Pattern recognition of simulated fake trigger events

B.Pastirčák, P. Bobík, M.Staroň, K.Kudela, M.Putiš, M.Bertaina, F.Fenu, K.Shinozaki

JEM-EUSO Simulation & Atmospheric Monitoring meeting 2013,
University of Sofia, *9. October 2013*

Short introduction

J

The goal of the trigger system is to detect the occurrence of scientifically valuable signal among very huge background noise detected by JEM-EUSO. The UV background registered is randomly distributed. We study if these random processes produce fake pattern, which could be mistakenly interpreted as extreme energy cosmic rays events. For this purpose very huge amount of measurements on one photo detection module with only detector noise were simulated. To distinguish between such simulated fake events and real extreme energy cosmic rays events we try to apply and develop PR methods.

Level	**Triggers rate PDM level [Hz]**	**Triggers rate FS level [Hz]**
Photon trig (channel)	$\sim 5.2 \times 10^8$	$\sim 7.8 \times 10^{10}$
Counting trig (EC)	$\sim 7.1 \times 10^5$	$\sim 1.1 \times 10^8$
1^{st} (PDM) Persistency trigger	~ 7	$\sim 10^3$
2^{nd} (PDM cluster) Linear track trigger	$\sim 6.7 \times 10^{-4}$	$\sim 10^{-1}$
EECR expected rate	$\sim 6.7 \times 10^{-6}$	$\sim 10^{-3}$

Short introduction

Very high statistics of simulated background needed
$\longrightarrow$ 10^5 events $\longrightarrow 10^{14}$ GTU's

Impossible to simulate by ESAF:
$\longrightarrow$ 10^3 slower then used code
$\longrightarrow$ cannot be computed parallely (mem. share)

Fast standalone code written in C++ was developed

- One PDM simulated
- Persistency Track Trigger algorithm → 1st level → 1Hz/PDM
- Linear Track Trigger algorithm → 2nd level → 1 mHZ/PDM
- Background source → Poisson distribution of average 500 photons (m^{-2} s^{-1} sr^{-1}) = 2.1 photons/pixel/GTU

Simulations

M36

BG = 2.1 ph/pix/GTU

PTT_integr = 43

LTT integration = 145

Consecutive GTU = 5

Yellow pixel th = 4

- **Accummulated:**
- 10^{12} GTU's $\rightarrow$ 3months CPU time on part of PC cluster
- 12000 LTT triggers $\rightarrow$ 1mHz/PDM
- 750000 PTT triggers $\rightarrow$ 0.1 Hz/PDM

M64

BG = 0.4 cts/ms

PTT_integr = 52

LTT integration = 115

Consecutive GTU = 5

Yellow pixel th = 2

- still running
- at present $5x10^{11}$ GTU's

Comment

However the trigger rates reached by massive simulation for M64 configuration do not match the expectation. It has to be inspected deeper this problem before starting new run of simulations

Pattern recognition

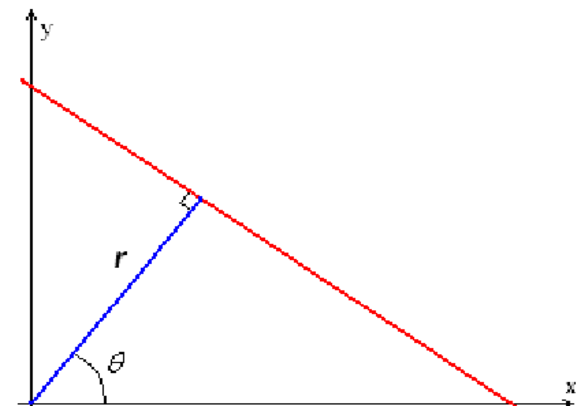
To verify whether the data obtained by simulation of random background could not contain random fake patterns whose can be mistaken as real events, we have applied pattern recognition methods for signal tracks.
The signal track on the FS contains information about the observed air shower and consequently about the primary EECR particle itself. It is a distribution of counts in space and time. There are possible several algorithms for the pattern recognition.

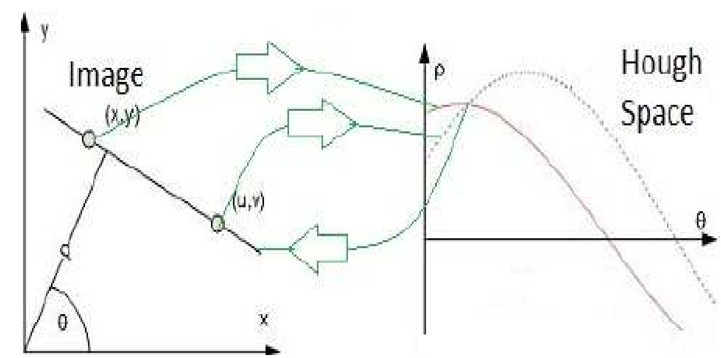
Hough transform

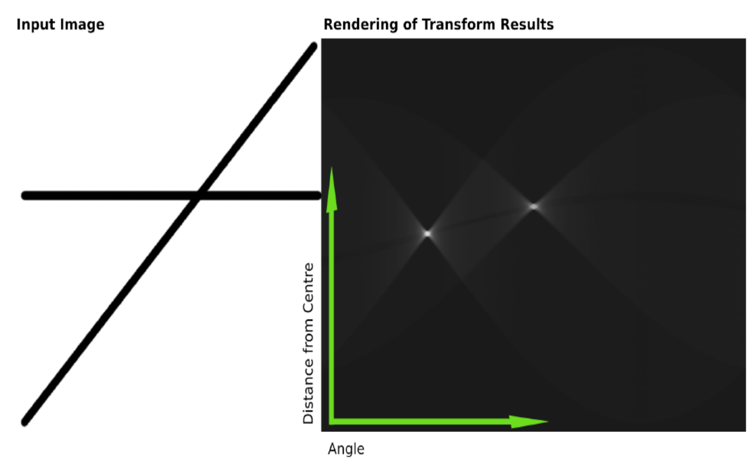
Developed to identify prefixed shapes within noise by transforming the relevant parameters to H space and back.

For each data point the HT assumes a number of lines passing through it. These lines can be parametrized by their distance from the origin of the coordinate system **r** and the angle **θ** between its normal and the x-axis.

Transformed into the HS, a 2dim parameter space spanned by **r** and **θ** each data point represents a sinusoidal curve. The intersection points of the many sinusoidals are summed up in accumulator. The intersection point that drews in most of the counts is then transformed back into the image space, where it corresponds to a straight line passing through as many data points as possible.







Method development

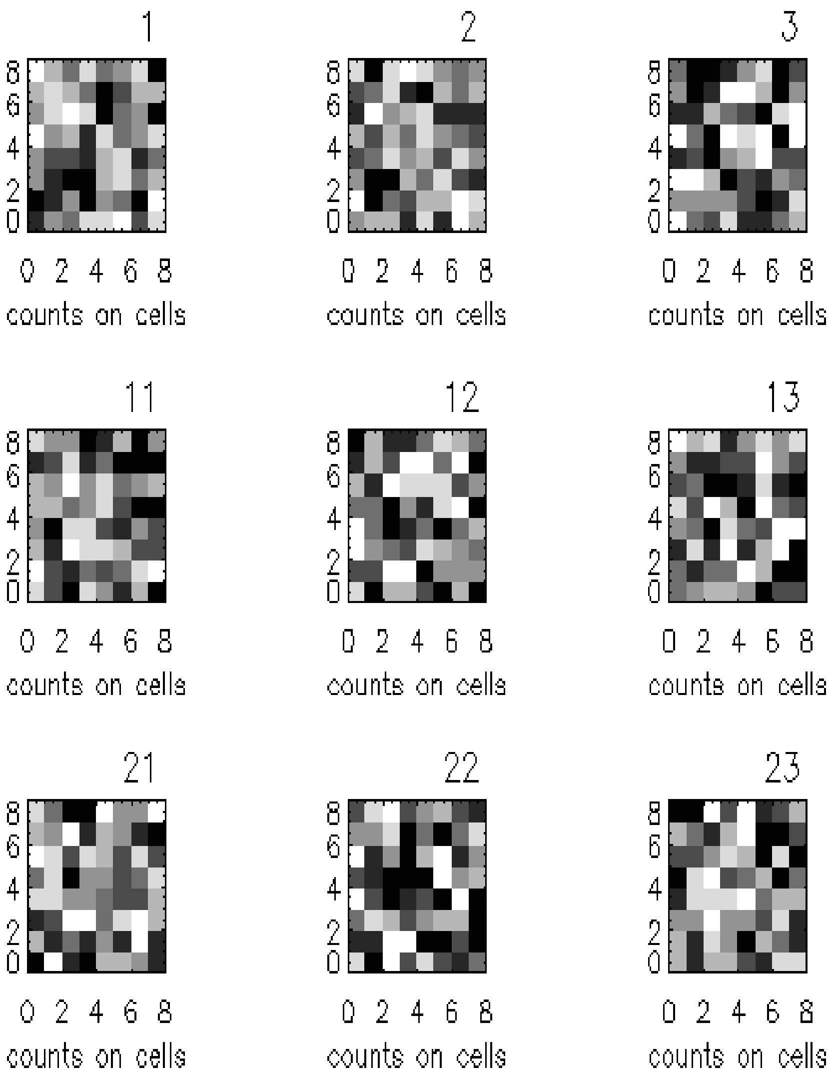
- **HT for purely uniformly distributed random values**

- **Modified HT for static patterns**

- **Modified HT for moving patterns (combined matrix method)**

- **Applied to poissonly distributed bckg filtered by trigger algorithms**

HT on uniformly distributed random values



Simulated results described analyzed by **HT** and consequently by **modified HT**

First step:
Method developed nad checked on
purely uniformly distributed random values

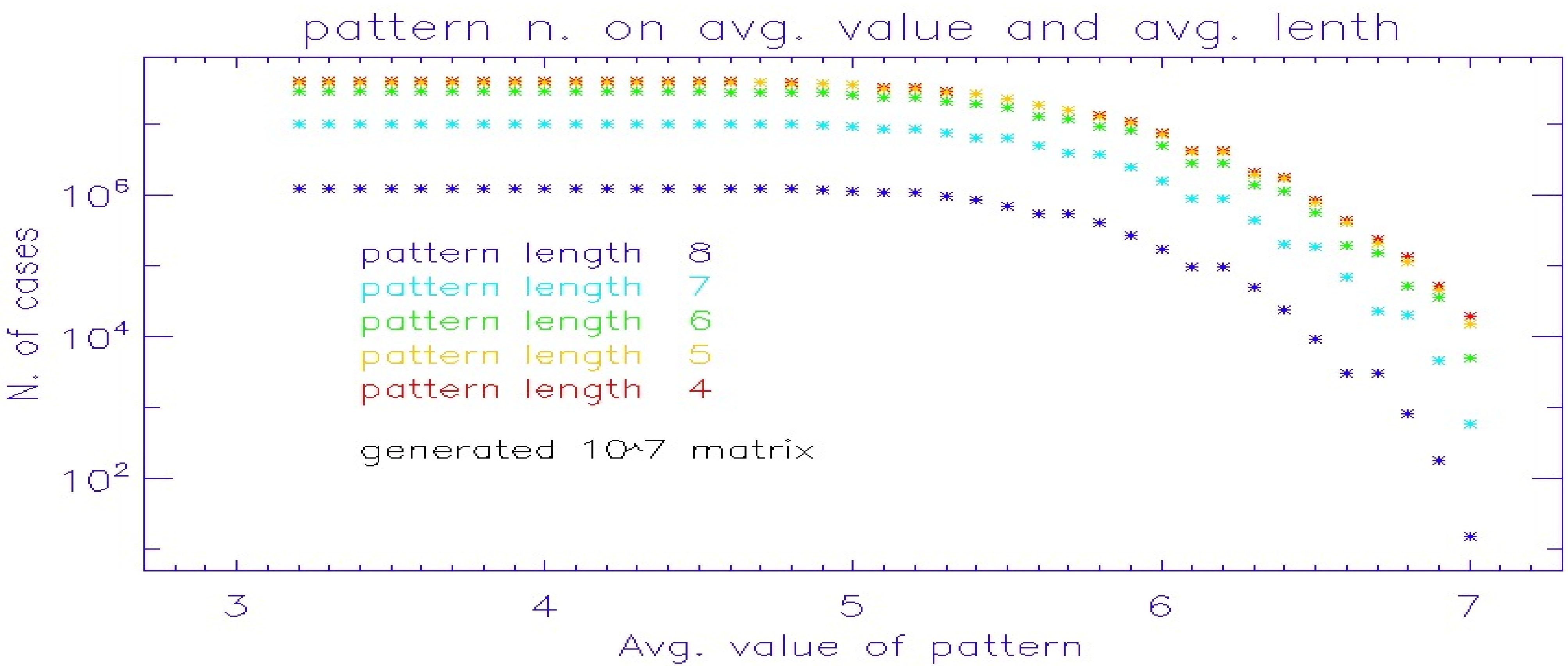
10^7 of matrices 8 x 8 (like PMT) were generated
Two pattern characteristics are of interest:

- $pattern\ length = No.\ of\ pixels\ over\ threshold$
- $avg\ pattern\ value = \sum\ pixel\ values / pattern\ length$

Testing the mehod

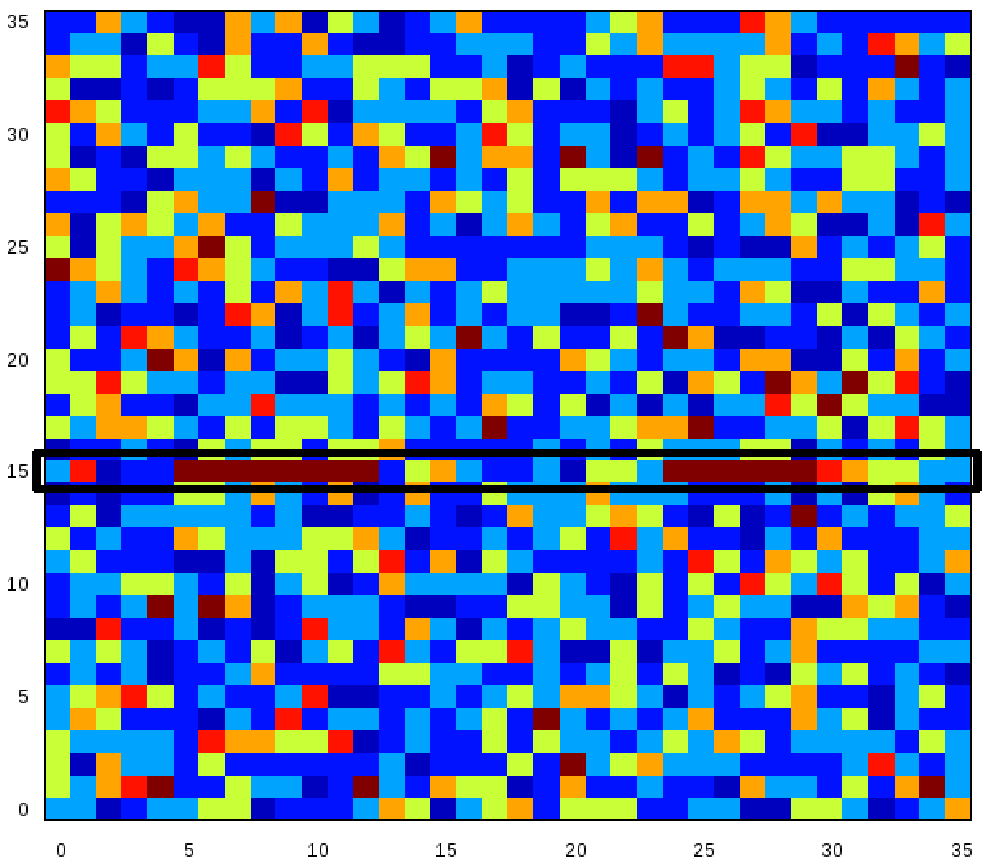
Tested by putting by hand small amount of patterns to huge amount of generated background. The method reliably detected artificial patterns.
Shown the number of detected patterns dependence over selected average pattern value for several pattern lengths (4 - 8).

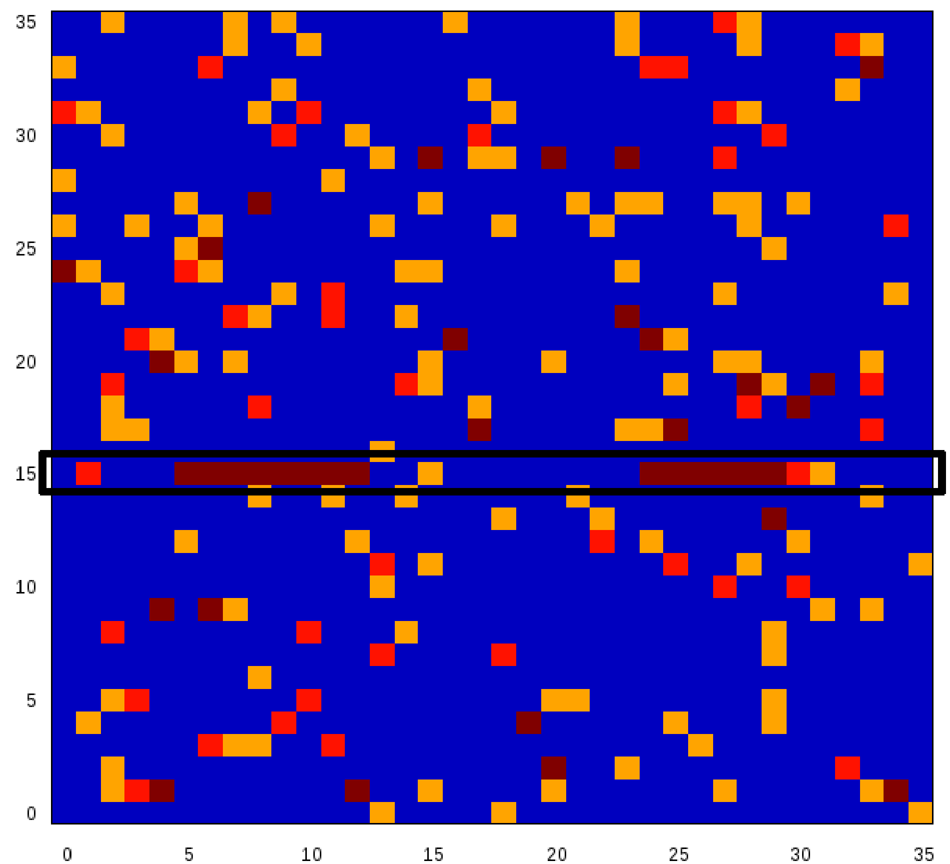
for 10^7 of matrices 8 x 8 matrices, around 20 matrices with fake pattern with the length of 8 pixels with average pixel value (all pixels at maximum) is found. The probabillity that matrix pixel has some value is 1/8. Any 8 pixel configuration, so lineal pattern 8 pixel long, too appears with a probability (1/8)^{8} = 5.96 x 10^{-8}. Such lineal patterns are 32, then the result is 19.07, compatible with the simulation result 20.



Modified HT

However, classic HT cannot distinguish between continuos and disconnected patterns. Thus the number of recognized patterns is overestimated. It was needed to improve the algorithm for the JEM-EUSO purpose to be able to differ between such patterns. It was done on the basis of pixel distance.

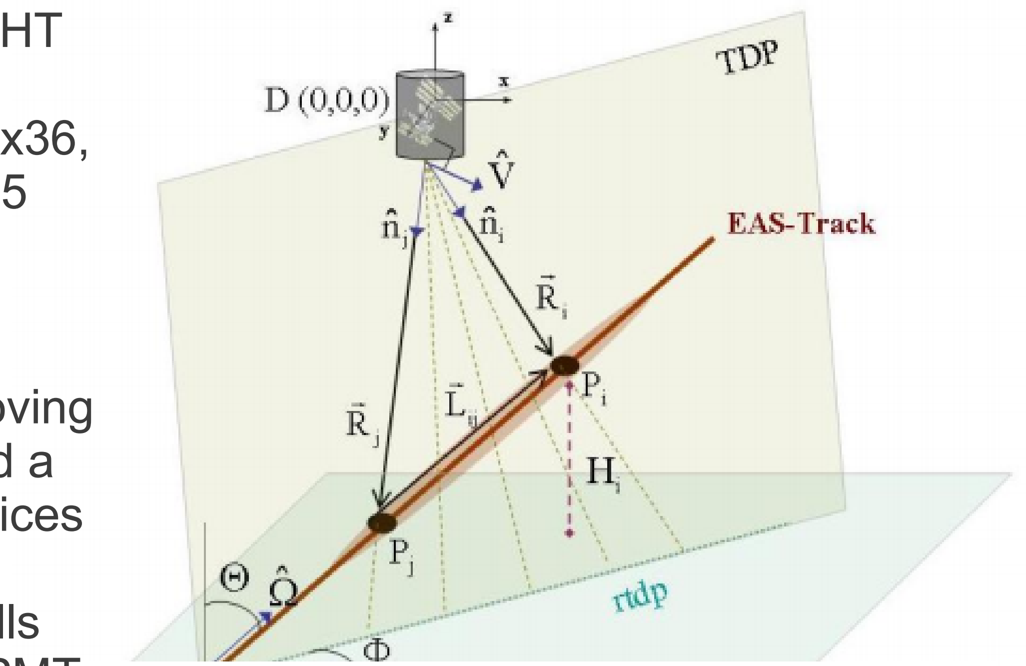


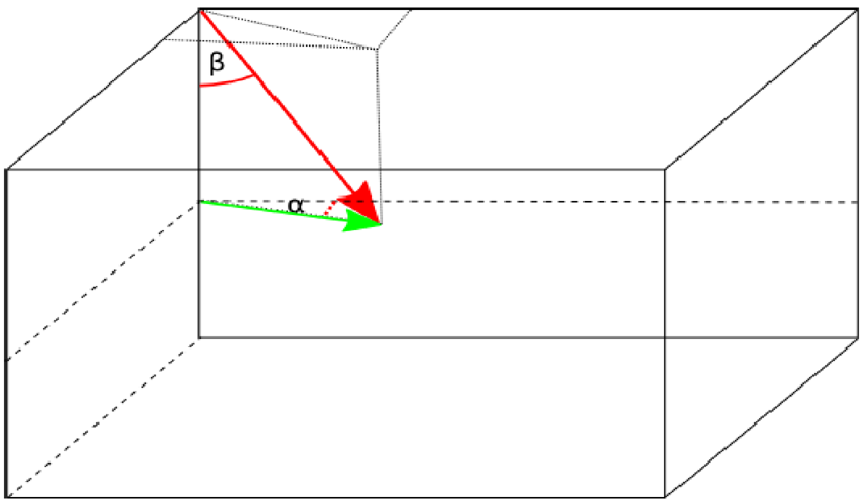


Modified HT for moving patterns

In the next step we have tested the modified HT algorithm on the obtained LTT triggers
For each LTT trigger we have 31 matrices 36x36, the actual snapshot and for 15 anterior and 15 posterior in GTU.

The real shower appears as a light speed moving point. On the basis of this we have developed a strategy of folding for above mentioned matrices to recognize the pattern created by moving point. We have divided atmosphere to the cells equivalent to pixel projection of JEM-EUSO PMT pixels on Earth surface. We evaluate a projection of moving light point created by shower on Earth and time when pixel is observed in GTU for a set of zenith and axial angles of incoming particle.

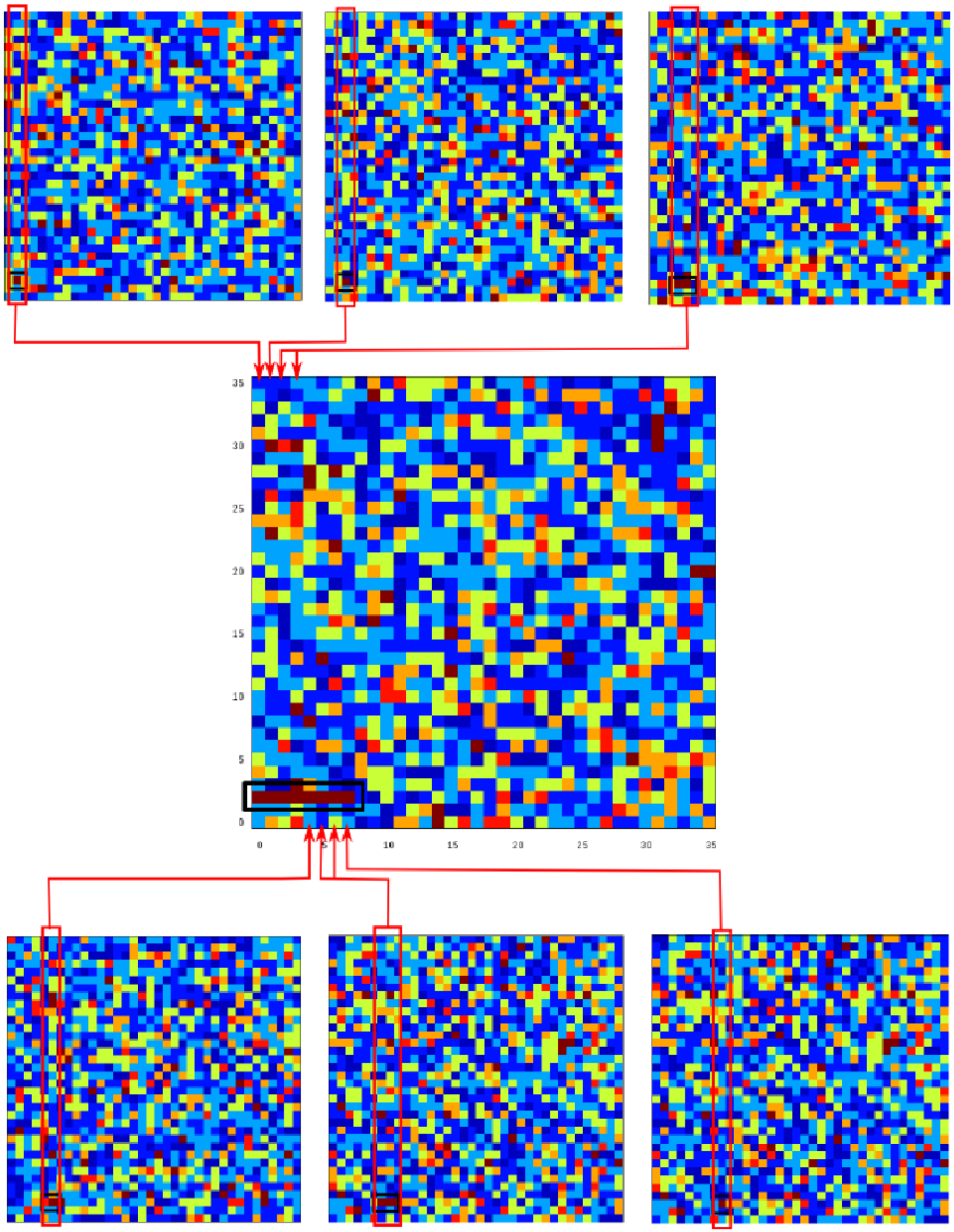




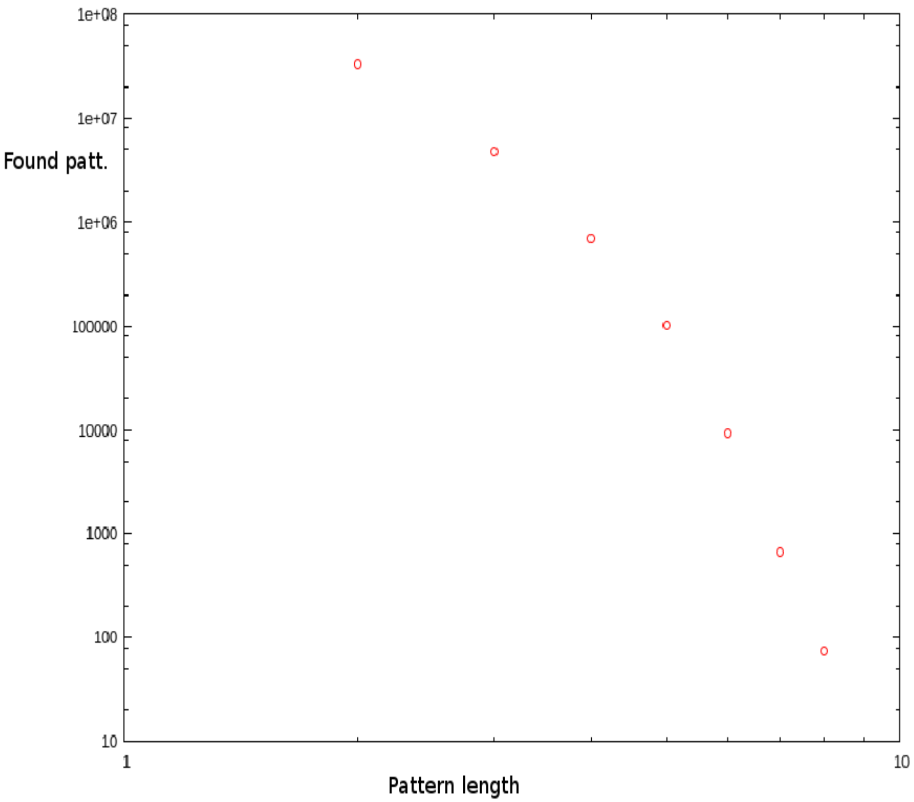
HT for moving patterns - implementation

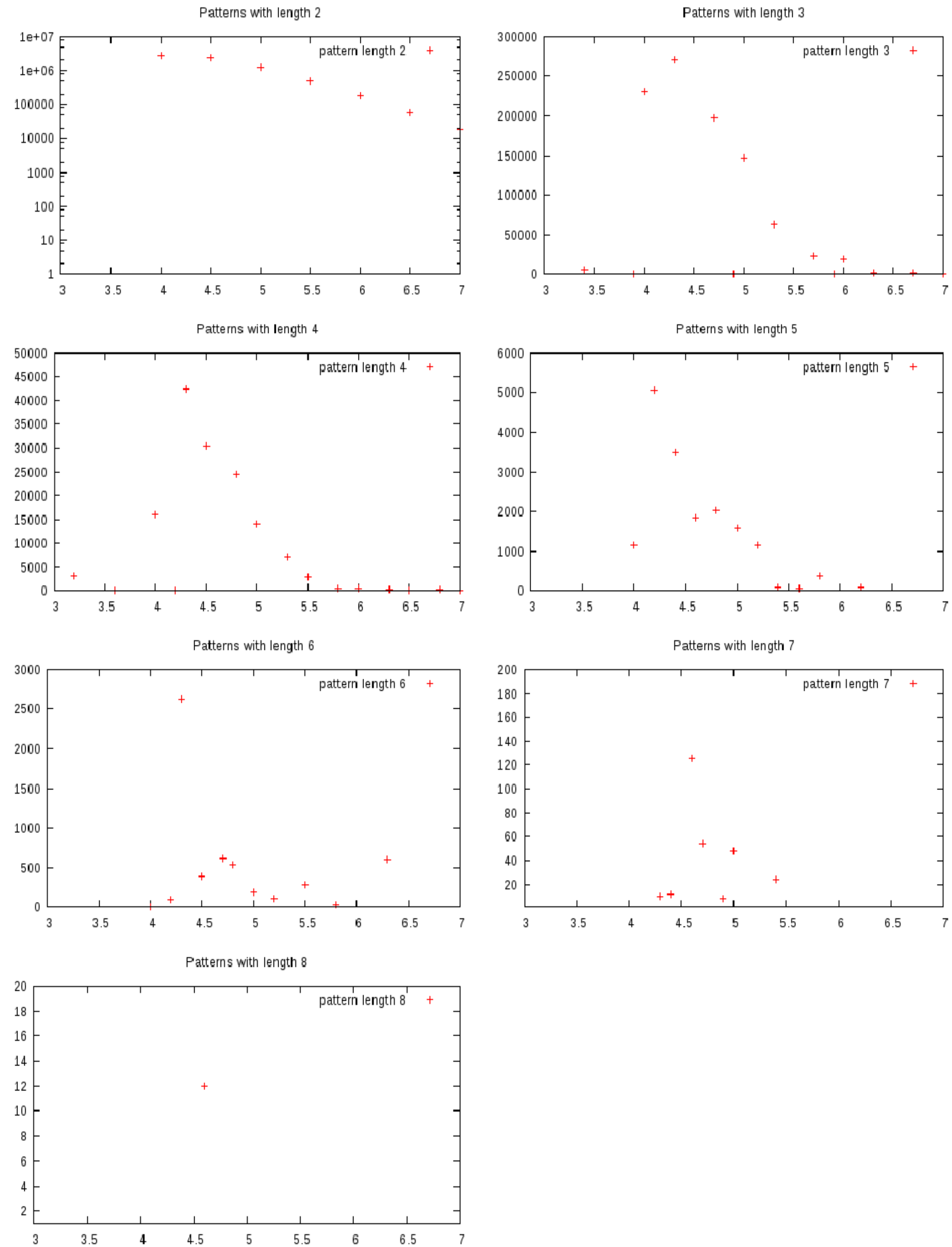
Every direction of incoming EECR particle is equivalent to a set of projections in consequetive GTUs. Then for one incoming direction we can take only columns where moving light point is visible. We combine a new matrix from stored 31 matrices from selected columns.

Pattern recognition method is then applied to this new matrix. We build over 31 stored matrices a set of new matrices for selected incoming angles of primary cosmic rays. Such an analysis is applied to all simulated sets of 31 matrices passed LTT trigger. The method validity was verified by artificial patterns with known incoming direction added to tested data set. All artificial patterns were found by the method.



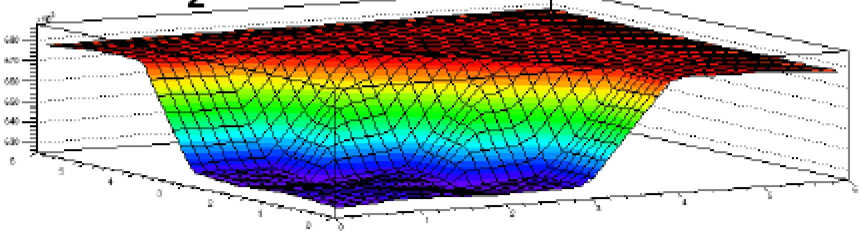
HT for moving patterns – results



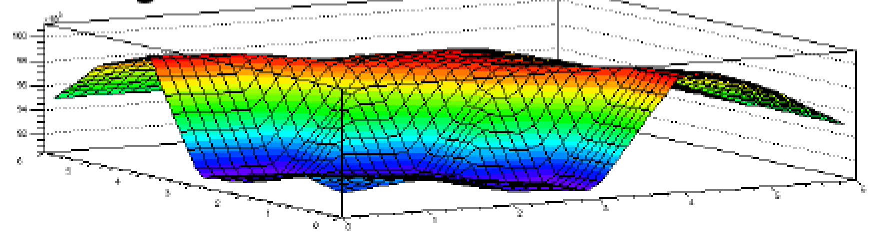


HT for moving patterns – results

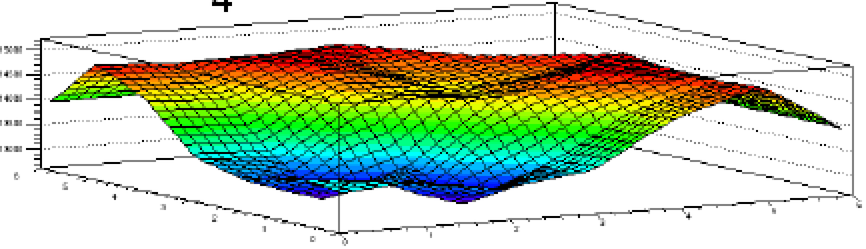
2



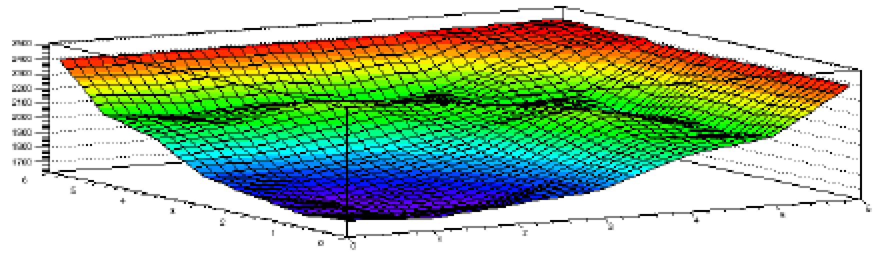
3



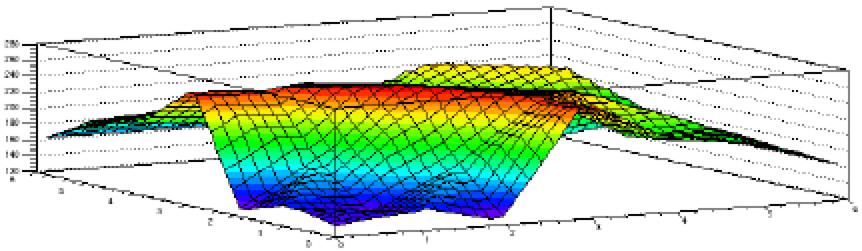
4



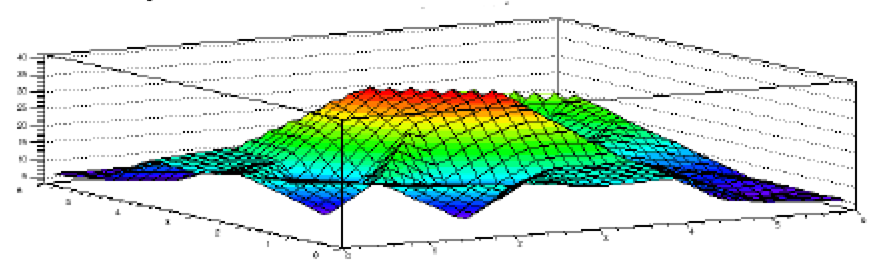
5



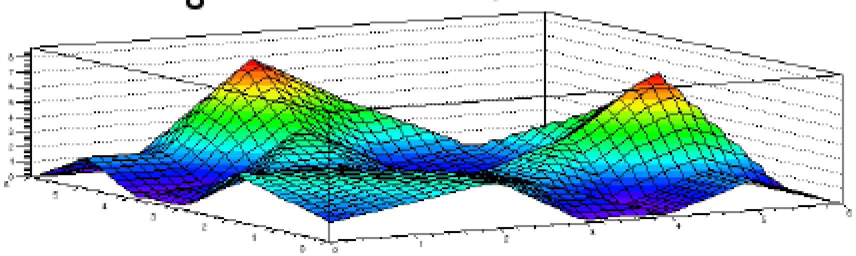
6

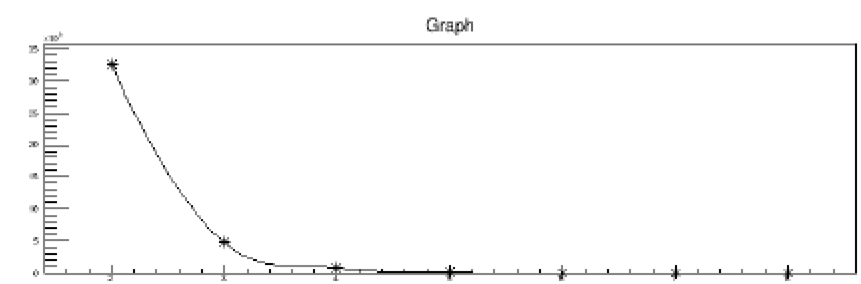


7



8



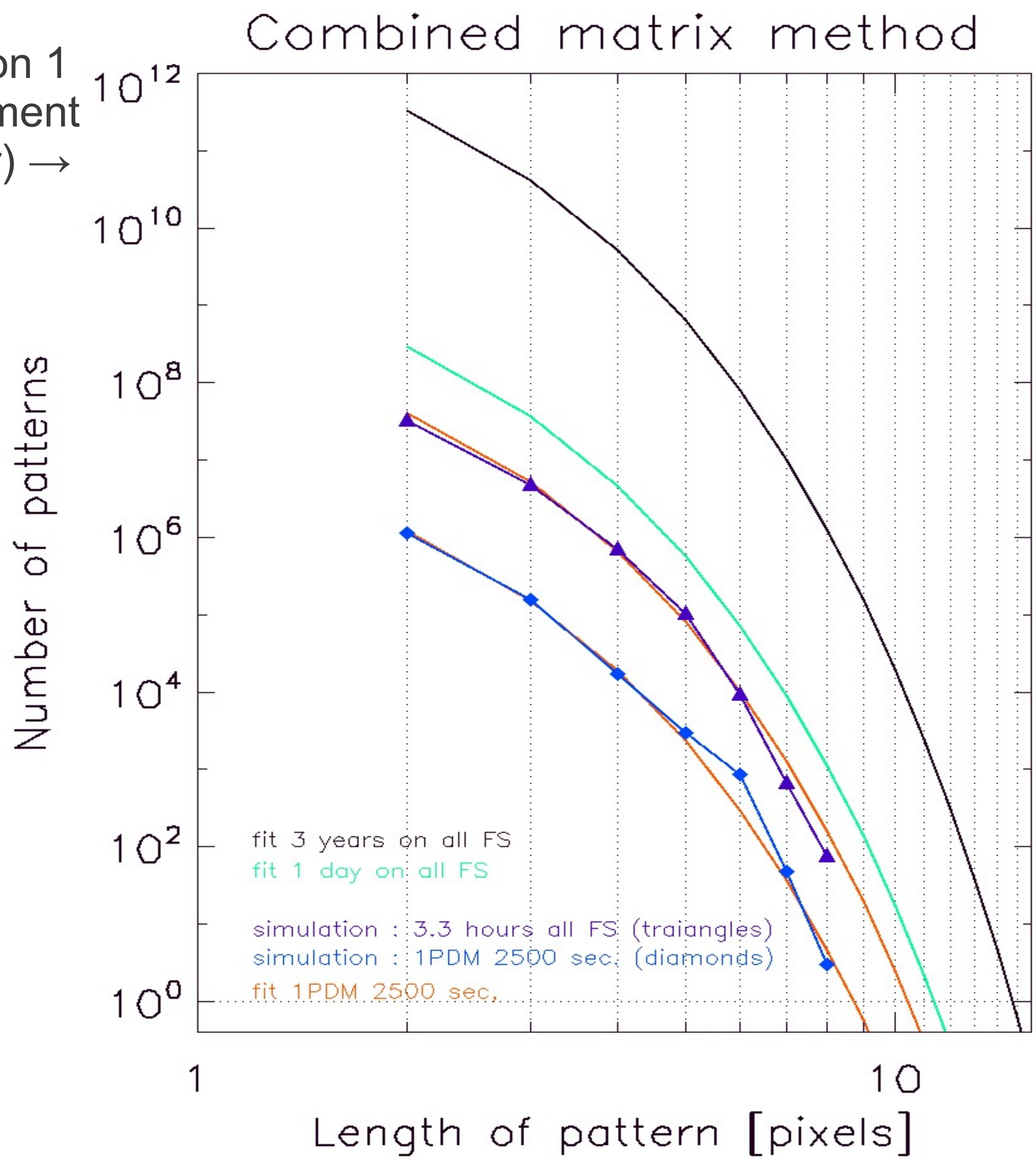


Combined matrix method result

Finally go through all simulated GTU's on 1 PDM (equivalent to 3.3 hours measurement of all 137 PDMs of JEM-EUSO detector) → magenta line with triangles. We fit by statistically motivated function:

$$N_p \,.\, (L_p) \sim \left(\frac{1}{N_{px}}\right)^{L_p}$$

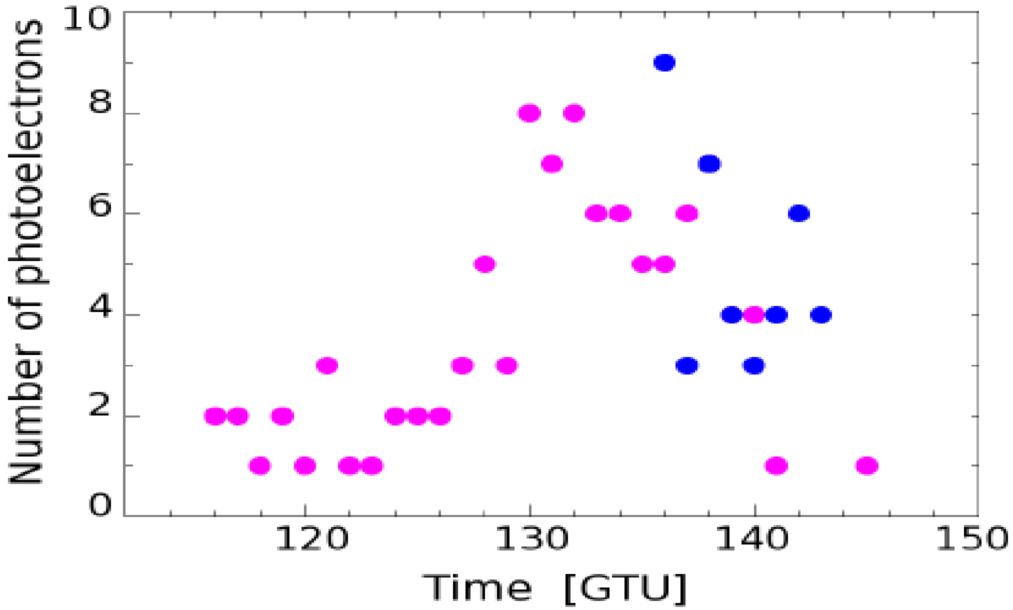
scalling to full planned 3 years of JEM-EUSO operation, we will find only one pattern with the length of 15 pixels. This 15 pixels on ground means 7.65 km long projection of shower. Showers created by more inclined and higher energetic particles are more easy to recognize and reconstruct. Worst case when we will have particle with energy 5 10^{19} eV and maximum zenith angle can create first pixel visible by detector at altitude 13 km. If fake pattern will be 7.65 km long with first visible point at 13 km, then zenith angle of primary particle is 30.5 degrees. Thus the fake pattern during 3 years of measurement can be mistaken by particle with zenith angle maximaly 30.5 degrees.

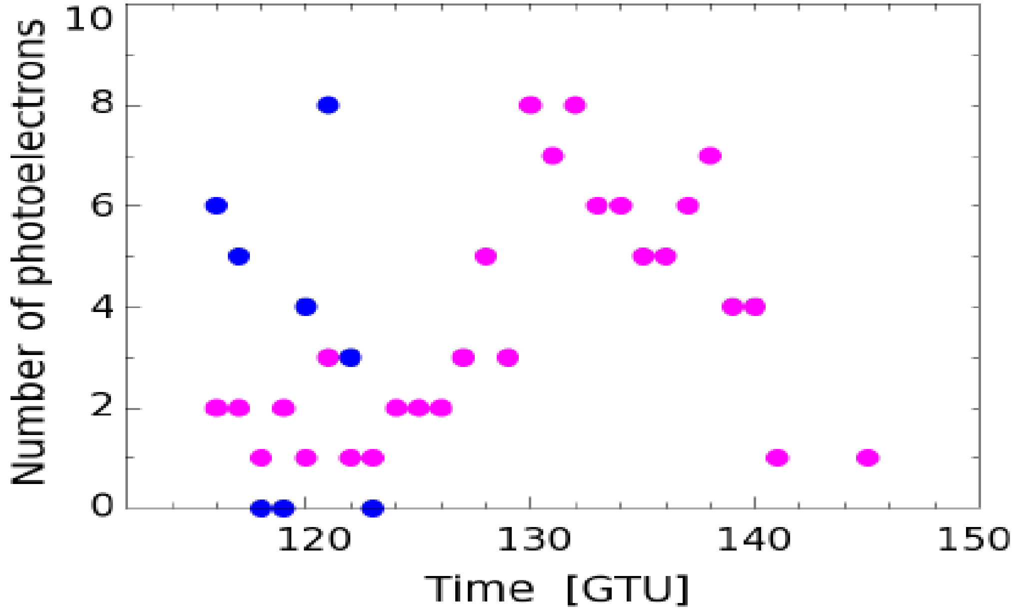


Shower development time dependence

Evaluate a sum of intensity around moving pixel in longest recognized patterns. Sum of 4 pixels - one from pattern and 3 from surrounding pixels (2x2 matrix) - in a triggered matrix was evaluated for every GTU when pattern was found.
Pattern intensity time evolution compared with a simulated showers created by primary particle with energy 5 x10^{19} eV.
While simulated shower evolution has a typical time profile with increasing intensity to maximum point H_{max} and then decreasing, the recognized patterns has different noisy shape and shorter length.

conclude that during 3 years measurement we cannot mistake
fake trigger to real event.





New analysis start

- after results from Hough algorithm presented at ICRC
- new diploma student Erik Gajdos
  - same analysis by second pattern recognition method (clustering)
  - verification of results from Hough method
  - later, aplication to M64 case and algorithm optimalisation
  - first results in Tokio

Summary

- **Modified HT method for PR of fake trigger bckg events developed**

- **Applied to M36 simulation results**

- **Adapted for M64, however fix the problem with rates for M64 needed**

- **Development of clustering method for verification started**