ML4Jets2024 07 November 2024

université **PARIS-SACLAY**

An implementation of Neural Simulation-Based Inference for Parameter Estimation in ATLAS

Arnaud Maury, on behalf of the ATLAS Collaboration

Run 2 analysis of the off-shell Higgs boson decaying into four leptons

- 1 analysis, 2 papers:
- A Physics measurement paper:

<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-016/>

• An ML-focused methodology paper (this talk): <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-015/>

The motivation for Neural Simulation-Based Inference (NSBI)

Typical LHC Workflow

- Detector has O(100 million) sensors
- Can't build 100M dimensional histogram
- ‣ Reconstruction pipeline, event selection
- ‣ Design sensitive one-dimensional observable

Density Estimation: What we're used to doing..

Measure signal strength *μ*

With histograms we can ask "Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?"

Density Estimation: What we're used to doing..

Measure signal strength *μ*

With histograms we can ask "Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?"

New challenge: Non-linear changes in kinematics (w.r.t. parameter of interest)

A histogram of any single observable is no longer optimal (see Ghosh et al: [hal-02971995\(p172\)\)](https://hal.science/hal-02971995v3/), but neural networks estimate high-dimensional likelihood ratios (see Cranmer et al: [arXiv:1506.02169](https://arxiv.org/abs/1506.02169)) !

Campbell et al: [arXiv:1311.3589](https://arxiv.org/abs/1311.3589)

New challenge: Non-linear changes in kinematics (w.r.t. parameter of interest)

A histogram of any single observable is no longer optimal (see Ghosh et al: [hal-02971995\(p172\)\)](https://hal.science/hal-02971995v3/), but neural networks estimate high-dimensional likelihood ratios (see Cranmer et al: [arXiv:1506.02169](https://arxiv.org/abs/1506.02169)) ! processes do so through their low-mass o-shell tails (see the Feynman diagrams for the main $\frac{1}{2}$ and $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ processes. The diagrams define the di effectival and <u>anxiv:1506.02169</u> the off-shell Higgs boson production signal strengths. <u>naruze resuperz</u>y, but nourar notwono-oomnato $eted$ and $erViv.$ 1506.02160

"Neural Simulation-Based Inference" channels, due to the different schemes used to derive the uncertainties. The hypothesis of systematic \mathbf{I} is the 4∪ and 2∪2 \mathbf{I} and \mathbf{I} and dominant sources of uncertainties, or the dominant sources of uncertainties, \mathbf{I}

The Control of Commission for the 4,000 days of the 4,000 days of the 4,000 days of the 4,000 days of the 4,000
The 4,000 days in the 1,000 days of the 4,000 days of the 4,000 days of the 4,000 days of the 4,000 days of th

Traditional framework: after the full fit to data with the total systematic uncertainty from the sources described in \mathcal{L} Section 3 are shown in the figure. The figure in the MOCK of the NN observables used in the 4∪ channel are 4∪ c

T distribution for the 2∑2a channel are shown in Figure 5.25 channel are shown in Figure 5.25 channel are shown in

 μ is flow off-shell signal regions in the $/$ $/$ μ is now arbitrary parameter of interest(s)

High-dim data

"Neural Simulation-Based Inference" channels, due to the different schemes used to derive the uncertainties. The hypothesis of systematic \mathbf{I} is the 4∪ and 2∪2 \mathbf{I} and \mathbf{I} and dominant sources of uncertainties, or the dominant sources of uncertainties, \mathbf{I}

<u>The neural inference framework:</u> neural inference framework[.] The fitted background in the factor of \sim <u>noural innoronoo irain</u>

The Control of Commission for the 4,000 days of the 4,000 days of the 4,000 days of the 4,000 days of the 4,000
The 4,000 days in the 1,000 days of the 4,000 days of the 4,000 days of the 4,000 days of the 4,000 days of th

T distribution for the 2∑2a channel are shown in Figure 5.25 channel are shown in Figure 5.25 channel are shown in

 μ is flow off-shell signal regions in the $/$ $/$ μ is now arbitrary parameter of interest(s)

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them into likelihood (ratio) model
- Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Open problems to extend to full ATLAS analysis:

- ‣ Robustness: Design and validation
- Systematic Uncertainties: Incorporate them into likelihood (ratio) model
- Neyman Construction: Sampling pseudo-experiments in a per-event analysis

 runs over different physics process *j* $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ ρ is the probability density of the event j runs over different ph
(Eq. $g g \to H^* \to 4l, g$ $1-2, 99$ rate for the 1.7 σ

Example use case captured using only the coefficients 5 178 9 178 9 178 9 178 9 178 9 178 9 178 9 178 9 178 9 178 9 178 9 178 9
Experimental rate and the total rate and the total rate of a decomposition is possible for a decomposition is p wide range of LHC analyses where the coefficients 5 9(\pm 179 \pm

Search-Oriented Mixture Model 174 states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

General Formula

¹⁷⁵ ⇠ different components, representing different physics processes,

$$
p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \ p_j(x_i)
$$

Search-Oriented Mixture Model ²¹¹ of NSBI to fully account for these non-linear effects. 174 states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

General Formula 22 **Changes of Street industry only (Seneral Formulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation including simulation**

$$
p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \ p_j(x_i)
$$

Example use case
\n
$$
p_{ggF}(x|\mu) = \frac{1}{v_{ggF}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_B(x) \right]
$$

 j runs over different physics process $(Eg. gg \to H^* \to 4l, gg \to ZZ \to 4l)$ ρ is the probability density of the event j runs over different ph
(Eq. $g g \to H^* \to 4l, g$ $1-2, 99$ rate for the 1.7 σ

nple use case

 $(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_B(x)$

¹⁷⁵ ⇠ different components, representing different physics processes,

Search-Oriented Mixture Model ²¹¹ of NSBI to fully account for these non-linear effects. 174 states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

22 **Changes of Street industry only (Seneral Formulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation including simulation** 2.31 interference effects 6.6 for the subscript indicates the subscript indicates that 2.5 for $p(x|u) = \frac{1}{\sqrt{2\pi}} \int f(u) du = \frac{1}{\sqrt{2\pi}} \int f(u) du$ $P^{(v_l|\mu)} = \nu(\mu)$ and $P^{(w)}$ in principle a coupling model with several scales that scales that scales the signal scales that scale 2 amplitude complex number, which would lead to a phase contributing to the interference term interference term interference term interference term in terms of the interference terms of the interference terms of the int 221 gebruik in Front Albert Annel 19 gebruik in deur Annel 19 gebruik in de Front Annel 19 gebruik in de Front
22 juil 19 gebruik in de Groep de Front Annel 20 gebruik in de Front Annel 20 gebruik in Front Annel 19 gebruik $p_{ggF}(x|\mu) =$ 1 $\nu_{\rm ggF}(\mu)$ ¹⁷⁵ ⇠ different components, representing different physics processes, $p(x_i|\mu) =$ 1 $\nu(\mu)$ $\breve{\bm{\nabla}}$ \overline{C} \boldsymbol{j} $f_j(\mu) \cdot \nu_j \, p_j(x_i)$ $\begin{array}{rcl} j \ \text{runs over different ph} \ \end{array}$ (Eq. $gg \rightarrow H^* \rightarrow 4l, g$ $1-2,00$ rate for the \sim known analy the cory model width range of LHC analyses where the case $\frac{1}{2}$ $p_{ggF}(x|\mu) = \frac{1}{\nu_{F}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_B(x) \right]$ $v_{\text{ggF}}(\mu)$ and the parameters of c

General Formula

Example use case

 $\left[\left(\mu - \sqrt{\mu} \right) v_S \, p_S(x) + \sqrt{\mu} \, v_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + \left(1 - \sqrt{\mu} \right) v_B \, p_B(x) \right]$

General Formula

Example use case

 $\left[\left(\mu - \sqrt{\mu} \right) v_S \, p_S(x) + \sqrt{\mu} \, v_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + \left(1 - \sqrt{\mu} \right) v_B \, p_B(x) \right]$ $p_{ggF}(x|\mu) = \frac{1}{v_{F}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_S(x) \right]$ $\frac{\nu_{\text{ggF}}(\mu)}{2}$

Search-Oriented Mixture Model of NSBI to fully account for these non-linear effects. states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

Known analytically from theory model und als alternatives and

Example use case

Search-Oriented Mixture Model ¹⁸³ physics processes can always be made, if the dependence on the parameter of interest is not analytically Search-Oriented Mixture Model ²¹¹ of NSBI to fully account for these non-linear effects.

$$
\underbrace{f_j(\mu)}_{p_{\text{ref}}(\chi_i)}(v_j) p_j(x_i) \qquad \qquad \underbrace{p(x_i|\mu)}_{p_{\text{ref}}(\chi_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot v_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}
$$

1991

 $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ J funs over different pnysics process
 $(Eq. \, gg \rightarrow H^* \rightarrow 4l, \, gg \rightarrow ZZ \rightarrow 4l).$

's interesting a collection
|-
|

$$
p_{ggF}(x|\mu) = \frac{1}{v_{ggF}(\mu)} \left[\left(\mu - \sqrt{\mu} \right) v_S \, p_S(x) + \sqrt{\mu} \, v_{SBI_1} \, p_{SBI_1}(x) + \left(1 - \sqrt{\mu} \right) v_B \, p_B(x) \right]
$$

Search-Oriented Mixture Model ¹⁸³ physics processes can always be made, if the dependence on the parameter of interest is not analytically Search-Oriented inixture inoder
Search-Oriented instructed instead to estimate the corresponding to 2018 in 1949 in 1949 in 1949 in 1949 in 19 ²¹¹ of NSBI to fully account for these non-linear effects. 216 Search-Oriented Mixture Model 174 states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

Event rates estimated from simulations \sim 1988 for the reference, the reference, the reference, the reference, the reference, the reference of signal processes, the reference of signal processes, the reference of signal j runs over different physics process
 $\overline{(Eq \rho \rho \rightarrow H^* \rightarrow 41 \rho \rho \rightarrow ZZ \rightarrow 41)}$ **CERTAL FORMULA** GENERAL FORMULA ¹⁸⁵ the following formalism. This paper defines a *search-oriented* mixture model, which is the probability $m(x|u)$ 1 22 **Changes of Street industry only (Seneral Formulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation including simulation** 2.31 interference effects 6.6 for the subscript indicates the subscript indicates that 2.5 for $p(x|u) = \frac{1}{\sqrt{f(u)}} \sum_{x} p(x)$ provided by $p(x|u) = \frac{C}{\sqrt{g(u)}}$ $P^{(v_l|\mu)} = (v(\mu)) \sum_{i=1}^{\infty} V^{(v_l)}$ in $P^{(v_l)} = \frac{1}{n} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_j(\mu) \cdot v_j \frac{1}{n}$ P respectively P respectively P respectively to P respectively to a phase contribution to the interference term P respectively to the interference term P respectively. Event rates estimated from simulations $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow$ X nown analytically from theory model 221 gebruik in Front Albert Annel 19 gebruik in deur Annel 19 gebruik in de Front Annel 19 gebruik in de Front
22 juil 19 gebruik in de Annel 19 gebruik in de Front Annel 19 gebruik in de Front Annel 19 gebruik in de Front $p_{ggF}(x|\mu) =$ 1 $\nu_{\rm ggF}(\mu)$ 218 General Formula require the computation. This would require the measurement of two independent parameters of interest, which can be done with \mathcal{C} $(\mathcal{V}(\mu))$ is the expression of $\mathcal{V}(\mu)$?ggF(G|`) = 1
1
1 ulation RNOWN ANALYCLEALLY FROM CHEORY MODEL
Captured using only the contribution from the cory model 223 represented using 223 represented using 223 represented using 223 representation in the non-linearity in the non- $\begin{array}{cc} \nu_{ggF}(\mu) \end{array}$ $p(x|\mu)$ $p_{\rm S}(x)$ = 1 $\nu(\mu)$ $p(x|\mu)$ 1 $\left[\begin{array}{ccc} 1 & \sqrt{1-x} & \sqrt$ ¹⁷⁵ ⇠ different components, representing different physics processes, $p(x_i|\mu) =$ 1 $\nu(\mu)$ $\breve{\bm{\nabla}}$ \overline{C} \dot{L} \widetilde{f} is over different photographic density \widetilde{f} correspending to the event \widetilde{g} (Eq. $gg \to H^* \to 4l$, g Event rates estimated from simulations \sim width range of LHC analyses where the case $\frac{1}{2}$ $p_{ggF}(x|\mu) = \frac{1}{v_{F}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_S(x) \right]$ $\frac{\nu_{\text{ggF}}(\mu)}{2}$ $p_S(x)$ $\nu(\mu)$ \lfloor μ μ ρ μ ρ μ ρ \lceil ρ ρ \lceil

10

Known analytically from theory model ⇠ signals in the City

Example use case

$$
x|\mu) = \frac{1}{v_{\text{eff}}(\mu)} \left[(\mu - \sqrt{\mu}) v_{\text{S}} p_{\text{S}}(x) + \sqrt{\mu} v_{\text{SBI}_1} p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) v_{\text{B}} p_{\text{B}}(x) \right]
$$

$$
u) = \frac{1}{\underbrace{\rho(u)}} \sum_{j} \underbrace{f_j(\mu)}_{j} \underbrace{\rho_j(x_i)}_{j} p_j(x_i) \qquad \qquad \frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}
$$

1991

 $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ J funs over different pnysics process
 $(Eq. \, gg \rightarrow H^* \rightarrow 4l, \, gg \rightarrow ZZ \rightarrow 4l).$

's interesting a collection
|-
|

$$
\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu}) v_S + \sqrt{\mu} v_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) v_B \frac{p_B}{p_S} \right]
$$

Search-Oriented Mixture Model ¹⁸³ physics processes can always be made, if the dependence on the parameter of interest is not analytically Search-Oriented inixture inoder
Search-Oriented instructed instead to estimate the corresponding to 2018 in 1949 in 1949 in 1949 in 1949 in 19 ²¹¹ of NSBI to fully account for these non-linear effects. 216 Search-Oriented Mixture Model 174 states with a coefficient that is some function of the parameter of 100 interest. If the decomposition is into the decomposition in the decomposition is interest. If the decomposition is interest. In the decomposit

Know Gene Examp ¹⁸⁵ the following formalism. This paper defines a *search-oriented* mixture model, which is the probability $m(x|u)$ 1 22 **Changes of Street industry only (Seneral Formulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation, and the combined simulation including simulation** 221 gebruik in Front Albert Annel 19 gebruik in deur Annel 19 gebruik in de Front Annel 19 gebruik in de Front
22 juil 19 gebruik in de Annel 19 gebruik in de Front Annel 19 gebruik in de Front Annel 19 gebruik in de Front $p_{ggF}(x|\mu) =$ 1 $\nu_{\rm ggF}(\mu)$ $\frac{p(x|\mu)}{\mu} = \frac{1}{\mu} \left[(\mu - \sqrt{\mu}) y_S + \sqrt{\mu} y_{SBL} \cdot \frac{p_{SBL}(x)}{p_{SBL}} + (1 - \sqrt{\mu}) y_B \right]$ $(\mathcal{V}(\mu))$ is the expression of $\mathcal{V}(\mu)$?ggF(G|`) = 1
1
1 $\begin{array}{cc} \nu_{ggF}(\mu) \end{array}$ $p(x|\mu)$ $p_{\rm S}(x)$ = 1 $\nu(\mu)$ $p(x|\mu)$ 1 $\left[\begin{array}{ccc} 1 & \sqrt{1-x} & \sqrt$ ¹⁷⁵ ⇠ different components, representing different physics processes, $p(x_i|\mu) =$ 1 $\nu(\mu)$ $\breve{\bm{\nabla}}$ \overline{C} \dot{L} \widetilde{f} is over different photographic density \widetilde{f} correspending to the event \widetilde{g} (Eq. $gg \to H^* \to 4l$, g Event rates estimated from simulations \sim known analy the cory model width range of LHC analyses where the case $\frac{1}{2}$ $p_{ggF}(x|\mu) = \frac{1}{v_{F}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{SBI_1} p_{SBI_1}(x) + (1 - \sqrt{\mu}) v_B p_S(x) \right]$ $\frac{\nu_{\text{ggF}}(\mu)}{2}$ $p_S(x)$ $\nu(\mu)$ \lfloor μ μ ρ μ ρ μ ρ \lceil ρ ρ \lceil

 k a parameterised network can be trained instead to estimate the trained instead to $\mathcal{S}(\mathcal{S})$ in $\mathcal{S}(\mathcal{S})$

General Formula
\n
$$
p(x_i|\mu) = \frac{1}{\sqrt{\mu(\mu)}} \sum_{j} \widehat{f_j(\mu)}(V_j) p_j(x_i)
$$
\n
$$
p_{ref}(x_i) = \frac{p(x_i|\mu)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu)} \sum_{j} \sum_{j} f_j(\mu) \cdot \nu \left(\frac{p_j(x_i)}{p_{ref}(x_i)}\right)
$$
\n
$$
j \text{ runs over different physics process}
$$
\n
$$
E
$$
\n
$$
F
$$
\n

$$
\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu}) \nu_S + \sqrt{\mu} \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B}{p_S} \right]
$$

A separate classifier per physics process j $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$

*H*_{ref}: Reference hypothesis Robust, parameterised classifier without parameterising ¹⁸³ physics processes can always be made, if the dependence on the parameter of interest is not analytically bust, parameterised classifier without parameterising $\overline{}$ ¹⁸⁵ the following formalism. This paper defines a *search-oriented* mixture model, which is the probability

 $p(x_i|\mu)$ $p_{ref}(x_i)$ = 1 $\nu(\mu)$ $\breve{\nabla}$ \overline{C} \boldsymbol{j} $f_j(\mu) \cdot \nu_j$ $p_j(x_i)$ $p_{ref}(x_i)$

A separate classifier per physics process j A separate classifier per priysics process j
 $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ $\frac{1}{8}$ freedom to make any choice for the reference, the reference, this paper defines it as a combination of signal processes, the reference, the reference, the reference signal processes of signal processes, the refe

H_{ref} : Reference hypothesis Robust, parameterised classifier without parameterising $\ddot{\bullet}$ **neterisi** 5 ⁹(`) · a ⁹ ? ⁹(G8)*,* (8)

$$
\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}
$$

where the contract of the contract of

A separate classifier per physics process j $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ \sim UUUU J $\therefore \rightarrow 4l, gg \rightarrow LL \rightarrow 4l$ 11 define the effective couplings ^⁺ used to define the off-shell Higgs boson production signal strengths. $\left(\begin{array}{c} 0 \\ 0 \end{array} \right)$ \rightarrow 40 A separate classifier per priysics process just $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ and the independent of $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ $\frac{1}{8}$ freedom to make any choice for the reference, the reference, this paper defines it as a combination of signal processes, the reference, the reference, the reference signal processes of signal processes, the refe

*H*_{ref}: Reference hypothesis Robust, parameterised classifier without parameterising \ddotsc **neterisi** 5 ⁹(`) · a ⁹ ? ⁹(G8)*,* (8)

$$
\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j f_j(\mu) \cdot \nu_j \frac{p_j(x)}{p_{\text{ref}}(x)}
$$

where the contract of the contract of

 $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ $\therefore \rightarrow 4l, gg \rightarrow LL \rightarrow 4l$ 11 define the effective couplings ^⁺ used to define the off-shell Higgs boson production signal strengths. $\left(\begin{array}{c} 0 \\ 0 \end{array} \right)$ \rightarrow 40 $\frac{1}{8}$ freedom to make any choice for the reference, the reference, this paper defines it as a combination of signal processes, the reference, the reference, the reference signal processes of signal processes, the refe

*H*_{ref}: Reference hypothesis Robust, parameterised classifier without parameterising \ddotsc **neterisi** 5 ⁹(`) · a ⁹ ? ⁹(G8)*,* (8)

$$
\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j f_j(\mu) \cdot \nu_j \frac{p_j(x)}{p_{\text{ref}}(x)}
$$

*H*_{ref}: Reference hypothesis Robust, parameterised classifier without parameterising \ddotsc **neterisi** 5 ⁹(`) · a ⁹ ? ⁹(G8)*,* (8)

$$
\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j f_j(\mu) \cdot \nu_j \frac{p_j(x)}{p_{\text{ref}}(x)}
$$

 $(Eq. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$ define the effective couplings ^⁺ used to define the off-shell Higgs boson production signal strengths. $\left(\begin{array}{c} 0 \\ 0 \end{array} \right)$ $\frac{1}{8}$ freedom to make any choice for the reference, the reference, this paper defines it as a combination of signal processes, the reference, the reference, the reference signal processes of signal processes, the refe

Reference Sample 164 expressed using the finite of the finite number of \sim 9 μ . While the independent density ratios, μ

A combination of signal samples, to ensure there's non-vanishing support in pre-selected region ⇒ In our dataset, $p_{\textit{ref}}(\ \cdot\) = p_{\textit{S}}(\ \cdot\)$ $p_{ref}(x_i) =$ 1 $\overline{\sum}$ k v_k Csignals We do not \boldsymbol{k} $v_k \cdot p_k(x_i)$

165 freedom to make any choice for the reference, the reference, this paper defines it as a combination of signal processes, the reference, the reference, the reference, the reference of signal processes, it as a combinati

$$
p_{\text{ref}}(x_i) = \frac{1}{\sum_{k} \nu}
$$

 $p_{ref}(\cdot)$ can be r

$$
t_{\mu} = -2 \ln \left(\frac{L_{\text{full}}(\mu, \widehat{\widehat{\alpha}})/L_{\text{ref}}}{L_{\text{full}}(\widehat{\mu}, \widehat{\alpha})/L_{\text{ref}}} \right)
$$

Choice of $p_{ref}(\,\cdot\,)$ can be made purely on numerical stability of training, as it drops out from the likelihood ratio The eqs. 9 G $p_{ref}(\cdot)$ can be made purely on numerical stability of training,
as it drops out from the likelihood ratio ¹⁶⁹ of \ which allows the final profile likelihood ratio constructed with this method to be independent of nade pure rely on r ro \mathbf{m} the like .
ike *Ierical Stability of training,*

 $\frac{\widehat{\alpha}}{\sqrt{2}}$ $\frac{1}{2}$ μ $\widehat{\alpha}$)/*L*_{ref} $\overline{}$ \setminus

Reweighting: Calculate weights w_i for events x_i in green sample to match blue sample

Validate quality of LR estimation with re-weighting task

$$
w_i = r_j(x_i) = \frac{p_j(x_i)}{p_{ref}(x_i)}
$$

Already estimated using an ensemble of networks

Open problems to extend to full ATLAS analysis: ✓ Robustness: Design and validation

-
-

‣ Systematic Uncertainties: Incorporate them into likelihood (ratio) model • Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Systematic uncertainties

Image: [arXiv:2105.08742](https://arxiv.org/abs/2105.08742)

Systematic uncertainties

ter
ter α and α is the mathematic α t to be called mass of the mass of the mission α_k

Image: [arXiv:2105.08742](https://arxiv.org/abs/2105.08742)

 10σ and have eimulations of 2 yer a and only have simulations at 0 w

Known interpolation strategies

Subsetz in Root available for normalization uncertainteering for normalization uncertainteering for normalization uncertainteering for normalization uncertainteering for normalization uncert

(OverallSys) in the subsequent patch releases. In future releases, this may become the default.

[See](#page-78-0) formula used

)α

 \Rightarrow Combine these traditional interpolation with neural network estimation of per-event likelihood ratios

=

 $p_{ref}(x_i)$

 $\overline{\text{See}}$ details of vertical interpolation for $G_{\!j}(\alpha_{\scriptscriptstyle{k}}), g_{\!j}^{}(x_{\scriptscriptstyle{i}},\alpha_{\scriptscriptstyle{k}})$

Probability density ratio including nuisance parameters (*α*)

 $p(x_i | \mu, \alpha)$

 $p(x_i | \mu, \alpha)$ *pref*(*xi*) = $\frac{\mu,\alpha}{\mu(\chi_i)} = \frac{1}{\nu(\mu,\alpha)}$ *C* ∑ *j f* $f_j(\mu) \cdot \nu_j \cdot$ $p_j(x_i)$ $p_{ref}(x_i)$ ⋅ N_{syst} ∏ *k* $G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$

 $p(x_i | \mu, \alpha)$ *pref*(*xi*) = $\frac{\mu,\alpha}{\mu(\chi_i)} = \frac{1}{\nu(\mu,\alpha)}$ *C* ∑ *j f* $f_j(\mu) \cdot \nu_j \cdot$ $p_j(x_i)$ $p_{ref}(x_i)$ ⋅ N_{syst} ∏ *k* $G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$

 $\frac{p(x_i | \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sqrt{\mu}},$ *pref*(*xi*) = *ν*(*μ*, *α*) *C* ∑ *j f* $f_j(\mu) \cdot \nu_j \cdot$ $p_j(x_i)$ $p_{ref}(x_i)$ ⋅ *Nsyst* ∏ *k* $G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$ We have this already $g_j(x_i, \alpha_k) =$ $p_j(x_i, \alpha_k)$ $p_j(x_i)$ α_1 α_2

Estimate from simulations and existing interpolation methods

 $\frac{p(x_i | \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sqrt{\mu}},$ *pref*(*xi*) = *ν*(*μ*, *α*) *C* ∑ *j* We have this already

Estimate from simulations and existing interpolation methods

$L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$

$$
\prod_{i}^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)
$$

$L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$

 $L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$

Prod over events and uncertainty associated with the source of systematic uncertainty associated with the number of systematic uncertainty associated with the source of systematic uncertainty associated with the number of

Rate term $L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$ \mathcal{A}_1 where the global observables \mathcal{A}_2 : and the auxiliary measurements and their associated and their associated and the auxiliary measurements and the auxiliary measurements and their associated and the auxilia $Proof over events$

Prod over events

Prod over events and uncertainty associated with the source of systematic uncertainty associated with the number of systematic uncertainty associated with the source of systematic uncertainty associated with the number of $L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$ \mathcal{A}_1 where the global observables of the values of the values of the auxiliary measurements and their associated \mathcal{A}_2

 $\frac{1}{\text{Im}(\mathcal{D})}$ = Pois($N_{\text{data}}|v(\mu,\alpha)$ $L_{\textrm{full}}(\mu, \alpha | \mathcal{D})$ $L_{\mathrm{ref}}(\mathcal{D})$ $= \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha))$ \mathcal{A}_1 where the global observables of the values of the values of the auxiliary measurements and their associated \mathcal{A}_2 $Proof over events$

Prod over events er events

Profiling:

$$
t_{\mu} = -2 \ln \left(\frac{L_{\text{full}}(\mu, \hat{\alpha}) / L_{\text{ref}}}{L_{\text{full}}(\hat{\mu}, \hat{\alpha}) / L_{\text{ref}}} \right)
$$

here is an example for a reference from the bibles file (see file (see file (see file (see file (see file (see

reference figures (see Figure 1). (` b b U) = argmax *,*

This is why we define
$$
p_{ref}
$$
 to be independent of μ

!ref

 422 Note that since r is does not affect the position of $\mathcal{L}_\mathcal{A}$ and $\mathcal{L}_\mathcal{A}$ affect the position of $\mathcal{L}_\mathcal{A}$

Negative Likelihood Ratio result \mathcal{L} likelihood is shown in Figure 8 and compared to a histogram analysis using the Ofixed observable. The Ofixed observable \mathcal{L} The system is syntaxid uncertainties result we see the measurement of the measurement, and measurement, and the measu

Negative Likelihood Ratio result \mathcal{L} likelihood is shown in Figure 8 and compared to a histogram analysis using the Ofixed observable. The Ofixed observable \mathcal{L} The system is syntaxid uncertainties result we see the measurement of the measurement, and measurement, and the measu

sono oriapo dao to hon mioar onooto rioni quantum intorioronoo Non-parabolic shape due to non-linear effects from quantum interference

Open problems to extend to full ATLAS analysis:

- ✓ Robustness: Design and validation
- ✓ Systematic Uncertainties: Incorporate them into likelihood (ratio) model
-

‣ Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Sampling (per-event) pseudo-experiments using bootstrapping

Traditionally:

NSBI:

$$
w_i^{toy} = Poisson(w_i^{Asimov})
$$

('Unweighted' events, i.e. integer weights)

- For each hypothesis:
	- Generate pseudo-experiments using bootstrapping
	- Compute the test statistic at the value of the considered hypothesis
	-

Neyman Construction

• Integrate up to 68.27% (95.45%) to determine $1\sigma(2\sigma)$ CI as a function of the parameter of interest

Confidence belts

Why does NSBI work better than traditional analyses?

Why does it work better than traditional analyses?

Big picture of the implementation of NSBI for Parameter Estimation in ATLAS

O(16) observables

Big picture of the implementation of NSBI for Parameter Estimation in ATLAS

Big picture of the implementation of NSBI for Parameter Estimation in ATLAS

- \triangleleft Train $O(10^3)$ networks on TensorFlow
- ✦ Computing resources provided by Google, SMU, other HPC clusters
- ◆ Fits with JAX

- Developed a complete statistical framework for high-dimensional statistical inference
	- Builds upon traditional methodology in ATLAS
	- Developed diagnostic tools for validation
- Such methods are crucial for analyses where kinematic distributions change non-linearly with the parameter of interest, eg. EFT studies
- Weaknesses: Same as traditional analyses, requires well trained networks

- Developed a complete statistical framework for high-dimensional statistical inference
	- Builds upon traditional methodology in ATLAS
	- Developed diagnostic tools for validation
- Such methods are crucial for analyses where kinematic distributions change non-linearly with the parameter of interest, eg. EFT studies
- Weaknesses: Same as traditional analyses, requires well trained networks

Thanks!

Backup

Building a 'Search-Oriented Mixture Model' ¹⁶⁰ of the task into individual physics processes can always be made, if the dependence on the parameter of Building a Search-Oriented Mixture Model

 $\sum_{l=1}^{n}$?ref(G8) $\frac{\nu(\mu)}{1}$ a(\) ⇠ 5 ⁹(\) · a ⁹ · $\sum_{l=1}^{n}$?ref(G8) Event rates $p(x_i|\mu) =$ 1 $\mathcal{Y}(\mu)$ \sum \overline{C} \boldsymbol{j} $f_j(\mu) \cdot \nu_j \; p_j(x_i)$, in a over different phy where ? ⁹(G8) is the probability density for the event G⁸ correspeonding to the process 9, and a ⁹ ¹⁷⁶ the inclusive 1777 rates are defined with process are defined with \sim can be full dependence on \sim can be function on \sim can be func

 runs over different physics process *j* $(Eg. gg \rightarrow H^* \rightarrow 4l, gg \rightarrow ZZ \rightarrow 4l)$

¹⁶² This paper defines a *search-oriented* mixture model, which is the probability density ratio between a

175 º DIFFERENT COMPONENTS, REPRESENTING DIFferent physics processes, representing a component physics processes, and the physics processes, and the physics processes, and the physics processes, and the physics processes,

Event rates
Comes from theory model chosen to interpret data ¹⁶⁵ freedom to make any choice for the reference, this paper defines it as a combination of signal processes, corres from cheory momentum coefficients free

> Define a 'reference' density with support over entire region of analysis Does not have to be physical !

$$
p_{\text{ref}}(x_i) = \frac{1}{\sum_{k} v_k} \sum_{k}^{C_{\text{signals}}} v_k \cdot p_k(x_i)
$$

Define a 'reference' density with support over entire re
Does not have to be physical !

 x_i is one individual event

 $\mathscr{L}(\mu | \mathscr{D}) = p(\mathscr{D} | \mu)$

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

 $\mathscr{L}(\mu | \mathscr{D}) = p(\mathscr{D} | \mu)$

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

 $\mathscr{L}(\mu | \mathscr{D}) = p(\mathscr{D} | \mu)$

 $s(x_i) =$ $p(x_i|S)$ $p(x_i|S) + p(x_i|B)$

 $p(\mathcal{D}|\mu)$ Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

 $p(\mathcal{D} | \mu_0)$

A neural network classifier trained on S vs B, estimates the decision function*:

* Equal class weights

$$
\mathscr{L}(\mu\,|\,\mathscr{D})=p(\mathscr{D})
$$

We want to compare likelihoods: $p(\mathscr{D}|\mu)$ $p(\mathcal{D} | \mu_0)$ Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

 $s(x_i) =$ $p(x_i|S)$ $p(x_i|S) + p(x_i|B)$ A neural network classifier trained on S vs B, estimates the decision function*:

$$
\frac{p(x_i|\mu)}{p(x_i|\mu=0)} = \frac{\mu \cdot \sigma_S \cdot p(x_i|S) + \sigma_B \cdot p(x_i|B)}{\sigma_B \cdot p(x_i|B)} = \mu \cdot \frac{\sigma_S}{\sigma_B} \cdot \frac{s(x_i)}{1 - s(x_i)} + 1.
$$

Same observable *s* is optimal to test all μ hypotheses! * Equal class weights **and the contract of the club of the club of the club of the club of the M** μ

Which contains all the information required for the likelihood ratio:

-
-

$$
\frac{p(x_i | \mu_1)}{p(x_i | ref)} = \frac{s(x_i)}{1 - s(x_i)}
$$

 $p(\mathscr{D} \,|\, \mu)$ *p*(|*ref*)

A neural network classifier trained on simulated samples from θ_1 vs simulated samples from *ref*, estimates the decision function:

Estimating high-dimensional density ratios

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

✴ Optimal statistic to test each value of *μ* ✴ We get the LR *per event (*unbinned)

 $\mathscr{L}(\mu | \mathscr{D}) = p(\mathscr{D} | \mu)$

 $s(x_i) =$ $p(x_i | \mu_1)$ $p(x_i|\mu_1) + p(x_i|ref)$

Which contains all the information required for the likelihood ratio:

 \overline{p}

reweighted reference sample and target sample and the target sample. The target sample and the target sample, w

Calibration Curves

PSBI $P_{SBI} + P_{ref}$

PB $P_B + P_{ref}$

$$
-2 \cdot \log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}
$$

$$
-2 \cdot \log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)} -2 \cdot \log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}
$$

${\sf Interpreability:}$ Which phase space favours one hypothesis over another?

log ?B/?(G|` = 1), as a discriminant, with 15 bins. The markers show the sensitivity for various histogram analyses

Negative Weighted Events

- 1. Start from a positive weighted reference sample instead
- 2. Re-weight to intended parameter point
- 3. Throw toys from this sample

$$
w_i^{\text{rwt-ref}} \rightarrow w_i^{\text{Asimov}}(\mu) =
$$

$$
\frac{\nu(\mu)}{\nu_{\text{rwt-ref}}} \cdot \frac{p(x_i | \mu)}{P_{\text{rwt-ref}}(x_i)} \cdot w_i^{\text{rwt-ref}}
$$

Image: [Source](https://www.lancaster.ac.uk/stor-i-student-sites/jack-trainer/bootstrapping-in-statistics/)

the mean

Estimating the variance on mean: Bootstrapping

Quantifying uncertainty on estimated density ratio

$$
w_i \rightarrow w_i \cdot Pois(1)
$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles $\overbrace{\hspace{2.8cm}}^{\text{mean}}$

Quantifying uncertainty on estimated density ratio

$$
w_i \rightarrow w_i \cdot Pois(1)
$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles $\overbrace{\hspace{2.8cm}}^{\text{mean}}$

$$
w_i \rightarrow w_i \cdot Pois(1)
$$

- e as
0.0
Train an ensemble of networks, each on a Poisson fluct $\frac{q}{\ddot{a}}^{2.5}$ the training dataset
- Ensemble average used as final prediction, estimate the variance on $\begin{array}{c} \bigcap_{i=1}^{\infty}$

$$
w_i \rightarrow w_i \cdot Pois(1)
$$

- es and the main an ensemble of networks, each on a Poisson fluct $\frac{25}{9}$ e.o.
Train an ensemble of networks, each on a Poisson fluct $\frac{25}{9}$ -2.5 the training dataset
- Ensemble average used as final prediction, estimate the variance on $\begin{array}{c} \bigcap_{i=1}^{\infty}$
- Propagate with spurious signal method

f j $(\mu) \rightarrow f$ *j* $(\mu + \alpha \cdot \Delta \hat{\mu}(\mu))$

Constraint term: *Gauss*(0,1)

Simulated Samples and the simulated as inputs to construct a matrix of the simulated samples \mathbf{a} decay kinematic variables are defined as the seven kinematic observables cos \mathbf{a} $\overline{1, 1, 2}$

- PoI: Signal strength *μ*
- Simplified, unphysical dataset:
	- Processes: S: $gg \rightarrow H^* \rightarrow 4l$ & B: $gg \to ZZ \to 4l$, SBI: full process
	- No VBF processes or qqZZ background
	- Two systematics: ggF NLO K-factor uncertainty (shape + norm) & luminosity uncertainty (norm only)

Input variables

Combination with histogram analyses

 $L_{\text{comb}}(\mu, \alpha)$ Lref = $L_{\mathrm{full}}(\mu, \alpha)$ Lref $L_{\text{hist}}(\mu, \alpha)$

⁴²⁹ When likelihood ratios are estimated with neural networks, an uncertainty may be introduced to account

Calculating pulls and impacts in JAX

U)

beautif

Hessian:

⇠=< =

1
1910 - 1910
1910 - 1910 - 1910

,

m
2. March 1990
2. March 1990

$$
\Gamma_k = \frac{\partial \widehat{\mu}}{\partial \alpha_k} \times \sqrt{C_{kk}}
$$

= $-\left[\frac{\partial^2 \lambda}{\partial^2 \mu}(\widehat{\mu}, \widehat{\alpha})\right]^{-1} \frac{\partial^2 \lambda}{\partial \mu \partial \alpha_k}(\widehat{\mu}, \widehat{\alpha}) \times \sqrt{C_{kk}},$

Hessian:
\n
$$
C_{nm} = \left[\frac{1}{2} \frac{\partial^2 \lambda}{\partial \alpha_n \partial \alpha_m}(\hat{\mu}, \hat{\alpha})\right]^{-1} \qquad \lambda(\mu, \alpha) = -2 \ln(L_{\text{full}}(\mu, \alpha) / L_{\text{ref}})
$$

(*`*

^U: ^U⁰

:

Pulls:

Post-fit Impact:

$$
\lambda(\mu, \alpha) = -2 \ln(L_{full}(\mu, \alpha) / L)
$$

, (25)

Vertical interpolation

$$
G_j(\alpha_k) = \begin{cases} \left(\frac{v_j(\alpha_k^+)}{v_j(\alpha_k^0)}\right)^{\alpha_k} & \alpha_k > 1\\ 1 + \sum_{n=1}^6 c_n \alpha_k^n & -1 \le \alpha_k \le 1\\ \left(\frac{v_j(\alpha_k^-)}{v_j(\alpha_k^0)}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}
$$

With some continuity requirements
 interpolation strategy and continuity requirements can be used to interpolate 69(G⁸ ⁶⁹⁰ *,* U:), **1 Analy requirements**

$$
g_j(x_i, \alpha_k) = \begin{cases} \left(g_j(x_i, \alpha_k^+)\right)^{\alpha_k} & \alpha_k > 1\\ 1 + \sum_{n=1}^6 c_n \alpha_k^n & -1 \le \alpha_k\\ \left(g_j(x_i, \alpha_k^-\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}
$$

