Computing and Informatics, Vol. 32, 2013, 1256–1271

A NEEDLE IN THE HAYSTACK: HIGGS BOSON SEARCHES IN THE ATLAS EXPERIMENT

Andrzej Olszewski, Marcin Wolter

Institute of Nuclear Physics PAN
ul. Radzikowskiego 152,
30-342 Kraków, Poland
e-mail: {andrzej.olszewski, marcin.wolter}@ifj.edu.pl

Abstract. A preliminary combination of Standard Model Higgs searches with the ATLAS experiment, in a dataset collected at energy $\sqrt{s} = 7$ TeV at the LHC in year 2011, is presented.[1] The Higgs boson mass ranges from 112.7 GeV to 115.5 GeV, 131 GeV to 237 GeV and 251 GeV to 468 GeV are excluded at the 95 % confidence level. An excess of events is observed for a Higgs boson mass hypothesis close to 126 GeV. This successful analysis was possible due to the excellent performance of the GRID distributed computing system, in which two Polish sites ACK Cyfronet from Kraków and PSNC from Poznań have participated. The ATLAS analysis chain and the usage of grid for ATLAS data processing are described in this paper.

Keywords: Higgs boson, ATLAS, LHC, grid computing

1 INTRODUCTION

In the 1970s physicists realized, that two out of four fundamental forces, gravitation, electromagnetism, weak and strong interactions, can be described by the same theory. The so called Standard Model unifies the electromagnetic and weak forces, which implies that electricity, magnetism, light and some types of radioactivity are all manifestations of a single underlying force called the electroweak force.

To make such a unification working mathematically a fundamental problem had to be solved: in this model all the intermediate particles, carrying the force, are

[1] The paper refers to the status of Higgs boson searches as from March 2012, before the Higgs discovery at LHC.

massless. This was obviously not true, as known from the experiments. In 1964 three independent groups of physicists François Englert and Robert Brout [1], Peter Higgs [2, 3] and also Gerald Guralnik, Carl Richard Hagen, and Tom Kibble [4] found a solution of this problem. The suggestion was that just after the Big Bang all particles had no mass. When later on the Universe cooled down the so called "Higgs field" together with the associated "Higgs boson" were formed. All particles that interact with it are given a mass, while particles that do not interact with the Higgs field, like photons or gluons, stay massless.

The idea agrees very well with observed phenomena. However, the Higgs boson has up to now never been observed in any experiment. It is the only remaining undiscovered piece of the Standard Model, albeit an essential one. The search for the Higgs boson is therefore considered to be one of the most important tasks for the LHC experiments.

The problem with Higgs boson searches is that the theory does not predict its mass. Therefore the search requires systematic scanning of a range of mass within which it is predicted to exist. The up to now unexplored region became accessible due to the Large Hadron Collider (LHC) at CERN, Geneva. Within the incoming year it should prove or disprove the existence of the Higgs boson. If it happens that it is not found, the new field will be opened for physicists to develop a completely new theory to explain the origin of particle mass.

In this article we present the results of the Higgs searches by the ATLAS experiment at LHC collider after analyzing of $4.9\,\mathrm{fb}^{-1}$ (corresponding to about 3×10^{14} collisions) of data taken in 2011 at the energy of $\sqrt{s} = 7\,\mathrm{TeV}$ available at the interaction point. The obtained limits are based on the combination of results from several decay channels.

The Higgs analysis was not an easy task from computational point of view. Finding a very small number of Higgs candidates after all selection cuts (like about 70 for $H \to \gamma\gamma$ channel) required processing of enormous amount of data. Computing of ATLAS and other LHC experiments dealt well with this challenge by using extensively the Worldwide LHC Computing Grid (WLCG) [5] infrastructure.

2 HIGGS BOSON SEARCHES IN THE ATLAS EXPERIMENT

In 2011, the LHC delivered an integrated luminosity of more than 5 fb^{-1} of proton-proton collisions at the energy of $7\,\mathrm{TeV}$. This outstanding performance allowed ATLAS to collect and analyze up to $4.9\,\mathrm{fb}^{-1}$ of useful data to update its previous searches for the Higgs boson.

Searches of the Higgs boson by the ATLAS experiment are based on the combination of results from various Higgs decay channels [6]. The three most sensitive channels are: $H \to ZZ^{(*)} \to l^+l^-l^+l^-$ (decay into two Z bosons and than into four leptons, i.e. muons or electrons), $H \to \gamma\gamma$ (decay into two photons) and $H \to WW^{(*)} \to l^+\nu l^-\bar{\nu}$ (decay into two W bosons and then into leptons and neutrinos). Also the other channels with W and Z bosons are taken into the combination:

$H \to ZZ \to l^+l^-\nu\bar{\nu}$, $H \to ZZ \to l^+l^-q\bar{q}$, ($q$ denotes the quark, which becomes an origin of a jet, i.e. a narrow cone of particles) and $H \to WW \to l\nu qq'$. The tau channels $H \to \tau\tau \to l\tau_{had}3\nu$ (where τ denotes the tau lepton and τ_{had} is the tau lepton undergoing the hadronic decay producing a jet) and $H \to \tau\tau \to l^+l^-4\nu$ are also considered.

The 95 % confidence level (C.L.) cross-section limits set by the individual channels are shown in Figure 1. The limits are presented in units of the Standard Model expectation set by the individual channels using the CL_s prescription described in [7, 8]. If the solid line in the plot drops below “1”, then this Higgs boson mass interval is excluded. When it is significantly above the dashed line, corresponding to the Standard Model predictions without Higgs, then it is a hint of the Higgs observation at a given mass.

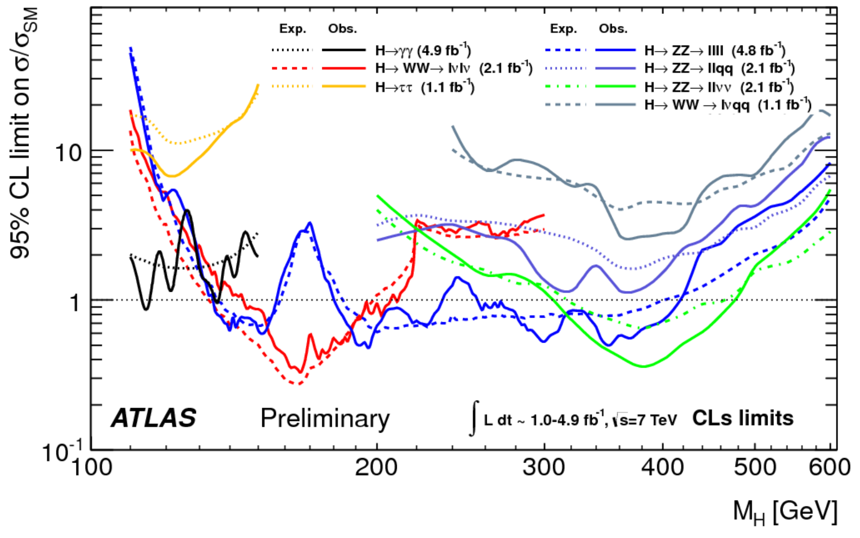


Figure 1. The expected (dashed) and observed (solid) cross-section limits for the individual search channels, normalized to the Standard Model Higgs boson cross section, as functions of the Higgs boson mass. These results use the profile likelihood technique with 95 % C.L. limits using the CL_s prescription.

All these individual channels are combined together to increase precision of the measurement. With the collected data, the Standard Model Higgs boson masses are excluded at the 95 % C.L. at ranges from 112.7 GeV to 115.5 GeV, 131 GeV to 237 GeV and 251 GeV to 468 GeV, while the expected Higgs boson mass exclusion in the absence of a signal ranges from 124.6 GeV to 520 GeV (see Figure 2). An excess of events is observed at Higgs mass around 126 GeV with statistical significance of 3.6 σ.

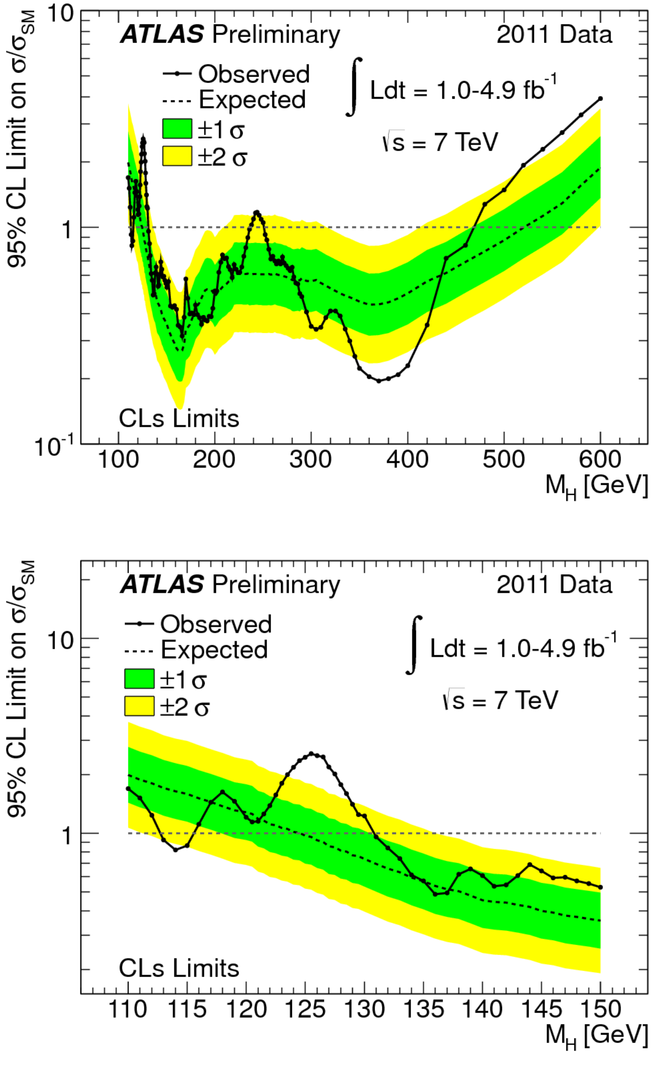


Figure 2. The combined upper limit on the Standard Model Higgs boson production cross section divided by the Standard Model expectation as a function of Higgs boson mass (solid line) for the Higgs mass range 100 GeV to 600 GeV (top plot) and 110 GeV to 150 GeV (bottom plot). This is a 95 % C.L. limit using the CL_s method in the entire mass range. The dotted line shows the median expected limit in the absence of a signal and the green and yellow bands reflect the corresponding 68 % and 95 % expected regions.

The deficit of events observed in the excluded mass range, and in particular between 300 GeV and 400 GeV is mainly due to the concordance of various small deficits in several high mass channels. The observed exclusion covers a large part of the expected exclusion range, except at low and high Higgs boson mass hypotheses where excesses of events are observed, and at around 245 GeV where the excess mostly seen in the $H \to ZZ^{()} \to l^+l^-l^+l^-$ channel is present.

The confidence level with which the Standard Model Higgs boson is excluded is shown in Figure 3. It should be noted that a signal of the strength predicted by the Standard Model is excluded at high confidence for 360 GeV, while an exclusion Confidence Level (CL_s) in excess of 99 % is observed in the regions between 133 GeV and 230 GeV and between 260 GeV and 437 GeV. The strongest exclusions have false exclusion rates at a level of one per million.

The local significance of an excess is estimated using a consistency test of the observation with the background-only hypothesis. It is estimated by the p_0 probability that a background-only experiment is more signal-like than the observed one. The probability p_0 is constructed to be equal to 50 % for downward fluctuations of the background and smaller than 50 % when more events are observed than expected. This probability is displayed as a function of the Higgs boson mass hypothesis in Figure 4. This excess of events over the background-only hypothesis is seen in $H \to ZZ^{(*)} \to l^+l^-l^+l^-$ and $H \to \gamma\gamma$ channels and is supported by a broad low-significance excess in the $H \to WW^{(*)} \to l^+\nu l^-\nu$ channel. The results from three strongest channels are consistent showing an excess of events for about 125 GeV Higgs boson mass hypothesis.

The local 3.6 σ significance of the excess observed at Higgs boson mass hypotheses around 126 GeV is corrected for the fact that the local excess could have appeared anywhere in the mass region in which the Higgs boson has been searched for. This is known as the *look-elsewhere effect* [9]. Taking this effect into account the global probability of such an excess to occur in the full search range is approximately 1 %, corresponding to 2.3 σ [6].

Such an excess is not significant enough to be a proof of the Higgs boson existence and it is too early to draw definite conclusions. More studies and more data are needed. But nevertheless already this very early analysis allowed for the exclusion of the Standard Model Higgs boson existence in the wide mass range.

3 GRID COMPUTING IN THE YEAR 2011 FOR HIGGS SEARCHES

Results of the analysis for Higgs searches are based on a full data sample collected in the year 2011 [10]. The high number of pp collisions sampled in 2011 was achieved at the cost of increased number of collisions per bunch crossing (pileup) that reached value of $\langle\mu\rangle = 16$ which is three times higher than in the year 2010 and close to the maximum expected at the full LHC design luminosity. The increased instantaneous luminosity and the pileup have been a challenge to ATLAS computing operations by increasing track density in the recorded events which in turn increased the data

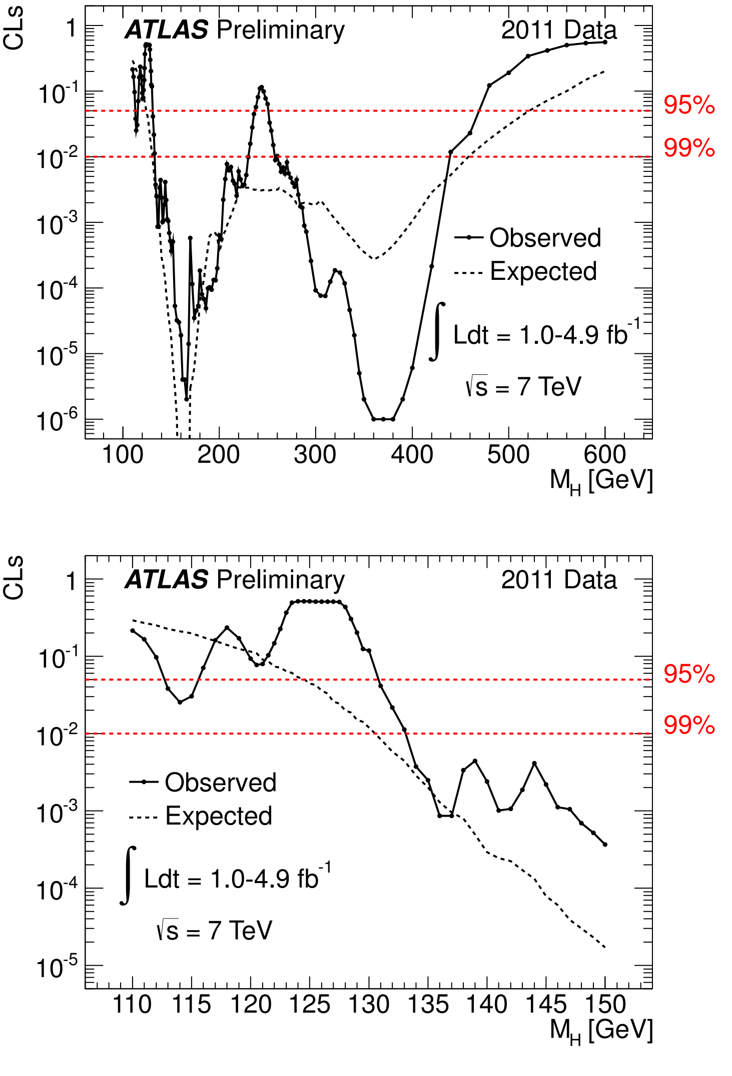


Figure 3. The value of the combined CL_s for $\mu = 1$ (testing the Standard Model Higgs boson hypothesis) as a function of Higgs boson mass in the full mass range of this analysis (upper plot) and in the low mass range (lower plot). By definition, the regions with $CL_s < \alpha$ are considered excluded at the $(1-\alpha)$ C.L. or stronger. When the best-fit value of the strength parameter exceeds the tested signal hypothesis $(\mu = 1)$ the observed CL_s is bound to be equal to $50\,\%$ by construction.

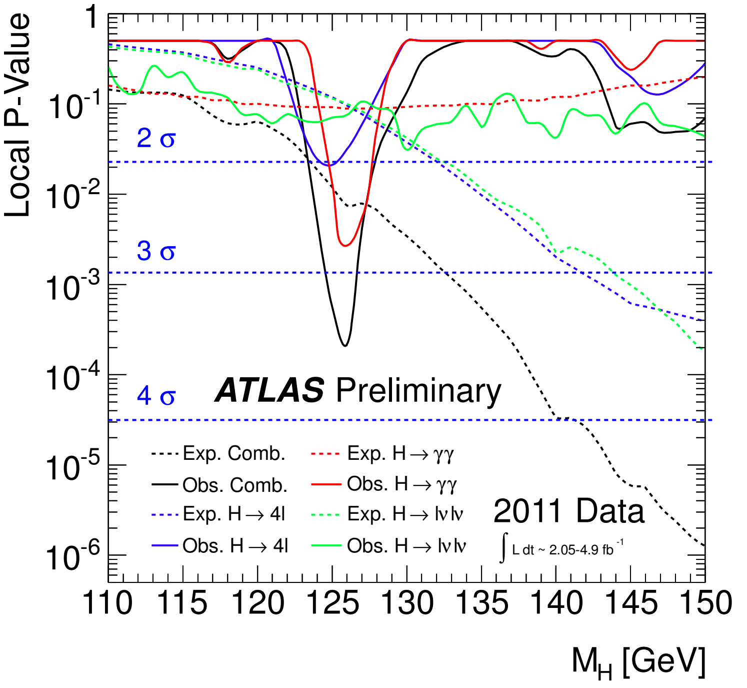


Figure 4. The consistency of the observed results with the background-only hypothesis for the three strongest channels and the combination in the low mass region. The dashed curves show the median expected significance in the hypothesis of a Standard Model Higgs boson production signal, which is about equal for all three of these channels near 125 GeV.

volume that had to be handled and required more computer time spent on reconstruction.

ATLAS computing dealt well with the challenge by using intensively and efficiently the WLCG Grid [5] computing infrastructure (see Figure 5). In 2011 ATLAS has been using on average 224 kHS06 (90 %) pledged resources at Tier1 sites and 324 kHS06 (115 %) at Tier2 sites with efficiency close to 90 %. The available disk space at Tier1 sites has been used at 82 % while only 44 % at Tier2 sites thanks to the modified policy to distribute data for analysis mostly on demand [11]. Altogether, ATLAS wrote 1 950 TB of RAW data to disk for the 1 852 M pp collision events collected after trigger selection. Reconstruction produced further 2 912 TB of intermediate data plus 112 TB of ROOT Ntuples for analysis.

In High Energy Physics (HEP) experiments, the analysis of real collision data is usually accompanied by a comparison with simulated collisions based on Monte Carlo models of the studied processes. Thus, in the ATLAS computing operations

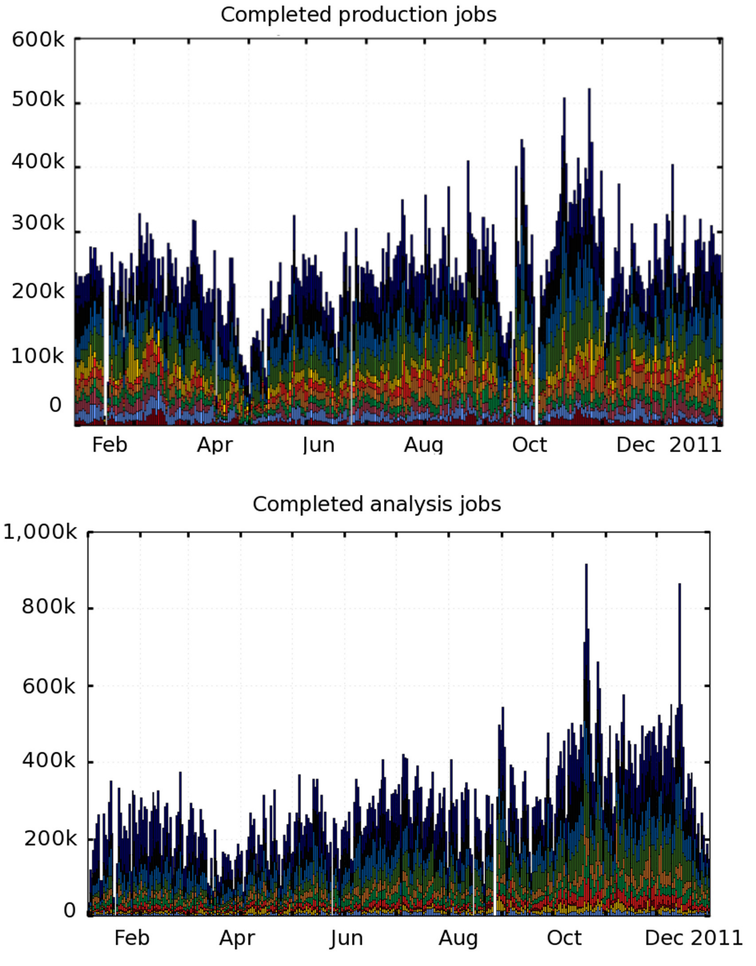


Figure 5. Number of completed jobs per day in ATLAS production (upper plot) and analysis (lower plot) on WLCG Grid in the year 2011

there were two chains of production running in parallel, for real and simulated data. The real data processing proceeded in several steps. First RAW data were promptly processed at Tier0 computing center at CERN and together with reconstructed data were subsequently distributed to 10 ATLAS Tier-1 centers all over the world. Late in the summer, a reprocessing with improved calibration information and reconstruction software had run mostly at Tier1 sites which processed over 900 million events collected during the first 6 month of data taking in just a month. After reconstruction, individual physics groups had run their analysis software in a special group version of central production still benefiting from the common software distribution and WLCG computing infrastructure. Intensity of this production grew strongly over the year, starting from 12 500 and increasing to 40 000 completed jobs per day towards the end of the year with peaks up to several hundred thousands jobs per day, taking up to 30 % of CPU time at Tier1 sites at times. Simulations of results of the passage of particles produced in the collision through the ATLAS detector were running continuously on the Grid at Tier1 and Tier2 sites, with an average of approximately 50 000 jobs running concurrently and with approximately 1.6 billion events simulated in 2011. These were the most time consuming calculations in Atlas production. After the simulation step the same version of reconstruction and production of group derived data (DPD, Ntuples) used on RAW detector data was run also on simulated collisions (Figure 6).

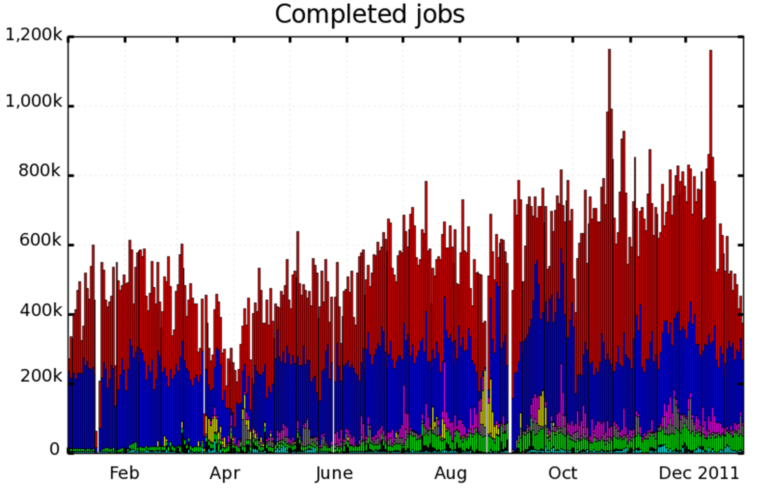


Figure 6. Development in the number of completed ATLAS jobs grouped by production type in year 2011. Most time consuming activities shown in the plot are user analysis (top part of histogram), simulations (middle part), group production (lower part).

The distribution of the number of jobs completed for different types of production and of the amount of CPU time used at the three Tier0, Tier1 and Tier2 site categories are shown in Figures 7 and 8. One can see that a big part of CPU power comes from 50 Tier2 sites. They contributed mainly to the production of collision simulations and participated in user analysis. In 2011 some 1 500 ATLAS scientists has been actively running their analysis jobs with increasing intensity. In total more than 105 million jobs has been submitted to the Grid, individual analysis using 50 % of pledged CPU resources (see Figure 8). In the end of 2011 the individual analysis and group production has become so intensive at Tier1 sites that ATLAS had to reduce the allowed share for analysis at these sites and introduce new fair-share mechanisms between various flavors of official production to preserve priority for the main central production work especially for the digitization and reconstruction of Monte Carlo simulations.

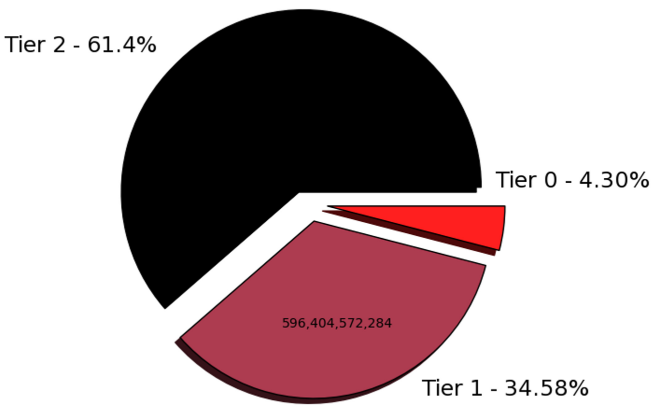


Figure 7. CPU used by ATLAS at Tier0, Tier1 and Tier2 computing sites on WLCG Grid

The entire ATLAS and WLCG computing and software chain has performed very well throughout 2011 data-taking (see Figure 5) and thanks to this fact ATLAS was able to update results of its analyses, like the results for Higgs particle searches presented in December 2011, soon after the last collision was delivered by LHC [12]. This success in using Grid resources was possible only after LHC experiments implemented their dedicated production systems that complement basic Grid tools and allowed to handle this massive task of submitting and handling in parallel thousands of production jobs. Without them the amount of manual work would be too much to handle. For example, in ATLAS, production system components like PanDA [13] and its helpers DQ2 and ProdSys serve to define and manage production tasks, prepare description, submit and manage jobs, coordinate computing and data transfer activities according to complex experiment computing models and specific task needs, aggregate, integrate and register output results in central catalogs. As practice taught us a large distributed computing system like Grid is extremely complex, some of the computing sites or services going out and back into the production, some simply not working properly. The most important role of a production system is to make it as much as possible resistant against such faults in Grid service operations. PanDA has been designed with this in mind and performs very well. It has a pilot job based submission system for "just in time" workload management. The pilot job is doing many sanity checks on worker node (WN) before pulling real job, checking properties and available resources that cannot be specified in a standard job description language (jdl). PanDA provides also insulation from grid specific

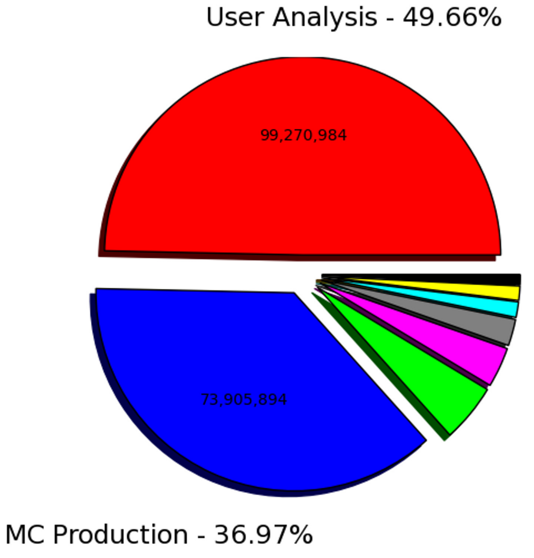
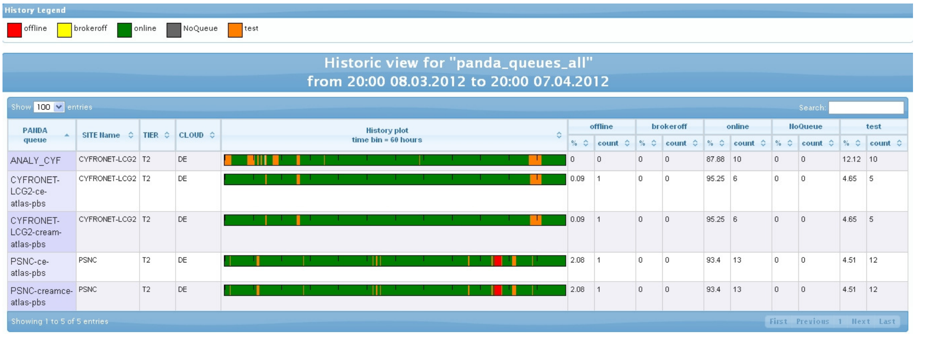


Figure 8. Completed ATLAS jobs grouped by production type

middleware deficiencies, submission latencies and failures. It has a fast brokerage, with ATLAS defined priorities, user and group level controls which allow to manage the use of Grid resources according to well defined policies and current needs.

Having a robust production system is not enough to achieve a stable and highly efficient computing operations on a distributed computing system. When a component of the system fails it is necessary to stop using it, otherwise production system spends most of the time on fixing failing operations. In 2011 ATLAS has very much improved services that implement such model. This was achieved by the improved monitoring, a system of experiment specific tests and fast, automatic reaction to job failures. The monitoring and historical data based on experiment defined metrics for computing sites supporting ATLAS is provided by Site Status Board (SSB) [14]. This tool is used for ranking sites according to their reliability and performance. Detailed information about progress of different types of production and data transfers are shown in Data Distribution and Production Dashboards. These tools are used for the moment in a manual mode by ATLAS members for monitoring and managing production operations on special shifts distributed worldwide so they are active 24 hours a day. An important component of the site monitoring system are so called HammerCloud (HC) tests running regularly on all sites testing their functionality for ATLAS production. Results of these tests are collected in SSB and used for quick automatic exclusion of the failing Tier2 sites (Figure 9). This mechanism proved so successful that it is currently being extended from managing only analysis queues to production queues.



PANDA queue	SITE Name	TIER	CLOUD	History plot time bin = 60 hours	offline		brokeroff		online		NoQueue		test	
					%	count	%	count	%	count	%	count	%	count
ANALY_CYF	CYFRONET-LCG2	T2	DE		0	0	0	0	87.88	10	0	0	12.12	10
CYFRONET-LCG2-ce-atlas-pbs	CYFRONET-LCG2	T2	DE		0.09	1	0	0	95.25	6	0	0	4.65	5
CYFRONET-LCG2-cream-atlas-pbs	CYFRONET-LCG2	T2	DE		0.09	1	0	0	95.25	6	0	0	4.65	5
PSNC-ce-atlas-pbs	PSNC	T2	DE		2.08	1	0	0	93.4	13	0	0	4.51	12
PSNC-creamce-atlas-pbs	PSNC	T2	DE		2.08	1	0	0	93.4	13	0	0	4.51	12

Figure 9. Example of monitoring of site availability in one of the SSB views

The illustration of the success of the implemented automatic tools for monitoring and management of queues was high efficiency and low failure rates in ATLAS central production. Introduction of Tier2 site categories ranking them according to reliability of their services and performance of data transfers has increased rate of success in user analysis at these sites. The combination of monitoring and analysis of historical data allowed to detect the already mentioned fierce competition for Tier1 resources among different types of production and analysis tasks in the end of 2011. The solution implemented by ATLAS is to assign more different production tasks, like simulation, digitization and reconstruction that used to run at Tier1 sites now also to Tier2 sites. However, in order to make this move successful one needs to assure a good data transfer capabilities and high reliability of Tier2 site operations. This is the direction of efforts that ATLAS and the WLCG Grid services support are working on.

4 ATLAS COMPUTING PRODUCTION IN POLAND

In Poland ATLAS experiment is supported by two computing centers, Academic Computer Centre CYFRONET AGH at Kraków, and Supercomputing and Networking Center at Poznań. Both are members of the Polish Tier2 in WLCG Grid organization. Together they provided more than 1.5 % of all the pledged computing resources for ATLAS in 2011, effectively delivering 2.9 % wall processing time in production and 2.2 % in analysis (see Figure 10). The sizable contribution from Polish sites to ATLAS production is illustrated on the plot of the total number of completed ATLAS jobs at Tier2 sites, where Poland is ranking the 8$^{\text{th}}$ among all Tier2 sites in the WLCG worldwide Grid (see Figure 11).

Among Polish sites, Cyfronet is the larger one, with a dedicated support for ATLAS. It is configured in ATLAS production system both for running simulations and other central data processing and it is also available for running user analysis

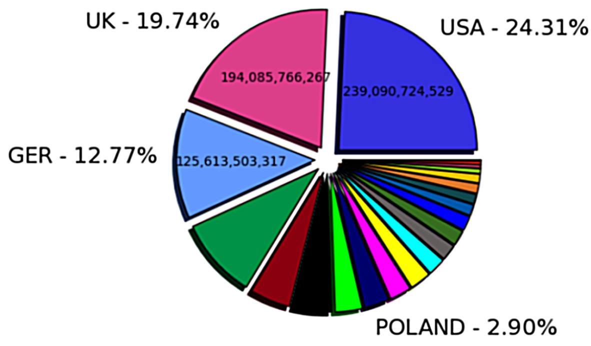


Figure 10. Wall time used by ATLAS jobs on WLCG Grid sites at different countries in 2011 (2.9 % in Poland)

jobs. The analysis activity generates large data transfers bringing input data to the site and transferring out results as illustrated by a plot in (Figure 12).

In 2011 Cyfronet has run more than 91 % of jobs for this experiment in Poland. The PSNC site is mostly dedicated to ALICE experiment, but provides limited share also for ATLAS. In 2011 it processed close to 300 000 jobs delivering 9 % of ATLAS processing power in Poland (see Figure 13).

5 SUMMARY

With the data collected in 2011 ATLAS has excluded the existence of Standard Model Higgs boson in a wide mass range. An excess of events is observed at Higgs mass hypotheses close to 126 GeV. The global probability of such an excess to occur in the full search range is approximately 1 %, corresponding to 2.3σ. If ATLAS collects in 2012 four times more data than in 2011 (about 20 fb^{-1}, as planned) we should be able to get a decisive answer, whether the Higgs boson exists and what is its mass.

The quick analysis of the enormous amount of collected data and obtaining of the presented results was only possible thanks to an excellent performance of the LHC Computing Grid in which Polish computing centers, the ACK Cyfronet AGH Krakow, ICM Warsaw and PSNC Poznan, also participated.

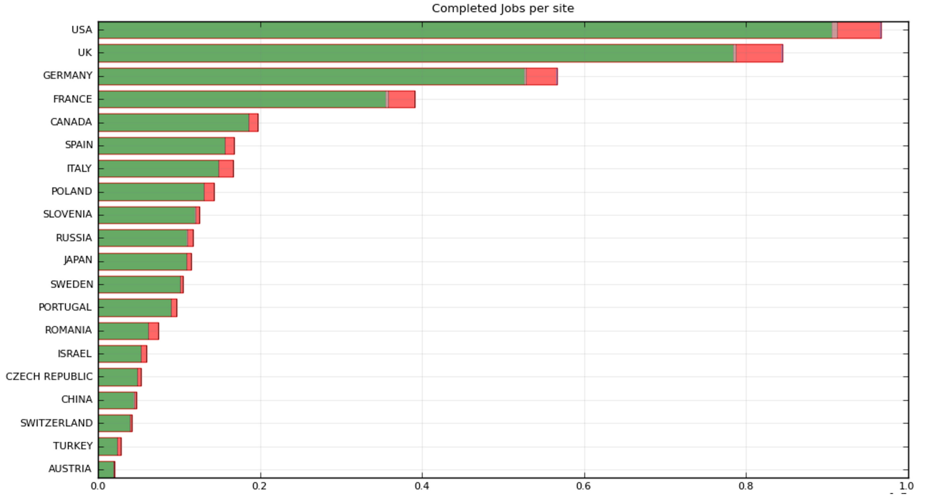


Figure 11. Completed ATLAS production jobs at Tier2 sites on WLCG Grid in 2011, classified by status of finished job: success (left part of the bin), aborted (right part), site-failed (middle part)

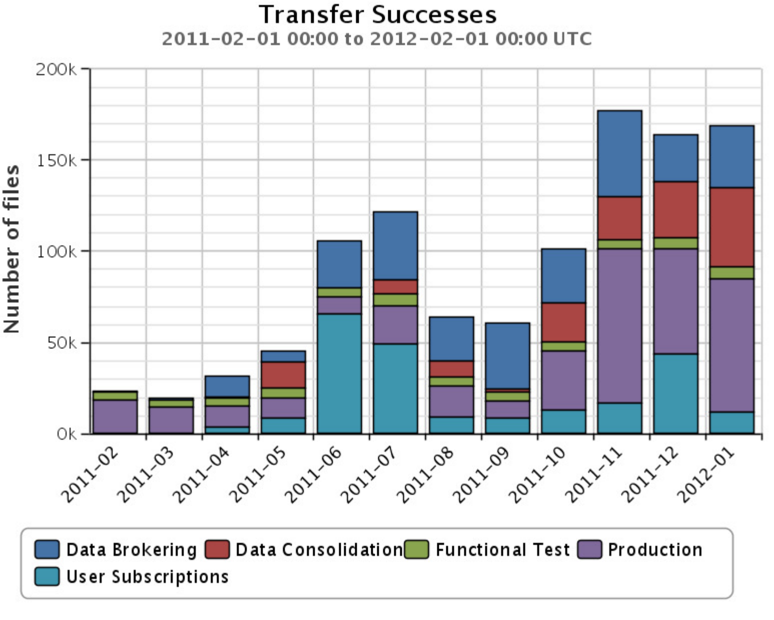


Figure 12. Evolution of the amount of data (number of files) transferred to Cyfronet in 2011

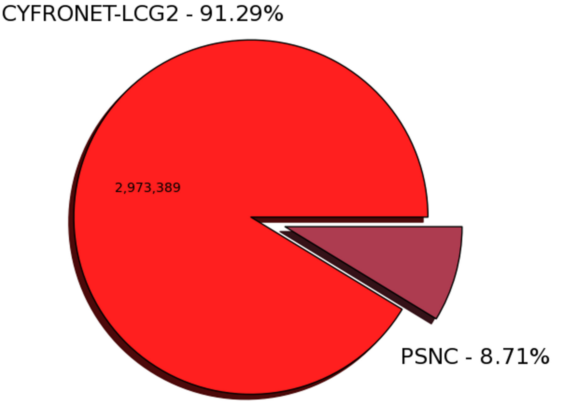


Figure 13. Completed ATLAS jobs at Polish sites in 2011 (CYF 91 %, PSNC 9 %)

Acknowledgement

The work was supported in part by Polish Government grants 665/N-CERN-ATLAS/2010/0, NN202 127937874 (years 2009-2011) and 764/N-DAAD/2010/0.

REFERENCES

[1] Englert, F.—Brout, R.: Broken Symmetry and the Mass of Gauge Vector Mesons. Phys. Rev. Lett., Vol. 13, 1964, pp. 321–323.

[2] Higgs, P. W.: Broken Symmetries and the Masses of Gauge Bosons. Phys. Rev. Lett., Vol. 13, 1964, pp. 508–509.

[3] Higgs, P. W.: Broken Symmetries, Massless Particles and Gauge Fields. Phys. Lett., Vol. 12, 1964, p. 132.

[4] Guralnik, G. S.—Hagen, C. R.—Kibble, T. W. B.: Global Conservation Laws and Massless Particles. Phys. Rev. Lett., Vol. 13, 1964, pp. 585–587.

[5] LCG project, `http://lcg.web.cern.ch/LCG`.

[6] ATLAS Collaboration, Combination of Higgs Boson Searches with Up to 4.9 fb^{-1} of pp Collisions Data Taken at a Center-of-Mass Energy of 7 TeV with the ATLAS Experiment at the LHC. ATLAS-CONF-2011-163, Dec. 2011, `https://cdsweb.cern.ch/record/1406358`.

[7] Read, A. L.: Modified Frequentist Analysis of Search Results (the CL_s Method). `http://cdsweb.cern.ch/record/451614/`.

[8] Read, A. L.: Presentation of Search Results: The CL(s) Technique. J. Phys., Vol. G28, 2002, pp. 2693–2704.

[9] Gross, E.—Vitells, O.: Trial Factors for the Look Elsewhere Effect in High Energy Physics. Eur. Phys. J., Vol. C70, 2010, pp. 525–530.

[10] Gianotti, F.: Presentation at the LHC C-RRB. CERN-RRB-2011-109, October 2011.

[11] Bozzi, C. et al.: Presentation of the CRSG Report to the LHC C-RRB. CERN-RRB-2011-076, October 2011.

[12] Gianotti, F.: Presentation at the CERN Public Seminar. `https://indico.cern.ch/getFile.py/access?contribId=0&resId=1&materialId=slides&confId=164890`, December 2011.

[13] Maeno, T. et al.: PanDA: Distributed Production and Distributed Analysis System for ATLAS. J. Phys.: Conf. Ser., Vol. 119, 2008, 062036.

[14] Site Status Board (SSB). `http://dashb-atlas-ssb.cern.ch/dashboard/request.py/siteviewhome`.

Andrzej Olszewski has received his Ph. D. in high energy physics in 1993 from the Institute of Nuclear Physics in Kraków, Poland. Since 2004 he is a member of ATLAS experiment on LHC accelerator, studying physics of heavy ion collisions and contributing to the experiment software development and computing operations on WLCG grid.

Marcin Wolter has obtained his Ph. D. degree in physics from the Swiss Federal Institute of Technology Zurich in 1996. Since then he has joined the ATLAS group of the Institute of Nuclear Physics in Kraków, Poland. He has been working on development of new detectors for ATLAS experiment and on the physics analysis of collected data. He has been involved in the development of multivariate analysis methods used in ATLAS tau identification.