Karlsruhe Institute of Technology

UNSAM
UNIVERSIDAD NACIONAL DE SAN MARTÍN

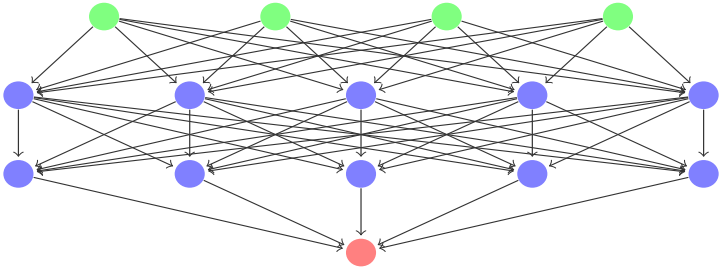
Double Doctoral degree in Astrophysics
DDAp

PIERRE AUGER OBSERVATORY

SD data analysis with Deep Learning

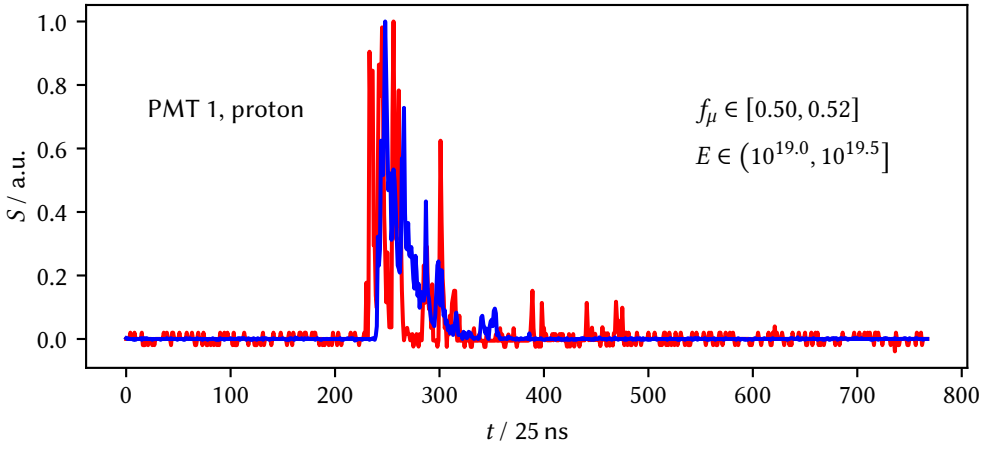
Steffen Hahn, Supvs. Prof. Dr. {Ralph Engel, Brian Wundheiler}, Dr. Markus Roth | September 22, 2019

KIT - IKP

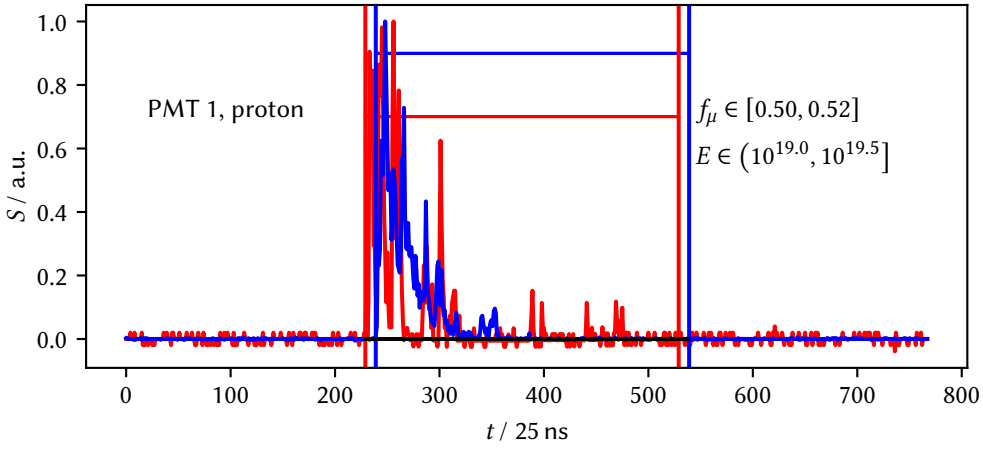


Outline

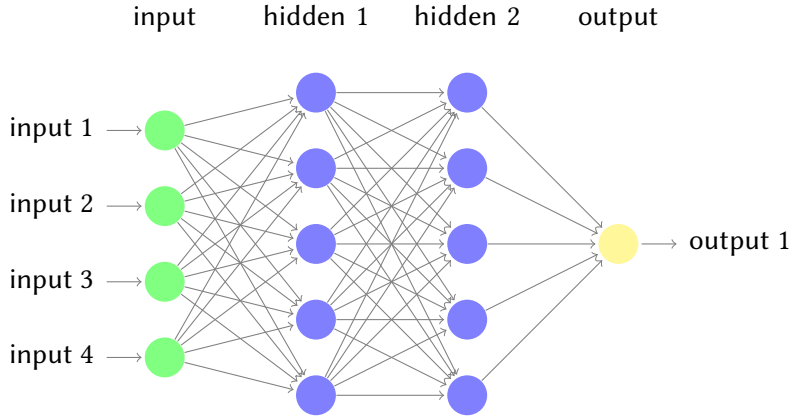
Motivation


PMT 1, proton
$f_\mu \in [0.50, 0.52]$
$E \in \left(10^{19.0}, 10^{19.5}\right]$
S / a.u.
t / 25 ns
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
700
800

Motivation

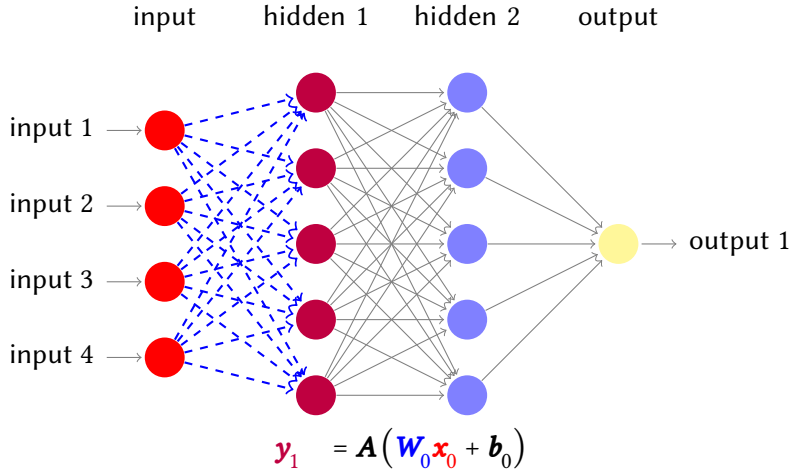

PMT 1, proton
$f_\mu \in [0.50, 0.52]$
$E \in \left(10^{19.0}, 10^{19.5}\right]$
S / a.u.
t / 25 ns
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
700
800

Neural Networks - I

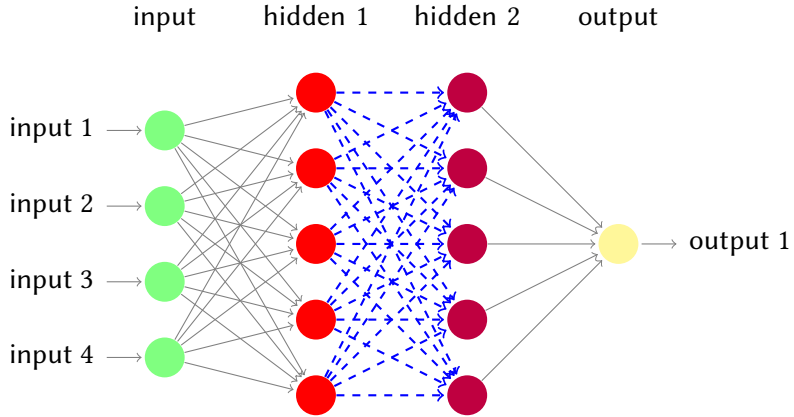


$$\boldsymbol{y}_{i+1} = \boldsymbol{A}\left(\boldsymbol{W}_i\boldsymbol{x}_i + \boldsymbol{b}_i\right)$$

Neural Networks - I

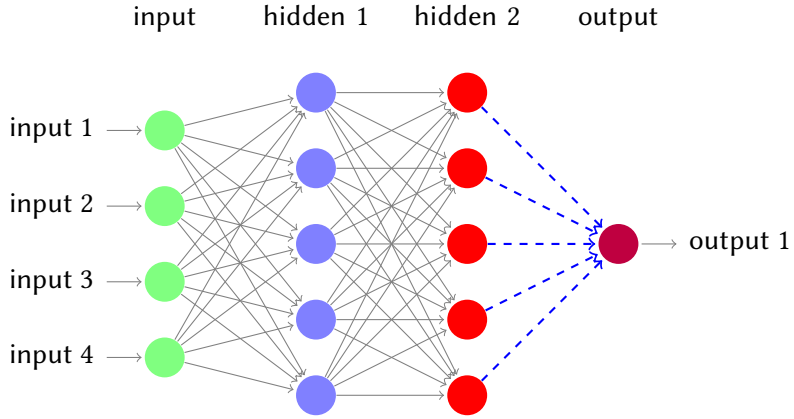


Neural Networks - I



$$y_2 = A\left(W_1 x_1 + b_1\right)$$

Neural Networks - I



Neural Networks - I

layer

$$\begin{pmatrix} y_0 \\ \vdots \\ y_4 \end{pmatrix} = A\left[\begin{pmatrix} w_{00} & \cdots & w_{03} \\ \vdots & \ddots & \vdots \\ w_{40} & \cdots & w_{43} \end{pmatrix}\begin{pmatrix} x_0 \\ \vdots \\ x_3 \end{pmatrix} + \begin{pmatrix} b_0 \\ \vdots \\ b_4 \end{pmatrix}\right]$$

e.g. $A = \max(0, x)$, $A = \tanh(x)$

- parallelizable
- framework: keras on top of tensorflow

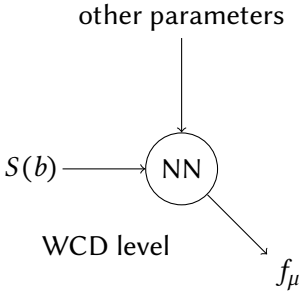
Neural Networks - I

layer

$$\begin{pmatrix} y_0 \\ \vdots \\ y_4 \end{pmatrix} = A\left[\begin{pmatrix} w_{00} & \cdots & w_{03} \\ \vdots & \ddots & \vdots \\ w_{40} & \cdots & w_{43} \end{pmatrix}\begin{pmatrix} x_0 \\ \vdots \\ x_3 \end{pmatrix} + \begin{pmatrix} b_0 \\ \vdots \\ b_4 \end{pmatrix}\right]$$

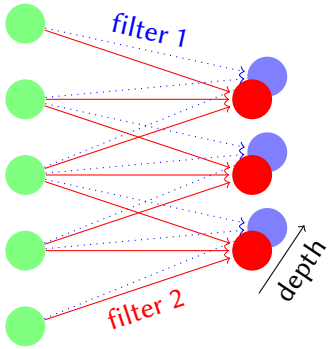
e.g. $A = \max(0, x)$, $A = \tanh(x)$

- parallelizable
- framework: keras on top of tensorflow



Neural Networks - II

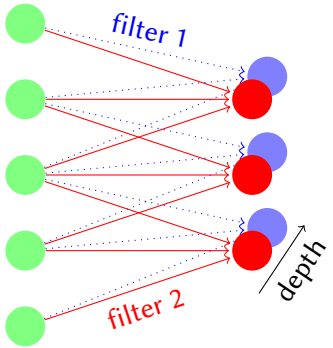
Convolutional Neural Network (CNN)



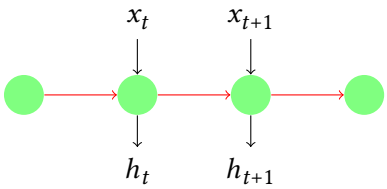
Recurrent Neural Network (RNN)

Neural Networks - II

Convolutional Neural Network (CNN)

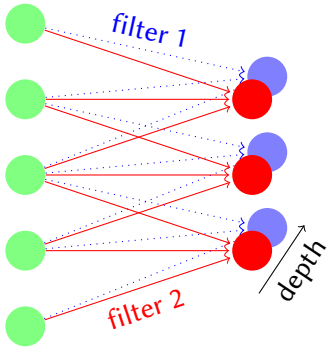


Recurrent Neural Network (RNN)

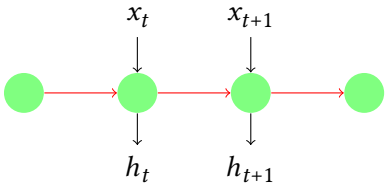


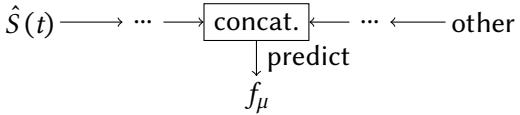
Neural Networks - II

Convolutional Neural Network (CNN)

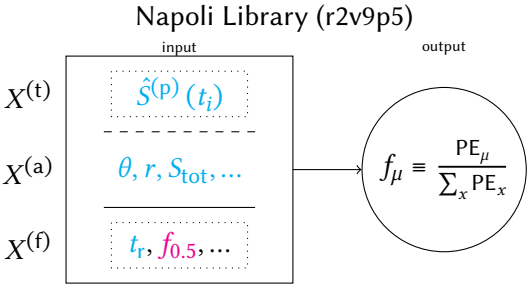


Recurrent Neural Network (RNN)





Data preparation



X_q (QGSJ-II.04)

3604075 samples

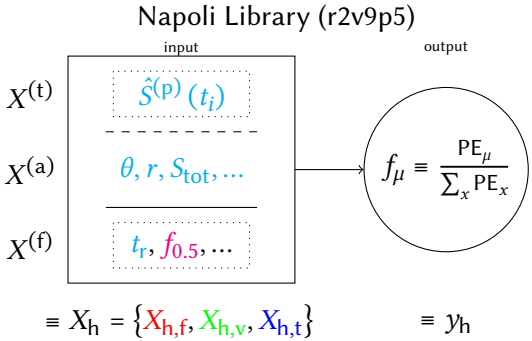
X_e (EPOS-LHC)

3714117 samples

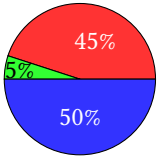
- cuts: no low-gain sat. traces
- 45:5:50 / training:validation:test
- $\hat{S}^{(p)} \in [0, 1]$; $X^{(a)}, X^{(f)}$ standardized

taken from Offline, feature engineered

Data preparation

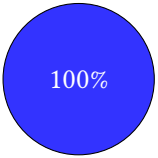


X_q (QGSJ-II.04)



3604075 samples

X_e (EPOS-LHC)

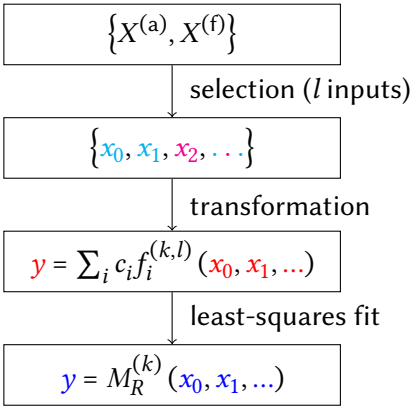


3714117 samples

- cuts: no low-gain sat. traces
- 45:5:50 / training:validation:test
- $\hat{S}^{(p)} \in [0, 1]$; $X^{(a)}, X^{(f)}$ standardized

taken from Offline, feature engineered

Reference models - I



no validation used

Definition:

$f_i^{(k,l)}(x_0, \ldots) \equiv (x_0)^{b_{0i}} (x_1)^{b_{1i}} \ldots, k \geq \sum_j b_{ji}$

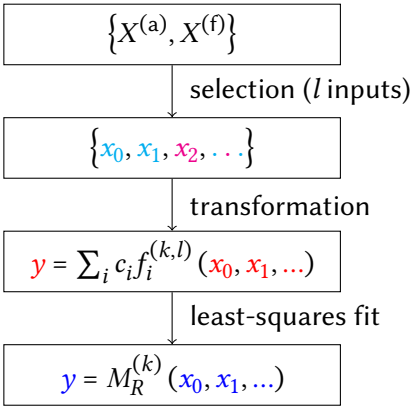
e.g. $f_1^{(2,2)} = x_0, f_3^{(2,2)} = x_0^2, f_4^{(2,2)} = x_0 x_1$

num of $\{c_0, \ldots\} = \binom{k+l}{k}$

$M_R^{(k)} \sim$ Shallow NN

$\sim$ one hidden layer

Reference models - I



no validation used

Definition:

$f_i^{(k,l)}(x_0, \ldots) \equiv (x_0)^{b_{0i}} (x_1)^{b_{1i}} \ldots, k \geq \sum_j b_{ji}$

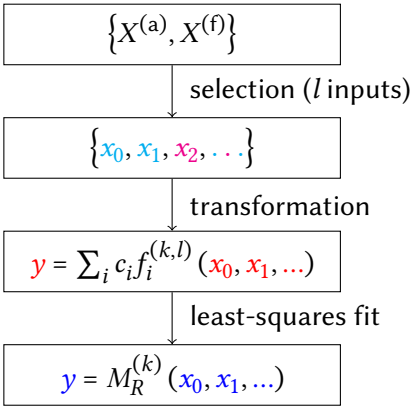
e.g. $f_1^{(2,2)} = x_0$, $f_3^{(2,2)} = x_0^2$, $f_4^{(2,2)} = x_0 x_1$

num of $\{c_0, \ldots\} = \binom{k+l}{k}$

$M_R^{(k)}$ ~ Shallow NN

~ one hidden layer

Reference models - I



no validation used

Definition:

$f_i^{(k,l)}(x_0, \ldots) \equiv (x_0)^{b_{0i}} (x_1)^{b_{1i}} \ldots, k \geq \sum_j b_{ji}$

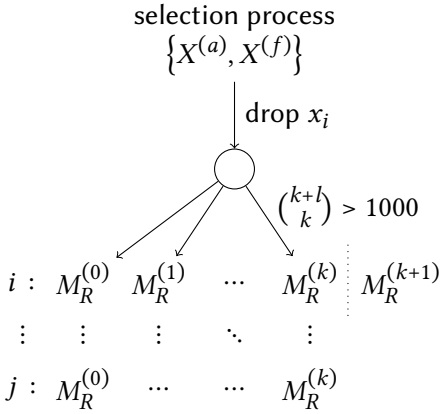
e.g. $f_1^{(2,2)} = x_0$, $f_3^{(2,2)} = x_0^2$, $f_4^{(2,2)} = x_0 x_1$

num of $\{c_0, \ldots\} = \binom{k+l}{k}$

$M_R^{(k)}$ ~ Shallow NN

~ one hidden layer

Reference models - II



$M_{R,1}^{(k)} \equiv M_1$

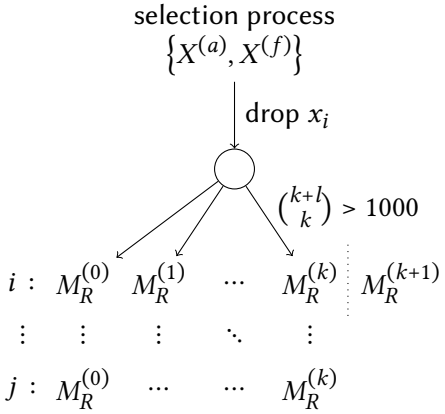
$\{X^{(a)}, X^{(f)}\}$

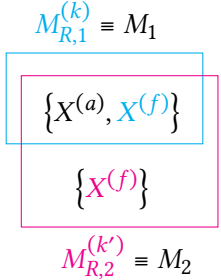
$\{X^{(f)}\}$

$M_{R,2}^{(k')} \equiv M_2$

- $M_{R,1}^{(k)}$: $k' = 3$, $l = 11$
- $M_{R,2}^{(k')}$: $k = 3$, $l = 16$
- highly subjective (!)

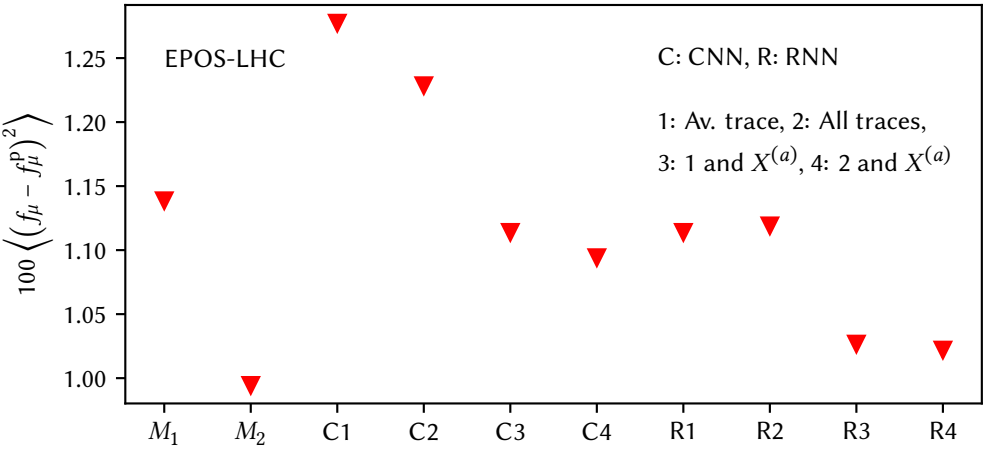
Reference models - II



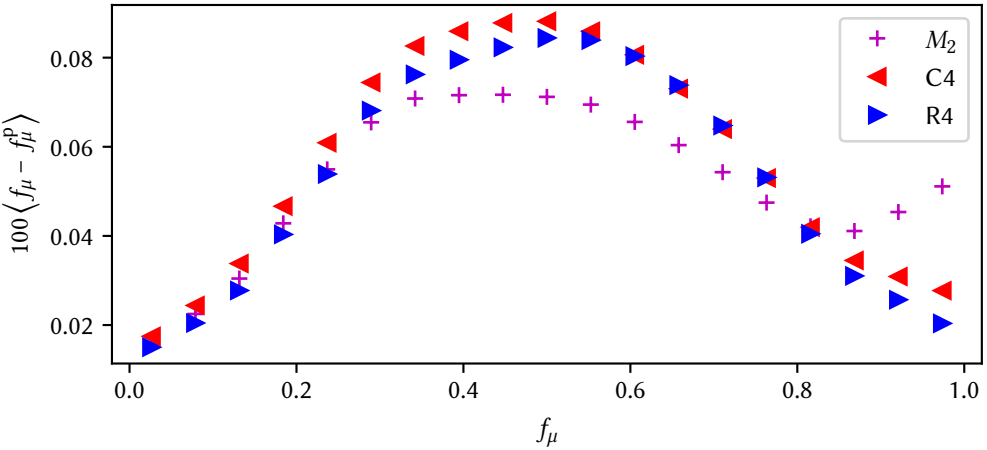


- $M_{R,1}^{(k)}$: $k' = 3$, $l = 11$
- $M_{R,2}^{(k')}$: $k = 3$, $l = 16$
- highly subjective (!)

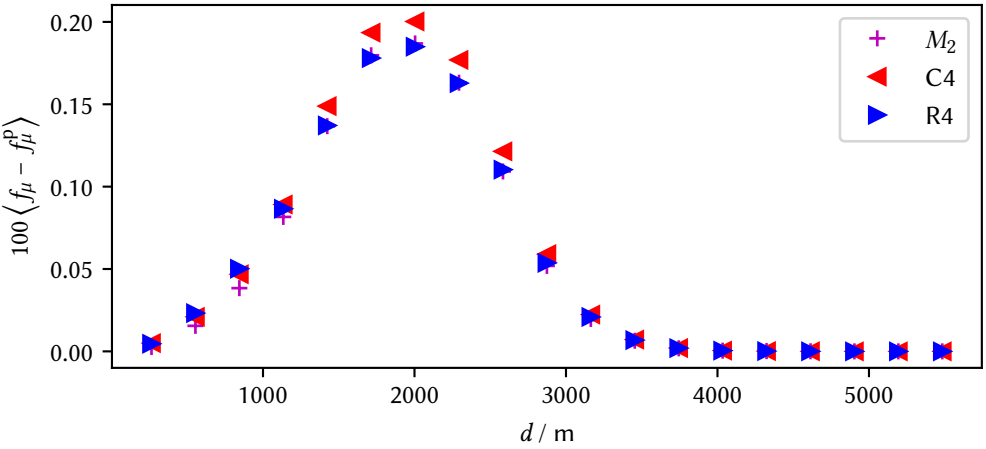
Preliminary results - I



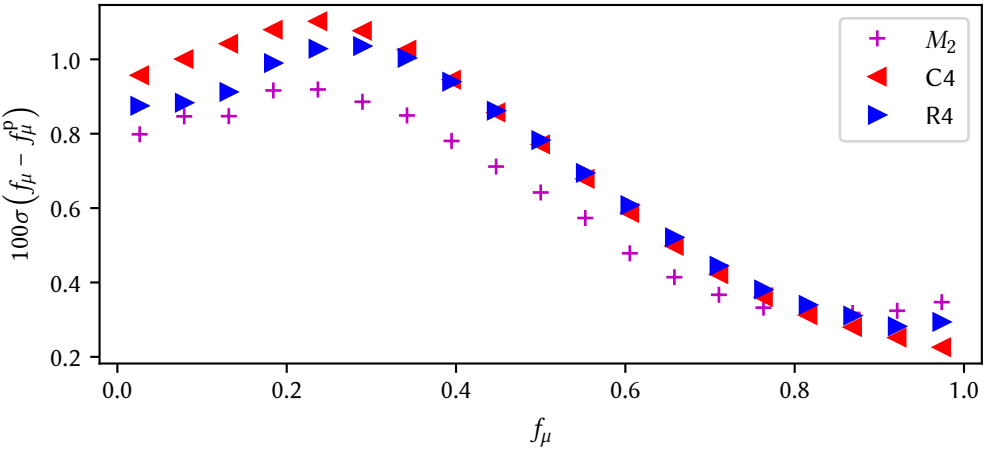
Preliminary result - II



Preliminary result - II



Preliminary result - III



Outlook

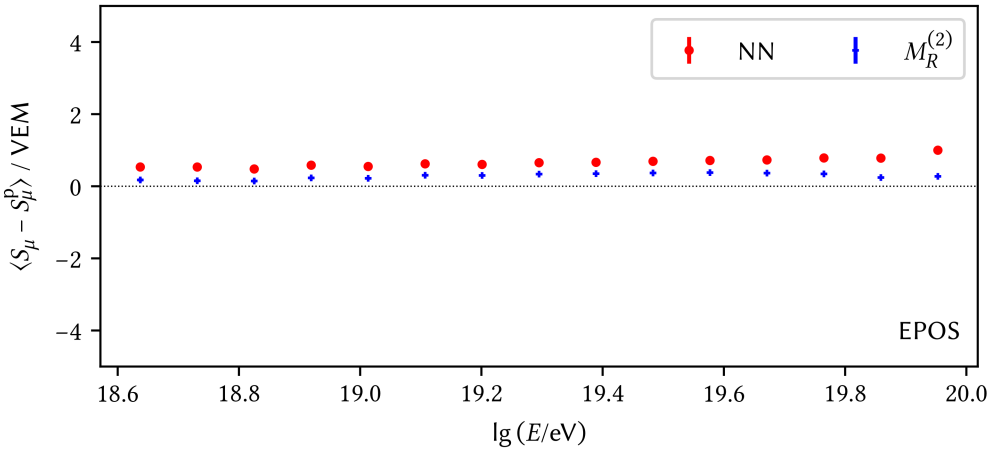
Preliminary results

- trace only is ‘surprisingly’ good
- avg. trace is comparable to three traces
- RNN beats CNN $\overset{?}{\Longrightarrow}$ global beats local

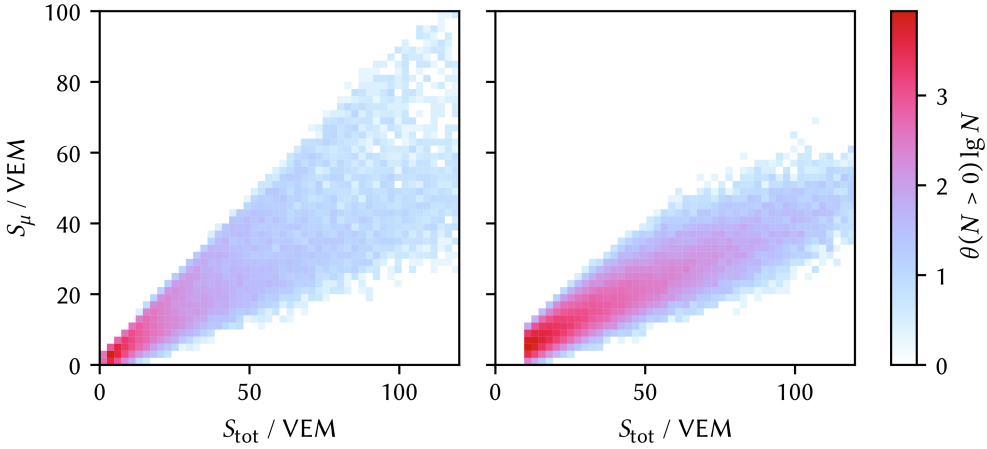
Outlook

- bagging?
- stacking?
- SSD and UUB (!)

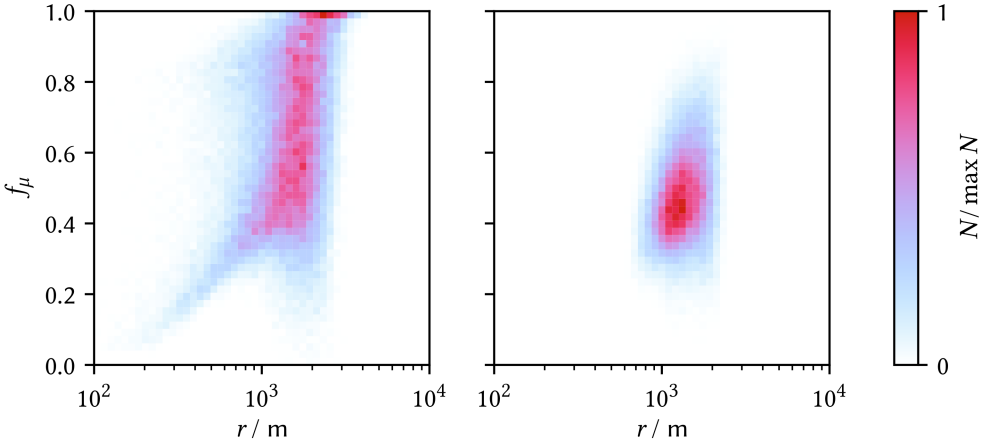
Performance



Effect of cuts



Effect of cuts



Additional parameters

From Offline

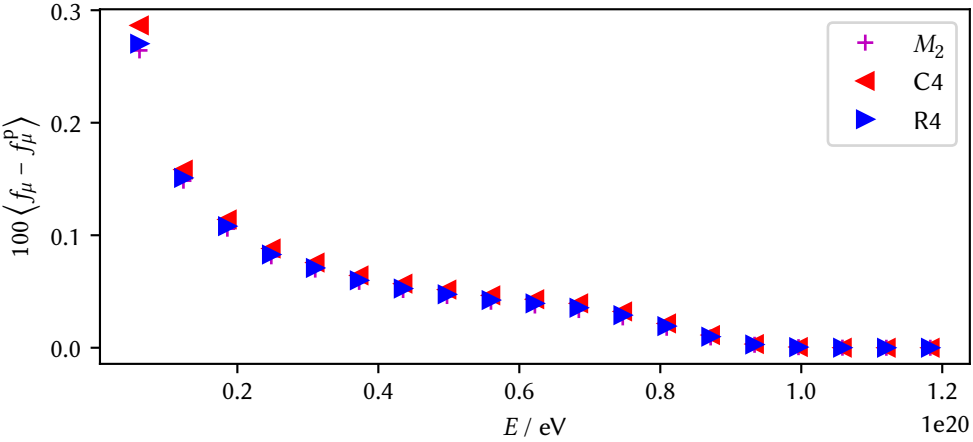
SdRecShower:Zenith, SdRecShower:ZenithError, SdRecStation:SPDistance, SdRecStation:SPDistanceError, SdRecStation:TotalSignal, SdRecStation:TotalSignalError, SdRecStation:SignalStartSlot, SdRecStation:SignalEndSlot, SdRecStation:TimeNSecond, SdRecStation:FallTime, SdRecStation:RiseTime

Feature engineered

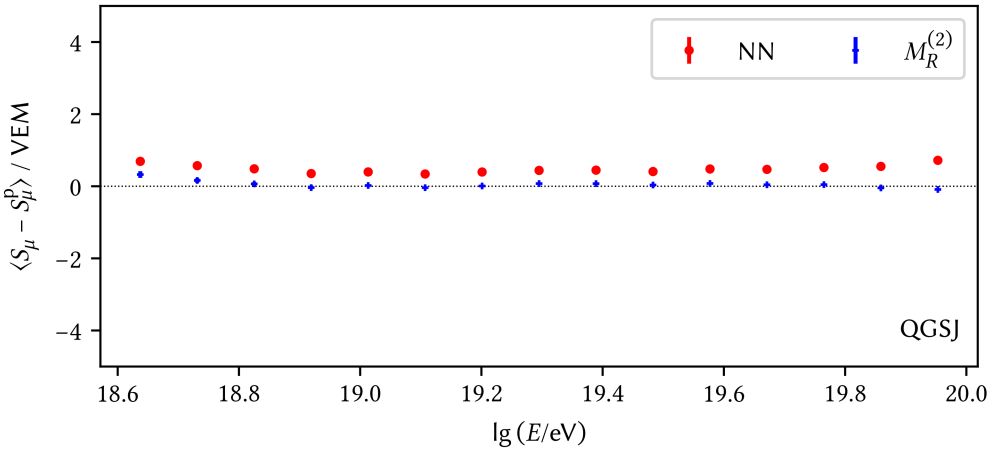
value of trace peak, $f_{0.05}$, $f_{0.50}$, $f_{5.00}$, p_0

$$f_x = \sum S(t)\Theta\left(S(t) > x\right), p_0 = \langle S \rangle^2 / \left\langle S^2 \right\rangle$$

Performance



Performance



Performance