A neural network to classify GRAND radio time traces

ARENA 2022
Santiago de Compostela, June 7 – 10

Sandra Le Coz (LPNHE, Paris)
with
Aurélien Benoit-Lévy (CEA, Paris)
Olivier Martineau (LPNHE, Paris)

Towards a giant radio array for ν detection

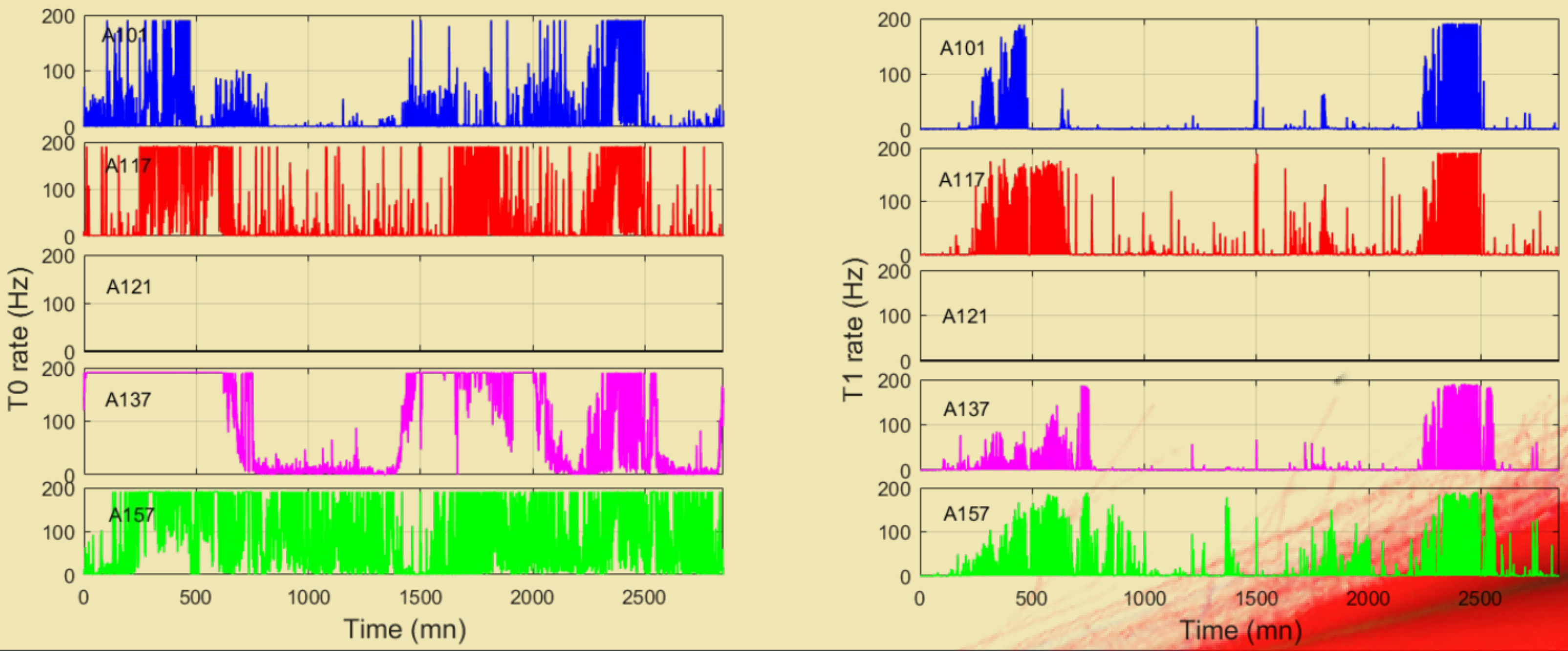
- Low interactions rate of UHE ν

→ low flux of ν - extensive air shower on Earth

→ huge area of detection

→ cheap detectors

→ antennas without external (particle) trigger



TREND data

- Anthropic transient rate >10Hz even in « quiet » places

→ **a smart trigger is required to avoid saturation of acquisition**

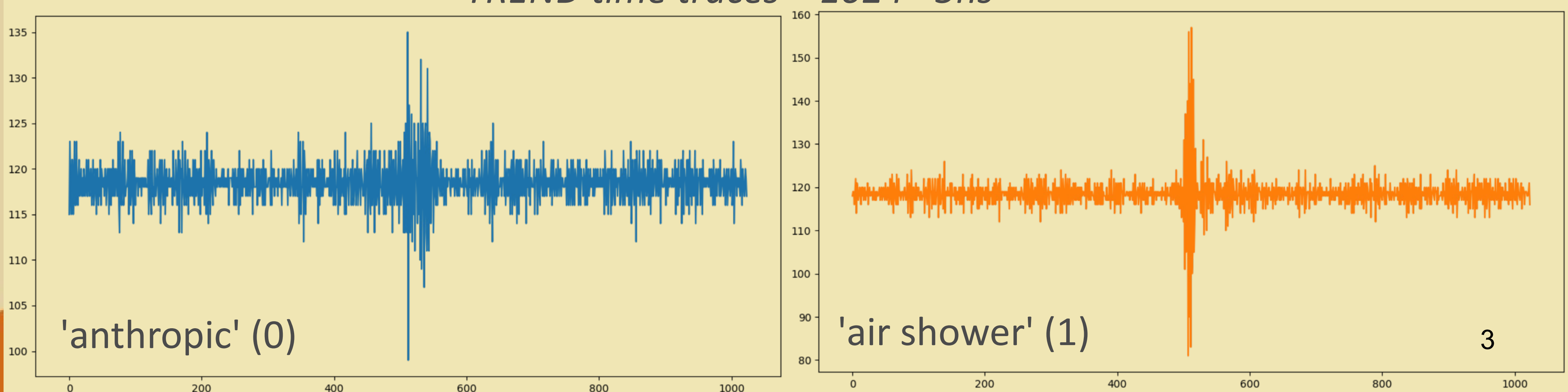
A neural network as a smart trigger

How to discriminate **air shower/anthropic** signals @ antenna level?
→ Use a neural network (NN)?

- NN input = a radio time trace
- NN target = 0/1 (anthropic origin/air shower origin)

On which data do train/test?
→ experimental data rather than simulations
→ data we already have, a subset of TREND data

TREND time traces – 1024 5ns*



TREND experiment

- 50 butterfly antennas
- Single-polarized
- 50-100 MHz
- Self-triggered
- Setup in XinJiang, China
- Between 2011 and 2014

- Trigger @ antenna level if :

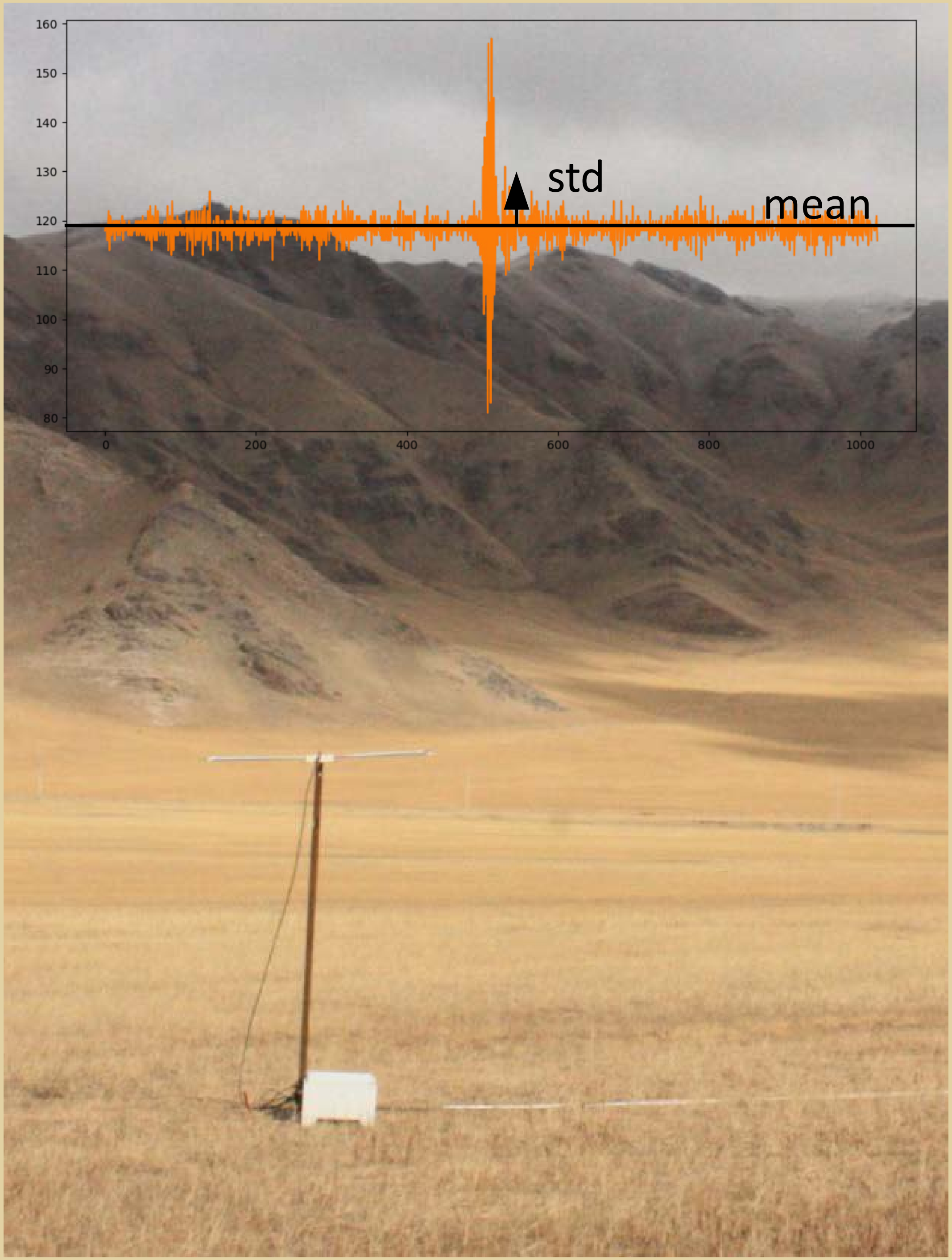
abs(amp$_i$ - μ(amp)) > 6σ(amp)

- Trigger @ array level if :

space-time correlations between 5+ antenna triggers

→ « coincidence » recorded

In this study we use a subset :
1.25e8 recorded coincidences
9e8 antenna time traces



TREND data labelisation

Before making predictions, NN must be trained on data for which label is known
→ We put labels ('anthropic'=0/'air shower'=1) on TREND radio time traces using TREND standard data analysis :

See [arXiv:1810.03070], 2018

Offline cuts were appplied to reject :

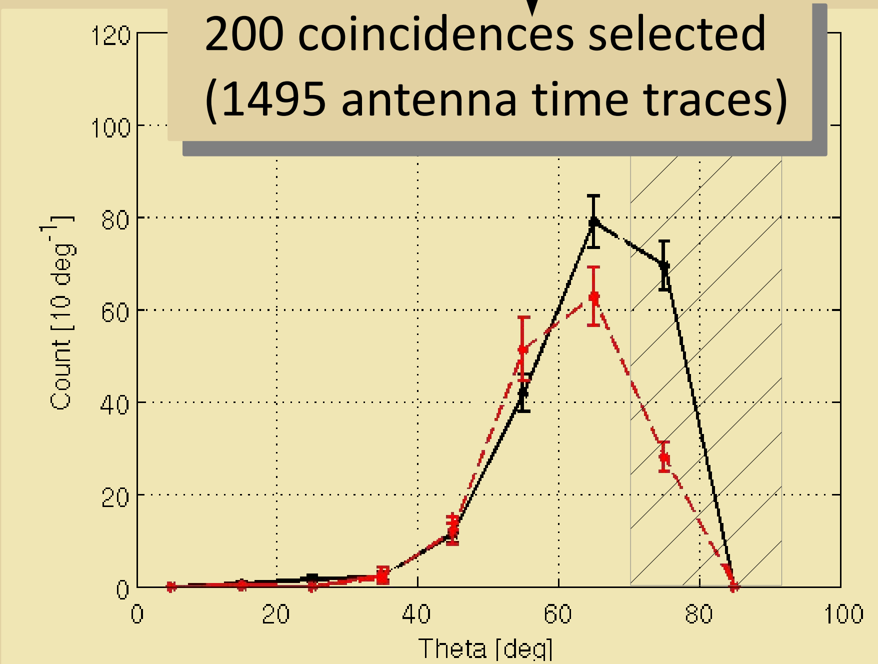
- noisy periods
- too long antenna signals
- angle/source reconstruction : hight χ^2, near source position, zenith>80°
- discontinuous trigger pattern at ground
- direction-time correlations between coincidences

TREND data labelisation

1.25e8 coincidences recorded
(9e8 antenna time traces)

cuts

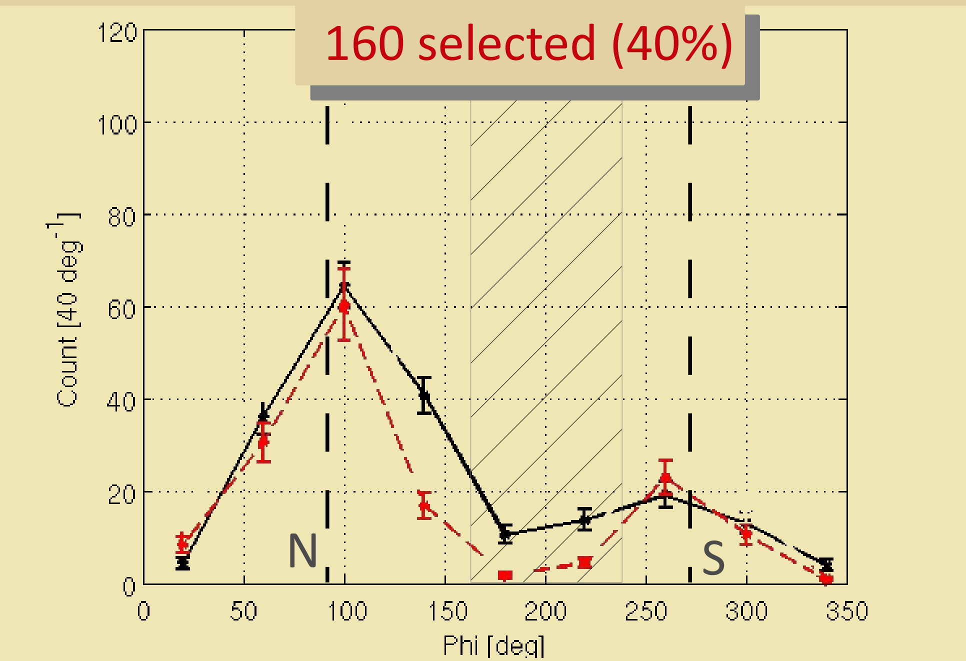
200 coincidences selected
(1495 antenna time traces)



Simulations of air showers + TREND response :
370 air showers would have been recorded

cuts

160 selected (40%)



- Consistency between selected **experimental data** & **simulations**

regarding shape (N/S geomagnetic expectations) & integral of angular distribution
→ The 200 selected coincidences might be air showers with a contamination ~20%

- Contamination is high for **θ>70 & 160<φ<240** → rejected of labelisation

→ **'air shower'** label for the 984 traces (139 surviving coincidences)
→ **'anthropic'** label for the remaining ~9e8 traces

Convolutional Neural Network

~ Dataset ~

984*2 = 1968 $\neq$ 'anthropic' traces (randomly picked)

984 $\neq$ 'air shower' traces (w=2 to keep balance)

split = 80% training/20% test

~ Preprocessing ~

times traces shifted to a mean of 0 + scaled to a range [-1;1]

('mean of the trace' is irrelevant feature + help training)

first layer input = FFT(time trace) (works better)

~ Layers~

convs (filters=8/16, kernel size=51, padding='same', activation='relu') + maxpoolings

60%-dropout + L2 regul = 2e-3 to avoid overfit

~ Fit ~

optimizer = adam

loss = cross entropy

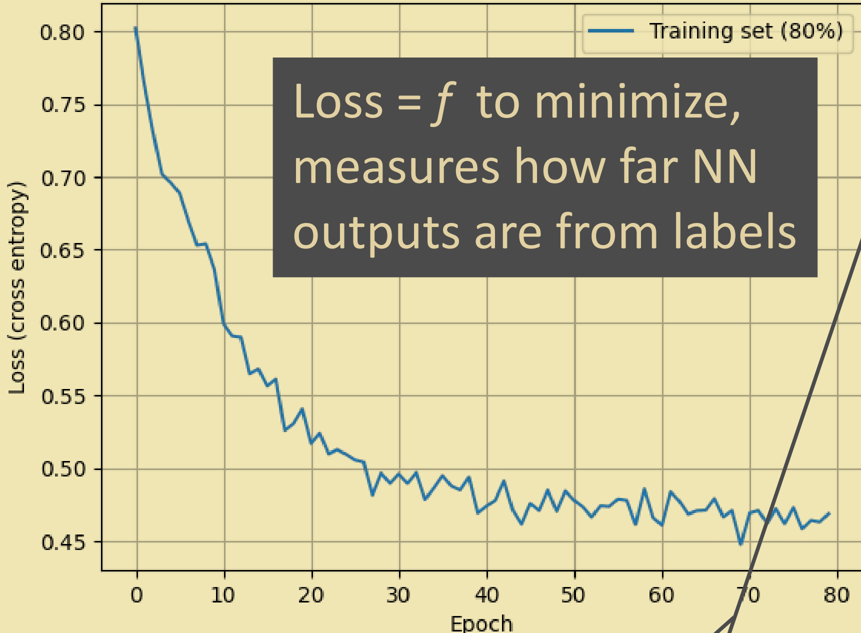
Model summary

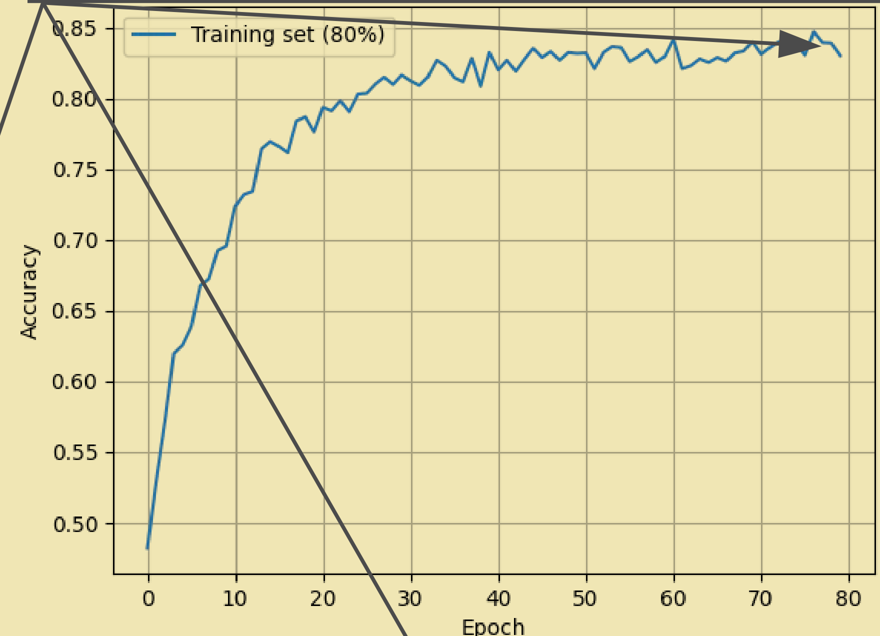
```
_________________________________________________________________
 Layer (type)              Output Shape            Param #
=================================================================
 input_1 (InputLayer)       [(None, 1024, 1)]        0
 conv1d (Conv1D)           (None, 1024, 8)         416
 max_pooling1d (MaxPooling1D)  (None, 512, 8)          0
 dropout (Dropout)          (None, 512, 8)          0
 conv1d_1 (Conv1D)          (None, 512, 16)         6544
 max_pooling1d_1 (MaxPooling 1D) (None, 256, 16)         0
 dropout_1 (Dropout)        (None, 256, 16)          0
 conv1d_2 (Conv1D)          (None, 256, 16)         13072
 flatten (Flatten)         (None, 4096)            0
 dropout_2 (Dropout)        (None, 4096)            0
 dense (Dense)              (None, 16)             65552
 dropout_3 (Dropout)        (None, 16)              0
 dense_1 (Dense)            (None, 2)              34
=================================================================
Total params: 85,618
Trainable params: 85,618
Non-trainable params: 0
```

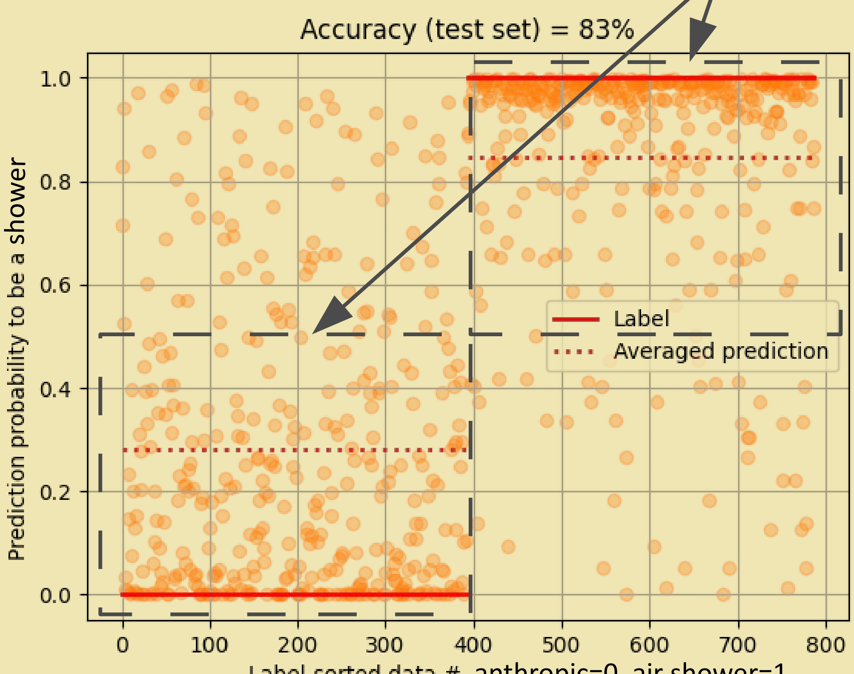
After training

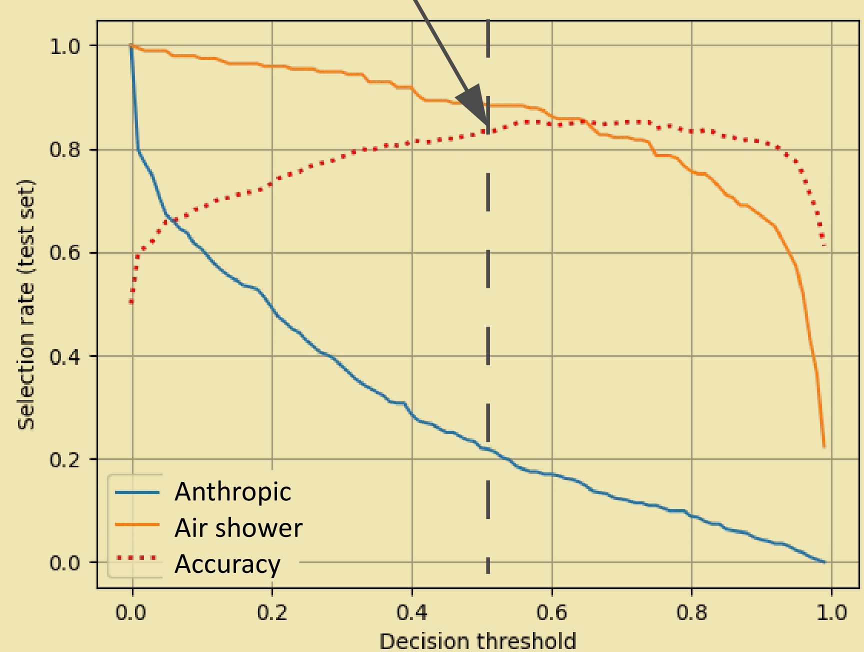
Accuracy = #well classified data / #data
If decision threshold = 0.5 :
accuracy = 83% on training & test sets

Loss = f to minimize, measures how far NN outputs are from labels

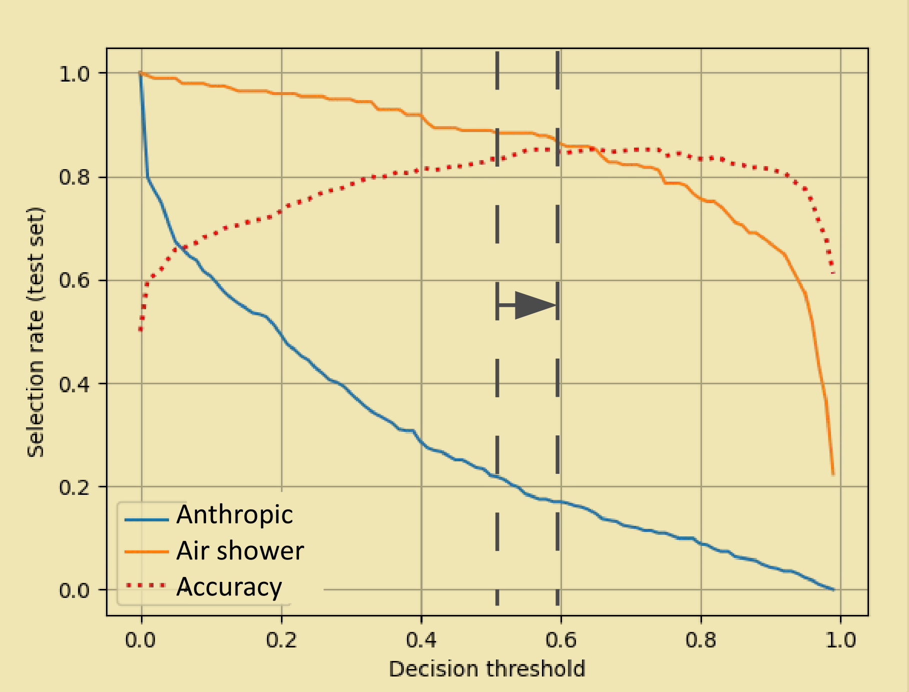








NN inferences on all data



NN is applied to each 9e8 recorded traces to simulate a 'NN trigger':

- NN trigger @ antenna level if :

NN output (probability to come from air shower) > *decision threshold*

- NN trigger @ array level if :

space-time correlations between 5+ antenna NN triggers → recorded

- Recorded coincidences with 'TREND std' (without NN) :

Anthropic	1.25e8
Air shower	370 (simulation estimation)

- Expected recorded coinc. with 'NN trigger' (if we chose a *decision threshold* = 0.6) :

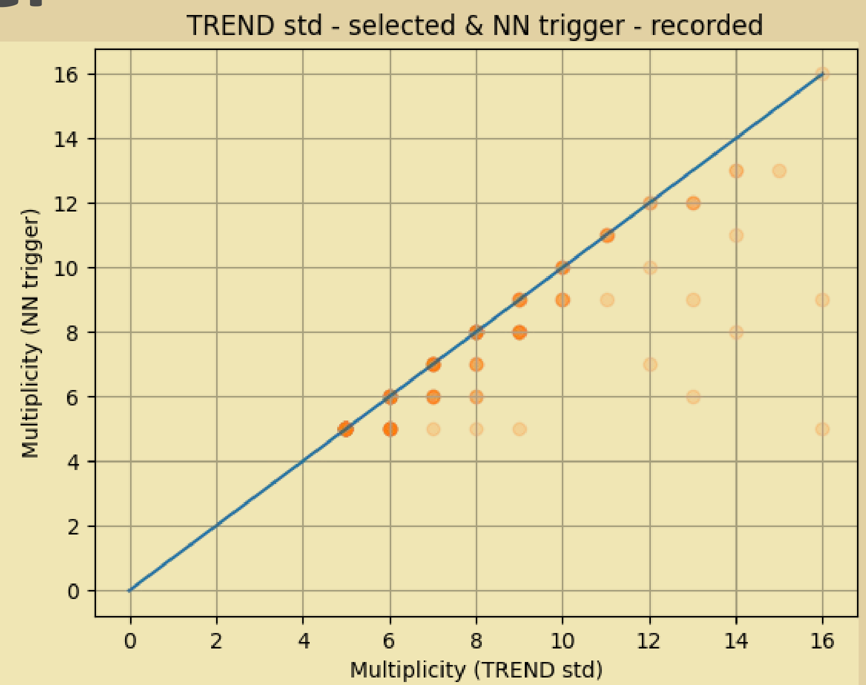
	Fully dep. traces	Indep. traces
Anthropic	18% (22.5e6)	0.07% (87 500)
Air shower	86% (318)	91% (337)

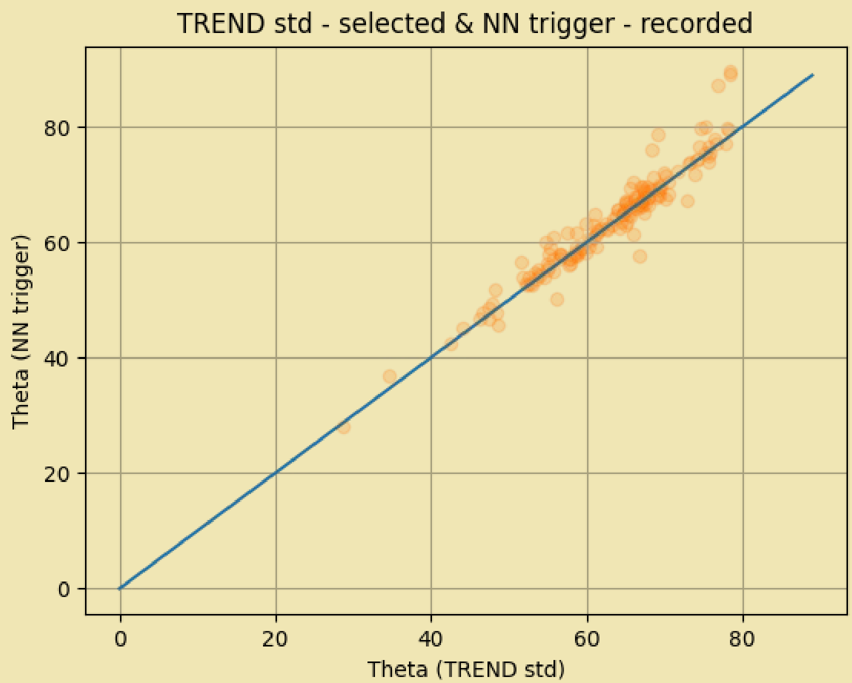
Actual recorded coincidences with 'NN trigger' : 3e6 (2.4%)

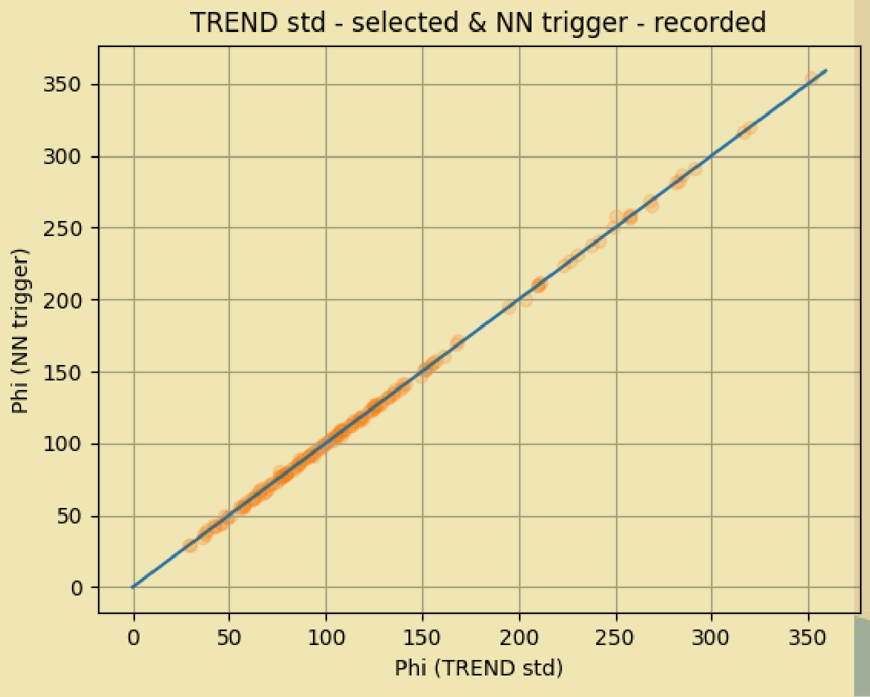
Recorded with NN trigger

selected 'TREND std' (200) ∩ recorded 'NN trigger' (3e6) = 159 coincidences

Loss of 'air shower' coincidences quality ?
<# antenna per coincidence> : -7%
→ $<|\Delta\theta|> < 3\%$
→ $<|\Delta\varphi|> < 1\%$







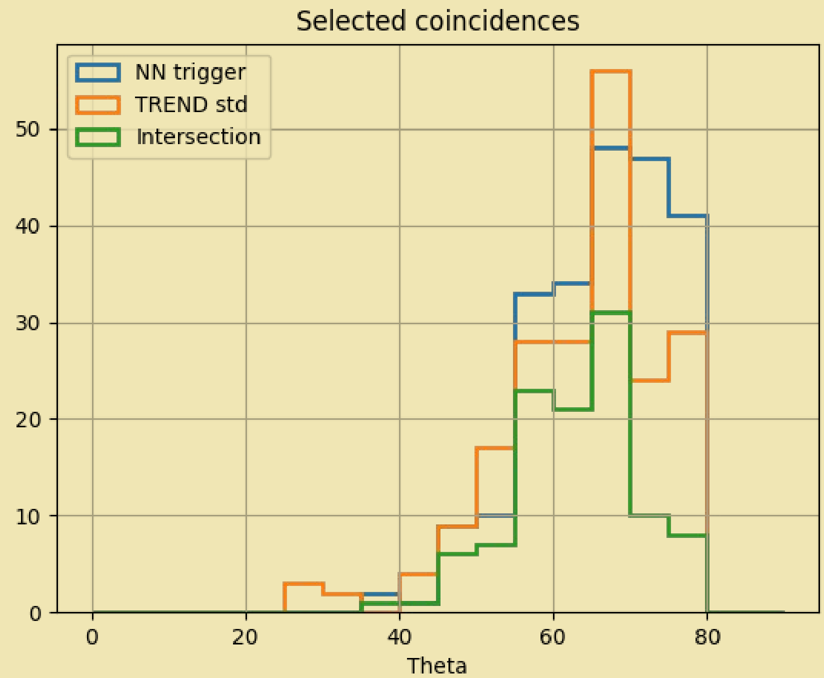
Selected with NN trigger

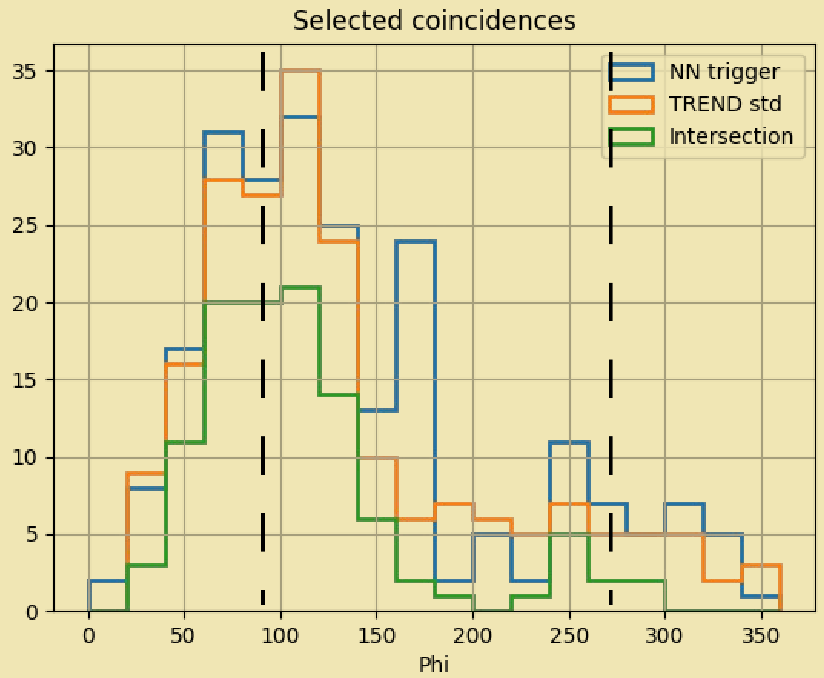
3e6 coincidences recorded

A few simplified offline cuts are applied to reject :

- noisy periods
- angle/source reconstruction : hight χ^2, near source position, zenith>80°
- direction-time correlations between coincidences

225 coincidences selected





selected 'TREND std' (200) ∩ selected 'NN trigger' (225) = 108 coincidences

→ Consistency between selected and 1st order expectations for air showers

→ **NN trigger may at once improve trigger purity & replace some offline cuts**

Coincidences summary

	TREND std		NN trigger
Recorded	1.25e8	∩ 159	3e6
Offline cuts			
Selected	200	∩ 108	225

Selected TREND std (200)

∩

Recorded NN trigger (3e6)

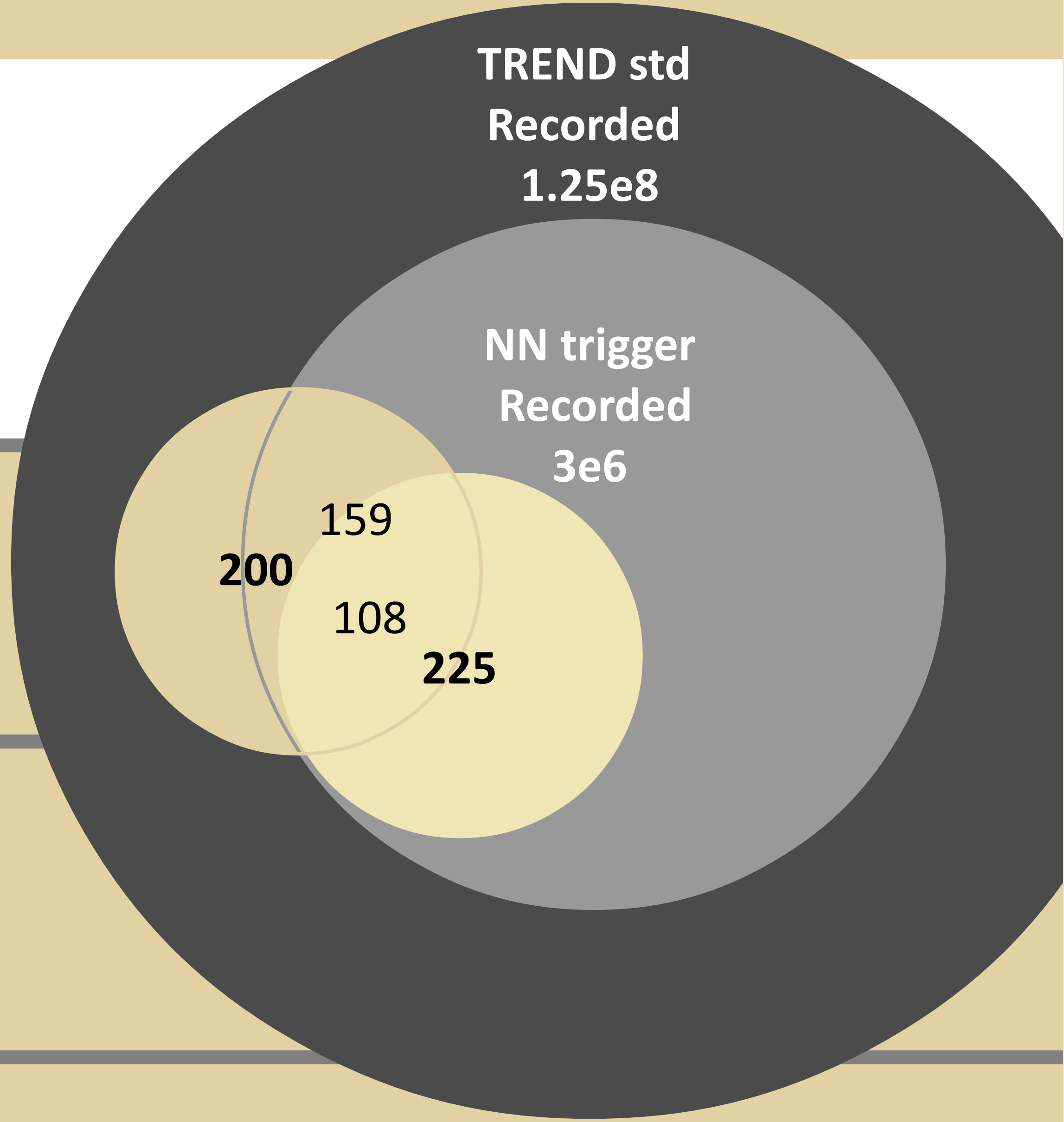
= 159

Selected TREND std (200)

∩

Selected NN trigger (225)

= 108



Neural Network trigger study summary

- Achievements :

Data-driven study

Triggered data : -82% @ antenna level

-98% @ array level

~90% efficiency on air showers

- Limitations :

Training dataset too small, not pure & biased

- Next :

Data from 3-polarizations antennas of GRANDproto300

→ more informations should improve NN performance