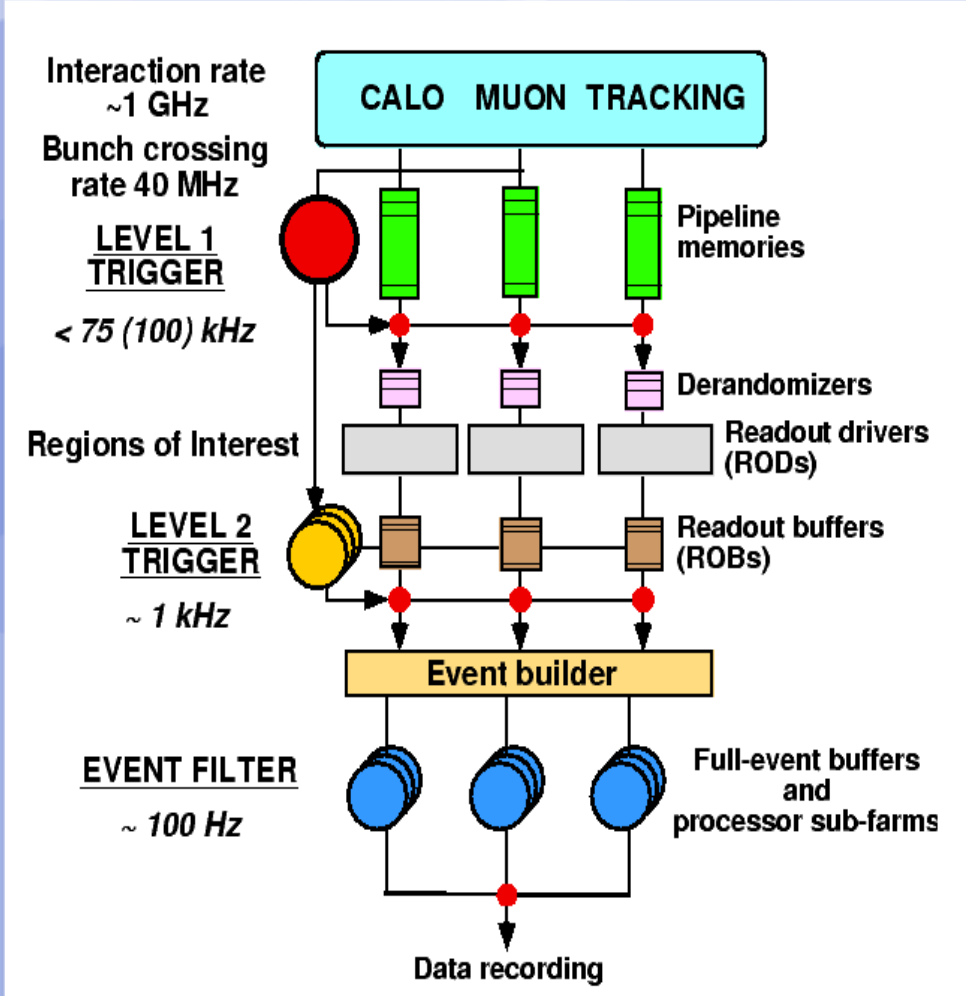
Software Validation Infrastructure for the ATLAS High-Level Trigger

Danilo Enoque Ferreira de Lima
(Federal University of Rio de Janeiro - COPPE/Poli)

On behalf of the
ATLAS Trigger Validation Group

Goals



- **Select interesting physics events while rejecting an overwhelming background (5 events are accepted per million background events)**
  - Trigger false negative should be kept to a minimum
  - Statistical bias and software implementation problems should not exist at all
- **Allow online monitoring and offline validation to check results with respect to previous running conditions**

HLT Validation

- The trigger code will evolve continuously and needs to be well tested offline before it goes online
- Validation effort should be automatized and allow for easy code debugging
- A common infra-structure, using specific checks provided by developers is encouraged
  - As a consequence, common tools can be used to check and make results promptly available

Development for the ATLAS software

- A set of new releases are made each night
  - There are different branches depending on specific project targets
- The NIghtly Control System (NICOS) controls the build of the packages
- Trigger Validation uses ATN and RTT which are general-purpose ATLAS validation machinery
  - ATN (Atlas Testing Nightly) tests can check code integrity as soon as a new release is built. It runs on all new releases to test for inconsistencies using simple scripts
  - RTT (Run Time Tester) tests can run longer jobs and check the physics output
- As soon as a problem is found, the developers are informed through an automatic email sent by the infra-structure

NICOS results

This web page shows the list of ATLAS nightlies (see twiki page for details). Some nightlies are not added to this list: experimental nightlies, doxygen builds, stable releases, tests of a new NICOS version.

Nightly testing summaries: ATN (rel_0, rel_1, rel_2, rel_3, rel_4, rel_5, rel_6, also accessible from nightly web pages), RTT, FCT (AID scans), TCT (AID scans).

New column "Kit" shows the results of kit builds and downloads (made with package PackDist). The red crosses or green ticks images are links to the corresponding logfiles

Message of a day about nightlies problems and usability

```
Oct 18, 2008: rel_6 14.X.0 32 bit opt nightly has missing libraries, dbg version is OK
Oct 11, 2008: 14.X.0 32 bit opt nightly is not completed: build machine problem
Oct 4, 2008: copy-to-AFS processes of many Sat. nightlies (incl. all MIG mightlies) are stuck because of AFS problems
```

Nightly Title	# Platforms	# Projects	Latest Rel.	Build	Date	Copy	Kit	Ave. Failed Builds	Ave. Test Success(%)
MAJOR NIGHTLIES									
14.X.0	4	10	rel_0	done	10/26 07:34	done	rel_0	1.4	48.8
14.X.0-VAL	2	10	rel_0	done	10/26 07:22	done	N/A	0.8	54.0
14.2.2X NIGHTLIES									
14.2.2X	2	10	rel_0	done	10/26 07:36	done	N/A	0	62.2
14.2.2X-VAL	2	10	rel_0	done	10/26 07:55	done	N/A	0	63.8
PATCH NIGHTLIES									
14.2.2X.Y-P1HLT	2	1	rel_0	done	10/26 07:25	done	rel_0	0	27.0

NICOS/ATN by Alex Undrus (BNL)

Package built

Package Name	Container	Build	Checkreq Test	Unit Test	Manager(s)
AtlasTriggerRelease	N/A		N/A	FAIL	emil.obreshkov at cern.ch,david.quarrie at cern.ch
TrigCaloRec	Trigger/TrigAlgorithms		N/A	N/A	cibran.santamarina.rios at cern.ch,xella at nbi.dk,delsart at lps.umontreal.ca,fer
TrigCaloRinger	Trigger/TrigAlgorithms		N/A	N/A	rodrigo.torres at cern.ch,daniloefl at gmail.com
TrigEFMissingET	Trigger/TrigAlgorithms		N/A	N/A	diego.casadei at cern.ch,kyle.cranmer at cern.ch,xiaowen.lei at cern.ch,ignacio.a
TrigT2CaloCommon	Trigger/TrigAlgorithms		N/A	N/A	xin.wu at physics.unige.ch,pilar.casado at cern.ch,kyle.cranmer at cern.ch,damaz
TrigT2CaloCosmic	Trigger/TrigAlgorithms		N/A	N/A	damazio at mail.cern.ch
TrigT2CaloEgamma	Trigger/TrigAlgorithms		N/A	N/A	xin.wu at ph
TrigT2CaloJet	Trigger/TrigAlgorithms		N/A	N/A	kyle.cran
TrigT2CaloTau	Trigger/TrigAlgorithms		N/A	N/A	olga.igonkina a
TrigT2MinBias	Trigger/TrigAlgorithms		N/A	N/A	w.bell at cern.ch,regina.kwee at cern.ch
TrigValAlgs	Trigger/TrigValidation		N/A	N/A	goncalo at pp.rhul.ac.uk,strom at physics.uoregon.edu
TrigmuIso	Trigger/TrigAlgorithms		N/A	N/A	stefano.giagu at cern.ch
TrigNavigation	Trigger/TrigEvent		N/A	FAIL	tomasz.bold at cern.ch,carlo.schiavi at ge.infn.it
TrigSteering	Trigger/TrigSteer		N/A	FAIL	nicolas.berger at cern.ch,till.eifert at cern.ch
TrigTimeAlgs	Trigger/TrigTools		N/A	FAIL	j.t.m.baines at rl.ac.uk,tomasz.bold at cern.ch
AnalysisTriggerAlgs	PhysicsAnalysis/AnalysisTrigger		N/A	N/A	wolfgang.ehrenfeld at cern.ch,tmaeno at bnl.gov,atw at hep.ph.bham.ac.uk,mon
AtlasTriggerRunTime	N/A		N/A	N/A	emil.obreshkov at cern.ch,david.quarrie at cern.ch

Build status for each package:
Clicking on the package,
the build log is shown.

Summary of ATN tests

An ATN test and its results:
The log file, error messages and other outputs are available.

Trigger ATN test results summary

Nightly test: 14X032BS4TrgOpt rel_5

Other nightlies: 0 1 2 3 4 5 6

Test name	Test script	Athena exit	Error Msgs	Reg. tests	Rootcomp	Exit code	Post cmd	Dir. link	Log link
AthenaModernBS_standalone	-	OK	OK	FAIL	MISMATCH [ps]	68	OK	dir	testAthenaModernBS_standalone_tail.log testAthenaModernBS_standalone_test.log
AthenaModernRDO	-	OK	OK	FAIL	MISMATCH [ps]	68	OK	dir	testAthenaModernRDO_tail.log testAthenaModernRDO_test.log
AthenaModernRDO_blackholes	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_blackholes_tail.log testAthenaModernRDO_blackholes_test.log
AthenaModernRDO_full	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_full_tail.log testAthenaModernRDO_full_test.log
AthenaModernRDO_full_no_Bphysics	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_full_no_Bphysics_tail.log testAthenaModernRDO_full_no_Bphysics_test.log
AthenaModernRDO_full_no_Bphysics_no_prescale	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_full_no_Bphysics_no_prescale_tail.log testAthenaModernRDO_full_no_Bphysics_no_prescale_test.log
AthenaModernRDO_full_no_prescale	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_full_no_prescale_tail.log testAthenaModernRDO_full_no_prescale_test.log
AthenaModernRDO_lumi0.01	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_lumi0.01_tail.log testAthenaModernRDO_lumi0.01_test.log.gz
AthenaModernRDO_lumi0.01_no_Bphysics	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_lumi0.01_no_Bphysics_tail.log testAthenaModernRDO_lumi0.01_no_Bphysics_test.log
AthenaModernRDO_lumi0.01_no_Bphysics_no_prescale	-	OK	FAIL	FAIL	MISMATCH [ps]	100	OK	dir	testAthenaModernRDO_lumi0.01_no_Bphysics_no_prescale_tail.log testAthenaModernRDO_lumi0.01_no_Bphysics_no_prescale_test.log

Done

Trigger tests results on ATN

http://fernando.web.cern.ch/fernando/TriggerSummaryPage/TriggerHLTP1TestPage.html

- A web page that summarizes the ATN test results for different building projects, like AtlasTrigger, AtlasAnalysis, HLTTesting and AtlasPoint1
  - The aim is to show every relevant information from NICOS in one single page for all available builds, each night
- Updated automatically every hour
- Quick and self-explanatory traffic light system (red, yellow, green)
- The summary is shown as a HTML table with colored cells
- When clicked, cell shows a menu with links to full information: log files, ROOT files, directory with results, ...

Trigger tests results on ATN

Show builds: ☑ 14.2.2X.Y ☑ 14.2.2X.Y-VAL ☑ bugfix

Package	Build	rel_0	rel_1	rel_2
HLT Testing				
HelloWorldMT_run-stop-run	14.2.2X.Y	0 X 0	0 X 0	0 X 0
	14.2.2X.Y-VAL	0 X 0	0 X 0	0 X 0
	bugfix	X X X	0 X 0	0 X 0
HelloWorldPT_run-stop-run	14.2.2X.Y	0 X 0	0 X 0	0 X 0
	14.2.2X.Y-VAL	0 X 0	0 X 0	0 X 0
	bugfix	X X X	0 X 0	0 X 0
HistOH	14.2.2X.Y	3 X 0	0 X 0	3 X 0
	14.2.2X.Y-VAL	0 X 0	0 X 0	0 X 0
	bugfix	X X X	0 X 0	0 X 0
Athena	14.2.2X.Y	0 0 0 R	0 0 0 R	0 0 0 R
	14.2.2X.Y-VAL	0 0 0 R	0 0 0 R	0 0 0 R
	bugfix	X X X	357 >99 X R	356 >99 X R

3 X 0

3 X 0

3 errors in log

Unknown Reg test differences

0 exitcode

test summary

Summary log file

Checklog file

Directory with results

Tail of log file

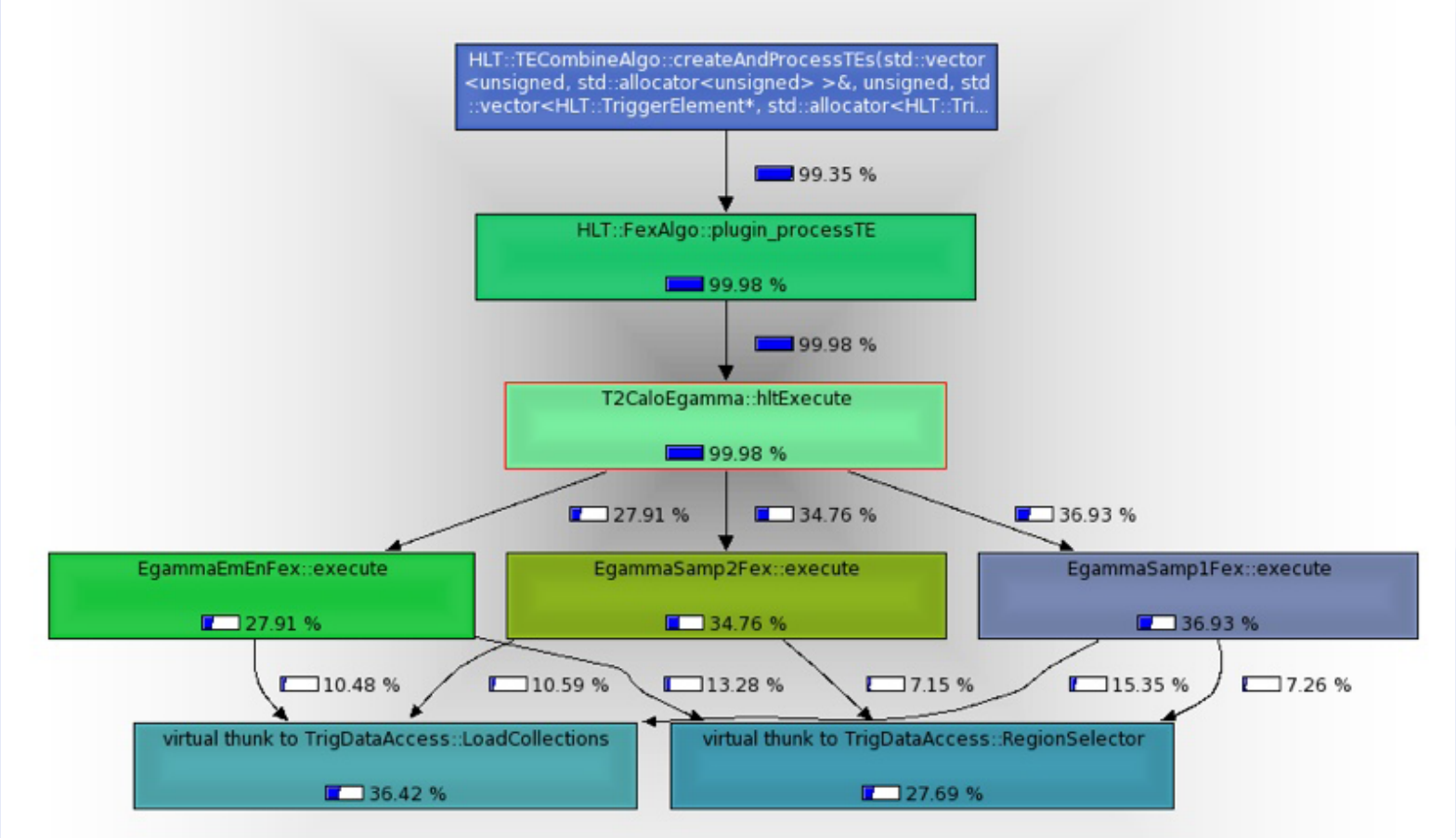
Complete log file

Memory monitoring

- Memory leaks must be identified
  - Continuous monitoring of memory usage
  - Memory leak checks are done on special jobs
  - Memory leak targets are 10 bytes/event on L2 and 1KB/event on EF (the less, the better)
  - PerfMon is a tool to monitor memory leaks
    - It generates graphs of the memory usage per event processed
    - It is enabled on all RTT jobs

Memory monitoring

- Valgrind is a set of very powerful tools for software quality and performance study
  - It has been used continuously to check the algorithms performance (such as memory leaks and CPU usage)



Memory monitoring

Trigger RTT PerfMon results - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://atlas-project-trigger-release-validation.web.cern.ch/atlas-project-trigger-release-valic

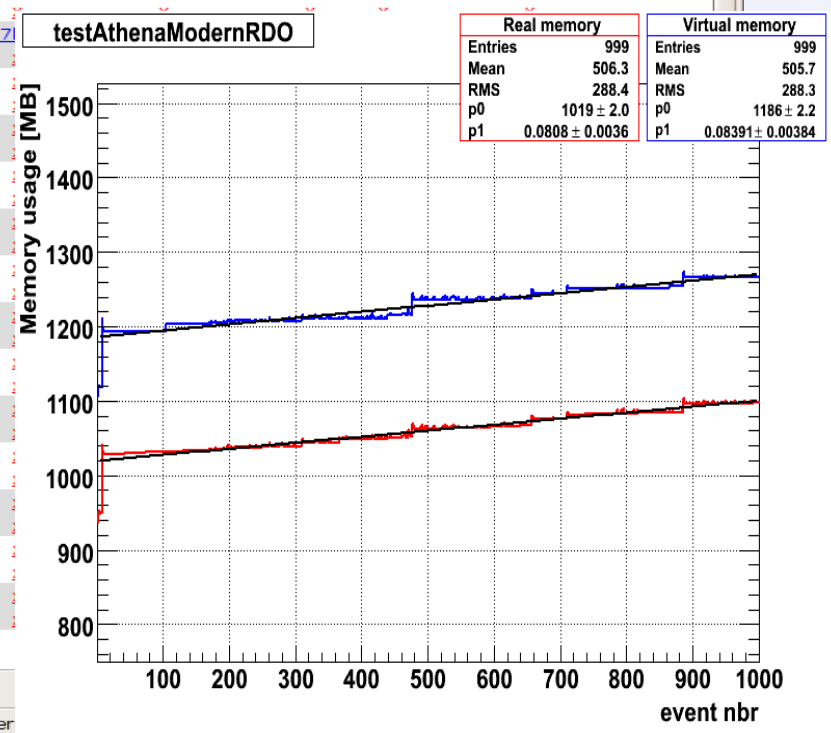
ATLAS HN PerfMon Trigger CVS Root BaBar Calendar ELOG MMM LEO LEO Spiegel Heise Interactions Wetter GOL

Summary of Trigger PerfMon RTT results

The tuple in each cell is the result of a linear fit to the vmem/event graph, i.e. the first number is the initial **virtual memory** consumption (in MB) and the second number the memory increase per event (in kB). **Left-click on the links for more options.**

Test		rel_0	rel_1	rel_2	rel_3	rel_4	rel_5	rel_6
testAthenaModernRDO	pcache(slc3)	1186M,85.9k	1190M,86.3k	x	x	x	?	1184M,85.4k
	dev(slc4)		x	x	x	x	x	x
testAthenaModernRDO_memTrack	pcache(slc3)		3.1k	1187M,93.8k	1187M,95.9k	1186M,94.6k	?	1186M,96.7k
	dev(slc4)		x					
testAthenaModernRDOtoBS	pcache(slc3)		6.9k	958M,122.7				
	dev(slc4)		x					
testAthenaOldConfigRDO	pcache(slc3)		x					
	dev(slc4)		x					
testBjetSliceAthenaModernRDO	pcache(slc3)	854M,10.7k	871M,10.3k					
	dev(slc4)	x	x					
testBphysicsSliceAthenaModernRDO	pcache(slc3)	990M,59.9k	991M,60.2k					
	dev(slc4)	x	x					
testCaloAthenaModernRDO	pcache(slc3)	808M,0.0k	809M,0.0k					
	dev(slc4)	x	x					
testCosmicsSliceAthenaModernRDO	pcache(slc3)	x	x					
	dev(slc4)	x	x					
testElectronSliceAthenaModernRDO	pcache(slc3)	977M,22.2k	975M,20.8k					
	dev(slc4)	x	x					
testIDAthenaModernRDO	pcache(slc3)	852M,23.5k	853M,23.2k					
	dev(slc4)	x	x					
testJetSliceAthenaModernRDO	pcache(slc3)	804M,18.9k	x					
	dev(slc4)	x	x					
testMETSliceAthenaModernRDO	pcache(slc3)	933M,31.8k	934M,31.0k					
	dev(slc4)	x	x					
testMuonSliceAthenaModernRDO	pcache(slc3)	923M,10.8k	923M,12.3k					
	dev(slc4)	x	x					
testPhotonSliceAthenaModernRDO	pcache(slc3)	860M,14.4k	924M,32.6k					
	dev(slc4)	x	x					
testTauSliceAthenaModernRDO	pcache(slc3)	874M,29.9k	876M,28.9k					
	dev(slc4)	x	x					

Memory graph
PerfMon summary
Log file
RTT result page
[All Memory graphs]
status: success
done: 07/10/14 05:54

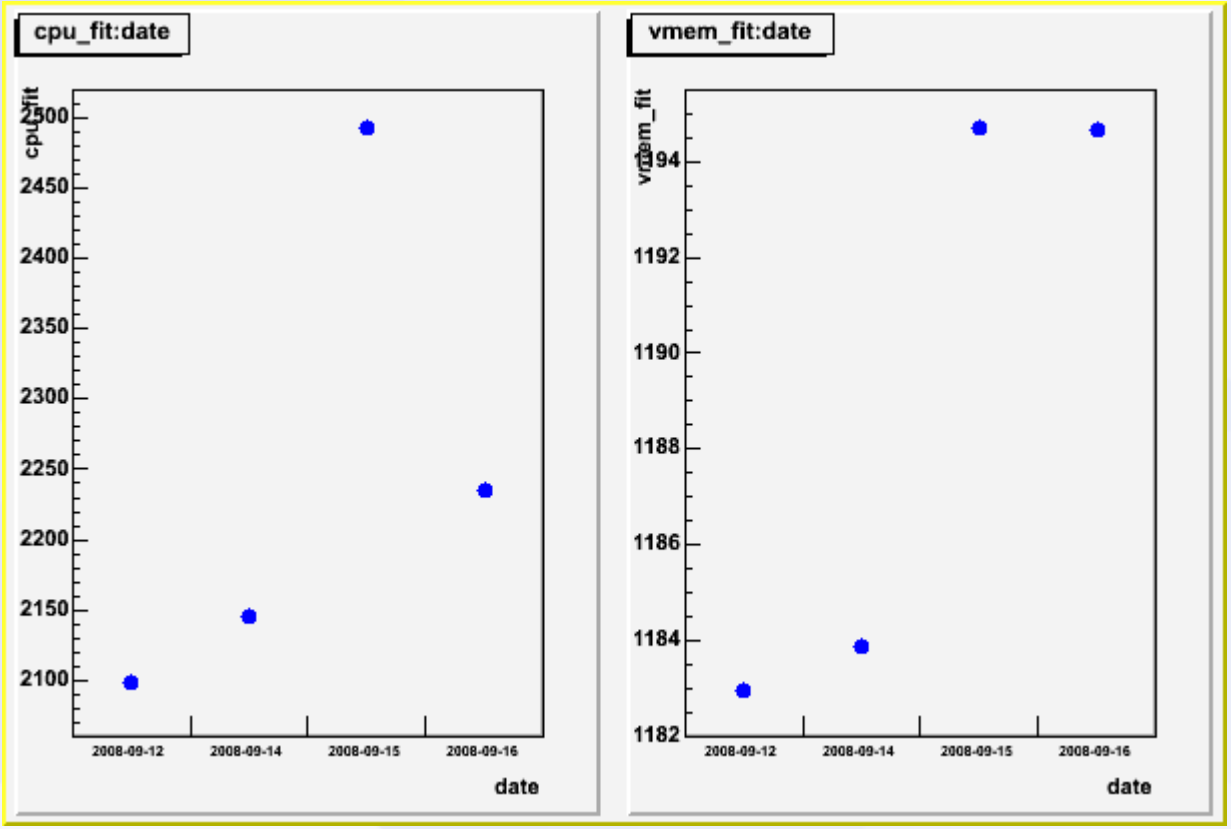


Find: Previous Next Highlight all Match case

http://atlas-project-trigger-release-validation.web.cern.ch/atlas-project-trigger-release-validation/www/per

Memory monitoring on the RTT

- Long-term memory usage plots can be done from RTT jobs
  - Useful to check memory leak in many algorithms



DCube

- An infra-structure is available to process and compare histograms generated by a test job
- Tests use an XML configuration file to determine which histograms to compare
- DCube has been integrated to run on RTT and shows the results on the web

DCube on RTT

TrigAnalysisTest[-00-00-98]
[done]

[Hide] [OfflineValidation] [Reco] [Trig-Common] **FAIL**

Name	JobGroup	Job status	Tests	Datasets	Hash
testTauSliceAthenaModernRDOtoESDAOD [id:470]	AthenaTrigAnalysisTestIOTests	success	error	[Show]	-412669765
testAthenaModernAOD_TrigDecTool_fixedAOD [id:476]	AthenaTrigAnalysisTestIOTests	success	no tests	[Show]	-755428664
testMuonSliceAthenaModernRDOtoESDAOD [id:469]	AthenaTrigAnalysisTestIOTests	success	error	[Show]	-865739601
testElectronSliceAthenaModernRDOtoESDAOD [id:467]	AthenaTrigAnalysisTestIOTests	error	no tests	[Show]	-168082837
testJetSliceAthenaModernRDOtoESDAOD [id:471]	AthenaTrigAnalysisTestIOTests	success	error	[Show]	1929607159
Enhanced bias RDOtoESDAOD 10**31 [id:465]	AthenaTrigAnalysisTestIOTests	error	no tests	[Show]	2127463886
testBphysicsSliceAthenaModernRDOtoESDAOD [id:473]	AthenaTrigAnalysisTestIOTests	success	error	[Show]	744186383
testMinBiasSliceAthenaModernRDOtoESDAOD [id:475]	AthenaTrigAnalysisTestIOTests	success	error	[Show]	1792798891
ttbar displaced beamspot [id:478]	AthenaTrigAnalysisTestIOTests	error	no tests	[Show]	1158855822
Top RDOtoESDAOD 10**31 [id:464]	AthenaTrigAnalysisTestIOTests	error	no tests	[Show]	-118628775
testPhotonSliceAthenaModernRDOtoESDAOD	AthenaTrigAnalysisTestIOTests	success	error	[Show]	-435983789

RTT page shows the validation test status

Description: --
Monitored file: /afs/cern.ch/atlas/project/RTT/Results/rel_2/bugfix/build/i686-slc4-gcc34-opt/offline/TrigAnalysisTest/AthenaTrigAnalysisTestIOTests/testMuonSliceAthenaModernRDOtoESDAOD/469/expert-monitoring.root
Reference file: /afs/cern.ch/atlas/project/RTT/Results/reference/offline/Trigger/TrigValidation/TrigAnalysisTest/expert-monitoring_muon.root
Date: 2008-09-23
File Descr.: --
View log file as plain text as HTML

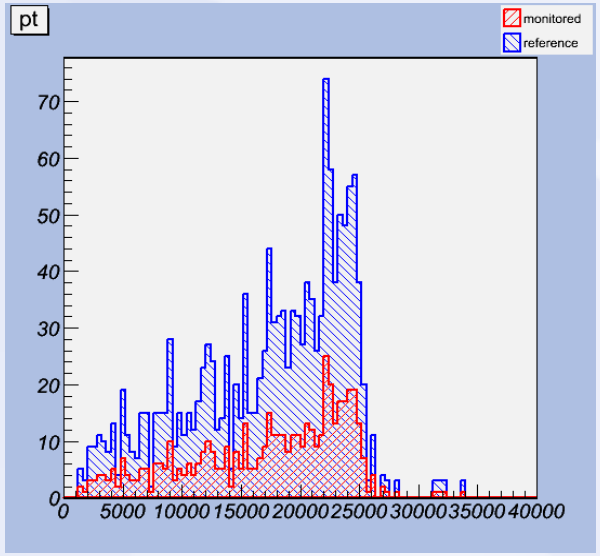
Legend

stats tests	OK (p-value >0.95)	WARN (0.7<p-value<0.95)	FAIL (p-value<0.7)	ABSENT (no test)
DCube status	✔ OK	? WARN	! FAIL	ABSENT (no status)

Normal view | Plots View | Summary

name	K-S test			χ² test			"bin-by-bin" test			DCube status
/	144	67	415	295	61	270	61	139	426	!
CTPSimulation	—	—	4	—	—	4	—	—	4	✔
InDetTrigDetailedTrackTruthMaker(...)FID	—	—	2	—	—	2	—	—	2	✔
InDetTrigPRD_MultiTruthMaker_Muon_EFID	—	—	3	—	—	3	—	—	3	✔
InDetTrigParticleCreation_Muon_EFID	—	2	19	16	2	3	—	1	20	✔
InDetTrigTrackParticleTruthMaker(...)FID	—	—	4	—	—	4	—	—	4	✔
MooHLTAlgo_CB_Muon	—	4	4	6	1	1	—	—	8	✔
MooHLTAlgo_MS_Muon	11	1	2	11	2	1	1	6	7	✔
MooHLTAlgo_SA_Muon	—	3	3	4	1	1	—	1	5	✔

DCube page: this shows the results



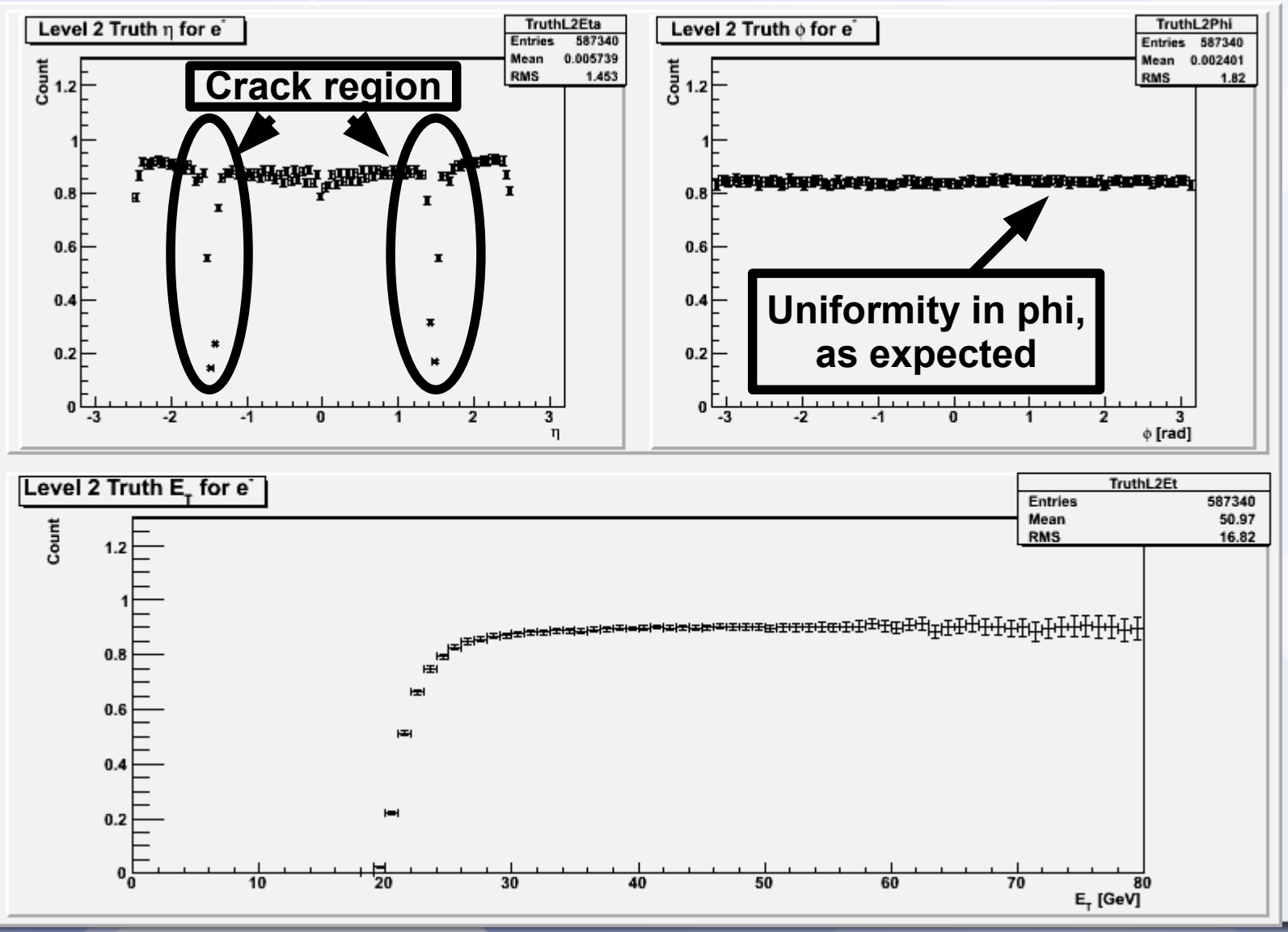
TrigEgammaValidation

- Systematic tool to validate Trigger electron/photon Studies
- Generates efficiency and jet rejection measurements, and plots relevant trigger cut variables for different input simulation datasets
- Uses statistical metrics to check discrepancies on histograms with respect to reference ones
- Automatically sends emails with results of the comparison for the release

TrigEgammaValidation method

- Generates histograms of variables used on Trigger decision
- Currently uses Kolmogorov-Smirnov statistical test to compare each histogram generated with the same one in a previous release
- Discrepancies are notified when they are found, allowing fast correction of problems
- Other metrics should be studied to improve identification of problems

TrigEgammaValidation plots



TrigEgammaValidation web visualization tool

TrigEgammaValidation Results for 14.2.20 (22/10/2008)

- List all releases
- Overall Results
- Z -> e + e RDO Dataset
  - Level 1 Results
  - Level 2 Results
  - Level 2 ID Scan Results
  - Level 2 ID Si Results
  - Event Filter Calo (after L2ID Scan Match) Results
  - Event Filter Calo (after L2ID Si Track Match) Results
  - Event Filter ID (after L2ID Scan Match) Results
  - Event Filter ID (after L2ID Si Match) Results
  - Perfmon Results
- Z -> e + e BS Dataset
  - Level 1 Results
  - Level 2 Results
  - Level 2 ID Scan Results
  - Level 2 ID Si Results
  - Event Filter Calo (after L2ID Scan Match) Results
  - Event Filter Calo (after L2ID Si Track Match) Results
  - Event Filter ID (after L2ID Scan Match) Results
  - Event Filter ID (after L2ID Si Match) Results
  - Perfmon Results
- Single Electron 25 GeV RDO Dataset
  - Level 1 Results

Zee [RDO]

Histogram name	KS test	JS test	KL metric
ZeeValidation/Dump/EFScanCaloEta	0.00%	99.80%	0.08160
ZeeValidation/Dump/EFScanCaloDphi	99.77%	99.94%	0.04028
ZeeValidation/Dump/EFScanCaloEt	86.11%	99.80%	0.02250
ZeeValidation/Dump/EFScanTrackD0	0.00%	99.11%	0.20401
ZeeValidation/Dump/EFScanTrackEta	46.87%	99.51%	0.02018
ZeeValidation/Dump/EFScanTrackPhi	78.92%	99.37%	0.02549
ZeeValidation/Dump/EFScanTrackPt	58.56%	99.27%	0.02955
ZeeValidation/Dump/EFScanTrackZ0	90.14%	99.52%	0.10830
ZeeValidation/Dump/EFSiCaloEta	0.00%	99.82%	0.07458
ZeeValidation/Dump/EFSiCaloDphi	99.35%	99.94%	0.03924
ZeeValidation/Dump/EFSiCaloEt	76.02%	99.79%	0.02289
ZeeValidation/Dump/EFSiTrackD0	0.00%	99.11%	0.22869
ZeeValidation/Dump/EFSiTrackEta	20.15%	99.49%	0.02060
ZeeValidation/Dump/EFSiTrackPhi	75.85%	99.37%	0.02531
ZeeValidation/Dump/EFSiTrackPt	49.01%	99.28%	0.02890
ZeeValidation/Dump/EFSiTrackZ0	94.04%	99.47%	0.12668
ZeeValidation/Dump/L1EmClus	91.37%	99.78%	0.01478
ZeeValidation/Dump/L1EmIsol	100.00%	99.97%	0.02061
ZeeValidation/Dump/L1HadCore	99.97%	99.97%	0.02283
ZeeValidation/Dump/L1HadIsol	81.17%	99.98%	0.01862
ZeeValidation/Dump/L2CaloEt	56.45%	99.50%	0.15332
ZeeValidation/Dump/L2CaloHadEt	87.81%	99.88%	0.10343
[illegible]	[illegible]	[illegible]	[illegible]

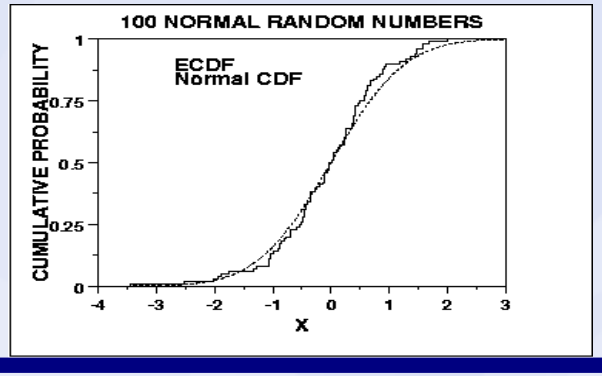
Color code	
	100% similarity
	90%-100% similarity
	70%-90% similarity
	less than 70% similarity

Symbol	Measure	Description
KS	Kolmogorov-Smirnov test	Similarity is calculated based on the maximum distance between the cumulative distribution functions of the random variables.
KL	Kullback-Leibler Symmetrized metric	A distance between the random variable is calculated based on information theory's Entropy concept (this is a value greater than zero and unbounded).
	[illegible]	A "normalized version" of the Kullback-Leibler Symmetrized metric, giving a value

Statistical metrics description

Kolmogorov-Smirnov

- The Kolmogorov-Smirnov Test can be used to compare two pdfs by measuring the maximum distance between their cumulative distribution function



Quadratic Negentropy

- Measures the distance between the pdfs bin-per-bin

$$D_{QN}(P\|Q) = \int_{-\infty}^{\infty} (p(x) - q(x))^2 \, dx$$

Kullback-Leibler

- Kullback-Leibler divergence uses the entropy concept from Information Theory to measure the *distance* between two pdfs
  - Such a method is naturally assymetric, but can be made symmetric

$$D_{\mathrm{KL}}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx$$

$$D_{\mathrm{KL}}(P\|Q) + D_{\mathrm{KL}}(Q\|P)$$

Jensen-Shannon

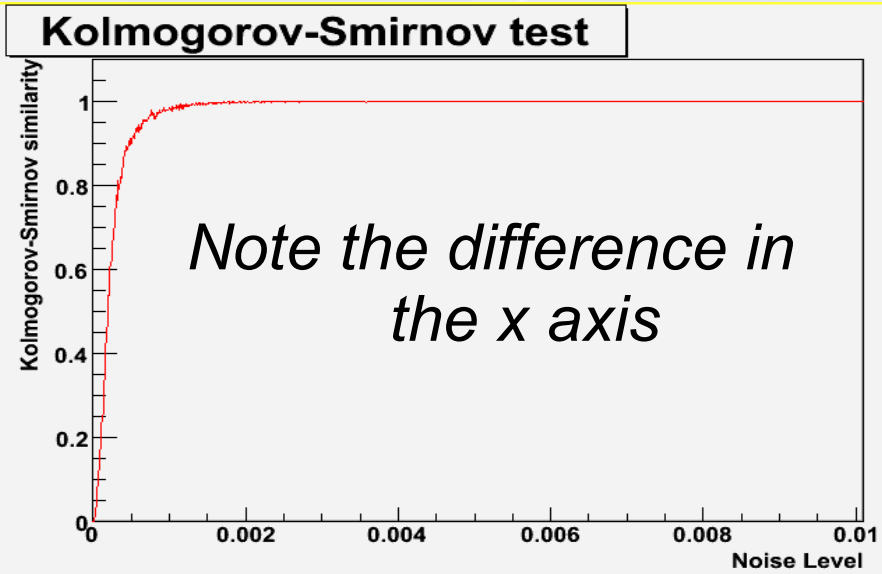
- Kullback-Leibler is unbounded. Using the lemma that the distribution closest to the other two is the mean distribution, Jensen-Shannon proposed a new metric.
  - Jensen-Shannon metric is always positive and less than 1.

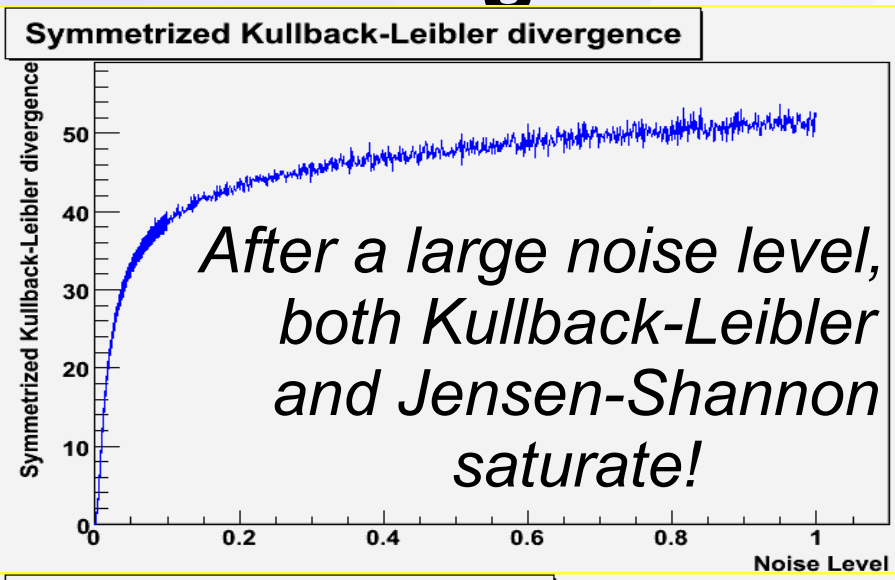
$$M = \frac{1}{2}(P + Q) \qquad JSD(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M)$$

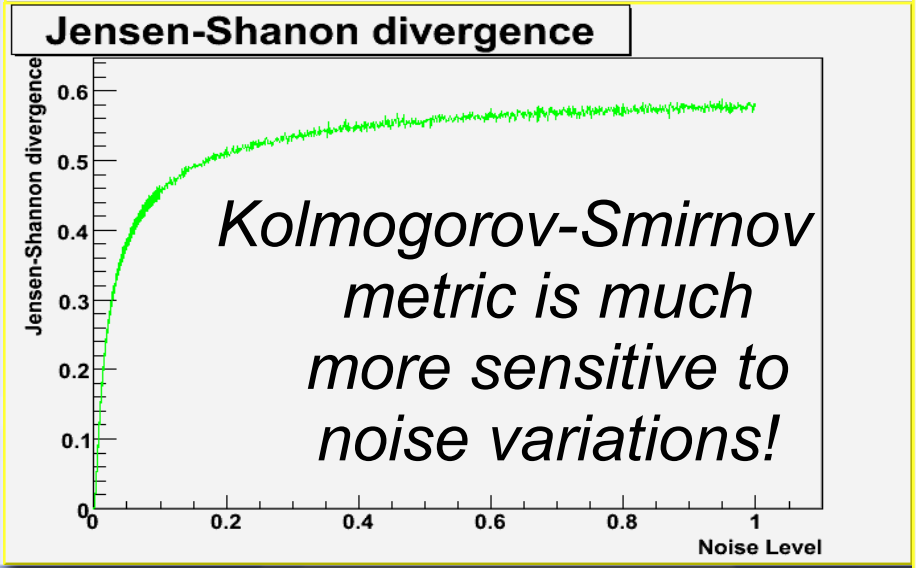
Metric comparison: Methodology

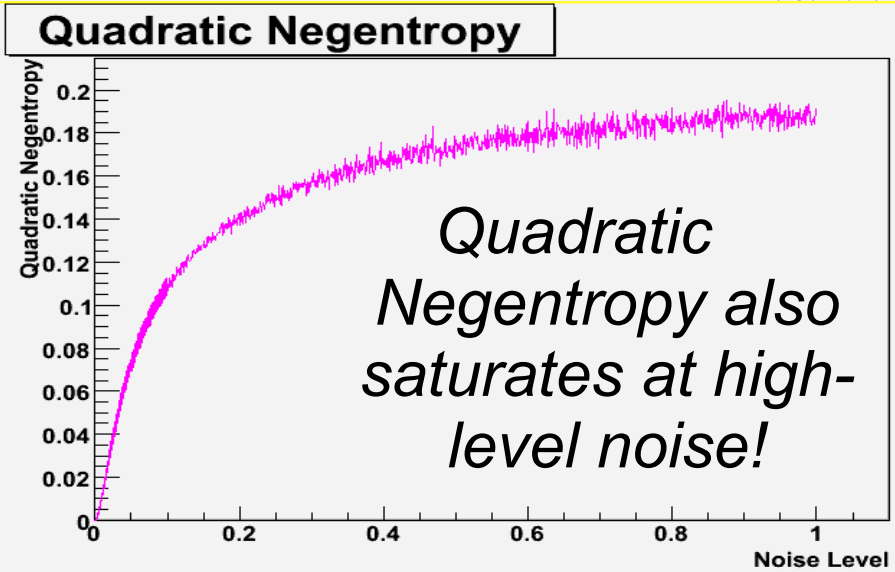
- To compare each metric's result and observe their sensitivity, two methods were tried:
  - Adding Gaussian noise to test histograms and comparing with the original histograms
  - Multiplying Gaussian noise to test histograms and comparing with the original histograms
- The standard deviation of the noise is stepwise increased to check the metric's results
- A mean of all TrigEgammaValidation histograms comparisons is taken to evaluate each metric's performance

Metric comparison:
Additive noise to the histograms

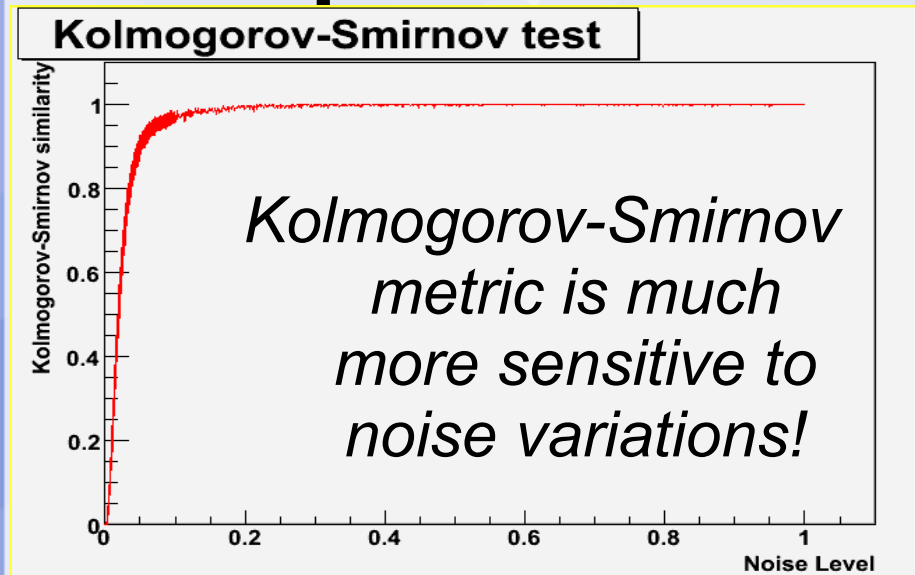


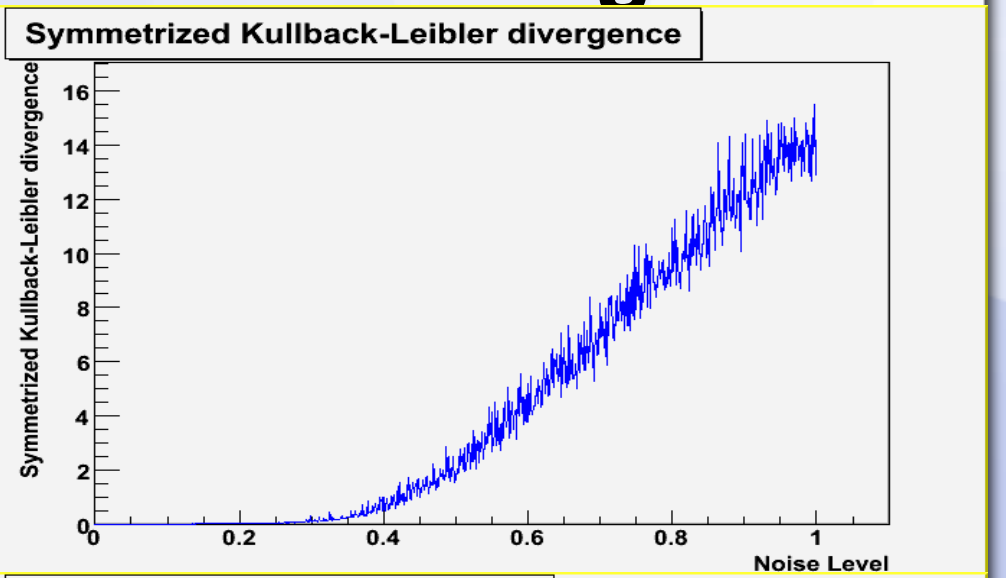


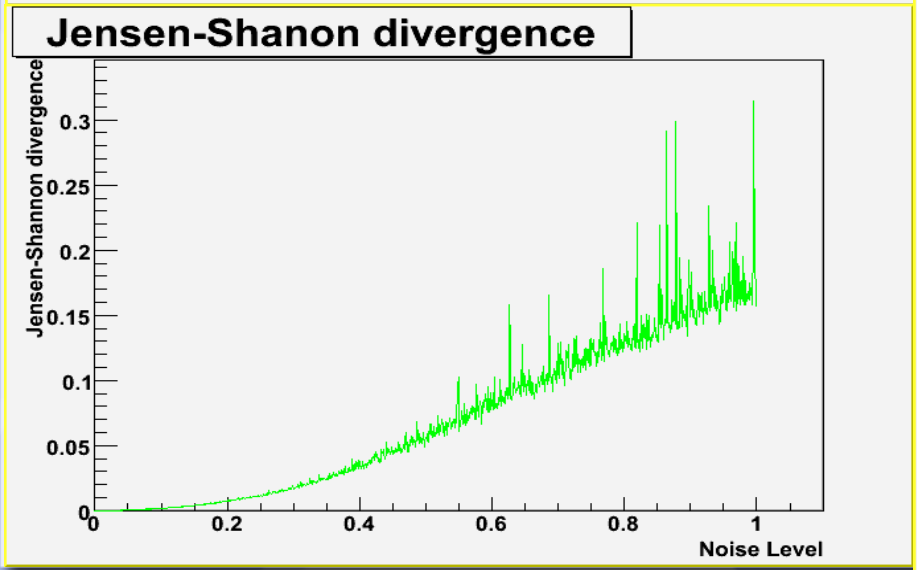


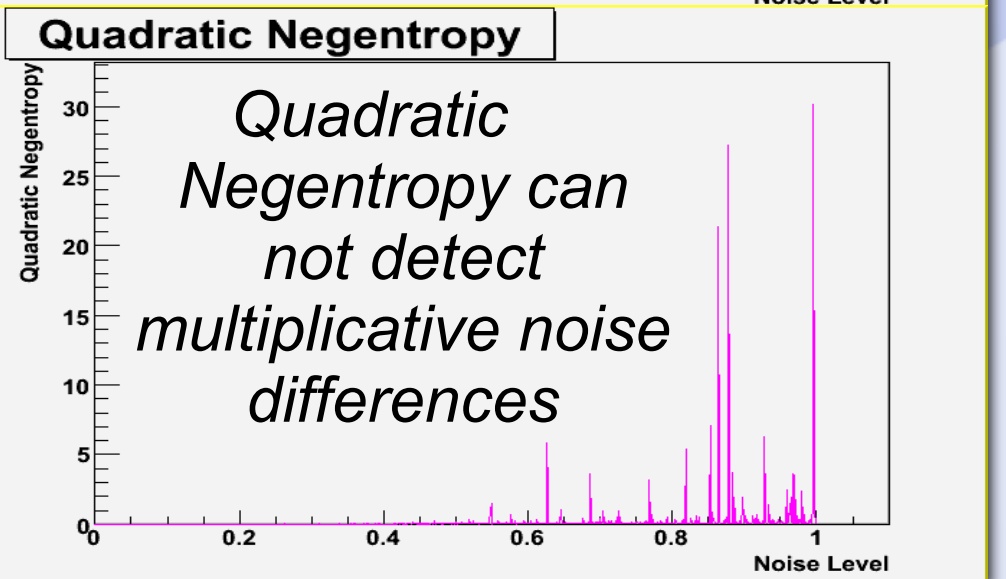


Metric comparison:
Multiplicative noise to the histograms









Metric comparison:
Results

- The Kolmogorov-Smirnov test is very sensitive to small variations of a random variable
- The Kullback-Leibler and Jensen-Shanon can't distinguish differences between histograms after a large noise level and are not as much sensitive to multiplicative noise as Kolmogorov-Smirnov
- Quadratic Negentropy can not detect multiplicative noise, and it also can not detect differences at high additive noise!
- To be investigated: The combination of such methods

Conclusion

- Validation on Trigger is being applied systematically, controlling the problems as closely as possible
- Software quality (memory leaks, time measurements, ...) on the code are tested nightly by PerfMon on RTT
- ATN tests are extremely useful for checking small code differences;
- Comparison of histograms allows systematic validation of the software through physics measurements, as in TrigEgammaValidation
- Thanks to the ATLAS Software Infrastructure Group for providing the testing frameworks in which our trigger tests run; particularly to Alex Undrus