Algorithm for OCDB Archive DB data reduction

Marian Ivanov

Motivation

- DCS data size
  - Enormous
  - TPC example:
    - 4000 temperature sensors, 1Hz update frequency x 4By ~ 60 MBy per hour (typical duration of run)
    - Offline usage – direct access ➔ full data volume in memory
- ➔Compression necessary

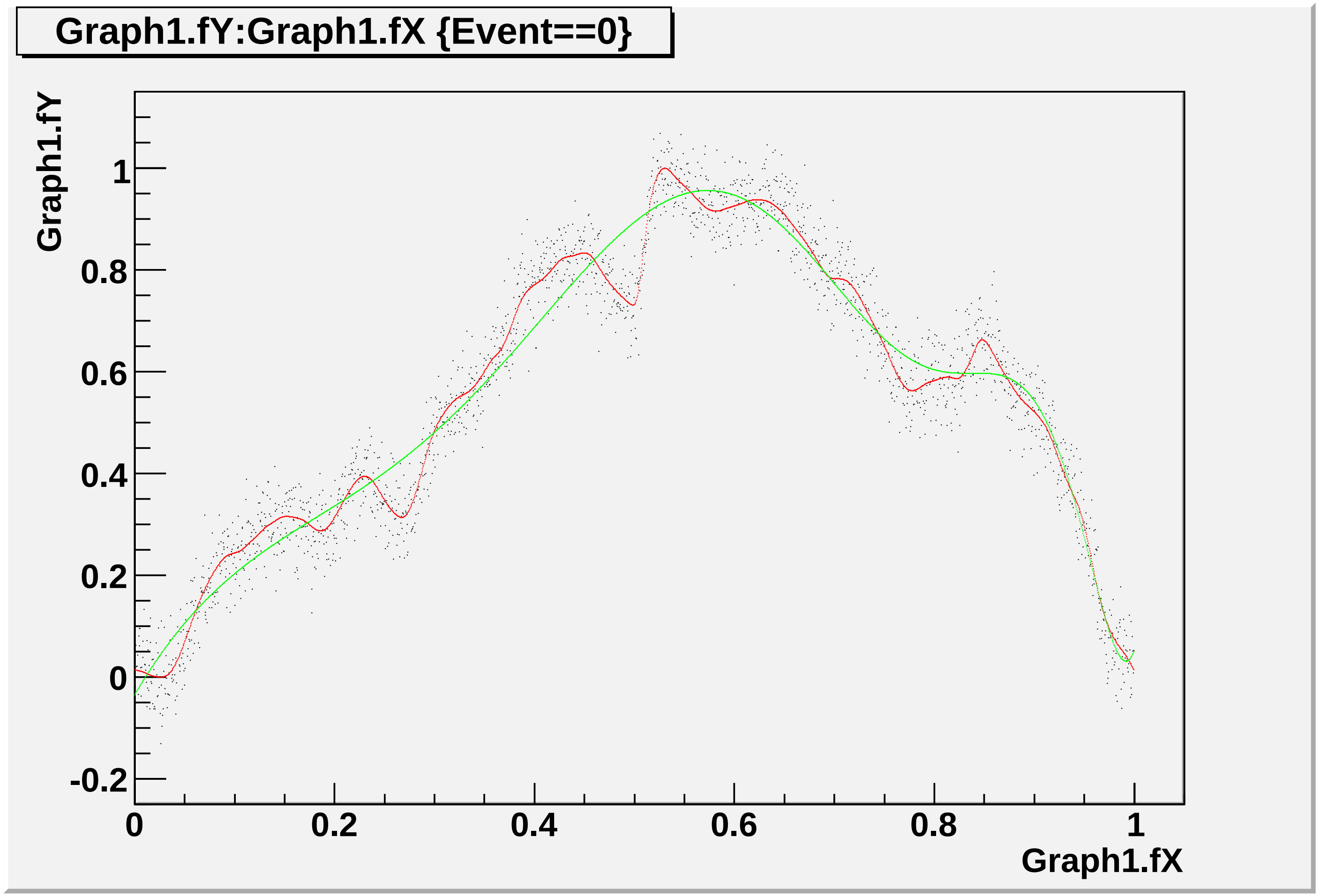
Compression technique

- Option 0
  - Reduce the value size by representing it by Byte, Short (according to required precision)
  - Problem – small reduction factor (2~4)

Compression technique

- Option 1
  - Perform a global fit on the data and use the fitted analytical function
  - Reduction factor ~ ?
  - Problem – unknown analytical behavior of the data (unless we can predict temperature)
    - This unknown function is not very well described by common global approximations i.e polynomial fit (see following slides)

Option 1 – Polynomial fit



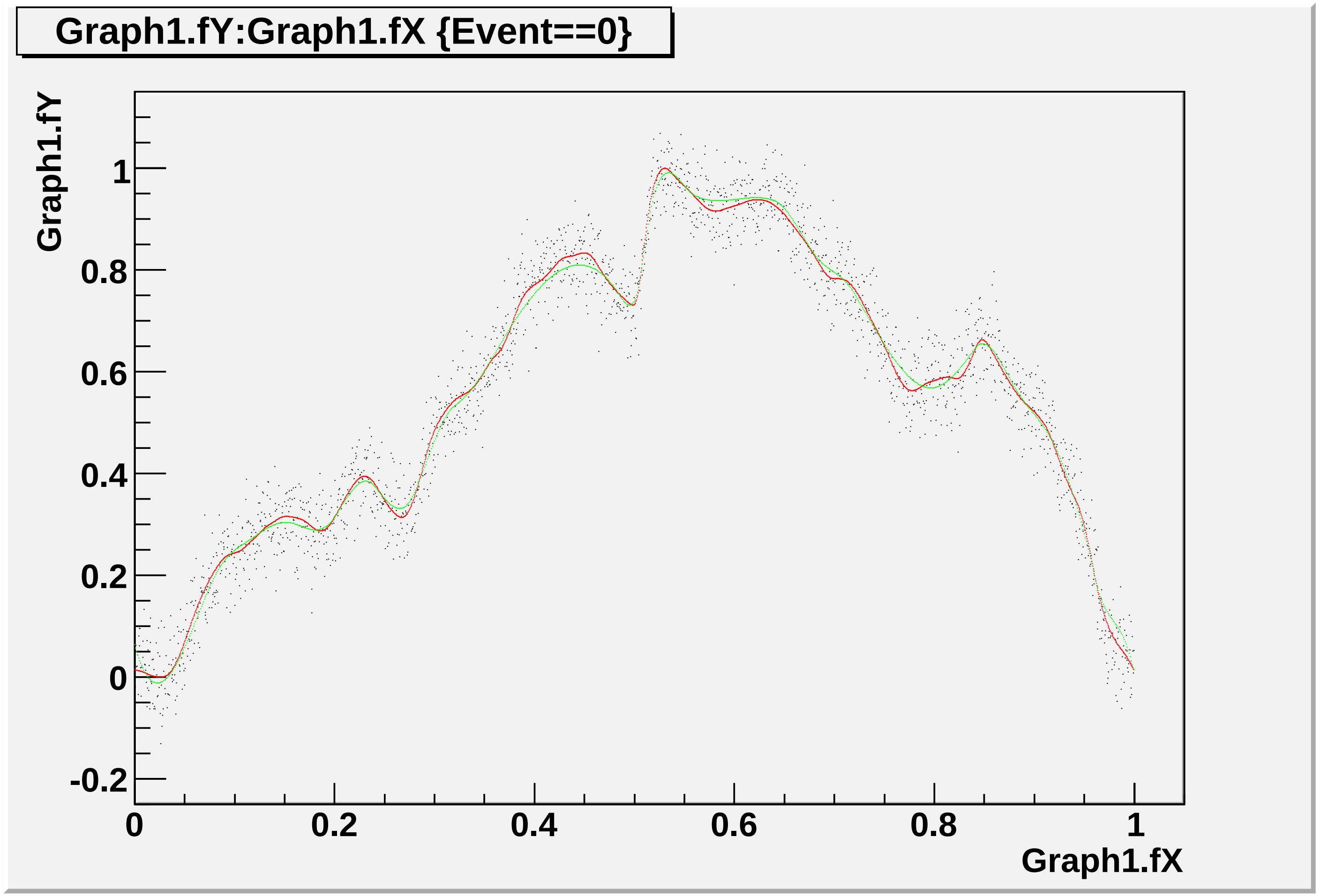
Compression technique

- Option 2
  - Local polynomial regression
  - Fit the data on the fixed size intervals with smoothness conditions on the ends
  - Problems:
    - Very small intervals might be needed to adequately describe data – The noise starts to seriously affect the fit

Compression technique

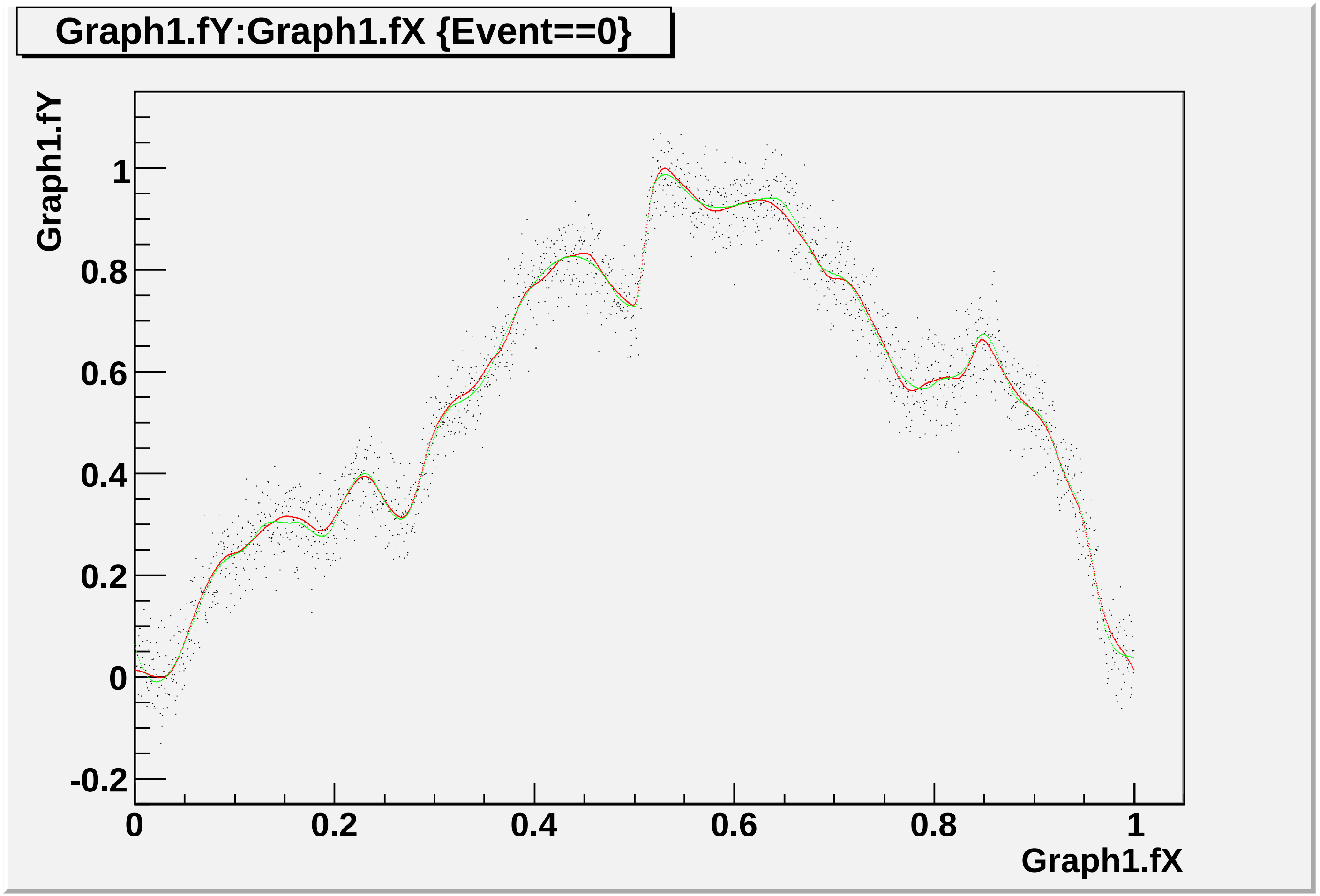
- Option 3
  - Local polynomial regression
  - Fit the data on the variable size intervals with smoothness conditions on the ends
  - The size of intervals is chosen depending on local function behavior
  - Compression factor
    - ~ 100 (for data like typical temperature behavior)

Option 3 – Local polynomial regression



Compression technique

- Option 4
  - Applying stronger smoothness condition
  - Equality of zero and first derivation on the intervals ends - Spline fit
  - Compression factor
    - ~ 100 (for data like typical temperature behavior)
    - Smaller residuals than in option 3


Graph1.fY:Graph1.fX {Event==0}
Graph1.fY
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.2
0.4
0.6
0.8
1
Graph1.fX

AliSplineFit

- Calculation of the intervals according desired precision
  - InitKnots(TGraph * graph, Int_t min, Double_t maxDelta);
  - SplineFit(Int_t nder) – fit of the data with spline (the last smooth derivation)
- Eval(Double_t x, Int_t deriv=0) const;

Performance

Option	Sigma of residuals
Option 1 - Polynom n	0.077
Option 3 – local regression	0.012
Box Kernel smooth	0.025
Option 4	0.008

Conclusion

- Local polynomial regression to compress and fit OCDB data implemented
  - Tested on the “simulated” data
  - Soon on CVS
- Next step
  - Use real calibration data