Updates pT Training: KNN

Wei Shi

EMTF Working Meeting

Euclidean distance

x: input variables(dimension d) of train event
y: input variables of test event

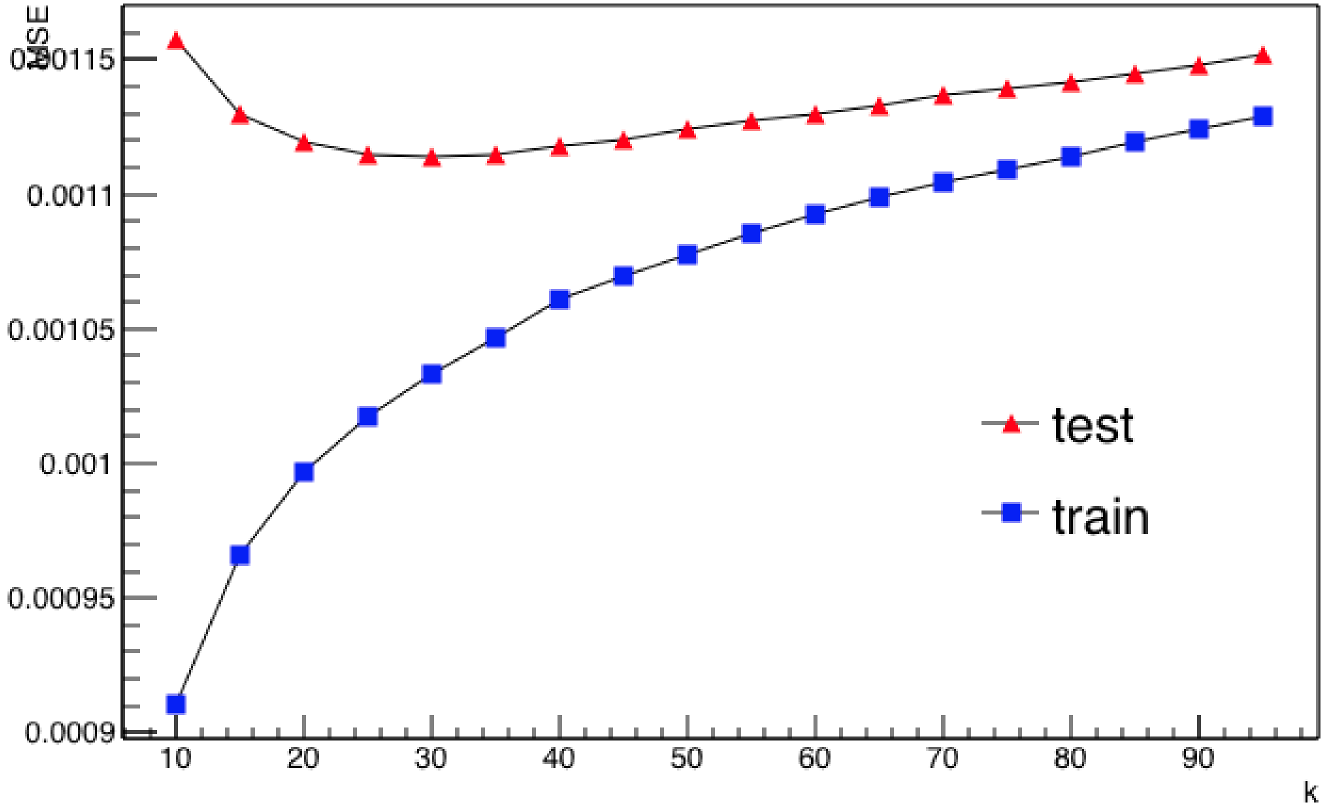
Mode: 15
Input variables: dphi12, dphi23, dphi34
Target variable: 1/Gen pT

$$R_{\text{rescaled}} = \left(\sum_{i=1}^{d} \frac{1}{w_i^2} |x_i - y_i|^2 \right)^{\frac{1}{2}}$$

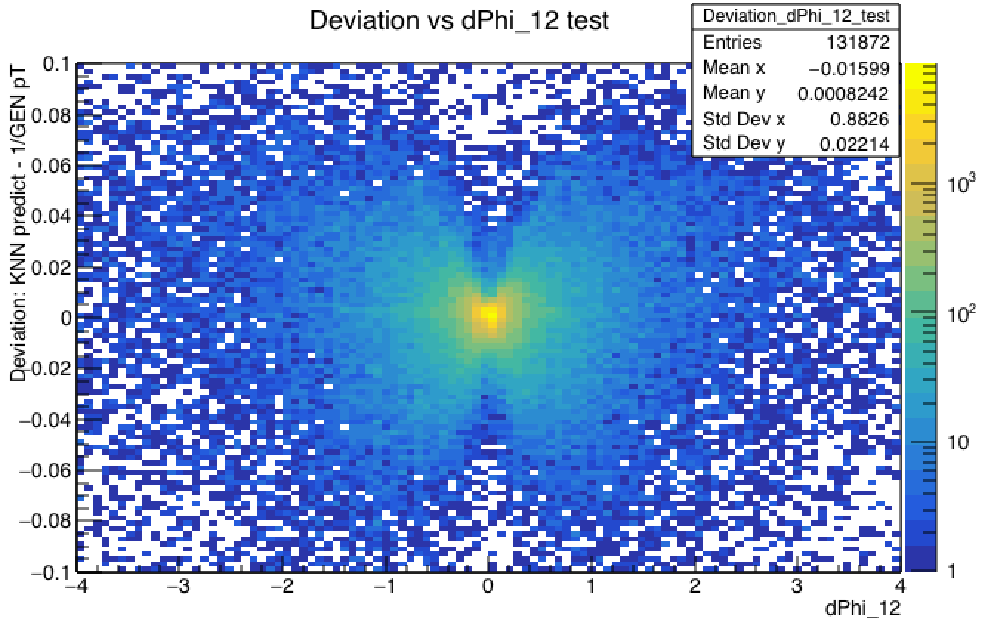
$$\langle t(i,V) \rangle = \frac{\sum_{j \in V} w_j t_j f(\text{dis}(i,j))}{\sum_{j \in V} w_j f(\text{dis}(i,j))}$$

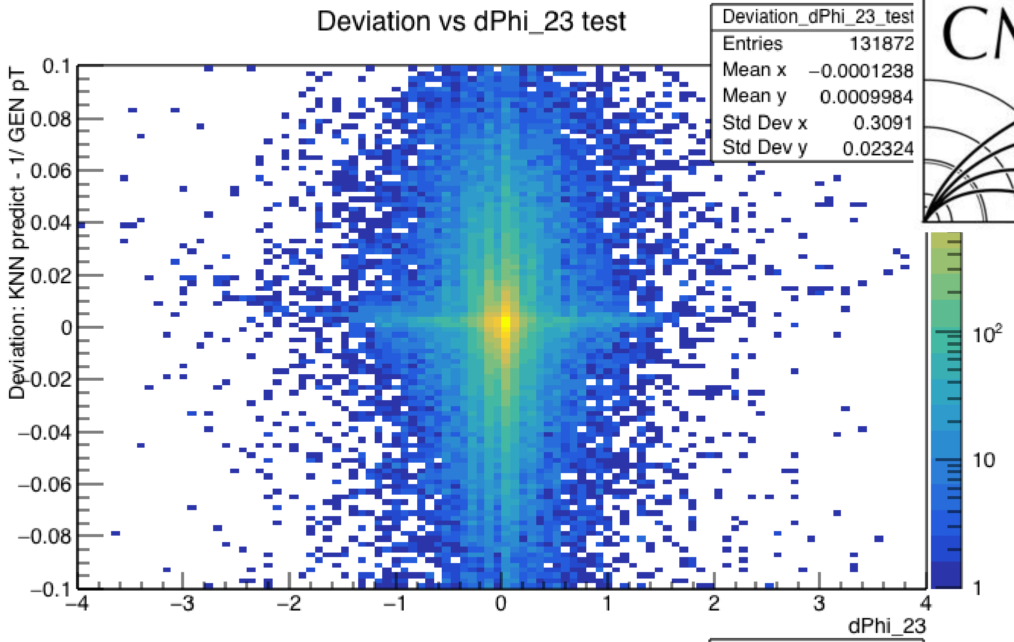
"nkNN=30:ScaleFrac=0.8:SigmaFact=1.0:Kernel=Gaus:UseKernel=F:UseWeight=T:!Trim"

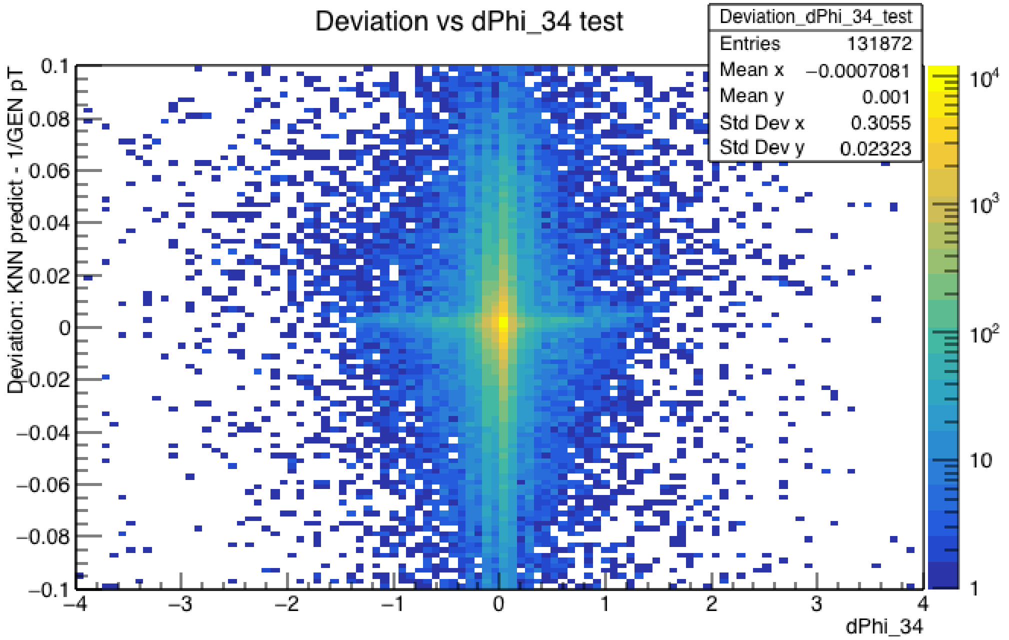
MSE = Deviation^2/N

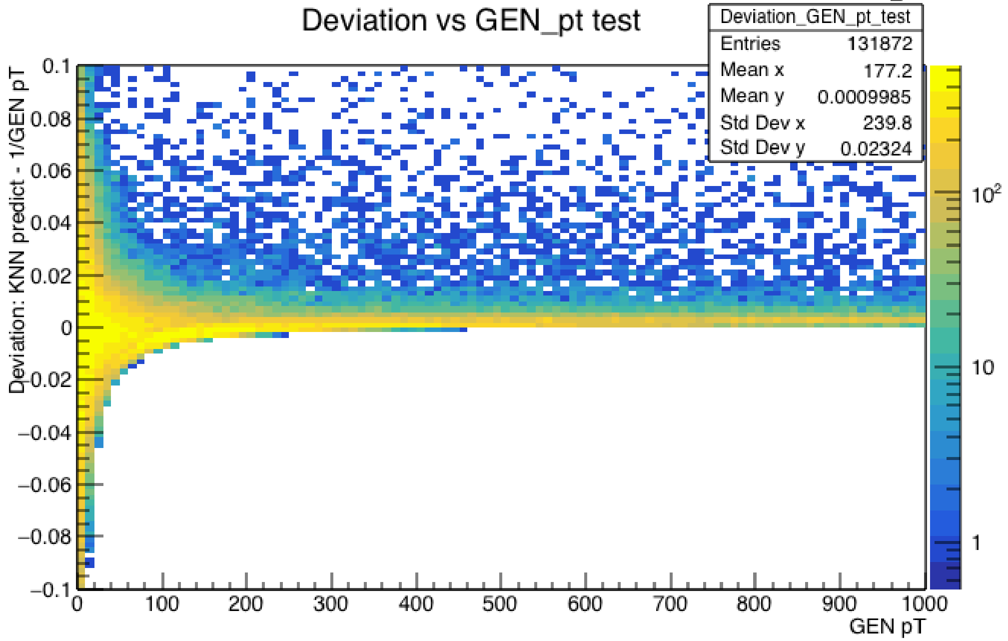


/afs/cern.ch/work/a/abrinke1/public/EMTF/Analyzer/ntuples/EMTF_MC_NTuple_SingleMu_noRPC_300k.root

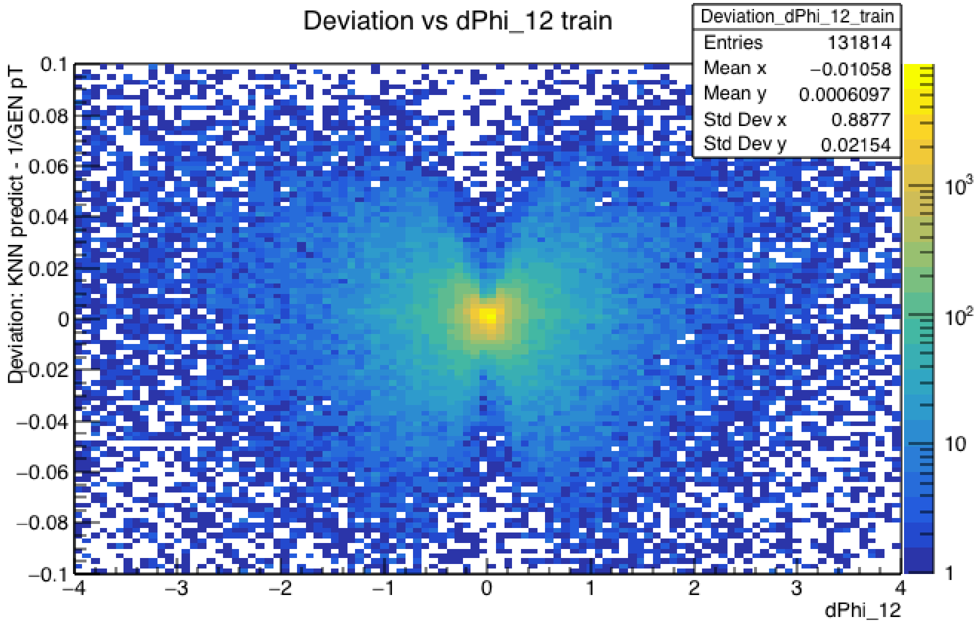


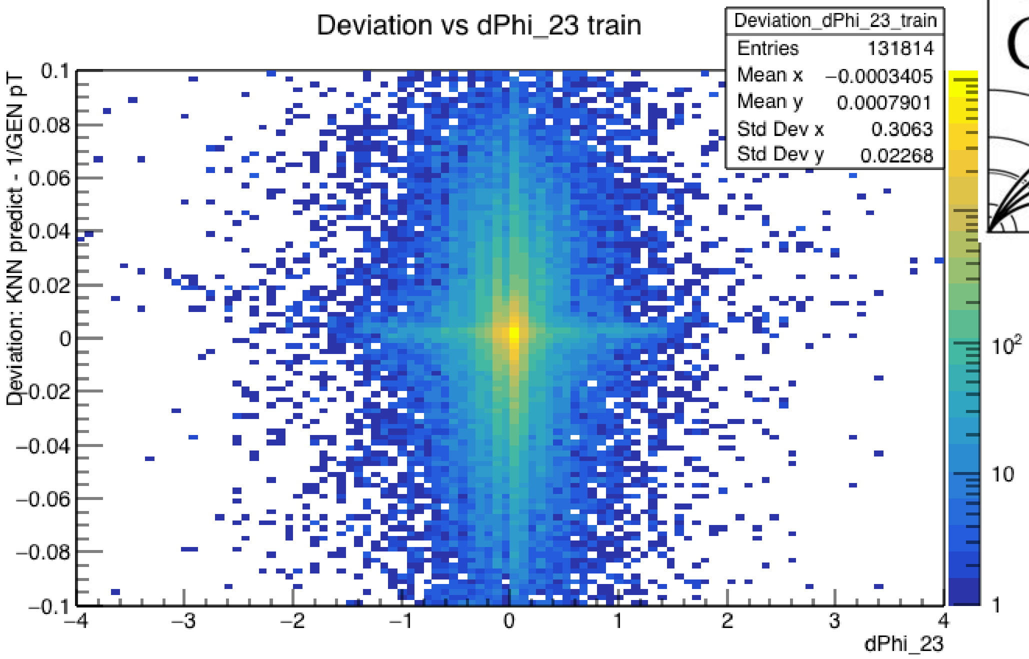


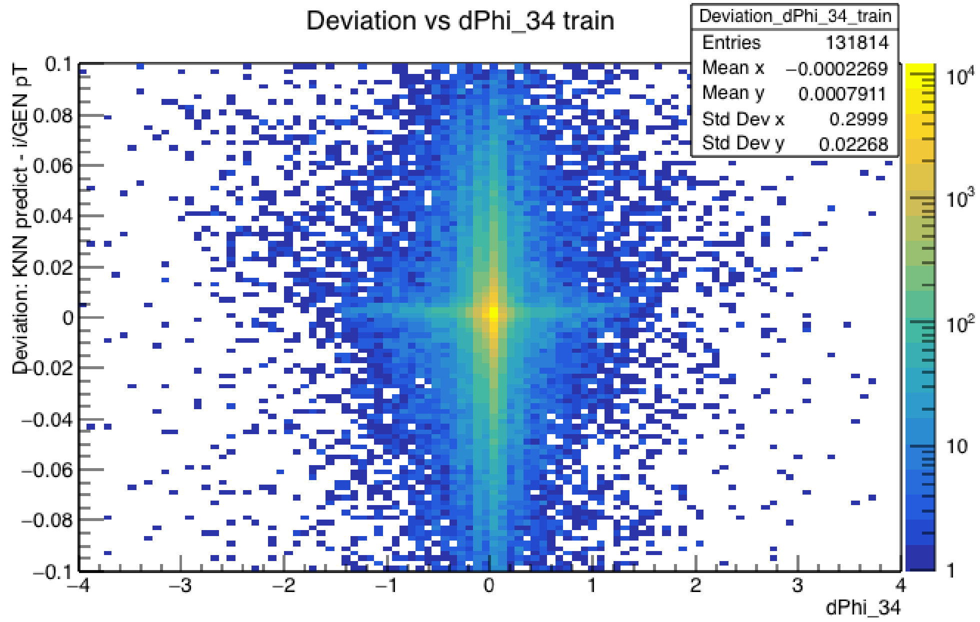


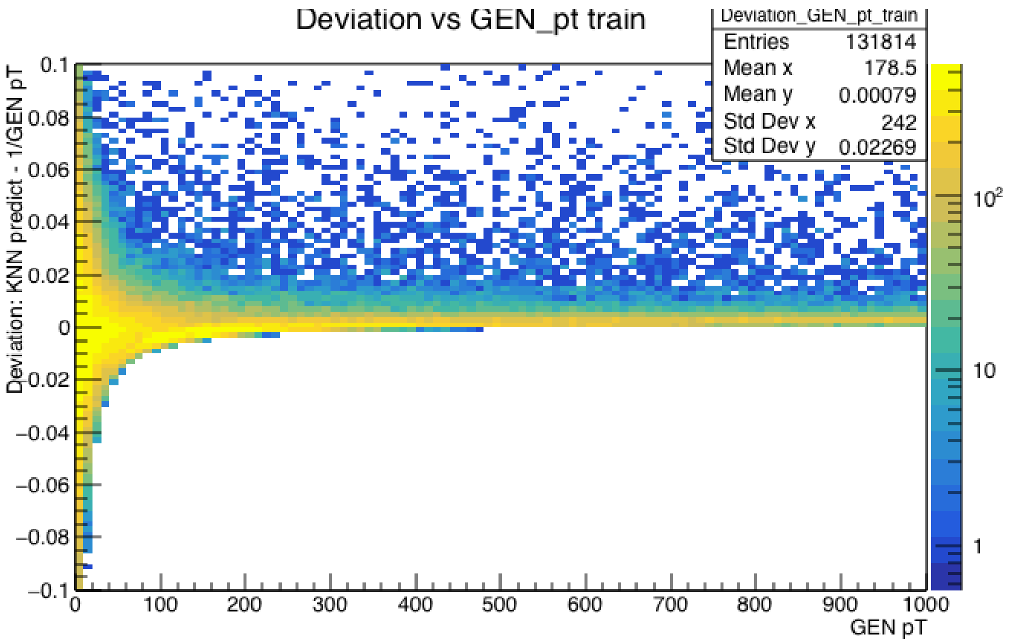


Last plot shows predicted pT tends to be smaller than GEN pT, for large pT


Deviation vs dPhi_12 train
Deviation_dPhi_12_train
Entries 131814
Mean x −0.01058
Mean y 0.0006097
Std Dev x 0.8877
Std Dev y 0.02154
Deviation: KNN predict - 1/GEN pT
dPhi_12


Deviation vs dPhi_23 train
Deviation_dPhi_23_train
Entries 131814
Mean x −0.0003405
Mean y 0.0007901
Std Dev x 0.3063
Std Dev y 0.02268
Deviation: KNN predict - 1/GEN pT
dPhi_23


Deviation vs dPhi_34 train
Deviation_dPhi_34_train
Entries 131814
Mean x −0.0002269
Mean y 0.0007911
Std Dev x 0.2999
Std Dev y 0.02268
Deviation: KNN predict - i/GEN pT
dPhi_34


Deviation vs GEN_pt train
Deviation_GEN_pt_train
Entries 131814
Mean x 178.5
Mean y 0.00079
Std Dev x 242
Std Dev y 0.02269
Deviation: KNN predict - 1/GEN pT
GEN pT

Comments

- What pT range are we caring about? Depends on L1 endcap muon trigger threshold?
- Maybe we don’t need to choose the best model, instead, we can use weighted models for final pT training
- Maybe we can use different model for different input variables, some model may perform better than others in certain parameter region(eta)

Back up

- For a test event, the algorithm finds the k-nearest neighbours using the input variables, where each training event contains a regression value. The predicted regression value for the test event is the weighted average of the regression values of the k-nearest neighbours

"nkNN=95:ScaleFrac=0.8:SigmaFact=1.0:Kernel=Gaus:UseKernel=F:UseWeight=T:!Trim"

- Test with 131872 events, Train with 131814 events

k	10	15	20	25	30	35	40	45	50
MSE(test)	0.00115753	0.00112952	0.00111974	0.0011147	0.00111398	0.00111496	0.00111774	0.0011205	0.00112462
MSE(train)	0.000910486	0.00096604	0.000997173	0.00101723	0.00103356	0.00104684	0.00106053	0.00106954	0.00107752
k	55	60	65	70	75	80	85	90	95
MSE(test)	0.00112702	0.00112989	0.00113308	0.00113655	0.00113934	0.00114193	0.00114449	0.00114828	0.00115224
MSE(train)	0.00108521	0.00109236	0.00109916	0.00110448	0.00110943	0.00111416	0.0011192	0.00112404	0.00112887