Parameter Inference from Event Ensembles and the Top-Quark Mass

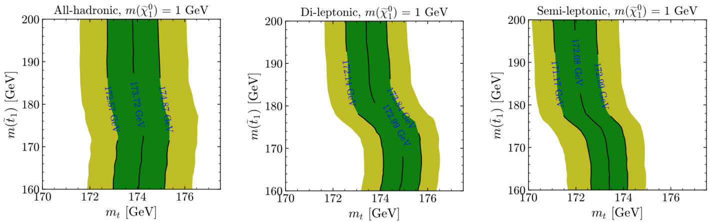
Katherine Fraser

Department of Physics
Harvard University
kfraser@g.harvard.edu

Phenomenology Symposium 2021
arXiv: 2011.04666, Submitted to JHEP
with F. Flesher, C. Hutchinson, B. Ostdiek, M. Schwartz

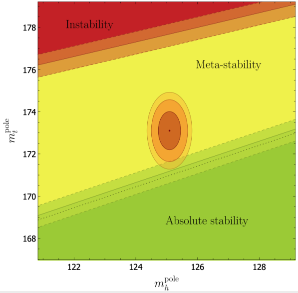
Motivation

Why precision top measurement?



[Cohen, Majewski, Ostdiek, Zheng: 1909.09670]

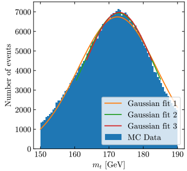
[Andreassen, Frost, Schwartz:1707.08124]



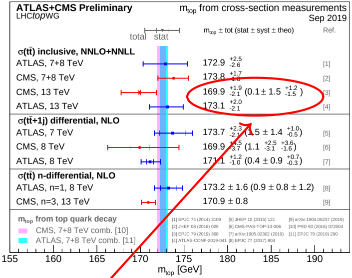
- Want to undertand the SM
- Can be used to search for BSM physics like SUSY
- Affects vacuum stability

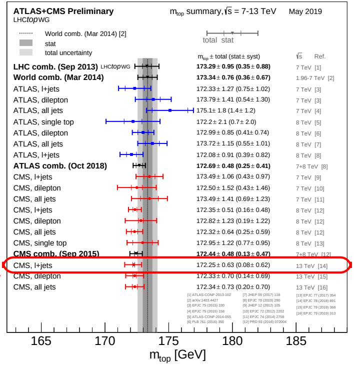
The Top Mass Measurement

- Measurements using the total cross section have a large error
- Best measurements fit the invariant mass peak in $t\bar{t}$ events



[LHCTopWG twiki]





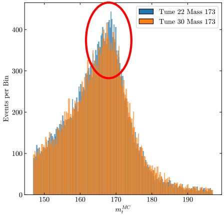
Large Error
$\mathcal{O}$(GeV)

Small Error
$\mathcal{O}$(600 MeV)

The Monte Carlo Mass

- Fit of the peak is the MC mass.
- Changes in fit due to dependence on unphysical simulation parameters (tunes) is the error.
- Goal: to eliminate dependence of the top quark MC mass on tune.

Peak Depends on tune



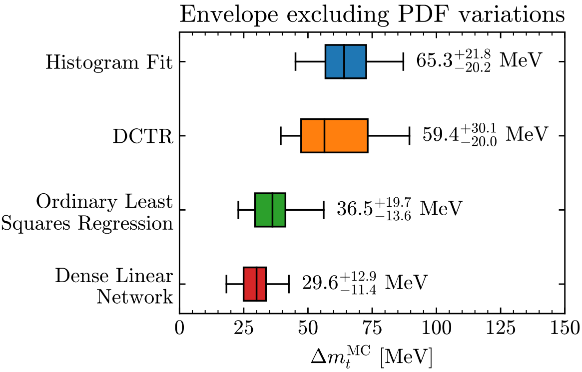
	2D approach δm_t^{2D} [GeV]	2D approach δJSF^{2D} [%]	1D approach δm_t^{1D} [GeV]	Hybrid δm_t^{hyb} [GeV]	Hybrid δJSF^{hyb} [%]
Experimental uncertainties					
Method calibration	0.05	<0.1	0.05	0.05	<0.1
JEC (quad. sum)	0.13	0.2	0.83	0.18	0.3
– InterCalibration	(−0.02)	(<0.1)	(+0.16)	(+0.04)	(<0.1)
– MPFInSitu	(−0.01)	(<0.1)	(+0.23)	(+0.07)	(<0.1)
– Uncorrelated	(−0.13)	(+0.2)	(+0.78)	(+0.16)	(+0.3)
Jet energy resolution	−0.20	+0.3	+0.09	−0.12	+0.2
b tagging	+0.03	<0.1	+0.01	+0.03	<0.1
Pileup	−0.08	+0.1	+0.02	−0.05	+0.1
Non-$t\bar{t}$ background	+0.04	−0.1	−0.02	+0.02	−0.1
Modeling uncertainties					
JEC Flavor (linear sum)	−0.42	+0.1	−0.31	−0.39	<0.1
– light quarks (uds)	(+0.10)	(−0.1)	(−0.01)	(+0.06)	(−0.1)
– charm	(+0.02)	(<0.1)	(−0.01)	(+0.01)	(<0.1)
– bottom	(−0.32)	(<0.1)	(−0.31)	(−0.32)	(<0.1)
– gluon	(−0.22)	(+0.3)	(+0.02)	(−0.15)	(+0.2)
b jet modeling (quad. sum)	0.13	0.1	0.09	0.12	<0.1
– b frag. Bowler–Lund	(−0.07)	(+0.1)	(−0.01)	(−0.05)	(<0.1)
– b frag. Peterson	(+0.04)	(<0.1)	(+0.05)	(+0.04)	(<0.1)
– semileptonic B decays	(+0.11)	(<0.1)	(+0.08)	(+0.10)	(<0.1)
PDF	0.02	<0.1	0.02	0.02	<0.1
Ren. and fact. scales	0.02	0.1	0.02	0.01	<0.1
ME/PS matching	-0.08 ± 0.09	+0.1	$+0.03 \pm 0.05$	-0.05 ± 0.07	+0.1
ME generator	$+0.15 \pm 0.23$	+0.2	$+0.32 \pm 0.14$	$+0.20 \pm 0.19$	+0.1
ISR PS scale	$+0.07 \pm 0.09$	+0.1	$+0.10 \pm 0.05$	$+0.06 \pm 0.07$	<0.1
FSR PS scale	$+0.24 \pm 0.06$	−0.4	-0.22 ± 0.04	$+0.13 \pm 0.05$	−0.3
Top quark p_T	+0.02	−0.1	−0.06	−0.01	−0.1
Underlying event	-0.10 ± 0.08	+0.1	$+0.01 \pm 0.05$	-0.07 ± 0.07	+0.1
Early resonance decays	-0.22 ± 0.09	+0.8	$+0.42 \pm 0.05$	-0.03 ± 0.07	+0.5
Color reconnection	$+0.34 \pm 0.09$	−0.1	$+0.23 \pm 0.06$	$+0.31 \pm 0.08$	−0.1
Total systematic	**0.75**	**1.1**	**1.10**	**0.62**	**0.8**
Statistical (expected)	0.09	0.1	0.06	0.08	0.1
Total (expected)	**0.76**	**1.1**	**1.10**	**0.63**	**0.8**

Tunes: coordinated changes of simulation parameters

[CMS: EPJC 78 (2018) 891 (1805.01428)]

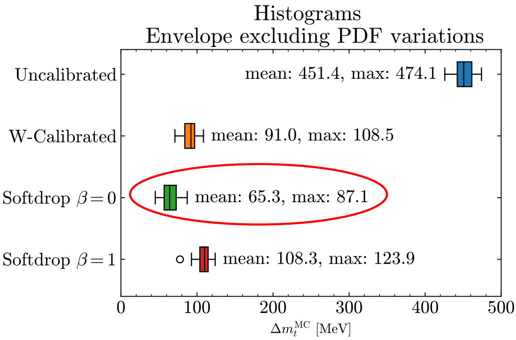
Results and Outline

Methods to decrease the uncertainty on the top quark monte carlo (MC) mass:



Previous Work - Histogram Fits with Grooming

- To extract the top-quark MC mass, iteratively fit histograms.
- Uncertainty $\sim 500 - 800$ MeV; ~ 100 MeV with W-calibration: ($m_{\text{calibrated}} = m_{3J}\frac{m_W}{m_{2J}}$)
- Error can be further reduced by grooming with soft drop ($\sim 30\%$) [Andreassen, Schwartz: 1705.07135]



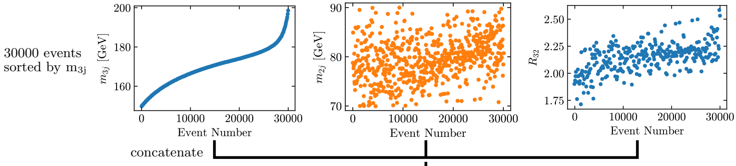
How can we do better?

Curve fitting reduces histograms to a very small number of parameters that are dependent on the parameterization.

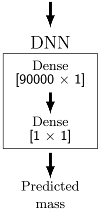
Can we use more of the distribution? Eliminate dependence on parameterization?

We can! Do regression on an ensemble of events without a binned histogram.

Regression on Event Ensembles



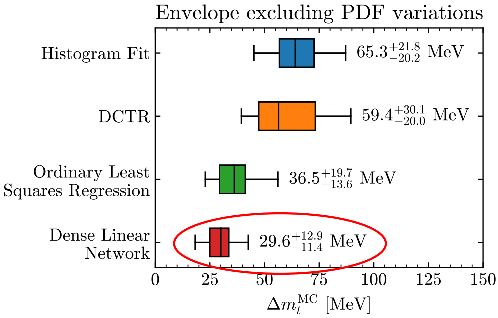
Plug 90k numbers directly into regression



Pick Method for regression.
Ex) NN

Regression on Event Ensembles

- Improves by a factor of ~ 2
- Modifications don't help, including:
  - Different inputs
  - Different network structures
  - Ordinary Least Squares regression

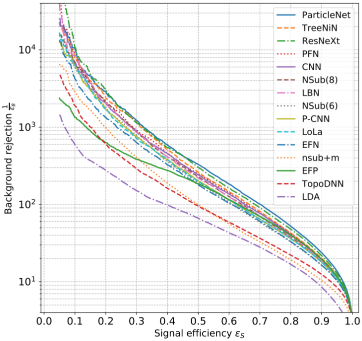


Can Modern ML Help?

- Low level info helps for other applications such as top tagging
- Can't easily include low level info directly into regression
- Other options: reduce to Energy Flow Polynomials, regression using classifiers, DCTR

  [Komiske, Metodiev, Thaler: 1712.07124]

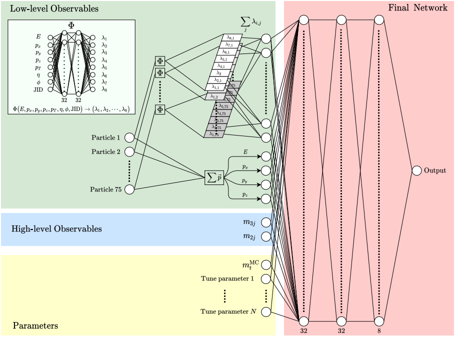
  [Nachman, Thaler: 2101.07263]



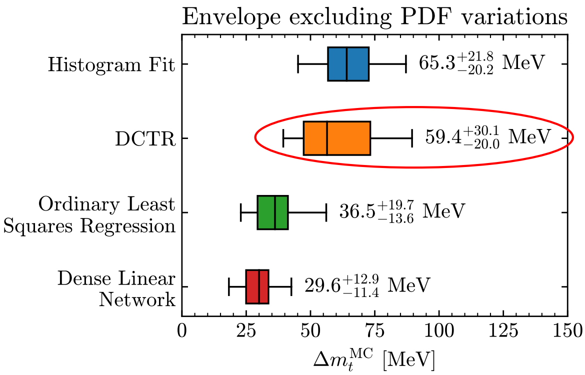
[Comparing ML Top Taggers Performance: Kasieczka et al, 1902.09914]

A DCTR Study

- DCTR is a modern ML method [Andreassen, Nachman: 1907.08209]
- Relies on parameterized classifiers, which can take as input HLV, low level input, and parameters
- Works because loss of a classifier = likelihood ratio



DCTR Results



Summary

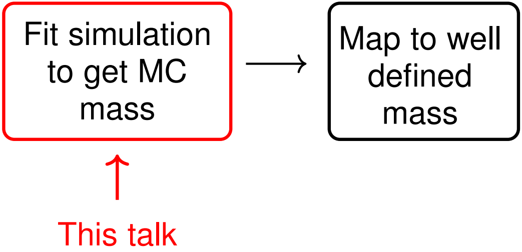
1. **The degeneracy between mass and tune is a significant source of error in measuring m_t.**
2. Three strategies for reducing this error: (1) Jet Grooming [1705.07135] (2) Regression on event ensembles (3) Modern ML DCTR method. **Regression on event ensembles reduces error by a factor of ~ 2 compared to other methods.**
3. **Regression on event ensembles could potentially be useful for other measurements.** [Ex. W polarization: Kim, Martin: 2102.05124]

Back Up Slides

A Two Step Process

Two steps in measuring the top mass:

1. Fit to the peak in simulation to get the Monte Carlo (MC) mass
2. Map the MC mass to a well-defined mass scheme like $\overline{ms}$ by treating it like the pole mass



ATLAS A14 7 TeV Tunes

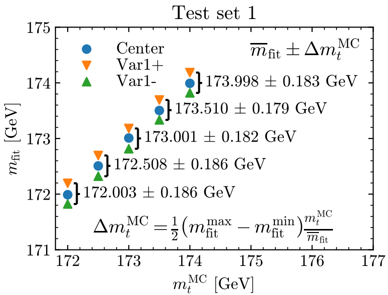
Variation	Tunes	ColorRec	α_S^{MPI}	p_{T0}^{Ref}
VarPDF	19-22	1.71	0.126	1.56
Var1	21, 23, 24	[1.69,1.73]	[0.121,0.131]	1.56
Var2	21, 25, 26	1.71	0.126	[1.50,1.60]
Var3a	21, 27, 28	1.71	[0.125,0.127]	[1.51,1.67]
Var3b	21, 29, 30	1.71	0.126	1.56
Var3c	21, 31, 32	1.71	0.126	1.56

Variation	$p_T^{\mathrm{dampFudge}}$	α_S^{FSR}	p_T^{maxFudge}	α_S^{ISR}
VarPDF	1.05	0.127	0.91	0.127
Var1	1.05	0.127	0.91	0.127
Var2	[1.04,1.08]	[0.124,0.136]	0.91	0.127
Var3a	[0.93,1.36]	[0.124,0.136]	[0.88,0.98]	0.127
Var3b	[1.04,1.07]	[0.114,0.138]	[0.83,1.00]	[0.126,0.129]
Var3c	1.05	0.127	0.91	[0.115,0.140]

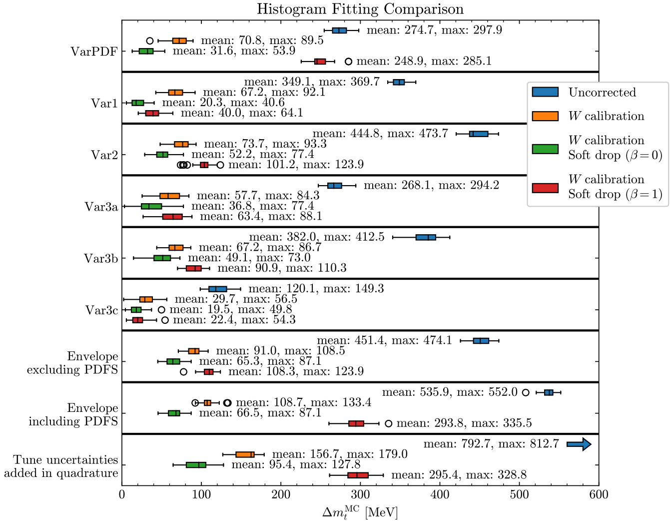
Defining the Error

For a set of tunes with fixed input mass, the error is defined to be

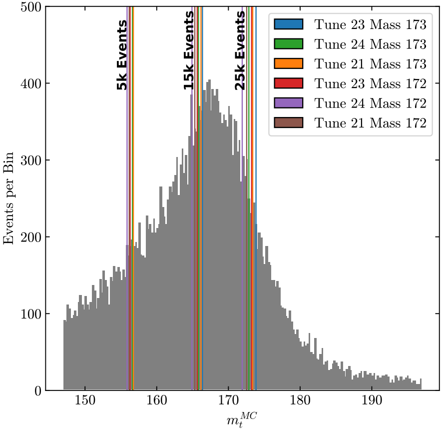
$$\Delta m_t^{MC} = \frac{1}{2}\left(m_{\text{fit}}^{\max} - m_{\text{fit}}^{\min}\right)\frac{m_t^{MC}}{\bar{m}_t^{\text{fit}}}.$$



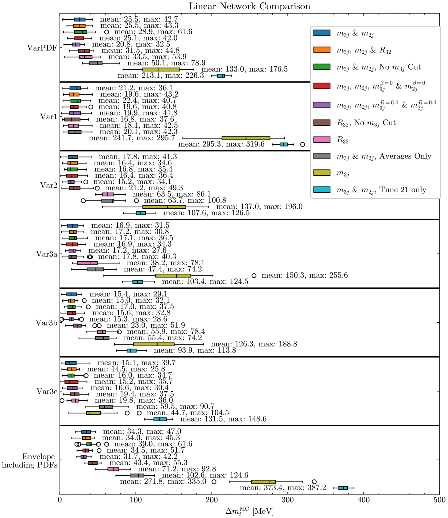
Histogram Full Results



Event Histograms

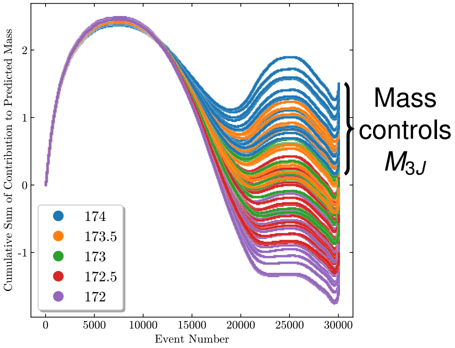


Linear Network Full Results



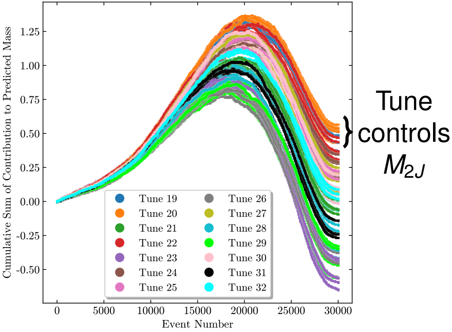
Understanding the Improvement

M_{3J} contribution:



14 lines of each color (one for each tune)

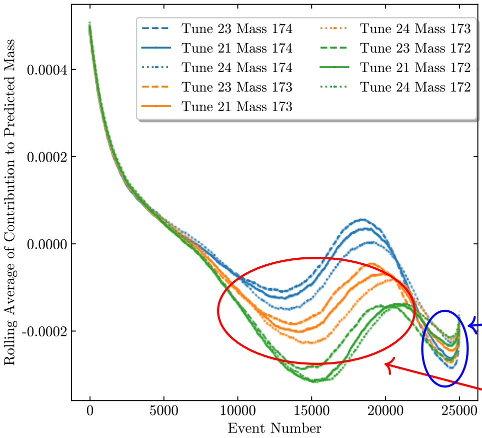
M_{2J} contribution:

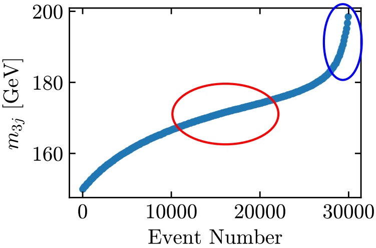


5 lines of each color (one for each mass)

A Closer Look at M_{3J}

M_{3J} contribution:



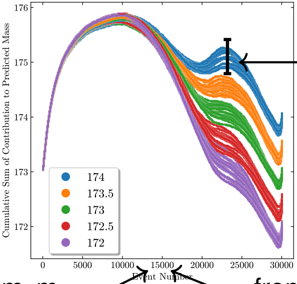


Correct for tunes away from peak

Learn the mass from the peak

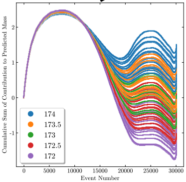
Understanding the Improvement

The full network can use tails to correct for differences in tune:

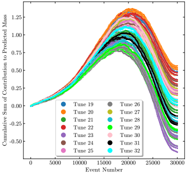


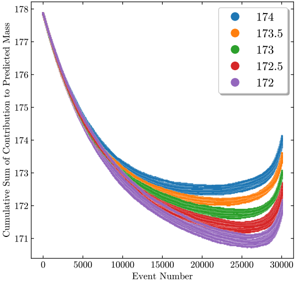
Width of color band = size of error

from m_{3J}



from m_{2J}





$m_{3J}^{Avg}, m_{2J}^{Avg}$

[14 lines of each color: one for each tune]

DCTR in Practice

Two Steps:

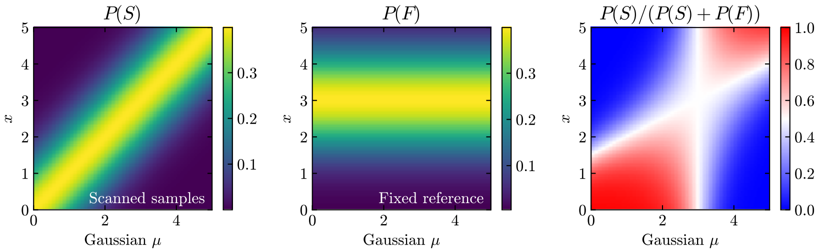
1. Train a parameterized classifier to learn the likelihood ratio
2. Evaluate classifier for an unknown sample and a reference sample, trying multiple input parameters for the unknown sample.

   Likelihood is maximized at true parameters used to generate the unknown sample

DCTR Gaussian Example: Training Step

- 2 Training Samples:
  1. Samples S from Gaussian with mean μ random, label μ
  2. Samples U from Gaussian with mean μ fixed, random label
- Output of Classifier gives ratio of probability densities:

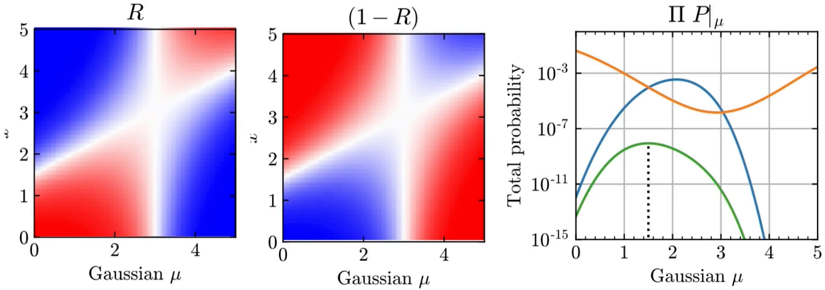
$$R(\mu, x) = \frac{P(S(\mu, x))}{P(S(\mu, x)) + P(U(\mu, x))}$$



More DCTR Results: Evaluation Step

To infer most probable parameter, maximize $C(\mu)$:

$$C(\mu) = \prod_{x \in T} R(\mu, x) \prod_{x \in F} (1 - R(\mu, x))$$



In Example: Reference $\mu = 3$, New $\mu = 1.5$