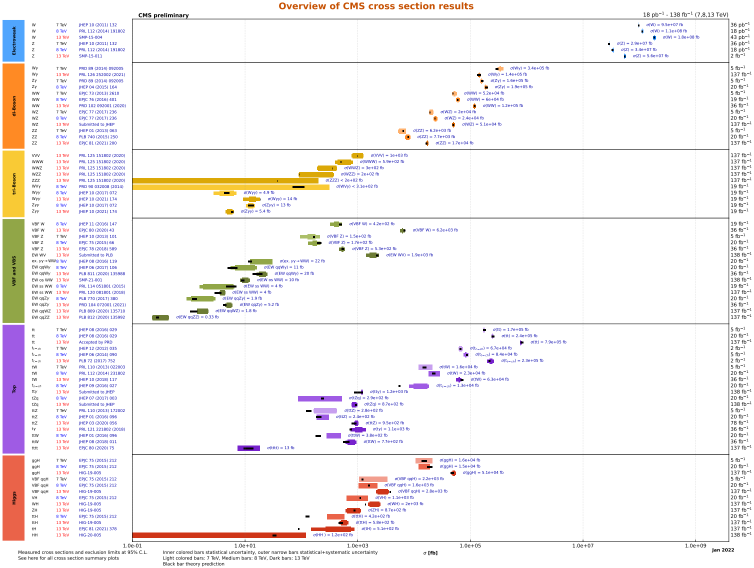
Machine-learning methods for polarisation tagging

Mathieu PELLEN

University of Freiburg

COMETA workshop on vector-boson polarisations
Toulouse, France

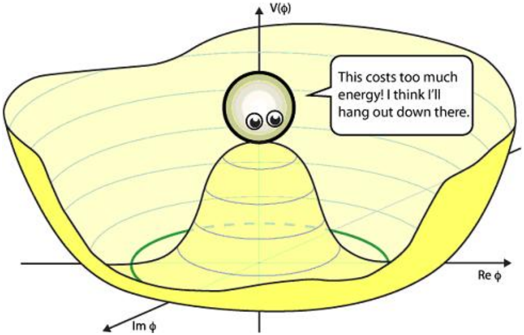
23rd of September 2024



- Triumph of the Standard Model ...

Motivation for polarisation studies

- Polarisation of gauge bosons (W and Z) related to Electroweak symmetry breaking
  → longitudinal polarisation
  "the Higgs mechanism is the conversion of Goldstone modes into the longitudinal polarisation mode of massive weak bosons" [Pelliccioli]
  → probe of new physics/extended Higgs sector



Definition

Master formula

$$\left|\mathcal{M}^{\mathrm{NWA}}(\mathrm{Zj})\right|^2 = \frac{\pi}{M_Z\Gamma_Z}\left|\sum_{h\in\Lambda}\mathcal{M}_h(\mathrm{pp}\to\mathrm{Zj})\cdot\mathcal{M}_h\left(\mathrm{Z}\to\ell^+\ell^-\right)\right|^2$$

with $\Lambda = \{+1,-1\}\,(\mathrm{Transverse}), 0\,(\mathrm{Longitudinal})$

Definition

Master formula

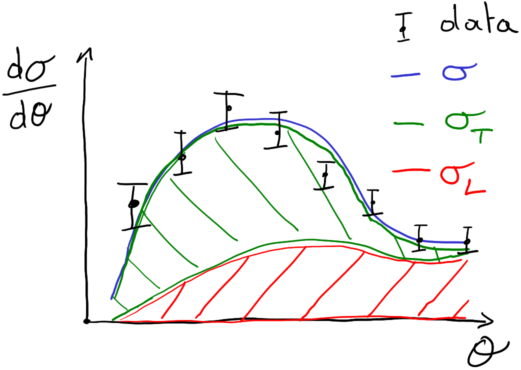
$$\left|\mathcal{M}^{\mathrm{NWA}}(\mathsf{Zj})\right|^2 = \frac{\pi}{M_Z\Gamma_Z}\left|\sum_{h\in\Lambda}\mathcal{M}_h(\mathsf{pp}\to\mathsf{Zj})\cdot\mathcal{M}_h\left(\mathsf{Z}\to\ell^+\ell^-\right)\right|^2$$

with $\Lambda = \{+1,-1\}\,(\mathrm{Transverse}), 0\,(\mathrm{Longitudinal})$

- Unpolarised cross section $\sigma_{\mathrm{unp.}} \sim |\mathcal{M}^{\mathrm{NWA}}|^2$ **(experimentally measured)**
- Polarised cross section: $\sigma_{\mathrm{L}} \sim |\mathcal{M}_0|^2 \cdot |\Gamma_0|^2$
- Polarisation fraction: $\boxed{f_{\mathrm{L}} = \sigma_{\mathrm{L}}/\sigma_{\mathrm{unp.}}}$

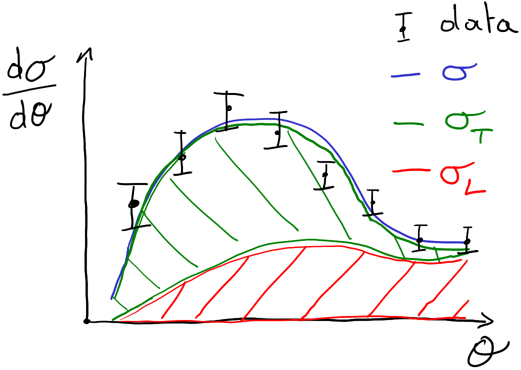
“Measuring polarisation”

- No measurement of polarisation: *template method*
  → extraction of parameters based on theory input



"Measuring polarisation"

- No measurement of polarisation: *template method*
  → extraction of parameters based on theory input



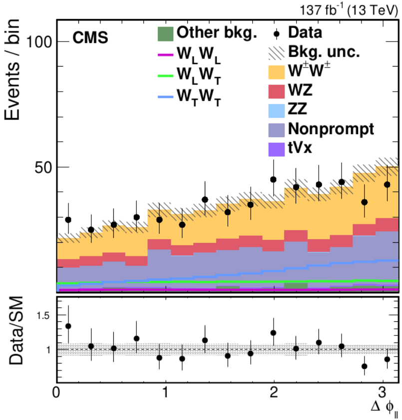
Shortcomings about polarisation extraction

⚠ Polarisation only defined for on-shell bosons

⚠ Only the unpolarised prediction is observed

Experimental analyses

- WZ: [ATLAS; 1902.05759, 2211.09435], [CMS; 2110.11231]
- Vector-boson scattering $W^\pm W^\pm$ [CMS; 2009.09429]



[CMS; 2009.09429]

Process	$\sigma\mathcal{B}$ (fb)	Theoretical prediction (fb)
$W^\pm_L W^\pm_L$	$0.32^{+0.42}_{-0.40}$	0.44 ± 0.05
$W^\pm_X W^\pm_T$	$3.06^{+0.51}_{-0.48}$	3.13 ± 0.35
$W^\pm_L W^\pm_X$	$1.20^{+0.56}_{-0.53}$	1.63 ± 0.18
$W^\pm_T W^\pm_T$	$2.11^{+0.49}_{-0.47}$	1.94 ± 0.21

Discussion

Drawbacks of the *template method*

- Restricted to given observables
- Unclear what are the *optimal* observables
- Interpretation at the integrated level

Discussion

Drawbacks of the *template method*

- Restricted to given observables
- Unclear what are the *optimal* observables
- Interpretation at the integrated level

Typical in experiment:

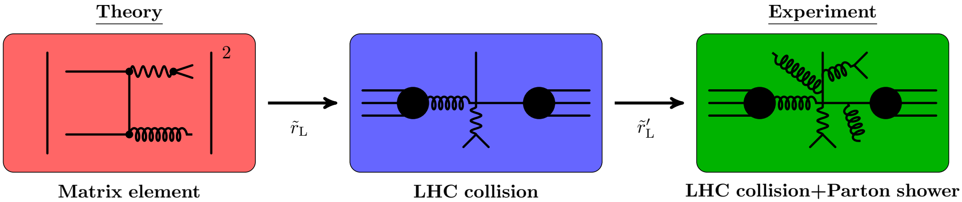
→ NN/BDT on samples of longitudinally polarised vs. background samples

Drawbacks:

- Difficult learning
  → large samples needed
- Ill defined discrimination between signal and background
  → Indirect link to polarisation definition
- Unclear what to feed (empirical and high-level variables)
  → possibly not optimal

- **New idea for polarisation extraction:**
  Amplitude-assisted tagging of longitudinally polarised bosons using wide neural networks

- **New idea for polarisation extraction:**
  Amplitude-assisted tagging of longitudinally polarised bosons using wide neural networks



General idea

Use amplitude to extract theory parameters / pseudo observables from data
$\rightarrow$ requires the use of neural network [Grossi, Incudini, MP, Pellicciolі; 2306.07726]

General idea

Use amplitude to extract theory parameters / pseudo observables from data
→ requires the use of neural network [Grossi, Incudini, MP, Pellicciolli; 2306.07726]

Advantages

- ~~Restricted to given observables~~ → No observable
- ~~Unclear what are the *optimal* observables~~ → optimal by definition
- ~~Interpretation at the integrated level~~
  → event-by-event interpretation (fully differential)

⚠ General method usable for other problems

General idea

Use amplitude to extract theory parameters / pseudo observables from data
→ requires the use of neural network [Grossi, Incudini, MP, Pellicciol; 2306.07726]

Advantages

- ~~Restricted to given observables~~ → No observable
- ~~Unclear what are the *optimal* observables~~ → optimal by definition
- ~~Interpretation at the integrated level~~
  → event-by-event interpretation (fully differential)

⚠ General method usable for other problems
→ Alternative approaches:

- *matrix-element method* [Kondo; J. Phys. Soc. Jap. 57 (1988) 4126-4140 / 60 (1991) 836–844.]
- *optimal-observable method* [Diehl, Nachtmann; Z. Phys. C 62 (1994) 397-412, hep-ph/9603207], [Janot; 1503.01325]
- *MELA* *(Matrix Element Likelihood Approach)* [Gao, Gritsan, Melnikov, Schulze, et al.; 1001.3396, 1208.4018,1309.4819, 1606.03107]

Simple observation

$$f_{\mathrm{L}}(\mathcal{O}) = \frac{\mathrm{d}\sigma_{\mathrm{L}}}{\mathrm{d}\mathcal{O}} \Bigg/ \frac{\mathrm{d}\sigma_{\mathrm{unp}}}{\mathrm{d}\mathcal{O}} \quad \text{with} \quad \sigma \propto \int \mathrm{d}\Phi |\mathcal{M}|^2$$

→ At the event-by-event/phase-space–point level, at leading order (LO), equivalent to

$$r_{\mathrm{L}} = \frac{|\mathcal{M}_{\mathrm{L}}|^2}{|\mathcal{M}|^2}$$

⚠ r_{L} is the probability for an event to be longitudinally polarised

Simple observation

$$f_{\rm L}(\mathcal{O}) = \frac{{\rm d}\sigma_{\rm L}}{{\rm d}\mathcal{O}} \Big/ \frac{{\rm d}\sigma_{\rm unp}}{{\rm d}\mathcal{O}} \quad \text{with} \quad \sigma \propto \int {\rm d}\Phi |\mathcal{M}|^2$$

→ At the event-by-event/phase-space–point level, at leading order (LO), equivalent to

$$r_{\rm L} = \frac{|\mathcal{M}_{\rm L}|^2}{|\mathcal{M}|^2}$$

⚠ $r_{\rm L}$ is the probability for an event to be longitudinally polarised

NB:

- All information about longitudinal polarisation contained in $r_{\rm L}$
- **If $r_{\rm L}$ was a physical observable,**
  only its measurement would be required to extract polarisation
  or
  Polarised predictions obtainable by reweighting unpolarised ones with $r_{\rm L}$

Application

- Z+j production at the LHC at $\sqrt{s} = 13.6\,\mathrm{TeV}$
- MadGraph5_aMC@NLO for checks

  [Alwall, et al.; 1405.0301], [Buarque Franzosi, Mattelaer, Ruiz, Shil; 1912.01725]
- Recola [Actis et al.; 1605.01090] for r_{L} computation
- Pythia [Sjöstrand et al.; 1410.3012] for PS

Application

- Z+j production at the LHC at $\sqrt{s} = 13.6$ TeV
- MadGraph5_aMC@NLO for checks

  [Alwall, et al.; 1405.0301], [Buarque Franzosi, Mattelaer, Ruiz, Shil; 1912.01725]
- Recola [Actis et al.; 1605.01090] for r_{L} computation
- Pythia [Sjöstrand et al.; 1410.3012] for PS

→ *Generation* set-up

$$p_{\mathrm{T,j}} > 10\,\mathrm{GeV}\,, \qquad |y_{\mathrm{j}}| < 5, \qquad \text{and} \qquad 76\,\mathrm{GeV} < M_{\mu^+\mu^-} < 106\,\mathrm{GeV}$$

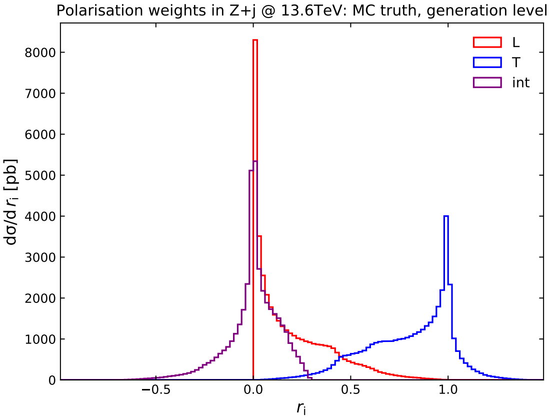
→ *Inclusive* set-up

$$p_{\mathrm{T,j}} > 20\,\mathrm{GeV}\,, \qquad |y_{\mathrm{j}}| < 4, \qquad \text{and} \qquad 81\,\mathrm{GeV} < M_{\mu^+\mu^-} < 101\,\mathrm{GeV}$$

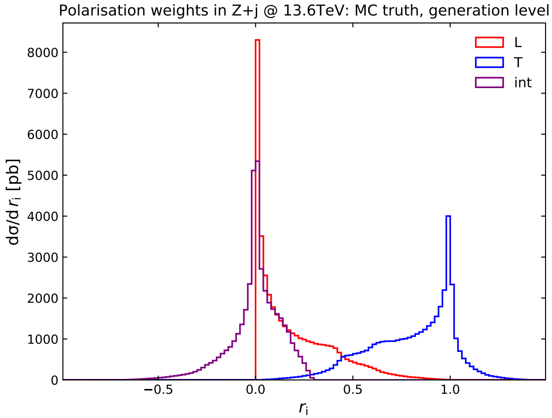
⚠ No cuts on Z-boson decay products

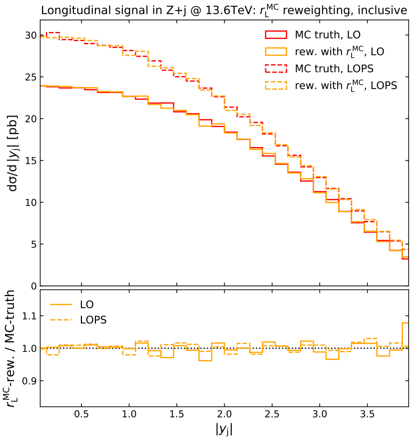
→ *Fiducial* set-up = *Inclusive* set-up +

$$p_{\mathrm{T},\mu^\pm} > 20\,\mathrm{GeV} \qquad \text{and} \qquad |y_{\mu^\pm}| < 2.7$$



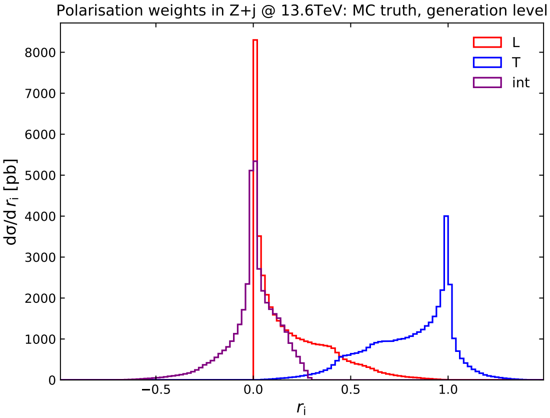
Polarisation weights in Z+j @ 13.6TeV: MC truth, generation level
L
T
int
dσ/d r_i [pb]
8000
7000
6000
5000
4000
3000
2000
1000
0
−0.5
0.0
0.5
1.0
r_i

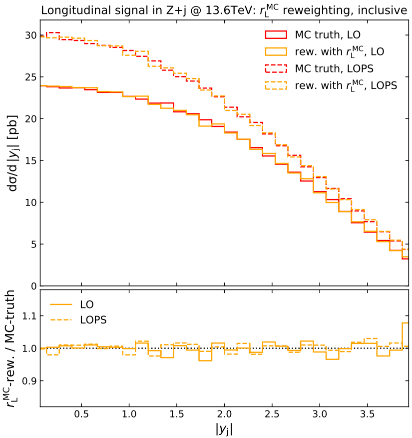




<u>NB:</u> For LO+PS, reweighting done on events before PS

→ assumption: polarisation is not influenced by PS





NB: For LO+PS, reweighting done on events before PS

→ assumption: polarisation is not influenced by PS

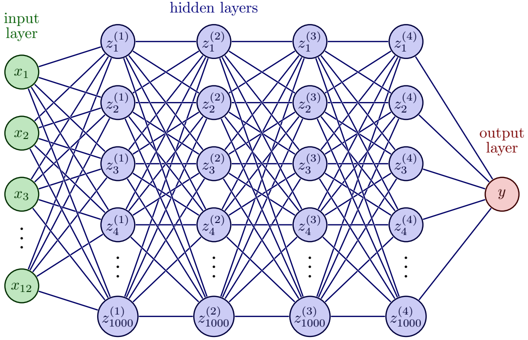
⚠ r_L requires knowledge of all momenta (initial and final)

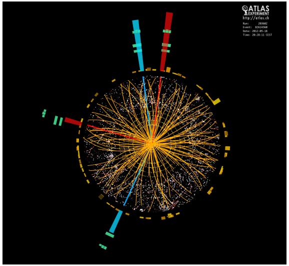
⚠ r_L requires knowledge of the partonic process and the PDF: ($qg \to q$, $q\bar{q} \to g$, ...)

Solution → Wide neural networks

IN:

- Training with Monte Carlo events
  → Truth
- Parameter to learn: r_{L}
- Input: all accessible information (leptons, jets, ...)

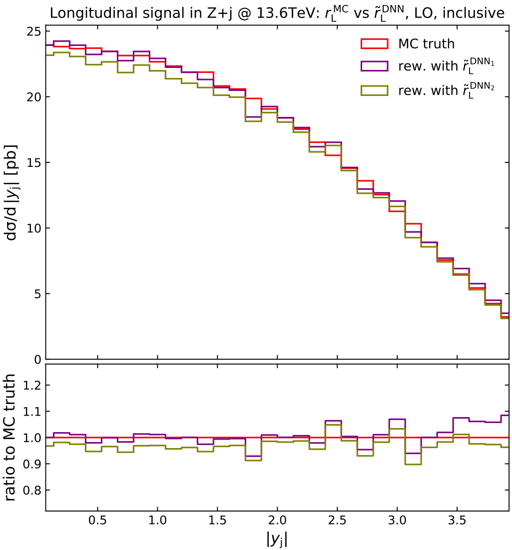




OUT:

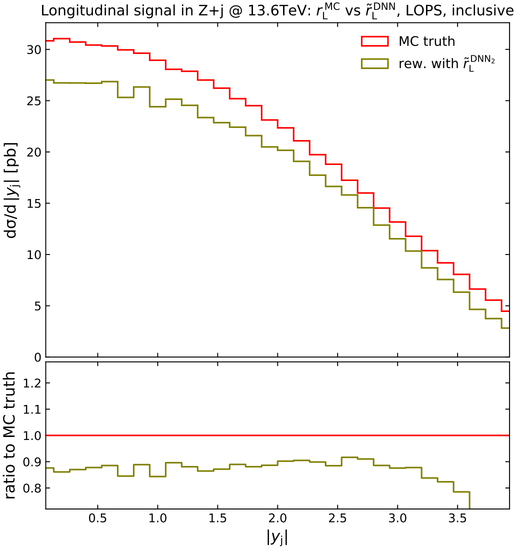
- Result: $\tilde{r}_{\mathrm{L}}$
  → Proxy of r_{L}
  → Relies only on accessible information

LO



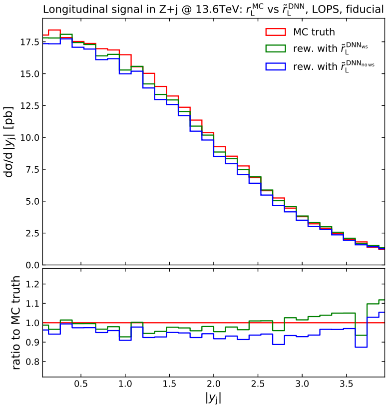
→ Method is working at LO at per-cent level

LO+PS using LO training



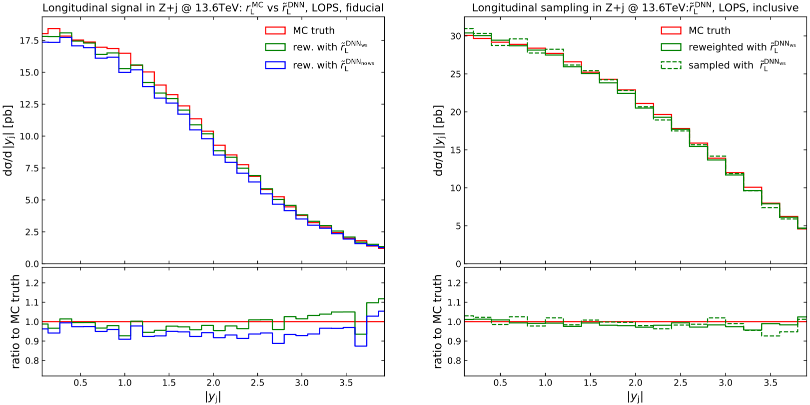
⚠ Failing!
(describing PS corrections instead of polarisation)
→ Retraining ...
... with r_L computed before PS!

LO+PS with warm-up



→ Warmup training gives better results
Initial conditions of the LO+PS learning is set by LO learning

LO+PS with warm-up



→ Warmup training gives better results
Initial conditions of the LO+PS learning is set by LO learning

NB: LO+PS is better reproduced than LO (less $\tilde{r}_L < 0$) → mitigation effect?

Experiment / Theory

- $\tilde{r}_{\rm L}$ is an approximation of $r_{\rm L}$ relying only on physical inputs
  $\rightarrow$ Use wide neural network
- Given a set of data/unpolarised sample
  $\rightarrow$ $\tilde{r}_{\rm L}$ allows to tag/reweight longitudinally-polarised events
- Can be used in experimental analysis/theoretical calculations
  $\rightarrow$ f_L extracted to be compared to theory predictions

Discussion

- Validity of the method
  - Training inclusive, to be used within phase-space
    $\longrightarrow$ out-of-support extrapolation possible?
  - One training per process at a given energy
    $\longrightarrow$ Applicable at different energy?

Discussion

- Validity of the method
  - Training inclusive, to be used within phase-space
    $\rightarrow$ out-of-support extrapolation possible?
  - One training per process at a given energy
    $\rightarrow$ Applicable at different energy?
- Error propagation
  - training statistics: inferred by training with different data sets
  - theory: scale variation (ratio $\rightarrow$ small)
  - experimental errors: training with pseudo data

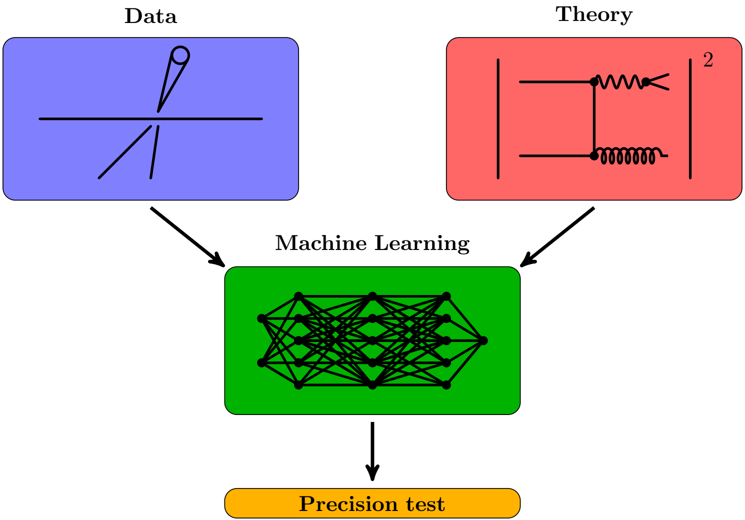
Discussion

- Validity of the method
  - Training inclusive, to be used within phase-space
    $\rightarrow$ out-of-support extrapolation possible?
  - One training per process at a given energy
    $\rightarrow$ Applicable at different energy?
- Error propagation
  - training statistics: inferred by training with different data sets
  - theory: scale variation (ratio $\rightarrow$ small)
  - experimental errors: training with pseudo data
- Model independence
  $\rightarrow$ Any model can be used *e.g.* EFT or simplified models

Outlook

Natural extensions

- Extension of the method beyond LO
  → NLO QCD first
- Test on multi-boson processes
  → di-boson, tri-boson, vector-boson scattering
- Application to other problems (castable into ratios)
  → irreducible backgrounds: ttbb, vector-boson scattering, ...


Data
Theory
2
Machine Learning
Precision test

Conclusion

New idea for extraction of boson polarisation:

[Grossi, Incudini, MP, Pelliccioli; 2306.07726]

Conclusion

New idea for extraction of boson polarisation:

[Grossi, Incudini, MP, Pelliccioli; 2306.07726]

- Decisive information for SM tests
  $\rightarrow$ Precision programme at the LHC
- Crucial interplay between theory and experiment
  $\rightarrow$ Big impact on physics results

Conclusion

New idea for extraction of boson polarisation:

[Grossi, Incudini, MP, Pelliccioli; 2306.07726]

- Decisive information for SM tests
  → Precision programme at the LHC
- Crucial interplay between theory and experiment
  → Big impact on physics results

Thank you

BACK-UP