Monday, January 29, 2024

Report

Readiness Review for Pass1 processing of the Hall-B/CLAS12 RG-C Summer-2022 data set

Review committee:
Nathan Baltzell
Marco Battaglieri (chair)
Larry Weinstein
Marco Mirazita
Cole Smith

The review of the readiness of RG-C to process a second pass of the Summer-2022 dataset with the latest improved reconstruction software available took place on January 26 on Zoom. The meeting agenda and presentations can be found on the review page:
https://clasweb.jlab.org/wiki/index.php/RGC_Summer2022_pass1_review

The review committee would like to thank the RG-C team for preparing the presentations and addressing the reviewer's questions.

We believe the RG-C team addressed all the charges of this review, suggesting starting **RG-C Summer-2022 data cooking as soon as recommendations are cleared.**

Details about the charges and responses are reported below.

Review Charges

Charge #1: Is the quality of detector calibration and alignment adequate to achieve the performance specifications foreseen for CLAS12 or achievable at the current time, given the "state-of-the-art" calibration, alignment, and reconstruction algorithms?

FINDINGS:

The RG presented the status of dataset calibration for the fraction of the run under review. Data were taken at different beam energies (2.21 GeV and 10.54 GeV). Besides production runs on NH3 and ND3 targets with different polarization alignments, many calibration and special runs (Carbon,

CH2, and empty target) were performed between July and August 2022. RG-C developed the procedure to calibrate the raster system, including a comparison with MYA DB, first used in this experiment, showing that results are stable and consistent.
Overall, the calibration constants that will be used to reconstruct data from the CLAS12 are well within the requested limits, appear to be stable over the whole run, and the results are consistent. The CALCOM cleared the calibration set. The CVT shows very good results after alignment.

COMMENTS:

Similar to other RGs, data from the RICH detector will require further post-processing before physics analysis.

RECOMMENDATIONS:

None

Charge #2 Is data quality as a function of run number or time for the data set proposed for cooking stable and understood? Have runs been classified in terms of type (empty target, calibration, special, production, …) and quality (golden run, known issues, …), and is a detailed list available? Based on validation studies, have all CLAS12 subsystem performances been understood and issues identified?

FINDINGS:

The RG-C timelines presented at the review demonstrate good stability (within the specs) of all CLAS12 subsystems as a function of time (or, equivalently, run numbers). Almost all sudden changes in the monitored quantities were explained by considering the different experimental conditions of the different data subsets but some were related to an unwanted change in the experimental conditions (see Comments and Recommendations sections). Some of the runs show behavior outside specifications: RG-C intends to process all runs (including outliers) postponing the decision of using them or not to the specific physics analysis they will be used for.

COMMENTS:

RG-C identified a particularly problematic period where a significant Hall-C beam bleed through affected some monitored parameters for about 60 runs. The Review Committee understands the rationale for including all possible runs in the Pass1 processing but the RG-C team should provide a clear strategy to address known issues in some of the runs that will be cooked (e.g. flagging the run quality or defining a post-processing procedure).

RECOMMENDATIONS:

A list of problematic runs with known issues should compiled and the information book kept in a DB to warn physics analyses.

While the effect of the Hall-C beam bleed through is visible in some monitored quantities (e.g. ECal start time), and to some extent could be corrected during the analysis, the implications for other variables and effects on measured physics observable are not clear. For instance, how would it affect the background merging necessary to compare simulation to data? A study about a mitigation procedure and a strategy to quote a reliable systematic error for runs affected by this issue, or a procedure to exclude problematic runs from the analyses should be defined.

Jumps in the FC QA timelines should be understood and a possible mitigation strategy for consequences in charge normalization and beam-charge asymmetry should be defined.

Charge #3: Has a Hardware (HW) status table (i.e., bad channel table) been compiled for use in the data and MC reconstructions? Has the efficiency versus beam current been studied? How does it compare to MC simulations with the merged

background? Are the DAQ translation tables correcting for all known cable swaps? At what stage(s) in the software?

FINDINGS:

HW status tables have been defined for the CLAS12 subdetectors and are ready in the 'default' variation with run-by-run dependencies (adopted for the first time). No DC cable swaps were identified. Validation has been performed.

COMMENTS:

None

RECOMMENDATIONS:

None

Charge #4: Are analysis plans for the data set developed at adequate levels? Is the list of planned skims defined and tested running the analysis trains on preliminary data? Is all ancillary information helicity, Faraday Cup, …) available and understood?

FINDINGS:

Several physics channels were analyzed and data from polarized NH3 and Carbon were thoroughly compared. It is worth noticing that in the exclusive two-pion electroproduction, the resolution on the missing mass of the proton is consistent with RG-A and RG-K data. Preliminary results on asymmetries calculated on SIDIS epX show good agreement with the theoretical predictions. pDVCS shows a target-spin asymmetry of the expected modulation and size. All physics analyses performed on a cooked subset of RG-C data, show promising results. As per the other RGs where the comparison is available, AI-assisted tracking outperformed the standard algorithm. A significant increase in the event yield (from 10% to 30%, depending on the reaction and the kinematic) was found. De-noising procedure was also demonstrated to be effective on the RG-C data set.

COMMENTS:

The AI algorithm was trained both on RG-B and RG-C data sets and results obtained with the two training sets were compared showing a slightly better performance when the RG-B data set was used. While a weak dependence on the training set may be reasonable, it is not understood why the best results were obtained with the 'unnatural' data set.

Directories with pass1 results should be named after the target type for easy access during the physics analysis.

RECOMMENDATIONS:

None

Charge #5: Are the data processing tools that will be used adequately for the proposed processing task? Is the data management plan (staging area, tape destination, directory structure, logs, …) defined and appropriate given the available resources? Is the estimate of processing time per event available and resources needed to complete the task sound?

FINDINGS:

RG-C presented estimates of the necessary disk space needed by pass-1 cooking. The cooked dataset size (estimated to be DST: 51 TB, and SKIMS: 15 TB) is compatible with the current disk resources allocated to CLAS12.

COMMENTS:

The farm is currently not used by any full-scale CLAS12-related data processing. The expected processing time has been estimated at 25 days at Hall-B priority fair share.

RECOMMENDATIONS:

None

Charge #6: Have the tools for monitoring the quality of the cooking output and identifying/correcting failures been defined and ready to be used?

FINDINGS:

QA timelines together with high-level physics analyses, will be used to monitor the pass1 data.

COMMENTS:

None

RECOMMENDATIONS:

None

Charge #7: Is the person-power identified and in place for the proposed data processing?

FINDINGS:

Personnel, including the coordinator, chef, calibrators, and physics analyzers were found to be adequate.

COMMENTS:

None

RECOMMENDATIONS:

None