Streamlining ATLAS Monte Carlo Generator Validation with PAVER

Mustafa Schmidt[1,*] on behalf of the ATLAS collaboration

[1]Bergische Universität Wuppertal

Abstract. Monte Carlo simulations are crucial in modern high-energy physics, providing theoretical predictions that are used for experimental analyses. Hence, efficient and systematic validation of these simulations is critical to ensure the reliability and reproducibility of results. The PAVER system has been developed to meet this requirement by centralizing and automating the validation workflow, incorporating statistical tools and collaborative review processes to ensure high-quality Monte Carlo production. This paper discusses its workflow, features, and impact, emphasizing the contribution to efficient validation and resource optimization within the ATLAS experiment at the LHC.

1 Introduction

High-energy physics experiments rely on extensive Monte Carlo simulations to model particle interactions and detector responses. These simulations must be validated to verify their accuracy and ensure alignment with theoretical expectations and experimental data. The ATLAS experiment at the Large Hadron Collider (LHC) [1] generates billions of events annually, allocating a large fraction of existing computing resources. Hence, Monte Carlo validation plays a crucial role not only in ensuring the accuracy and reliability of physics analyses but also in optimizing computational resources.

A previously used framework was **JEM** (Jet Event Monitoring) which was originally designed to monitor and validate key aspects of Monte Carlo event generation. While JEM provided basic capabilities for automated validation, the growing scale and complexity of ATLAS simulations required a more robust and centralized solution. As a result, the **PAVER** (PMG Architecture for Validating Evgen with Rivet) [2] system was developed by the ATLAS **PMG** (Physics Modelling Group), incorporating modern tools like **RIVET** (Robust Independent Validation of Experiment and Theory) [3] and enhancing automation, scalability, and statistical analysis.

PAVER introduces a fully automated workflow that systematically compares new Monte Carlo samples to validated reference datasets for mitigating the limitations of manual and decentralized validation methods. This approach significantly reduces the computational and human resources required for validation, while ensuring the reliability of large-scale Monte

*e-mail: muschmidt@uni-wuppertal.de

Carlo production by identifying potential issues in generator updates before large-scale production campaigns. As a result, efficient validation workflows contribute to energy savings and promote sustainability, aligning with the broader goals of resource-conscious scientific computing.

2 Monte Carlo Validation in ATLAS

Monte Carlo simulations are used for exploring a wide range of physics processes, from Standard Model measurements to searches for new phenomena. The precision and accuracy of these simulations depend on periodic updates to the underlying generators, which incorporate theoretical advancements and improved models.

However, each generator update also introduces potential changes that must be carefully validated. Without a robust validation mechanism, inconsistencies can easily propagate into large-scale production campaigns, leading to resource-intensive reprocessing. Hence, PAVER addresses these challenges by providing a centralized system that automates and streamlines the validation workflow. All samples are created within the ATLAS offline software framework Athena [4] making it possible to identify additional integration issues including job options or modification in the event generation.

2.1 The PAVER Workflow

The PAVER system follows a structured validation workflow including high automation and expert review:

1. **Sample Preparation and Reference Comparison:** New Monte Carlo production samples are generated in the official ATLAS production system. These samples are then compared to validated reference datasets, typically the last approved version of the generator or physics process under study.

2. **Histogram Generation and Statistical Analysis:** Using RIVET, PAVER generates hundreds of histograms per sample. These histograms are subjected to statistical tests, including Kolmogorov-Smirnov, to identify deviations.

3. **Collaboration with Generator Experts:** Validation results are shared with generator authors and ATLAS members during regular round-table discussions. These collaborative sessions provide a forum for addressing identified issues and ensuring alignment across different teams.

2.2 Implementation

The PAVER system is designed to handle the increasing complexity and scale of Monte Carlo validation in the ATLAS experiment. Its implementation combines automation, user accessibility, and advanced statistical tools, ensuring a simple and efficient validation process.

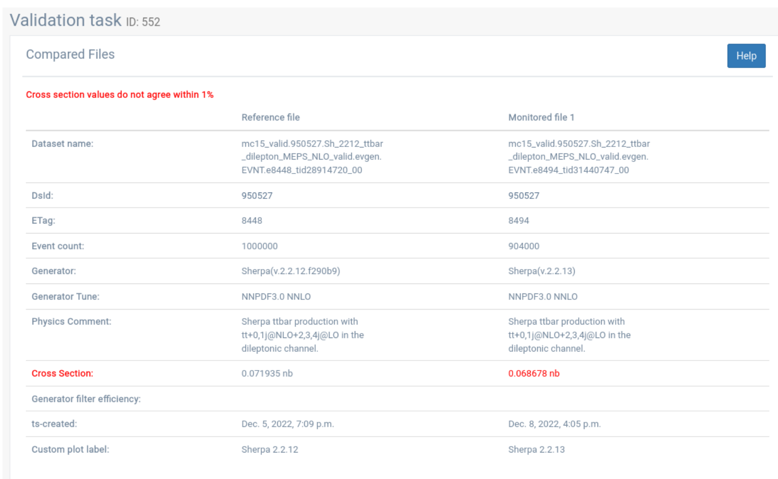
2.2.1 Histogram Generation and Filling

The core of the PAVER system contains the ability to automatically generate and populate histograms using the RIVET framework. For each Monte Carlo sample, 200–400 histograms are generated, capturing key observables relevant to the physics processes under investigation. These observables include distributions such as transverse momentum, invariant mass, and jet multiplicities.

The histogram generation process is fully automated and can be triggered via the PAVER web interface which then submits the jobs to the Worldwide LHC Computing Grid (WLCG). This website can only be accessed by users with a CERN account and for submitting any job, a separate PAVER account is required. Users can select relevant RIVET analyses, typically between 10 and 20 per sample, to ensure comprehensive coverage of the physics under study. This automates job submission, monitoring, and result retrieval, streamlining the entire validation workflow. It also allows PAVER to handle the massive data volumes and computational requirements of modern Monte Carlo validation.

2.2.2 Metadata and Filtering Options

The metadata of each sample includes the selected physics processes covered by that sample and the statistical test results as well as the significance levels. An example of a PAVER validation output can be found in the screenshot of the website in Figure 1 for a Sherpa [5] validation performed in 2022 showing all important metadata and a cross-section comparison.


Validation task ID: 552

Compared Files

Help

Cross section values do not agree within 1%

	Reference file	Monitored file 1
Dataset name:	mc15_valid.950527.Sh_2212_ttbar _dilepton_MEPS_NLO_valid.evgen. EVNT.e8448_tid28914720_00	mc15_valid.950527.Sh_2212_ttbar _dilepton_MEPS_NLO_valid.evgen. EVNT.e8494_tid31440747_00
DsId:	950527	950527
ETag:	8448	8494
Event count:	1000000	904000
Generator:	Sherpa(v.2.2.12.f290b9)	Sherpa(v.2.2.13)
Generator Tune:	NNPDF3.0 NNLO	NNPDF3.0 NNLO
Physics Comment:	Sherpa ttbar production with tt+0,1j@NLO+2,3,4j@LO in the dileptonic channel.	Sherpa ttbar production with tt+0,1j@NLO+2,3,4j@LO in the dileptonic channel.
Cross Section:	0.071935 nb	0.068678 nb
Generator filter efficiency:		
ts-created:	Dec. 5, 2022, 7:09 p.m.	Dec. 8, 2022, 4:05 p.m.
Custom plot label:	Sherpa 2.2.12	Sherpa 2.2.13

Figure 1. PAVER Website showing all relevant metadata of a Sherpa validation from 2022 [2]

The PAVER web interface additionally provides powerful filtering and sorting options, enabling users to quickly focus on specific samples, physics observables, or regions of interest. This feature is particularly useful when dealing with large-scale validation campaigns involving numerous samples and configurations. In addition, the interface supports uploading and managing new Monte Carlo samples, viewing histograms, statistical test results, and metadata, or downloading results for offline analysis.

The full website is secured using CERN Single Sign-On (SSO), ensuring that only authorized users can access the system. This accessibility encourages collaboration among validation teams and generator developers, fostering a more transparent and interactive validation process.

2.2.3 Review Process

Once validation results are available, they are shared with generator authors and ATLAS members through the PAVER platform. Regular round-table discussions provide an opportunity for collaborative review, enabling experts to discuss identified discrepancies and their potential causes, suggest improvements to generator configurations, and formulate decisions on whether the new samples are ready for production. This collaborative approach ensures that issues are addressed systematically, and validated samples meet the highest quality standards.

3 Validation Statistics and Results

3.1 Validation Campaigns

In 2023, the ATLAS collaboration produced approximately 80 billion events for official production analyses and 0.5 billion events specifically for validation purposes. These validation samples covered a wide range of physics processes and generator versions, including complex scenarios such as $t\bar{t}$ production, W/Z+jets processes, or supersymmetry (SUSY) signals and other beyond standard model (BSM) processes.

Typically, seven validation samples are generated for each new generator version included in the latest Athena release, with identical generator configurations determined by so-called dataset identifiers (DSIDs). This approach ensures that PAVER can evaluate a diverse set of scenarios, providing a robust validation framework.

3.2 Identified Issues and Their Resolution

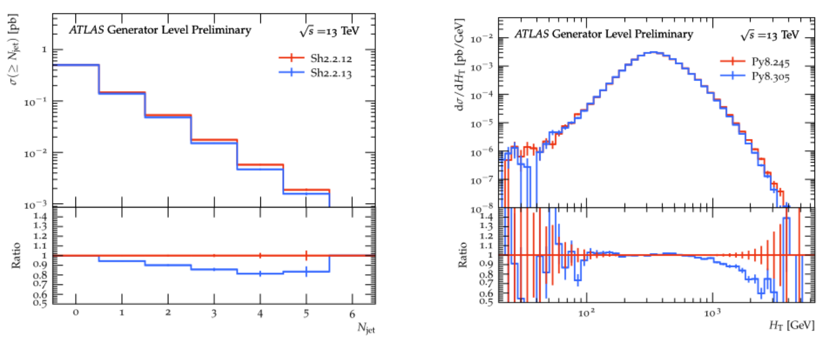


Figure 2. Side effects of performance optimization in Sherpa (left) and an issue with CKKW-L merging in Pythia (right) [2].

Some notable findings from recent validation campaigns include the following issues:

- **CKKW-L Merging Issue in Pythia8:** PAVER identified a discrepancy in the merging algorithm used in Pythia8 [6], which could have resulted incorrect jet multiplicity distributions. This issue was promptly reported to the generator authors and resolved in subsequent updates.
- **Performance Optimization in Sherpa 2.2.13:** While the new version of Sherpa showed improved computational performance, PAVER detected unintended side effects on physics results, specifically in inclusive jet multiplicity. This finding highlighted the importance of balancing performance enhancements with physics accuracy.

- **Compatibility Issues in Herwig7 with HepMC3:** A unit mismatch in the Herwig7 [7] interface to HepMC3 was uncovered, which could have introduced systematic biases. The issue was escalated and corrected in newer versions of the generator.
- **Runtime Errors in MadGraph with LHAPDF:** PAVER flagged runtime errors caused by compatibility issues between MadGraph and specific versions of the LHAPDF library. These errors were mitigated by updating the library and adjusting configuration parameters.

The validation outputs related to the performance optimization of Sherpa and the CKKW-L merging in Pythia can be found in Figure 2.

A timeline for the Monte Carlo validation campaigns in 2022 and 2023 is illustrated in Figure 3. The red boxes indicate the generator versions showing the above-mentioned issues that resulted in disproving this specific version during the validation process. As a result, a new version or fix had to be provided by the generator experts, which could finally be approved at a later time, since only fully validated generators can be used in Monte Carlo production campaigns.

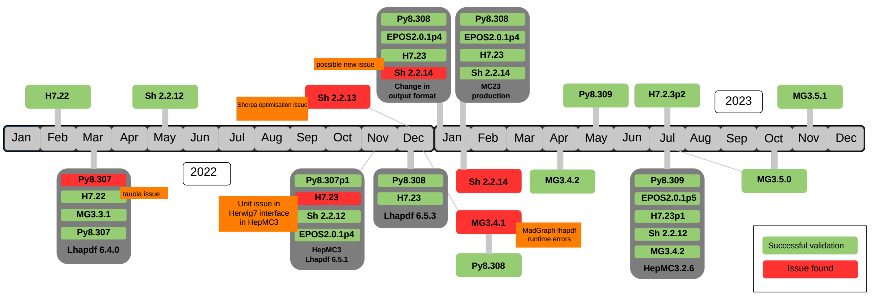


Figure 3. Timeline for Monte Carlo generator validations in 2022 and 2023 [2].

3.3 Statistical Insights from Validation

The PAVER system uses advanced statistical tools to provide quantitative insights into the performance of new Monte Carlo production samples. Statistical tests, including the Kolmogorov-Smirnov test, compare histograms generated from new and reference samples. Key metrics derived from these tests include the agreement scores for quantifying the similarity between distributions and the significance levels (p-values) which illustrate the likelihood of observed discrepancies being statistically significant.

The results are visualized using a dedicated color-coded scheme, enabling users to interpret statistical outcomes and prioritize further investigation quickly. Histograms with significant deviations are flagged in red, drawing immediate attention to potential issues. One example for each color taken from a validation of Sherpa 2.2.13 is shown in Figure 4. The red and blue lines within each histogram indicate the monitored and reference samples respectively, whereas the black points represent data to additionally compare the generator observables with measurements provided by RIVET.

The proactive validation approach enabled by PAVER has a transformative impact on the efficiency and sustainability of ATLAS Monte Carlo production. These include resource optimization, since an early identification of issues reduces the need for reprocessing, saving computational time and energy. It also includes sustainability by avoiding unnecessary reruns, contributing to the reduction of the ATLAS experiment's carbon footprint. Finally, it

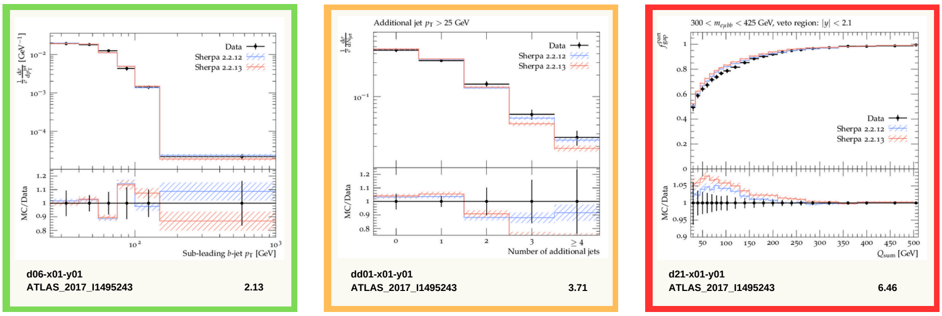


Figure 4. Validation histograms including the color-scheme of PAVER with the frames in green, yellow, and red [2]. The red and blue lines represent the results of the monitored and validated samples, whereas the black points show the comparison with experimental data.

improves physics accuracy as validated samples ensure that ATLAS analyses are based on accurate and reliable input, enhancing the robustness of scientific results.

During the 2023 validation campaigns, the detection and resolution of the Sherpa optimization issue alone prevented millions of incorrect events from being generated, saving significant resources and preserving analysis quality.

4 Conclusion

PAVER represents a significant advancement in the validation of Monte Carlo simulations for the ATLAS experiment. By automating key aspects of the validation process and maintaining collaboration between physicists and generator developers, PAVER ensures the reliability and efficiency of large-scale Monte Carlo production. Its continuous development will play a critical role in supporting the scientific goals of ATLAS and the broader high-energy physics community.

References

[1] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, Journal of Instrumentation **3**, S08003 (2008). 10.1088/1748-0221/3/08/S08003

[2] ATLAS Collaboration, Monte Carlo Validation in ATLAS with PAVER (2024), all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-013, `http://cds.cern.ch/record/2904943`

[3] A. Buckly et al., Rivet user manual, Computer Physics Communications **184**, 2803–2819 (2013). 10.1016/j.cpc.2013.05.021

[4] ATLAS Collaboration, Software and computing for Run 3 of the ATLAS experiment at the LHC, arXiv:2404.06335 (2024).

[5] E. Bothmann et al., Event generation with sherpa 2.2, SciPost Physics **7** (2019). 10.21468/scipostphys.7.3.034

[6] C. Bierlich et al., A comprehensive guide to the physics and usage of PYTHIA 8.3, SciPost Phys. Codebases p. 8 (2022). 10.21468/SciPostPhysCodeb.8

[7] J. Bellm et al., Herwig 7.0/herwig++ 3.0 release note, The European Physical Journal C **76** (2016). 10.1140/epjc/s10052-016-4018-8