well the charged energy flow for inclusive minimum bias, hard and non-diffractive enriched event classes, slightly overestimating the measurements in the last two bins. It underestimates the charged energy flow for diffractive processes in the forward region.

EPOS LHC predictions are very similar to those of PYTHIA 8.219 for all event classes except the diffractive enriched class, where similarly to the previous ver-

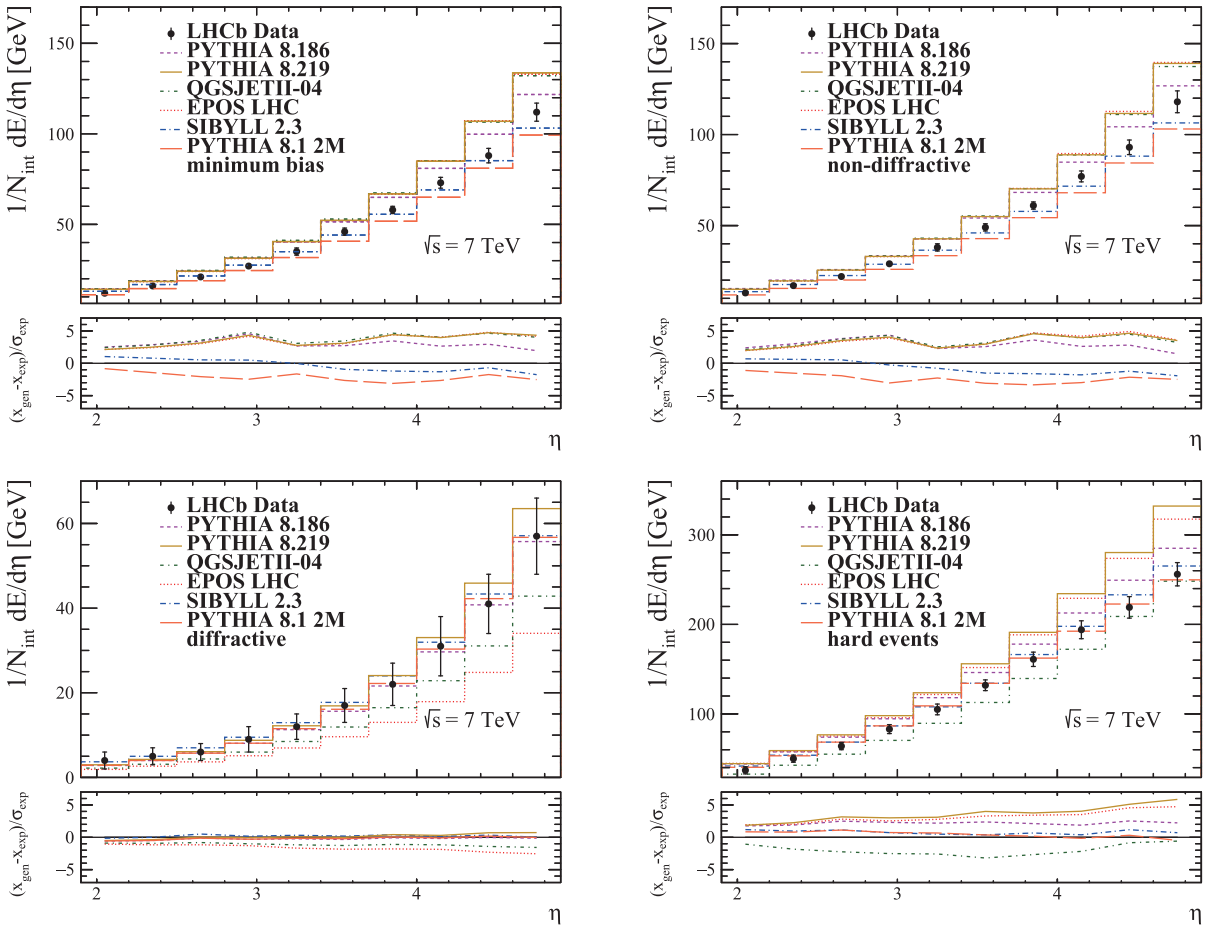

Fig. 2. (color online) Charged energy flow for different classes of events from pp collisions at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the systematic uncertainty (the statistical uncertainty is negligible) [36].

sion, it underestimates the charged energy flow. As one can see in the restricted minimum bias plot, the apparent overestimate of the soft process component is similar to PYTHIA 8.219. Compared to the previous version, we observe that the predictions are worse (except for diffractive events). Overall, EPOS LHC overestimates the measurements, with an increasing trend in the forward pseudorapidity region.

The predictions of QGSJET01 and QGSJETII-03 from [36] are similar for the inclusive minimum bias class and they overestimate the charged energy flow. QGJSET01 gives a better description of the diffractive and hard event classes in the central region, but tends to overestimate the measurements for hard events and underestimates them for the diffractive events in the forward region. The general trend of QGSJETII-03 is that of underestimation for the hard events class.

The predictions of QGSJETII-04 are similar to the previous versions for the inclusive minimum bias event class. The description of the charged energy flow for hard events is underestimated even more than in the case of QGSJETII-03. The diffractive component description is similar to QGSJETII-03, but with a slightly larger tendency to underestimate. For the other event classes the differences with respect to the measured LHCb charged energy flow are significant. Although the absolute values are clearly far from the experimental values, the shapes are well described. QGSJETII-04 is very similar to EPOS LHC and PYTHIA 8.219 in its description of the charged energy flow for inclusive minimum bias and non-diffractive enriched event classes.

SIBYLL 2.1 prediction [36] describes very well the measurements for inclusive minimum bias events. It also gives a reasonably good description of the diffractive events, as the values are within the error bars, although a tendency to underestimate them can be seen. The hard event component is well described in the central region, but it is overestimated in the forward region.

SIBYLL 2.3 seems to give the best prediction for all event classes (on par with PYTHIA 8.186 for the diffractive enriched class). It can be seen that it has a slight tendency to underestimate in the forward region in the case of inclusive minimum bias and non-diffractive enriched event classes.

As one can see in Table 1, PYTHIA 8.219 and EPOS give similar ratios of hard events, but the number of visible events and the ratio of diffractive events are smaller for EPOS. PYTHIA 8.186 ratio of hard events is larger than for version 8.219, but the ratio of diffractive events is close, indicating that the mechanisms of diffractive processes are similar. QGSJETII-04 ratio of hard events is clearly larger than the others, while the ratio of diffractive events is smaller, so that the hard process component seems to be larger for this generator. Likewise, SIBYLL hard process component is larger than in PYTHIA and EPOS.

As one can see in the transverse momentum plot, Fig. 3, PYTHIA and QGSJET predictions are similar in shape. There is no major difference between PYTHIA LHC-tuned versions. PYTHIA and EPOS predictions are rather similar in the interval 0.5-1.5 GeV/c. QGSJET prediction seems closest to the LHCb measurements, but for all generators there are visible differences in absolute scale, especially in the hard part of the spectrum. SIBYLL-generated spectrum has a shape which approaches the experimental one, but the absolute values differ significantly. The shapes of the spectra generated with QGSJET, EPOS and both versions of PYTHIA are close to the experiment.

In the pseudorapidity plot, shown in Fig. 3, one can see that all predictions cluster together at low values, as the models were tuned using the measurements from the central region of the LHC experiments. QGSJET, EPOS and PYTHIA 8.2 underestimate the measurements for values below $\eta = 3.5$ and overestimate them in the forward region (where they also remain clustered together). PYTHIA 8.1 also underestimates the measurements in the central region, but the prediction in the forward region seems to be reasonably good. SIBYLL largely underestimates the measurements across the whole range.

For the (probability density of) multiplicity distribution in Fig. 3, the closest prediction seems to be that of EPOS. All LHC-tuned generators reproduce the measurements well for this distribution, except SIBYLL which deviates significantly. One can see that EPOS prediction clusters together with PYTHIA estimates in the medium-high multiplicity region. For values below $n_{ch} = 10$, EPOS seems to be better than PYTHIA. QGSJET prediction is close to EPOS and PYTHIA, but the underestimation at low multiplicities in the interval $n_{ch} = 10$-20 is larger, the deviations from the measurements ranging between $\sim 3-5\ \sigma$. SIBYLL prediction strongly favors low multiplicities, but gets closer to the measured values towards high multiplicities.

The pseudorapidity distribution in Fig. 4 is best described by PYTHIA 8.186. PYTHIA 8.219 prediction is close too. EPOS and QGSJET estimates are a bit further away from the experimental values. SIBYLL prediction is significantly different, in absolute value as well as in the shape of the distribution. With the exception of SIBYLL, the clustering of the predictions can be seen in the central pseudorapidity region, indicating that the tuning was done using similar measurements. The prediction of EPOS describes the measurements reasonably well in the central region ($2 < \eta < 2.5$), but it diverges upwards from the measured values in the forward region. This effect of overestimation in the forward region is similar to Fig. 3. QGSJET slightly underestimates the meas-

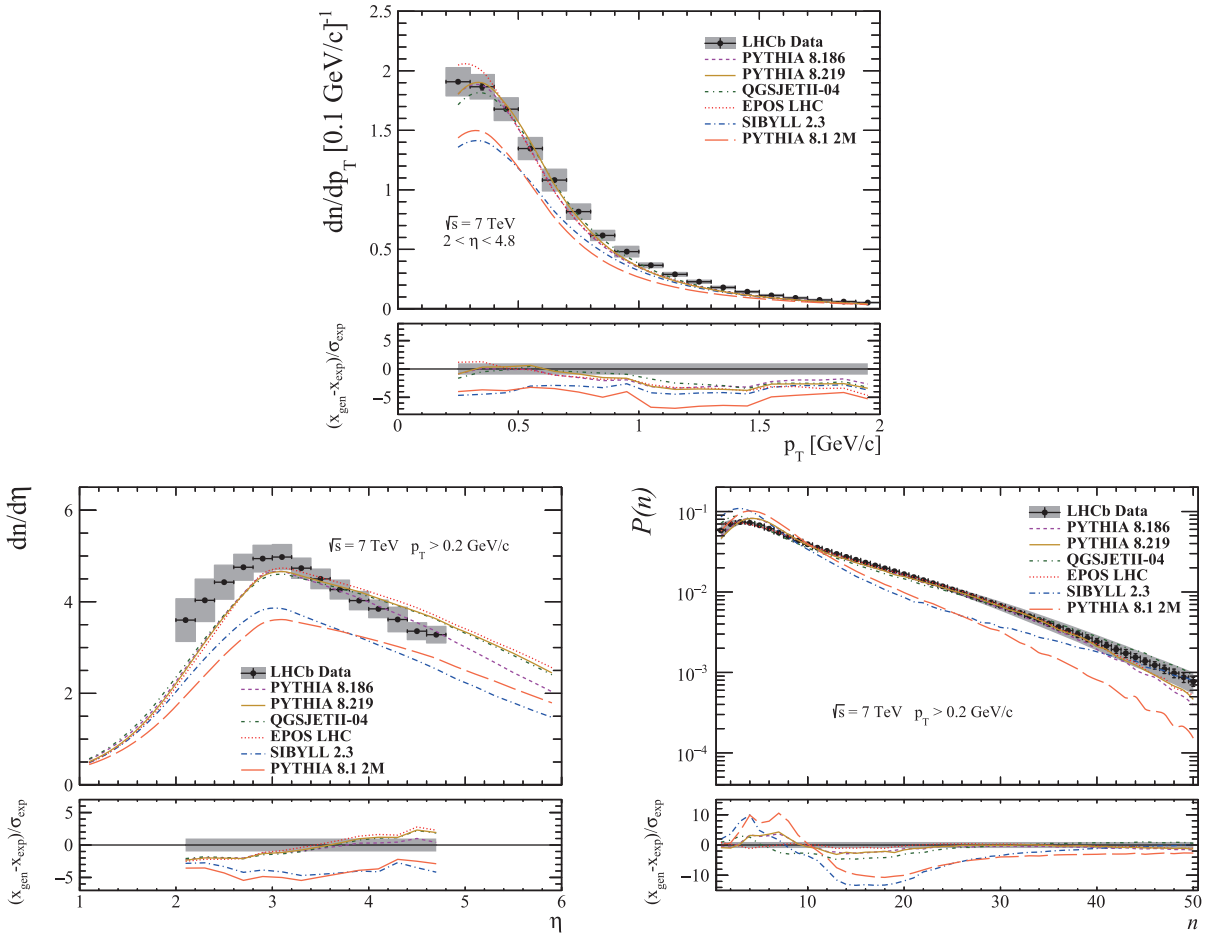

Fig. 3. (color online) Transverse momentum, pseudorapidity and multiplicity distributions for prompt charged particles in the kinematic region $2 < \eta < 4.8$, $p \geqslant 2$ GeV/c and $p_T > 0.2$ GeV/c , at $\sqrt{s} = 7$ TeV. The vertical bars represent the statistical error and the grey bands represent the combined uncertainties (statistical and systematic) [37].

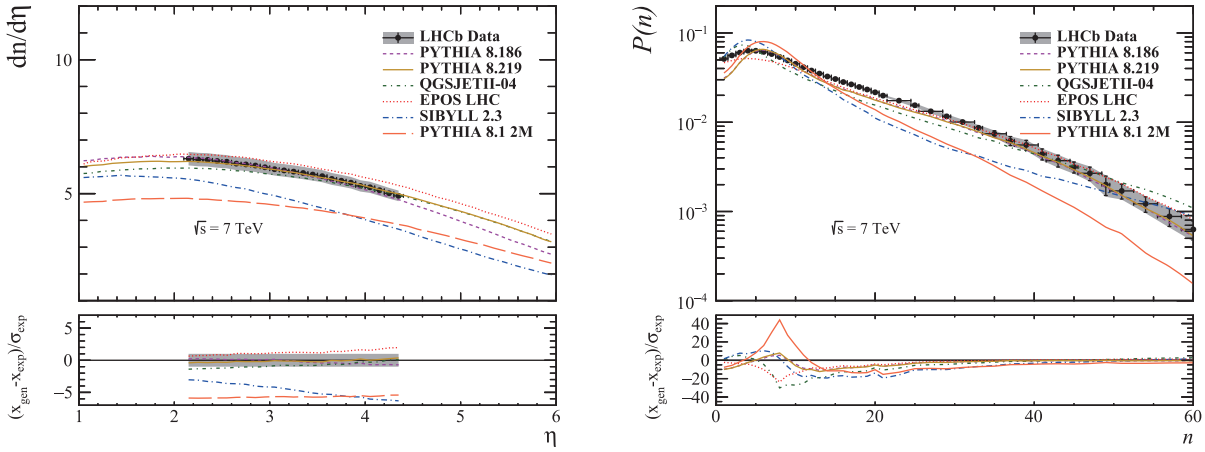

Fig. 4. (color online) Pseudorapidity and multiplicity distributions for prompt charged particles in the kinematic region $2 < \eta < 4.5$ at $\sqrt{s} = 7$ TeV. The vertical bars represent the statistical error and the grey bands represent the combined uncertainties (statistical and systematic) [38].

urements in the central region, but gets closer in the forward region (overlapping with PYTHIA 8.219).

The multiplicity distribution is not perfectly described by any of the generators, but one can see that the predictions of EPOS and PYTHIA seem to get better at higher multiplicities, as we have also seen for the multi-

plicity distribution. The distributions generated with SIBYLL and QGSJET are significantly different from the experimental ones.

The pseudorapidity plot in Fig. 5 shows a good agreement between PYTHIA versions and the LHCb measurements. EPOS also gives a good description of the measurements in the central region, but diverges upwards in the forward region. SIBYLL prediction is similar to QGSJET at low rapidity, but they both diverge in the forward region and are far from the experimental distribution. The discontinuity at $\eta = 2.5$ is due to the hard event selection criterion of a minimum of one particle with $2.5 \leqslant \eta \leqslant 4.5$ and $p_T \geqslant 1$ GeV/c [38].

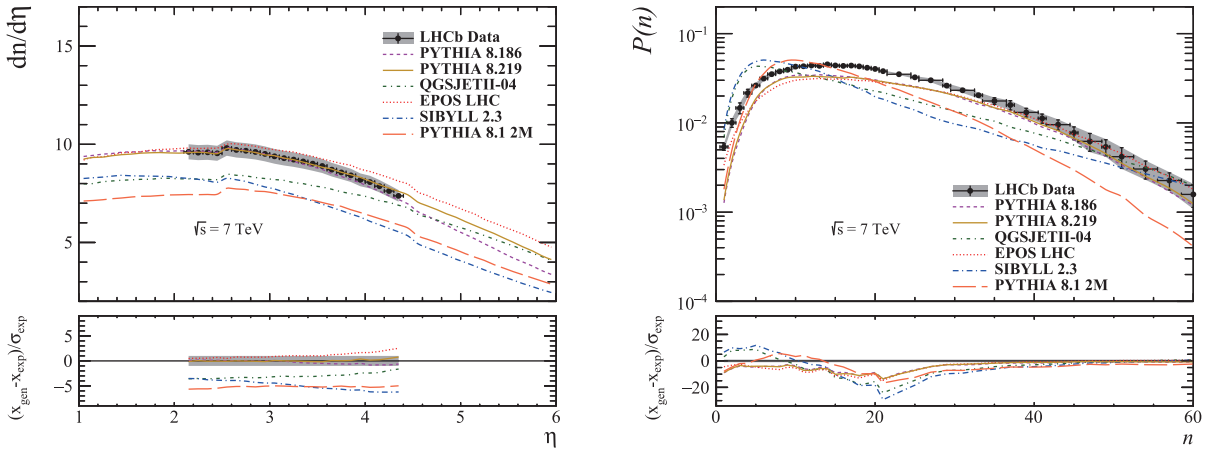

Fig. 5. (color online) Pseudorapidity and multiplicity distributions for prompt charged particles in the kinematic region $2 < \eta < 4.5$ from "hard" events at $\sqrt{s} = 7$ TeV. The vertical bars represent the statistical error and the grey bands represent the combined uncertainties (statistical and systematic) [38].

As in Fig. 4, the multiplicity distribution is not well described by the generators, with PYTHIA and EPOS closest to the measurements.

As can be seen in Fig. 6, the best predictions are given by QGSJET, EPOS and PYTHIA 8.219. All generated shapes and spectrum slopes agree well with the experimental distribution.

From the pseudorapidity plots in Figs. 4-6 it can be seen that the predictions of PYTHIA 8.1 2M largely underestimate the measurements. The differences between the predictions of PYTHIA with Tune 2M and the two

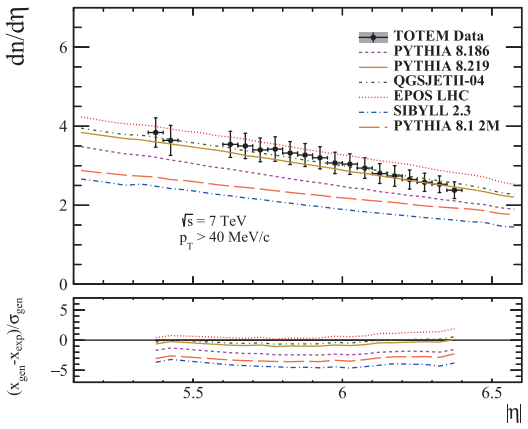

Fig. 6. (color online) Prompt charged particle pseudorapidity distribution in the kinematic region $p_T > 40$ MeV/c and $5.3 \leqslant |\eta| \leqslant 6.5$, at $\sqrt{s} = 7$ TeV. The error bars represent the combined statistical and systematic errors [39].

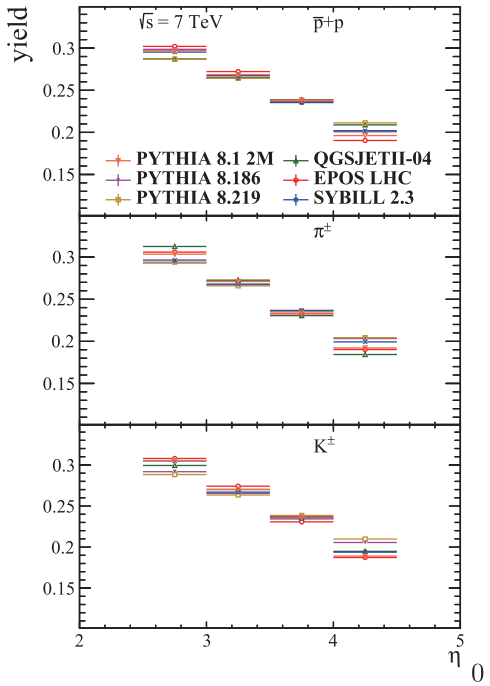

Fig. 7. (color online) Yields (normalized to 1) of protons, pions and kaons with $p_T \geqslant 1.2$ GeV/c and $p \geqslant 5$ GeV/c.

LHC tunes are large in the central region and exhibit a converging trend towards higher pseudorapidity. The multiplicity plots in Figs. 3-6 are clearly not well reproduced by PYTHIA 8.1 2M, which favors very low multiplicities.

The ratios of hard events for PYTHIA and EPOS, given in Table 2, are close, suggesting a similarity between the descriptions of hard processes. SIBYLL ratio is slightly higher. QGSJET ratio of hard events is considerably higher than for the other generators, so again one can see that it favors the hard processes.

The plot of the $\bar{p}/p$ ratio is shown in Fig. 9. All predictions have the same trend of apparent decrease towards the beamline and it can be said that the ratio is reasonably well described. The π^-/π^+ ratio, shown in the same figure, is also well described by all generators, with the exception of QGSJET for the high p_T region, where it seems to show a charge asymmetry between π^+ and π^-. Also, all predictions seem to cluster together, again with the exception of QGSJET at high p_T. The K^-/K^+ ratio, shown in Fig. 10 , is fairly well described by all generators.

The closest prediction of $(K^+ + K^-)/(\pi^+ + \pi^-)$ (shown in the same figure) seems to be that of SIBYLL followed by EPOS. However, overall all generators fail to describe this measurement. In the high p_T range, QGSJET underestimates the measurements and has a pronounced ascending trend.

A clustering of the predictions in the low p_T plot for $(p+\bar{p})/(\pi^+ + \pi^-)$ (shown in Fig. 11) is observed. Here, all generators give a good description of the measurements. For the high p_T range the closest predictions are from EPOS and PYTHIA 8.1, while for the middle p_T range no generator seems to correctly describe the ratio. In the high p_T range, the ratio is again underestimated by QGSJET, which again has an ascending trend, while SIBYLL largely overestimates the ratio.

The $(p+\bar{p})/(K^+ + K^-)$ ratio is shown in the same figure. The best prediction overall is from EPOS LHC. SIBYLL and QGSJET give a good description of this ratio in the low p_T range. In the middle p_T range, SIBYLL prediction overlaps with EPOS LHC. In the high p_T range, PYTHIA 8.219 and QGSJET also give a reason-

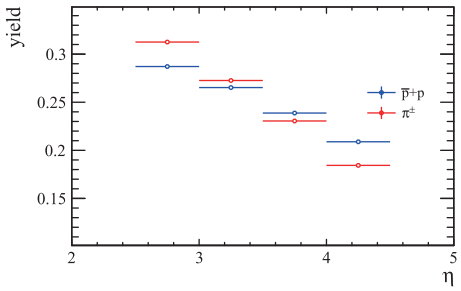

Fig. 8. (color online) Yields (normalized to 1) of protons and pions with $p_T \geqslant 1.2$ GeV/c and $p \geqslant 5$ GeV/c generated with QGSJETII-04.

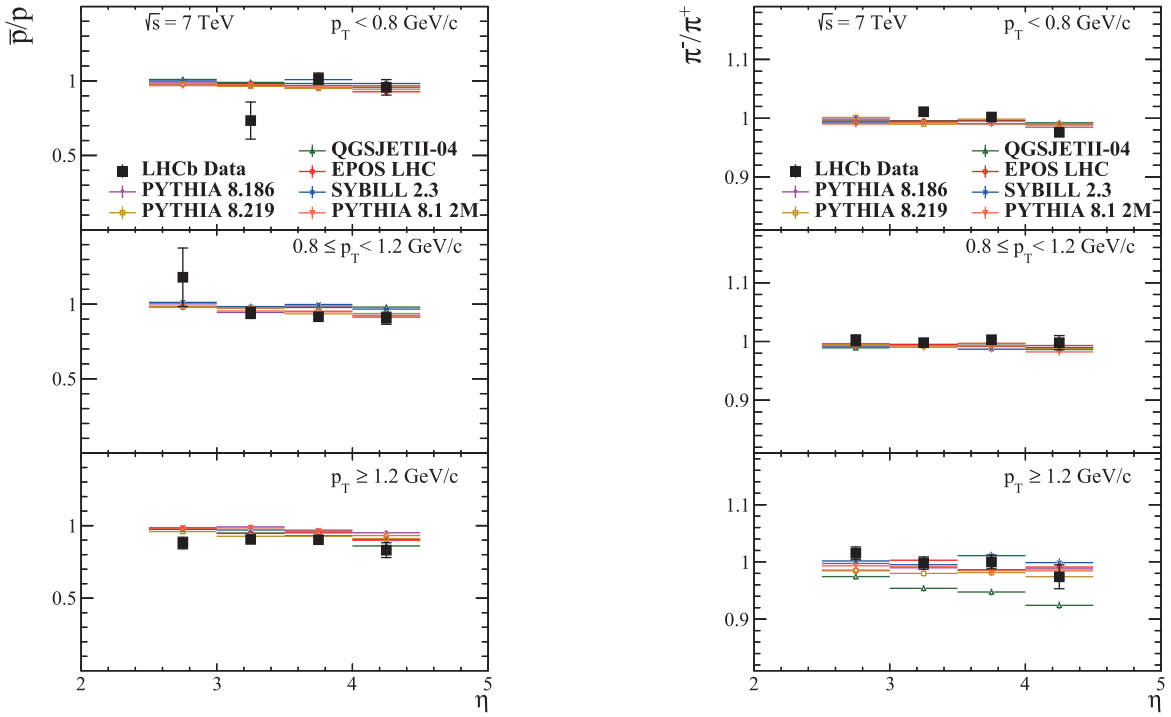

Fig. 9. (color online) Prompt charged hadron ratios as a function of pseudorapidity in the kinematic region $2.5 \leqslant \eta \leqslant 4.5$ and $p \geqslant 5$ GeV/c , in various p_T intervals, at $\sqrt{s} = 7$ TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties [40].

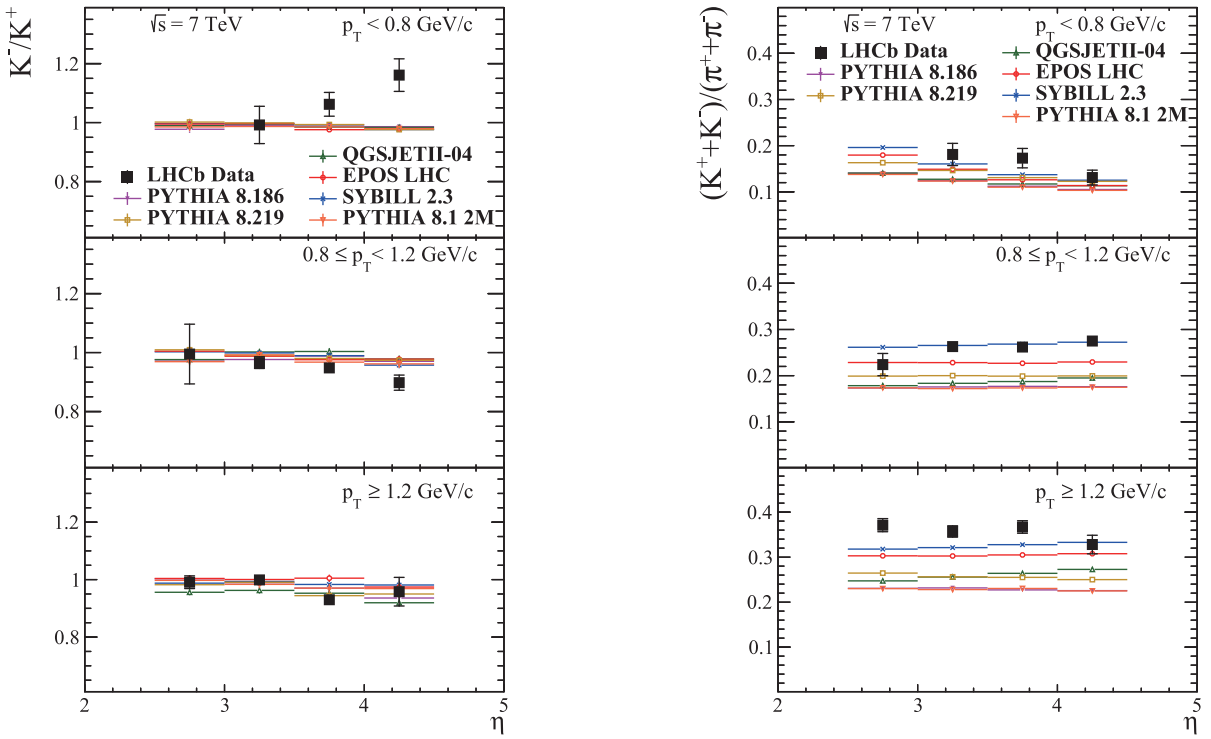

Fig. 10. (color online) Prompt charged hadron ratios as a function of pseudorapidity in the kinematic region $2.5 \leqslant \eta \leqslant 4.5$ and $p \geqslant 5$ GeV/c , in various p_T intervals, at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties [40].

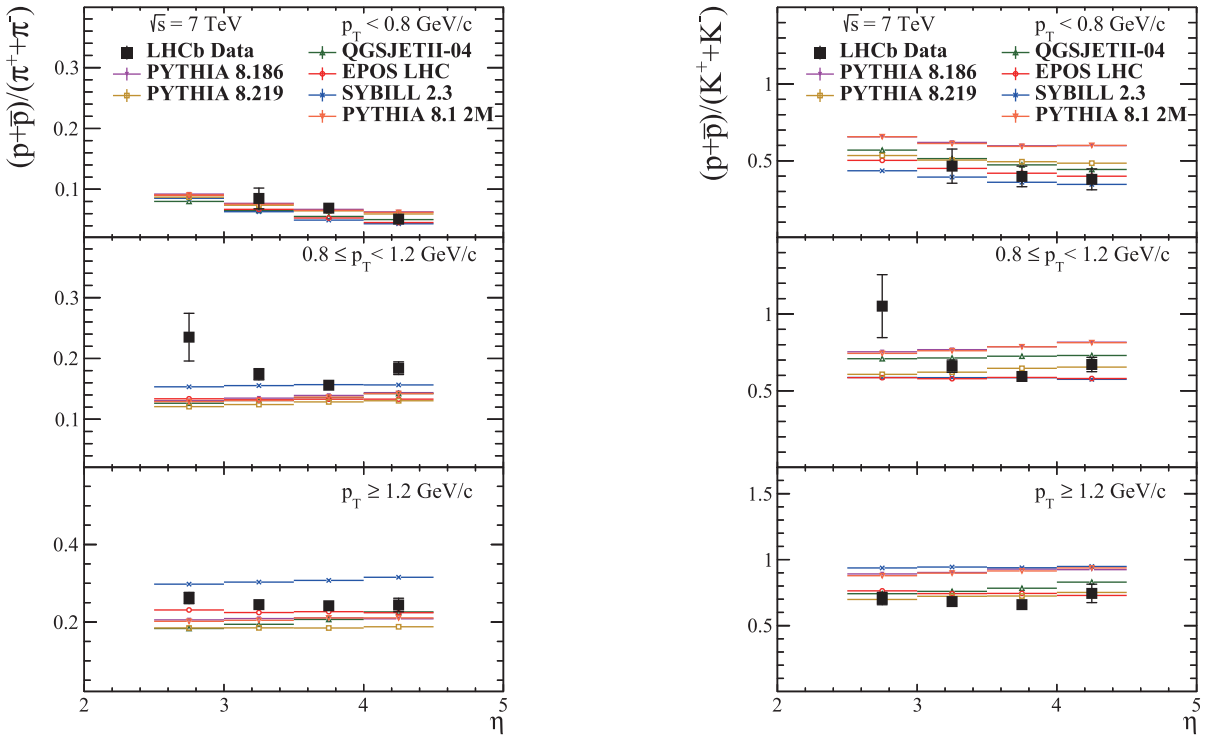

Fig. 11. (color online) Prompt charged hadron ratios as a function of pseudorapidity in the kinematic region $2.5 \leqslant \eta \leqslant 4.5$ and $p \geqslant 5$ GeV/c , in various p_T intervals, at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties [40].

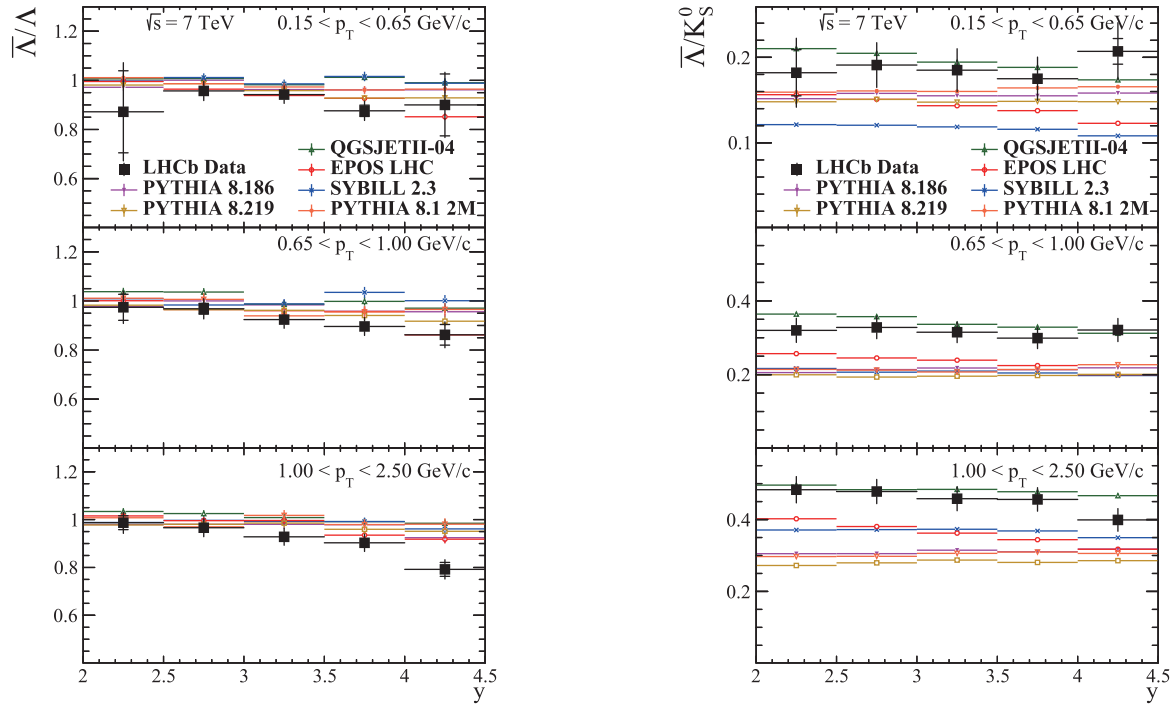

Fig. 12. (color online) Prompt V^0 particle ratios as a function of rapidity in the kinematic region $2 \geqslant y \geqslant 4.5$, in various p_T intervals, at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties, and the small horizontal ones show the statistical component [41].

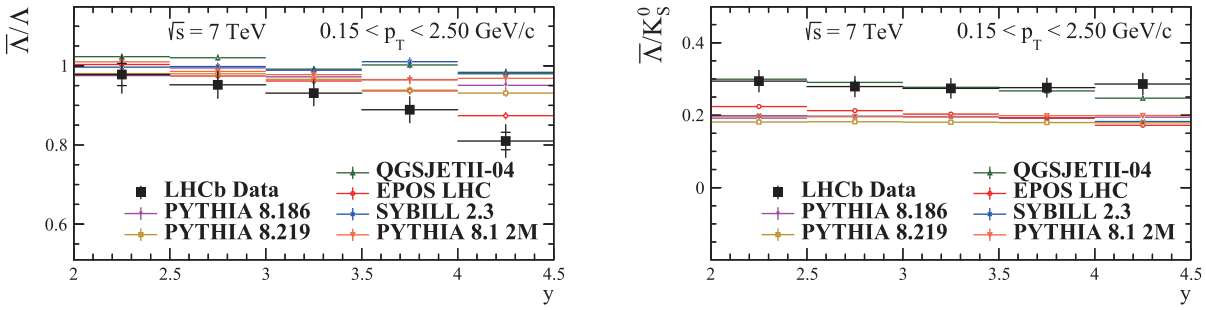

Fig. 13. (color online) Prompt V^0 particle ratios as a function of y in the kinematic region $2 \geqslant y \geqslant 4.5$ and $0.15 < p_T < 2.50$ GeV/c , at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties, and the small horizontal ones show the statistical component [41].

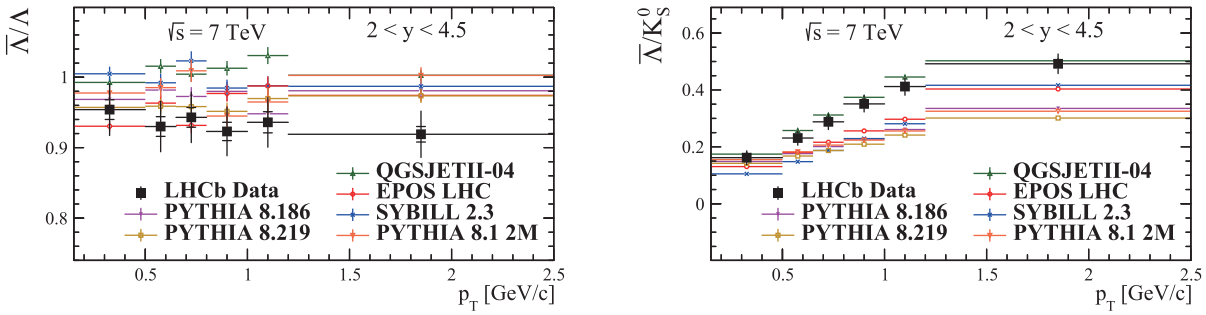

Fig. 14. (color online) Prompt V^0 particle ratios as a function of p_T in the kinematic region $2 \geqslant y \geqslant 4.5$ and $0.15 < p_T < 2.50$ GeV/c , at $\sqrt{s}$ = 7 TeV. The LHCb data vertical bars represent the combined statistical and systematic uncertainties and the small horizontal ones show the statistical component [41].

ably good description, although QGSJET exhibits again an ascending trend. SIBYLL again largely overestimates the ratio in this range together with PYTHIA 8.1. The predictions of PYTHIA for the proton/kaon and kaon/pion ratios are clearly improved by the strangeness enhancement from the Monash 2013 tune.

The yields of protons and pions in the high p_T region obtained with QGSJET are shown in Fig. 8 . It is rather clear that the decreasing slope for pions is higher than for protons. The yields of protons, pions and kaons in the same p_T region for all generators are shown in Fig. 7. It can be seen that the slope of the proton yield distribution from QGSJET is the lowest, while for the pion yield it is the highest. The slope of the kaon yield is in between the slopes from the other generators. These observations, together with the observed ascending trend of the QGSJET predictions for the proton/pion, kaon/pion and proton/kaon ratios in the high p_T range, while the data and the predictions of the other generators do not show such a trend, suggest that the proton multiplicity decreases too slowly and the pion multiplicity decreases too quickly for high pseudorapidities.

As can be seen in Figs. 12-14, the $\bar{\Lambda}/\Lambda$ ratio is best described by EPOS LHC and PYTHIA 8.219, pointing to a good baryon number transport. Nonetheless, all predictions have more or less the same trend. The $\bar{\Lambda}/K^0_S$ ratio seems to be reasonably well described by QGSJET, while the other generators largely underestimate it.

5 Conclusions

In this paper we presented a comparative study of the generators EPOS LHC, QGSJETII-04, SIBYLL 2.3 and versions 8.186 and 8.219 of PYTHIA. The observables for which the study was conducted were the charged energy flow, charged particle multiplicities and densities, charged hadron production ratios, V^0 ratios and the other strange particle distributions. It is reasonably clear that no generator reproduces the data for all observables studied, but rather that one generator describes well only a particular set of observables or aspects of particle production. As a general trend, the predictions are better in the central region. The tuning using the data from the central rapidity range of the general purpose LHC detectors is visible and clearly improves the estimations even for the forward region, although the effect of extrapolation to higher rapidities is in clear disagreement with the experimental data.

It was observed that the charged energy flow, which can be regarded as a global event observable, is relatively well described by all generators, at least in terms of shape. The best prediction overall for the charged energy flow is from SIBYLL 2.3, a generator tuned specifically to reproduce correctly this type of observable. PYTHIA 8.186 gives the best description of the other LHC-tuned generators.

EPOS and PYTHIA, especially version 8.219, are very similar in their description of the observables. The similarity between these generators may arise from the partonic approach and similar perturbative calculations that are used for hard parton collisions.

QGSJET is similar to EPOS in the description of some observables, like the charged energy flow (except for the hard event class) and charged particle densities, but also shares some similarities with SIBYLL.

The multiplicity distributions are generally not well reproduced by the generators. EPOS and PYTHIA give the best predictions overall. Also, they seem to get better with the increasing hardness of the processes, but exhibit a similar effect as the other generators, i.e. favoring either very low or high multiplicity events, albeit at a much lower level than SIBYLL, which has the most polarizing behavior.

SIBYLL has a few notable successes in describing some particle ratios. Also, its predictions for charged particle pseudorapidity and transverse momentum distributions have a good shape.

The best baryon transport mechanism seems to be that of EPOS, followed by PYTHIA, while the $\bar{\Lambda}/K^0_S$ ratio is best described by QGSJET.

Most of the observed differences seem to be an effect of extrapolation in the forward region, and the extrapolation uncertainties seem to be rather large. Nonetheless, in the majority of cases, the measurements fall within a band defined by the most extreme predictions.

The relative contributions of particle production processes differ between the central and forward regions. In the central pseudorapidity region, there is a significant contribution of hard parton-parton scatterings (with high squared momentum transfer), to which high multiplicity events and high p_T jets are associated. In the forward region, on the other hand, the underlying events (multi-parton interactions and beam remnants), as well as diffractive processes, give a considerable contribution. The event generators usually have different sets of parameters for each process, and as such, when tuning using the measurements from one pseudorapidity region or the other, different parameters are constrained, so that each tune is applicable for studies in its respective region. As shown in this paper, the predictions in the forward region are improved by tuning the generators using the measurements from the central region, but it seems that a dedicated tuning procedure is still necessary. As a result, the effectiveness of each tune is somewhat limited when extrapolating from the central to the forward region and vice versa. Ideally, the measurements from both the forward and central regions should be used simultaneously when tun-

ing a generator, but this is seldom done. In many cases there are intrinsic limitations of the generators, or of the models they are based on, which prevent a simultaneous tune in both regions, and hence a more consistent overall description of the processes. Difficulties related to such a tuning procedure also arise from the different experimental conditions in each region.

As we have seen in this paper, it seems that modeling of the soft processes is still open to improvement, and a tuning of generators is required to improve the precision in the forward rapidity range. Hence, it may prove useful for tuning to take into account the measurements from LHCb and TOTEM, the LHC experiments in the forward region, where the soft component is considerably larger than in the central region, the baryon transport is different, and the multi-parton collisions might give a different signal.

We would like to thank the authors of PYTHIA/EPOS/QGSJET/SIBYLL generators and the authors of the CRMC interface (T. Pierog, C. Baus, and R. Ulrich).

References

1 K. A. Olive et al, (Particle Data Group), Chin. Phys. C, **38**: 090001 (2014)
2 T. Pierog et al, Phys. Rev. C, **92**: 034906 (2015)
3 S. Ostapchenko, Phys. Rev. D, **83**: 014018 (2011)
4 F. Riehn, R. Engel et al, in *34th International Cosmic Ray Conference* (ICRC 2015), The Hague, Netherlands, C15-07-30, PoS ICRC2015(2016) 558, arXiv: 1510.00568
5 The CRMC project homepage: https://web.ikp.kit.edu/rulrich/crmc.html (Last visited: Sept 29, 2017)
6 T. Sjöstrand et al, Computer Physics Communications, **178**(11): 852-867 (2008)
7 T. Sjöstrand et al, Computer Physics Communications, **191**: 159-177 (2015)
8 G. Aad et al (ATLAS Collaboration), JINST, **3**: S08003 (2008)
9 K. Aamodt et al (ALICE Collaboration), JINST, **3**: S08002 (2008)
10 S. Chatrchyan et al (CMS Collaboration), JINST, **3**: S08004 (2008)
11 A. Augusto Alves, Jr. et al (LHCb Collaboration), JINST, **3**: S08005 (2008)
12 G. Anelli et al, JINST, **3**: S08007 (2008)
13 T. Sjöstrand, S. Mrenna, and P. Skands, JHEP, **0605**: 026 (2006)
14 P. Skands, Boston Jets Workshop, MIT, January 21-23, 2014, http://skands.physics.monash.edu/slides/14/14-SoftQCD-BostonJets.pdf
15 S. Ostapchenko, Nuclear Physics B - Proceedings Supplements, **151**(1): 143-146 (2006)
16 T. Sjöstrand, Lecture notes, Lund University, Lund, Sweden, 24 February 2015, http://home.thep.lu.se/~torbjorn/ppp2015/lec5.pdf
17 T. Sjöstrand, LHC and the Standard Model: Physics and Tools, CERN, 16 June 2017, http://home.thep.lu.se/~torbjorn/talks/cern17pdf.pdf
18 S. Navin, LUTP-09-23, MCnet/10/09, arXiv: 1005.3894
19 H.J. Drescher et al, Physics Reports, **350**(2-4): 93-289 (2001)
20 C. Ewerz, HDTHEP-02-35, arXiv: hep-ph/0306137
21 S. Porteboeuf, T. Pierog, and K. Werner, Proceedings of the XLVth Rencontres de Moriond (La Thuile, Italy, 2010) http://moriond.in2p3.fr/QCD/2010/Proceedings10/Porteboeuf.pdf
22 T. Pierog and K. Werner, Nuclear Physics B - Proceedings Supplements, **196**: 102-105 (2009)
23 S. Ostapchenko and D. Heck, in *29th International Cosmic Ray Conference Pune* 7 (2005), p. 135-138
24 G.I. Lykasov and M.N. Sergeenko, Z. Phys. C, **70**: 455-462 (1996)
25 S. Ostapchenko, AIP Conf. Proc., **928**: 118-125 (2007)
26 E. Ahn et al, Phys. Rev. D, **80**: 094003 (2009)
27 R. Engel et al, ISVHECRI 2016- XIX International Symposium on Very High Energy Cosmic Ray Interactions, EPJ Web of Conferences, **145**: 08001 (2017)
28 V. Khachatryan et al, Eur. Phys. J. C, **76**: 155 (2016)
29 PYTHIA Collaboration, http://home.thep.lu.se/~torbjorn/pythia82html/Welcome.html (Last visited: Sept 29, 2017)
30 R. Corke and T. Sjöstrand, J. High Energ. Phys., **2011**: 32 (2011)
31 P. Skands et al, Eur. Phys. J. C, **74**: 3024 (2014)
32 S. Ostapchenko, EPJ Web Conf. 52 (2013) 02001, C12-08-10.1, DOI: 10.1051/epjconf/20135202001
33 S. Ostapchenko, Progress of Theoretical Physics Supplement, **193**: 204211 (2012)
34 F. Riehn et al, PoS ICRC, 2017 (2017) 301, arXiv: 1709.07227
35 F. Riehn et al, EPJ Web of Conferences 99, 12001 (2015) DOI: https://doi.org/10.1051/epjconf/20159912001.
36 R. Aaij et al, Eur. Phys. J. C, **73**: 2421 (2013)
37 R. Aaij et al (LHCb Collaboration), Eur. Phys. J. C, **74**: 2888 (2014)
38 R. Aaij et al (LHCb Collaboration), Eur. Phys. J. C, **72**: 1947 (2012)
39 G. Antchev et al (TOTEM Collaboration), Europhys. Lett., **98**: 31002 (2012)
40 R. Aaij et al, Eur. Phys. J. C, **72**: 2168 (2012)
41 R. Aaij et al, JHEP, **1108**: 034 (2011)