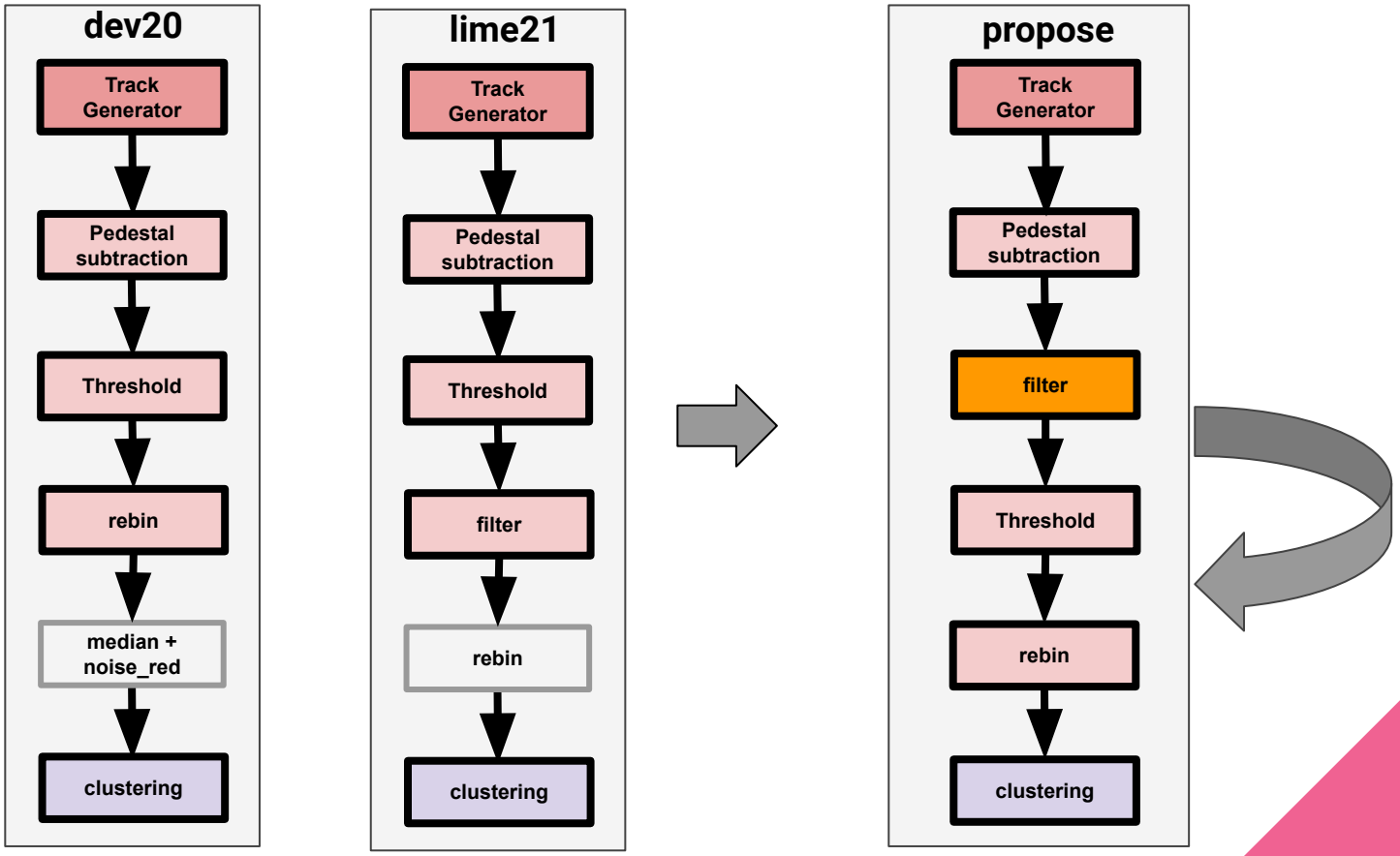
Impact of filtering on Cygno’s images

Guilherme Lopes, Igor Abritta, Rafael Nóbrega

Propose overview

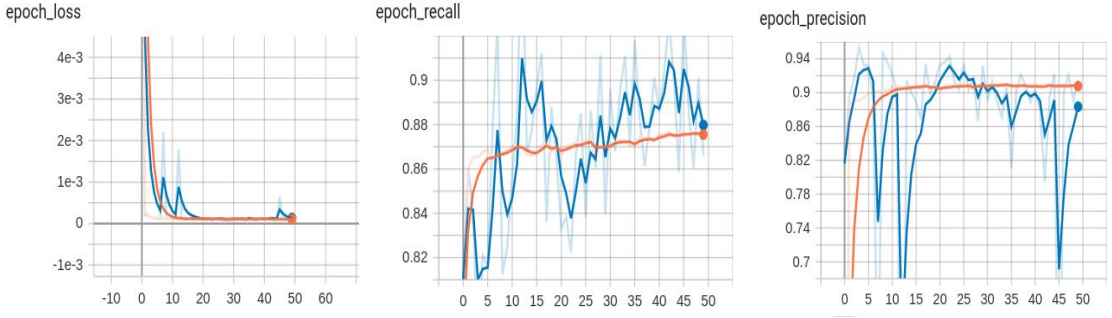
*until April 2021



Setup

- ***Input***
  - ***ER and He from 1 to 60 keV;***
- ***Noise simulation using ecdf algorithm for run 2054;***
- ***Filters used:***
  - ***U-Net***
  - ***Mean using window from 3 to 23;***
  - ***Median using window from 3 to 23;***
  - ***Gaussian using window from 3 to 23;***
  - ***Cygno (n*sigma threshold using std map);***

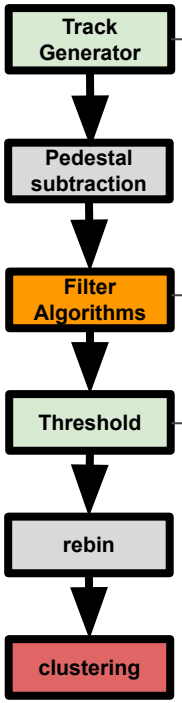
U-Net training



Evaluating

Compare algorithms

Inputs



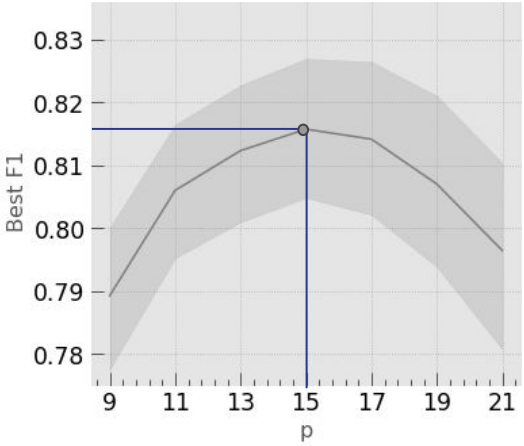
Track (Energy, particle)

from simulation

Filter algorithm

Points after threshold

Choosing filters
F1-score

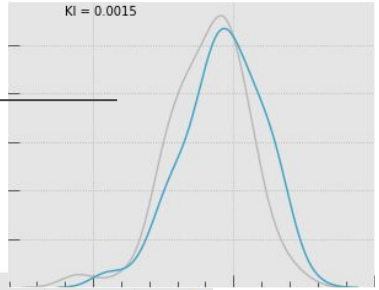


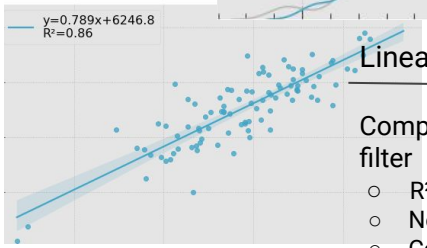
Algorithm	Parameter	Best F1 ($\mu \pm \sigma$)
Median	15	0.820 ±0.127
Mean	17	0.800 ±0.184
Gaussian	17	0.771 ±0.225
Cygno	None	0.137 ±0.127
U-Net	Many	0.873 ±0.060

Energy reconstruction

PDF's similarity

KL-Divergence





Linear model analysis

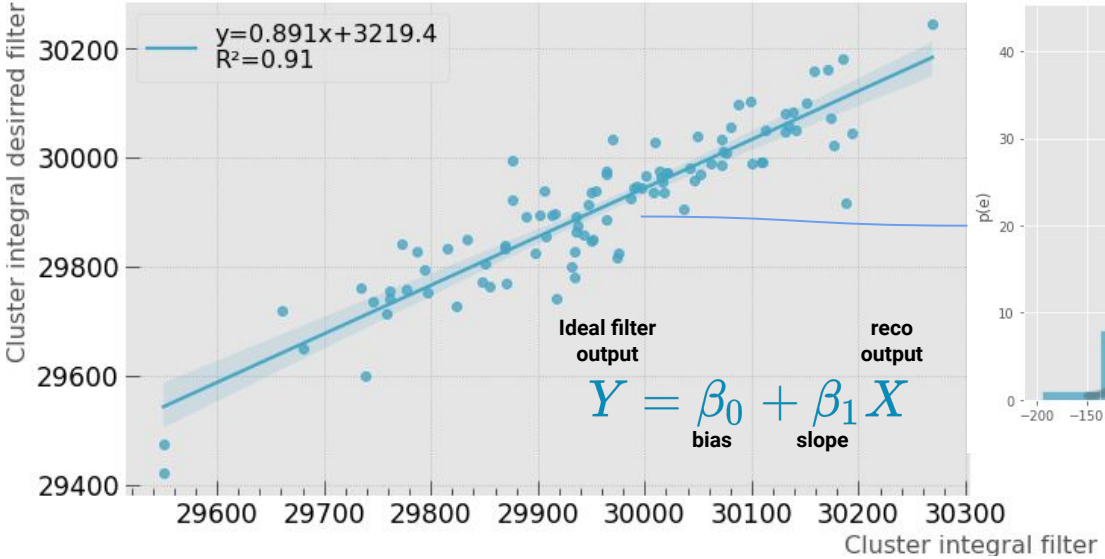
Compare with **ideal** filter

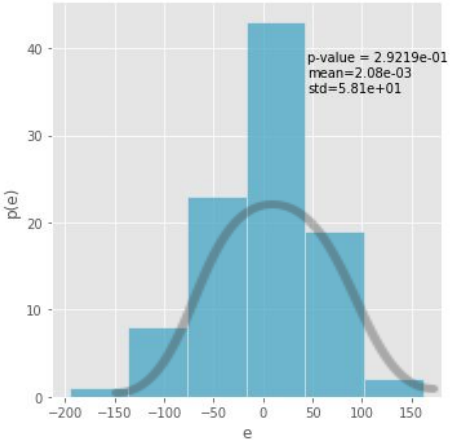
- R^2;
- Normallity test;
- Coeficients;

Simulation results

Evaluation of energy reconstruction

Linearity





Apply normality test to check if error distribution is Gaussian

It will be analyzed only if pass to test (p-value > 0.5)

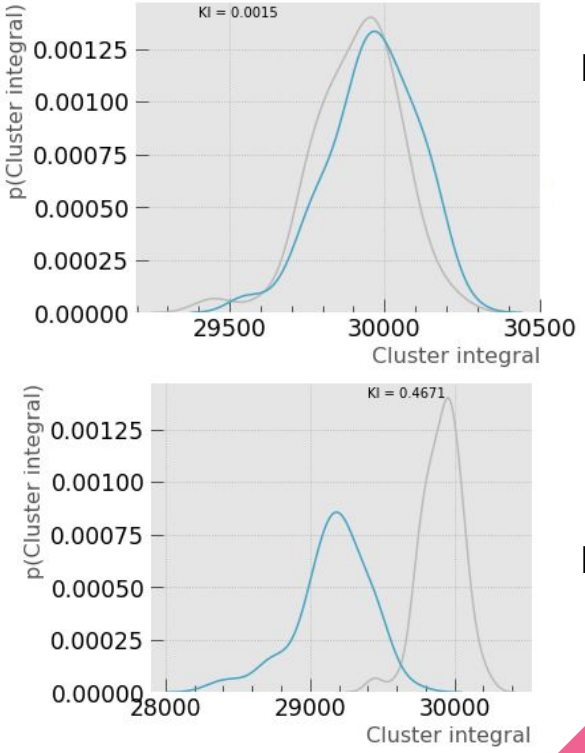
Evaluation of energy reconstruction

Probability density functions dissimilarity

Kullback-leibler Divergence

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$D_{KL}(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx$$



Low

High

$$D_{KL}(P||Q)$$

Compare algorithms

Energy = 3 keV

	energy	particle	filter_name	n_pts	r_square	p_value_err	a	b	kl
66	3	He	median	400	0.939029	0.311357	0.864374	13.720401	0.000508
816	3	ER	median	500	0.911126	0.131334	0.891460	11.555502	0.000261
52	3	He	dev20	30000	0.715373	0.906026	0.627167	49.258771	0.155139
811	3	ER	dev20	200000	0.541629	0.975362	0.384164	63.117425	0.008927
22	3	He	unet	400	0.946940	0.629968	0.870462	9.488900	0.002439
771	3	ER	unet	300	0.918860	0.272915	0.932048	7.014407	0.000233
14	3	He	mean	700	0.931253	0.214432	0.836683	15.175868	0.000431
767	3	ER	mean	700	0.879747	0.265393	0.777244	22.477410	0.000257
48	3	He	gaussian	700	0.936856	0.044891	0.831365	15.745751	0.000501
801	3	ER	gaussian	700	0.874114	0.317448	0.764113	23.874448	0.000423
52	3	ER	lime21	30000	0.761061	0.848375	0.744488	48.838904	0.156627
811	3	ER	lime21	200000	0.576221	0.913299	0.456027	62.579430	0.009013

Compare algorithms

Energy = 6 keV

	energy	particle	filter_name	n_pts	r_square	p_value_err	a	b	kl
400	6	He	median	500	0.985310	0.247963	0.942287	5.662906	0.000087
741	6	ER	median	400	0.915257	0.245449	0.866524	14.234140	0.001802
397	6	He	dev20	3000	0.817429	0.581769	0.731591	46.498102	0.162786
732	6	ER	dev20	200000	0.754846	0.207815	0.503195	50.270701	0.007794
355	6	He	unet	400	0.985794	0.444885	0.979809	1.330613	0.000266
689	6	ER	unet	500	0.893699	0.142038	0.864360	12.419921	0.002052
335	6	He	mean	500	0.982764	0.791792	0.935082	6.274182	0.000101
675	6	ER	mean	400	0.914275	0.067423	0.856766	15.251263	0.001768
369	6	He	gaussian	500	0.981765	0.496541	0.942535	5.376574	0.000062
709	6	ER	gaussian	400	0.909196	0.102328	0.897782	11.074089	0.001376
397	6	ER	lime21	3000	0.877041	0.493618	0.834358	45.974296	0.182883
732	6	ER	lime21	200000	0.809893	0.176327	0.573879	49.704396	0.008756

Compare algorithms

Energy = 30 keV

	energy	particle	filter_name	n_pts	r_square	p_value_err	a	b	kl
245	30	He	median	700	0.974028	0.082776	0.960865	3.867601	0.000041
214	30	He	dev20	70000	0.781361	0.196032	0.557094	45.116781	0.046761
194	30	He	unet	700	0.965520	0.816794	0.968402	2.809840	0.001851
177	30	He	mean	700	0.973122	0.294100	0.959444	3.976287	0.000125
199	30	He	gaussian	500	0.964482	0.104664	0.990610	1.234626	0.001540
214	30	ER	lime21	70000	0.785592	0.163890	0.645301	38.142871	0.047778

Compare algorithms

Energy = 60 keV

	energy	particle	filter_name	n_pts	r_square	p_value_err	a	b	kl
667	60	He	median	700	0.972474	0.076592	0.970193	3.001877	0.000068
645	60	He	dev20	300000	0.817156	0.708603	0.762554	23.584996	0.001060
616	60	He	unet	700	0.972411	0.809420	0.989745	0.930593	0.000606
593	60	He	mean	1000	0.921755	0.300152	0.924704	7.391445	0.000798
633	60	He	gaussian	700	0.973504	0.550620	0.989531	1.074815	0.000088
645	60	ER	lime21	300000	0.826531	0.687086	0.791976	20.566608	0.001127

Qualitative analysis using real data

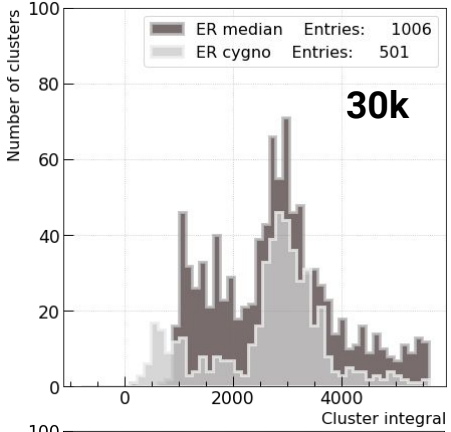
Compare algorithms *(55Fe histograms)*

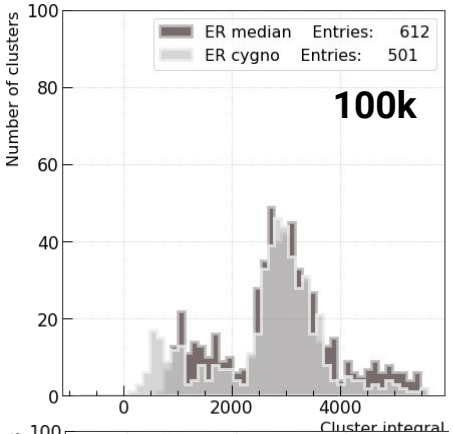
Number of points (Cygno threshold = 1.3 sigma fixed)

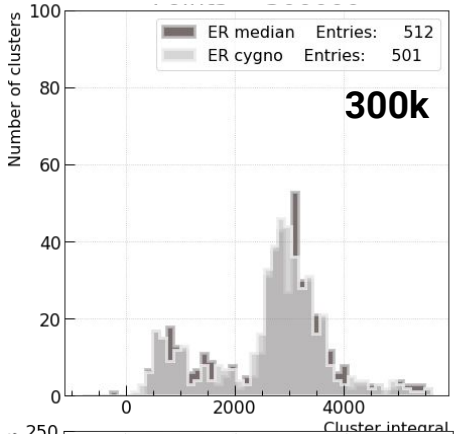
Filter

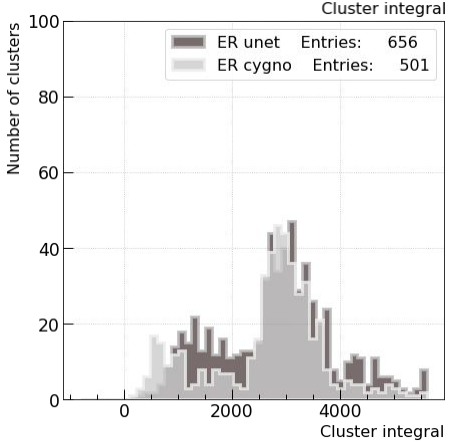
Median

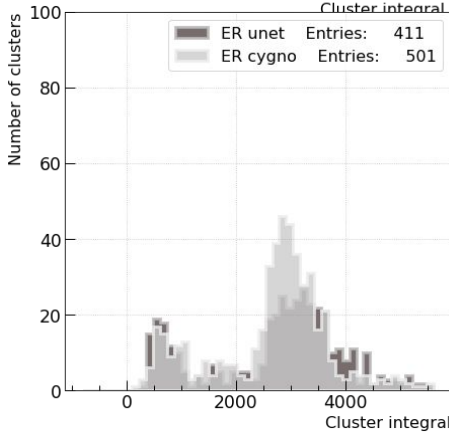
U-Net

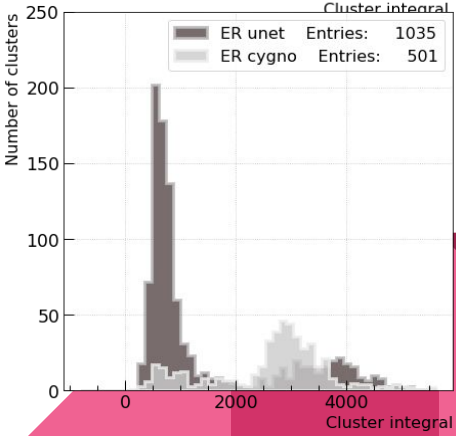












Compare algorithms

Algorithm	Total time (869)(1 core)
cygno	55h 30m
median	**12h 04m**
U-Net (1x GPU P4 16 Gb)	**5h 32m**

~ 4.5x faster
using median

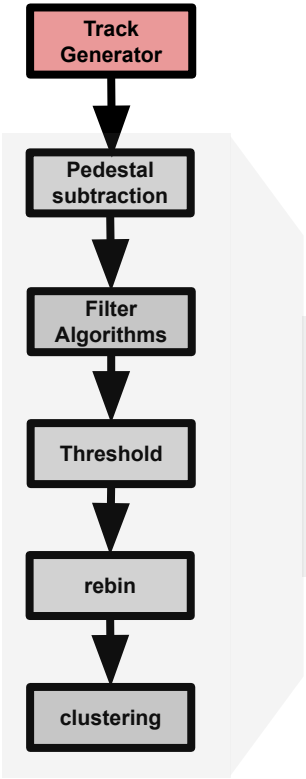
~ 10.0x faster
using U-Net(gpu)

Conclusions and next steps

Resume

Partial conclusions

- ***For the simulation, the filters are able to reduce the number of points sent to the clustering algorithm.***
- ***Filters can improve the processing time of the reconstruction algorithm maintaining the energy distribution similar to the output of the algorithm used by the collaboration in Fe55 region;***
- ***Clustering parameters have been dimensioned for the cygno algorithm***
    - ***optimizing them for the filters might improve their performances.***



Next steps (Under working)

Instance segmentation models (DNN)