Two New Ideas for ML-Theory

Tilman Plehn

Universität Heidelberg

Milano, December 2021

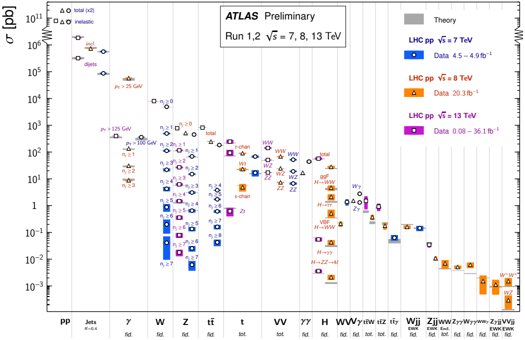
LHC goals

Fundamental questions

- particle nature of dark matter?
- origin of the Higgs mechanism? [hierarchy problem?]
- matter-antimatter asymmetry? [CP-symmetry]
- Standard Model all there is?

Rate measurements

- many processes
- vastly different rates
- high precision
- predicted by theory



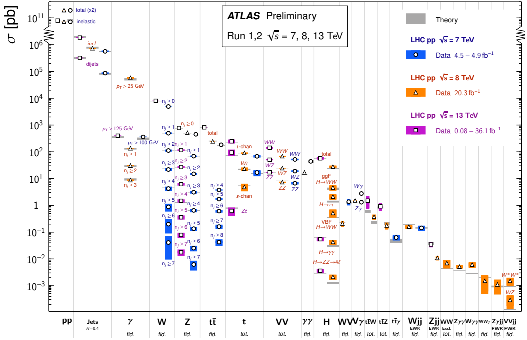
LHC goals

Fundamental questions

– particle nature of dark matter?
– origin of the Higgs mechanism? [hierarchy problem?]
– matter-antimatter asymmetry? [CP-symmetry]
– Standard Model all there is?

Rate measurements

– many processes
– vastly different rates
– high precision
– predicted by theory
– but completely useless!

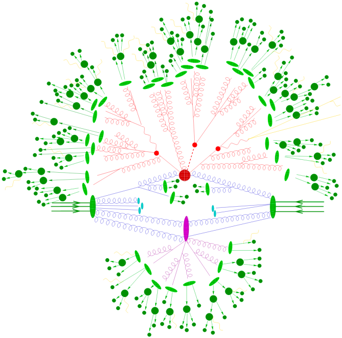


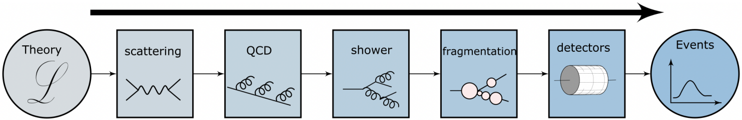
1– First-principle simulations

Simulation-based inference [likelihood-free inference]

- start with Lagrangian, perturbative QFT
- simulate events [Sherpa, Madgraph, Pythia, Powheg]
- simulate detectors

$\Rightarrow$ LHC events in virtual worlds





1– First-principle simulations

Simulation-based inference [likelihood-free inference]

– start with Lagrangian, perturbative QFT

– simulate events [Sherpa, Madgraph, Pythia, Powheg]

– simulate detectors

⇒ LHC events in virtual worlds

Forward LHC simulations

– HL-LHC: preparing for 25-fold data set

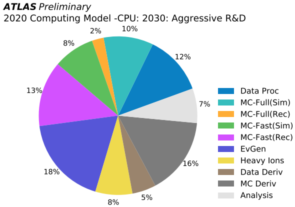
– simulated event numbers ∼ expected events

– statistics requiring 1%-2% uncertainty [NNLO/N^{3}LO]

– flexible signal hypotheses [time-dependent]

– low-rate high-multiplicity backgrounds

⇒ Event generation limiting factor



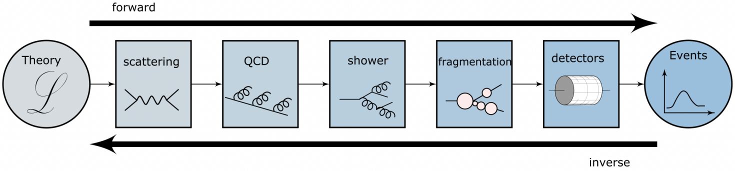
1– First-principle simulations

Simulation-based inference [likelihood-free inference]

– start with Lagrangian, perturbative QFT
– simulate events [Sherpa, Madgraph, Pythia, Powheg]
– simulate detectors
⇒ LHC events in virtual worlds

Inverted LHC simulations

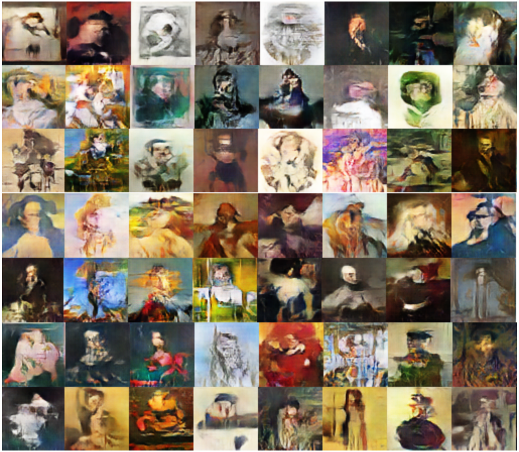
– unfolding QCD-shower to hard parton standard [jet algorithm]
unfolding detector common
unfolding top-quark decays useful
matrix element method complete unfolding
⇒ Maybe benefit from NN-concepts [Omnifold, cINN]



Generative networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

- can networks create new pieces of art?
  map random numbers to image pixels
- train on 80,000 pictures [organized by style and genre]
- generate portraits



Generative networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

– can networks create new pieces of art?
  map random numbers to image pixels
– train on 80,000 pictures [organized by style and genre]
– generate portraits

Edmond de Belamy [Caselles-Dupre, Fautrel, Vernier (2018)]

– trained on 15,000 portraits
– sold for $432.500

⇒ ML often marketing and sales

Generative networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

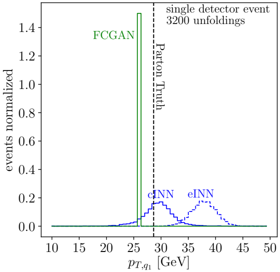
– can networks create new pieces of art?
  map random numbers to image pixels
– train on 80,000 pictures [organized by style and genre]
– generate portraits

Edmond de Belamy [Caselles-Dupre, Fautrel, Vernier (2018)]

– trained on 15,000 portraits
– sold for $432.500
⇒ ML often marketing and sales

LHC applications

– jets [de Oliveira (2017), Carrazza-Dreyer (2019)...]
– LHC events [Butter (2019), Review (2020)...]
– inversion/unfolding [Omnifold, cGAN, cINN (2019/2020)]
– inference [QCD splittings (2020)...]
– Parton density compression [Rabemananjara (2021)]
  ...
⇒ Science from the beginning

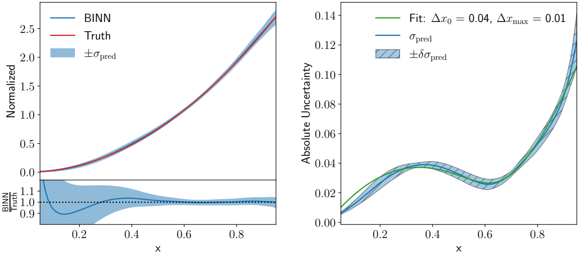


Generative networks with error bars

Nothing Bayesian about Bayesian INNs [Bellagente, Haußmann, Luchmann, TP]

- network with weight distributions [thesis Yarin Gal (2016)]
  sample for network output [including error bar]
  working for regression, classification [Haußmann, Kasieczka, TP,...]
  frequentist: efficient ensembling
- new: generate events with error bars [density & uncertainty maps]
- technically: normalizing flow — INN [Köthe]
  bijective mapping
  known Jacobian
  sampling from Gaussian latent space
- 2D toy models: wedge ramp, kicker ramp, Gaussian ring

$\Rightarrow$ Side remark: see how INN learns



Precision generator

Challenging ML-event generators [useful playground]

– training from event samples
  no energy-momentum conservation
  no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets [resonance peaks]

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]

Precision generator

Challenging ML-event generators [useful playground]

– training from event samples
  no energy-momentum conservation
  no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets [resonance peaks]

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]

INN-generator [Butter, Heimel, Hummerich, Krebs, TP, Rousselot, Vent]

– challenging ΔR_{jj} features

– monotouous function with weights [opposite of importance sampling]

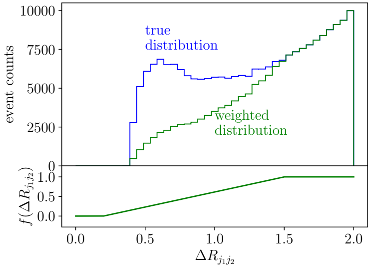
$$w^{(\text{1-jet})} = 1$$

$$w^{(\text{2-jet})} = f(\Delta R_{j_1, j_2})$$

$$w^{(\text{3-jet})} = f(\Delta R_{j_1, j_2}) f(\Delta R_{j_2, j_3}) f(\Delta R_{j_1, j_3})$$

with

$$f(\Delta R) = \begin{cases} 0 & \text{for } \Delta R < R_- \\ \dfrac{\Delta R - R_-}{R_+ - R_-} & \text{for } \Delta R \in [R_-, R_+] \\ 1 & \text{for } \Delta R > R_+ \end{cases}$$



Two Ideas

Tilman Plehn

Generative

Uncertainties

Optimal Obs

SRegression

Precision generator

Challenging ML-event generators [useful playground]

– training from event samples
  no energy-momentum conservation
  no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets [resonance peaks]

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]

INN-generator [Butter, Heimel, Hummerich, Krebs, TP, Rousselot, Vent]

– challenging ΔR_{jj} features

– monotouous function with weights [opposite of importance sampling]

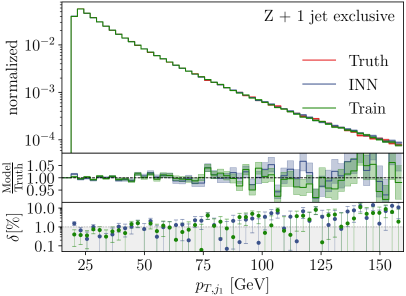
$$w^{(1\text{-jet})} = 1$$

$$w^{(2\text{-jet})} = f(\Delta R_{j_1, j_2})$$

$$w^{(3\text{-jet})} = f(\Delta R_{j_1, j_2}) f(\Delta R_{j_2, j_3}) f(\Delta R_{j_1, j_3})$$

with

$$f(\Delta R) = \begin{cases} 0 & \text{for } \Delta R < R_- \\ \dfrac{\Delta R - R_-}{R_+ - R_-} & \text{for } \Delta R \in [R_-, R_+] \\ 1 & \text{for } \Delta R > R_+ \end{cases}$$



$\Rightarrow$ Per-cent precision possible

Precision generator

Challenging ML-event generators [useful playground]

– training from event samples
no energy-momentum conservation
no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets [resonance peaks]

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets [Z-peak, variable jet number, jet-jet topology]

INN-generator [Butter, Heimel, Hummerich, Krebs, TP, Rousselot, Vent]

– challenging ΔR_{jj} features

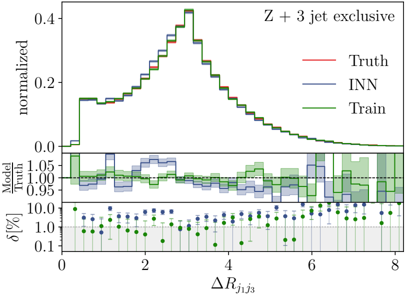
– monotouous function with weights [opposite of importance sampling]

$$w^{(1\text{-jet})} = 1$$
$$w^{(2\text{-jet})} = f(\Delta R_{j_1, j_2})$$
$$w^{(3\text{-jet})} = f(\Delta R_{j_1, j_2}) f(\Delta R_{j_2, j_3}) f(\Delta R_{j_1, j_3})$$

with

$$f(\Delta R) = \begin{cases} 0 & \text{for } \Delta R < R_- \\ \dfrac{\Delta R - R_-}{R_+ - R_-} & \text{for } \Delta R \in [R_-, R_+] \\ 1 & \text{for } \Delta R > R_+ \end{cases}$$

$\Rightarrow$ Per-cent precision possible



Controlled precision generator

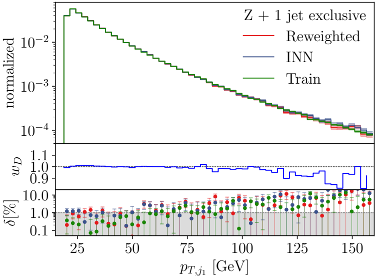
Additional discriminator: training vs generated

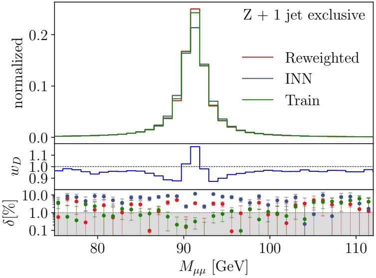
– input $\{p_T, \eta, \phi, M, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$

– decent generator training $D \approx 0.5$

– additional event weight $w_D = D/(1 - D) \to 1$

⇒ Dual purpose: control and reweight





Controlled precision generator

Additional discriminator: training vs generated

- input $\{p_T, \eta, \phi, M, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
- decent generator training $D \approx 0.5$
- additional event weight $w_D = D/(1-D) \to 1$

$\Rightarrow$ Dual purpose: control and reweight

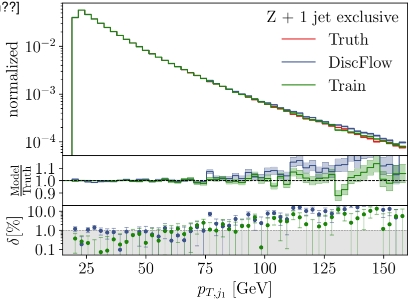
Joint DiscFlow training [GAN inspiration]

- GAN-like training unstable [Nash equilibrium??]
- coupling through weights

$$L_{\text{DiscFlow}} = -\sum_{i=1}^{B} w_D(x_i)^\alpha \log \frac{P(x_i)}{P_{\text{ref}}(x_i)}$$

$$\approx -\int dx \, \frac{P_{\text{ref}}^{\alpha+1}(x)}{P^\alpha(x)} \log \frac{P(x)}{P_{\text{ref}}(x)}$$

$\Rightarrow$ Unweighted, controlled events



Controlled precision generator

Additional discriminator: training vs generated

- input $\{p_T, \eta, \phi, M, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
- decent generator training $D \approx 0.5$
- additional event weight $w_D = D/(1-D) \to 1$

$\Rightarrow$ Dual purpose: control and reweight

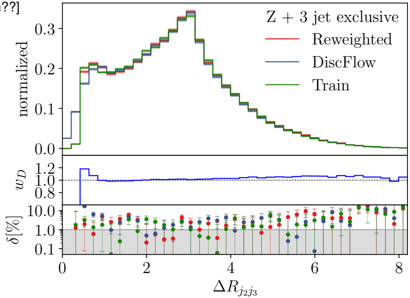
Joint DiscFlow training [GAN inspiration]

- GAN-like training unstable [Nash equilibrium??]
- coupling through weights

$$L_{\text{DiscFlow}} = -\sum_{i=1}^{B} w_D(x_i)^\alpha \ \log \frac{P(x_i)}{P_{\text{ref}}(x_i)}$$

$$\approx -\int dx \ \frac{P_{\text{ref}}^{\alpha+1}(x)}{P^\alpha(x)} \ \log \frac{P(x)}{P_{\text{ref}}(x)}$$

$\Rightarrow$ Reweightable controlled events

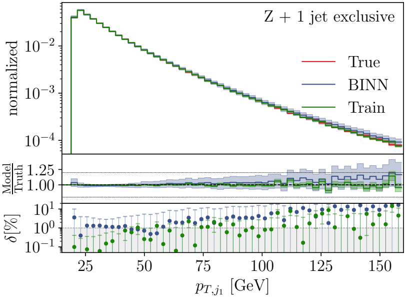


Precision generator with uncertainties

BINN generator

- Bayesian precision generator
- uncertainty over phase space
- training statistics leading source

⇒ Training-related error bars



Precision generator with uncertainties

BINN generator

- Bayesian precision generator
- uncertainty over phase space
- training statistics leading source

⇒ Training-related error bars

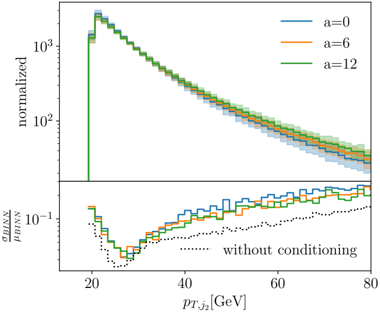
Theory uncertainties

- BNN regression/classification: systematics from data augmentation
- systematic uncertainties in tails

$$w = 1 + a \left(\frac{p_{T,j_1} - 15\ \text{GeV}}{100\ \text{GeV}} \right)^2$$

- augment training data [$a = 0 \dots 30$]
- train conditionally on smeared a error bar from sampling a

⇒ Systematic/theory error bars

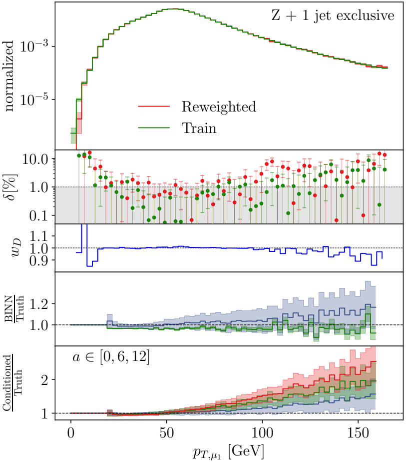


Precision generator with uncertainties

BINN generator

- Bayesian precision generator
- uncertainty over phase space
- training statistics leading source

$\Rightarrow$ Training-related error bars



Theory uncertainties

- BNN regression/classification: systematics from data augmentation
- systematic uncertainties in tails

$$w = 1 + a \left(\frac{p_{T,j_1} - 15 \ \text{GeV}}{100 \ \text{GeV}} \right)^2$$

- augment training data $[a = 0 \dots 30]$
- train conditionally on smeared a error bar from sampling a

$\Rightarrow$ Systematic/theory error bars

$\Rightarrow$ Generative networks for LHC standards

2– Optimal observables

Measure model parameter θ optimally [Atwood-Soni, Diehl-Nachtmann, Davier etal]

– single-event likelihood [from Monte Carlo]

$$p(x|\theta) = \frac{1}{\sigma_{\text{tot}}(\theta)} \frac{d^d \sigma(x|\theta)}{dx^d}$$

– expanded locally in θ, define score [just taylor log]

$$\log \frac{p(x|\theta)}{p(x|\theta_0)} \approx (\theta - \theta_0) \left. \nabla_\theta \log p(x|\theta) \right|_{\theta_0} \equiv (\theta - \theta_0) \, t(x|\theta_0) \equiv (\theta - \theta_0) \, \mathscr{O}^{\text{opt}}(x)$$

– parton level, as used in ATLAS [CPV, Schumacher]

$$p(x|\theta) \approx |\mathcal{M}|_0^2 + \theta |\mathcal{M}|_{\text{int}}^2 \qquad \Rightarrow \qquad t(x|\theta_0) \sim \frac{|\mathcal{M}|_{\text{int}}^2}{|\mathcal{M}|_0^2} ,$$

$\Rightarrow$ Easy at parton level, LEP physics...

2– Optimal observables

Measure model parameter θ optimally [Atwood-Soni, Diehl-Nachtmann, Davier etal]

– single-event likelihood [from Monte Carlo]

$$p(x|\theta) = \frac{1}{\sigma_{\text{tot}}(\theta)} \frac{d^d\sigma(x|\theta)}{dx^d}$$

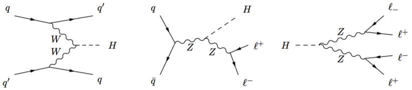
– expanded locally in θ, define score [just taylor log]

$$\log \frac{p(x|\theta)}{p(x|\theta_0)} \approx (\theta - \theta_0) \, \nabla_\theta \log p(x|\theta)\Big|_{\theta_0} \equiv (\theta - \theta_0) \, t(x|\theta_0) \equiv (\theta - \theta_0) \, \mathcal{O}^{\text{opt}}(x)$$

– parton level, as used in ATLAS [CPV, Schumacher]

$$p(x|\theta) \approx |\mathcal{M}|_0^2 + \theta |\mathcal{M}|_{\text{int}}^2 \qquad \Rightarrow \qquad t(x|\theta_0) \sim \frac{|\mathcal{M}|_{\text{int}}^2}{|\mathcal{M}|_0^2} \, ,$$

$\Rightarrow$ Easy at parton level, LEP physics...



Discrete symmetry [Brehmer, Kling, TP, Tait]

– CPV at dimension-6

– unique CP-observable [C-even, P-odd, $\hat{T}$-odd]

$$t \propto \epsilon_{\mu\nu\rho\sigma} \; k_1^\mu \; k_2^\nu \; q_1^\rho \; q_2^\sigma \; \text{sign}\,[(k_1 - k_2) \cdot (q_1 - q_2)] \xrightarrow{\text{lab frame}} \sin \Delta\phi_{jj}$$

$\Rightarrow$ Computable, modulo prefactor from D6-operator

Optimal observables after detector

Computing score using MadMiner [Brehmer, Kling, Espejo, Cranmer]

– likelihood ratio at detector level

$$\log \frac{p(x_d|\theta)}{p(x_d|\theta_0)} = \log \frac{\int dx_p \, T(x_d|x_p) \, p(x_p|\theta)}{\int dx_p \, T(x_d|x_p) \, p(x_p|\theta_0)}$$

– minimization problem for

$$F(x_d) = \int dx_p \; |g(x_d, x_p) - \hat{g}(x_d)|^2 \; T(x_d|x_p) \, p(x_p|\theta)$$

smart choice

$$g(x_d, x_p) = \frac{p(x_p|\theta)}{p(x_p|\theta_0)} \qquad \Rightarrow \qquad \hat{g}_*(x_d) = \frac{p(x_d|\theta)}{p(x_d|\theta_0)}$$

– same for unobservable phase-space directions [joint score $t(x, z|\theta)$]

$\Rightarrow$ Minimization means ML, function as NN

Optimal observables after detector

Computing score using MadMiner [Brehmer, Kling, Espejo, Cranmer]

– likelihood ratio at detector level

$$\log \frac{p(x_d|\theta)}{p(x_d|\theta_0)} = \log \frac{\int dx_p \; T(x_d|x_p) \; p(x_p|\theta)}{\int dx_p \; T(x_d|x_p) \; p(x_p|\theta_0)}$$

– minimization problem for

$$F(x_d) = \int dx_p \; \left| g(x_d, x_p) - \hat{g}(x_d) \right|^2 \; T(x_d|x_p) \; p(x_p|\theta)$$

smart choice

$$g(x_d, x_p) = \frac{p(x_p|\theta)}{p(x_p|\theta_0)} \qquad \Rightarrow \qquad \hat{g}_*(x_d) = \frac{p(x_d|\theta)}{p(x_d|\theta_0)}$$

– same for unobservable phase-space directions [joint score $t(x, z|\theta)$]

$\Rightarrow$ Minimization means ML, function as NN

Going back to formulas [Brehmer, Butter, TP, Soybelman]

– detector-level score from MadMiner

– parton-level score analytically

– good enough formula for controlled use?

$\Rightarrow$ Symbolic regression

PySR

Analytic formula for score [M Cranmer (2020)]

– function to approximate $t(x|\theta)$

– order-one phase space parameters $x_p = p_T/m_H, \Delta\eta, \Delta\phi$ [node]

– operators $\sin x, x^2, x^3, x+y, x-y, x*y, x/y$ [node]

– represent formula as tree [complexity = number of nodes]

$\Rightarrow$ figures of merit

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\Big[g_i(x) - t(x,z|\theta)\Big]^2$$

$$\text{score} \approx \text{MSE} + \text{parsimony} \cdot \text{complexity}$$

Simulated annealing

– combine trees to populations

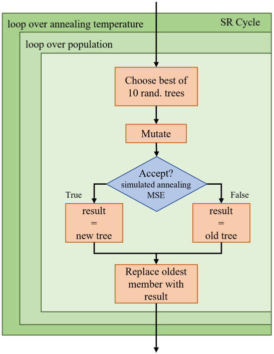
– mutate trees exchange, add, delete nodes

– acceptance probability

$$p = \exp\left(-\frac{\text{score}_\text{new} - \text{score}_\text{old}}{\alpha T \; \text{score}_\text{old}}\right)$$

– added: proper fit of pre-factors

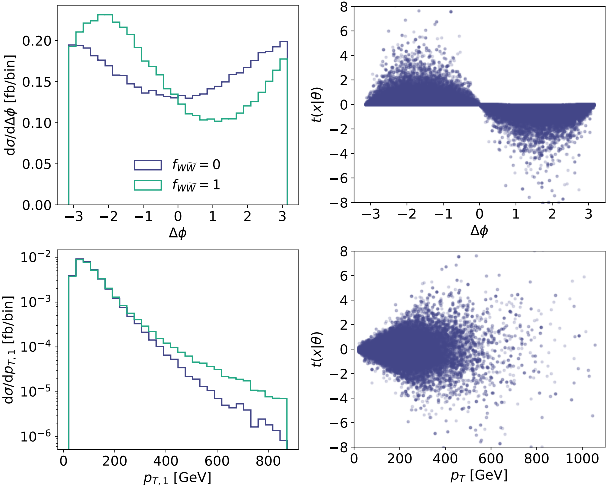
$\Rightarrow$ Hall of Fame: best equation per complexity



Score around Standard Model

Score around Standard Model

– shift in distributions, reflected in score [parton level]
CP-effect in $\Delta\phi_{jj}$
D6-effect in $p_{T,j}$



Score around Standard Model

Score around Standard Model

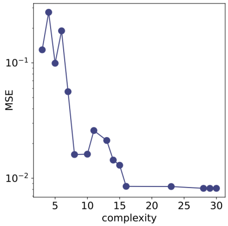
- shift in distributions, reflected in score [parton level]
  CP-effect in $\Delta\phi_{jj}$
  D6-effect in $p_{T,j}$
- best 4-parameter formula including $\Delta\eta$ [without/with detector]

$$t = -x_{p,1}\,(x_{p,2}+c)\,(a - b\Delta\eta)\sin(\Delta\phi + d)$$

with $a = 1.086(11)$ $b = 0.10241(19)$ $c = 0.24165(8)$ $d = 0.00662(32)$
$a = 0.926(2)$ $b = 0.08387(35)$ $c = 0.3542(20)$ $d = 0.00911(67)$

$\Rightarrow$ Mostly expected formula

compl	dof	function	MSE
3	1	$a\,\Delta\phi$	$1.30 \cdot 10^{-1}$
4	1	$\sin(a\Delta\phi)$	$2.75 \cdot 10^{-1}$
5	1	$a\Delta\phi x_{p,1}$	$9.93 \cdot 10^{-2}$
6	1	$-x_{p,1}\sin(\Delta\phi + a)$	$1.90 \cdot 10^{-1}$
7	1	$(-x_{p,1} - a)\sin(\sin(\Delta\phi))$	$5.63 \cdot 10^{-2}$
8	1	$(a - x_{p,1})x_{p,2}\sin(\Delta\phi)$	$1.61 \cdot 10^{-2}$
14	2	$x_{p,1}(a\Delta\phi - \sin(\sin(\Delta\phi)))(x_{p,2}+b)$	$1.44 \cdot 10^{-2}$
15	3	$-(x_{p,2}(a\Delta\eta^2 + x_{p,1}) + b)\sin(\Delta\phi + c)$	$1.30 \cdot 10^{-2}$
16	4	$-x_{p,1}(a - b\Delta\eta)(x_{p,2}+c)\sin(\Delta\phi + d)$	$8.50 \cdot 10^{-3}$
28	7	$(x_{p,2}+a)(bx_{p,1}(c - \Delta\phi)$ $-x_{p,1}(d\Delta\eta + ex_{p,2} + f)\sin(\Delta\phi + g))$	$8.18 \cdot 10^{-3}$



Score around Standard Model

Score around Standard Model

– shift in distributions, reflected in score [parton level]
  CP-effect in $\Delta\phi_{jj}$
  D6-effect in $p_{T,j}$

– best 4-parameter formula including $\Delta\eta$ [without/with detector]

$$t = -x_{p,1}\left(x_{p,2} + c\right)\left(a - b\Delta\eta\right)\sin(\Delta\phi + d)$$

with $a = 1.086(11)$ $b = 0.10241(19)$ $c = 0.24165(8)$ $d = 0.00662(32)$
$a = 0.926(2)$ $b = 0.08387(35)$ $c = 0.3542(20)$ $d = 0.00911(67)$

$\Rightarrow$ Mostly expected formula

Score away from Standard Model

– saturation effect of squared term

$$|\mathcal{M}(\theta)|^2 \sim p_0 + a\theta + b\theta^2 \qquad \Rightarrow \qquad t \sim \frac{\nabla_\theta |\mathcal{M}(\theta)|^2}{|\mathcal{M}(\theta)|^2} = \frac{a + 2b\theta}{p_0 + a\theta + b\theta^2}$$

– regression including division [rational functions]

$\Rightarrow$ Optimal observables more complex

cmpl	dof	function	MSE
3	1	$ax_{p,\times}$	0.124
12	2	$ax_{p,\times}/(x_{p,\times}/\Delta\eta + \Delta\eta + b)$	0.116
15	2	$(s_\phi + a)(-s_\phi + x_{p,\times} - b)/(-s_\phi + x_{p,\times} + \Delta\eta/x_{p,\times})$	0.054
26	4	$a/(b - (s_\phi - c - d/(s_\phi^2 - s_\phi\Delta\eta - s_\phi/x_{p,\times} + ex_{p,\times}^2))/x_{p,\times})$	0.048
31	7	$a/(b - (s_\phi + (cs_\phi^2 - d)/(es_\phi^2 x_{p,\times}^2 - s_\phi\Delta\eta + f) - g)/x_{p,\times})$	0.039

Score around Standard Model

Score around Standard Model

– shift in distributions, reflected in score [parton level]
CP-effect in $\Delta\phi_{jj}$
D6-effect in $p_{T,j}$

– best 4-parameter formula including $\Delta\eta$ [without/with detector]

$$t = -x_{p,1}\,(x_{p,2} + c)\,(a - b\Delta\eta)\sin(\Delta\phi + d)$$

with $a = 1.086(11)$ $\quad b = 0.10241(19)$ $\quad c = 0.24165(8)$ $\quad d = 0.00662(32)$
$\quad a = 0.926(2)$ $\quad b = 0.08387(35)$ $\quad c = 0.3542(20)$ $\quad d = 0.00911(67)$

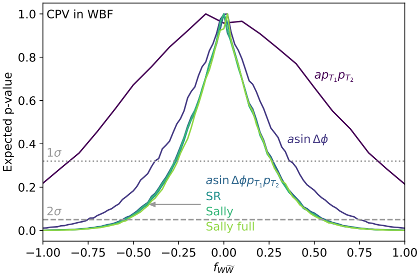
$\Rightarrow$ Mostly expected formula

So what does the formula buy us?

– expected limits:

very wrong formula
wrong formula
right formula
MadMiner

$\Rightarrow$ Statistically limitated for Run II...



ML for LHC Theory

ML-applications in LHC analysis and theory

– just another numerical tool for a numerical field

– driven by money from data industry, medical research

– goals are...

...improve established tasks
...develop new tools for straightforward tasks
...come up with new ideas, now possible

1- example: controlled forward/backward simulation with uncertainties

2- example: recovering formulas from numerics

$\Rightarrow$ Opportunity for young people to make a difference!