Thursday, February 20, 2025

Report

Readiness Review for Pass1 processing of the Hall-B/CLAS12 RG-C Spring/Winter-2023 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 first pass of the Spring/Winter 2023 dataset with the latest reconstruction software available took place on February 19 on Zoom. The meeting agenda and presentations can be found on the review page:

https://clasweb.jlab.org/wiki/index.php/RGC_Spring2023_pass1_review.

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

The RG-C team addressed all the charges of this review, suggesting starting **RG-C Spring/Winter 2023 data cooking as soon as the RG chooses which raster calibration procedure they want to adopt.**

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 10.5 GeV beam energy. 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 January and March 2023. No luminosity scans were performed. For this data set, RG-C tested both algorithms developed to calibrate the raster: the standard procedure (with calibration constants in CCDB) and the new method on a run-by-run basis proposed

for the RG-C Fall 22 data set cooking, known as *my raster*. The new method uses the detached vertexing technique identifying (e- pi+ X) events to identify the beam position with high accuracy. From direct comparison on a data subset, both methods were demonstrated to be adequate. A systematic difference of 0.5mm was observed. The advantage of my raster vs standard demonstrated in Fall 2022 calibration (in particular in correcting outlier runs) was less evident for this data set. The effects of the two methods on final observables used in physics analysis do not allow us to select one over the other. Overall, the calibration constants 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, requiring minimal variation (and iteration) from the calibration constants used for the previous RGC data set.

COMMENTS:

The committee commends the effort of the RG-C team to study in detail the effect of the two raster calibration methods, recognize that both are adequate, and leave the final decision on which to use to the RG.

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). Sudden changes in the monitored quantities were explained by considering the different experimental conditions (e.g., in-bending vs. out-bending torus set). A careful analysis of the HWP position was performed, showing consistent results in the corresponding timeline.

COMMENTS:

We noticed that the RICH detector calibration constants are pretty stable in time, but since no QA limits are defined, it is not possible to judge the quality of the extracted parameters.

RECOMMENDATIONS:

None

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. No DC cable swaps were identified.

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 (elastic, pDVCS, TCS, SIDIS with detected proton, pion, and kaon) were analyzed, and results obtained from Fall 2022 and this data set were thoroughly compared.

COMMENTS:

It was noticed that, in several physics channels, the normalized yield obtained with this data set is 10-20% higher than the one obtained by Summer 22 cooking. The increase was related to the various improvements (in particular the tracking) of the reconstruction software. We also noticed that the background merging procedure shows a better comparison between MC and data (even if the agreement is not perfect).

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: 31TB, and SKIMS: 11 TB) is compatible with the current disk resources allocated to CLAS12 and similar in size to the previously cooked RG-C Summer-22 data set. The time necessary to cook the whole data set (assuming the same RG-C Summer-22 data set fair share) estimated by the RG is ~ 7 days.

COMMENTS:

The processing time per event (estimated to be ~336ms/ev) was found to be significantly smaller than in the previous cooking.

Considering that processing the RGC Summer 22 dataset took approximately three weeks, the RG time estimate appears to be overly optimistic.

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