Calorimeter Cell Monitoring Tool

Viatcheslav Shary*
Laboratoire de l'Accélérateur Linéaire

August 4, 2003

Contents

1 Introduction

This work is devoted to assess the calorimeter data quality at the level of calorimeter cells signals from the calorimeter data chunk. This work has two main goals. The first is to detect the problems which were not found in the previous stages of the data quality check (e.g. at the calorimeter shift or global monitoring shift level) and determine the quality of the data at the run level from the calorimeter point of view. The second is to provide the data quality information for further use: for example, list of BLSs with unusual large or small signal that can result in "warm" or "cold" zones appearance, local inefficiency or shifts in missing E_T. This information is of interest for the reconstruction and simulation programs and, of course, for the data analysis.

The thumbnails, raw files or event from online distributor can be used. In order to limit the size of output root trees and be independent from effects like triggers modifications, luminosity increasing

*e-mail: shary@lal.in2p3.fr

and so on, only "zero bias" events are used. The two types of data quality information are collected: the summary information for the each event (like number of cells, total cell energy etc) and summary information for each run but separately for each calorimeter cell (like number of entries in cell, total signal in cell etc) from which more condensed information, at the BLS or ADC level, can be derived. Using cuts on these parameters, the data quality (good or bad) of each event, each cell and BLS is determined.

As a results of this work, the calorimeter cell monitoring tool for offline and online mode was created.

The structure of this document is the following. In section 2 the analysing scheme, the list of all accumulated parameters and the root trees structure are described. In section 3 the quality criteria are listed. The typical distributions on parameters in use are also shown. In section 4 the typical run quality problems are described. The first results and problematic runs are quoted. In section 5 the influence on the missing E_T distribution of the bad cells/BLSs and a possible correction are discussed. In section 6 issues and perspectives are discussed. Programs manuals are added to the appendixes A,B.

2 The analysing scheme.

The created scheme to analyse calorimeter data quality was realized as *dq_calo* package and can be run online and offline (see fig. 1). The outputs of the package are three files: the text report file with data quality information and bad cells/BLSs lists; the input file for the data quality data base; the file with root trees for further analysis. The DQCalo class reads a calorimeter data chunk and creates two trees. The first tree is named "dq_events". Each entry corresponds to one "zero bias" event. Leaves are:

- runNum — the run number,
- eventNum — the event number,
- nCells — the number of cells hits in the event (N)
- e — the total energy in the event, $\sum_{i=1}^{N} E_i$, where E_i — energy in one cell.
- e2 — the $\sum_{i=1}^{N} E_i^2$ in the event,
- set — the scalar E_T in the event,
- met — the missing E_T in the event.

In the second tree "dq_cells", each entry corresponds to one cell in the calorimeter, with summary information for a given run. Leaves are:

- address - the cell calorimeter address,
- entriesNum - number of times the cell was hit (N),
- e - total energy in the cell $\sum_{i=1}^{N} E_i$, where E_i – energy of one entry.
- e2 - $\sum_{i=1}^{N} E_i^2$ in the cell.

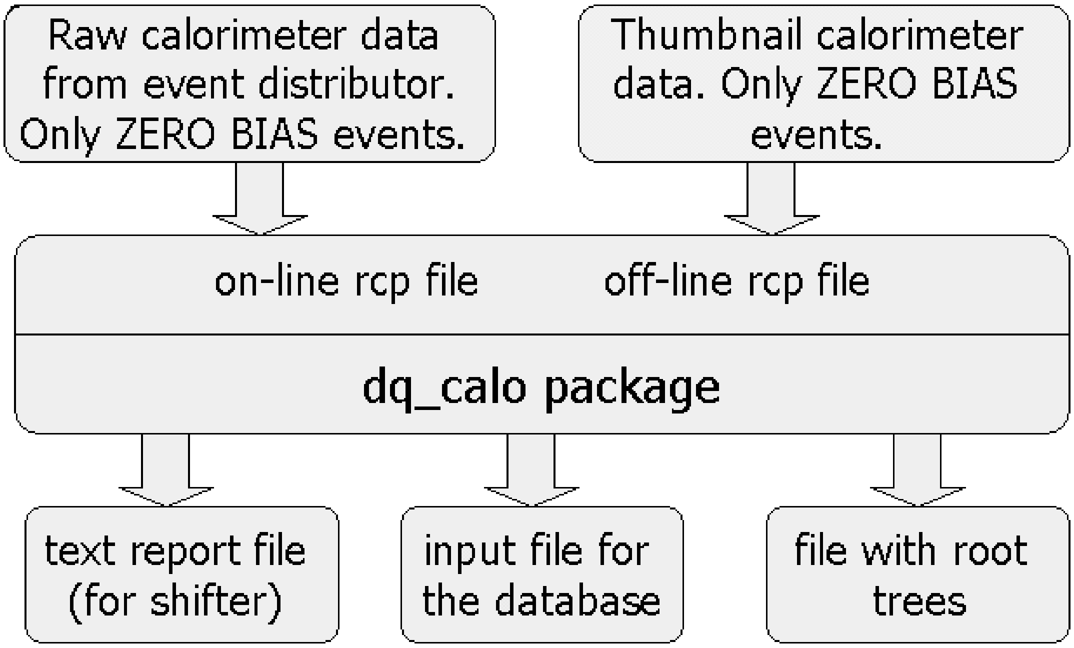


Figure 1: The analysing scheme.

The event with address equal to 0 is special and placed before the cells information for the given run. The leaves are:

- address == 0,
- entriesNum == run number,
- e == number of events selected by "zero bias" trigger in this run,
- e2 == 0

The *root file has a size near 900 kbyte and has a weak dependence from the number of events.

3 Trees analysis and selection criteria

In order to determine the fraction of "bad" events in the run, the following criteria are applied to the "event" tree to define good events (the corresponding distributions are shown on fig. 3 – fig. 8 for a set of good runs in the 170000 – 178000 range):

1. number of cells satisfying the condition $150 \leq N_{cells} \leq 10000$
2. the total energy satisfying the condition $-1000\ GeV \leq E \leq 1500\ GeV$
3. the average energy per cell satisfying the condition $-0.4\ GeV \leq E_{mean} \leq 0.5\ GeV$
4. the cell energy RMS satisfying the condition $0.1\ GeV \leq \sigma_E \leq 2.0\ GeV$
5. the scalar E_T satisfying the condition $40\ GeV \leq SET \leq 300\ GeV$
6. the missing E_T satisfying the condition $0\ GeV \leq MET \leq 40\ GeV$

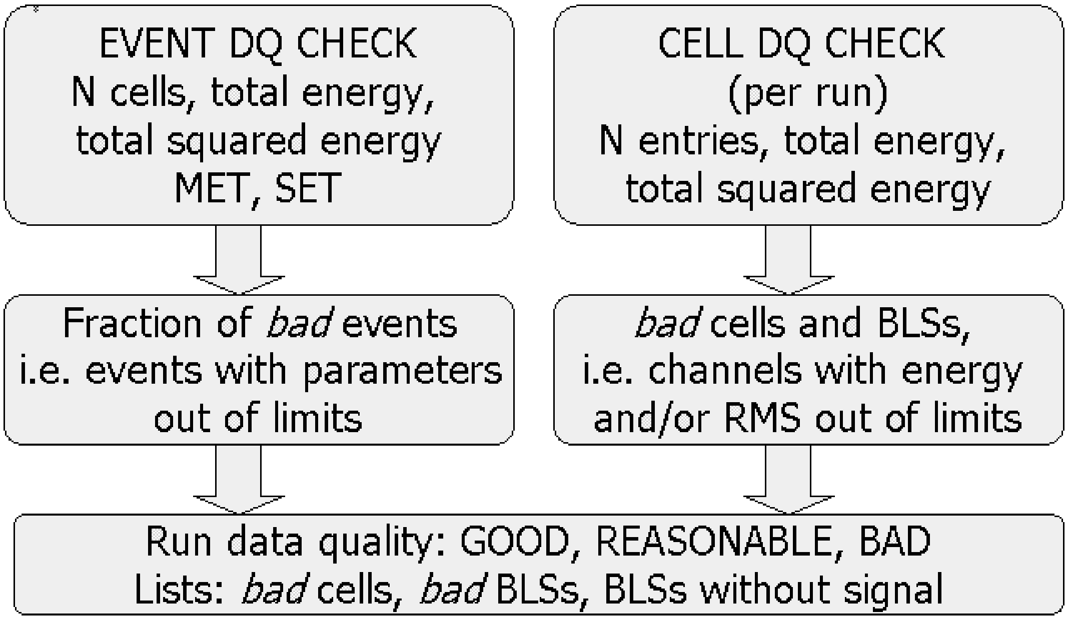


Figure 2: The analysing scheme.

To check the electronics status in the run, the following criteria are applied to the the “cell” tree to define good cells/BLS (the corresponding distributions are shown on fig. 9 – fig. 17 for a set of runs from 170000 – 174484 range):

1. the average energy per entry in cell satisfying the condition $-4\ GeV \leq E \leq 5\ GeV$,
2. the energy RMS in cell satisfying the condition $0.03\ GeV \leq \sigma_E \leq 15.0\ GeV$,
3. the BLS average energy satisfying the condition $-1.5\ GeV \leq E \leq 2\ GeV$,
4. the BLS energy RMS satisfying the condition $0.05\ GeV \leq \sigma_E \leq 2.50\ GeV$,

A cell is declared “bad” if it is out of limits 1, 2 and has more than 9 entries for the run under consideration. A BLS board is declared “bad” if it has more than 5 channels with quality “bad”, or it is out of limits 3, 4. The BLS boards with no signal (“killed” or not powered) are also detected.

The run quality (good, reasonable, bad or unknown) is determined according to the following rules:

- if the fraction of ”bad” events $> 10\%$ the quality is BAD ;
- else if the number of cells out of energy limits > 24 the quality is BAD ;
- else if the number of ”zero bias” events < 100 the quality is UNKNOWN ;
- else if the number of BLSs without signal > 1 the quality is BAD ;
- else if the number of BLSs without signal $== 1$ the quality is REASONABLE ;
- else if the number of BLSs ”out of limits” > 0 the quality is REASONABLE ;
- else if number of ”bad” events $> 2\%$ the quality is REASONABLE ;

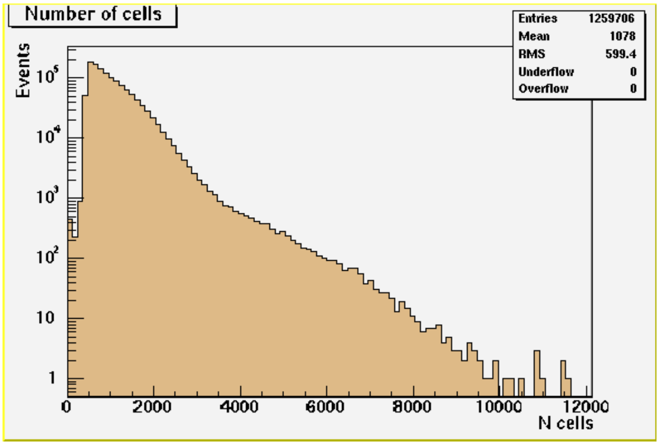


Figure 3: Distribution of the number of cells in "zero bias" events.

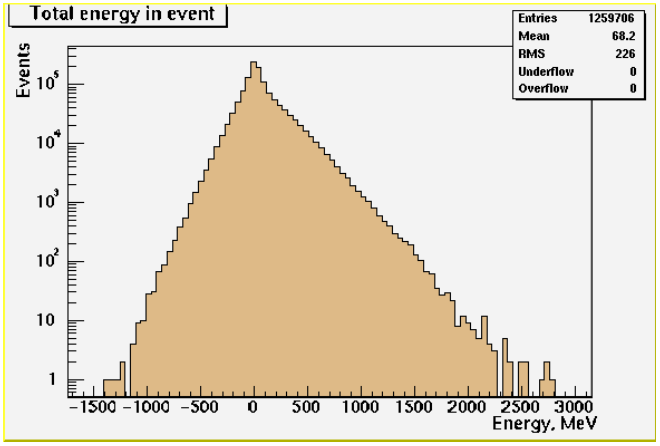


Figure 4: Distribution of the total energy in "zero bias" events.

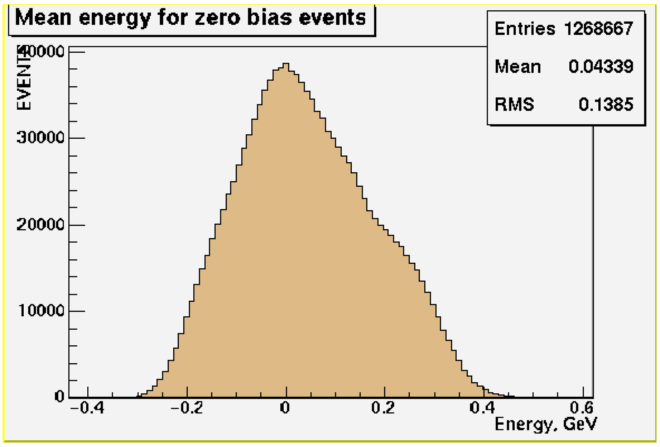


Figure 5: Distribution of the average energy per cell in "zero bias" events.

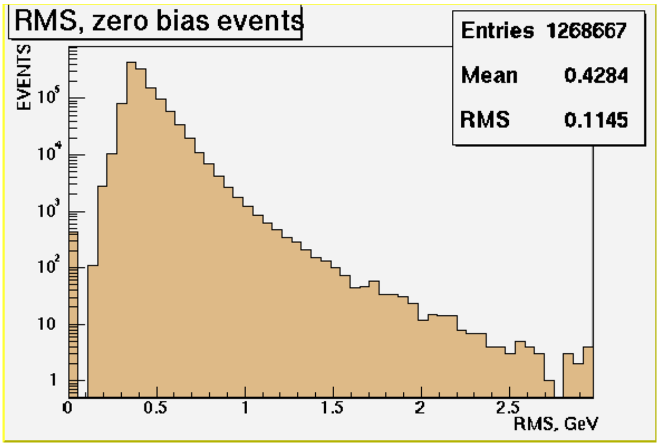


Figure 6: Distribution of the cell energy sigma in "zero bias" events.

- else the quality is GOOD.

An example of distribution for the 2d histograms in the $\eta - \phi$ plane for the mean energy per cell for each layer in calorimeter are presented on fig. 18 and 19. The same kind of histograms for the electronics channels is presented at fig. 20. To visualise the histograms for the recent runs the "Monitor" application was designed. For more details see appendix B. This application also allows to generate an html page with runs data quality information [1].

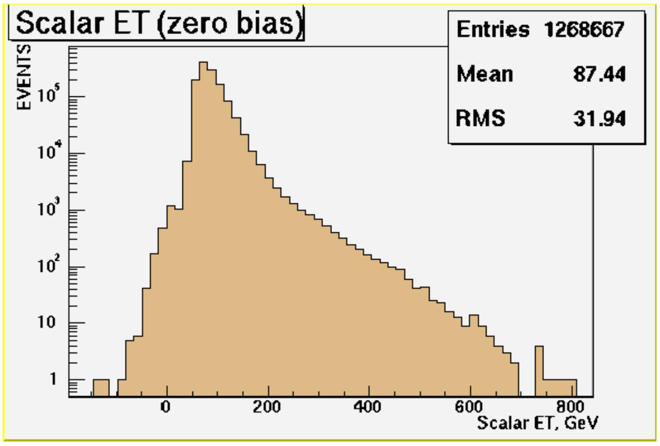


Figure 7: Distribution of the scalar E_T in "zero bias" events.

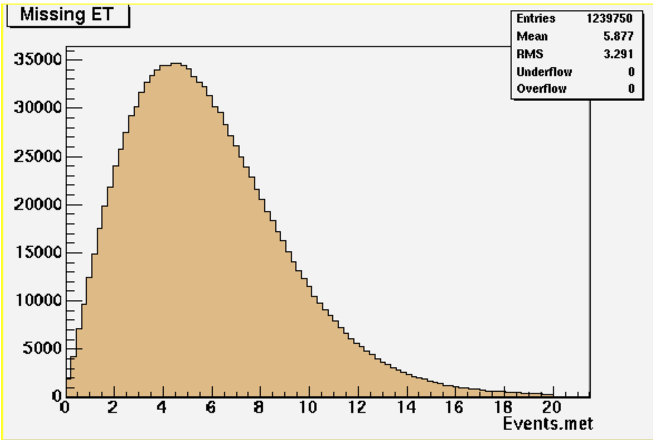


Figure 8: Distribution of the missing E_T in "zero bias" events.

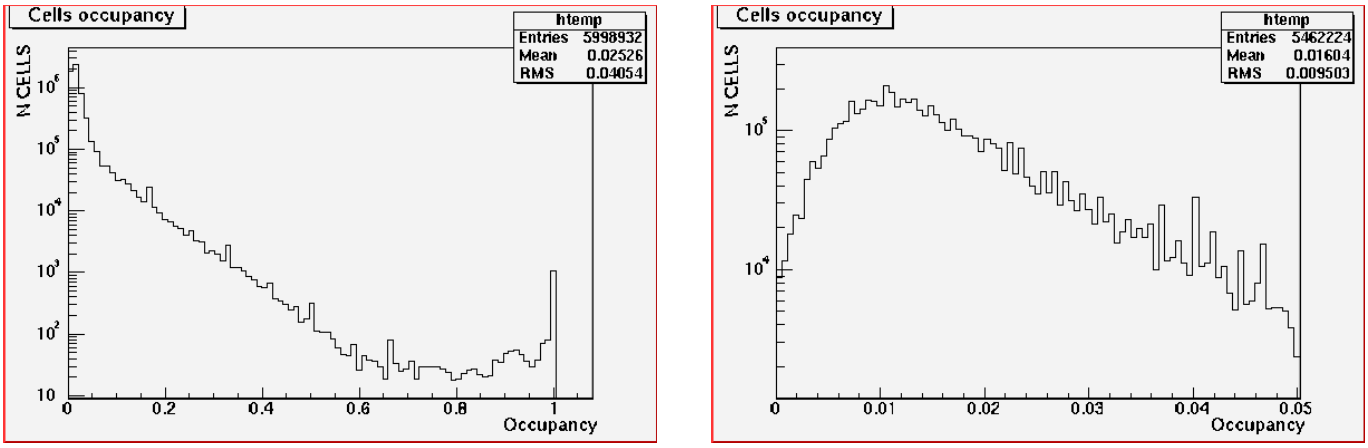


Figure 9: Cells occupancy distribution.

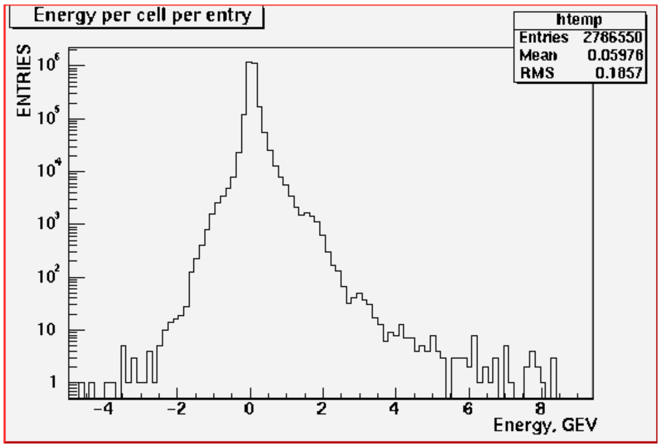


Figure 10: Distribution of the mean energy per entry per cell.

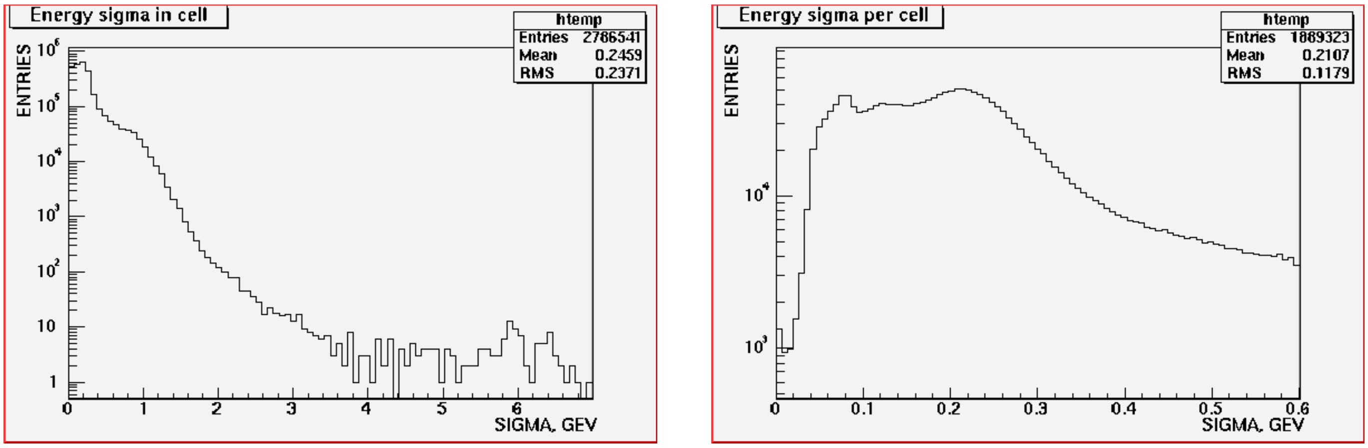


Figure 11: Energy sigma distribution.

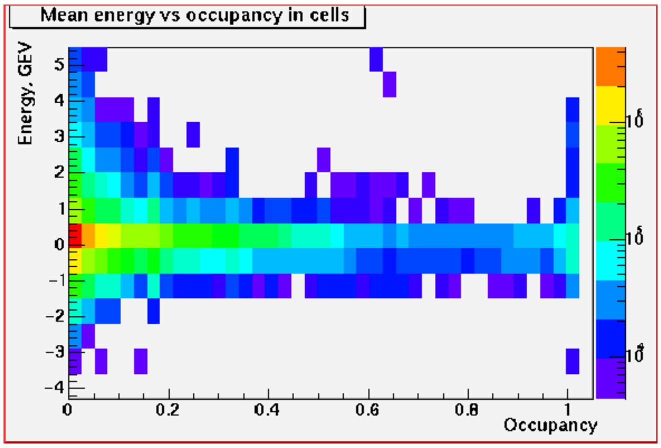


Figure 12: Mean energy per cell vs cell occupancy.

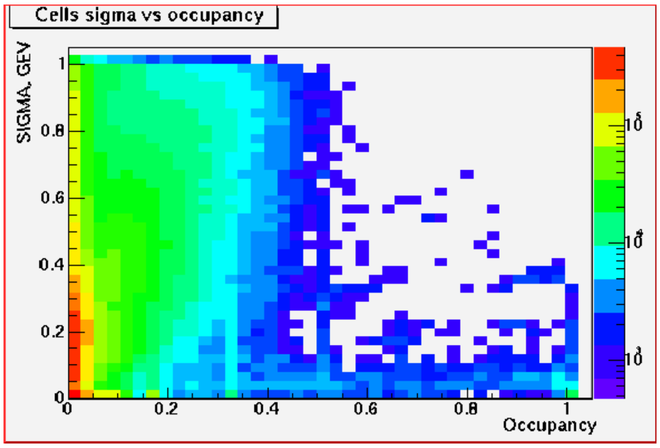


Figure 13: Energy sigma per cell vs cell occupancy.

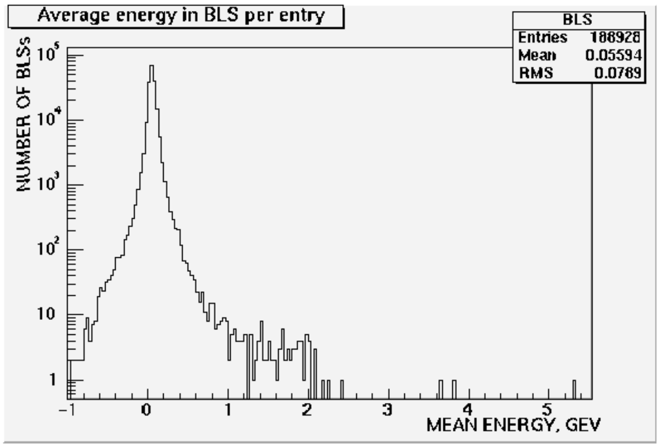


Figure 14: Distribution of the mean BLS energy per entry.

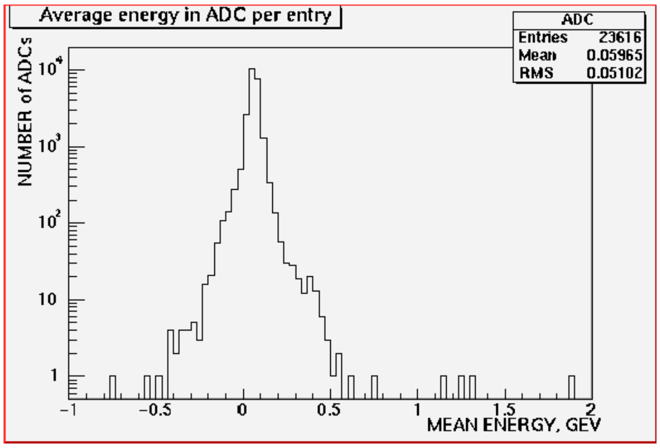


Figure 15: Distribution of the mean ADC energy per entry.

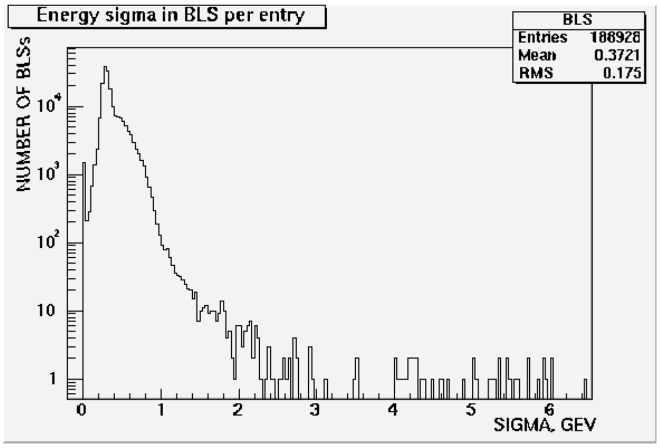


Figure 16: Distribution of the BLS energy sigma.

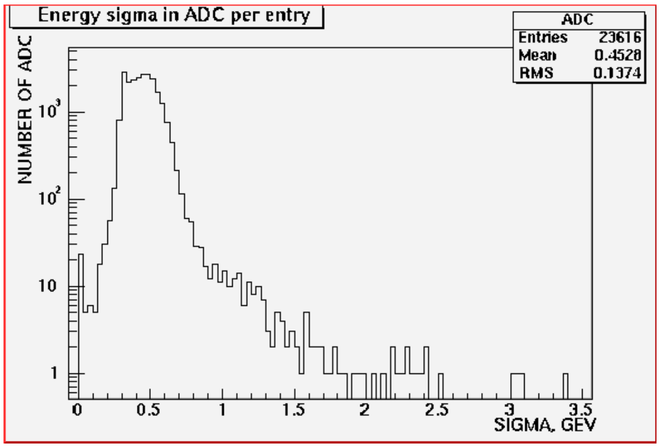


Figure 17: Distribution of the ADC energy sigma.

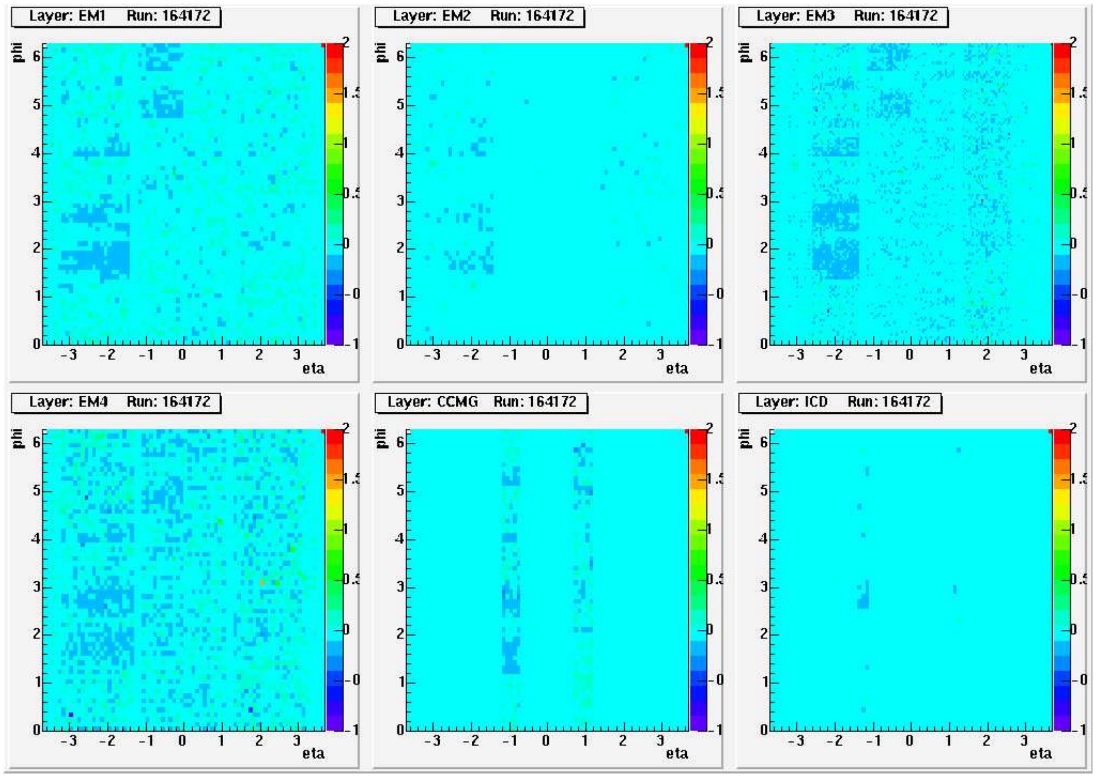


Figure 18: An example of the 2d ($\phi - \eta$) calorimeter layer histograms for run 164172. The Z axis corresponds to the mean energy per entry in the cell (GeV). Layer from EM1 to ICD.

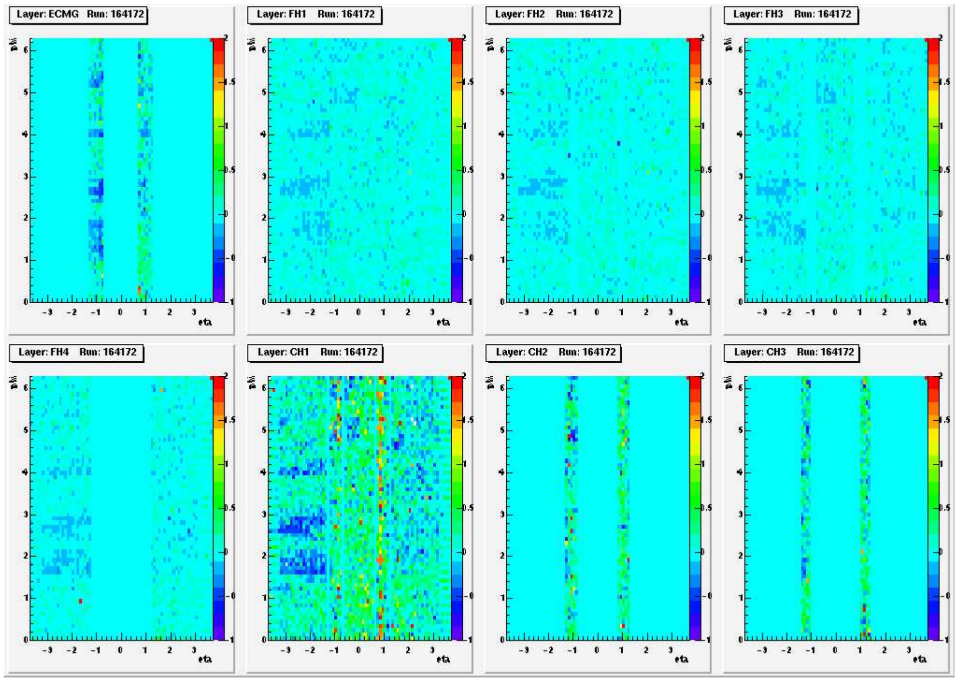


Figure 19: An example of the 2d ($\phi - \eta$) calorimeter layer histograms for run 164172. The Z axis corresponds to the mean energy per entry in the cell (GeV). Layers from ECMG to CH3.

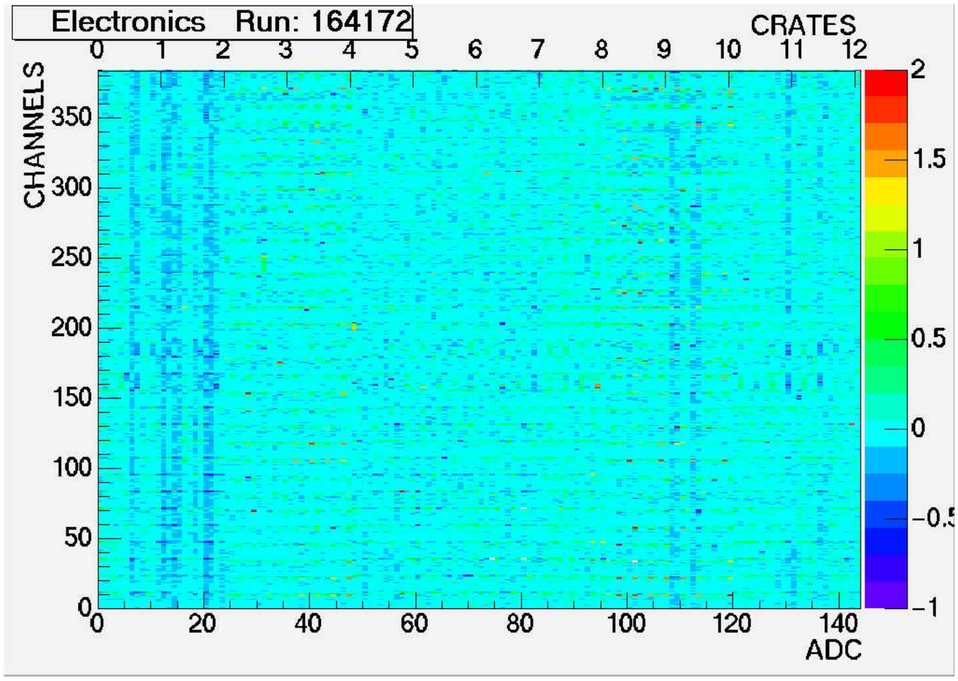


Figure 20: An example of the electronics 2d histogram for run 164172. X axis — serial ADC number or crate number. Y axis — serial channel number in ADC. The Z axis corresponds to the mean energy per entry in the cell (GeV).

4 Results

One can find the list of runs with corresponding qualities and 2d histograms for each layer at [1]. The following results were obtained.

Runs	N of runs	N *good*	N *reasonable*	N *bad*
170000–178852 (January – June 2003)	677	392 (59%)	203 (30%)	21 (3%)
160000–169999 (August – December 2002)	601	123 (20%)	413 (69%)	23 (4%)

Several types of problems were detected. The runs are declared *bad* because

- 52% of runs (after January 2003) have no signal from several BLSs due to BLS power supply failure or because BLS was killed by an expert.
- 38% of runs have too many (more than 24) channels with big signal.
- 10% of runs have several channels with huge signal. The total signal is so high that a big fraction of events (more than 10%) should be declared "bad".

The runs are declared *reasonable* because

- 84% of runs have 1 or more BLSs with too big or too small mean energy and/or big RMS. In 50% of cases there are 1 or more channels in these BLSs with the out of limit mean energy. In other cases such channels probably also exists, but the signal and/or occupancy are not so high to pass the limits for an individual channel, but sufficient to pass the limits for the BLSs. To reduce the number of such runs, the online monitoring can be very useful. To correct physic quantities offline the input/output interface should be done for the data quality data base. Some of the good runs also have hot cells, but the average energy in BLS is in the limits.
- 16% of runs have no signal from 1 BLS. The usual reason — shifter kill the the corresponding hot cells.

5 Data quality influence on missing E_T for zero bias events.

The typical parameters for the missing E_T distribution in zero bias events for different algorithms are shown in table 1. One can say that these values show the effect of the calorimeter noise on the missing E_T distribution in physical events. For further analysis the calculation **D** (all cells with signal $E > 0$ but not coarse hadronic) is used (this is the algorithm recommended by the Jet/MET group). If one compares the missing E_T distribution for *good* and *reasonable* runs the difference in the distributions can be clearly seen (fig. 21). The width of the distribution for *good* runs near 3 GeV, but 3.5 GeV for the *reasonable* runs. Because the difference in widths is small, it should not affect the missingE_T parameter for the physical events. More important is the difference in the tails (see fig. 22). In spite of the fact that the probability for an event from the *reasonable* runs to have a big missing E_T is less than 10^{-3} it can be important in the rare process measurements and in the new phenomena searches.

In order to improve the calorimeter data quality and the accuracy in missing E_T calculation, a correction procedure is proposed. In this procedure all cells and/or BLSs marked as bad by dq_calo program are removed from the calculation. The results of this correction are shown on fig. 21,22. One can see that the difference between *good* and *reasonable* runs almost disappeared. The average missing E_T is not stable in time (see fig. 23). For the runs with *good* quality the shift in missing E_T less than 3 GeV. For the *reasonable* runs less then 7 GeV. The *bad* runs can have arbitrary

Table 1: The missing E_T distribution parameters (GeV) for the zero bias events and different types of calculations. **A**: only cells with $E > 0$ taking into account. **B**: only cells with $E > 100MeV$. **C**: all cells but without Coarse Hadronic (CH) cells. **D**: only cells with $E > 0$ and NOT CH. **E**: only cells with $E > 100MeV$ and NOT CH.

parameters	all cells	**A**	**B**, **C**	**D**,**E**
Peak	4.5 GeV	3.5 GeV	2.5 GeV	2.0 GeV
FWHM	7.5 GeV	5.5 GeV	4.5 GeV	3.0 GeV
FWHM/2.36	3.2 GeV	2.3 GeV	1.9 GeV	1.3 GeV

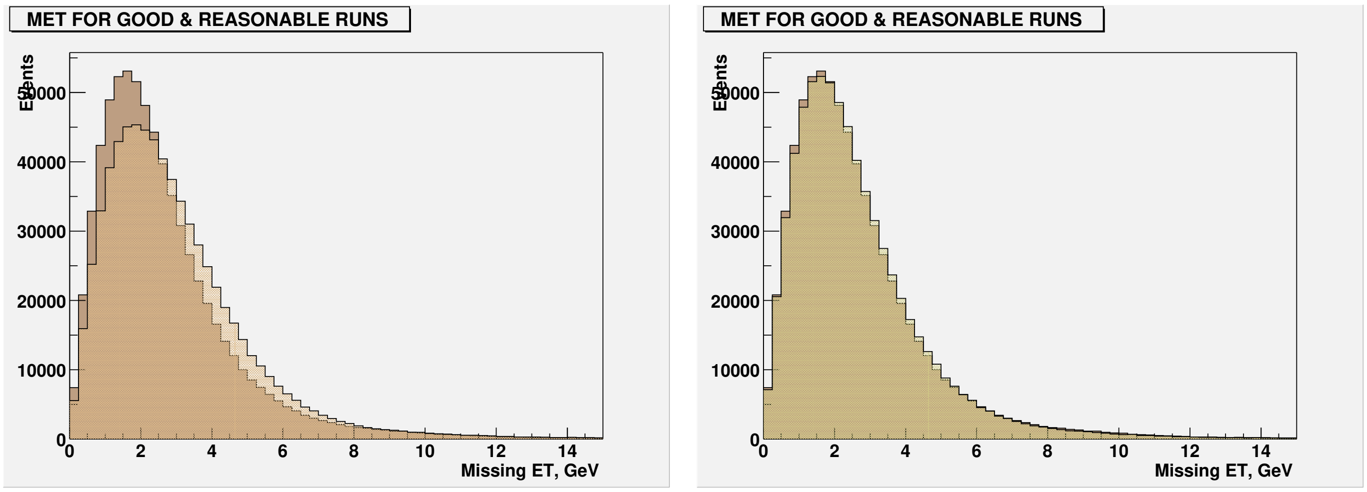


Figure 21: Left: missing E_T distribution for the *reasonable* (white histogram) and *good* runs. Right: the same distributions after correction, i.e. all bad cells and BLSs were removed from missing E_T calculation.

large missing E_T. The correction procedure reduce the shift in average missing E_T up to 3 GeV for all runs. From this one can conclude that the reason of bad behaviour of missing E_T are bad cells and/or BLSs. Probably the same reasons explain the missing E_T behaviour for non zero bias events (see [2]).

After correction the average missing E_T is not completely stable (fig. 23). The possible reason of it that not all hot cells are detected, e.g. if a channel has parameters near but inside the limits. Another possible reason that this approach do not allow detect all problems in calorimeter. E.g. the "shared" energy problem, coherent noise can also shift the average missing E_T.

Another problem in the distribution is the long tail in missing E_T for the zero bias events. On fig. 25 one can clearly see that events with missing $E_T > 20$ GeV have in average more number of cells, but smaller RMS then cells with missing $E_T < 20$ GeV. This is a strong indication that these events exhibit coherent noise. On fig. 27 the RMS versus missing E_T dependence is shown. For missing E_T less than 15 GeV the average RMS is growing with missing E_T, probably due to the fact that the bigger missing E_T the bigger number of cell is necessary. For missing E_T more then 15 GeV the average RMS decreases up to 0.3 GeV, these can be an indication that these events mainly have a coherent noise origin. On fig. 27 one can see that the RMS dependence from the number of cells is completely different for events with missing E_T more and less 20 GeV.The fraction of the events with missing $E_T > 20$ GeV is near 1 – 3 % for *good* runs (fig. 28).

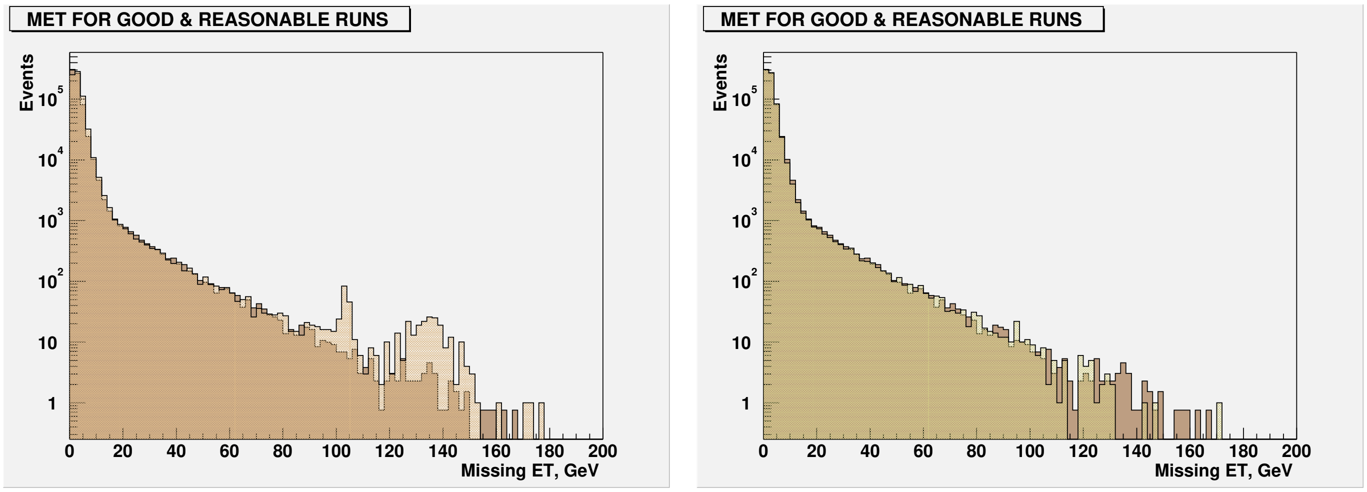


Figure 22: Left: missing E_T distribution for the *reasonable* (white histogram) and *good* runs. Right: the same distributions after correction, i.e. all bad cells and BLSs were removed from missing E_T calculation.

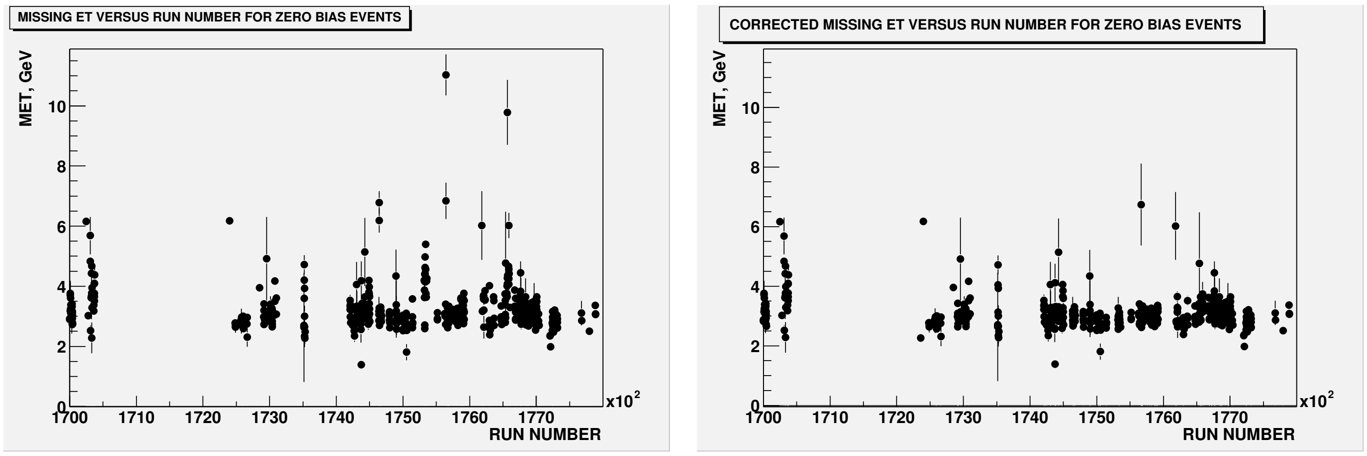


Figure 23: Left: missing E_T mean value versus run number for all runs. 5 runs with quality *bad* are not shown, because they have missing E_T more than 20 GeV. Right: the same distributions for the corrected missing E_T (see text). All runs are shown!

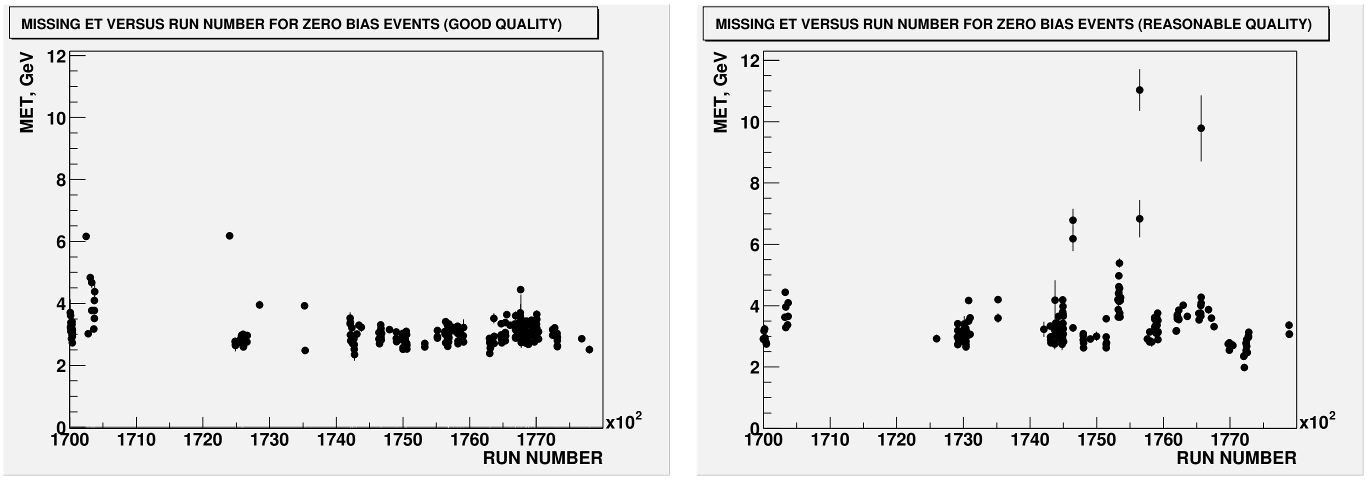


Figure 24: Left: missing E_T mean value vs run number for runs with *good* quality. Right: the same distributions for runs with *reasonable* quality.

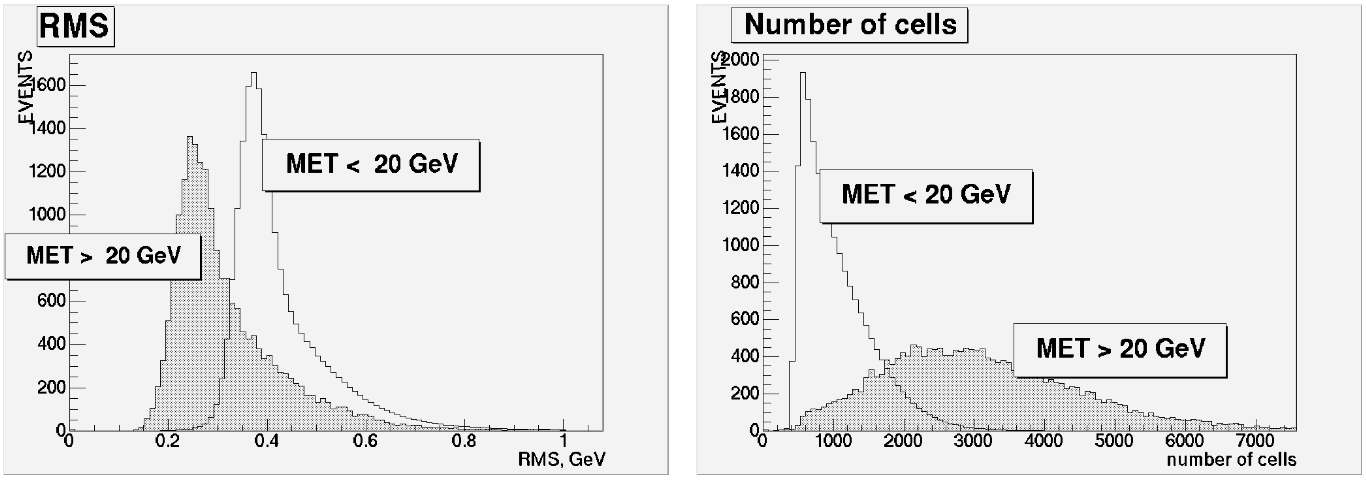


Figure 25: RMS and cells number distributions for events with missing E_T more/less 20 GeV. The histograms scale were adjusted.

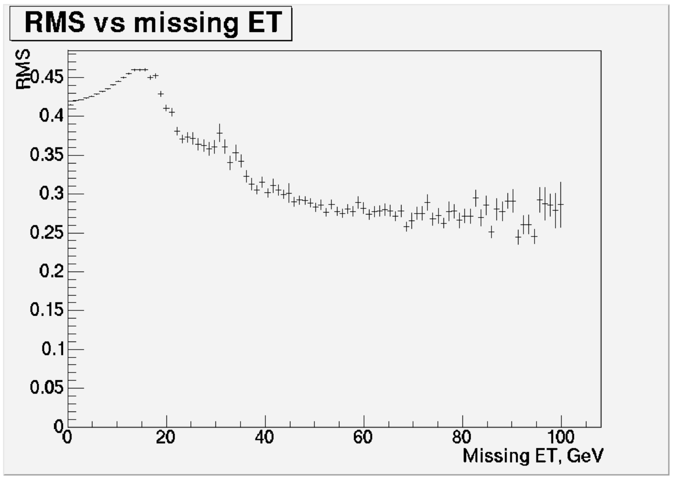


Figure 26: RMS versus missing E_T.

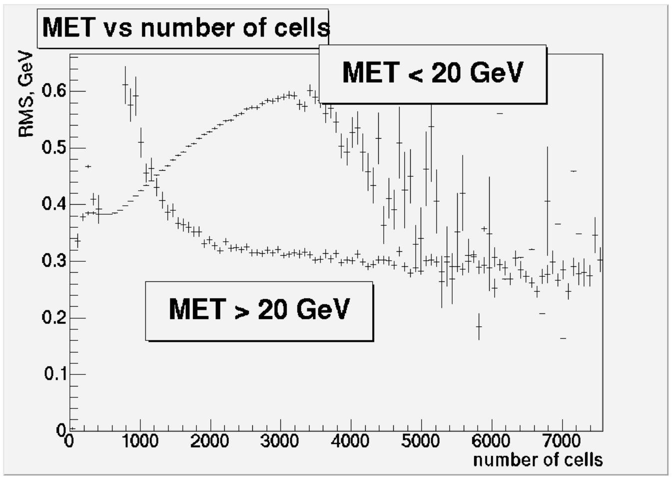


Figure 27: RMS versus number of cells for events with missing E_T more/less 20 GeV.

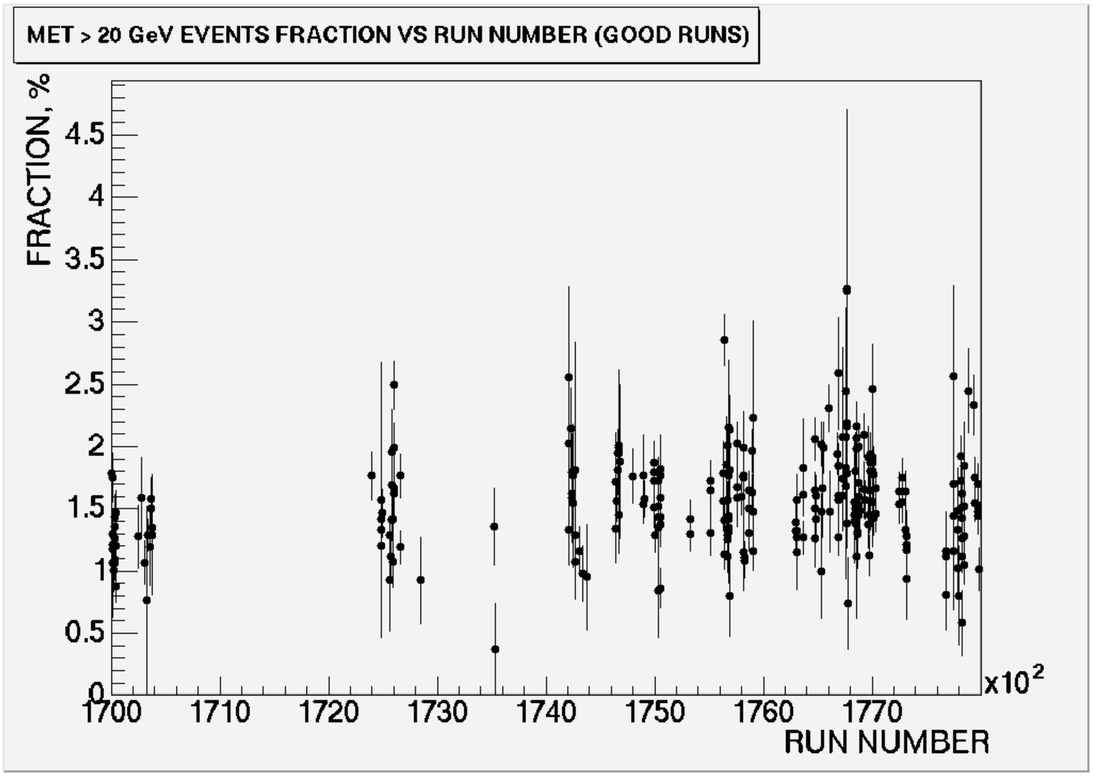


Figure 28: Fraction of events with missing E_T more than 20 GeV versus run number.

6 Issues and perspectives

Once the data quality criteria are agreed upon, there are still several issues that need to be solved.

The first issue is to organise the permanent monitoring and feedback with shifters. Since the program was successfully tested online and the special Root based GUI was created to view the output histograms, all is ready for the permanent online monitoring.

The second issue is to better document hardware changes in the calorimeter electronics and make the information more easily available. It is important not only to catch problems, but to understand them and keep track of when they are solved. The person(s) responsible for the calorimeter run quality should not have to scan the CAL logbook every day to retrieve that information. There could be a notification system for events that potentially affect the data quality.

The next issue is to make available the data quality information. We plan to use the runs quality database to save "bad" cells and/or BLSs like $< name, value >$ pairs. Special classes was created to read and make use of this information. The main problem here, that there is no interface to read the data base information from the user program. As one can see from section 5 this information is very important for missing E_T calculation and most likely for physics object reconstruction.

References

[1] http://d0-france.in2p3.fr/WORKING_GROUPS/DQ

[2] http://www-d0.fnal.gov/Run2Physics/np/d0_private/meetings/NP-07-15-03-YG.pdf

A dq_calo manual

To install dq_calo package the usual framework setup should be done. I.e.

- setup D0RunII p##.##.## (the package was tested with p13.06.01 and p14.03.00 versions).
- setup d0cvs
- if you don't have any release directory: newrel p##.##.## release_directory_name
- cd release_directory_name
- addpkg dq_calo current_version[1]
- srt_setup SRT_QUAL=maxopt
- d0setwa
- gmake dq_calo.all

The package can be started online or offline. To start it in the online mode the run_dq_calo_online.rcp file should be used. To start it in the offline mode the run_dq_calo_offline.rcp file should be used. The bool variable ONLINE in DQCalo.rcp file should be *true* for the online mode and *false* for the offline one[2]. Before starting the correct output directories should be setup in the DQCalo.rcp file.

- string RootDirName = "full_directory_name/dq_calo" — the files with root trees will be saved as "full_directory_name/dq_calo_runNumber.root"
- string TxtDirName = "full_directory_name/dq_calo" — the report text file will be saved as "full_directory_name/dq_calo_runNumber.txt"
- string DBDirName = "full_directory_name/CAL" — the files with an information for the data quality data base will be saved as "full_directory_name/CAL_runNumber.dat"

It is better to have one output directory for all 3 types of output files. **Don't change any other parameter in rcp file if you are not an expert!**

To start the program online or offline with raw data files 3 files should be available in the working directory. They are cal_nlc_146225.coef, cal_zsup_limits.dat, cal_zsup_noise.dat. You can make a link to these files. The files location is $D0RUNII_DIR/rundata/d0reco/ [3].

To start program — run the dq_calo_x. After each 100[4] zero bias events the text report file will be created/updated. In the beginning of the run the *root and *dat files for the previous run will be saved. To view the histograms you can use Monitor application (dq_calo_analysis package, see section B)[5].

[1] now: v01-00-03

[2] More precisely this variable should be true to read raw data

[3] In version p15.02.01 the file cal_zsup_noise.dat located only in $D0RUNII_DIR/caltables/data/

[4] you can change this number via rcp parameter Min_number_of_events

[5] only histograms for the runs before current can be viewed

B dq_calo_analysis manual

The dq_calo_analysis package contains following classes.

- Classes to decode the data quality information from CAL_runNumber.dat files. These classes can be used to access the lists of bad cells/BLSs. The main class — "DecodeDB".
- Example how to calculate corrected missing E_T from thumbnails. To correct the missing E_T the bad cells/BLSs are removed using "DecodeDB" class. The program name is "DQCaloAnalysis".
- The "Monitor" program to view histograms created from the dq_calo_runNumber.root files. This program can create html page with information about runs data quality.

The package can be install in the same way as dq_calo[6] (see section A).

B.1 Monitor program

Before starting the "Monitor" program you should defined the parameter DirName in the Monitor.rcp file. This parameter indicates where dq_calo package output files are located. All files (*root, *txt, *dat) should be in the same directory. The parameters FileName, DBFileName should not be changed in the rcp file. If you are an expert and want to create the html report, you should specify the HtmlDirName parameter — the destination directory for the report (don't forget to make necessary corrections in ReportHtml.rcp file). If you want to see the reference histograms you should specify the location of the file with reference histograms. The file with reference histogram available at http://d0-france.in2p3.fr/WORKING_GROUPS/DQ/reference.root To start program simply run Monitor_x. At least one *root file should be available at the specified location.

The Monitor program is a root-based GUI (see screen-shots 29–31). You have all root functionality for the histograms. The following histograms are available.

1. 2d histograms (channel/BLS number,ADC number) for the channel occupancy.
2. 2d histograms (ϕ, η) for the cells occupancy for each calorimeter layer (see screen-shot 29).
3. 2d histograms (channel/BLS number,ADC number) for the channel mean energy.
4. 2d histograms (ϕ, η) for the cells mean energy for each calorimeter layer.
5. 2d histograms (channel/BLS number,ADC number) for the channel RMS.
6. 2d histograms (ϕ, η) for the cells RMS for each calorimeter layer.
7. Missing E_T distribution (see screen-shot 30). The reference histogram available.
8. Scalar E_T distribution (see screen-shot 30). The reference histogram available.
9. Number of hitted cells in event distribution (see screen-shot 30). The reference histogram available.
10. Mean energy distribution (see screen-shot 30). The reference histogram available.
11. RMS distribution. The reference histogram available.

[6] current version: v01-00-02

12. Missing E_T versus run number (see screen-shot 31).
13. Scalar E_T versus run number (see screen-shot 31).
14. Number of cells in event versus run number.
15. Mean energy in event versus run number.
16. RMS versus run number.

The histograms 1–6 available when only 1 run is loaded. The histograms 12–16 available only when more than 1 runs are loaded.

Menus **File**, **Histograms**, **Runs**, **Html** are available. See menu-bar at the top of the window.

- **File**. You can **Exit** from program.
- **Histograms**. Choose **Hide** submenu. The small window with list of histograms will appear. Select several histograms and press **OK**. Selected histograms will be invisible. You can make invisible histograms visible using **Show** submenu.
- **Runs**. You can choose the quality of the runs to be shown: **All** runs, **Good only**, **Reasonable only**, **Good and reasonable** runs, **Bad only**.
- **Html**. For expert only! You can create an html page with runs quality report. See screen-shot 32.

Tools bar allow to choose runs to be loaded, to choose number of histograms to show at one page, show the report window.

- If the box **last run** selected only the last available run will be loaded.
- In the **from** box you can choose the first run to show.
- In the **to** box you can choose the last run to show. If the last run equal the first run, only one run will be shown and histograms 1–6 are available. The last run number can not be less then the first one.
- **Reload** button allow to reload list of available runs.
- **X,Y** boxes allow to choose number of histograms to show.
- If **Report** check box is selected, the report window will be shown. If only one run is loaded, in report window the dq_calo_runNumber.txt file will be shown. If several runs are loaded only one line with quality information for each run will be shown (see screen-shots 31).

In the bottom of the window the navigation bar is located.

- In the histogram box (at the left side) you can choose a histogram to show. If more then one histograms can be shown at one page, the selected histogram will be the first histogram (at the left-top corner of the main window).
- The left arrow on the right side will show the first page with histograms.
- The right arrow on the right side will show the next page with histograms.

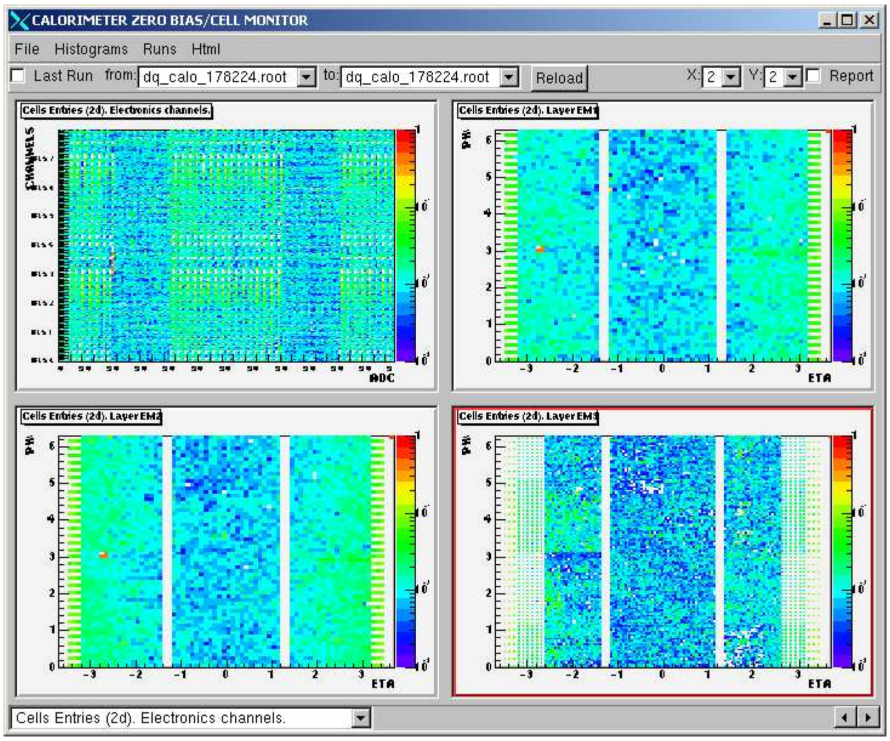


Figure 29: "Monitor" screen shot. The occupancy for the electronics channels, EM1, EM2, EM3 layer are shown. At (ϕ, η) near (3,-2.65) the noisy tower can be seen.

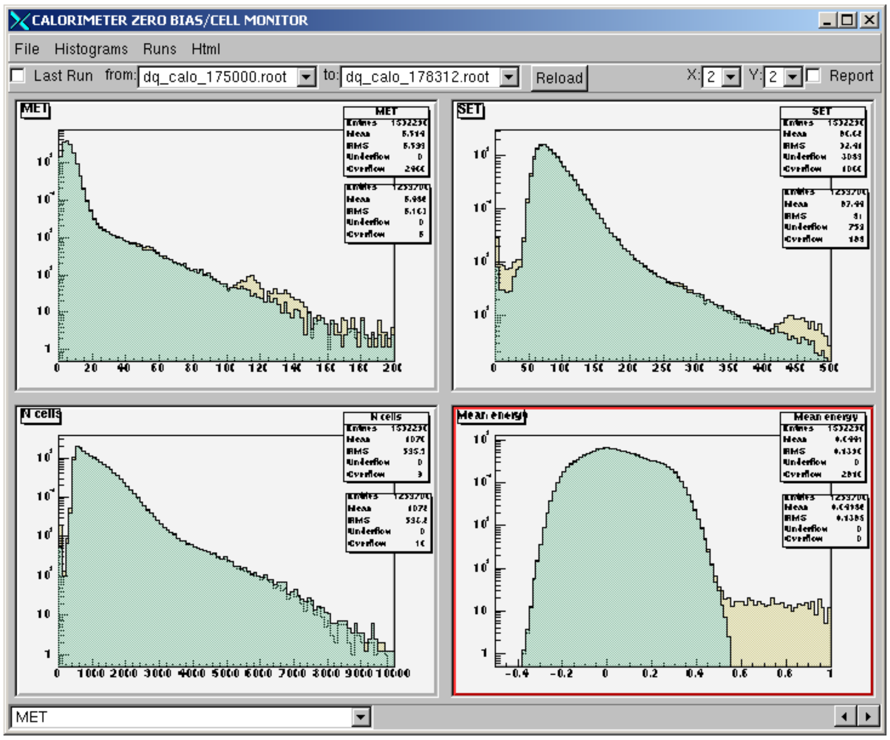


Figure 30: "Monitor" screen shot. MET, SET, number of cells, mean energy distributions are shown. The reference histograms are green and the current histograms — light-brown.

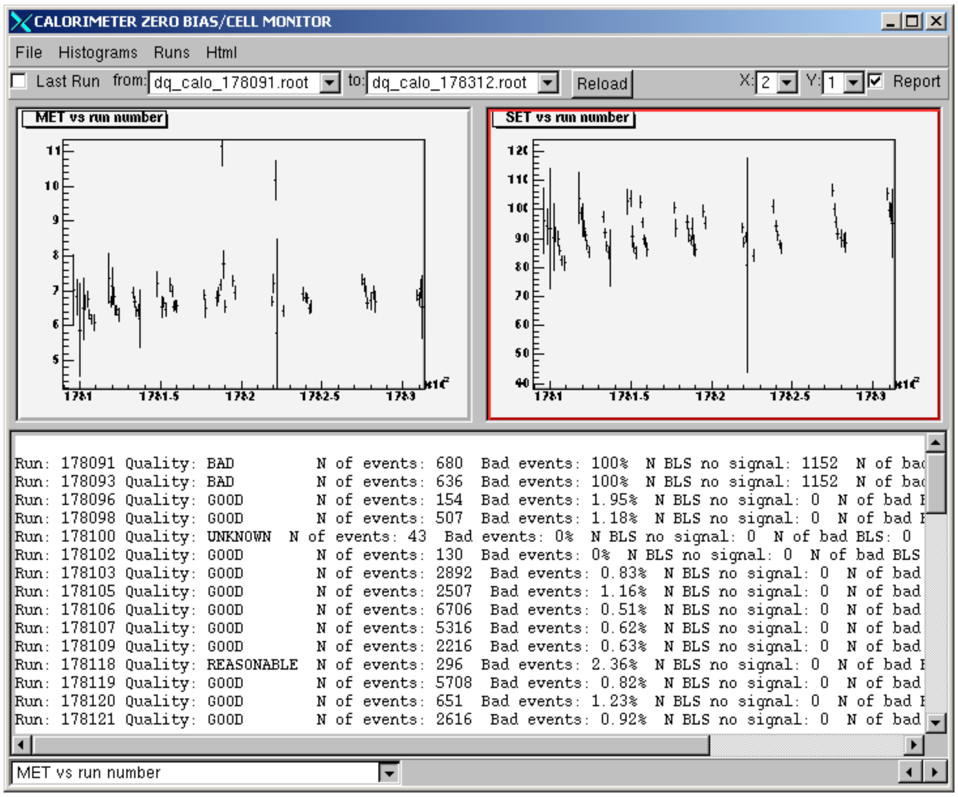
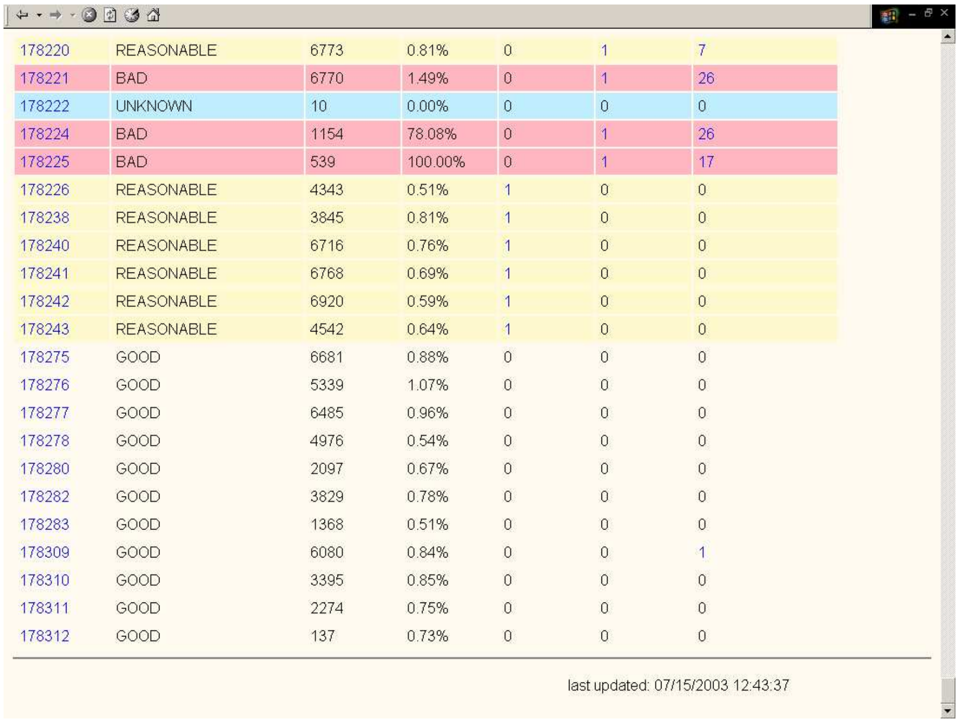


Figure 31: "Monitor" screen shot. The MET, SET vs run number graphics presented. The report window show the quality information for the loaded runs.



178220	REASONABLE	6773	0.81%	0	1	7
178221	BAD	6770	1.49%	0	1	26
178222	UNKNOWN	10	0.00%	0	0	0
178224	BAD	1154	78.08%	0	1	26
178225	BAD	539	100.00%	0	1	17
178226	REASONABLE	4343	0.51%	1	0	0
178238	REASONABLE	3845	0.81%	1	0	0
178240	REASONABLE	6716	0.76%	1	0	0
178241	REASONABLE	6768	0.69%	1	0	0
178242	REASONABLE	6920	0.59%	1	0	0
178243	REASONABLE	4542	0.64%	1	0	0
178275	GOOD	6681	0.88%	0	0	0
178276	GOOD	5339	1.07%	0	0	0
178277	GOOD	6485	0.96%	0	0	0
178278	GOOD	4976	0.54%	0	0	0
178280	GOOD	2097	0.67%	0	0	0
178282	GOOD	3829	0.78%	0	0	0
178283	GOOD	1368	0.51%	0	0	0
178309	GOOD	6080	0.84%	0	0	1
178310	GOOD	3395	0.85%	0	0	0
178311	GOOD	2274	0.75%	0	0	0
178312	GOOD	137	0.73%	0	0	0

last updated: 07/15/2003 12:43:37

Figure 32: A web page example created with "Monitor" program.

C For person on shift

- To use the program please login d0ol06.
- Check that dq_calo process is working. To do this type *ps -C dq_calo_x*. You should see several processes dq_calo_x, but in reality it is one process. If no such process is running, start it. For this:
  - cd to the release directory /home/shary/DQ/p15.01.00/
  - setup D0RunII p15.01.00
  - srt_setup SRT_QUAL=maxopt
  - d0setwa
  - dq_calo_x -rcp dq_calo/rcp/run_dq_calo_online.rcp -log results/dq_calo.log -out results/dq_calo.out &
- You can find a recent results in the directory /scratch/DQ/results/ (only on this computer). To see the run data quality status look in the dq_calo_runNumber.txt file. The file with the biggest number is current (the file will appear or updated after 100 zero bias events). The results has a sense after $\sim$ 300 events. If you see the cell marked as bad and this cell has a big number of entries — this is a hot cell.
- to see the histograms for the files before current type Monitor_x in the release (/home/shary/DQ/p15.01.00/ directory. The root GUI will started (see section B.1).