Algorithm for data analysis of CREDO smartphone application

Presenter: *K. Almeida Cheminant*

CREDO Week 2018

1st – 5th October 2018

The science behind the CREDO app

$\gamma_{\mathbf{UHE}}$
(e.g. 10^{20}eV)

Citizens browse the data looking for „improbable" **time-space coincidences**

1) $\mathbf{t_n - t_1 < \sim 1\ \mu s}$
2) $\mathbf{t_1 < \ldots < t_n}$**,**

→ **indirect search for New Physics manifestations!**
→ **verification of „classic" QED predictions (preshower @ Sun)**

The science behind the CREDO app

CREDO App: *Smartphone application* to turn cameras into particle detectors
→ available on *Androïd*

top 5 users

Login	Detections
kilo	86,167
Mafia75 7	56,188
mates	33,949
Bogdan51	31,588
Krzysztof	22,295

last 5 registered users

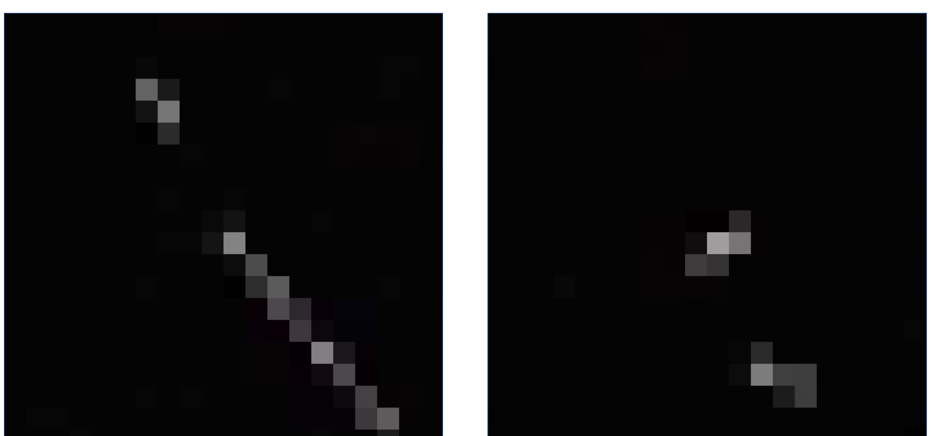
Login	Detections
filipfcb42	1
Grzegorz	0
kris	0
Hibiskus	226
prawdziwytomasz	0

Last 20 detections

sort : date by : descending Envoyer

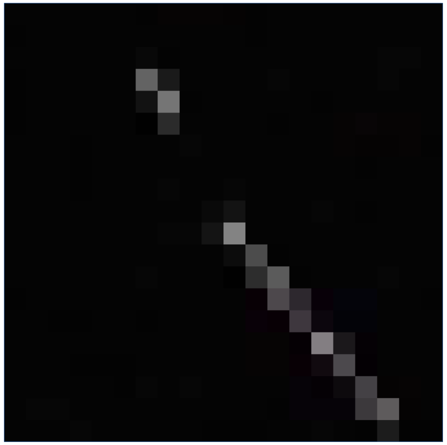
date	login	team	img
NEW 1 Hour ago	Piotr J. Piotrowski	no team	
1 Hour ago	Piotr J. Piotrowski	no team	
1 Hour ago	Piotr J. Piotrowski	no team	
4 hours ago	Piotr J. Piotrowski	no team	

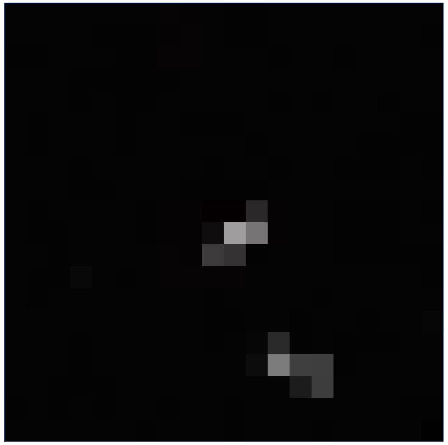
Image examples:



The science behind the CREDO app

Image examples:

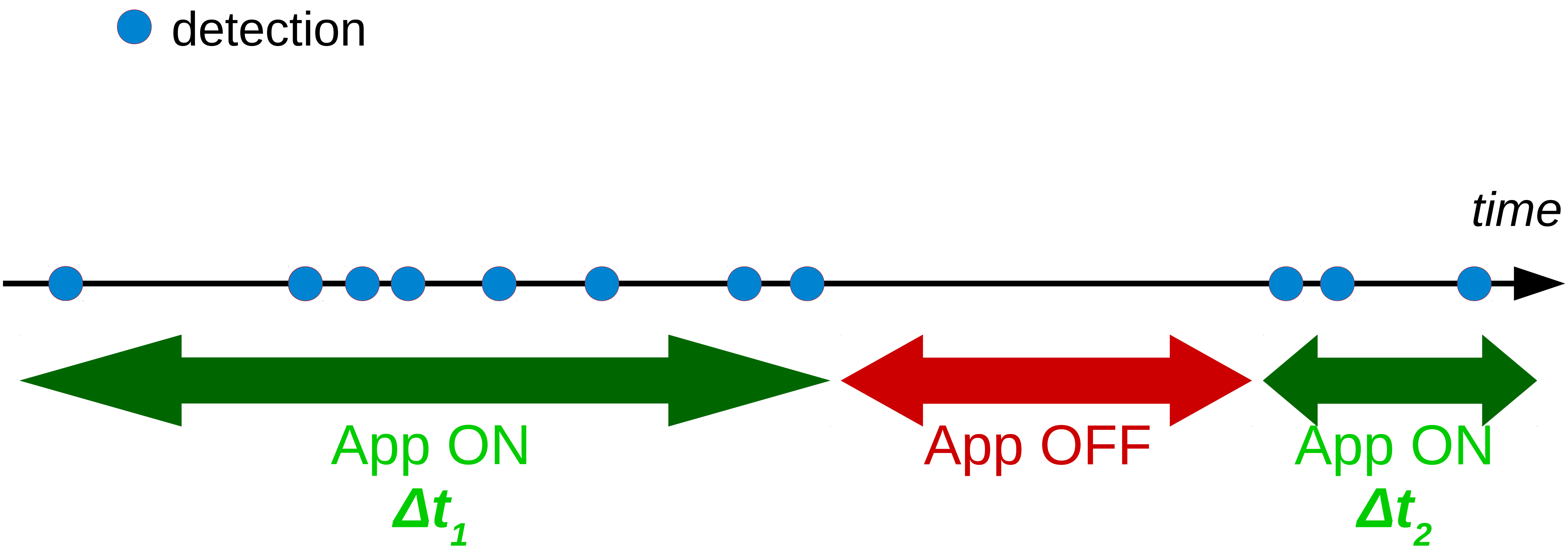




Informations obtained about the data:

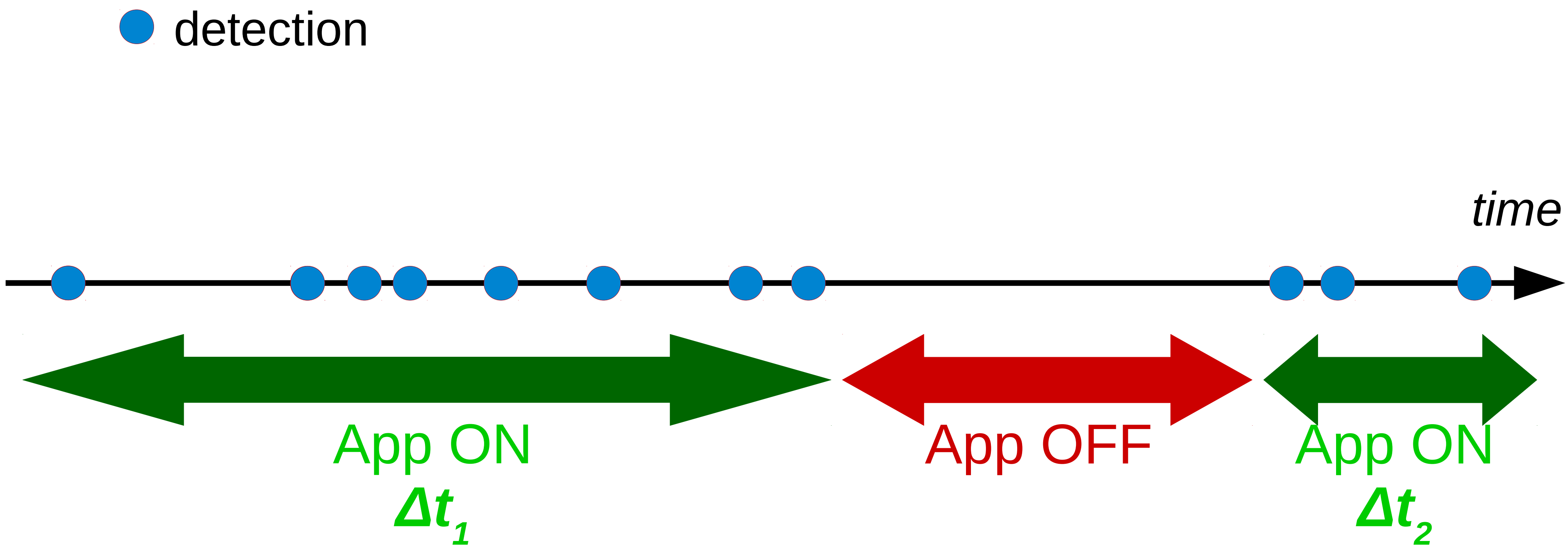
Timestamps and *GPS location*

Data acquisition and pre-processing



- If $\boldsymbol{\Delta t_1}$ + $\boldsymbol{\Delta t_2}$ + … + $\boldsymbol{\Delta t_n}$= 24h → timestamps are saved in a file.

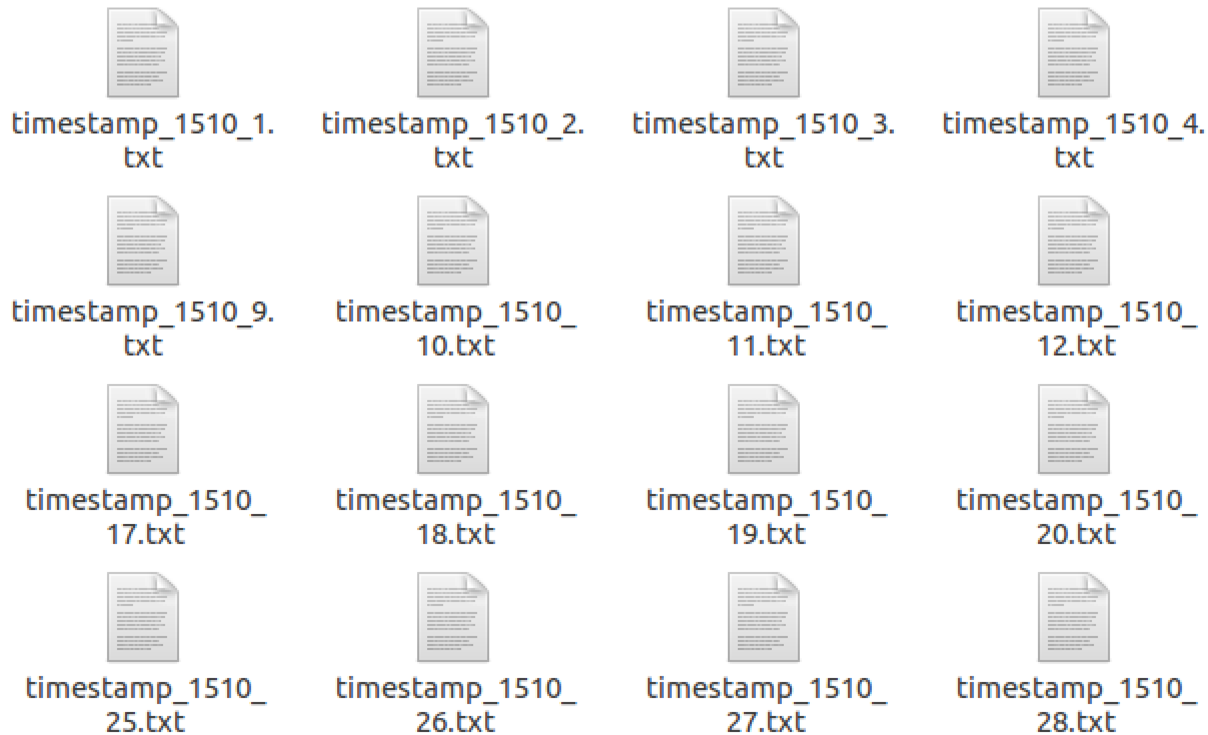
Data acquisition and pre-processing



- For each user, we obtained a file containing the timestamps for 24h periods:

  timestamps_<userID>_1.txt, timestamp_<userID>_2.txt, timestamp_<userID>_3.txt, etc...

Data acquisition and pre-processing




```
1529324517747
1529325363714
1529326915006
1529326915039
1529327416692
1529327430452
1529327430452
1529327799978
1529328562431
1529328592824
1529328624409
1529328729567
1529328739873
1529328784549
1529329109491
1529329940873
1529330466254
1529330595181
1529331273051
1529331883964
1529333849364
1529334153055
1529334217083
1529334217083
1529334799244
1529335003646
1529335139004
1529335139004
1529335561395
1529335561395
1529335765857
```

What are we looking for?

VOLUME 50, NUMBER 26 PHYSICAL REVIEW LETTERS 27 JUNE 1983

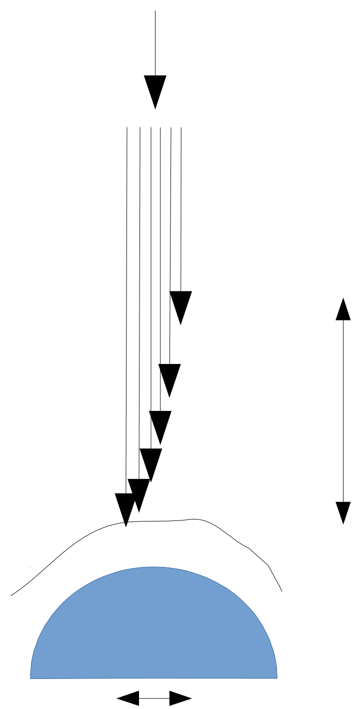
Possible Observation of a Burst of Cosmic-Ray Events in the Form of Extensive Air Showers

Gary R. Smith, M. Ogmen, E. Buller, and S. Standil
Physics Department, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

(Received 7 April 1983)

A series or burst of 32 extensive air showers of estimated mean energy 3×10^{15} eV was observed within a 5-min time interval beginning at 9:55 A.M. (CST) on 20 January 1981 in Winnipeg, Canada. This observation was the only one of its kind during an experiment which recorded 150 000 such showers in a period of 18 months between October 1980 and April 1982.

PACS numbers: 94.40.Pa, 94.40.Rc, 95.30.-k



Year = 1981
N_{obs} = 32
N_{exp} = 1
E = 3×10^{15}eV

Δt ~5 min.

Δx >= small

What are we looking for?

VOLUME 50, NUMBER 26 PHYSICAL REVIEW LETTERS 27 JUNE 1983

Possible Observation of a Burst of Cosmic-Ray Events in the Form of Extensive Air Showers

Gary R. Smith, M. Ogmen, E. Buller, and S. Standil
Physics Department, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

(Received 7 April 1983)

A series or burst of 32 extensive air showers of estimated mean energy 3×10^{15} eV was observed within a 5-min time interval beginning at 9:55 A.M. (CST) on 20 January 1981 in Winnipeg, Canada. This observation was the only one of its kind during an experiment which recorded 150 000 such showers in a period of 18 months between October 1980 and April 1982.

PACS numbers: 94.40.Pa, 94.40.Rc, 95.30.-k

First analysis:

Looking for how many times two consecutive detections happen within 5 minutes time windows in the data and compare to background expectations!

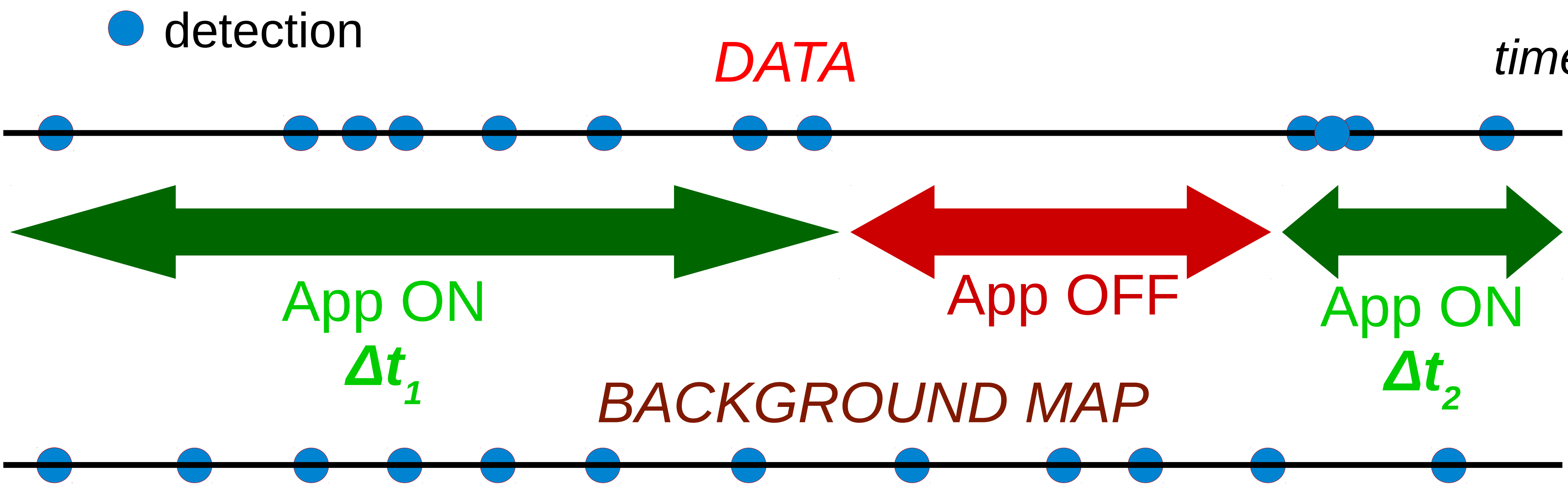
What does the algorithm look like?

1) **Extract data** from *timestamp* file.

What does the algorithm look like?

1) Extract data from *timestamp* file.

2) **Simulate MULTIPLE background maps** based on a uniform distribution of detections and the number of detections in the data.



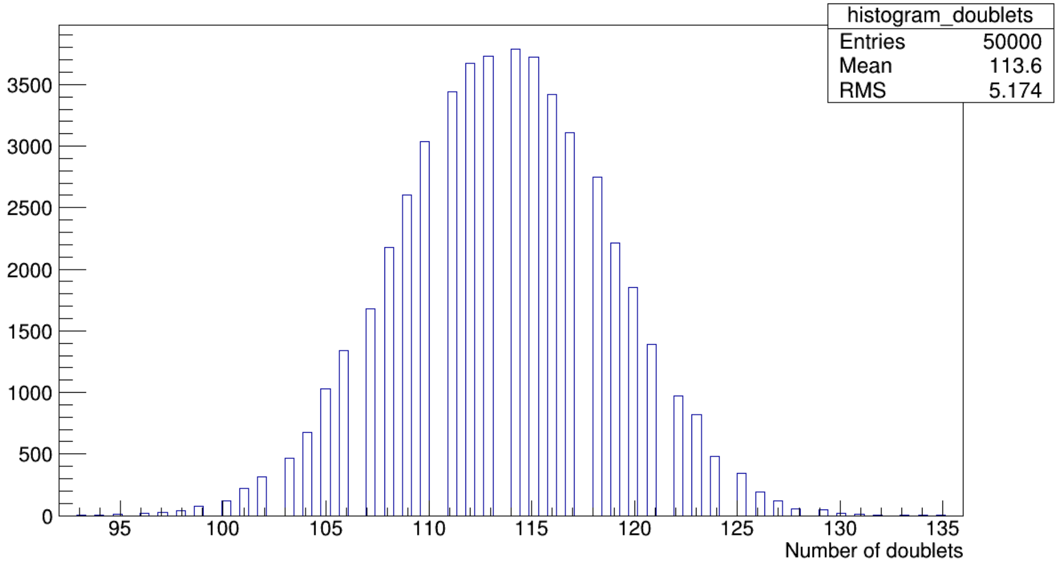
$$\Delta t_B = \Delta t_1 + \Delta t_2 = 24H$$

What does the algorithm look like?

1) Extract data from *timestamp* file.

2) Simulate multiple background maps based on a uniform distribution of detections and the number of detections in the data.

3) Count **how many times two consecutive detections happen within 5 minutes time windows in each background map** to obtain distribution and 3-sigma/5-sigma values:



What does the algorithm look like?

1) Extract data from *timestamp* file.

2) **Simulate MULTIPLE background maps** based on a uniform distribution of detections and the number of detections in the data.

3) Count how many times two consecutive detections happen within 5 minutes time windows in each background map to obtain distribution and 3-sigma/5-sigma values.

4) Count **how many times two consecutive detections happen within 5 minutes time windows <u>in data</u>** to compare do background distribution and **obtain sigma/p-value**.

What does the algorithm look like?

1) Extract data from *timestamp* file.

2) Simulate multiple background maps based on a uniform distribution of detections and the number of detections in the data.

3) Count how many times two consecutive detections happen within 5 minutes time windows in each background map to obtain distribution and 3-sigma/5-sigma values.

4) Count how many times two consecutive detections happen within 5 minutes time windows in data to compare do background distribution and obtain sigma/p-value.

5) Save **expected (background) and observed (data) values, 3 and 5 sigma bands, and significance of observed** in output file.

What does the algorithm look like?

1) Extracting data from *timestamp* file.

2) Simulate multiple background maps based on a uniform distribution of detections and the number of detections in the data.

REPEAT PROCESS FOR EACH TIMESTAMP FILE

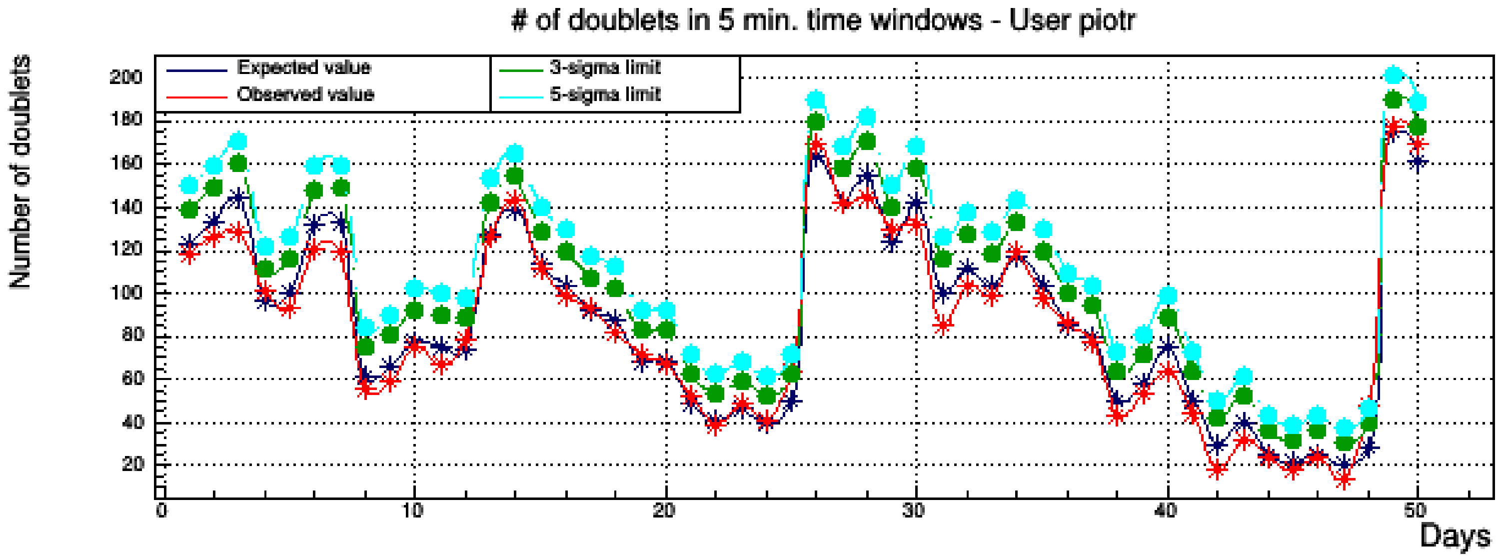
3) C … n
wi … obtain
dis …

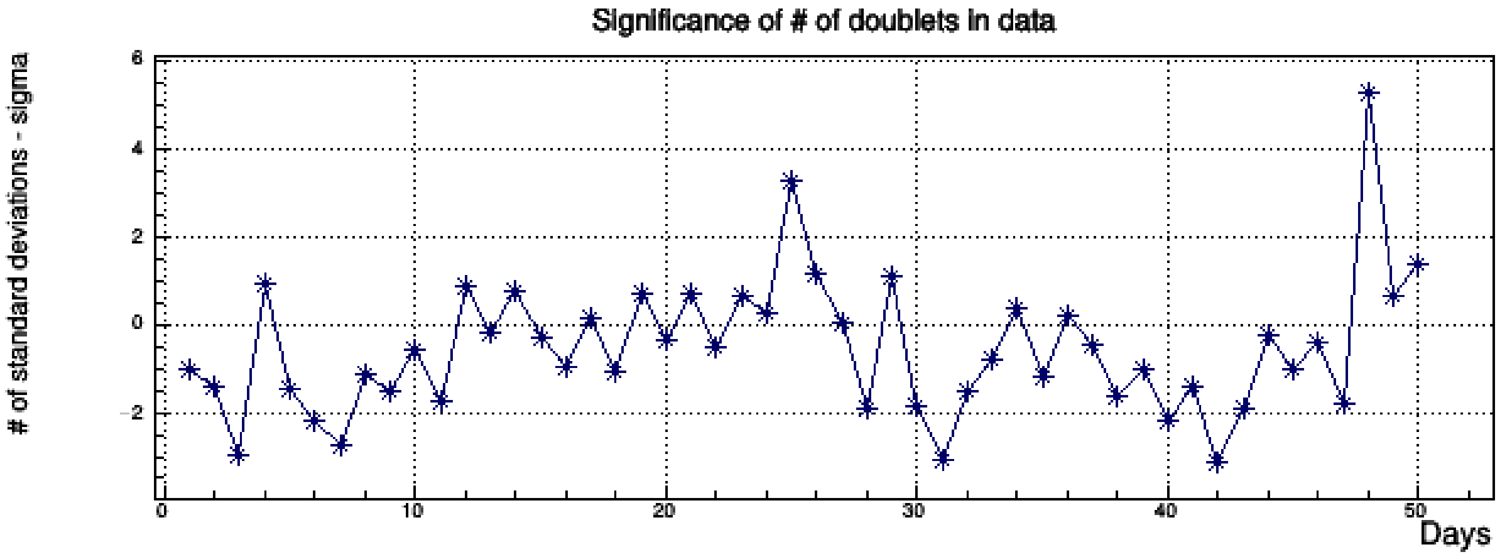
4) Count how many times two consecutive detections happen within 5 minutes time windows in data to compare do background distribution and obtain sigma/p-value.

5) Save expected (background) and observed (data) values, 3 and 5 sigma bands, and significance of observed in output file.

What do the results look like?

- Each point correspond to one timestamp file <=> one 24h period





How to run the algorithm?

- **3 FILES:**

  - *Analysis.cpp*: algorithm written in c++ performing the previously mentioned analysis.
    → OUTPUT: txt file with values used for plots.

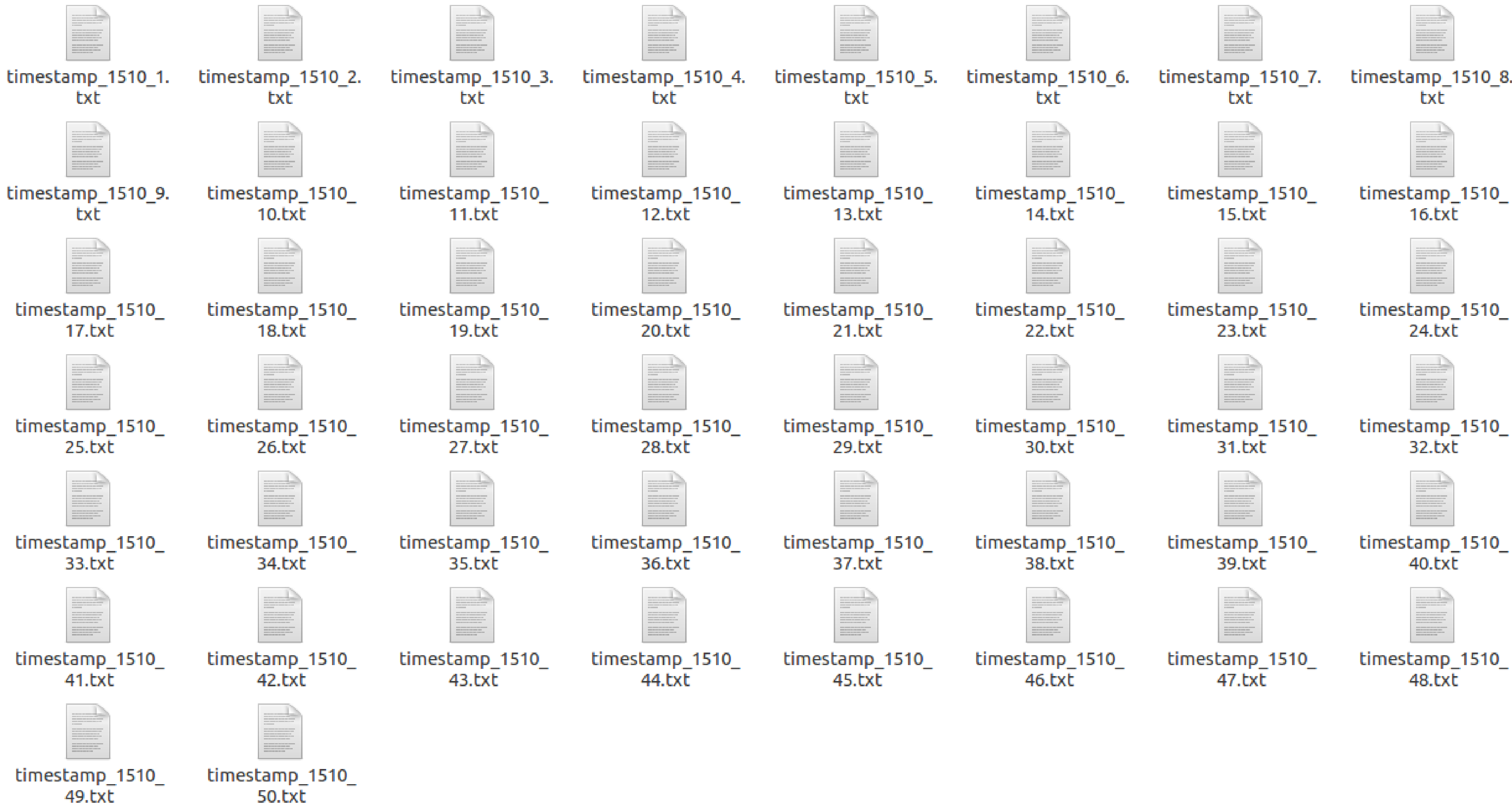
  - *Plot4user.C*: ROOT macro plotting the results obtained from the analysis.
    → OUTPUT: plots.

  - *Run.sh*: bash script compiling *analysis.cpp* file and looping over all timestamp files.

  README file contains extra informations!

PRACTICE!

Practice


timestamp_1510_1.txt
timestamp_1510_2.txt
timestamp_1510_3.txt
timestamp_1510_4.txt
timestamp_1510_5.txt
timestamp_1510_6.txt
timestamp_1510_7.txt
timestamp_1510_8.txt
timestamp_1510_9.txt
timestamp_1510_10.txt
timestamp_1510_11.txt
timestamp_1510_12.txt
timestamp_1510_13.txt
timestamp_1510_14.txt
timestamp_1510_15.txt
timestamp_1510_16.txt
timestamp_1510_17.txt
timestamp_1510_18.txt
timestamp_1510_19.txt
timestamp_1510_20.txt
timestamp_1510_21.txt
timestamp_1510_22.txt
timestamp_1510_23.txt
timestamp_1510_24.txt
timestamp_1510_25.txt
timestamp_1510_26.txt
timestamp_1510_27.txt
timestamp_1510_28.txt
timestamp_1510_29.txt
timestamp_1510_30.txt
timestamp_1510_31.txt
timestamp_1510_32.txt
timestamp_1510_33.txt
timestamp_1510_34.txt
timestamp_1510_35.txt
timestamp_1510_36.txt
timestamp_1510_37.txt
timestamp_1510_38.txt
timestamp_1510_39.txt
timestamp_1510_40.txt
timestamp_1510_41.txt
timestamp_1510_42.txt
timestamp_1510_43.txt
timestamp_1510_44.txt
timestamp_1510_45.txt
timestamp_1510_46.txt
timestamp_1510_47.txt
timestamp_1510_48.txt
timestamp_1510_49.txt
timestamp_1510_50.txt

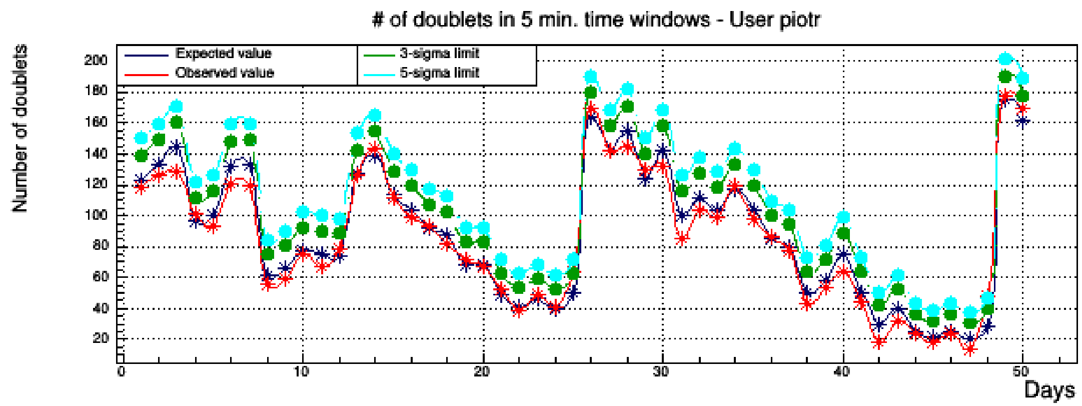
Practice

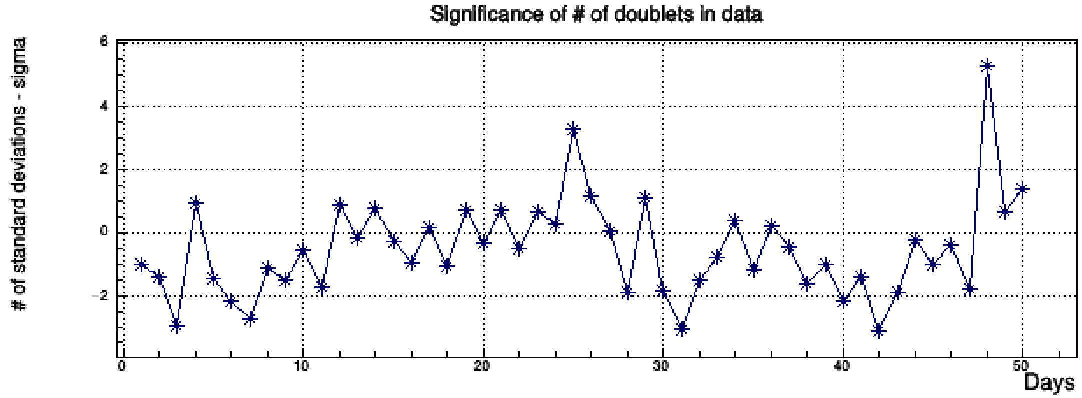
```
asava@asava-ubuntu:~/CREDO/Time_Clustering - API/v4$ bash run.sh
Processing data from user data - period 1...
Number of events in data: 252
Number of events in data after removing events with same timestamps: 227
Time covered by data = 140301.176 sec
Real ontime = 86400 sec
----- Doublet analysis -----
Expected number of doublets = 123.37118 || Number of doublets in data = 118 || pvalue = -0.694091016 || sigma = -1.023843958
3 sigma at 139.1094572 || 5 sigma at 149.601642
Elasped time is 2.00 seconds.

Processing data from user data - period 2...
Number of events in data: 278
Number of events in data after removing events with same timestamps: 238
Time covered by data = 88047.369 sec
Real ontime = 86400 sec
----- Doublet analysis -----
Expected number of doublets = 133.42482 || Number of doublets in data = 126 || pvalue = -0.8432097339 || sigma = -1.415949584
3 sigma at 149.1559309 || 5 sigma at 159.6433382
Elasped time is 1.00 seconds.

Processing data from user data - period 3...
Number of events in data: 273
Number of events in data after removing events with same timestamps: 250
Time covered by data = 99952.467 sec
Real ontime = 86400 sec
----- Doublet analysis -----
Expected number of doublets = 144.64178 || Number of doublets in data = 129 || pvalue = -0.9970569934 || sigma = -2.973629014
3 sigma at 160.4222757 || 5 sigma at 170.9426062
Elasped time is 2.00 seconds.

Processing data from user data - period 4...
Number of events in data: 216
Number of events in data after removing events with same timestamps: 196
Time covered by data = 74366.055 sec
Real ontime = 86400 sec
----- Doublet analysis -----
Expected number of doublets = 96.39774 || Number of doublets in data = 101 || pvalue = 0.635533989 || sigma = 0.9068880314
3 sigma at 111.6220891 || 5 sigma at 121.7716552
Elasped time is 2.00 seconds.
```

Practice

1	123.371180	118.000000	-1.023844	139.109457	149.601642	-5.371180
2	133.424820	126.000000	-1.415950	149.155931	159.643338	-7.424820
3	144.641780	129.000000	-2.973629	160.422276	170.942606	-15.641780
4	96.397740	101.000000	0.906888	111.622089	121.771655	4.602260

Each line corresponds to the analysis of one 24h period (one timestamp file)





Real data vs. simulated data