The CMS data analysis school experience

N De Filippis¹, L Bauerdick², J Chen³, E Gallo⁴, B Klima², S Malik⁵, M Mulders⁶, F Palla⁷ and G Rolandi⁸

¹ Dipartimento Interateneo di Fisica, Politecnico and INFN Bari, 70122 Bari, IT
² Fermilab, PO Box 500, Batavia, IL, U.S.A
³ National Taiwan University, Taipei, Taiwan
⁴ DESY, Hamburg, Germany
⁵ University of Puerto Rico, Mayaguez, Puerto Rico
⁶ CERN-PH, 1211 Geneva 23, Switzerland
⁷ Istituto Nazionale di Fisica Nucleare, Pisa, Italy
⁸ Scuola Normale Superiore di Pisa, Pisa, Italy

E-mail: Nicola.Defilippis@ba.infn.it

Abstract. The CMS Data Analysis School is an official event organized by the CMS Collaboration to teach students and post-docs how to perform a physics analysis. The school is coordinated by the CMS schools committee and was first implemented at the LHC Physics Center at Fermilab in 2010. As part of the training, there are a number of “short” exercises on physics object reconstruction and identification, Monte Carlo simulation, and statistical analysis, which are followed by “long” exercises based on physics analyses. Some of the long exercises go beyond the current state of the art of the corresponding CMS analyses.

This paper describes the goals of the school, the preparations for a school, the structure of the training, and student satisfaction with the experience as measured by surveys.

1. Introduction

The CMS experiment [1] at the CERN laboratory in Geneva, Switzerland is one of the four main experiments at the Large Hadron Collider (LHC). It was designed to search for the Standard Model (SM) Higgs boson (discovered in 2012 [2]), as well as for evidence of physics beyond the SM [3, 4, ?], such as supersymmetry and extra dimensions.

The CMS collaboration includes over 3500 physicists, engineers, and students from institutes all over the world. Managing such a large number of people in order to operate the experiment, to develop the analysis software, and to ensure high-quality and timely scientific results requires a highly efficient organization including a large number of individuals who are capable of analyzing data collected by the experiment.

One of the keys to the success of the collaboration is the training of people. The CMS Data Analysis schools (CMSDAS), initially conceived and implemented at the LHC Physics Center (LPC) at Fermilab in 2010, are the most important event organized by CMS for teaching participants how to handle physics objects and perform a physics analysis within the CMS software and computing environment, using both data collected by the experiment and Monte Carlo (MC) simulations. During the school, the participants also learn about the publication process. A full description of the CMSDAS model can be found in Ref. [6].
2. Pre-school setup
Preparations for a CMSDAS require coordination between facilitators at the host institute and the international CMS Schools Committee. A CMSDAS has been organized at the Fermilab LPC every year since 2010; others have been organized in Europe (CERN, Pisa, DESY, Bari) and in Asia (Taipei, Kolkata, Daegu) in order to locally serve the different scientific communities collaborating on the CMS experiment.

The period in which the school is held is carefully chosen in order to avoid overlaps with major physics conferences. The host institute must possess a computing infrastructure that can handle hundreds of analysis jobs running in parallel and terabytes of disk, both for real and simulated data, as well as provide individual accounts for the participants. The host institute should also have a large conference room for the plenary sessions and a number of smaller rooms, all with internet connectivity and standard white/blackboards. The smaller rooms are used by the different analysis teams working in parallel on the exercises.

Preliminary exercises on computing and software are prepared by experts and should be completed by the students before the beginning of the school. These pre-exercises are useful for ensuring that the students do not lose time with basic technical problems when they arrive, such as the site-specific configurations of CMS software environment, and that they have some experience with CMS computing tools.

“Short” and “long” exercises are prepared before the beginning of the school and documented in CMS twiki pages. All the exercises are pre-tested by small groups of experienced users. Each student is asked to provide preferences for both short and long exercises prior to the start of the school. Once the long exercises are assigned, the students are assigned (typically) three short exercises as preparation for their long exercise. Students are grouped in analysis teams based on the assignments of the short and long exercises.

3. Operations
Five full days are allocated to a DAS. The morning of the first day is devoted to plenary lectures including an introduction to the “big picture” in particle physics, with a summary of the SM as well as beyond-the-SM physics. The latest CMS experimental physics results are summarized as well as short and long-term prospect. These talks are followed by an introduction to Accelerator physics, including the status of the LHC, and an overview of the status of CMS detector as well as plans for upgrading it. Finally, the last lecture introduces CMS software and computing tools.

Two days are then devoted to ten or more parallel short exercises on physics objects, MC generators, and statistical analysis. Each short exercise is supervised by two or three experts from the collaboration. The short exercises are repeated two times with different teams and last 2–3 hours.

On the evening of day three the students attend a lecture on scientific writing and the publication process. This gives the students an appreciation for the level of effort required between the approval of an analysis and the final publication of a paper.

The last two and a half days of the school are devoted to specific physics analyses (long exercises). Teams of about six-eight students are formed to work on these exercises. The exercises include a hands-on tutorial run by facilitators, who are experienced in the analysis of the selected topic (e.g., Higgs, Exotica, Standard Model, Top, etc.). The facilitators guide the students through the main steps of the analysis up to final plots and numerical results. Within a team, the students divide themselves into smaller groups in order to accomplish the many tasks required for an analysis.

A mini-symposium is organized during the last half a day of the school, introducing an element of competition among the analysis teams. Each team is given 20 minutes to present their analysis and another 5 minutes is allocated for a team of judges to probe the understanding
of the students. Following the presentations, the judges deliberate and assign a prize to the best team based on the quality of the analysis and of the presentation as well as the ability of the students to respond to judges’ questions.

4. Results and survey
To date there have been sixteen CMS analysis schools worldwide and over one thousand participants. At the end of each school the attendees are given a survey on the experience. The latest survey data show that, on average, 15–20% of the participants are undergraduates, 75–80% are graduate students, and 5% are post-docs (Fig. 1). About 50% of the students had less than one year experience with CMS software, while 20% had no prior exposure to CMS software (Fig. 2). These average percentages come from the most recent CMSSDAS schools, at Fermilab and DESY; they are consistent with results from previous schools.

![Figure 1. Level of education of the school attendees.](image1)

![Figure 2. Level of expertise with CMS software.](image2)

It is also interesting to note that even though Fermilab, DESY and CERN are major host sites for physicists working on CMS physics analyses, the primary research location of the students is their home institute, as shown in Fig. 3.

The pre-exercises were completed by 94% of the students and were found to be very helpful in dealing with the exercises during the schools (Fig. 4).

The students were also asked to evaluate the balance between talks and exercises during the DAS experience; for the most part they felt the balance was about right. The quality and the relevance of the talks were judged to be appropriate, as shown in Fig. 5. Almost 50% of the respondents thought that 5 days was too short (Fig. 6); maybe an extension of one day could be an added value to the future schools.

The distribution of the students responses to the short and long exercises is shown in Figs. 7 and 8. The make-up of the exercises varies a bit from school to school, but typically the students express similar preferences independent of the school they attend. The short exercise on statistical analysis and long exercise on the discovery of the Higgs boson and the searches for physics beyond the SM are particularly well regarded. Most of the students participated to the exercise about the publication process.

The long exercises were much appreciated, as shown in Fig. 9. The students succeeded in producing meaningful plots and tables for the physics analyses, and the results they obtained are in line with the most recent results from CMS. In some cases the students actually improved on the current results through access to a larger data set or by modifying selection requirement
to improve the sensitivity! Most of the students expressed an intention to work on another long exercise following the CMSDAS.

The CMSDAS was found to be a valuable experience by all the participants; one that they would recommend to their colleagues. More than 60% of the students would also consider either attending another school or serving as a facilitator in the future (Fig. 10).

Through the CMSDAS experience the students become aware of the many on-going analyses within CMS; should they chose to join these the training goes a long way toward helping the students jump-start their involvement.

Within a large collaboration, such as CMS, the social aspects of working with others is an important part of a productive experience, and for many of the CMSDAS participants the school is a first major opportunity for networking within the collaboration.

5. Summary

The CMSDAS training program represents a successful model, proven to be a key tool for the preparation of young physicists and even older physicists who are new to the collaboration. To
date, sixteen schools have been held in the US, Europe, and Asia, and more than one thousand students have participated in the schools. The schools have also drawn on a large number of experts from within the collaboration. The organization of the schools, the lectures, and the short and long exercises are very much appreciated.

References
[1] CMS Collaboration 2008 The CMS experiment at the CERN LHC JINST 3 S08004
[2] CMS Collaboration 2012 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC Phys. Lett. B 716 30
[3] Glashow S. L. 1961 Partial Symmetries of Weak Interactions Nucl. Phys. 22 579-588
[4] Weinberg S 1967 A Model of Leptons Phys. Rev. Lett. 19 1264-1266
[5] Salam A, 1968 Weak and Electromagnetic Interactions in Elementary particle theory Proc. of the Eighth Nobel Symposium ed. N. Svartholm N. (Stockholm: Almquist and Wiksell) pp 367-377
[6] Malik S. et al 2014 CMS Analysis School Model J. Phys.; Conf. Series 513 062029