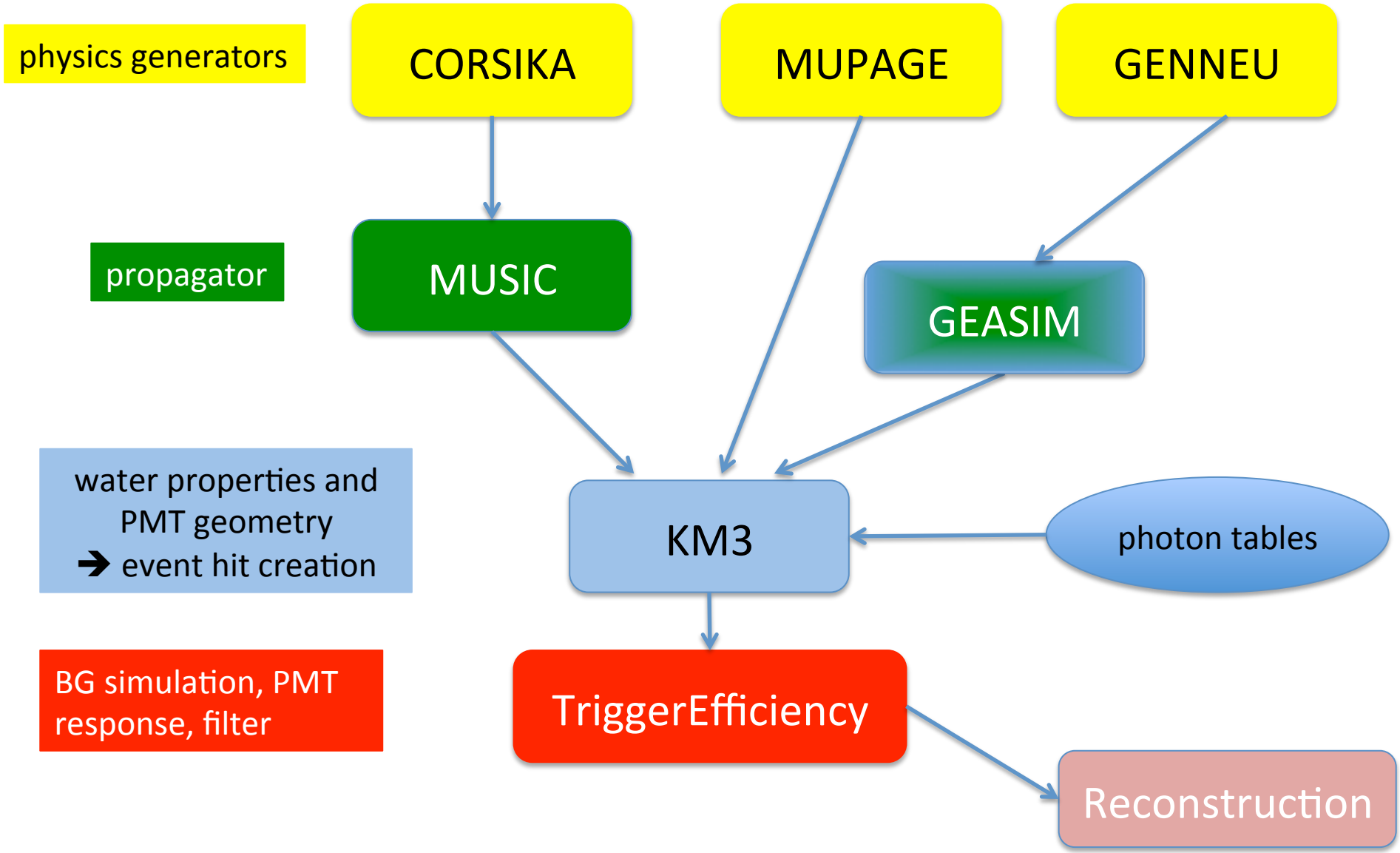
Data/MC comparison the Run-By-Run way

A. Margiotta

24 – Sep – 2011

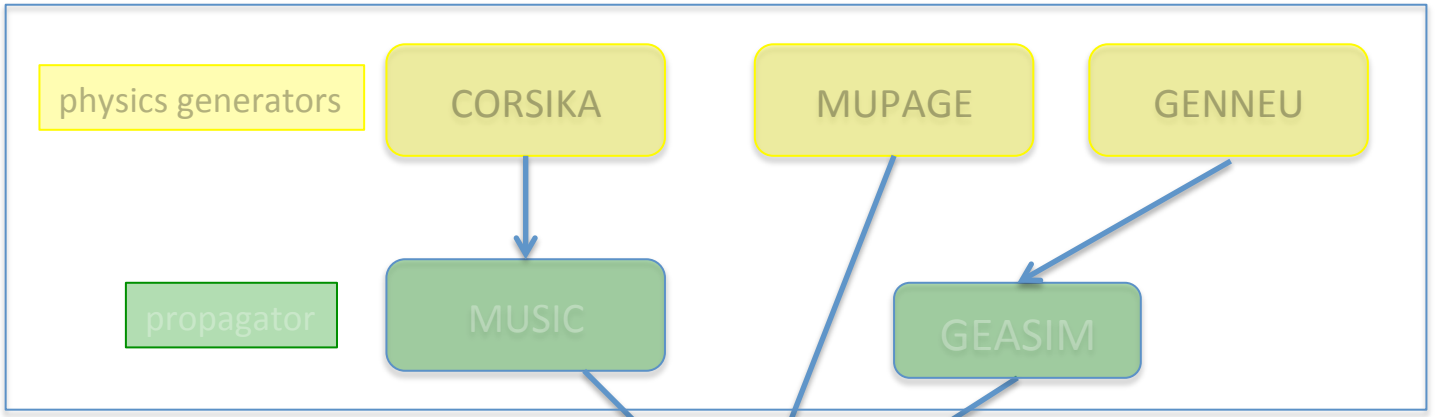
MANTS 2011 - Uppsala

ANTARES simulation chain

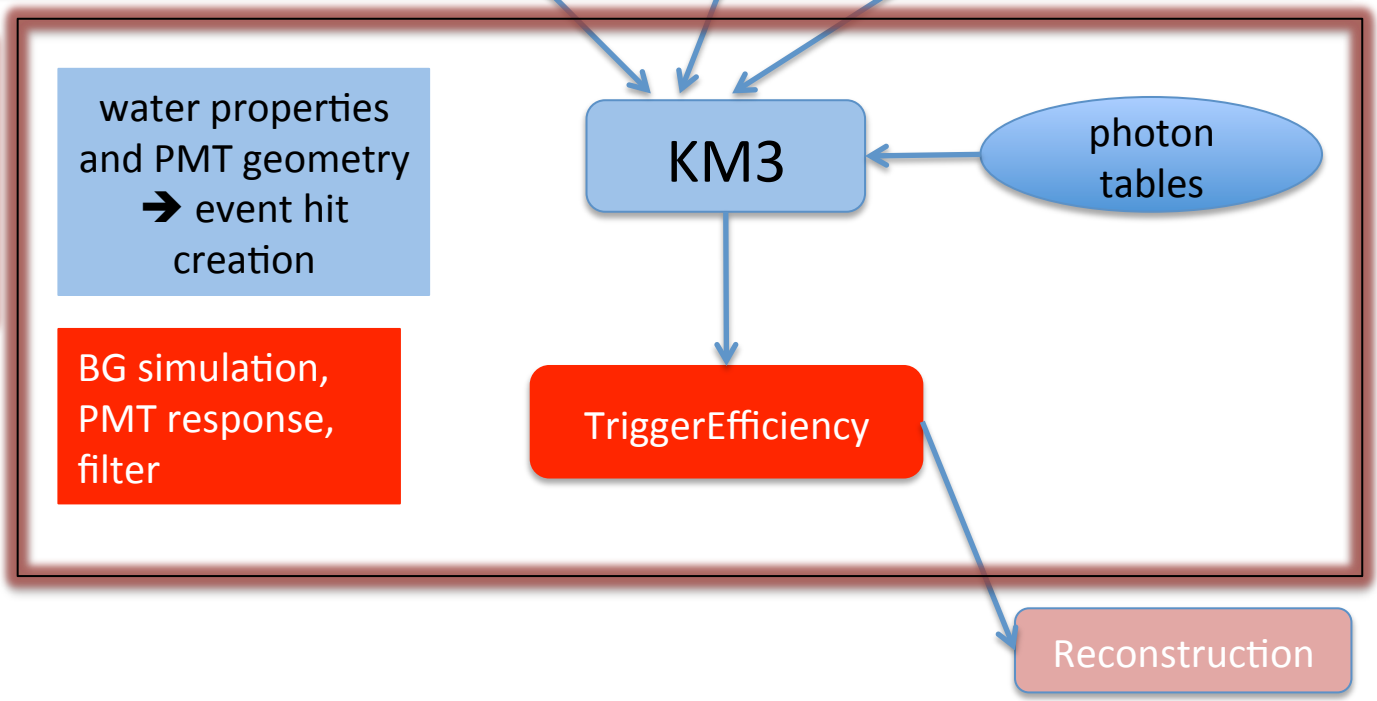


ANTARES simulation chain

uncertainties due to physics models



critical steps for a realistic description of the detector and of the environment



water properties and
PMT geometry
➔ event hit creation

- **“static”** characteristics: periodically to be revised, but constant for long periods

- present input parameters seem to represent quite well ANTARES site water (within 10-15%) ➔ new measurements are being analized

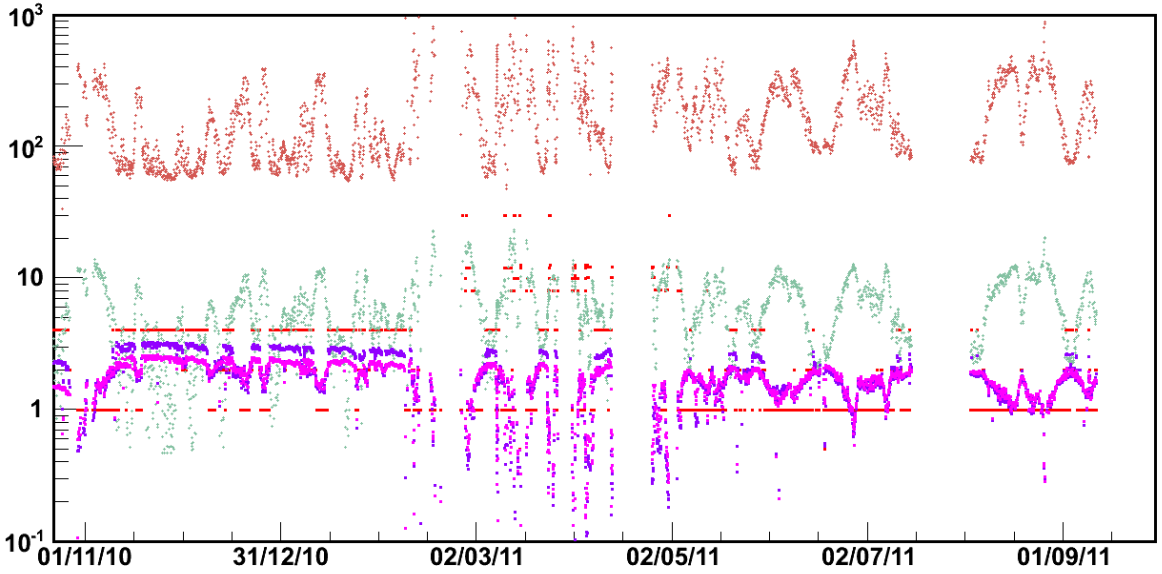
- photon propagation simulation ➔ improving present codes + development of new codes ➔ study of the effect on the reconstruction in progress.

- geometrical characteristics of the PMT ➔ uncertainties around 20%

our “best” values at the next MANTS ???

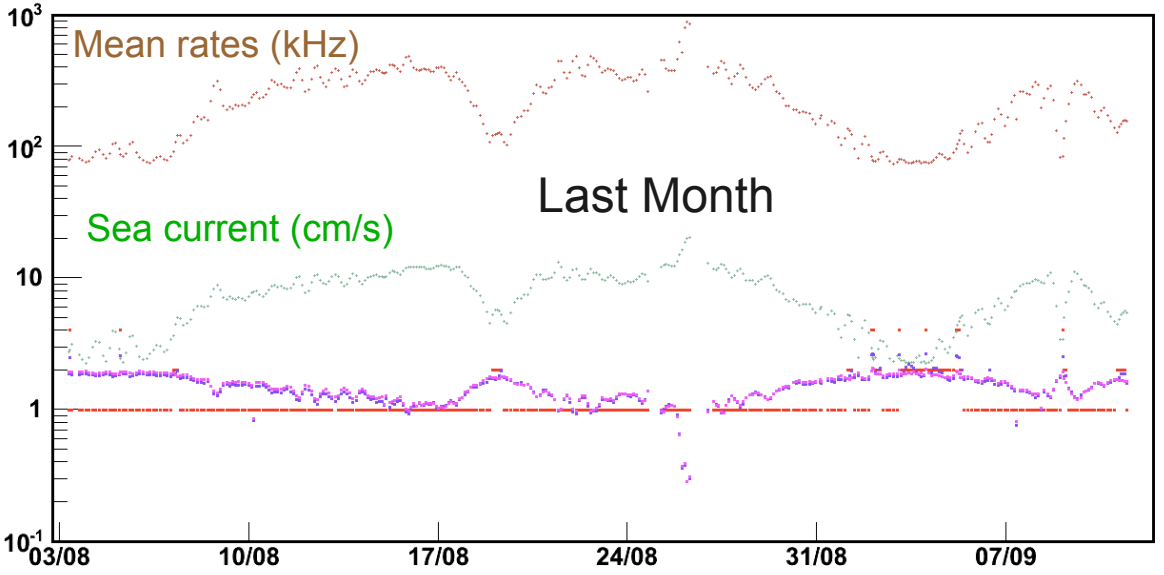
environmental conditions affect strongly ANTARES data acquisition

Single-line reconstructed muons (Hz)
Multi-line reconstructed muons (Hz)
Sea current (cm/s)
Mean rates (kHz)



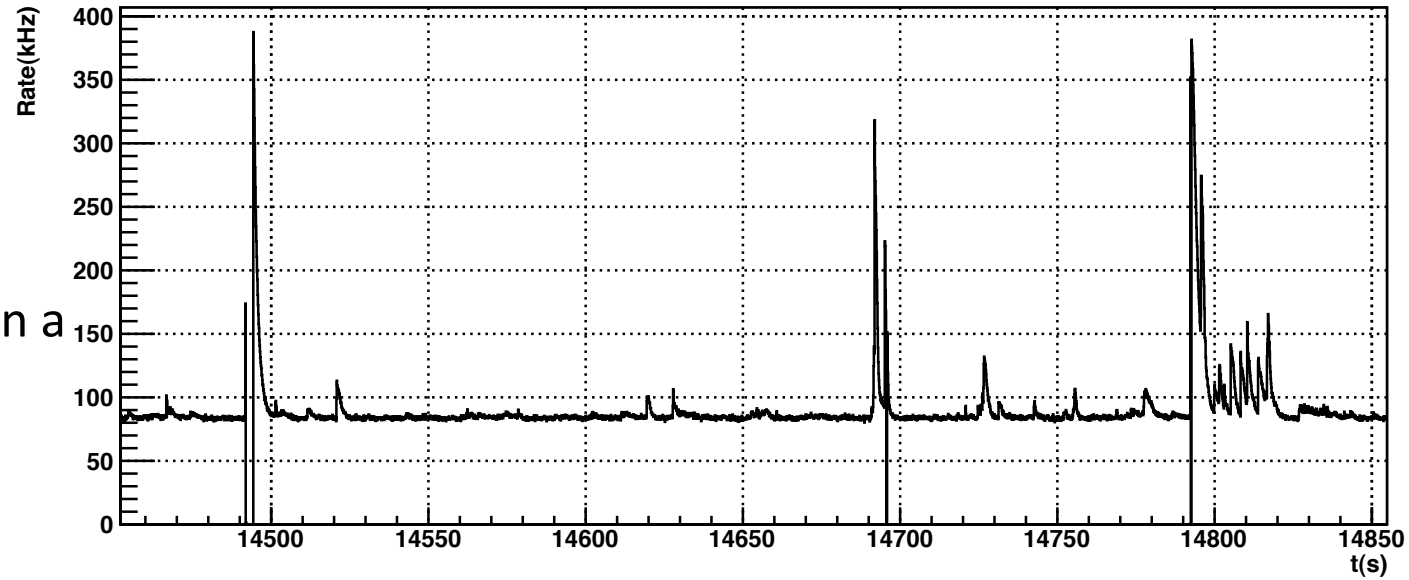
BG simulation,
PMT response,
filter

Single-line reconstructed muons (Hz)
Multi-line reconstructed muons (Hz)



BG simulation, PMT response, filter

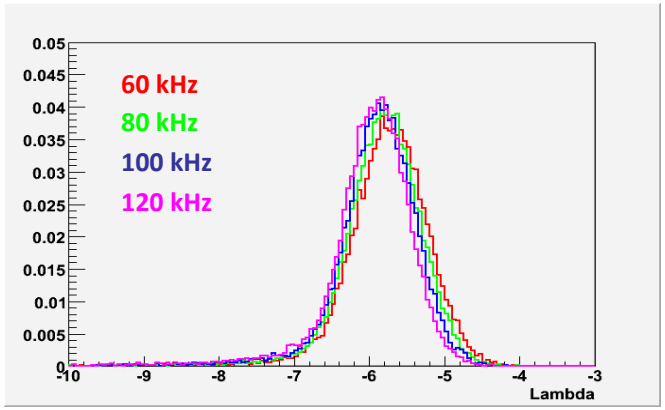
optical BG changes on a short time scale



optical BG = baseline (BL) + bursting events (BF) – a few seconds long

How to proceed for BG evaluation?

1. constant BL
   - burst events not accounted for.
2. hit counting extracted from a unique run representative of a period
   - variability of BL and BF during long data taking period not accounted for



reco quality parameter
NB: old reco strategy

3. Summary-of-summaries approach:
   - selection of a list of runs according to a set of quality criteria
   - reading information (nb of hits in the timeslice, triggers, active Oms…) directly in the timeslices randomly extracted from the runs included in the list:
   - good approximation for long list of runs
   - only few minutes extracted from each run
   - requires a new reprocessing in case of different analysis/run selection

FINALLY

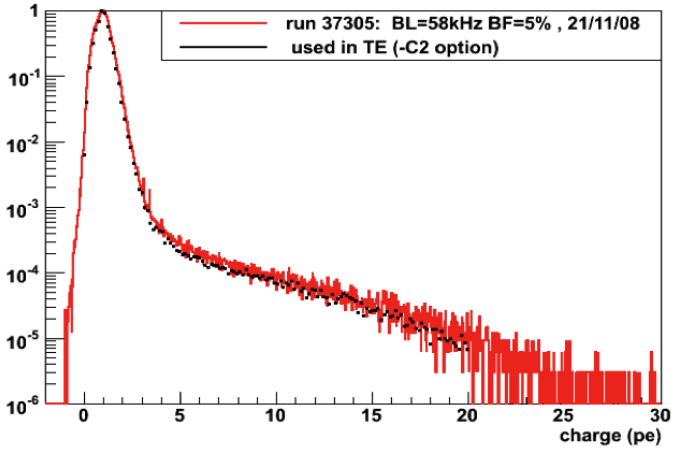
RUN-BY-RUN APPROACH

General ideas -1

- **NEED OF CHECKING TIME EVOLUTION OF DATA ACQUISITION**
- **ONE DATA RUN⇐⇒ ONE MC RUN**
- **ALL RUNS PROCESSED – NO QUALITY SELECTION**
- **MC LIVETIME :**
  - neutrinos : fixed number of events (5.10^8 per run)
  - atmospheric muons : a fraction of the data LT (10 and 20% respectively for MUPAGE and CORSIKA)
- **For CORSIKA:** 5 mass groups, different zenith angle intervals, 3 energy ranges, E^{-2} spectrum, to be weighted according to a physics model.

How to add the BG?

- **LOOP OVER MC EVENTS:**
- random position selected along the data run
- extraction of a group of timeslices
- information concerning nb of hits, triggers, active OMs…. read from timeslices and used for optical background definition and data filtering for a group of events (1 timeslice ←→ 1 MC event)
- charge for each hit extracted from the measured charge distribution
- BG hits added to MC events (only active OMs in that timeslice are considered)
- data filtering according to the active triggers in the run



Disadvantages:

- large CPU time requirement, but affordable.

Advantages:

- “puntual” representation of the BG (baseline and burst fraction)
- easy checking of the time evolution of data taking
- easy reprocessing : updated codes, new values of parameters,etc…
- good for all analyses /run selection etc.
- easy to use: events reconstructed with different strategies directly available
- moderate storage requirements: sea level showers produced with CORSIKA stored + all final files with reconstructed events ➔ MUPAGE + neutrinos 2007-2011 → ≈ 1TB

PRESENT SITUATION
(RBR V0.1)

MUPAGE+ NEUTRINOS production for 2007-2011

10014 runs ➔ real LT = 1108,57 days ➔ MC LT = 111 days

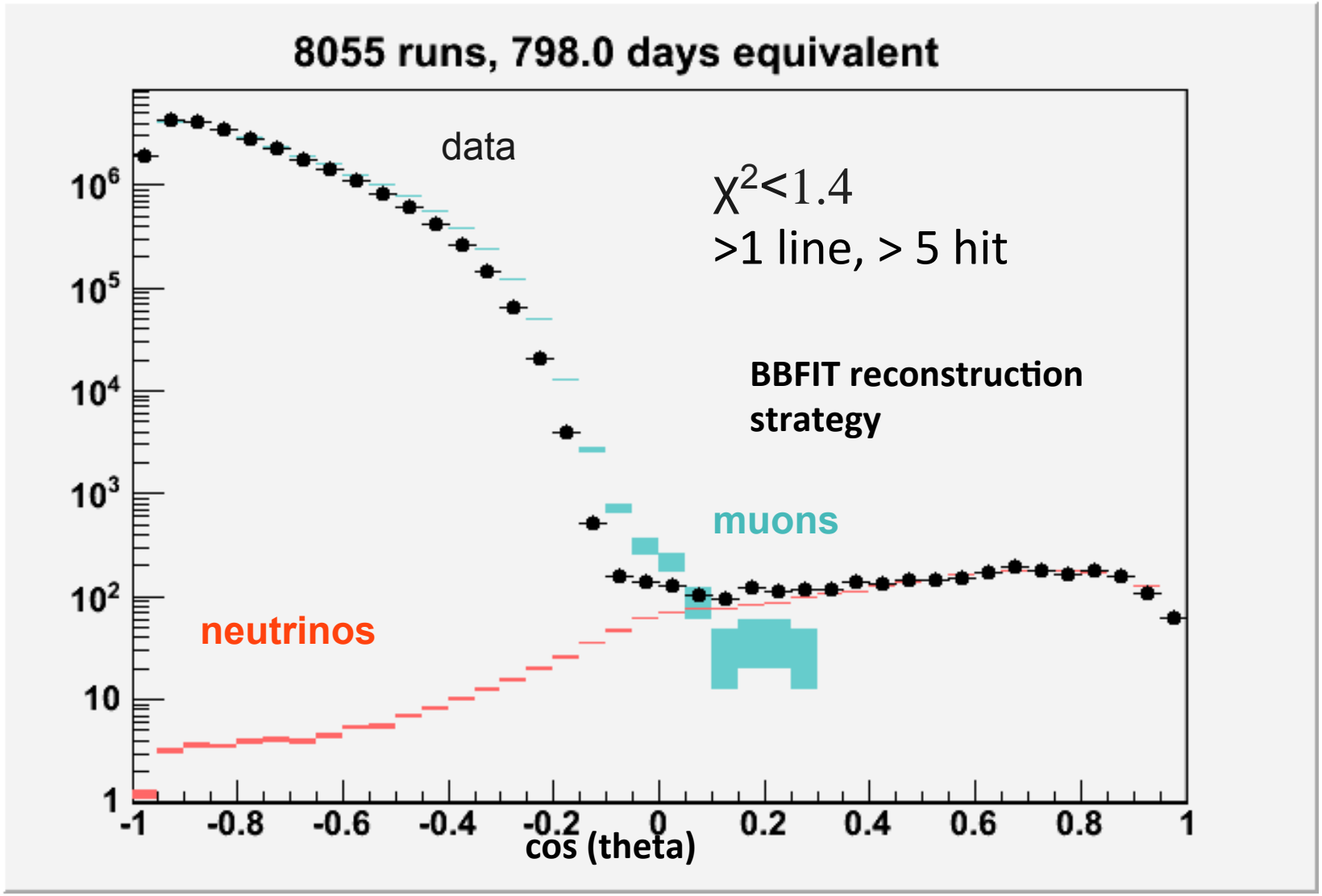
3 weeks at Lyon CC

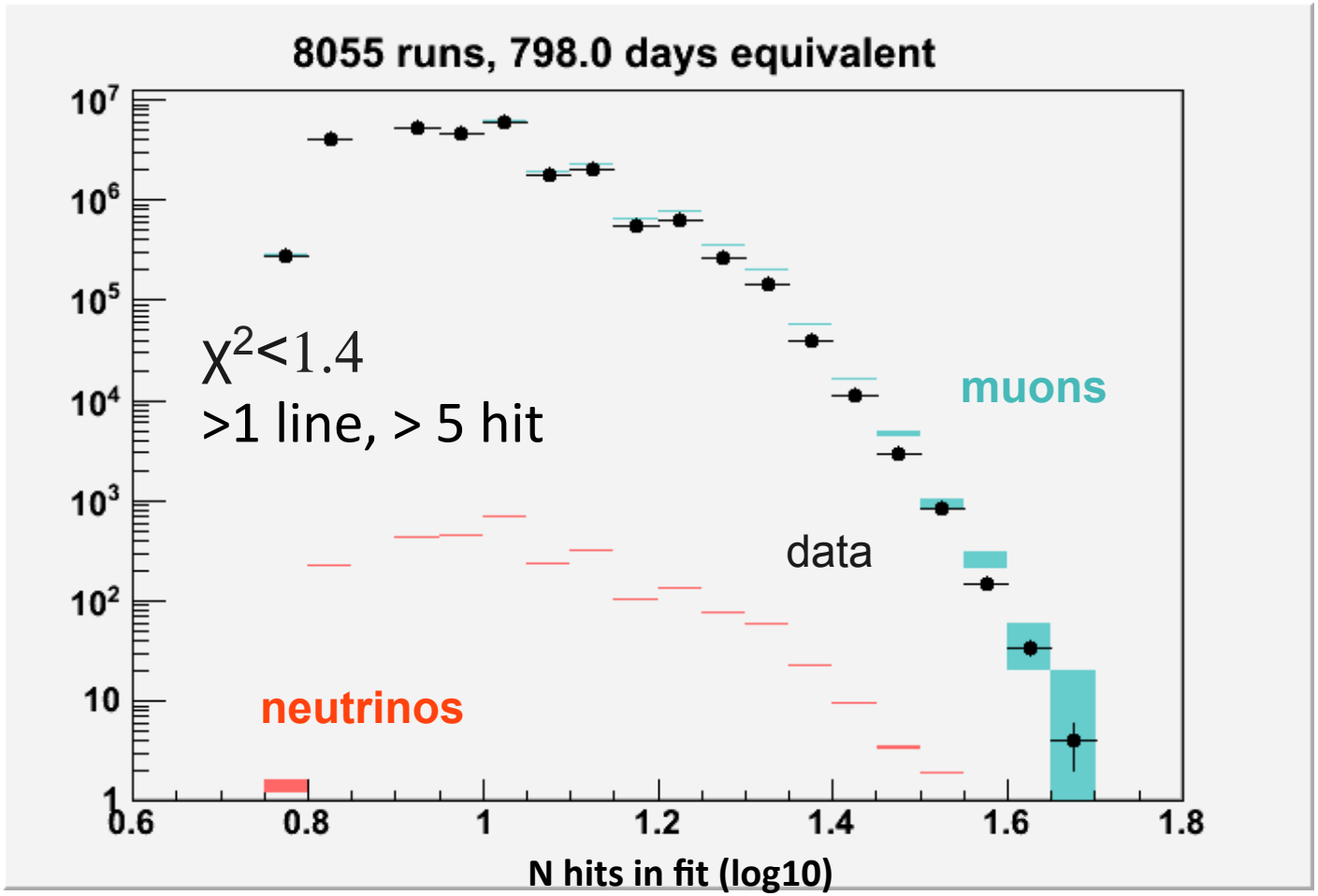
hard to give an absolute reference for CPU time requirement, strongly dependent on the available hardware

CORSIKA ➔ not available yet, shower production in progress

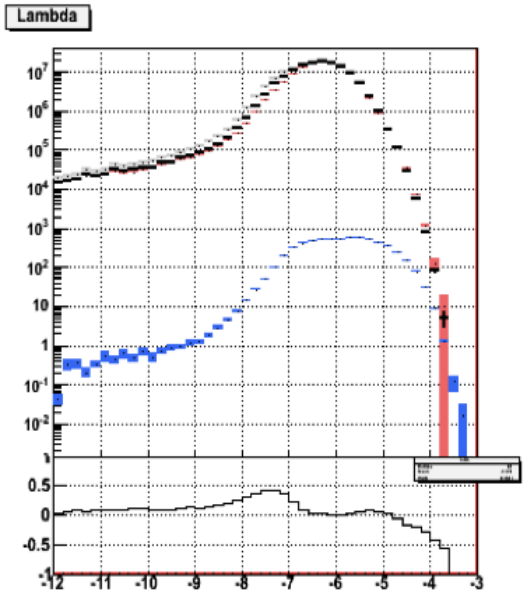
SOME PLOTS

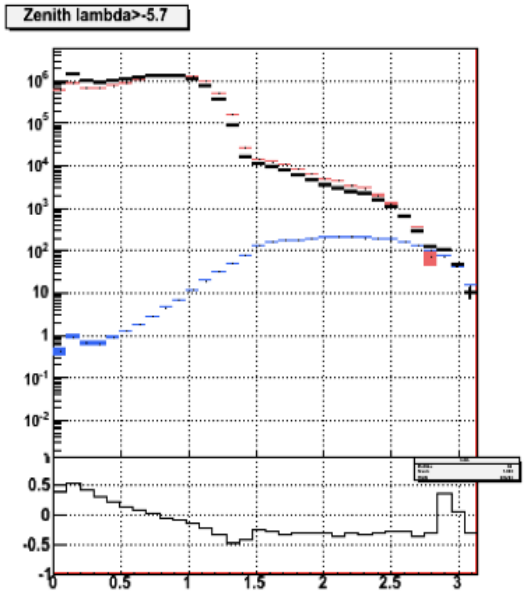
2008 - 2011

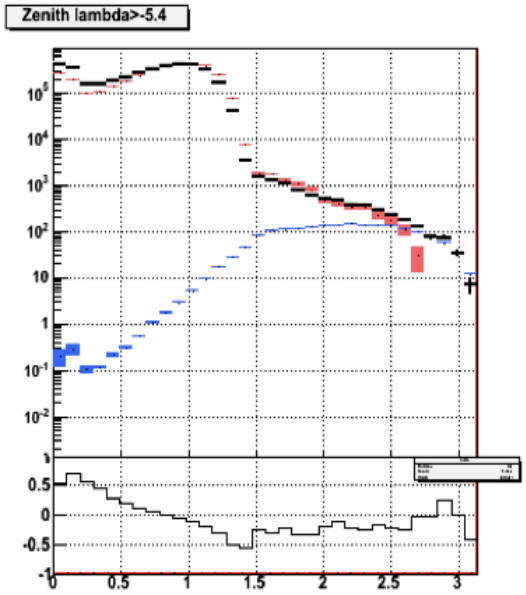

8055 runs, 798.0 days equivalent
data
χ^2<1.4
>1 line, > 5 hit
BBFIT reconstruction strategy
muons
neutrinos
10^6
10^5
10^4
10^3
10^2
10
1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
cos (theta)

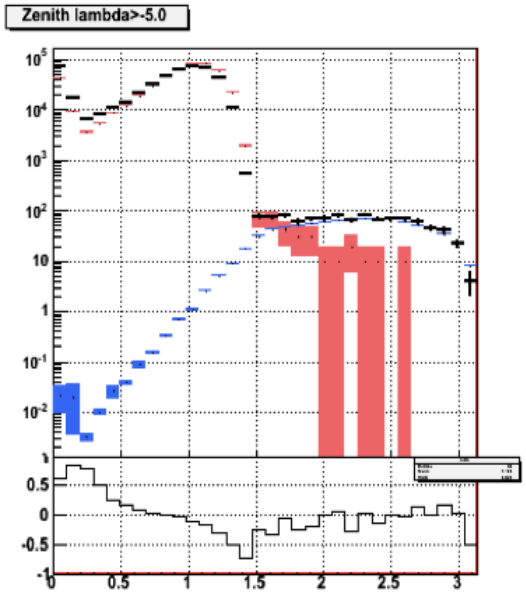

8055 runs, 798.0 days equivalent
10^7
10^6
10^5
10^4
10^3
10^2
10
1
χ^2<1.4
>1 line, > 5 hit
muons
data
neutrinos
0.6
0.8
1
1.2
1.4
1.6
1.8
N hits in fit (log10)

AAFIT reconstruction strategy

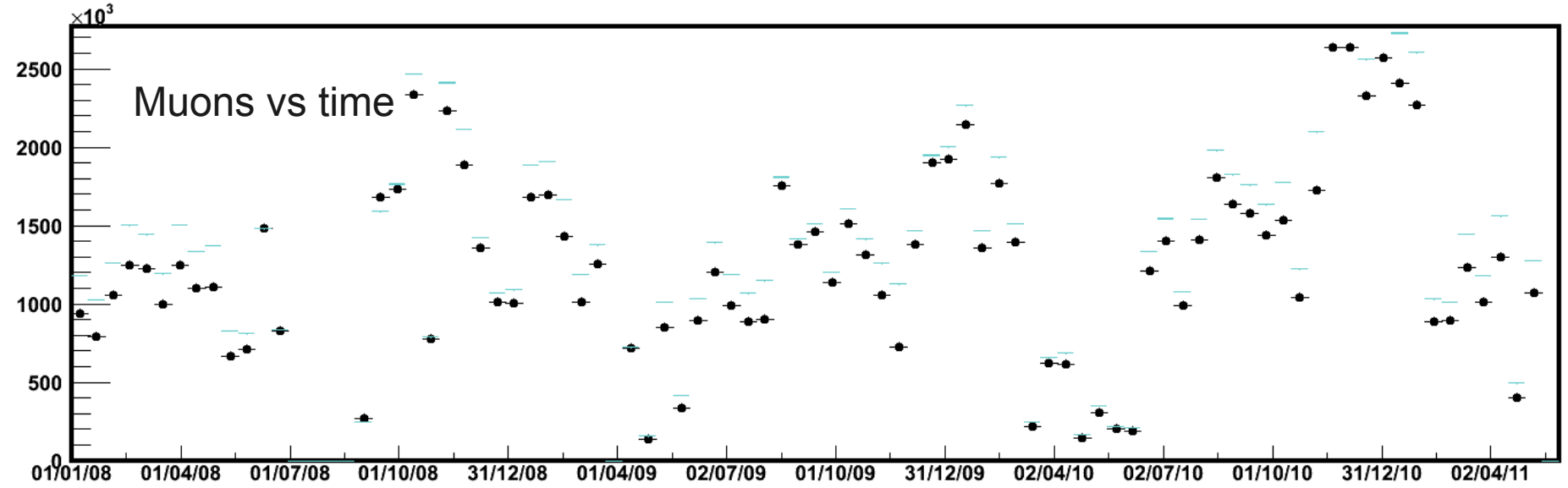


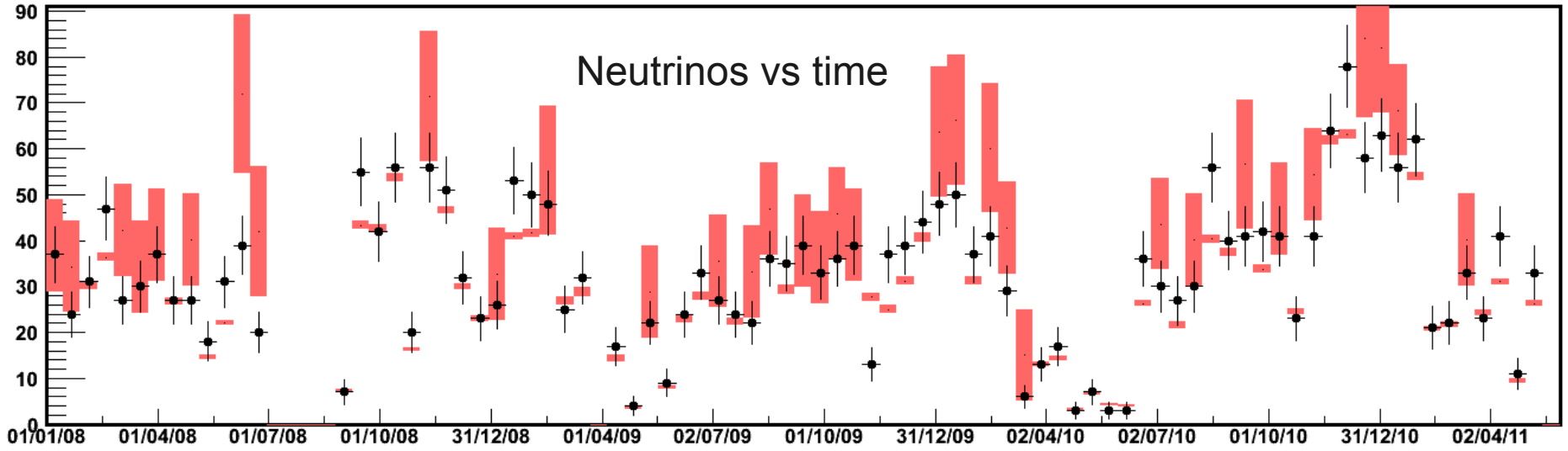




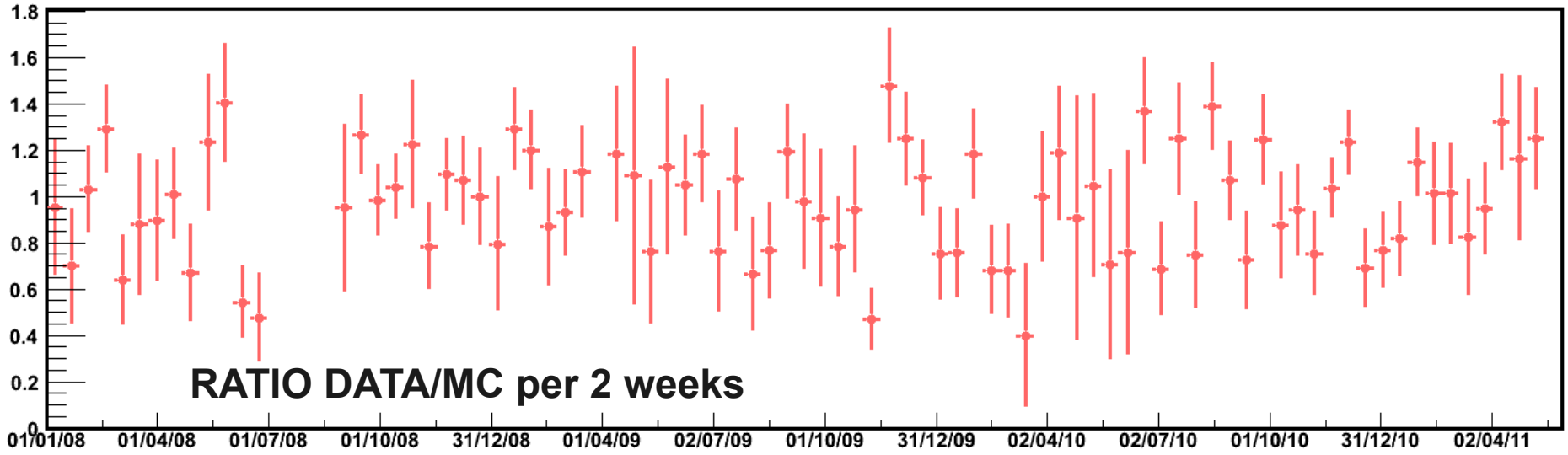


Comparison data/MC per 2 weeks

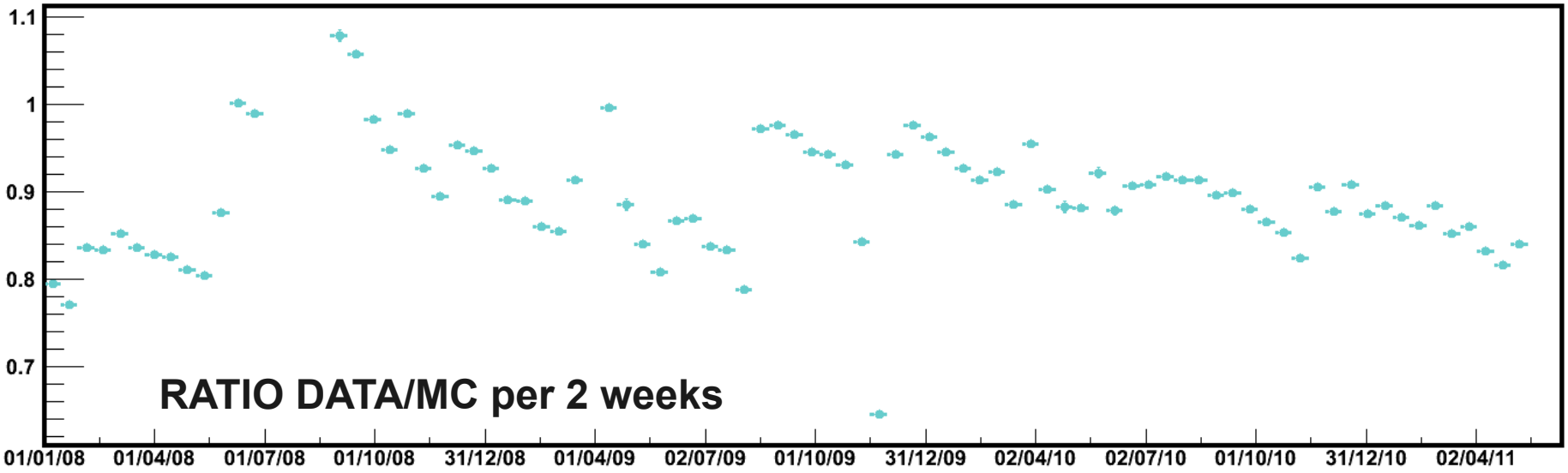




Neutrinos

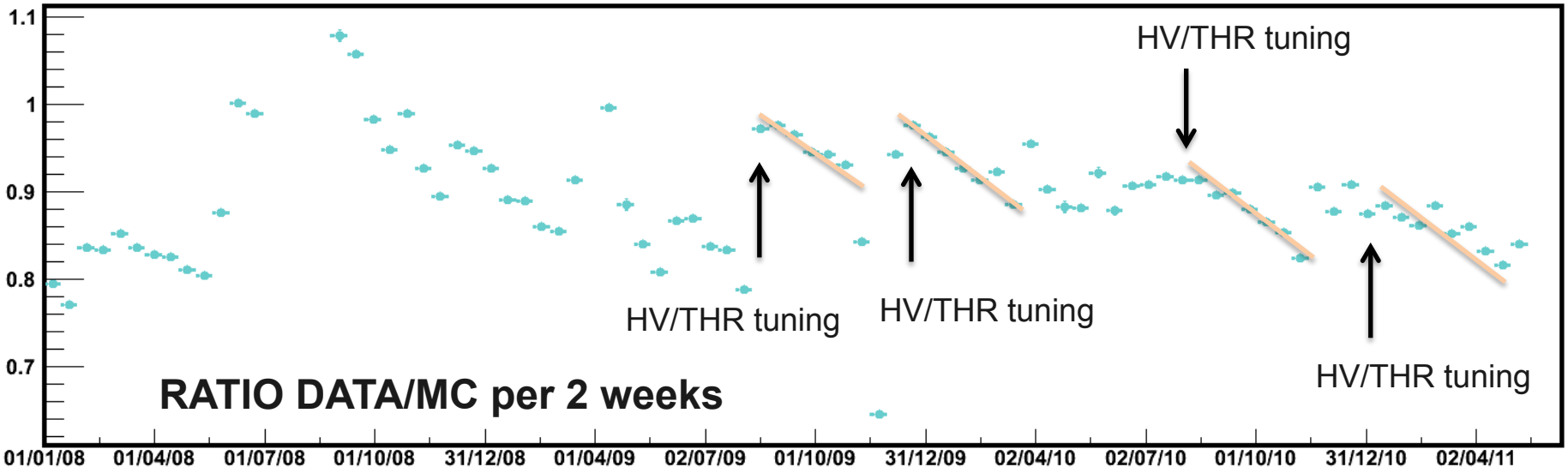

RATIO DATA/MC per 2 weeks
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
01/01/08
01/04/08
01/07/08
01/10/08
31/12/08
01/04/09
02/07/09
01/10/09
31/12/09
02/04/10
02/07/10
01/10/10
31/12/10
02/04/11

Muons


1.1
1
0.9
0.8
0.7
RATIO DATA/MC per 2 weeks
01/01/08
01/04/08
01/07/08
01/10/08
31/12/08
01/04/09
02/07/09
01/10/09
31/12/09
02/04/10
02/07/10
01/10/10
31/12/10
02/04/11

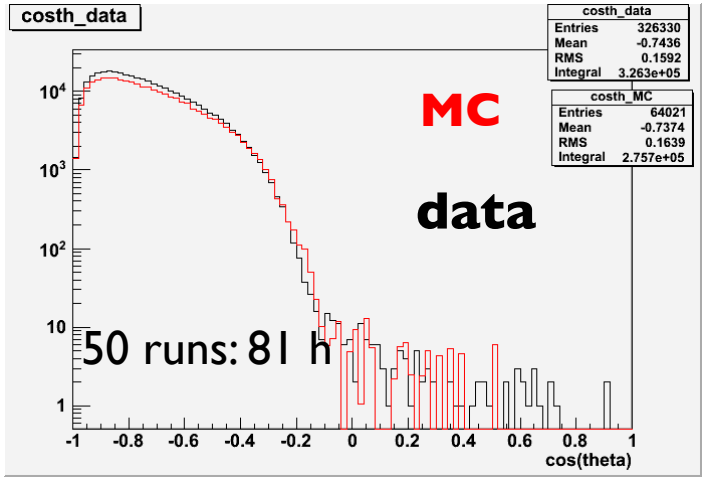
Muons

sensitive tool for checking the detector
in future versions thresholds in MC will be read from database, following “true” time evolution



todo list

- complete the CORSIKA simulation



- direct access to the database for thresholds, calibration, geometry file……
- inclusion of a realistic hit time distribution
- increase of statistics ➔ 30% of real LT
- periodic reprocessing with:
  - updated vrs of the software codes
  - more accurate description of the environment and of the detector

conclusions

- Run-by-run simulation → time evolution of data acquisition conditions
- increased CPU time requirement → feasible
- easy to use
- user-friendly approach
- immediate comparison data/MC