

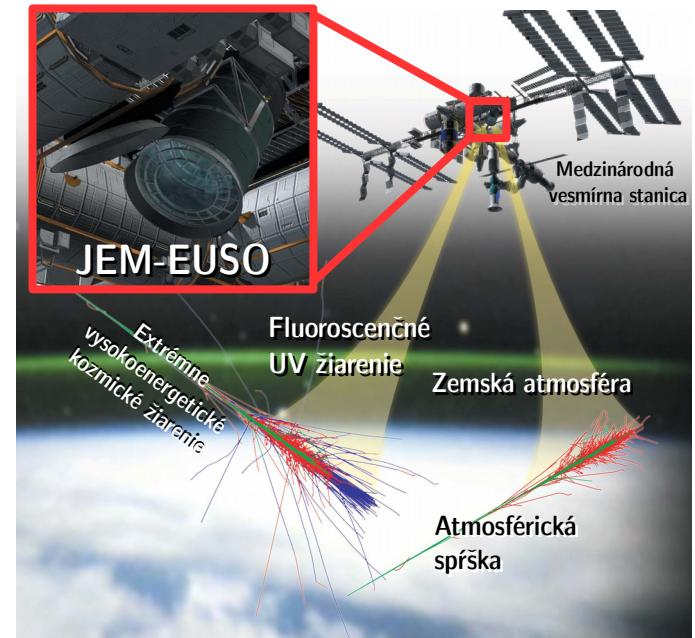
Hľadanie atmosférických spŕšok v dátach experimentu EUSO-SPB

Michal Vrábel
Katedra počítačov a informatiky, FEI TUKE

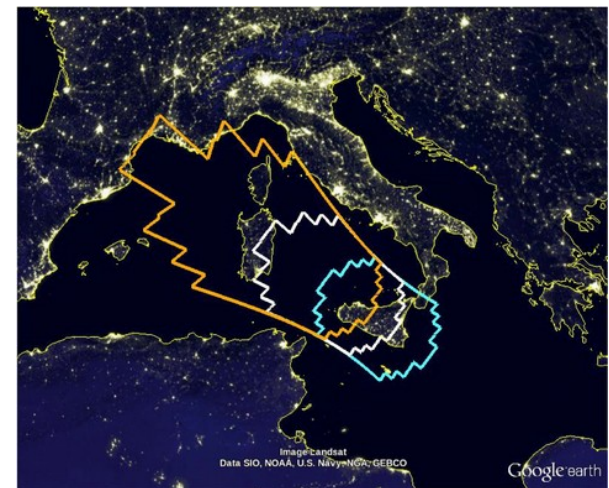
Doktorandský seminár ÚEF SAV
27. jún 2018

JEM-EUSO experiment

- Cieľ: Identifikácia zdrojov vysoko-energetických častíc kozmického žiarenia
- Kolaborácia zahŕňa 16 krajín a skoro 90 inštitúcií (TUKE člen od roku 2015)
- Pozorovanie atmosferických spŕšok z obežnej dráhy pomocou fluorescenčného detektora
- Detektor
 - pozoruje oblasť s polomerom 380 km, $4.5 \times 10^5 \text{ km}^2$ (ISS orbit, nadir), 1 bod odpovedá $0.56 \times 0.56 \text{ km}^2$
 - Priestorové rozlíšenie 320 000 obrazových bodov
 - Temporálne rozlíšenie $2.5 \mu\text{s} = 1 \text{ Gate Time Unit (GTU)}$.
 - Intenzita obrazového bodu počet pulzov produkovaných fotomultiplikátorom za 1 GTU.



Obr. 1: Princíp JEM-EUSO experimentu.



Obr. 2: Plocha pozorovaná JEM-EUSO detektorom (side-cut verzia).

EUSO-SPB

- štart 24. apríla 2017
- predpokladaná dĺžka letu 50 dní, pokus o 100 dňový rekordný let
- pád balónu po 12 dňoch
- detektor fungoval až do pádu do oceánu dobre
- prebieha analýza dát

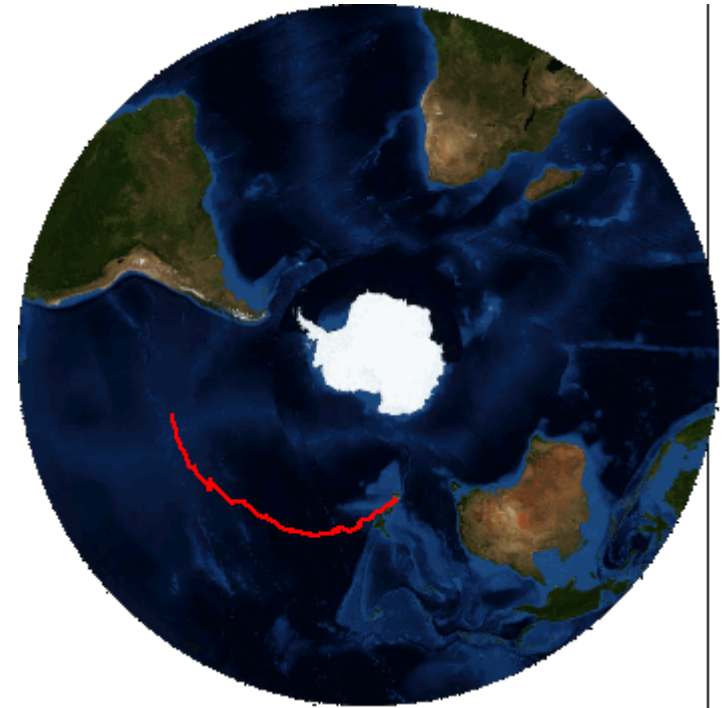


Obr. 3: EUSO-SPB.

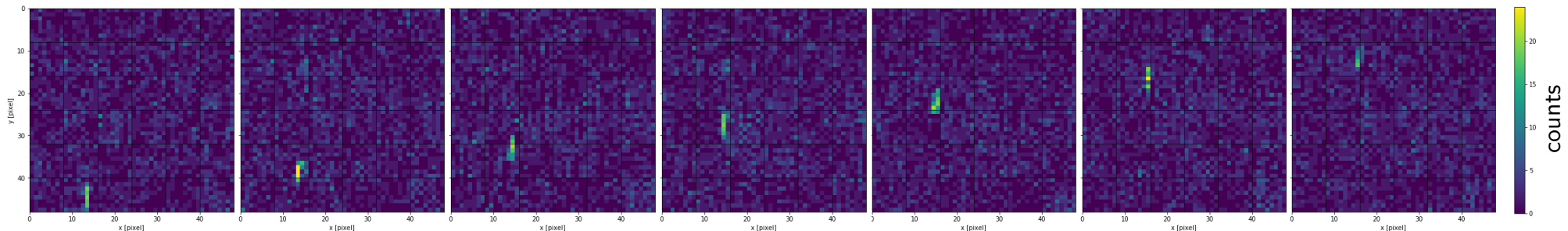


Hľadanie spršok v EUSO-SPB dátach

- Dátach z letu – prvotný cieľ bol redukovať množinu na zaujímavé udalosti
- Dáta z pozemných testov detektora v Utahu – spršky simulované laserom
- Simulované dáta s reálnym pozadím – využité pre vyhodnocovanie metód
- Čistý signál – kontrola metód, príprava „označenej množiny“

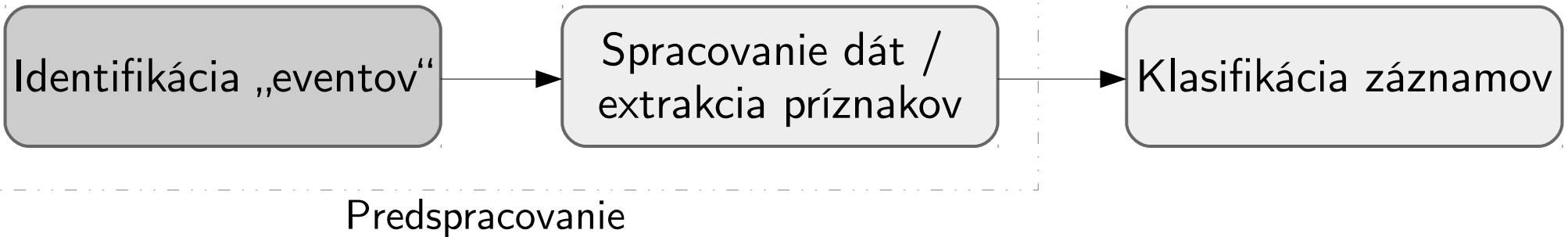


Obr. 4: Dráha letu EUSO-SPB.

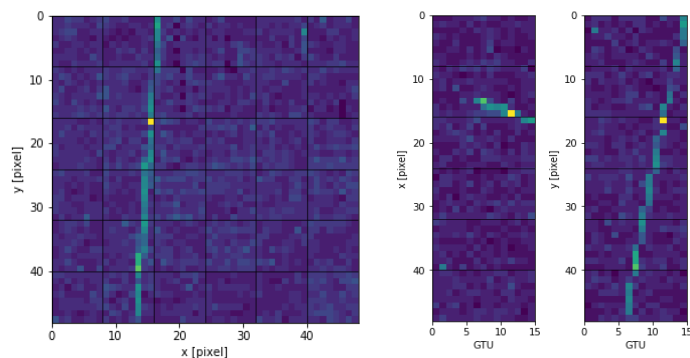


Obr. 5: Príklad simulovanej atmosferickej spršky pozorovanej detektorom EUSO-SPB.

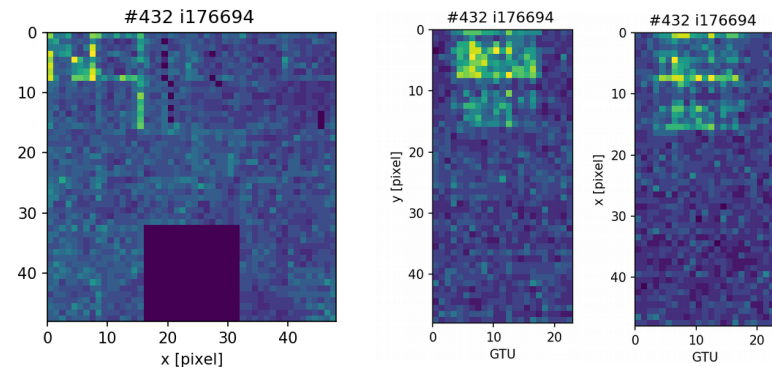
Metóda analýzy dát



- Identifikácia udalostí („eventov“) na základe kontinuálnej prítomnosti signálu spúšťača
- Samotný signál spúšťača automaticky neznamená, že bola pozorovaná sprška
- Spúšťač (L1 trigger) produkuje signál ak v skupine pixelov pretrváva signál dlhšie ako stanovený limit

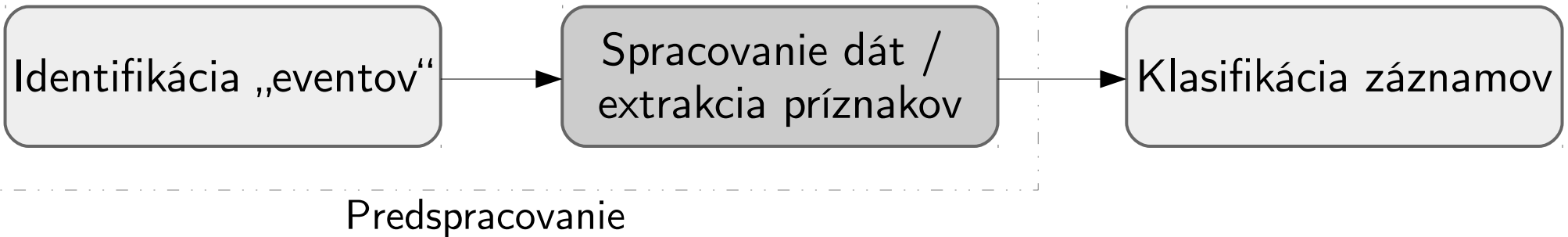


Obr. 6: Príklad simulovanej spršky.

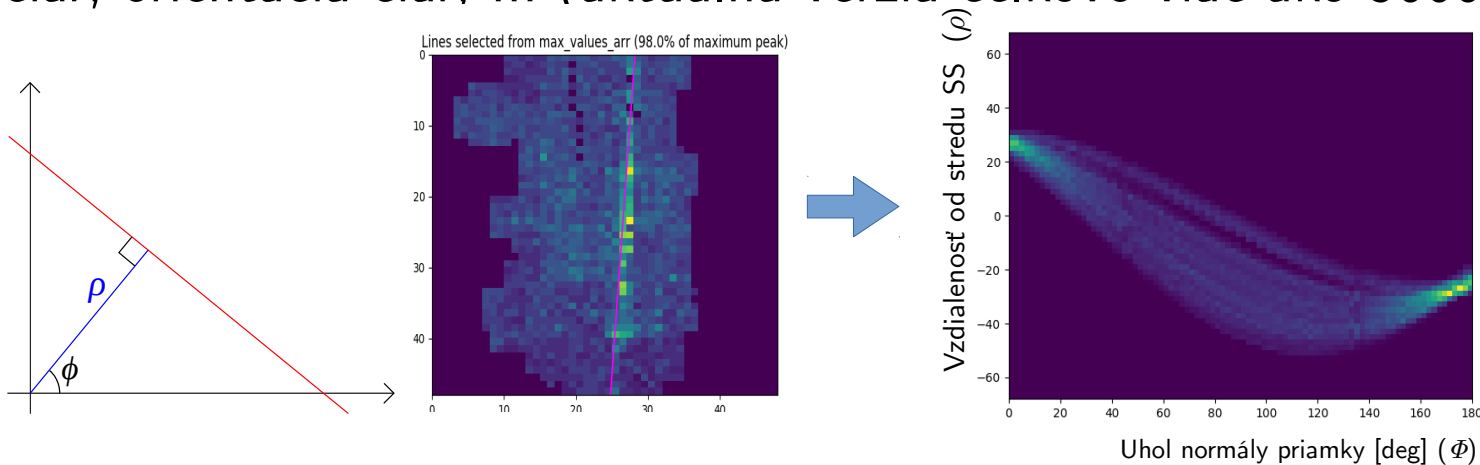


Obr. 7: Príklad udalosti v dátach z letu

Metóda analýzy dát

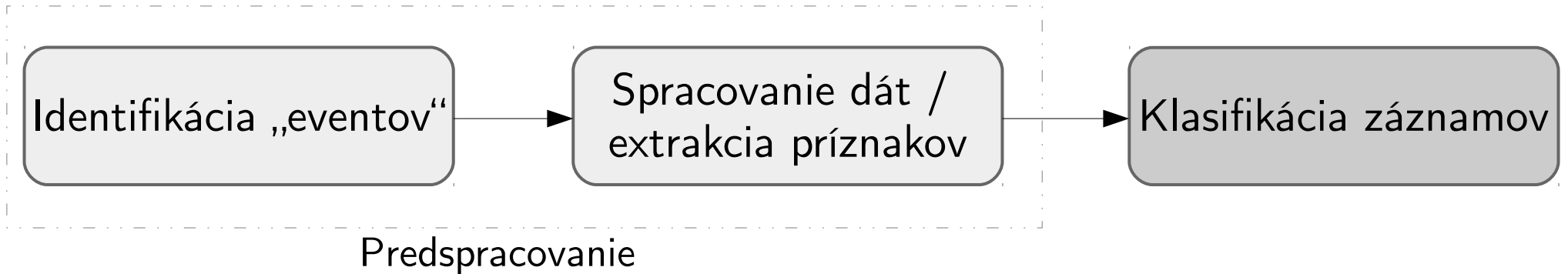


- Spříška – zjednodušené čiara
- Hľadanie čiar využitím Houghovej transformácie – hypotéza: veľkosť klastorov v Houghovom priestore opisuje pravdepodobnosť existencie spříšky (čiary) v dátach umožňujúc jednoduchú klasifikáciu
- Ďalšie extrahované parametre – veľkosť a počet klastrov, priemerné intenzity, pozície čiar, orientácia čiar, ... (aktuálna verzia celkovo viac ako 5000)



Obr. 8: Houghová transformácia - príklad.

Metóda analýzy dát



- Klasifikácia – manuálna (SQL požiadavky) alebo využitím strojového učenia
- Predchádzajúca verzia algoritmu s manuálnym / intuitívnym postupom bola $\approx 75\%$ efektívna (detekovateľné simulované spŕšky)
- Aktuálne prebieha práca na klasifikácii využitím strojového učenia (MLP, Boosted Decsion Tree, ...), práca na dobrej trénovacej množine

Príklad manuálneho výberu

Obmedzená dĺžka udalosti

Obmedzený počet pixelov aktivovujúcich spúšťač

Aspoň jedna z projekcií GTU-X a GTU-Y musí obsahovať naklonenú čiaru

Obmedzená šírka oblasti okolo vrchola v Houghovom priestore (rozsah náklonu čiar) a obmedzenie počtu vrcholov.

Rôzne podmienky pre pixely aktivujúce spúšťač a všetky pixely.

Obmedenie počtu vrcholov / klastrov v Houghovom priestore.

```
num_gtu BETWEEN 14 AND 40
AND num_triggered_pixels BETWEEN 10 AND 150

AND abs(gtu_in_packet - 40) <= 10

AND (abs(gtu_y_hough__peak_thr2_avg_phi) > 0.084 OR abs(gtu_x_hough__peak_thr2_avg_phi) > 0.084)
AND (abs(gtu_y_hough__peak_thr3_avg_phi) NOT BETWEEN 1.48353 AND 1.65806 OR abs(gtu_x_hough__peak_thr3_avg_phi) NOT BETWEEN 1.48353 AND 1.65806)

AND (
  (
    ((trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 50 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 4) OR
    ((trigg_gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 45) OR
    (trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 35 AND ((trigg_gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 50 AND trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 35 AND trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 35 AND trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (trigg_gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (trigg_gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 4)
  )
)

AND (
  (
    ((gtu_y_hough__dbscan_num_clusters_above_thr1 = 1 AND gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 40) OR
    (gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 35 AND ((gtu_x_hough__dbscan_num_clusters_above_thr1 < 4) OR
    (gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 35 AND gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 50)) OR
    (gtu_x_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND gtu_y_hough__peak_thr1_max_cluster_counts_sum_width <= 15 AND x_y_hough__peak_thr1_max_cluster_counts_sum_width <= 20)
  )
)

AND (
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 4 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 6) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 4 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 5 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 5 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 5 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 4) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 2 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 2) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr1 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 2 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr1 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr1 < 2)
)

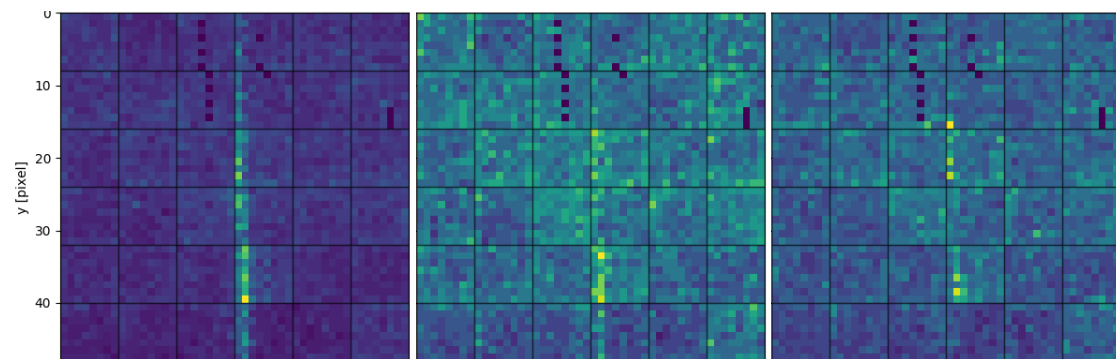
AND gtu_y_hough__dbscan_num_clusters_above_thr1 < 4 AND gtu_x_hough__dbscan_num_clusters_above_thr1 < 4 AND x_y_hough__dbscan_num_clusters_above_thr1 < 4

AND (
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr2 < 3 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 3 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 3) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr2 < 3 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 3 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 3 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 3) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr2 < 5 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 5 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 5) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 2) OR
  (trigg_gtu_y_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_gtu_x_hough__dbscan_num_clusters_above_thr2 < 2 AND trigg_x_y_hough__dbscan_num_clusters_above_thr2 < 2)
)

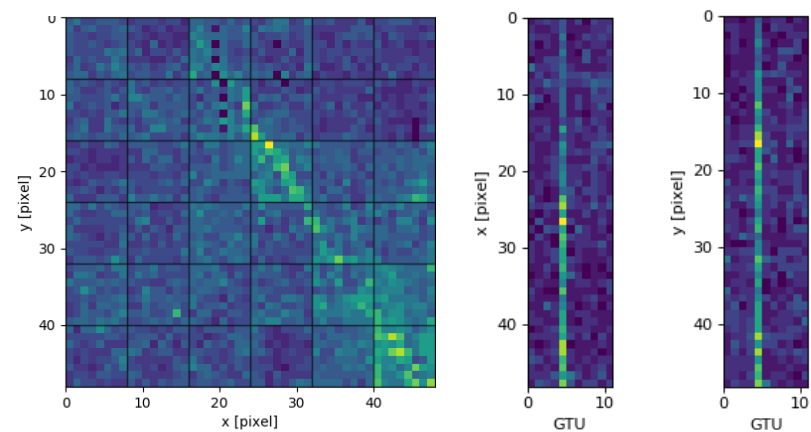
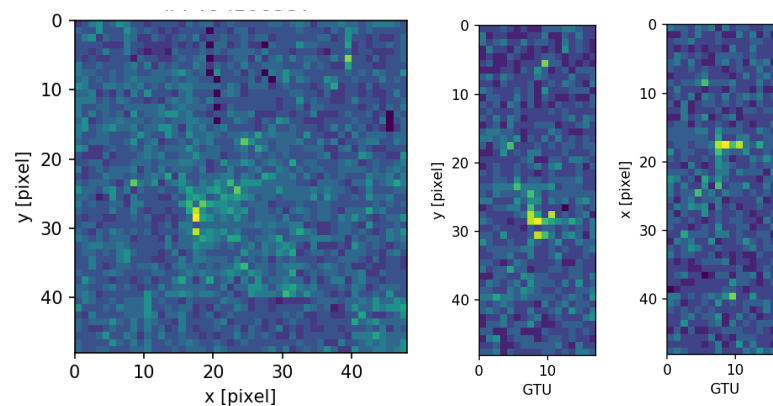
AND (
  (gtu_y_hough__dbscan_num_clusters_above_thr2 < 3 AND gtu_x_hough__dbscan_num_clusters_above_thr2 < 3 AND x_y_hough__dbscan_num_clusters_above_thr2 < 5) OR
  (gtu_y_hough__dbscan_num_clusters_above_thr2 < 3 AND gtu_x_hough__dbscan_num_clusters_above_thr2 < 4 AND x_y_hough__dbscan_num_clusters_above_thr2 < 3) OR
  (gtu_y_hough__dbscan_num_clusters_above_thr2 < 4 AND gtu_x_hough__dbscan_num_clusters_above_thr2 < 3 AND x_y_hough__dbscan_num_clusters_above_thr2 < 3)
)
)
```

(precise values and parameters depend on used method)

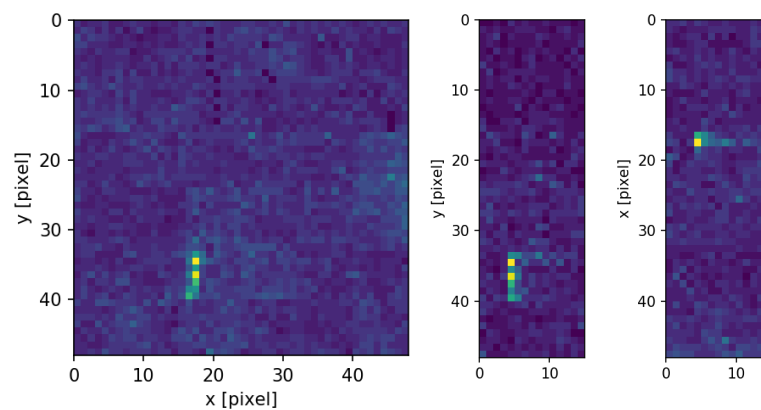
Príklady analyzovaných dát



Obr. 9: Rozpoznané spršky na detektore simulované laserom (intenzita lasera sa znižuje zľava)



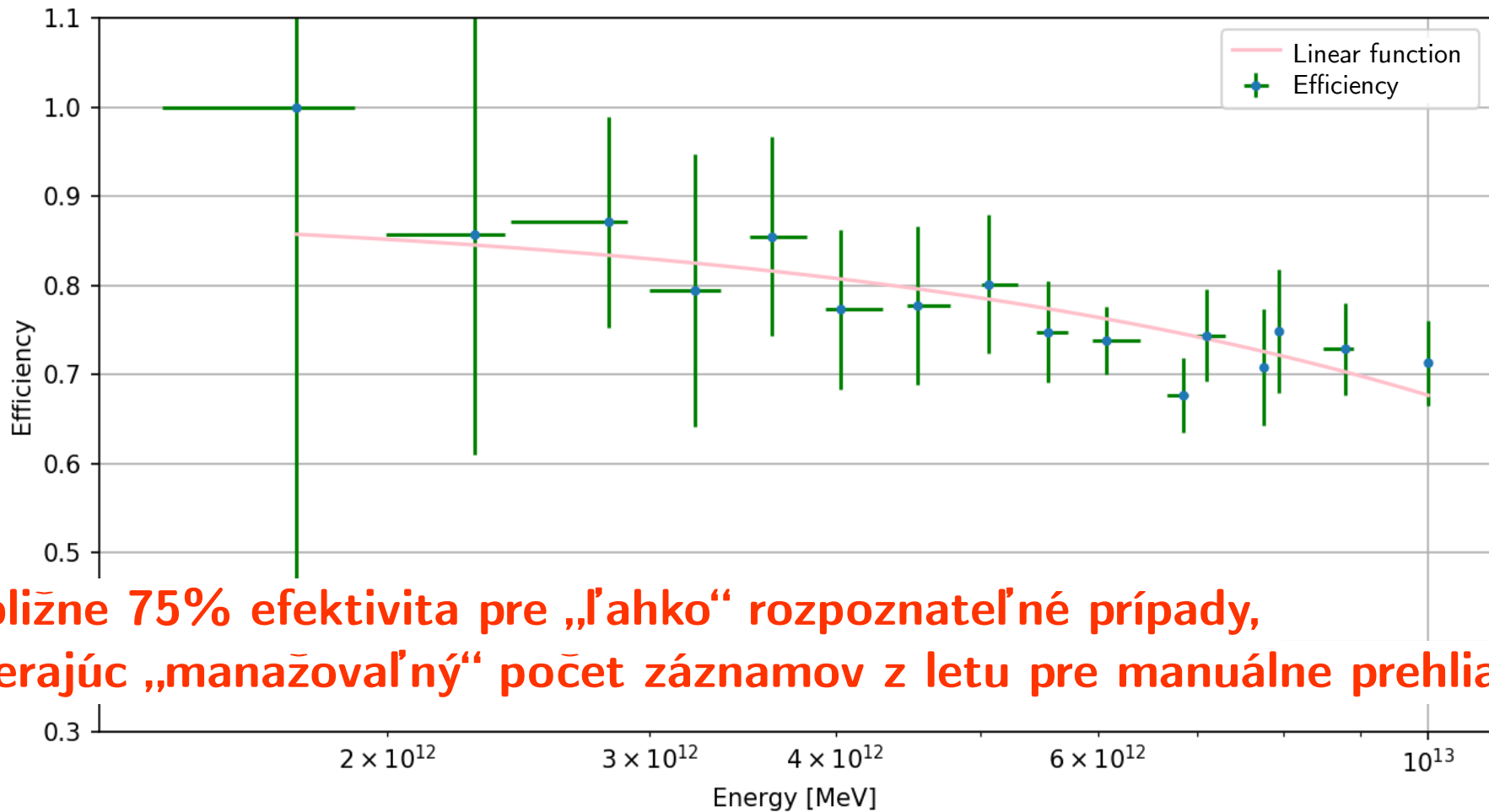
Obr. 10: „Falošná sprška“ zapríčinená nízko-energetickou časticou prelietajúcou cez detektor (správne rozpoznaná algoritmom)



Obr. 11: „Falošné pozitívne“ klasifikácie

Efektivita postupu

Podiel počtu vybraných spŕšok ku všetkým spŕškam, ktoré by aktivovali spúšťač a boli mimo “problematickej” časti detektora.



Približne 75% efektivita pre „ľahko“ rozpoznateľné prípady, vyberajúc „manažovateľný“ počet záznamov z letu pre manuálne prehliadnutie

Efficiency = **num. of filtered selected** / num. of all

(fit lineárnou funkciou je len pre ilustráciu)

Error X ... Range of energies for a datapoint

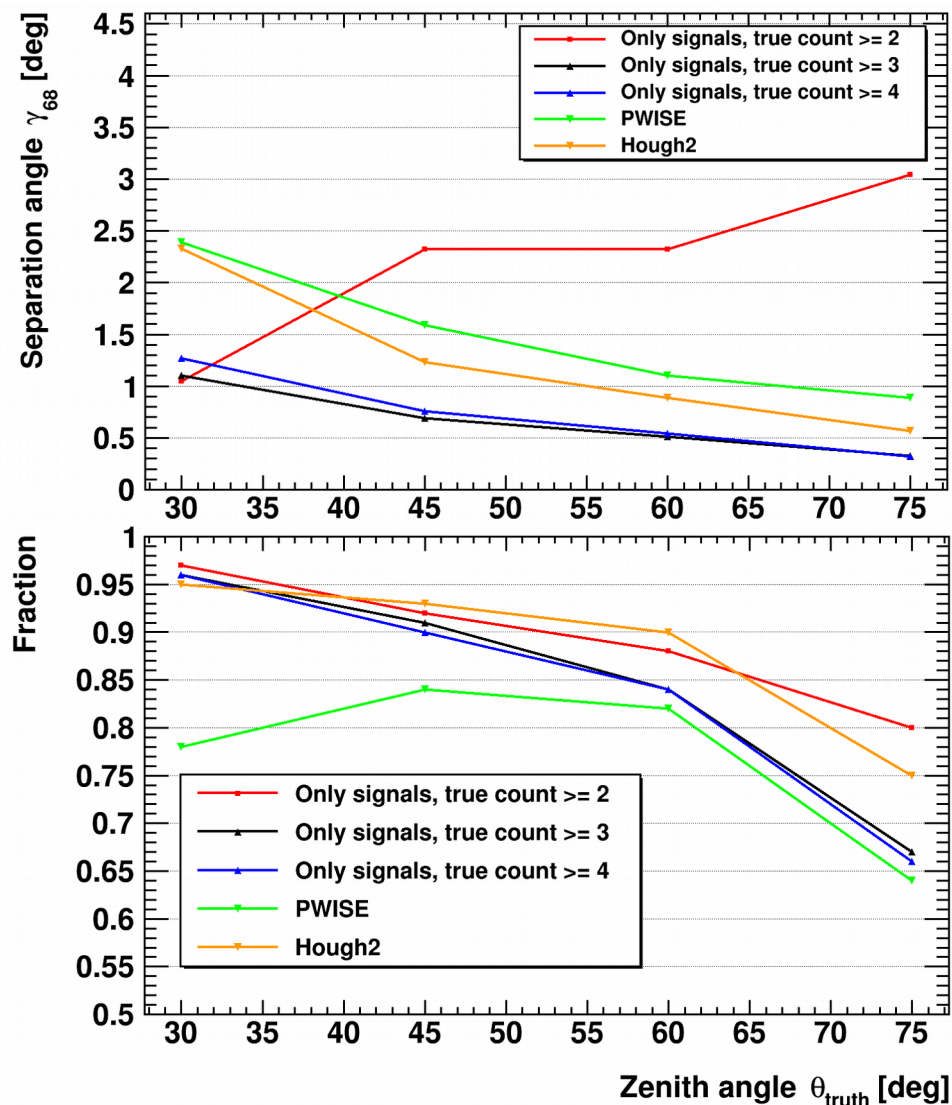
Error Y ... $\sqrt{((1-(n1/n2))/n1) + (1/n2)} * (n1/n2)$

Záver

- Metóda extrakcie príznakov umožňuje jednoduchú parametrizáciu záznamov
- Možná aj manuálna klasifikácia (na základe predpokladov sprške) na dátach, ktoré by mali byť rozpoznané detektorom
- Problémy vznikajú pri prípadoch „málo viditeľných a krátkych“ spršok
- Komplikovanejšie prípady vyžadujú strojové učenie (na príznakoch)
- **Budúcnosť pre klasifikáciu a rekonštrukciu – konvolučné neurónové siete, tréning priamo na pixeloch obrazu**

Metódy segmentácie spŕšky

- Vylepšenie presnosti rekonštrukcie UHECR skrz presnejší výber pixelov spŕšky.
- **Pokus o odstránenie limitácie presnosti rekonštrukcie UHECR spôsobenej segmentáciou pixelov.**
- Preverenie, že aktuálny proces rekonštrukcie neignoruje podstatné informácie využiteľné pri rekonštrukcii, ktoré zároveň nie sú stratené v šume.
- Návrh metód prispôsobujúcich sa meniacej sa intenzite UV pozadia.
- „Nová expozícia“ bude využitá pre vyhodnotenie algoritmov
- **Testovanie metód na dátach z EUSO-SPB experimentu.**



Obr. 7: Porovnanie rekonštrukcie uhla pre rôzne metódy segmentácie.