Physics and software validation for ATLAS

D Costanzo¹, A Pacheco²,³ and I Vivarelli⁴,⁵,⁶

¹University of Sheffield
²Institutt de Fisica d’Altes Energies
³CERN
⁴Istituto Nazionale di Fisica Nucleare
⁵Universita degli Studi di Pisa
E-mail: iacopo.vivarelli@cern.ch

Abstract. The validation procedure for the ATLAS software releases is reviewed. The software validation, whose aim is to provide functional releases, is first discussed. Then the procedure to verify the correctness of the output files produced by the release, the physics validation, is described.

1. Introduction
The ATLAS experiment is finally entering the data taking phase. The focus of the software community is shifting from software development to validation and reliability. The ATLAS software has to be both robust to process large datasets coming from the detector and from the simulation production, and produce the high quality output needed for the experiment scientific exploitation. The validation of the software is done in two distinct phases. The first step includes testing robustness and software functionalities, the second step includes the production of a relatively large sample of MonteCarlo data, to test the quality of the output quantities. The infrastructure for each stage of validation is described.

2. From the nightly build to the production release
At a given time, the ATLAS software consists of a limited number of open releases, typically one to be used for the data taking and one which is foreseen for major simulation production. Every 24 hours, all open releases are automatically built by the Nightly Control System (NICOS) [1]. NICOS collects from CVS the tags indicated by a database called Tag Collector and performs a number of modular steps on them: compilation of the code, testing of the compiled code, analysis of errors generated in the compilation testing phase, and finally creation of web pages reporting the results of the testing. Figure 1 shows a sketch of the components used by NICOS to build the software release.

The first stage of testing of the ATLAS software is the ATN. The ATLAS Testing Nightly (ATN) is a framework that allows quality assurance checks, unit tests and integration tests. The configuration of the test system is done through XML files. The quality assurance tests consist in the check of the requirement file of each package. Based on the requested package versions, ATN checks the consistency of the set of packages (and package dependencies) in a release.

⁶Now at Albert Ludwigs Universität, Freiburg.
The unit tests consist in the check of basic functionalities of a package, based on simple
executables. The check is done on the exit code of the executable itself. The integration test
consist in the check that packages can work together and that they can communicate with other
software components.

The results of the tests are returned to NICOS for processing and publication in the build
summary page for every nightly.

![Diagram showing the NICOS use of the components when building a nightly release.](image)

**Figure 1.** Diagram showing the NICOS use of the components when building a nightly release.

3. **The Run Time Tester (RTT)**

The second step of the software validation is the ATLAS Real Time Tester (RTT). This is
a testing framework written in python, which sets up relatively complex jobs (running few
g4 simulation events, running digitization with pileup, running full reconstruction on few
events, for example), runs them, performs post-execution actions and tests and finally publishes
the results on a web accessible location (a snapshot is available in fig. 2). A wide variety
of the allowed ATLAS software configurations are tested, spanning from event generation to
subdetector specific and combined simulation and reconstruction, to physics analysis.

Typically an RTT job requires the definition of:

- A validated software release whose results are used as a reference for the nightly under
testing
- A list of jobs to be run, configured through an XML file.

The large number of software tests that need to be run and the relatively limited number of
CPUs involved in this software test is actually limiting the total number of events that can be
used in the testing. For practical purposes, the RTT jobs rarely run on more than few tens of
events. Also, the files used as an input to the different RTT jobs have usually been produced in
a previously validated release. This implies that the capability of release to run on input files
produced in the same release is not checked (e.g., there is no explicit check that a release is
able to reconstruct its own simulated files). These are the main reasons that made the ATLAS
collaboration define a further layer of testing.
4. The Full Chain Test (FCT)
For every nightly, the full ATLAS software chain, from event generation up to the production of the final Analysis Object Data, is tested in the Full Chain Test. A large scale test using the same scaffolding as the FCT and running at the Tier0 (Tier0 Chain Test, TCT) is used for the validation of the release to be used for the (cosmic rays) data taking.

The Full Chain Test has been defined to overcome the practical limitations of the RTT. It performs a limited number of tests (only a fraction of those performed by the RTT). On the other hand, it runs the whole ATLAS software chain, from the event generation, to simulation, to digitization and reconstruction. This is done for few physics processes, from basic Standard Model processes ($Z \rightarrow ee, \mu\mu$ production), to exotic (and very complex) events (black holes production).

Given the relatively limited number of tests done by the FCT and the allocated resources, the statistics that can be considered is larger than in the RTT. As part of the FCT tests, the reconstruction of about 1000 events of top pair production is done. This is the largest scale test before the release is deployed to the grid. This is also the first step for the physics validation (described in detail in the next section). In fact, once the output of the reconstruction is available, an analysis jobs is run on it, the resulting histograms are automatically compared to a reference, and the results are published on a web page. Figure 3 shows the web page resulting from the automatic histogram comparison done using the ATLAS Data Quality Monitoring Framework [2].

The FCT has been designed to run the test jobs on dedicated batch queues, but using the same scripts that are used on the grid for production. This allows to test also the scripts themselves before their use on the grid.

The use of the FCT allows to certify that a release can run on $10^3$ events without producing a failure. It also provides a first, statistically significant, hint on the quality of the output produced by the reconstruction.

5. The physics validation
Once the RTT and FCT test are successful, the release is deployed on the GRID for its final validation step. This consist in the production of a sample of approximately 250K events. The composition of the sample has been chosen carefully to test all aspects of simulation and
reconstruction. It includes single particles, standard model processes (W and Z leptonic decays, top pair production, Higgs production) and non standard model processes (SUSY, black holes production). The sample is produced every time a major or bugfix release is distributed to the grid. The production of such a sample is resource demanding, and that is the main reason to produce it on the grid after the release is distributed.

![Configuration Parameters](image)

**Figure 3.** Histogram comparison using the ATLAS Data Quality Monitoring Framework.

![M_Z Resolution](image)

**Figure 4.** The reconstructed $Z \rightarrow e^+e^-$ peak (minus the true value) is compared for two ATLAS releases.

Every combined reconstruction performance and physics group in ATLAS is responsible to cross check the results of the sample analysis with those from a validated release. In case problems are found, the process is iterated: a new, fixed release is built, tested with RTT and FCT and finally distributed to the grid, where it is used to produce again the validation sample.
When a release passes the physics validation tests, it is declared validated, and it can be finally used for data or simulation processing (or reprocessing).

One example of the validation plots produced during the physics validation step is shown in fig. 4: two releases are compared concerning the reconstruction of the $Z \rightarrow ee$ peak.

The relatively large statistics of the physics validation sample allows:

• to discover software bugs that appear with a rate of $10^{-5}$ ($10^{-4}$ if they appear in a specific final state object reconstruction). This is considered an acceptable failure rate in production.

• to find minor problems in the output of the ATLAS software (e.g., a few percent shift in the jet energy scale).

6. Conclusions
ATLAS has a well established software and physics validation procedure. The different validation layers ensure software reliability and high quality output for physics analysis. The experiences with simulation production show that the validation procedure is reliable and fulfils the collaboration needs.

Acknowledgments
We would like to acknowledge the ATLAS SIT group and D. Quarrie for the software components development and the ATLAS physics validation group for providing the plots.

References
[1] E. Obreshkov et al., Nucl. Instrum. Meth. A584 (2008) 244251.1737
[2] A. Corsu-Radu et al., IEEE Trans. Nucl. Sci., Vol 55, Issue 1 (2008) 417-420