Rita Roque acknowledges the FCT PhD studentship (ref. SFRH/BD/143355/2019).

Data information

General conditions

Drift field = 940 V/cm, GEMs = 440 V, transfer field = 2.5 kV/cm.

Dark lab, He-40%CF_4

Z scan

Runs 5861 - 5911, taken on 04/11, at LNF

Exposure time of 50 ms, water cooled

Scan in z with the ^{55}Fe source: z=[5, 10, 15, 20, 25, 30, 35, 40, 45] cm

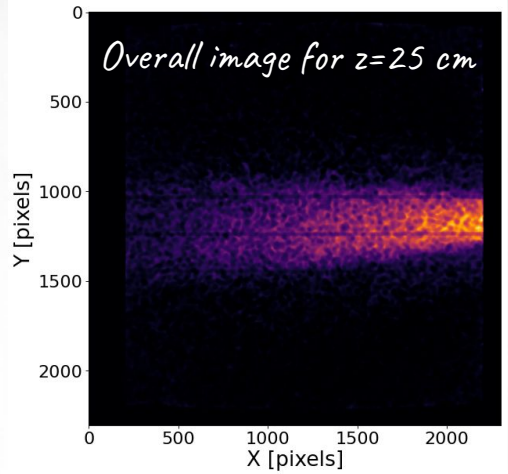
Z + VGEM1 scan

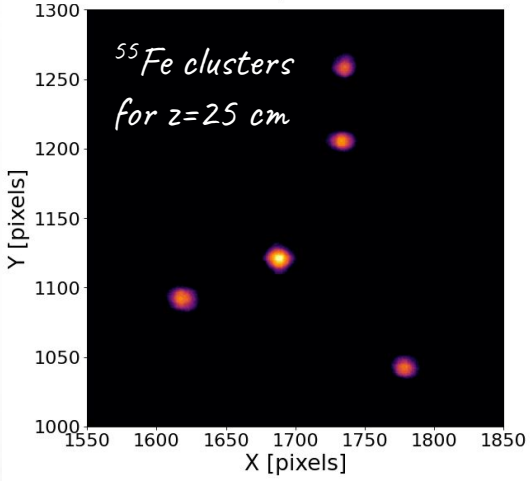
Runs 4432 - 4469, taken on 30/07, at LNF

Exposure of 200 ms, air cooled

Scan in z with the ^{55}Fe source: z=[5, 25, 35, 45] cm

Scan in GEM 1 voltage: V=[320, 350, 386, 406, 420, 431, 440] V

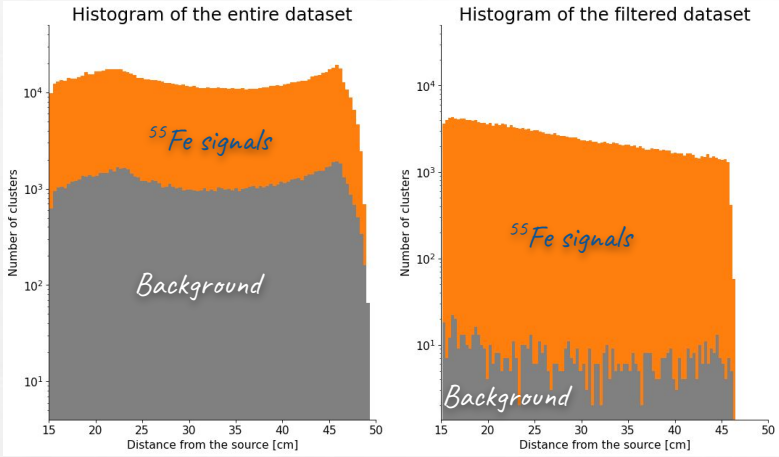


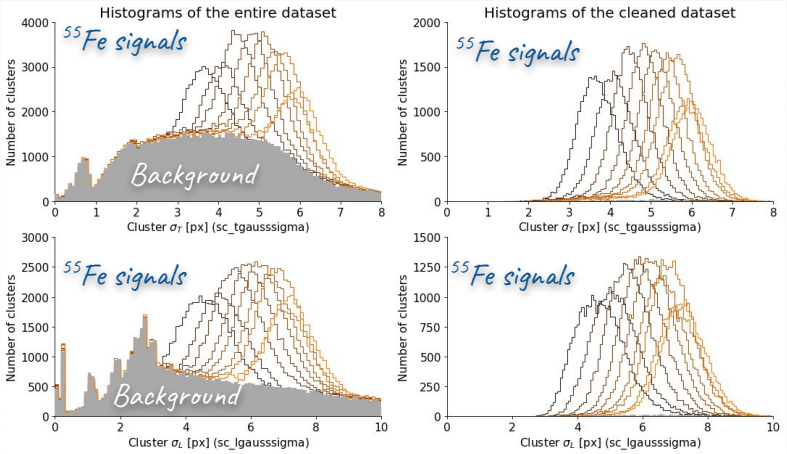


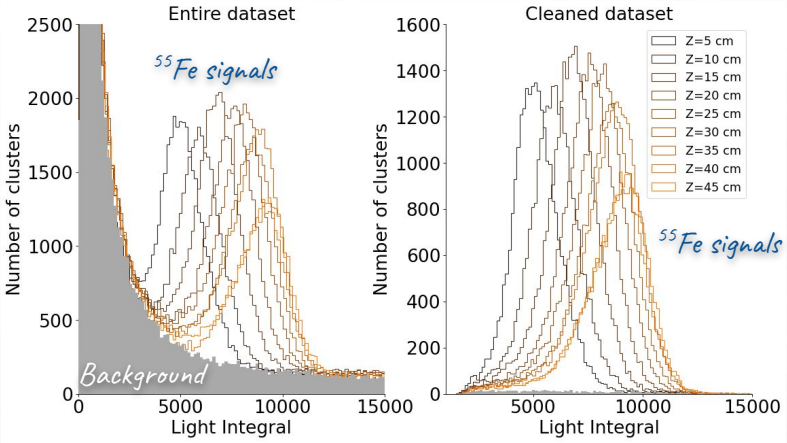
Data Cleaning

Several cuts were applied to the data, to reduce the number of background clusters and improve the SNR: **the original data was reduced to 19.6%.**

All the following analysis were performed with the cleaned data.

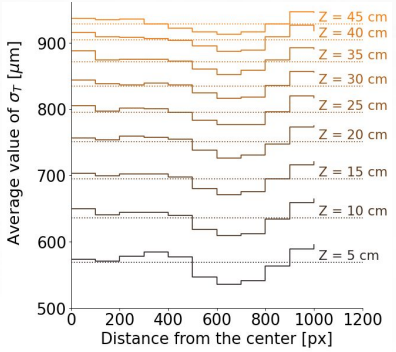


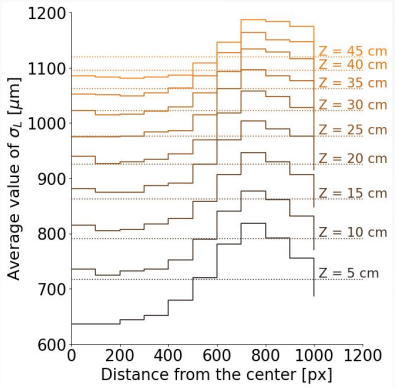


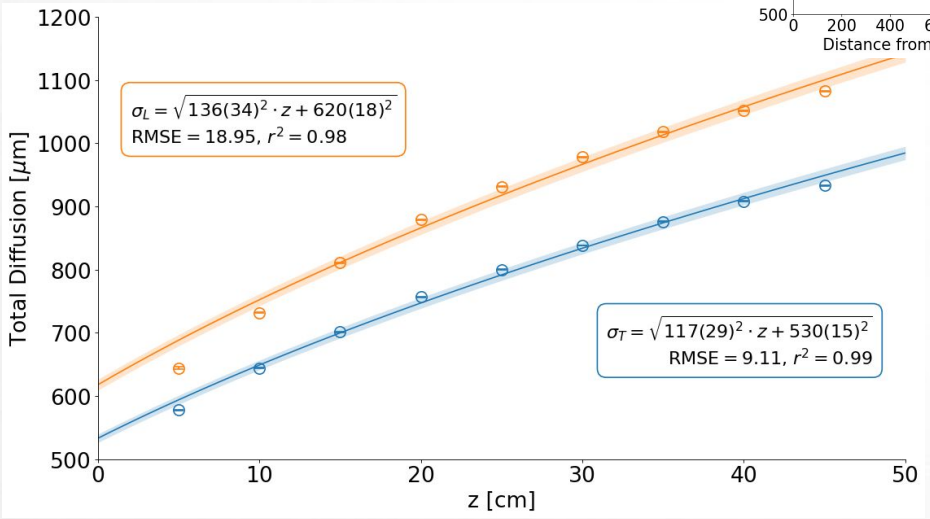


Longitudinal and transverse diffusion

Spatial uniformity study: Clusters located within 400 px from the centre of the image give the most uniform results.







The σ_T and σ_L dependence with z should be the same.

σ_T	LEMOn (MIP)	LIME (^{55}Fe)
D	129.7(31)	117(29)
σ_0	292(12)	530(15)

Absorption Length

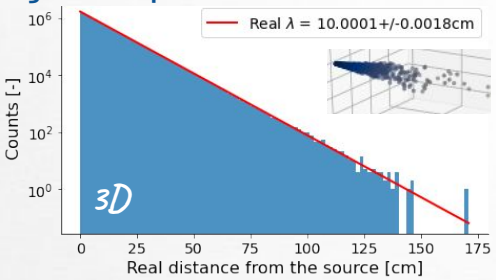
$$d = \sqrt{(x_0 - x_c)^2 + (y_0 - y_c)^2}$$

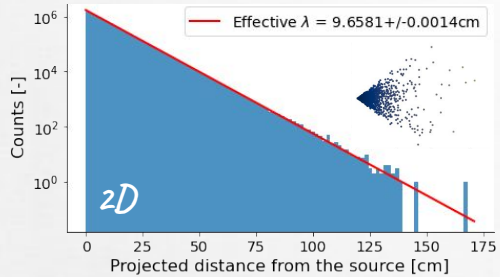
To calculate the λ, we need to calculate the distance between each cluster and the ^{55}Fe source.

Source position: (x_0, y_0, z_0) = (-135, 178.6, z) mm **Cluster position:** (x_c, y_c, z_c) = (sc_xmean, sc_ymean, ?)

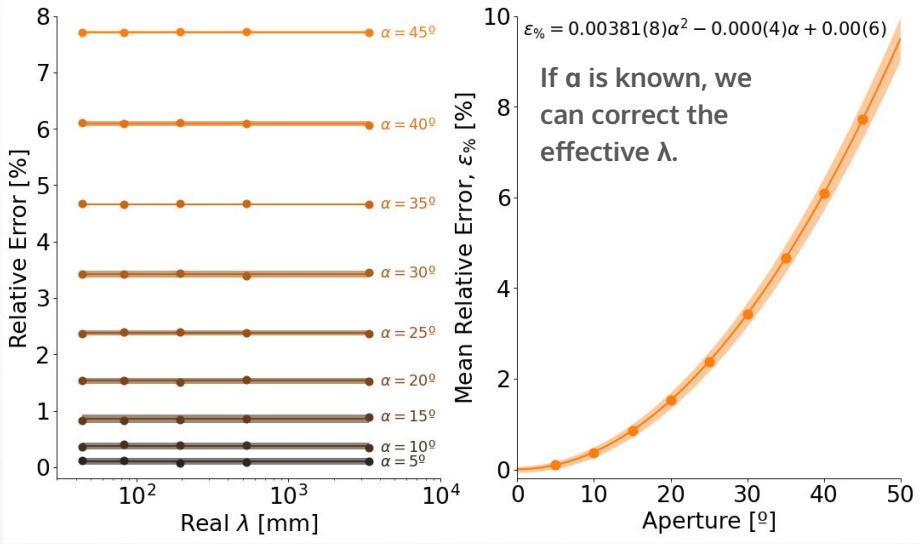
Simple MC study: random clusters where generated in a spherical cap (θ,ϕ), distributed as: A = exp(-ρ/λ)

Eg. for an aperture of 20º:



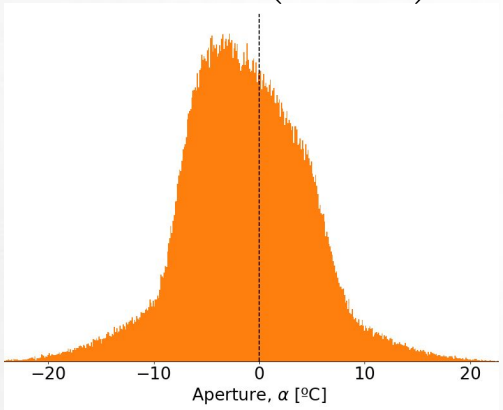


Overall results:

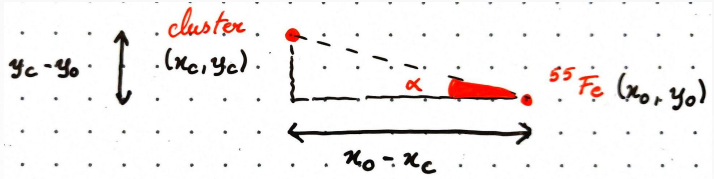


Absorption Length

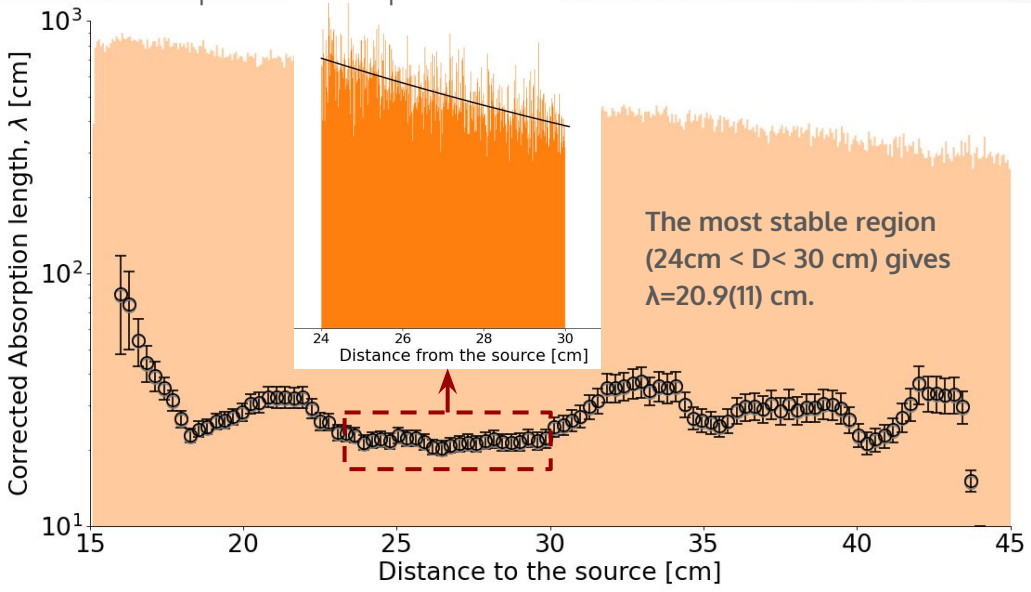
$$\alpha = \arctan\left(\frac{y_c - y_0}{x_c - x_0}\right)$$



Experimental data shows that the aperture of the source is ~20°, corresponding to a relative error of 1.534(14)%.



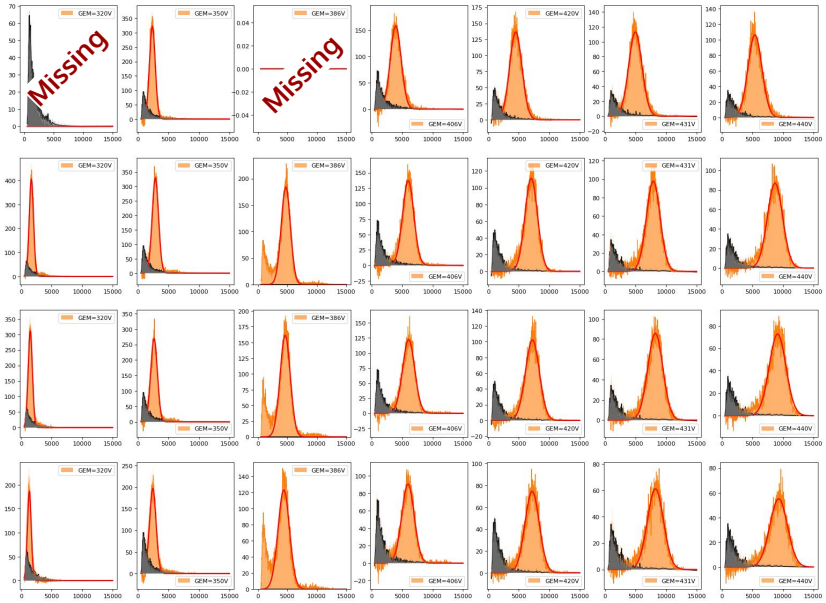
The experimental λ depends a lot on the area that we choose for the fit.

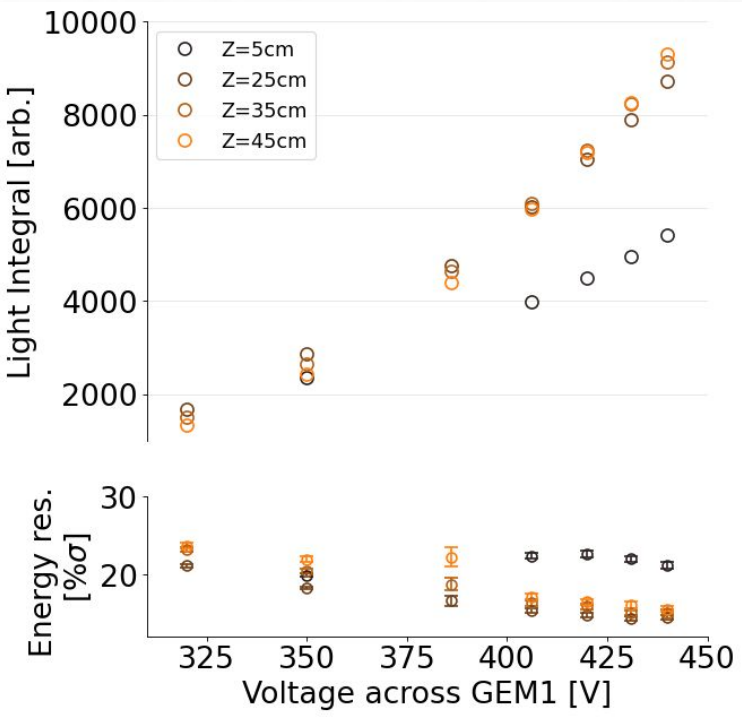


The theoretical λ for 5.93 keV x-rays in He-40%CF_4 is 19.268 cm

Scan across the voltage of GEM1

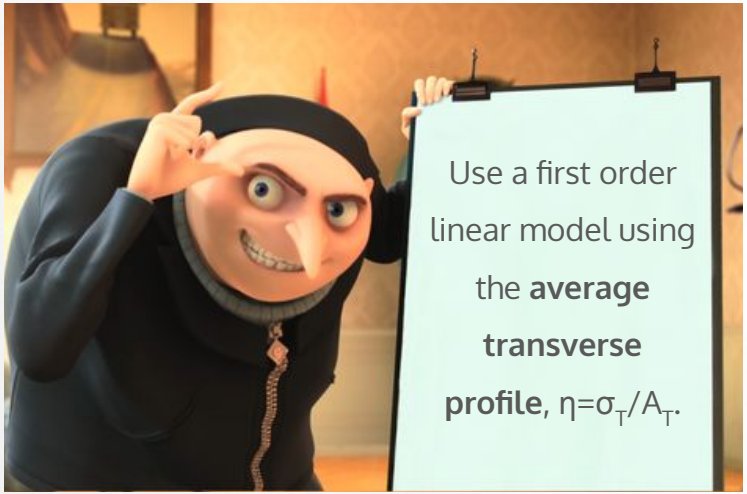
The light integral spectra of each VGEM1 and z was subtracted from the background and fitted to a Gaussian.





Prediction of z

Can we predict the cluster's z using only the 2D camera images?



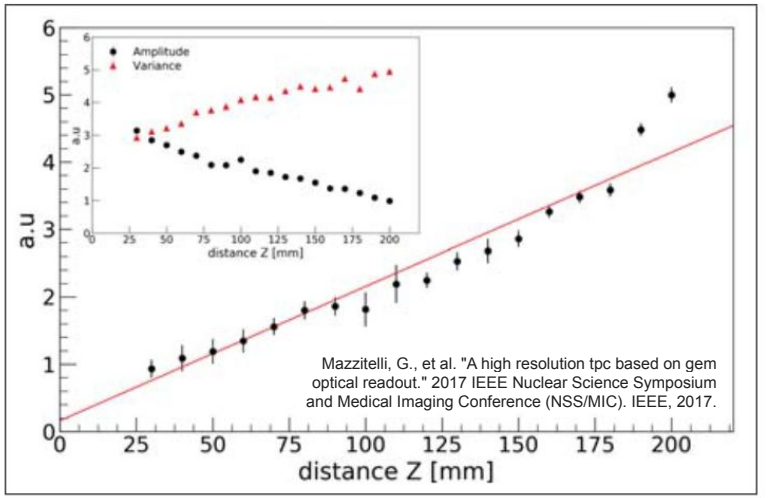
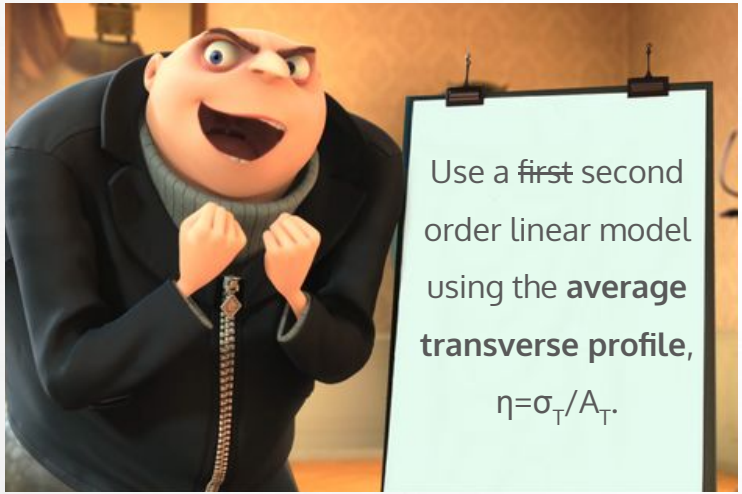
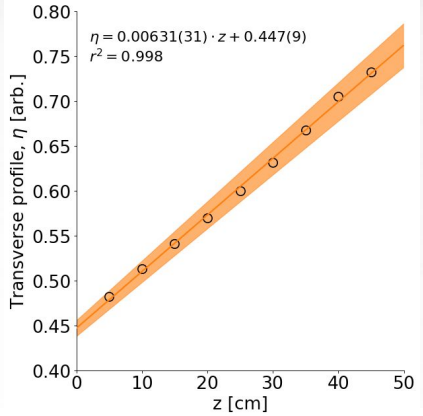


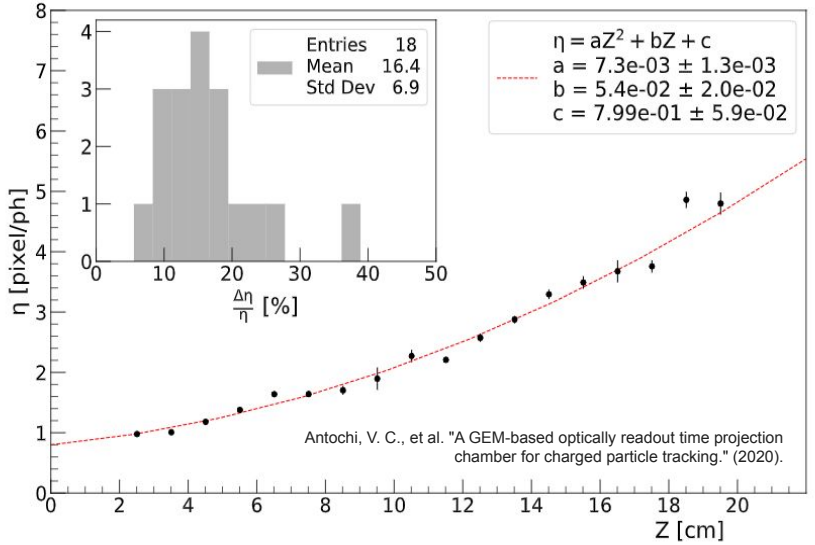
Fig. 9. The quantity, defined as $\eta = \frac{\sigma}{A}$, is a good estimator of the longitudinal particle track position, the expected and measured relative error being $\simeq 20\%$.

Prediction of z

Can we predict the cluster's z using only the 2D camera images?

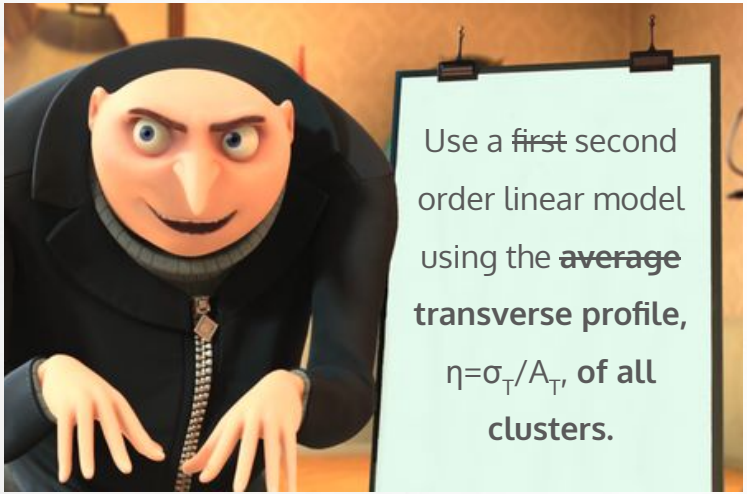






Prediction of z

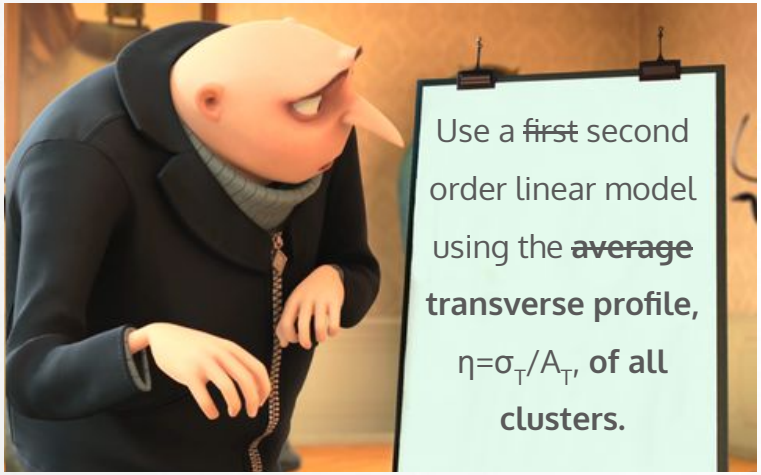
Can we predict the cluster's z using only the 2D camera images?



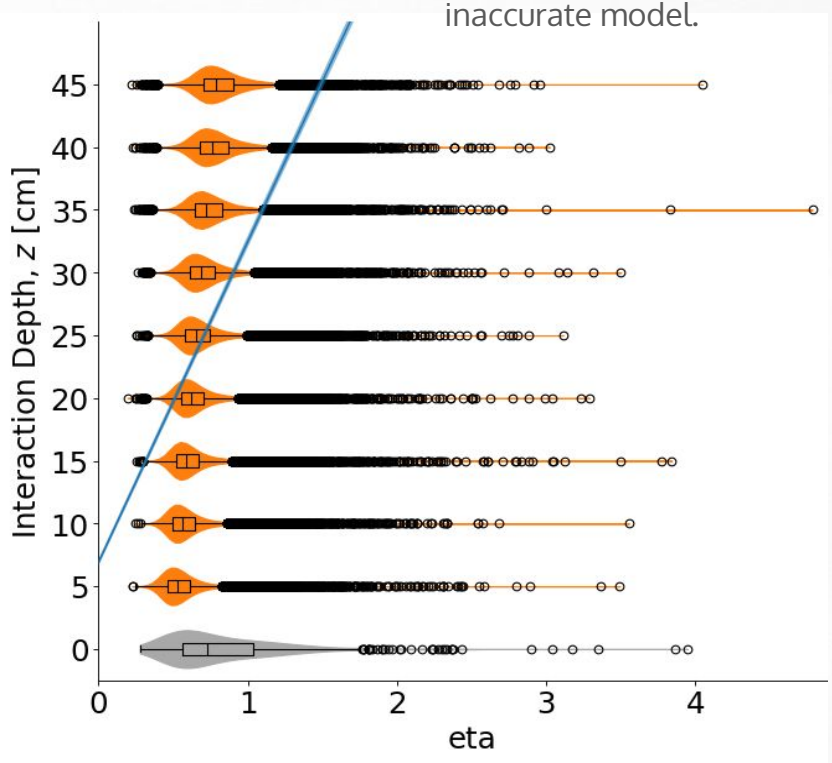
Prediction of z

Can we predict the cluster's z using only the 2D camera images?

Gives a clearly biased and inaccurate model.



	1st order	2nd order
RMSE [cm]	11.25(14)	10.73(24)



Prediction of z: feature engineering

To build a model that is valid for other energies,we discarded the energy-dependent features:

- integral
- corrintegral
- tgaussamp
- size
- nhits
- length

We also discarded quasi-constant features:

- energy
- pathlength
- lstatus
- slimness
- pearson
- tstatus

We included first-order interactions between the original features.

The following showed a promising relation with z:

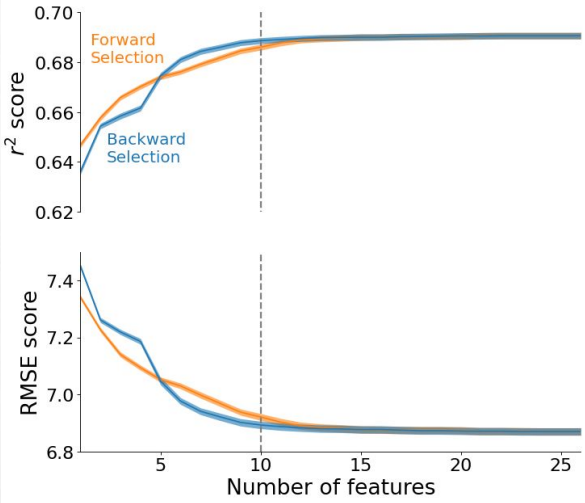
- lgausssigma*tfullrms
- lfullrms*tfullrms
- longrms*tfullrms
- lgausssigma*tgausssigma
- tgausssigma*lfullrms
- tgausssigma*longrms
- tfullrms/sc_rms

Based on f-statistic, r-statistic, and mutual information gain

They are mostly a combination of the longitudinal and transversal shape of the clusters.

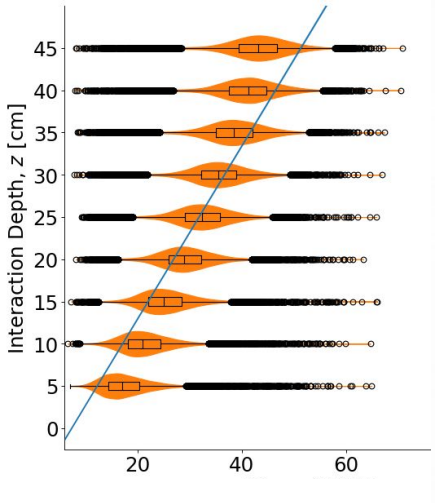
Prediction of z: Linear Models

Feature Selection



Backward Selection shows a good result for 10 features.

Univariate Linear Model



z =-7.52(5)+1.0233 TRMS x σ_L

RMSE = 7.34(13) cm

Multivariate Linear Model

Intercept = 36.5(8)

σ_T x lfullrms = 1.110(12)

tchi2 = 0.1949(24)

RMS = -2.126(29)

η = -7.95(27)

lfullrms = -5.41(9)

lp0fwhm = 1.215(21)

σ_L = 1.95(4)

width = -0.262(7)

lp0prominence = 0.00363(13)

tp0fwhm = 0.530(22)

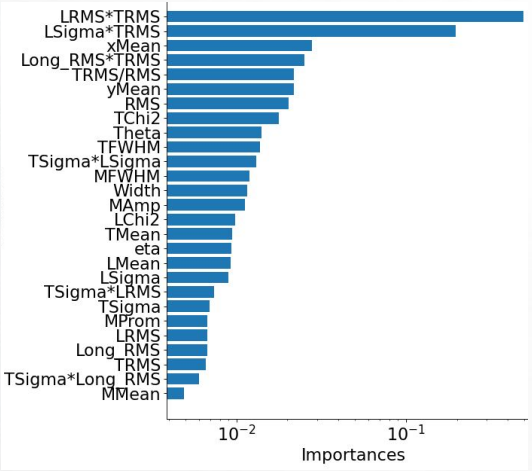
RMSE = 6.89(13) cm

Prediction of z: Preliminary Non-Linear Models

Still to do: feature selection and hyperparameter tuning.

Random Forest

sklearn.ensemble.RandomForestRegressor

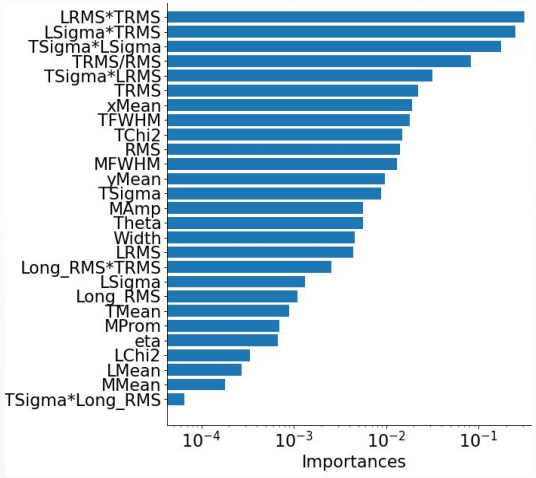


With some optimization:

RMSE = 6.11(7) cm

Boosted Decision Trees

sklearn.ensemble.GradientBoostingRegressor



With default parameters:

RMSE = 6.28(7) cm

Neural Network

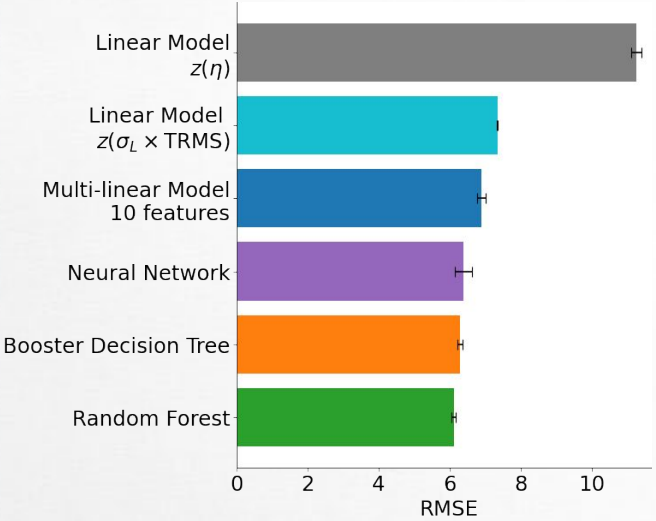
sklearn.neural_network.MLPRegressor

```
Iteration 128, loss = 20.62032726
Iteration 129, loss = 20.50199455
Iteration 130, loss = 20.46869279
Iteration 174, loss = 20.28440027
Iteration 175, loss = 20.22996556
Iteration 176, loss = 20.26293995
Iteration 177, loss = 20.28989738
Iteration 178, loss = 20.27266916
Iteration 179, loss = 20.25847311
Iteration 180, loss = 20.22230141
Iteration 181, loss = 20.26131107
Iteration 182, loss = 20.16544232
Iteration 183, loss = 20.20816011
Iteration 184, loss = 20.18866935
Iteration 185, loss = 20.18491165
Iteration 186, loss = 20.17318603
Iteration 187, loss = 20.20616052
Iteration 188, loss = 20.20560822
Iteration 189, loss = 20.22013667
Iteration 190, loss = 20.17777993
Iteration 191, loss = 20.13649429
Iteration 192, loss = 20.18145923
Iteration 193, loss = 20.13846137
```

With default parameters:

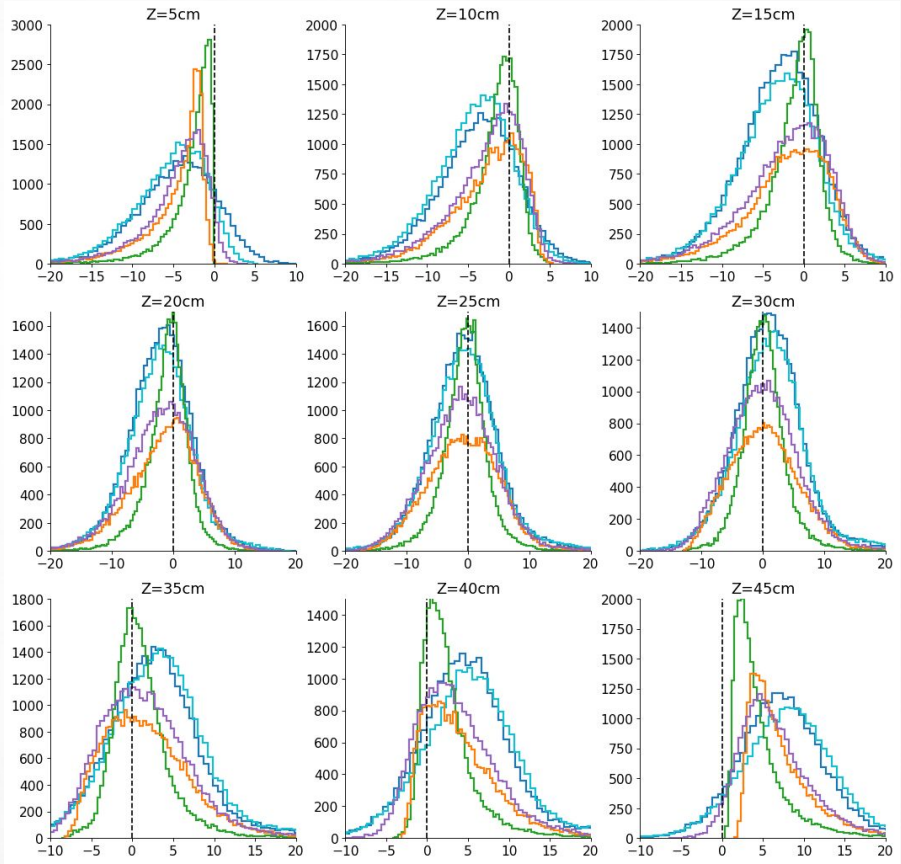
RMSE = 6.39(24) cm

Prediction of z: Performance (preliminary)



Non-linear models have a higher accuracy and lower bias than linear ones.

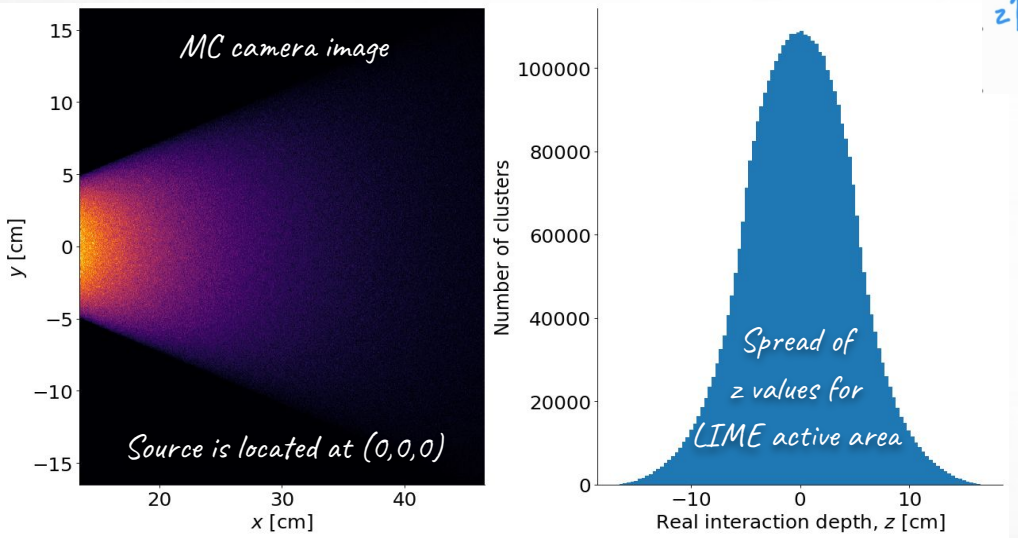
Residual distribution for each z

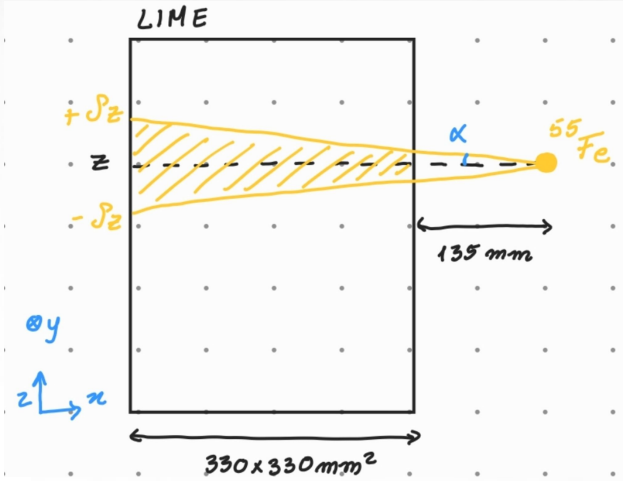


Prediction of z: data limitations

The target for the models, z, is the height of the ^{55}Fe source, but the clusters actually interact in a region z±δz, defined by the source aperture.

MC to estimate δz, for □=20°.





Data Limitations:

λ=	19.268 cm	21 cm
RMSE	4.68 cm	5.10 cm

The lowest RMSE we have now is 6.11(7) cm, for the Random Forest.

Conclusions

Data Cleaning was fundamental for analysing the LIME results:

The computational time was greatly reduced and the background profile became almost insignificant for most fits.

The measurement of the longitudinal and transverse diffusion needs to be improved:

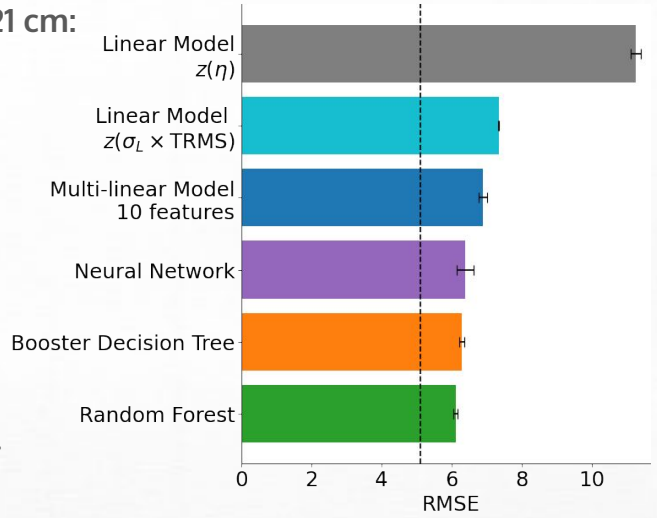
An optical effect is probably affecting the measurement of diffusion in the gas.

The absorption length for ^{55}Fe x-rays in He-40%CF_4 is around 21 cm:

This is close to the theoretical value of 19.268 cm.

We can predict the interaction depth from the 2D images:

Non-linear models have an RMSE ~6 cm, very close to the limits of the data.



Non-linear models still need to be tuned .

The codes are ready: it is just a matter of running time.

Then the models will be tested on MC data and on other energies.