1D → 3D correction

ND-linear model, derivative and RandomForest with FFT
and theoretical consideration

Work in progress

Marian
Input from Ernst, Matthias, Caitie
Goal - compare “physical model” (approximation) with the ML - RandomForest

- current(ROC) → density(drift volume) → E field → distortion/correction
- I → ρ - vdrift ?, epsilon ?
- ρ → E - boundary condition ?

Input MC:
- 3D - Delta of the space charge distortion map
- 3D - Local derivative of the distortion
- 1D - FFT of the space charge density

Output:
- ML model:
  - Linear fits convolution as a linear fit (convolution theorem)
  - Random forest
- Exported variables:
  - Local mean, local median, rms and quantiles of distribution
ML techniques suggested to correct for the density (current, ion drift $v$) fluctuation

- MC not aware of many details (epsilon, $v$ drift) → correction to be data driven

Using linear model as a pre-filter:

- Relying on the white-noise approximation and on the convolution theorem (approximation)
- used as pre-filter reducing linear obvious part of distortion
- **Coefficients extracted from data (not MC)**
- **Too deep RF needed to approximate data with sufficient precision** - see slide 7 reducible and irreducible error estimates

https://en.wikipedia.org/wiki/White_noise

Data driven derivative approximation - FFT 0 - constant term

Reference scaling approximation for constant term

\[
\frac{d\Delta R}{dI_0}(I) \approx \frac{\Delta R}{I_0}
\]

\[
\frac{d\Delta R}{dI_0}(I) \approx k(I) \frac{\Delta R}{I_0}
\]

Could we make ML training easier by subtracting “trivial linear” part?

In first approximation - in low distortion limit, homogeneous epsilon limit, non boundaries effects, derivative of distortion equal to ratio of distortion:

- non linearity, epsilon modification, ... B! = 0
- derivative I and position dependent → could we fit it?
- alternative way of v drift calibration
Data driven derivative approximation - FFT higher harmonics

\[ \frac{d \Delta R}{d I_0} (I) \approx k(I) \frac{\Delta R}{I_0} \int \sin(n d \rho \pi) \]

FFT are independent \(\text{(white noise vector approximation)}\)

E field convolution \(\rightarrow\) multiplication in frequency domain \(\text{(Convolution theorem)}\)

Approximation - response to higher harmonic - integral of the \(\sin(n \omega^* \text{pi} \star \text{drift})\)

- decreasing as \(1/k\)
- additional contribution due electric field convolution \(\rightarrow\) convolution in space \(\rightarrow\) multiplication of fourier coefficients

https://en.wikipedia.org/wiki/Convolution_theorem#:%3A%2F%2Fen.wikipedia.org%2Fwiki%2FConvolution_theorem%23%3A%3A:text=In%20mathematics%2C%20the%20convolution%20theorem%20states%20that%20under%20suitable%20conditions%20the%20Fourier%20transform%20of%20a%20convolution%20of%20two%20functions%20(or%20signals)%20is%20the%20pointwise%20product%20of%20their%20Fourier%20transforms%2C%20More%20generally%2C%20convolution%20in%20one%20domain%20(e.g.%2C%2C%20time%20domain)%20equals%20point-wise%20multiplication%20in%20the%20other%20domain%20(e.g.%2C%20frequency%20domain)%2C%20Other%20versions%20of%20the%20convolution%20theorem%20are%20applicable%20to%20various%20Fourier-related%20transforms.

https://en.wikipedia.org/wiki/White_noise
Distortion expressed as a linear fit response to individual harmonics

- first 20 FFT component used (real, imaginary)
- the relative contribution above 6 - 2 order magnitude smaller than 2 orders of magnitude
- compatible with statistical error - more statistic needed (I used 10^6 points)
- the dependence to be fitted?

```
from sklearn.linear_model import LinearRegression
```
Local probability density function estimator PDF

Book RF regressor

Fit 3 use case (full, LinearDeltBase, LinearDelta)

Extract PDF for RF residuals (RootInteractive wrapper)

export RF stat and PDF (generalized formula under preparation for the ML workshop)

- To interpret mean, median, standard and quantities, some assumptions have to be made -
- in the following case we use 2 different RF with max_depth=None and with max_depth

max_depth=None PDF approximate irreducible +reducible error

max_depth=N PDF approximate reducible error and model error

https://indico.cern.ch/event/1135398/contributions/4764024/subcontributions/370737/attachments/2402314/4114280/fitFFT.ipynb
Linear fits residuals

Linear fits residual compared to RandomForest regressor

3 versions:
- RF full
- RF after FFT0 (mean delta charge) - DeltaLinearBase correction
- RF after FFTn (mean delta charge) - DeltaLinear correction
Mean, median and std of the linear prediction for region at radius<100 cm
- to illustrate bias data points with c1_real<0 and c1_imag < 0 were chosen
Residual histograms definition (FFT1 c1_real<0)

- Mean and median residuals for selected data as function of drift length (0->1)
  - using 20 FFT components - residual (std and bias) significantly smaller

Residuals median in case and std << 0.01 cm even at the region of biggest distortion
Backup
Distortion derivative

in MC distortion derivative - could be expressed from distortion maps (without boundary)
mapping function non linear - as expected
to be considered as approximation in case no derivative available
Variables

Like in current cache tree - FFT coefficients

- + all mean distortions (currently only R)
- + all derivatives of distortion (currently only R)
- mean current