

Toy Model For Hyperfine Measurement

Adriano Del Vincio, Germano Bonomi,
Simone Stracka

University of Brescia, INFN Pisa

December 7, 2023

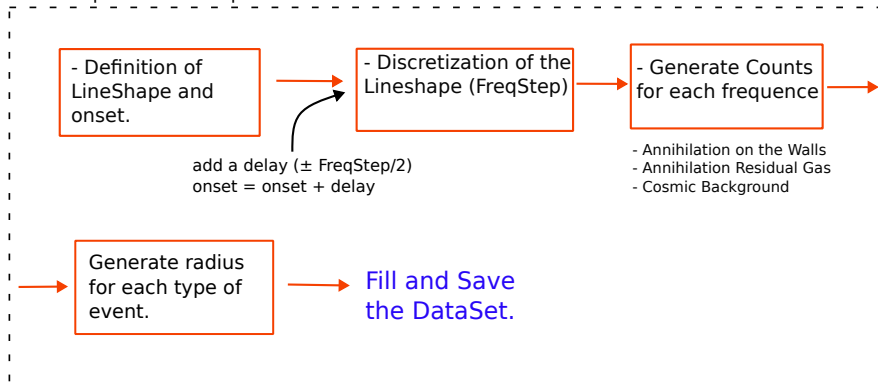


Scheme of the Simulation

Scheme of the Monte Carlo Toy for generating the events

Loop from 0 to Ntrials

Inner Loop: from 0 to Repetition



In this simulation, the data are created and analyzed using
RDataFrame framework.



A brief introduction about the Monte Carlo

We have developed a Monte Carlo Toy that produces two `.root` dataset files. The variables are columns of values that are shown in the figure below:

* Row *	* runNumber *	* random.ra *	* delay.del *	* frequence *	* type.type *	* radius.ra *
*****	*****	*****	*****	*****	*****	*****
0 *	0 *	0.4849736 *	2.4987087 *	-25 *	2 *	2.8792768 *
1 *	1 *	0.2899349 *	-1.685450 *	15 *	0 *	2.0739069 *
2 *	1 *	0.0197818 *	-1.685450 *	15 *	0 *	1.8959179 *
3 *	1 *	0.2412478 *	-1.685450 *	15 *	0 *	2.8919173 *
4 *	1 *	0.3846191 *	-1.685450 *	15 *	0 *	3.3842529 *
5 *	1 *	0.4549068 *	-1.685450 *	15 *	0 *	1.9130180 *
6 *	1 *	0.3739825 *	-1.685450 *	15 *	0 *	1.6047382 *

Figure: Structure of the dataset.

- *runNumber*: identifies which run the event belong to (from 0 to *Repetition* – 1)
- *random*: values uniform distributed from 0 to 1, can be used to randomize the selection or for sub-sampling in the data
- *shift*: store the onset shift
- *frequence*: the frequency of the event
- *type*: type of the event: 0 annihilation on the walls, 1 residual gas annihilation, 2 cosmic event
- *radius*: radius of the annihilation vertex.



A brief introduction about the Monte Carlo

The Annihilation on the walls are generated as function of the frequency, using the two line-shapes of the transitions ($c \rightarrow b$) and ($d \rightarrow a$). The Annihilation on the residual gas and the cosmic background are generated uniformly on the frequency spectrum. All the parameters of the simulation are loaded from the `ToyConfiguration.txt` file. The parameters are chosen to reproduce the runs 4b.

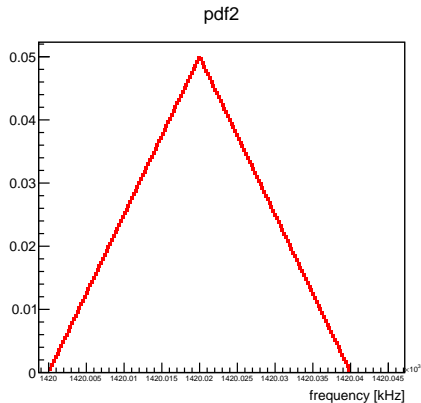
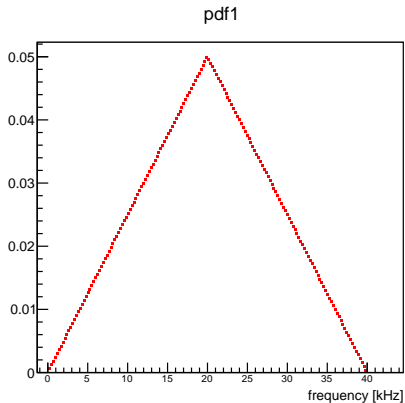
```
Nstack = 20
NHbar = 14
Repetition = 5
# cosmic rate is expressed in event/second
CosmicRate = 0.051028571
Efficiency = 1
pwall_cb = 1
pwall_ad = 1
C = 0.5
FrequencyStep = 5
TimeStep = 8
SweepStep = 24
# The following are in kHz units
x_cb_start = 0
x_cb_end = 40
x_cb_peak = 20
x_da_start = 1420000
x_da_peak = 1420020
x_da_end = 1420040
delay = 2.5
```

Figure: Parameters of the Toy.



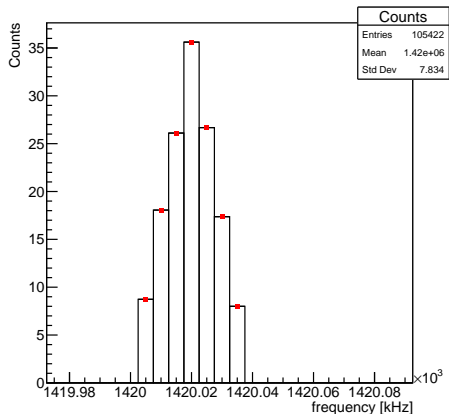
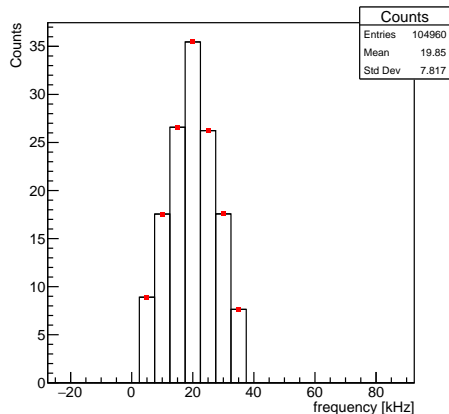
Triangular Line-shape Pdfs

For this first use of the toy, we have chosen simple line-shapes, triangular with a symmetric rise and fall.



Triangular Line-shapes Simulation

We sample at the given frequency step of 5 kHz the Pdfs, to simulate the experimental line-shapes. We applied the onset finding algorithm to this distribution.



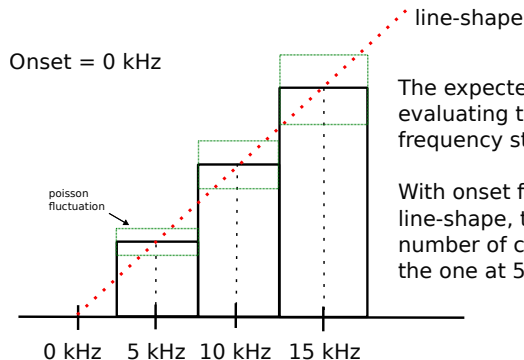
The onset is fixed at $f = 0$ kHz. In this sample the cosmic background is set to zero.



A simple Onset finding Algorithm

The first algorithm that is tested is quite simple: **the onset is identified by the first bin with a content over a given threshold ($> N\mu_{cosmic}$)**¹

Before showing the plot with the simulated data, it is useful to remind how this algorithm deals with the frequency step:



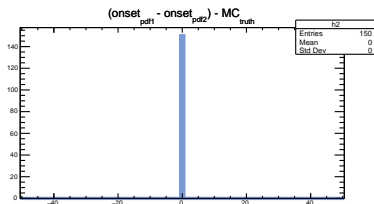
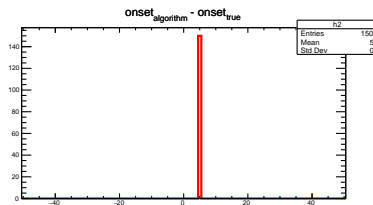
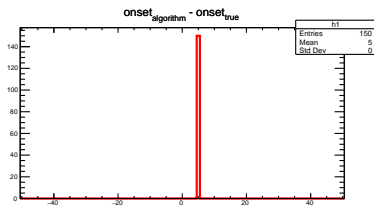
The expected bin content is given evaluating the lineShape at each frequency step (+ 5 kHz, +10kHz ...).

With onset fixed at 0, and a linear line-shape, the first bin with an expected number of counts different from 0 will be the one at 5 kHz (bias).

¹Where the μ_{cosmic} is computed from the Poisson distribution of the cosmic counts expected per bin.

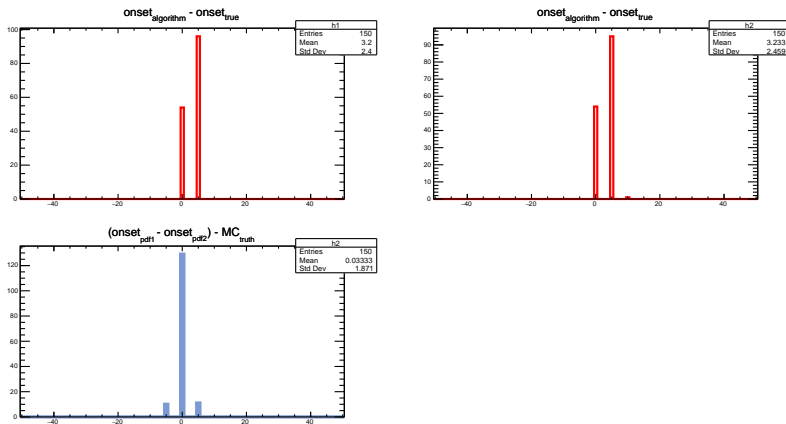
Consistency Check 1

We have tested the algorithm with a dataset without cosmic background and shift fixed to zero. The algorithm identifies the onset at frequency 5 kHz.



Consistency Check 2

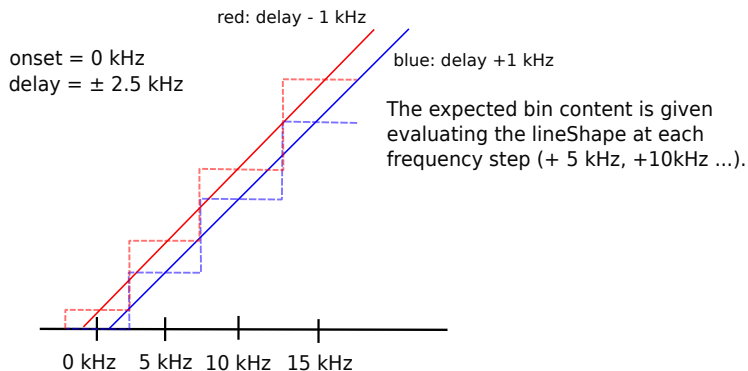
We have tested the algorithm with a dataset without cosmic background. The shift is uniform distributed in -2.5 kHz and 2.5 kHz.



With the shift, two bins ($frequency = 0$ kHz and $frequency = 5$ kHz) are populated.



Consistency Check 2

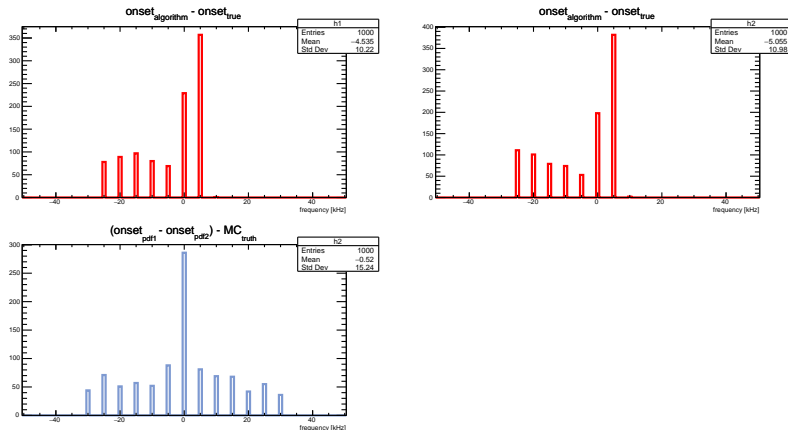


The frequency step are fixed, and the lineshape is shifted accordingly to the extracted value of the shift. In few words:

- step 1): fix the frequency steps (the frequency effectively used in the experiment)
- step 2): define the new value of the onset adding a uniform distributed shift
- step 3): shift the lineshape accordingly to the new value of the onset, discretize the lineshape and proceed with the generation of the data.

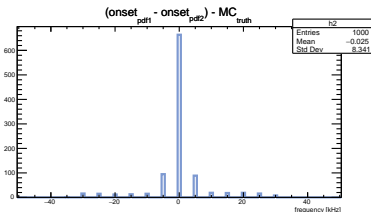
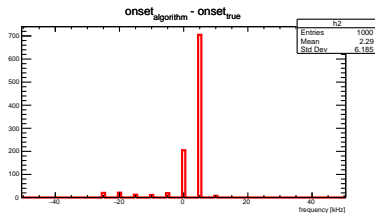
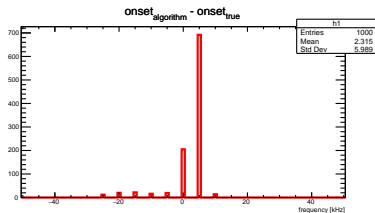
algorithm test: threshold $> 3\mu_{\text{cosmic}}$

In this case, we have applied the algorithm to simulated data with cosmic background (using a rate of $0.051 \frac{\text{event}}{\text{s}}$, from passcut1). Each bin has an expected cosmic background of $\text{dwelltime} \cdot \text{rate} = 0.408$. The shift is uniform distributed in -2.5 kHz and 2.5 kHz .



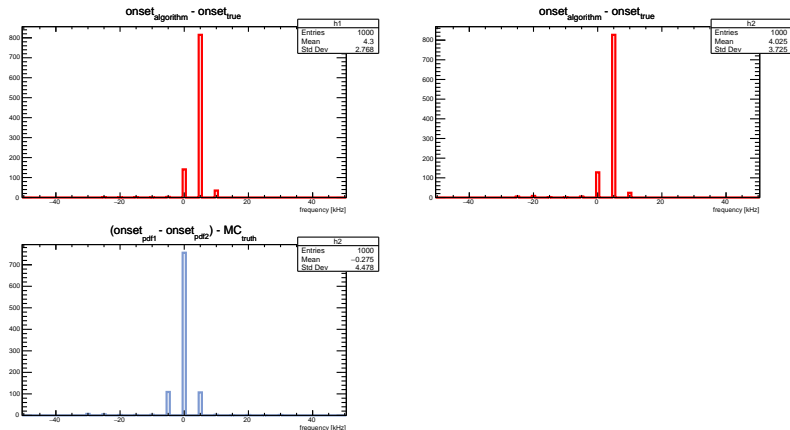
algorithm test: threshold $> 5\mu_{\text{cosmic}}$

In this case, we have applied the algorithm to simulated data with cosmic background (using a rate of $0.051 \frac{\text{event}}{\text{s}}$, from passcut1). Each bin has an expected cosmic content of $\text{dwelltime} \cdot \text{rate} = 0.408$. The shift is uniform distributed in -2.5 kHz and 2.5 kHz .



algorithm test: threshold $> 8\mu_{\text{cosmic}}$

In this case, we have applied the algorithm to simulated data with cosmic background (using a rate of $0.051 \frac{\text{event}}{\text{s}}$, from passcut1). Each bin has an expected cosmic content of $\text{dwelltime} \cdot \text{rate} = 0.408$. The shift is uniform distributed in -2.5 kHz and 2.5 kHz .



Improvements of the week (30/11 - 07/12)

- Implemented the onset-finding algorithm of 2017 (*first* > 0, *second* > 1)
- Simulation with a lineShape following the `run 69373` (lineShape with high statistics).
- Implementation and test of different onset finding algorithms.

Fit to the data of run 69373

The lineShape is fitted using a Cruiff function, which takes into account the asymmetry of the left-right tails, ($model = N \cdot \exp(\frac{-(x-x_0)^2}{2\sigma_{0,1} + k_{0,1}(x-x_0)^2})$)

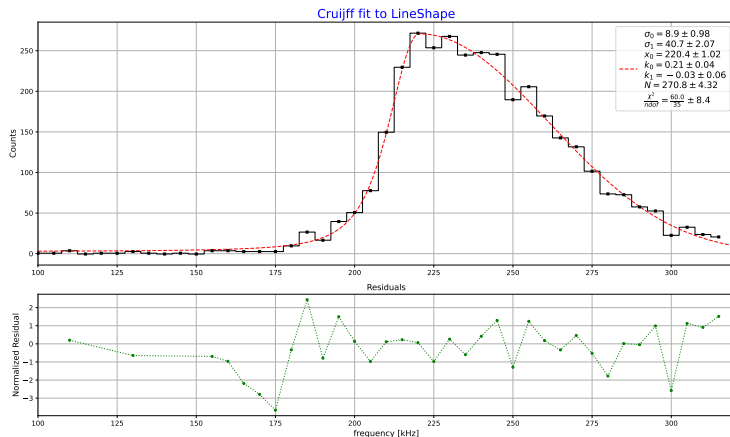
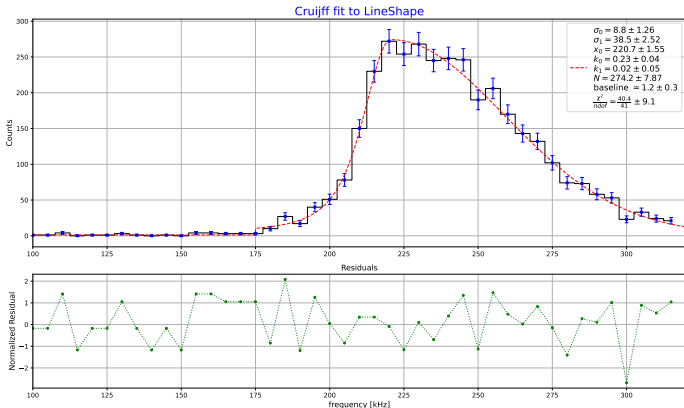


Figure: On top plot, the black line represents data and the red line the fit with the Cruiff function.

The Cruiff function is used in the simulation to generate the data. **The Cruiff is truncated at $f_0 = 175$ kHz.** In this way the onset of the lineshape is unambiguously determined. The new model is:

$$model = \begin{cases} baseline & f \leq 175 \text{ kHz} \\ N \cdot \exp\left(\frac{-(x-x_0)^2}{2\sigma_0+k_0(x-x_0)^2}\right) & 175 \text{ kHz} < f \leq x_0 \\ N \cdot \exp\left(\frac{-(x-x_0)^2}{2\sigma_1+k_1(x-x_0)^2}\right) & f > x_0 \end{cases} \quad (1)$$

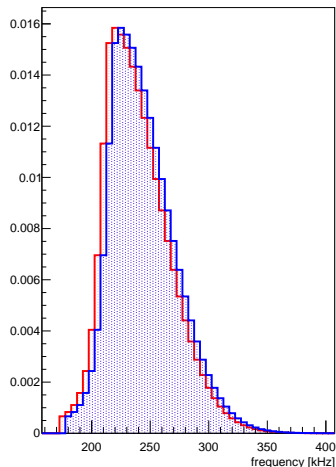


Is the baseline compatible with the Cosmic Background?

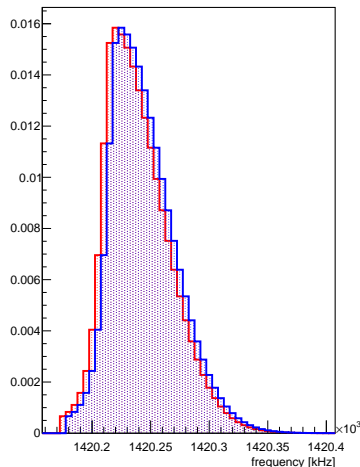
Discretization of the lineshape

Here we plot the discretized lineshape, obtained sampling the Cruijff function with the optimal parameters by the fit of the previous slide. The lineshape is discretized into a series of 50 points, with a increment step of 5 kHz. The starting frequency is $f = 160$ kHz

lineshape c to b

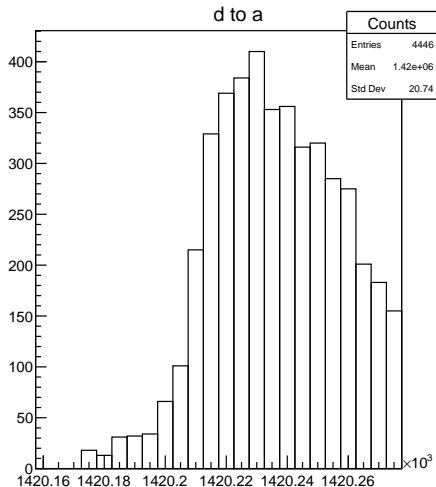
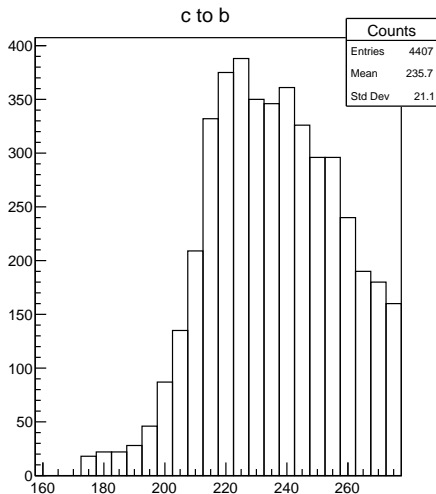


lineshape d to a



Generate Counts for each frequency

For illustration purposes, in this plot we show the simulated distribution (24 steps) for a single run with an artificially high number of antihydrogen (5000 in the following plot), without adding the cosmic background:



Parameters of the Simulation

We have studied the case of the series of run 4b. The parameters of the simulation are:

- $N_{stack} = 20$.
- $N_{\bar{H}} \text{ per stack} = 14$.
- $SweepSteps = 24$.
- $Repetition = 5$ (not yet used in the following).
- $TimeStep = 8 \text{ s}$
- $FrequencyStep = 5 \text{ kHz}$.
- $\mu_{cosmic} = 0.051 \text{ s}^{-1}$
- $onset_{1,true} = 0 \text{ kHz}$; $onset_{2,true} = 1\,420\,000 \text{ kHz}$
- $shift = \pm 2.5 \text{ kHz}$.

The percentage of events of annihilation to residual gas is set to zero. The amount of anti-hydrogen is divided equally for the two transition c-b and d-a.

The first algorithm that we test is the one applied to the data taking of 2017. The onset is estimated taking the frequency which fulfills the criteria

$$f_i : bin(i) > 0; bin(i + 1) > 1 \quad (2)$$

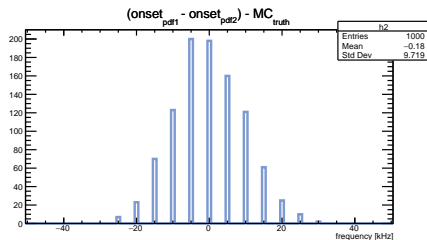
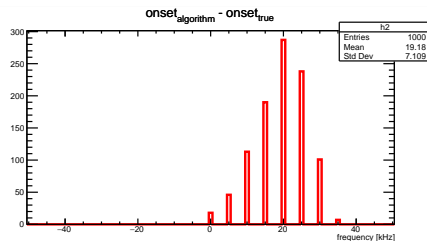
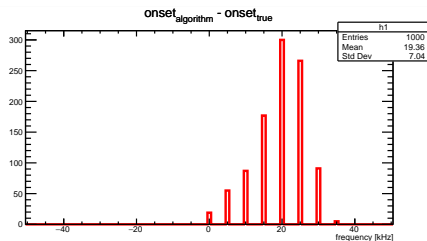
a second version of the same algorithm is implemented, analyzing the frequencies in decreasing order (reversed algorithm):

$$f_i : bin(i) < 3; bin(i - 1) < 2$$



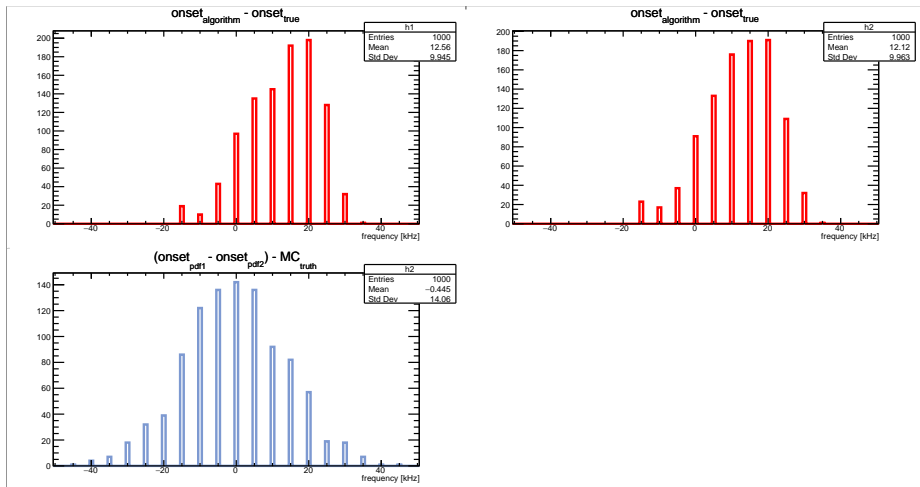
test 2017 algorithm

The algorithm is tested for $N_{\text{trial}} = 1000$. The amount of events per transition is expected to be poissonian distributed with mean $\simeq 130$. In this first scenario the cosmic events are removed from the data.



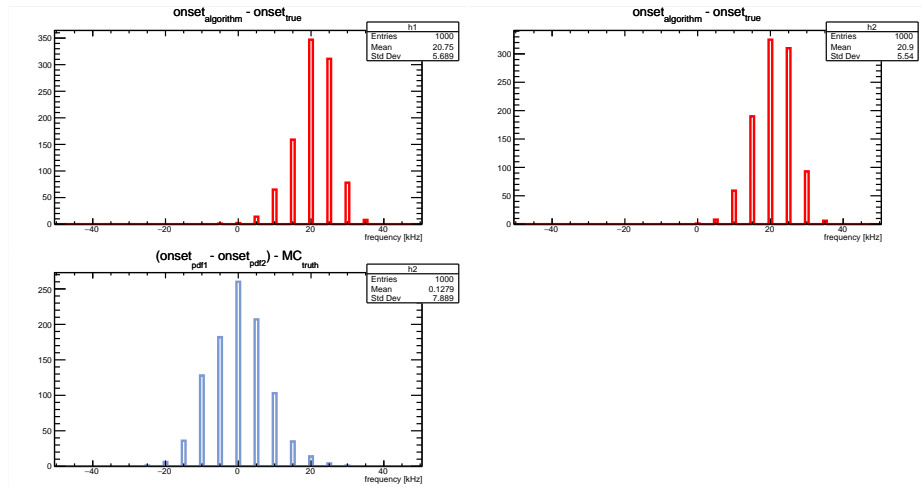
test 2017 algorithm

The algorithm is tested for $N_{\text{trial}} = 1000$. The amount of events per transition is expected to be poissonian distributed with mean $\simeq 130$. The cosmic background is fixed to 0.41 events per frequency.



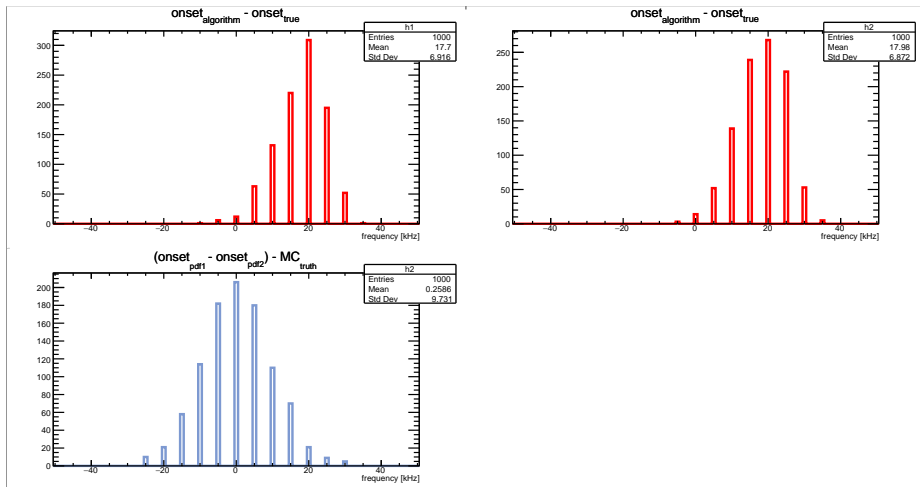
test 2017 algorithm (reversed)

The algorithm is tested for $N_{\text{trial}} = 1000$. The amount of events per transition is expected to be poissonian distributed with mean $\simeq 130$. In this scenario the cosmic background is removed.



test 2017 algorithm (reversed)

The algorithm is tested for $N_{\text{trial}} = 1000$. The amount of events per transition is expected to be poissonian distributed with mean $\simeq 130$. The cosmic background is fixed to 0.41 events per frequency.



Other strategies

We have tested other 2 different algorithms, that can be useful to identify the onset. The first algorithm identifies the onset as the first frequency with counts over threshold:

$$f_i : bin_i > Threshold \quad (3)$$

The second algorithm is a constant fraction discriminator. It works similarly to the previous algorithm, except for the fact that the threshold is computed each time as:

$$threshold = p \cdot \max\{bin_i\} \quad (4)$$

where p is a parameter of the algorithm, in the range $(0,1)$.

In the end we have tested another algorithm (*sumNeighbors*), defined in this way:

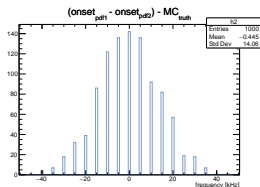
$$f_i : bin_i + bin_{i+1} + bin_{i+2} > 3 \cdot \mu_{cosmic} + \sqrt{3N} \cdot \sqrt{\mu_{cosmic}} \quad (5)$$

Where N is a parameter of the algorithm. This algorithm is directly derived from the t-student test.

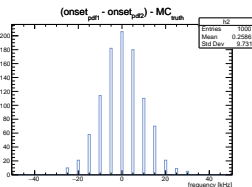
Results

algorithms	μ [kHz]	σ [kHz]	μ [kHz]	σ [kHz]
	without cosmic background		with cosmic background	
2017 (> 0 ; > 1)	-0.18	9.7	-0.445	14.1
2017 reversed (< 2 ; < 1)	+0.13	7.89	+0.25	9.73
threshold (> 1)	+0.025	11.5	+0.005	16.72
threshold (> 2)	-0.06	9.56	+0.67	14.1
threshold (> 3)	-0.1	7.48	-0.07	9.96
const. fraction ($p = 10\%$)	-0.02	11.56	+0.19	16.66
const. fraction ($p = 20\%$)	-0.215	9.05	+0.55	13.06
const. fraction ($p = 30\%$)	+0.08	7.43	+0.05	8.84
sumNeighbors ($N = 1$)	+0.01	9.47	+0.055	13.03
sumNeighbors ($N = 2$)	-0.425	8.83	+0.005	12.69
sumNeighbors ($N = 3$)	-0.47	7.85	+0.115	11.66

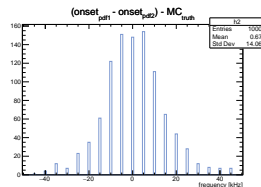
Table: Result of the simulation for $N_{\text{trials}} = 1000$. In the first two column the cosmic background is removed from the data. The last two columns contains the result of the simulation adding the cosmic background. The μ and σ are the quantities computed from the distribution of $(\text{onset}_2 - \text{onset}_1) - MC_{\text{truth}}$.



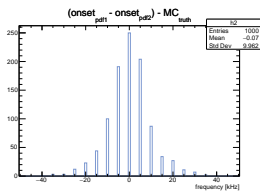
(a) algorithm 2017



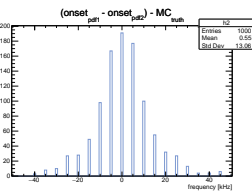
(b) algorithm 2017 reversed



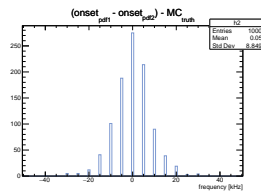
(c) threshold (> 2)



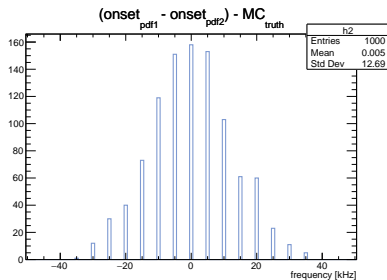
(d) threshold (> 3)



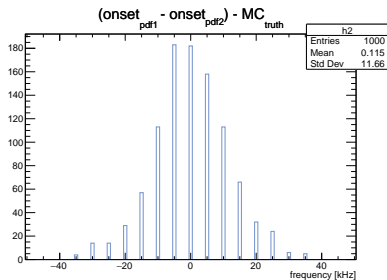
(e) const fraction (> 20%)



(f) const fraction (> 30%)



(g) *sumNeighbors* ($N > 2$)



(h) *sumNeighbors* ($N > 3$)

Next steps

- Add to the simulation the magnetic field drift
- Systematic studies of the relative shape asymmetry for the two transitions and its effect on the onset determination.