




# Confronting experimental data with heavy-ion models: RIVET for heavy ions

Christian Bierlich<sup>1,2,a</sup> , Andy Buckley<sup>3</sup>, Christian Holm Christensen<sup>1</sup>, Peter Harald Lindenov Christiansen<sup>4</sup>, Cody B. Duncan<sup>5,6</sup>, Jan Fiete Grosse-Oetringhaus<sup>7</sup>, Przemyslaw Karczmarczyk<sup>7,8</sup>, Patrick Kirchgaesser<sup>6</sup>, Jochen Klein<sup>7,9</sup>, Leif Lönnblad<sup>2</sup>, Roberto Preghenella<sup>10</sup>, Christine O. Rasmussen<sup>2</sup>, Maria Stefaniak<sup>8,11</sup>, Vytautas Vislavicius<sup>1</sup>

<sup>1</sup> Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark

<sup>2</sup> Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, 223 62 Lund, Sweden

<sup>3</sup> School of Physics and Astronomy, University of Glasgow, G12 8QQ Glasgow, UK

<sup>4</sup> Department of Physics, Lund University, Professorsgatan 1, 223 62 Lund, Sweden

<sup>5</sup> School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia

<sup>6</sup> Institute for Theoretical Physics, Karlsruhe Institut für Technologie, Wolfgang-Gaede-Straße 1, 76131 Karlsruhe, Germany

<sup>7</sup> CERN, Esplanade de Particules 1, Geneva, Switzerland

<sup>8</sup> Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland

<sup>9</sup> Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Turin, Italy

<sup>10</sup> Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Bologna, Italy

<sup>11</sup> Subatech-IMT Atlantique, 4 Rue Alfred Kastler, 44307 Nantes, France

Received: 10 March 2020 / Accepted: 8 May 2020

© The Author(s) 2020

**Abstract** The RIVET library is an important toolkit in particle physics, and serves as a repository for analysis data and code. It allows for comparisons between data and theoretical calculations of the final state of collision events. This paper outlines several recent additions and improvements to the framework to include support for analysis of heavy ion collision simulated data. The paper also presents examples of these recent developments and their applicability in implementing concrete physics analyses.

## 1 Introduction

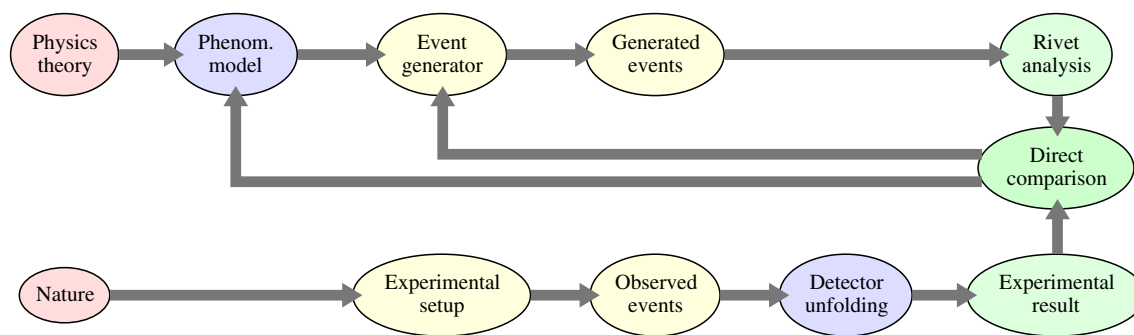
High energy collisions of hadrons have, with access to data from colliders at the energy frontier over the past 20 years, been the main avenue for precision studies of the phenomenology of the strong nuclear force (QCD). Since such collisions involve a vast amount of different phenomena – some calculable from first principles, other relying on models – development has tended towards large calculation packages, which attempt to include as many effects as possible, with the aim of simulating full “events” resembling the collisions observed by experiments. While such event generators are very useful, their individual model components are difficult to validate in a systematic way. This is because adding

new model components may compromise existing agreement with some results while improving agreement with other results. To overcome this challenge, mass comparison with current and past data is needed. Such comparisons are, for the most widespread proton–proton (pp) event generators, facilitated by the software package RIVET [1].

Heavy-ion physics has entered the precision era and quantitative and comprehensive model comparisons have become crucial [2]. Technical difficulties – similar to those faced earlier for collisions of protons – now also surface in this area of research. Inclusion of several new model components on the theoretical side, for example in the JETSCAPE generator package [3], introduces similar difficulties of validation, well known from hadron collisions. Furthermore, the discovery of QGP-like behaviour in collisions of small systems, i.e. pp and proton–ion (pA) [4, 5], suggests that similar physics is at play in these collisions. These points strengthen the need for a generalized and common method of performing systematic comparisons of theoretical models to data for proton–proton, nucleon–nucleus, and ion–ion (AA) collisions. Such a method will not only be practically useful but would also enable new avenues for discovering similarities between two previously quite detached fields of subatomic physics [6].

This paper introduces the new features of the RIVET framework to allow one to obtain a direct comparison between Monte–Carlo event generators and experimental results from

<sup>a</sup> e-mail: [christian.bierlich@thep.lu.se](mailto:christian.bierlich@thep.lu.se) (corresponding author)



**Fig. 1** Outline of how RIVET connects experimental analyses to theory and validation, with the typical workflow of a physics program shown. On the top row, a typical theoretical workflow is depicted, where physics

theory informs a phenomenological model implemented in an event generator. On the bottom row, the experimental workflow is shown in a similar fashion

pA and AA collisions. In Sect. 2 a brief introduction to the framework is given, along with additions which are motivated by requirements of heavy ion analyses, but are of more general nature. In Sect. 3 two further additions, more specialized to specific heavy ion analyses are introduced: (1) estimation of centrality in Sect. 3.1) and (2) the Generic Framework for flow observables in Sect. 3.2. In Sect. 4, a simple analysis example is given, with guidelines for the user to run the analysis and obtain a simple comparison. For the two latter additions, Sect. 5 presents use-cases where the need for calculating final-state observables in the same way as the experiment are obvious. For centrality estimation (in Sect. 5.1) we show how different centrality estimators on final state and initial state level change results on basic quantities like multiplicity at mid-rapidity, in particular in pA, but to some degree also in AA. For flow observables (in Sect. 5.2) we show how important non-flow contributions are captured using the generic framework implementation, as opposed to a simple calculation based on the initial state reaction plane.

## 2 The RIVET framework and new features of general nature

RIVET is a computational framework,<sup>1</sup> written as a C++ shared library supplying core functionality, and a large number (949 at the time of writing) of C++ plugin libraries implementing collider analysis routines.

In Fig. 1, the RIVET frameworks' connection to experiment and theory is outlined. The figure illustrates how comparison of experimental results to results from event generators provide a feed-back loop to the phenomenological models and the concrete event generator development. This type of feed-

back loop is widely used for development of Monte-Carlo event generators [8–10], as well as their validation and tuning [11–13].

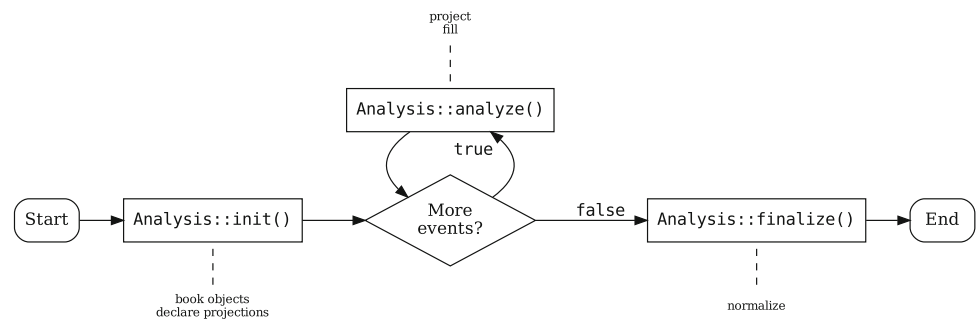
RIVET analyses are performed as comparisons to data unfolded to “particle level”, meaning that the experimental data should be processed to a level where it is directly comparable to Monte-Carlo output, as a list of stable or unstable particles.

After a short technical introduction in Sect. 2.1, the following sections present additions to RIVET motivated by heavy ion development, but of a general nature. In Sect. 2.3 an example of implementing experimental particle definitions is shown. In Sect. 2.4 the ability to generate cross-system observables by running the same analysis on Monte-Carlo for both AA/pA and pp, followed by the ability to run analyses in different “modes” specified by analysis options in Sect. 2.5. Finally, the possibility to perform event mixing is presented in Sect. 2.6.

### 2.1 Technical introduction

The core RIVET library provides tools to process simulated events in the HEPMC format [14] with the aim of calculating observable quantities. The `Projection` paradigm is one of the key RIVET concepts. Each projection calculates a specific observable from the full, generated event. This can, for example, be all charged particles in a certain fiducial volume, all possible reconstructed Z-bosons decaying leptonically, event shapes, or, as introduced in the Sect. 3, event centralities or  $Q$ -vectors. There are two key advantages to projections. The first is computational. Since a projection can be reused across analyses, several analyses can be run simultaneously without the need to repeat a “particle loop” for each analysis. Secondly, and most importantly, projections serve as convenient shorthands when implementing an experimental analysis. There is no need to repeat the implementation

<sup>1</sup> A brief review of the framework, with emphasis on functionality necessary for later sections, is given here. For a full description of the RIVET framework, the reader is referred to the original manual [1] or the manual of RIVET version 3 [7].

**Fig. 2** Execution flow of an Analysis class

to select stable particles, or of a certain experimental trigger definition if it already exists as a pre-defined projection.

A RIVET analysis is split into three parts, implemented as separate member functions of an analysis class, inheriting from the Analysis base class. The execution flow of an analysis is sketched in Fig. 2.

**Initialization:** The `init()` method is called once in the beginning of the analysis, and is intended for the user to initialize (or “book”) objects such as projections or histograms used throughout the analysis.

**Analysis:** The `analyze(const Rivet::Event&)` method is called once per event, and is intended for the user to implement calculation of single event quantities, and fill histograms.

**Finalization:** The `finalize()` method is called once at the end of the analysis, and is intended for the user to apply the correct normalization of histograms, compute all-event averages and possibly construct ratios of aggregated quantities. In Sect. 2.4 the possibility to run this method for a merged sample is outlined.

## 2.2 The HEPMC heavy ion container

A RIVET analysis has access to heavy-ion specific generator information, through the heavy-ion container of a HEPMC input event, provided that it is filled by the event generator. The heavy-ion container contains several data members, which in general are model dependent generation information, in most cases linked to the Glauber model [15, 16], such as number of wounded nucleons, number of hard sub-collisions and number of spectator neutrons/protons. The heavy-ion record also holds more general information, such as the impact parameter, the reaction plane angle, and from HEPMC version 2.6.10 onward, it is possible to set a value for generator-level centrality. Detailed information about data members available in RIVET, and for implementing a HEPMC heavy-ion container in an event generator, is provided in the projects’ respective online documentation. It is important to note the following for analyses implemented using information from the HEPMC heavy-ion container:

- There is no guarantee that an event generator implements the heavy-ion container, and one should perform appropriate tests to avoid code failure.
- There is no guarantee that the generator definition of a certain model dependent quantity is equivalent to the same as used in the experimental analysis.

## 2.3 Implementing experimental primary-particle definitions

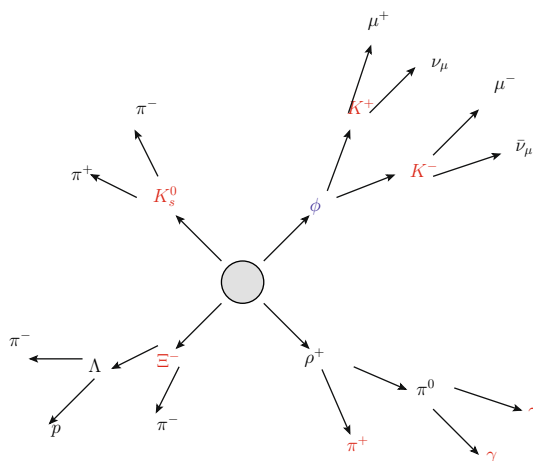
The full event record of the Monte–Carlo event generator cannot be compared directly to experimental results as these are often corrected to a certain definition of primary particles, illustrated in Fig. 3. In the context of the heavy-ion developments, pre-defined Projections have been developed for the primary-particle definitions of various experiments. As an example, the definition by the ALICE experiment [17] has been implemented in the projection `ALICE::PrimaryParticles`, where a particle is considered primary if it:

1. has a proper lifetime of  $\tau \geq 1$  cm/c, and
2. is *either* produced directly in the interaction;
3. *or* through the decay of a particle with mean proper lifetime  $\tau < 1$  cm/c.

This provides a reusable selection of primary particles for the analysis of a given experiment. In addition, this approach allows one to clearly identify cases where a requirement has been relaxed for a specific analysis. For example, in Ref. [18], the  $\phi$  as well as its kaon decay daughters have been considered primary.

## 2.4 Support for multiple runs through reentrant finalize

In order to produce statistically significant results with RIVET analyses, it is often desirable to parallelize the Monte–Carlo generation step over multiple computational units, and merge the individual results into a whole. If the final result includes anything but simple objects such as distributions represented as histograms or mean values, a simple merging will not yield the correct result if the `finalize` method (see Sect. 2.1) has already been executed. Furthermore, a large number



**Fig. 3** Visualization of the experimental primary particle definition by ALICE [17], with primary particles in red, particles not considered primary in black and resonances which can possibly be added in addition to primary particles in blue

of heavy-ion analyses requires a cross-system merging, to construct, for example, the nuclear-modification factors (see Sect. 4), which are ratios of distributions obtained in AA to the same distribution in pp, scaled by an appropriate factor.

For this version of RIVET, both of these requirements have been addressed. This solution is called re-entrant finalize, and is implemented in the script `rivet-merge`. The procedure for re-entrant finalization is as follows:

1. Before `finalize`, all output objects are saved in raw (un-normalized) form.<sup>2</sup>
2. `Finalize` is run as always, and both finalized and raw objects are saved in the normal output. Note that raw objects are written *before* `finalize` is executed, and any modifications done in `finalize` will thus not affect the raw objects.
3. By running `rivet-merge` on one or more produced output files, all used analyses are loaded and all raw objects will be initialized with summed content of all runs to be merged, and `finalize` will be run again.

For this procedure to work, it is important that all information needed for finalizing the analysis is properly booked in the analysis setup. This implies that *only* YODA objects booked in the analysis `init` method can be used in the analysis' `finalize`, thus *excluding* plain C++ member objects (e.g., `int` or `std::string`) as these cannot be made persistent by the framework.

It is worth mentioning that `rivet-merge` can be run in two different modes. If the files to be merged contain different

processes (e.g. one with pp events and another one with AA events for an  $R_{AA}$  analysis) the program is run in default mode. If, on the other hand, the files contain the result from several identical runs, the weight handling is different and the flag `-e` or `--equiv` should be given to indicate this.

## 2.5 Analysis options

Normally, the implementation of a RIVET analysis is done such that the user is left with no possibility to influence or modify analysis logic. In some cases, for example when selecting centrality (see Sect. 3.1), this dogma would severely restrict the usability of the framework: A specific combination of analysis and Monte-Carlo event generator may require a user to inform the analysis about the nature of the generated events. For that purpose, the possibility to implement *analysis options* has been implemented. The value of an analysis option is specified for each individual analysis at run-time, and is then used throughout the analysis pass. Analysis options can have any type that can be constructed from a string.

In practice, the options are given as one or more suffixes to the analysis name, as in

```
rivet -a {analysis name}:{option 1}={value 1}:\
{option 2}={value 2}
```

and the corresponding values for a given option is then available in the analysis class by calling the

`std::string getOption(std::string)` function. It is possible to load the same analysis several times, with different option values, and each instance will then be executed with the corresponding optional values. If analysis options are given to an analysis, the output histogram title(s) will be postfixed with a string `[Opt=Val]` to allow the user to distinguish between histograms generated with different settings.

## 2.6 Event mixing

The technique of event mixing is often used when studying correlations of two (or more) particles. By mixing data from distinct events, one can correct for acceptance and efficiency effects in a data-driven fashion. Data which is compared by RIVET should already be efficiency corrected. However, the effect of limited acceptance on observables needs to be estimated and is dependent on the distributions of particles and therefore on the Monte-Carlo generator used. A simple example is a rectangular acceptance cut (e.g.  $|\eta| < 1$ ). Sampling two particles from a uniform distribution in this region and calculating their pseudorapidity difference, leads to a triangular shape, which has no relation to the underlying physics. While this particular case can be analytically calculated, the distribution has a non-trivial shape as soon as

<sup>2</sup> The raw (un-normalized) histograms are prefaced with the string `RAW` in the output file.

particle production as a function of  $\eta$  is not uniform. Nature, and all state-of-the-art MC generators, falls into the latter category.

The implementation of event mixing in RIVET is based on the usual strategy at LHC experiments [19]. In this approach the studied observable is not only built from pairs of particles from a given event, but also from pairs where the particles are from different events. In these mixed events all physical correlations are presumably not present while those originating from the detector effects remain and are thus estimated in a data-driven way.

One way to illustrate event-mixing analyses is to consider some  $n$ -particle correlation measurement  $\Delta$

$$\Delta(E) = \text{Correlation of } n \text{ particles in } E. \quad (1)$$

The *signal* is then defined as the event-averaged measurement

$$S = \langle \Delta(E_i) \rangle_E, \quad (2)$$

while the *background* is defined as the average over mixed events

$$B = \langle \Delta(E_i \cap E_j) \rangle_E, \quad (3)$$

where  $E_i \cap E_j$  is the random mixture of events  $E_i$  and  $E_j$ . In the final result

$$R = \frac{S}{B}, \quad (4)$$

the random correlations are factored out.

The RIVET implementation allows for booking a projection which supplies the mixed event sample, by storing events in memory. The mixed event sample can be divided into bins based on any event property such that the mixed event sample is taken from events which are similar to the signal event in question. Common use cases are centrality and multiplicity, but it is possible to use other properties, for example event shapes (such as eg. transverse sphericity), should the need be.

Finally the possibility that the Monte–Carlo event generator supplies weighted events is considered. If a sample contains weighted events, the particle content of the mixed sample will be weighted as well, with particles being selected to the mixed sample with a probability given by  $w_e / \sum_i w_i$ , where  $w_e$  is the weight of the event containing the signal, and  $w_i$  are the weights off all events in the ensemble of events that form the mixed event.

### 3 Specialized features for heavy-ion collisions

Heavy-ion analyses commonly make use of concepts often not considered relevant for high energy particle physics and, thus, has been not available in RIVET. In this section two new features motivated by physics requirements are outlined. In Sect. 3.1 the framework for centrality is described, and in Sect. 3.2 the generic framework for flow measurements is briefly introduced, and the implementation outlined.

#### 3.1 Centrality framework

The impact parameter ( $b$ ) is the distance between the centers of the colliding nuclei. Its value determines on average the size and the transverse shape of the interaction region of a heavy-ion collision. Observables' dependence on the geometry of the interaction region, provides insight into the effect of the size and geometry on the underlying physical mechanisms.

Experimentally, centrality classifies collisions on the basis of some measurement (or measurements) which is assumed to depend monotonically<sup>3</sup> on the impact parameter  $b$  of a given collision. For a single measurement  $X$  which is assumed large for small  $b$  and small for large  $b$ , we can define

$$c_{\text{exp}} = \frac{1}{\sigma_{\text{vis}}} \int_X^\infty dX \frac{d\sigma_{\text{vis}}}{dX'} = \frac{1}{N} \int_X^\infty dX' \frac{dN}{dX'}, \quad (5)$$

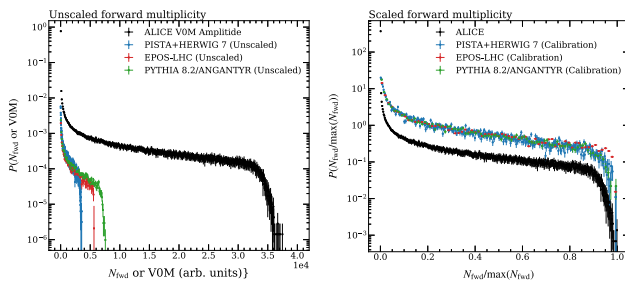
where  $d\sigma_{\text{vis}}/dX$  is the full *measured* distribution of  $X$  over the visible cross-section  $\sigma_{\text{vis}}$  (typically defined in terms of minimum-bias trigger conditions) of the experiment, and  $X$  is the single collision measurement. From a *theoretical* point of view, the centrality of a single collision can be defined as

$$c_b = \frac{1}{\sigma_{\text{inel}}} \int_0^b db' \frac{d\sigma_{\text{inel}}}{db'}. \quad (6)$$

If  $\sigma_{\text{vis}} \approx \sigma_{\text{inel}}$  and  $X(b)$  is reasonably well behaved (smooth, positive, monotonically decreasing, with appropriate limiting behaviour) [20],  $c_b \approx c_{\text{exp}}$ .

Different experiments use different observables for  $X$ . In RIVET, several experimental definitions are available, corresponding to different definitions listed in the following. The ALICE collaboration [21] uses the integrated signal in the V0 scintillator systems covering  $-3.7 < \eta < -1.7$  and  $2.8 < \eta < 5.1$ . The STAR collaboration [22] uses the number of charged tracks in the central pseudorapidity interval spanning  $-0.5 < \eta < 0.5$ , while BRAHMS used the charged-particle multiplicity at mid-rapidity [23, 24]. The ATLAS collaboration [25] uses transverse energy deposited in the forward calorimeters located at  $4.9 < |\eta| < 3.2$ .

<sup>3</sup> The monotonic dependence may be over a limited range, for example in *zero-degree calorimeters*.



**Fig. 4** Comparison of generated centrality measure generated by HERWIG+PISTA, EPOS-LHC and PYTHIA 8.2/ANGANTYR to V0M Amplitude from ALICE. To the left a direct comparison, and to the right a rescaled one (see text)

In central heavy ion collisions  $c_b = c_{\text{exp}}$  to good precision [20]. In peripheral collisions, or in pA and pp collisions the correspondence is less obvious [26]. Moreover, biases introduced by experimental trigger conditions for measurements of  $X$ , are by construction not included in Eq. (6). When possible, it should therefore always be preferred to use Eq. (5) (with appropriate trigger conditions) for estimation of centrality in a simulated event, essentially comparing experimentally obtained  $X$  directly to  $X$  obtained in a Monte-Carlo. By construction, such an approach also includes biases not related to the detector used, but to the definition of  $X$  [27].

In some cases, neither using Eq. (6) nor Eq. (5), is an option. For example, the event generator JEWEL [28] does not calculate the underlying event, but considers only a jet quenched in a medium at a certain impact parameter. Other generators, for example HERWIG+PISTA [9, 29] attempt to generate the forward-going particle multiplicity but do not obtain good agreement with experimental results. In other cases still, it is not possible to compare to the experimentally measured  $X$  without taking into account the detector response. This is for example the case for ALICE, where  $X$  is defined as a detector response (affected by secondary particles), not unfolded to particle level. This latter case is illustrated in Fig. 4, where three soft inclusive approaches (HERWIG+PISTA, EPOS-LHC and PYTHIA 8.2/ANGANTYR) are shown together with centrality data from ALICE. In Fig. 4 (left) the raw detector response is overlaid with the centrality measure implemented in RIVET, and in Fig. 4 (right), a rescaled, normalized version of the centrality measure is shown, comparing to the three MCs. It is clear that direct comparison of MC to experimentally obtained measure (as in Fig. 4 (left)) has no physical meaning at all, since the scale of the compared quantities are different, and in order to refrain from potential errors, this should be avoided in analyses. In such cases, a better choice is to use Eq. 5 on Monte-Carlo and compare the extracted  $c_{\text{exp}}$ . Note, however, that it is important then to implement the appropriate criteria for  $\sigma_{\text{vis}}$  in an equivalent fashion to the experiment compared to. The large excess of experimental measurements at low  $X$  are due to

contamination from electromagnetic interactions. However, as the analyses often ignore this region, it is less of a problem in direct comparisons to Monte-Carlo results. Remnant contributions from such electromagnetic interactions in more central events, are accounted for by appropriate systematic uncertainties on the experimental results.

In consequence, the slicing from data cannot be naively applied to Monte-Carlo, making the calibration step necessary. In total, four possibilities for centrality estimation can be used:

1. Direct comparison to, and binning in, measured  $X$ . This, to some extent, circumvents the notion of centrality altogether, but requires that the simulated and measured  $X$  are roughly equally distributed. In RIVET, this is selected by the analysis option `cent=REF`.
2. Comparison to a user-implemented  $X$ , which should be a particle level, final state observable to get *generator calibrated* centrality. As noted above, it is important to impose the same requirements as used by the experiment to determine  $\sigma_{\text{vis}}$ .<sup>4</sup> In RIVET, this is selected by the analysis option `cent=GEN`.
3. Centrality binning of Monte-Carlo according to Eq. (6) –  $b$  centrality. In RIVET, this is selected by the analysis option `cent=IMP`.
4. Provide a centrality percentile number in the HEPMC heavy-ion record (available from HEPMC 3.0). In RIVET, this is selected by the analysis option `cent=RAW`.

Possibilities 2 and 3 require the user to perform a calibration run, to generate the distribution of either  $X$  or  $b$  and then evaluate the integrals in Eqs. (5) or (6). The result of the calibration pass can then be used in subsequent passes with the analysis calculating the desired centrality-dependent observable. Concrete examples of how to perform such tasks using RIVET, are given in Sect. 4.

### 3.2 Flow measurements

In heavy-ion collisions, the study of azimuthal anisotropy of particle production is used to understand the transport properties of the produced medium, as well as the geometry of the initial state [30]. This is quantified in flow coefficients  $v_n$ 's, obtained by a Fourier expansion of the particle yield with respect to the reaction plane  $\Psi_n$

$$E \frac{d^3 N}{d^3 p} = \frac{1}{2\pi} \frac{d^2 N}{p_T d p_T dy} \left( 1 + 2 \sum_{n=1}^{\infty} v_n \cos([n(\phi - \Psi_n)]) \right), \quad (7)$$

<sup>4</sup> For example, the projection `ALICE::V0AndTrigger` selects the ALICE  $\sigma_{\text{vis}}$ .

where  $E$  is the energy of the particle,  $p_T$  is the transverse momentum,  $\phi$  is the azimuthal angle and  $y$  is the rapidity. As indicated, the coefficients are in general dependent on  $y$  and  $p_T$ . The present RIVET implementation allows one to perform a simple calculation of coefficients differentially with respect to  $p_T$ , as well as integrated coefficients.

Since the reaction plane cannot be determined experimentally, flow coefficients are estimated from two- or multi-particle correlations. For example, using two-particle correlations:

$$\langle v_n^2 \rangle \approx \langle \cos(n(\phi_1 - \phi_2)) \rangle = \langle \exp(in(\phi_1 - \phi_2)) \rangle, \tag{8}$$

where  $\phi_1 = \phi_2$  is excluded from the average. A more advanced technique – with several benefits, which are introduced in the following – calculates multi-particle cumulants from  $Q$ -vectors [31] ( $Q_n = \sum_{k=1}^M w_k \exp(in\phi_k)$  for an event with  $M$  particles),<sup>5</sup> which are combined into  $m$ -particle correlators denoted  $\langle m \rangle_{n_1, n_2, \dots, n_m}$ . Expressing  $m$ -particle correlators in terms of  $Q$ -vectors follows the Generic Framework [32], where the formula for any  $m$ -particle correlator is written as:

$$\langle m \rangle_{n_1, n_2, \dots, n_m} = \frac{N \langle m \rangle_{n_1, n_2, \dots, n_m}}{N \langle m \rangle_{0, 0, \dots, 0}} \quad \text{with} \tag{9}$$

$$N \langle m \rangle_{n_1, n_2, \dots, n_m} = \sum_{\substack{k_1, k_2, \dots, k_m=1 \\ k_1 \neq k_2 \neq \dots \neq k_m}}^M \exp(i(n_1\phi_{k_1} + n_2\phi_{k_2} + \dots + n_m\phi_{k_m})). \tag{11}$$

This expression allows one to easily calculate any correlator, since  $N \langle m \rangle_{n_1, n_2, \dots, n_m}$  can be calculated from  $Q$ -vectors, with the concrete expressions obtained using a recursion relation [32], for any values of  $m$  and  $n_i$ . As an example, for  $m = 2$ , one obtains:

$$N \langle 2 \rangle_{n_1, n_2} = Q_{n_1} Q_{n_2} - Q_{n_1+n_2}. \tag{12}$$

For the special case of  $n_1 = n_2 = n$ , the two-particle correlator directly becomes:

$$\langle 2 \rangle_n = \frac{|Q_n|^2 - M}{M(M-1)}. \tag{13}$$

Evaluation of the product in the numerator  $|Q_n|^2 = \sum_{k,l} \exp(in(\phi_k - \phi_l))$  reveals that the correlator reduces to Eq. (8). However, when evaluating using  $Q$ -vectors, the process requires only a single pass over data, as

<sup>5</sup> In most experimental treatments, the  $Q$ -vectors, and following the correlators are introduced with a particle weight, to correct for detector inefficiencies. Since RIVET compares Monte-Carlo, which suffers no detector inefficiencies, to corrected data, the weight is left out in this summary. The implementation does, however, allow for weights.

opposed to nested passes. Without the use of  $Q$ -vectors, the complexity of an  $m$ -particle correlator is obviously  $O(M^m)$  while the above approach is at most  $O(M \log M)$ .<sup>6</sup> In a heavy-ion event where  $M$  can be 10,000 tracks, this is absolutely crucial. From the correlators, a large number of flow observables can be defined. In most analyses, cumulants  $c_n\{m\}$  (where curly braces represent the order of the  $m$ -particle correlation) are calculated from correlators with  $n_1 = n_2 = \dots = n_m = n$ . The most common use cases are implemented along with uncertainty estimation (see below):

$$c_n\{2\} = \langle \langle 2 \rangle \rangle_{n,-n} \tag{14}$$

$$c_n\{4\} = \langle \langle 4 \rangle \rangle_{n,n,-n,-n} - 2 \langle \langle 2 \rangle \rangle_{n,-n}^2 \tag{15}$$

$$c_n\{6\} = \langle \langle 6 \rangle \rangle_{n,n,n,-n,-n,-n} - 9 \langle \langle 2 \rangle \rangle_{n,-n} \langle \langle 4 \rangle \rangle_{n,n,-n,-n} + 12 \langle \langle 2 \rangle \rangle_{n,-n}^3 \tag{16}$$

$$c_n\{8\} = \langle \langle 8 \rangle \rangle_{n,n,n,n,-n,-n,-n,-n} - 16 \langle \langle 6 \rangle \rangle_{n,n,n,-n,-n,-n,-n} \langle \langle 2 \rangle \rangle_{n,-n} - 18 \langle \langle 4 \rangle \rangle_{n,n,-n,-n}^2 + 144 \langle \langle 4 \rangle \rangle_{n,n,-n,-n} \langle \langle 2 \rangle \rangle_{n,-n}^2 - 144 \langle \langle 2 \rangle \rangle_{n,-n}^4. \tag{17}$$

The double averages denote an all-event average of event averages. Even higher-order cumulants, or other combinations of correlators, such as Symmetric Cumulants [33] can be easily added by the user, as the framework allows the user to fully access the correlators. The cumulants are finally transformed into flow coefficients

$$v_n\{m\} = (k_{m/2} c_n\{m\})^{\frac{1}{m}}, \tag{18}$$

where the coefficient  $k_m$  is written here for convenience for  $m < 4$ :

$$k_m = \left\{ 1, -1, \frac{1}{4}, -\frac{1}{33} \right\}. \tag{19}$$

Differential flow observables are constructed using an additional correlator, limited to a certain phase space. In the case of the implementation in RIVET, this is restricted to observables differential in  $p_T$ . The differential correlator can be constructed in a similar fashion [32] as the integrated ones in Eq. (9), but the transformation to differential flow coefficients differs. Denoting the differential  $m$ -particle correlator  $\langle m' \rangle$ , the differential two-particle cumulant and the corresponding differential flow coefficient are given by:

$$d_n\{2\} = \langle \langle 2' \rangle \rangle \quad \text{and} \quad v'_n\{2\} = \frac{d_n\{2\}}{\sqrt{c_n\{2\}}}, \tag{20}$$

where  $c_n\{2\}$  is a reference cumulant calculated for the full phase space as above. In the same way, the differential four-

<sup>6</sup> Note, however, that the evaluation of the  $Q$ -vectors themselves are generally  $O(mM)$ , and considerable computing time is spent building the event  $Q$ -vectors.

particle cumulant and the corresponding differential flow coefficient are:

$$d_n\{4\} = \langle\langle 4' \rangle\rangle - 2\langle\langle 2 \rangle\rangle\langle\langle 2' \rangle\rangle \quad \text{and} \quad v'_n\{4\} = -\frac{d_n\{4\}}{(-c_n\{4\})^{3/4}}. \tag{21}$$

The statistical uncertainty on cumulants and flow coefficients is calculated by a simple variance method, often denoted as bootstrap [34] method in experimental literature. Instead of propagating the sampling uncertainty on the individual correlators through the (long) expressions for cumulants and flow coefficients, the calculation is internally split up into sub-samples.<sup>7</sup> The default behaviour is to display the square root of the sample variance as statistical uncertainty, but the user can choose to use the envelope instead.

### 3.2.1 Suppressing non-flow with an $\eta$ -gap

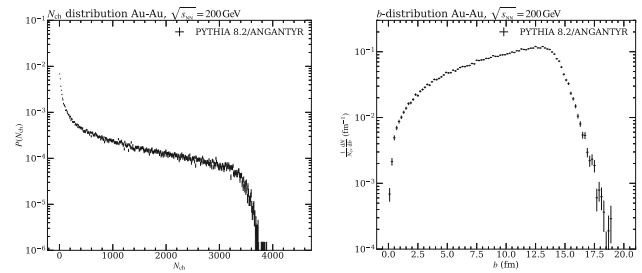
The approximate equality in Eq. (8) becomes a full equality in the absence of non-flow effects, which are correlations arising from other sources than typically attributed to collective expansion [35]. Non-flow effects can arise from several sources, such as jet production or resonance decay, and are few-particle correlations as opposed to flow which is correlation of all particles to a common symmetry plane. The idea behind the  $\eta$ -gap or sub-event method is to suppress sources of non-flow which are local in (pseudo)rapidity. The non-flow is then suppressed by requiring that particle measurements contributing to a correlator *must* come from different parts of the detector acceptance, usually separated in pseudorapidity. Using Eq. (8), an  $\eta$ -gap can be implemented simply by requiring that the two particles come from two different sub-events, here labeled *A* and *B*:

$$\langle v_n^2 \rangle \approx \langle \exp(in(\phi_1^A - \phi_2^B)) \rangle. \tag{22}$$

Translating to *Q*-vectors, it is first noted that  $\phi_1 = \phi_2$  does not need to be excluded from the average in this case, as it is excluded by construction. Equation (12) reduces to  $N\langle 2 \rangle_{n_1, n_2} = Q_{n_1}^A Q_{n_2}^A$ , where superscripts *A* and *B* indicates that *Q*-vectors are calculated in that specific sub-event only, as  $Q_n^X = \sum_{k=1}^{M^X} \exp(in\phi_k^X)$ . This procedure is generalized to multi-particle correlations, by removing terms from the correlator where cross-terms do not share a common sub-event.

In RIVET, correlators with two sub-events are implemented, with the possibility of calculating both integrated and differential flow.

<sup>7</sup> An alternative to the experimental practise of using “bootstrap” could be replaced by error propagation as outlined in ref. [32], and its reference implementation. In RIVET, the choice is taken to stay as close to the performed experimental analyses as possible.



**Fig. 5** Centrality observables from the analyses BRAHMS\_2004\_AUAUCentrality with PYTHIA 8.2/ANGANTYR model [26] displaying the generated calibration curve (left) and the impact parameter calibration curve (right)

## 4 Examples of implemented heavy-ion analyses

In this section, some of the implemented RIVET analyses using new features are presented, along with detailed instructions for a user to run the analyses and plot the resulting figures.

The *centrality framework* introduced in Sect. 3.1 is used by many analyses, each one implementing the centrality estimator of the respective experiment. In these cases the analysis implementation relies on an analysis for centrality calibration, called e.g.,:

ALICE\_2015\_PBPBCentrality, BRAHMS\_2004\_AUAUCentrality, and STAR\_BES\_CALIB.

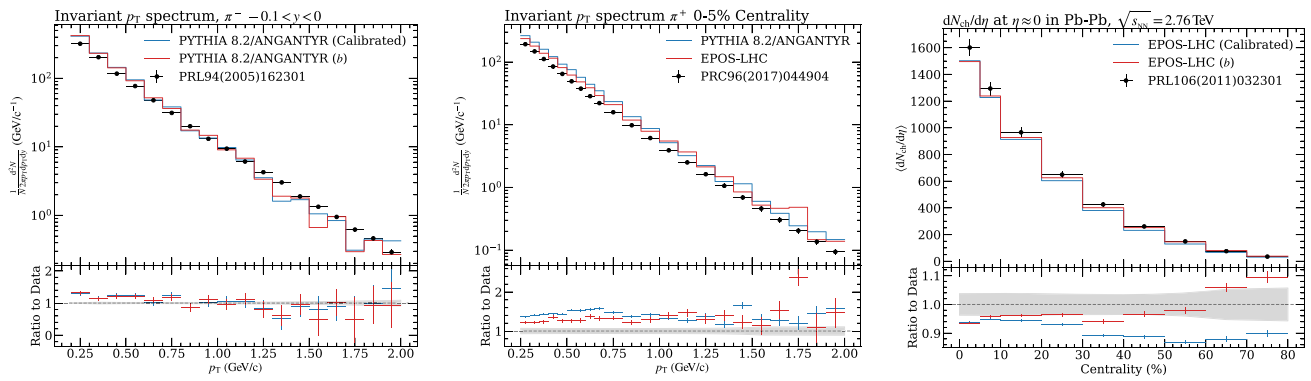
These can be reused by many analyses sharing the same centrality estimator.

Executing the centrality calibration analysis fills histograms for each defined centrality estimator, for example the impact parameter and the number of charged particles or their energy in the rapidity region where the relevant detector is situated. Figure 5 exemplifies the calibration for BRAHMS, for which an unfolded measure does not exist, and the options 2 or 3 (GEN and IMP) must be used.

In order to use the generated calibration histograms in the actual analyses, they must be pre-loaded when running the analysis code. The analysis will then calculate the percentile of the calibration curve for the value determined in the newly generated event and, thus, allow the user to retrieve the centrality for the current event.

Figure 6 presents results from BRAHMS, STAR and ALICE, using separate centrality calibrations. Comparison to PYTHIA 8.2/ANGANTYR and EPOS-LHC is performed. In the considered analysis by BRAHMS (BRAHMS\_2004\_I647076 [23]), only the 0–5% centrality bin is used, meaning that all other events are vetoed by the analysis code. The centrality definition is selected at runtime using, in this case, either the analysis option (see Sect. 2.5) GEN (for generator calibrated) or IMP (for impact parameter). In the STAR analysis STAR\_2017\_I1510593





**Fig. 6** Results from the analyses BRAHMS\_2004\_I647076 (left; invariant  $p_T$  of  $\pi^-$  in  $0 < y < 0.1$ ), STAR\_2017\_I151059 (middle; invariant  $p_T$  for  $K^+$  in  $|y| < 0.1$ ) and ALICE\_2010\_I880049 (right;  $dN_{ch}/d\eta$  vs. centrality) utilizing centrality calibration

[36], several centrality bins are analysed. The ALICE analysis ALICE\_2010\_I880049 [37] calibrates using the centrality analyses presented in Sect. 3.1 (cf. Fig. 4).

To reproduce those results based on a HEPMC file containing AA collisions generated by the considered generators, the following steps need to be followed for the example of the BRAHMS analysis:

1. Run the calibration analysis

```
rivet events.hepmc \
-a BRAHMS_2004_AUAUCentrality
-o calibration.yoda
```

2. Plot the calibration results from Fig. 5

```
rivet-mkhtml calibration.yoda
```

3. Run the analysis with the calibration file pre-loaded, for both centrality selections:

```
rivet events.hepmc -p calibration.yoda \
-a BRAHMS_2004_I647076:cent=IMP \
-a BRAHMS_2004_I647076:cent=GEN \
-o result
```

4. Plot the analysis result<sup>8</sup> from Fig. 6

```
rivet-mkhtml result.yoda
```

Note that in this case (and in fact in all currently implemented heavy ion analyses), the calibration and analysis steps are carried out on the same event sample. This would not be the

<sup>8</sup> Note that the figures in this paper are produced using Python/Matplotlib, to allow for rebinning and rescaling of results for illustrative purposes. Equivalent figures (apart from those requiring such steps) can be produced using `rivet-mkhtml`.

case for a signal analysis considering a signal process (eg. Z-production). Here the calibration step would be carried out on a minimum bias sample, and the signal step on a signal sample.

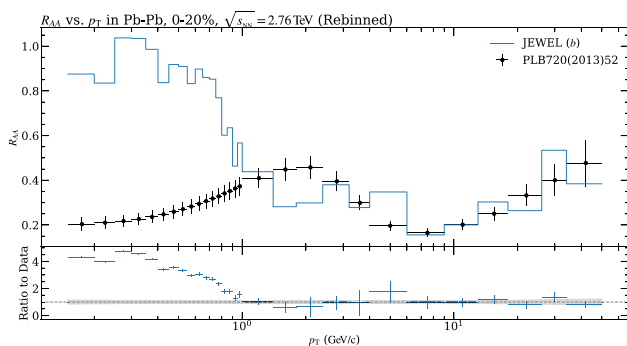
The real power of RIVET analyses is, however, not to compare a single generator at a time, but the ability to compare several generators to the same experimental data at once. This use case is shown in Fig. 6 (middle), where both EPOS-LHC and PYTHIA 8.2/ANGANTYR are compared to Au–Au data at  $\sqrt{s_{NN}} = 27.6$  GeV obtained by STAR [36]. In this case, the calibration must be run for both generators, as the output will in general be different. The shown figure can thus be obtained by slightly extending the workflow outlined above:

```
rivet epos-events.hepmc -a STAR_BES_CALIB \
-o epos-calib
rivet pythia-events.hepmc -a STAR_BES_CALIB \
-o pythia-calib
rivet epos-events.hepmc -p epos-calib.yoda \
-a STAR_2017_I1510593 -o epos-result
rivet pythia-events.hepmc -p pythia-calib.yoda \
-a STAR_2017_I1510593 -o pythia-result
rivet-mkhtml epos-result.yoda pythia-result.yoda
```

Note from the last line that comparison plots can be produced directly by the `rivet-mkhtml` script. Such comparisons are performed using large scale voluntary computing resources as part of the MCPLOTS project [38]. MCPLOTS is currently being updated and extended to include heavy-ion results and Monte–Carlo by members of the ALICE collaboration.<sup>9</sup>

The feature `reentrant_finalize` introduced in Sect. 2.4, is for example used in the analysis ALICE\_2012\_I1127497. The main observable is the nuclear-modification factor  $R_{AA}$  defined as the ratio of the yield in AA collisions over pp collisions scaled by an appropriate factor  $N_{coll}$  to accommodate

<sup>9</sup> The original MCPLOTS can be found at <http://mcplots.cern.ch>, and a preview of the heavy-ion update at <http://mcplots-alice.cern.ch>.



**Fig. 7** Nuclear-modification factor  $R_{AA}$  measured by ALICE [39] compared to JEWEL model calculations. The region below 1 GeV/c is not attempted to be modelled correctly in JEWEL resulting in the large discrepancy. Note, that data and MC have been rebinned to allow a reasonable comparison

for the trivial scaling of binary collisions [16]:

$$R_{AA} = \frac{d^2 N_{ch}/d p_T dy|_{AA}}{N_{coll} d^2 N_{ch}/d p_T dy|_{pp}} \quad (23)$$

In order to compute this quantity, RIVET is executed three times, first for a generator with AA settings, second with pp settings, and a third and final time to compute the ratio. A comparison of this observable to a calculation with the JEWEL generator is shown in Fig. 7, and can be performed with the following steps, given generated HEPMC files of pp and AA events by JEWEL.

1. The desired centrality range is set when events are generated in JEWEL. The centrality value as well as the impact parameter are saved in the HepMC output. The usage of the raw centrality values is supported only from HepMC 3.0 onwards (see Sect. 3.1), therefore the impact parameter centrality calibration is needed to be applied<sup>10</sup> after events have been generated in the full centrality interval (0–100%):

```
rivet aa-sample.hepmc --ignore-beams \
-a ALICE_2015_PBPBCentrality \
-o calibration.yoda
```

2. The RIVET analysis is run on the pp sample:

```
rivet pp-sample.hepmc --ignore-beams \
-a ALICE_2012_I1127497 -o jewel-pp.yoda
```

<sup>10</sup> Note, that JEWEL produces a single nucleon–nucleon collision instead of a full AA collision, and therefore the events are marked as being pp, np, or nn collisions. As the RIVET routine expects pp or AA events, the `ignore-beams` flag needs to be used.

3. Next the RIVET analysis is run on the AA sample. Note the beam option signifying that this is an AA run. Since the JEWEL event generator does not supply this information in the HEPMC file, but denotes all beams as pp beams, the analysis requires this additional information:

```
rivet aa-sample.hepmc --ignore-beams \
-p calibration.yoda \
-a ALICE_2012_I1127497:beam=HI:cent=IMP \
-o jewel-aa.yoda
```

4. The two samples must be merged, and the ratio plots constructed, with `rivet-merge`, which loads histograms from the two runs above, and runs finalize on the preloaded histograms:

```
rivet-merge --merge-option beam \
--merge-option cent jewel-pp.yoda \
jewel-aa.yoda -o jewel-merged.yoda
```

The optional argument `merge-option` removes the analysis options `beam` and `cent` from the heavy ion sample.

5. The merged sample can be plotted. There is no need to plot the unmerged histograms, as all histograms are contained in the merged file.

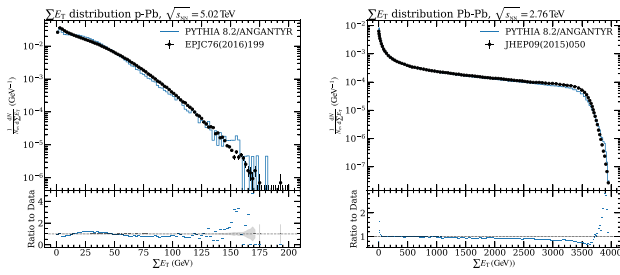
```
rivet-mkhtml jewel-merged.yoda
```

## 5 Understanding biases in simulation and data

Apart from direct validation of a given Monte–Carlo event generator, the analysis capabilities of RIVET are useful as tools to reveal inherent biases in representation of data. In this section the use of the physics motivated extensions presented in Sect. 3 are exemplified by presenting short analyses of centrality biases introduced by using different centrality observables for theory and data (Sect. 5.1) and from inclusion of non-flow effects in flow measurements (Sect. 5.2). These two effects are well known, and chosen to show the capabilities of the framework, rather than presenting an independent physics point.

### 5.1 Centrality observables

In pA collisions, such as those performed at the LHC with p–Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV, and recorded by the ATLAS experiment [40], the notion of centrality is used as in AA collisions, shown in Sect. 4, but without the clear relation to the physical impact parameter. Since the physical impact parameter is well defined and useful from a theoretical point



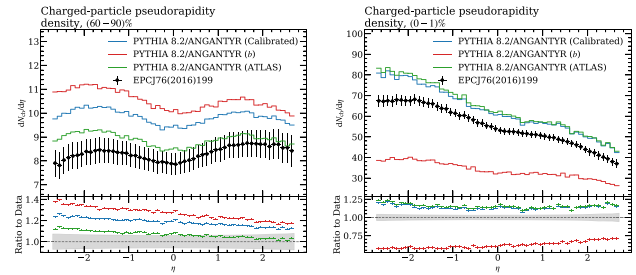
**Fig. 8** Centrality measure by ATLAS for p–Pb at  $\sqrt{s_{NN}} = 5.02$  TeV [40] (left) and Pb–Pb at  $\sqrt{s_{NN}} = 2.76$  TeV [41] (right), compared to PYTHIA 8.2. Note that data points are not unfolded, and taken from the figures in the paper

of view (as it can be used to infer properties of a created QGP phase), the correlation between modelled impact parameter and measured centrality is worth having direct access to, as the effects of possible differences between the two in a measurement is important to understand.

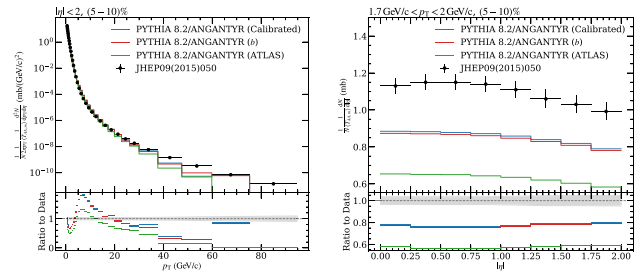
The RIVET centrality implementation offers the possibility to study such effects directly in specific simulations, by comparing results from measured centrality to Monte–Carlo using different centrality definitions. Consider the centrality measure used by ATLAS, shown in Fig. 8, which is the  $\sum E_T$  in the forward (lead) going direction [40], overlaid with a comparison to PYTHIA 8.2/ANGANTYR. This can be contrasted with the similar centrality measure used by ATLAS in Pb–Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV [41], which is  $\sum E_T$  in both forward and backward directions. For both centrality measures, it should be noted that data points are read-off from the paper. Furthermore the distributions are not unfolded. This adds an unknown source of uncertainty to the direct comparison to the centrality, as well as to all derived results, when the experimental centrality estimate is used.

In Fig. 9, the  $dN_{ch}/d\eta$  distributions for the three different centrality selections, experimental reference calibration (Experiment, option REF), generator calibrated (Calibrated, option GEN), and impact parameter ( $b$ , option IMP) in p–Pb are shown for peripheral (60–90%) (left) and central (0–1%) (right) events [40]. We see a small, but systematic, overestimate of the charged-particle pseudorapidity density when using either generator or experimentally calibrated centrality estimators. The impact parameter based centrality estimator has overall larger deviations from experimental results (also in centrality classes not shown here, see Ref. [26] for more comparisons), but without any clear systematics in the deviations.

In Fig. 10 the charged-particle multiplicity in Pb–Pb differential in  $p_T$  (left) and  $\eta$  for  $1.7 \text{ GeV}/c < p_T < 2.0 \text{ GeV}/c$  (right) in the 5–10% centrality bin are shown [41], again for the three different centrality estimators. Here, the results from the generated and impact parameter centrality estimators fully coincides, thus in agreement with the assumption



**Fig. 9** Charged-particle pseudorapidity density as a function of pseudorapidity in p–Pb at  $\sqrt{s_{NN}} = 5.02$  TeV for peripheral (60–90%) (left) and central (0–1%) (right) events [41], compared to PYTHIA 8.2



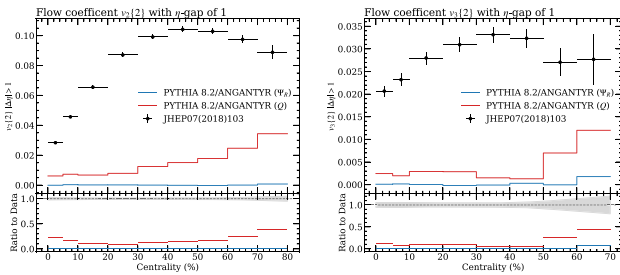
**Fig. 10** Charged-particle multiplicity as function of  $p_T$  (left) and  $\eta$  for  $1.7 \text{ GeV}/c < p_T < 2.0 \text{ GeV}/c$  (right) in Pb–Pb at  $\sqrt{s_{NN}} = 2.76$  TeV, in the 5–10% centrality bin [41], compared to PYTHIA 8.2

that the measured centrality (Eq. (5)) coincides with theoretical centrality (Eq. (6)) in AA collisions. It is also seen that comparison using the measured experimental centrality measure, deviates from the two others, as expected.

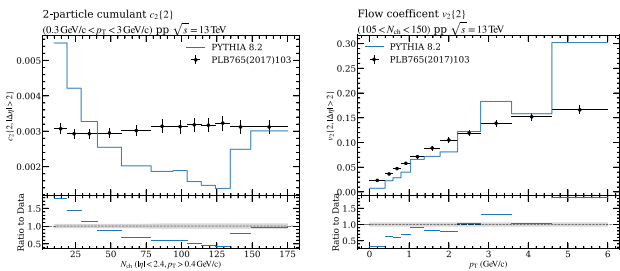
### 5.2 Non-flow effects

The  $v_2$  and  $v_3$  measured by ALICE in Pb–Pb collisions, at  $\sqrt{s_{NN}} = 2.76$  TeV as a function of centrality, is shown in Fig. 11 together with comparisons to PYTHIA 8.2/ANGANTYR simulations [42]. The simulations were obtained using two different approaches. First the reaction plane method, where Eq. (7) is evaluated directly, giving  $\langle v_n \rangle = \langle \cos(n(\phi - \Psi)) \rangle$ . The used value for the reaction plane angle  $\Psi$ <sup>11</sup> can be evaluated in simulations, but not in data. Secondly the Generic Framework (GF) implementation in RIVET, described in Sect. 3.2, is used to estimate  $v_2$  and  $v_3$  with two-particle correlations as it is done in the experiment. In the former case, the PYTHIA 8.2 simulations are consistent with 0. This is expected, as the typical mechanisms associated with the non-flow (jets, resonance decays) are not correlated with the reaction plane, and thus any coincidental correlations are averaged out. On the other hand, this is not true for the simplest case of two-particle correlations. In particular, back-to-back jets give rise to strong correlations that are unrelated

<sup>11</sup>  $\Psi$  is the the angle between the impact parameter vector and the beam axis.



**Fig. 11** Flow coefficients  $v_2\{2\}$  (left) and  $v_3\{2\}$  (right) in Pb–Pb collisions, at  $\sqrt{s_{NN}} = 2.76$  TeV as a function of centrality, evaluated with the reaction plane method ( $\Psi_R$ ) and the Generic Framework ( $Q$ ) implementation in RIVET. Data from ALICE [42] compared to PYTHIA 8.2



**Fig. 12** Example flow observables from pp collisions [51], compared to simulation with PYTHIA 8.2. Analysis performed with the implementation of the Generic Framework for flow measurements [32] presented in Sect. 3.2

to the hydrodynamical evolution of the system. As a result, PYTHIA 8.2 predicts non-zero  $v_2$  and  $v_3$  values calculated with two-particle correlations, as seen in Fig. 11. We note that the non-flow effects are more pronounced for  $v_2$  and increase for peripheral Pb–Pb collisions, where the  $dN_{ch}/d\eta$  is lower and therefore the different non-flow sources get less averaged out. It is noted that the non-flow contribution is here shown to be sizeable for  $v_2\{2\}$  even with a required minimum separation in  $\eta$  between the two correlated particles.

## 6 Collective effects in pp collisions

The discovery of heavy-ion-like effects in pp collisions has spurred a lot of interest from the proton-proton Monte–Carlo authors to introduce models for such effects [43–50]. The inclusion of heavy-ion features in RIVET will aid this development, as such features can be used directly in new analyses for pp as well. As an example, the seminal results from the CMS experiment on two- and multi-particle correlations [51] and from the ALICE experiment on strangeness enhancement [5] has been implemented in RIVET analyses CMS\_2017\_I1471287 and ALICE\_2016\_I1471838. Results from the former shown in Fig. 12.

The analysis is implemented with the flow framework presented in Sect. 3.2. In Fig. 12 the cumulant  $c_2\{2\}$  as a func-

tion of charged-particle multiplicity (left, Eq. (14)), and the corresponding flow coefficient as a function of transverse momentum (right) are shown.

## 7 Conclusion and future work

Extensions to the RIVET framework to facilitate comparisons between Monte–Carlo event generators and data have been presented. The RIVET framework is a standard tool for preservation of LHC analyses of proton–proton collisions, and with the presented extensions, the framework is ready to start taking a similar role for heavy-ion physics.

There are, however, areas where the updated framework is foreseen to be further extended in order to allow for better comparisons to jet physics studies, reproduction of global fits of QGP properties, as well as allowing for comparison of simulation tools developed for heavy-ion physics.

One avenue of further development is jet observables. RIVET includes an interface to the popular FastJet package [52] for jet reconstruction, which already includes some functionality for sub-jet observables geared towards heavy-ion collisions [53]. A more substantial problem for preservation of jet analyses is that of handling the large “underlying event” of a heavy-ion collision. A preliminary RIVET implementation to perform background subtraction and QGP back-reactions for specific models is available [54], but no truly model independent prescription exist. Due to large differences between models, it may even be such that experimental background subtraction cannot be fully reproduced for models unless the model also attempts to describe the full underlying event. In such case a better option might be to publish data without background subtraction for model comparisons.

Another future direction involves expanding the possibility for post-processing beyond what re-entrant finalization allows. Specifically to add the possibility to perform fits used in experimental analyses, such as for example reinterpretation in terms of simple thermal models like Boltzmann–Gibbs Blast Wave for AA [55] or Lévy–Tsallis for pp [22,56]. A technically similar development for high energy physics is realized in the CONTUR package [57], which constrains BSM physics using Standard Model measurements implemented in RIVET. Reinterpretation of heavy-ion analyses could be performed in a technically similar way, possibly also as an additional software package. Such an extension could also cater to theoretical models not suitable for direct event-by-event analysis.

**Acknowledgements** We thank the Discovery Center at the Niels Bohr Institute for hosting a successful workshop in August 2018, consolidating the efforts described in this paper, and in particular the NBI ALICE group for providing organizational support. The workshop was further supported by the Marie Skłodowska-Curie Innovative Training

Network MCnetITN3 (Grant Agreement 722104). The authors would further like to acknowledge support from individual funding bodies: CB was supported by the Swedish Research Council, Contract Number 2017-003. AB was supported by The Royal Society via University Research Fellowship Grant UF160548. COR and LL were supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 668679), and in part by the Swedish Research Council, Contract Number 2016-05996. LL was supported in part by the Swedish Research Council, Contract Number 2016-03291, and the Knut and Alice Wallenberg foundation, contract no 2017.0036. PaKi acknowledges support by the BMBF under Grant Number 05H18VKCC1.

**Data Availability Statement** This manuscript has no associated data or the data will not be deposited. [Authors' comment: Program core code, analysis plugins as well as reference data is available at <https://rivet.hepforge.org/>.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP<sup>3</sup>.

## A Analyses with heavy ion features

This appendix contains a list of all RIVET analyses implementing one or more heavy ion features, or heavy ion beams, at the time of writing, shown in Table 1. It includes a mention of features used, as a handy guide to anyone looking for inspiration to implement a specific feature into their analyses. An up-to-date list of all RIVET analyses can be found at <https://rivet.hepforge.org/>.

**Table 1** All RIVET analyses implementing one or more heavy ion features, or heavy ion beams. An up-to-date list can be found at <https://rivet.hepforge.org/>

Analysis name	System	$\sqrt{s_{NN}}$	Heavy ion features	Reference
ALICE_2010_I880049	Pb–Pb	2.76 TeV	CCentrality, primary particles, heavy ion container	[37]
ALICE_2012_I930312	Pb–Pb	2.76 TeV	Centrality, heavy ion container, re-entrant finalize	[58]
ALICE_2012_I1127497	Pb–Pb	2.76 TeV	Centrality, heavy ion container, re-entrant finalize	[39]
ALICE_2012_I1126966	Pb–Pb	2.76 TeV	Centrality, primary particles	[59]
ALICE_2013_I1225979	Pb–Pb	2.76 TeV	Centrality, primary particles	[60]
ALICE_2014_I1243865	Pb–Pb	2.76 TeV	Centrality, primary particles	[61]
ALICE_2014_I1244523	p–Pb	5.02 TeV	Centrality, primary particles	[62]
ALICE_2016_I1394676	Pb–Pb	2.76 TeV	Centrality, primary particles	[63]
ALICE_2016_I1419244	Pb–Pb	5.02 TeV	Centrality, generic framework	[64]
ALICE_2016_I1507090	Pb–Pb	5.02 TeV	Centrality, primary particles	[65]
ALICE_2016_I1507157	pp	7 TeV	Event mixing	[66]
ATLAS_2015_I1360290	Pb–Pb	2.76 TeV	Centrality	[41]
ATLAS_2015_I1386475	p–Pb	5.02 TeV	Centrality	[40]
ATLAS_PBPB_CENTRALITY	Pb–Pb	2.76 TeV	Centrality	[41]
ATLAS_pPb_Calib	p–Pb	5.02 TeV	Centrality	[40]
BRAHMS_2004_I647076	Au–Au	200 GeV	Centrality, primary particles	[23]
CMS_2017_I1471287	pp	13 TeV	Generic framework	[51]
LHCF_2016_I138587	p–Pb	5.02 TeV	–	[67]
STAR_2016_I1414638	Au–Au	7.7 – 200 GeV	Centrality	[68]
STAR_2017_I1510593	Au–Au	7.7 – 39 GeV	Centrality	[36]

## References

1. A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk, H. Schulz, F. Siegert, Rivet user manual. *Comput. Phys. Commun.* **184**, 2803–2819 (2013)
2. J. Schukraft, QM2017: status and key open questions in ultra-relativistic heavy-ion physics. *Nucl. Phys. A* **967**, 1–10 (2017)
3. J.H. Putschke et al., The JETSCAPE framework (2019)
4. V. Khachatryan et al., Observation of long-range near-side angular correlations in proton-proton collisions at the LHC. *JHEP* **09**, 091 (2010)
5. J. Adam et al., Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. *Nat. Phys.* **13**, 535–539 (2017)
6. Z. Citron et al., Report from Working Group 5, in *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC* ed. by A. Dainese, M. Mangano, A.B. Meyer, A. Nisati, G. Salam, M.A. Vesterinen, pp. 1159–1410 (2019)
7. C. Bierlich et al., Robust Independent Validation of Experiment and Theory: Rivet version 3 (2019)
8. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191**, 159–177 (2015)
9. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C* **76**(4), 196 (2016)
10. G. Bíró, G. G. Barnaföldi, G. Papp, M. Gyulassy, P. Lévai, X.-N. Wang, B.-W. Zhang, Introducing HIJING++: the heavy ion Monte Carlo Generator for the high-luminosity LHC Era. *PoS Hard-Probes2018*, 045 (2019)
11. P. Skands, S. Carrazza, J. Rojo, Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, **74**(8), 3024 (2014)
12. V. Khachatryan et al., Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C* **76**(3), 155 (2016)
13. Summary of ATLAS Pythia 8 tunes (2012)
14. M. Dobbs, J.B. Hansen, The HepMC C++ Monte Carlo event record for high energy physics. *Comput. Phys. Commun.* **134**, 41–46 (2001)
15. R.J. Glauber, Cross-sections in deuterium at high-energies. *Phys. Rev.* **100**, 242–248 (1955)
16. M.L. Miller, K. Reygers, S.J. Sanders, P. Steinberg, Glauber modeling in high energy nuclear collisions. *Ann. Rev. Nucl. Part. Sci.* **57**, 205–243 (2007)
17. S. Acharya et al., The ALICE definition of primary particles (2017)
18. S. Acharya et al., Multiplicity dependence of light-flavor hadron production in pp collisions at  $\sqrt{s} = 7$  TeV. *Phys. Rev. C* **99**(2), 024906 (2019)
19. S. Chatrchyan et al., Observation of long-range near-side angular correlations in proton-lead collisions at the LHC. *Phys. Lett. B* **718**, 795–814 (2013)
20. S.J. Das, G. Giacalone, P.-A. Monard, J.-Y. Ollitrault, Relating centrality to impact parameter in nucleus-nucleus collisions. *Phys. Rev. C* **97**(1), 014905 (2018)
21. B. Abelev et al., Centrality determination of Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with ALICE. *Phys. Rev. vol. C* **88**(4), 044909 (2013)
22. B.I. Abelev et al., Strange particle production in p+p collisions at  $s^{*(1/2)} = 200$ -GeV. *Phys. Rev. C* **75**, 064901 (2007)
23. I.G. Bearden et al., Charged meson rapidity distributions in central Au+Au collisions at  $s(\text{NN})^{*(1/2)} = 200$ -GeV. *Phys. Rev. Lett.* **94**, 162301 (2005)
24. M. Adamczyk et al., The BRAHMS experiment at RHIC. *Nucl. Instrum. Methods A* **499**, 437–468 (2003)
25. G. Aad et al., Measurement of the pseudorapidity and transverse momentum dependence of the elliptic flow of charged particles in lead-lead collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with the ATLAS detector. *Phys. Lett. B* **707**, 330–348 (2012)
26. C. Bierlich, G. Gustafson, L. Lönnblad, H. Shah, The Angantyr model for heavy-ion collisions in PYTHIA8. *JHEP* **10**, 134 (2018)
27. J. Adam et al., Centrality dependence of particle production in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. *Phys. Rev. C* **91**(6), 064905 (2015)
28. K.C. Zapp, JEWEL 2.0.0: directions for use. *Eur. Phys. J. C* **74**(2), 2762 (2014)
29. J. Bellm, C. Bierlich, PISTA: posterior ion STACKing (2018)
30. J.-Y. Ollitrault, Anisotropy as a signature of transverse collective flow. *Phys. Rev. D* **46**, 229–245 (1992)
31. A. Bilandzic, R. Snellings, S. Voloshin, Flow analysis with cumulants: direct calculations. *Phys. Rev. C* **83**, 044913 (2011)
32. A. Bilandzic, C.H. Christensen, K. Gulbrandsen, A. Hansen, Y. Zhou, Generic framework for anisotropic flow analyses with multiparticle azimuthal correlations. *Phys. Rev. C* **89**(6), 064904 (2014)
33. J. Adam et al., Correlated event-by-event fluctuations of flow harmonics in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Rev. Lett.* **117**, 182301 (2016)
34. L. Wasserman, All of Statistics. Springer Texts in Statistics, 1st edn. Springer, New York (2004)
35. S.A. Voloshin, A.M. Poskanzer, R. Snellings, Collective phenomena in non-central nuclear collisions. *Landolt-Bornstein* **23**, 293–333 (2010)
36. L. Adamczyk et al., Bulk properties of the medium produced in relativistic heavy-ion collisions from the beam energy scan program. *Phys. Rev. C* **96**(4), 044904 (2017)
37. K. Aamodt et al., Centrality dependence of the charged-particle multiplicity density at mid-rapidity in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Rev. Lett.* **106**, 032301 (2011)
38. A. Karneyeu, L. Mijovic, S. Prestel, P.Z. Skands, MCPLOTS: a particle physics resource based on volunteer computing. *Eur. Phys. J. C* **74**, 2714 (2014)
39. B. Abelev et al., Centrality dependence of charged particle production at large transverse momentum in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Lett. B* **720**, 52–62 (2013)
40. G. Aad et al., Measurement of the centrality dependence of the charged-particle pseudorapidity distribution in proton-lead collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the ATLAS detector. *Eur. Phys. J. C* **76**(4), 199 (2016)
41. G. Aad et al., Measurement of charged-particle spectra in Pb+Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with the ATLAS detector at the LHC. *JHEP* **09**, 050 (2015)
42. S. Acharya et al., Energy dependence and fluctuations of anisotropic flow in Pb-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  and 2.76 TeV. *JHEP* **07**, 103 (2018)
43. C. Bierlich, G. Gustafson, L. Lönnblad, Collectivity without plasma in hadronic collisions. *Phys. Lett. B* **779**, 58–63 (2018)
44. S. Gieseke, P. Kirchgaeßer, S. Plätzer, Baryon production from cluster hadronisation. *Eur. Phys. J. C* **78**(2), 99 (2018)
45. C.B. Duncan, P. Kirchgaeßer, Kinematic strangeness production in cluster hadronization. *Eur. Phys. J. C* **79**(1), 61 (2019)
46. S. Gieseke, P. Kirchgaeßer, S. Plätzer, A. Siodmok, Colour reconnection from soft gluon evolution. *JHEP* **11**, 149 (2018)
47. C. Bierlich, J.R. Christiansen, Effects of color reconnection on hadron flavor observables. *Phys. Rev. D* **92**(9), 094010 (2015)
48. C. Bierlich, G. Gustafson, L. Lönnblad, A. Tarasov, Effects of overlapping strings in pp collisions. *JHEP* **03**, 148 (2015)
49. J.R. Christiansen, P.Z. Skands, String formation beyond leading colour. *JHEP* **08**, 003 (2015)
50. N. Fischer, T. Sjöstrand, Thermodynamical string fragmentation. *JHEP* **01**, 140 (2017)
51. V. Khachatryan et al., Evidence for collectivity in pp collisions at the LHC. *Phys. Lett. B* **765**, 193–220 (2017)

52. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. *Eur. Phys. J. C* **72**, 1896 (2012)
53. H.A. Andrews et al., Novel tools and observables for jet physics in heavy-ion collisions (2018)
54. R. Kunnawalkam Elayavalli, K.C. Zapp, Medium response in JEWEL and its impact on jet shape observables in heavy ion collisions. *JHEP* **07**, 141 (2017)
55. E. Schnedermann, J. Sollfrank, U.W. Heinz, Thermal phenomenology of hadrons from 200-A/GeV S+S collisions. *Phys. Rev. C* **48**, 2462–2475 (1993)
56. C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988)
57. J.M. Butterworth, D. Grellscheid, M. Krämer, B. Sarrazin, D. Yallup, Constraining new physics with collider measurements of Standard Model signatures. *JHEP* **03**, 078 (2017)
58. K. Aamodt et al., Particle-yield modification in jet-like azimuthal di-hadron correlations in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Rev. Lett.* **108**, 092301 (2012)
59. B. Abelev et al., Pion, kaon, and proton production in central Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Rev. Lett.* **109**, 252301 (2012)
60. E. Abbas et al., Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Lett. B* **726**, 610–622 (2013)
61. B.B. Abelev et al., Multi-strange baryon production at mid-rapidity in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Lett. B* **728**, 216–227 (2014). [Erratum: *Phys. Lett. B* **734**, 409 (2014)]
62. B.B. Abelev et al., Multiplicity dependence of pion, kaon, proton and lambda production in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. *Phys. Lett. B* **728**, 25–38 (2014)
63. J. Adam et al., Centrality evolution of the charged-particle pseudorapidity density over a broad pseudorapidity range in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV. *Phys. Lett. B* **754**, 373–385 (2016)
64. J. Adam et al., Anisotropic flow of charged particles in Pb-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. *Phys. Rev. Lett.* **116**(13), 132302 (2016)
65. J. Adam et al., Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. *Phys. Lett. B* **772**, 567–577 (2017)
66. J. Adam et al., Insight into particle production mechanisms via angular correlations of identified particles in pp collisions at  $\sqrt{s} = 7$  TeV. *Eur. Phys. J. C* **77**(8), 569 (2017). [Erratum: *Eur. Phys. J. C* **79**(12), 998 (2019)]
67. O. Adriani et al., Measurements of longitudinal and transverse momentum distributions for neutral pions in the forward-rapidity region with the LHCf detector. *Phys. Rev. D* **94**(3), 032007 (2016)
68. L. Adamczyk et al., Beam energy dependence of the third harmonic of azimuthal correlations in Au+Au collisions at RHIC. *Phys. Rev. Lett.* **116**(11), 112302 (2016)