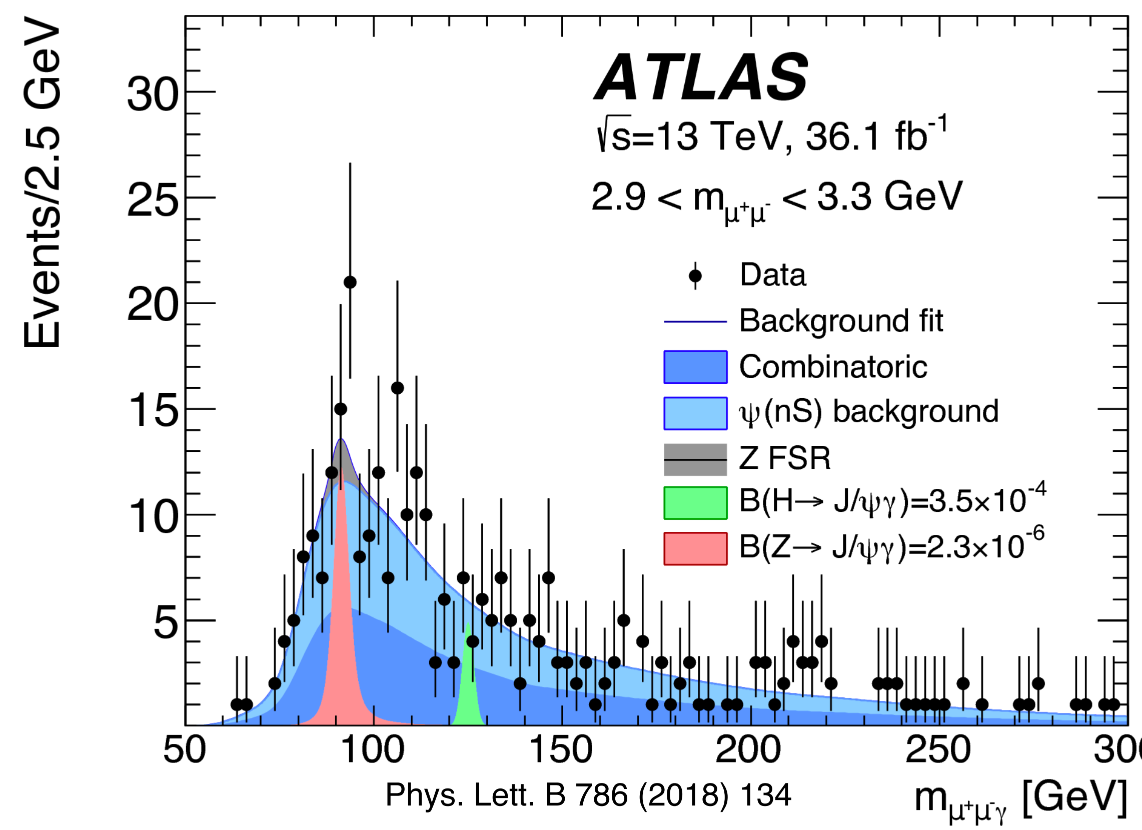
Non-Parametric Data-Driven Background Modelling using Conditional Probabilities

Konstantinos Nikolopoulos

University of Birmingham

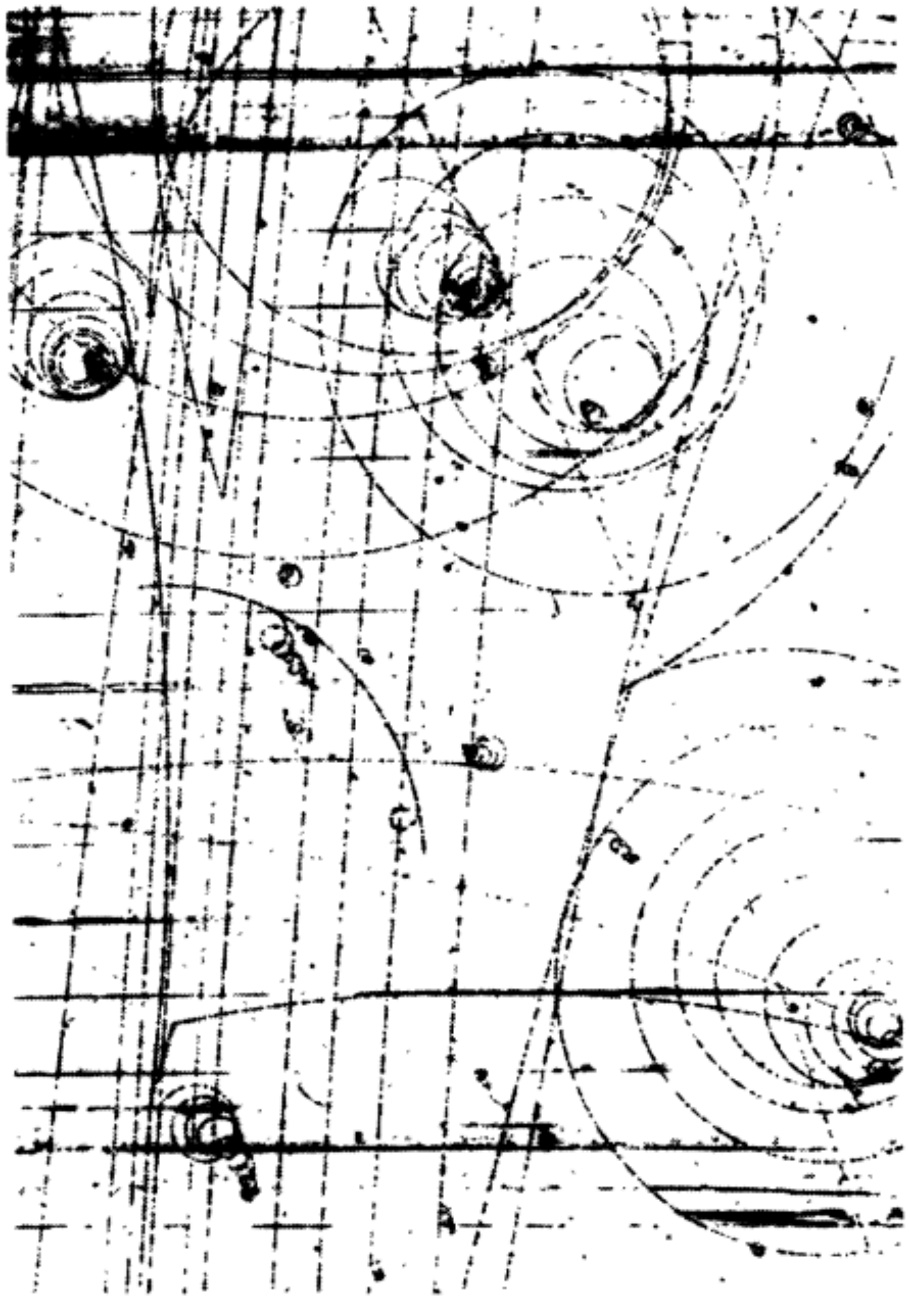


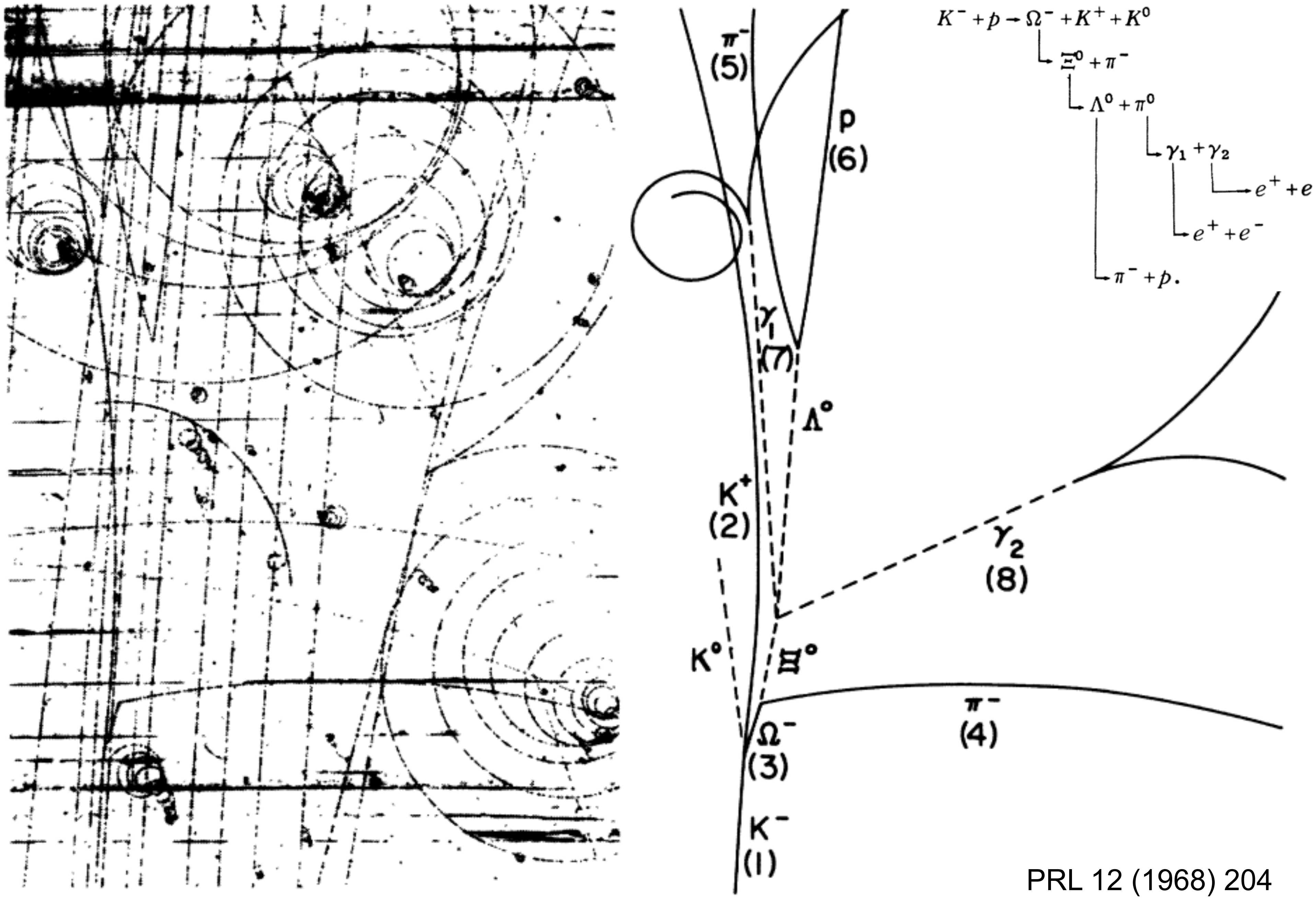
UNIVERSITY OF BIRMINGHAM

University of Birmingham, Particle Physics Seminar
January 26, 2022, Birmingham, UK

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement 714893 (ExclusiveHiggs) and under Marie Skłodowska-Curie agreement 844062 (LightBosons)

European Research Council
Established by the European Commission





PRL 12 (1968) 204

Discoveries of new signals

…are all about controlling the backgrounds

Discoveries of new signals

…are all about controlling the backgrounds

Construct the **profile likelihood ratio** test statistic: $\lambda\left(\mu\right) = \dfrac{L\left(\mu, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$

Discoveries of new signals

…are all about controlling the backgrounds

Construct the **profile likelihood ratio** test statistic: $\lambda\left(\mu\right) = \dfrac{L\left(\mu, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$

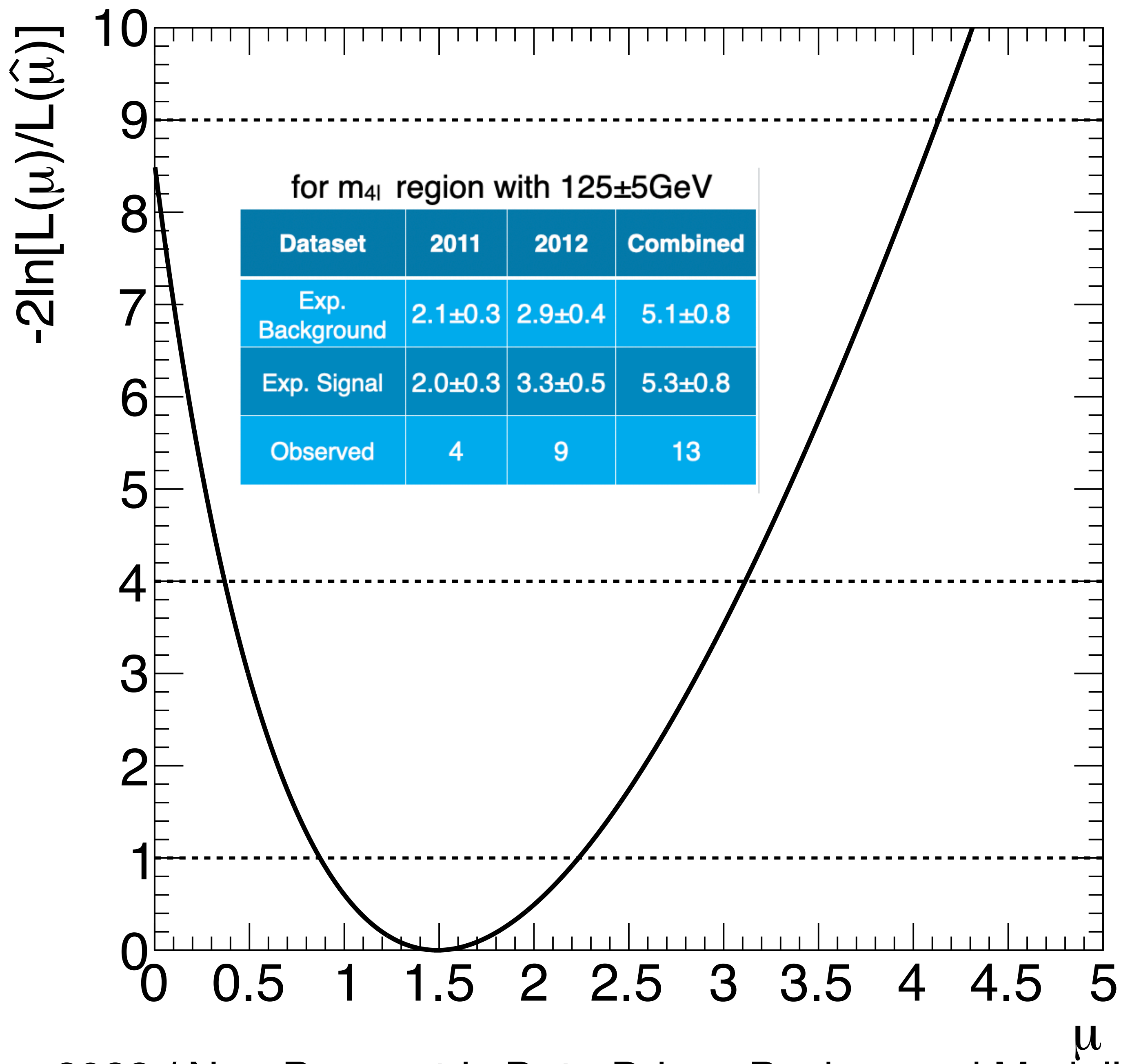
and test the **background-only** hypothesis ($\mu = 0$): $\lambda\left(0\right) = \dfrac{L\left(0, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$

Discoveries of new signals

…are all about controlling the backgrounds

Construct the **profile likelihood ratio** test statistic: $\lambda\left(\mu\right) = \dfrac{L\left(\mu, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$

and test the **background-only** hypothesis ($\mu = 0$): $\lambda\left(0\right) = \dfrac{L\left(0, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$



for m_{4l} region with 125±5GeV

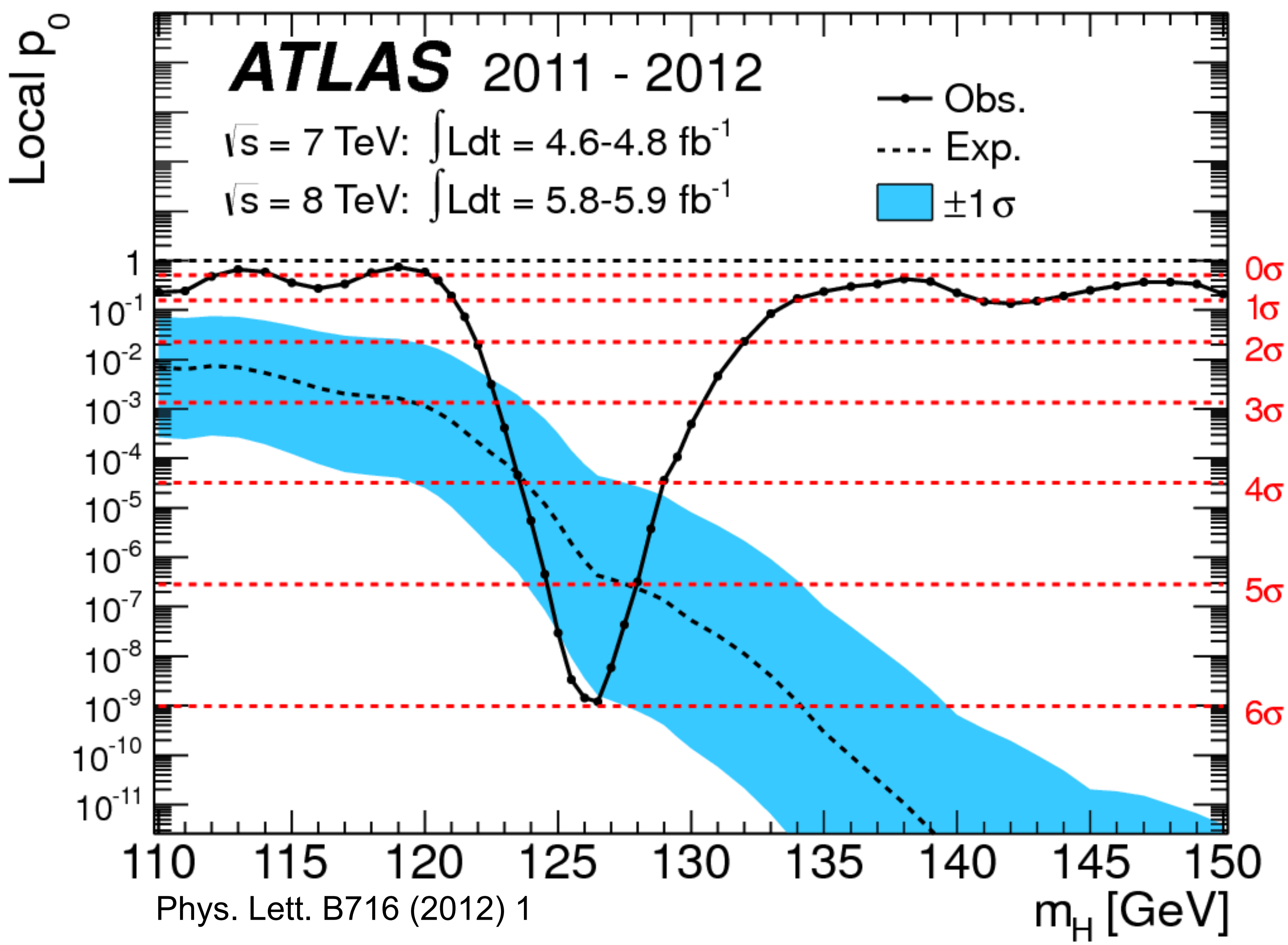
Dataset	2011	2012	Combined
Exp. Background	2.1±0.3	2.9±0.4	5.1±0.8
Exp. Signal	2.0±0.3	3.3±0.5	5.3±0.8
Observed	4	9	13

Discoveries of new signals

…are all about controlling the backgrounds

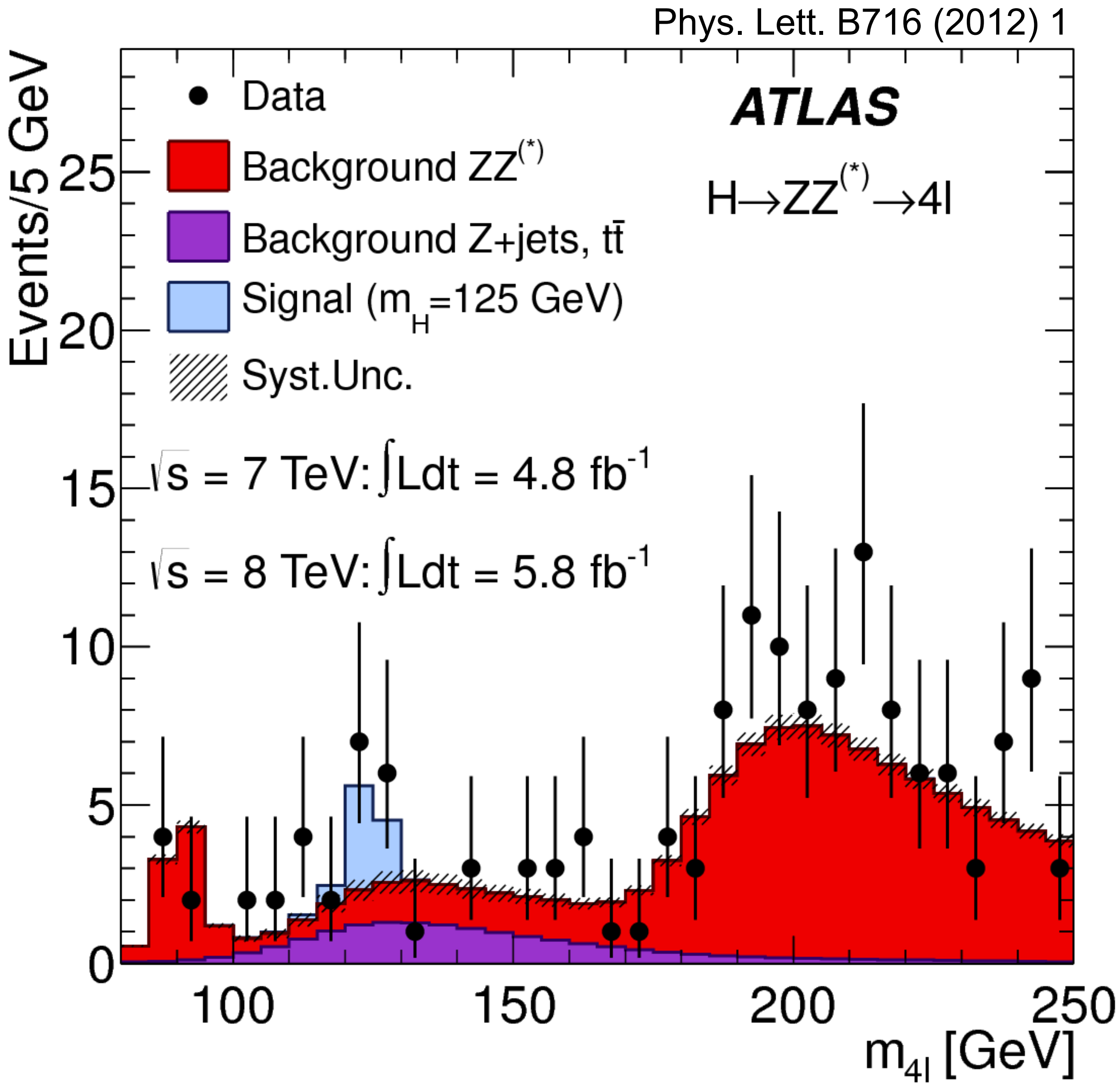
Construct the **profile likelihood ratio** test statistic: $\lambda(\mu) = \dfrac{L\left(\mu, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$

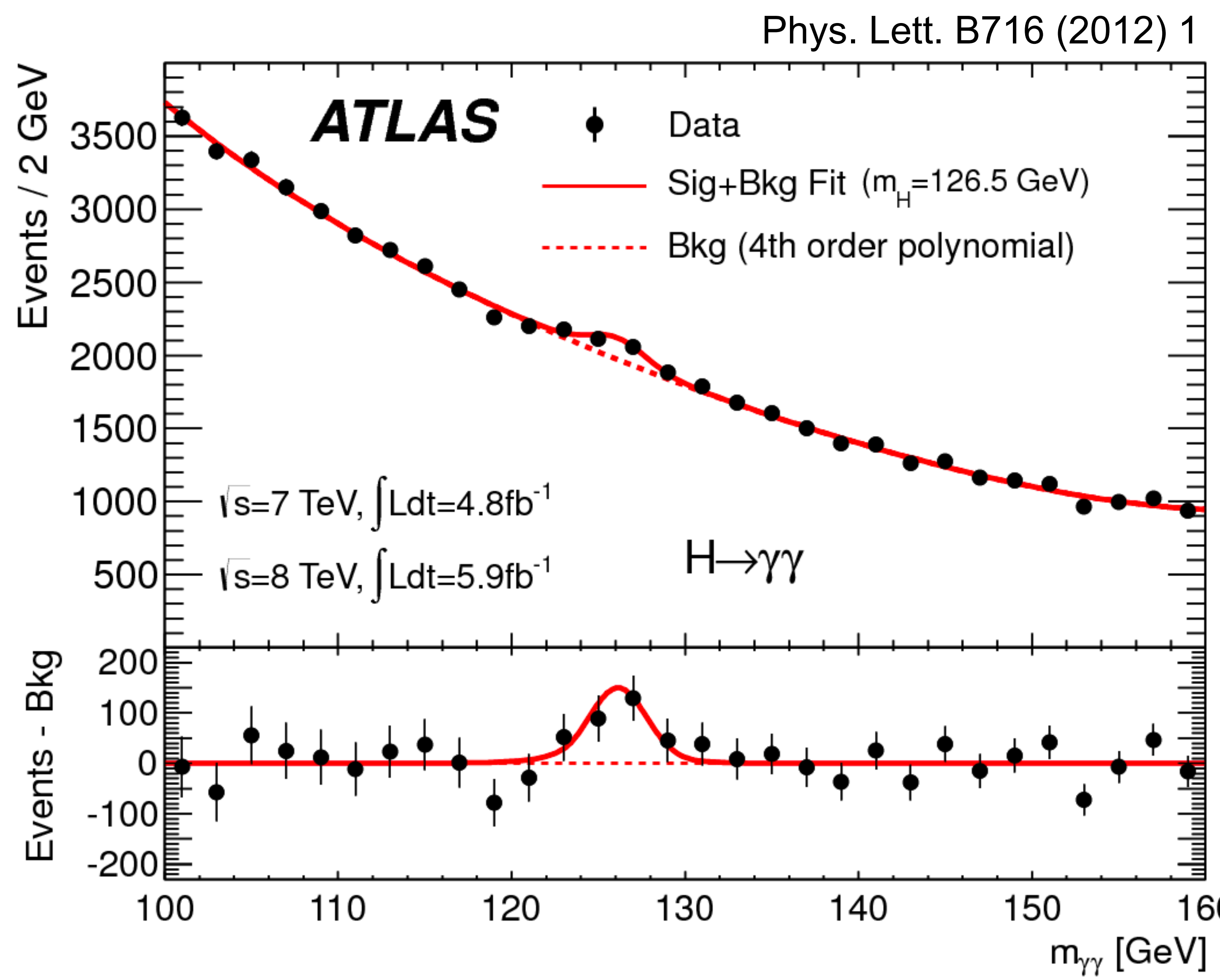
and test the **background-only** hypothesis ($\mu = 0$): $\lambda(0) = \dfrac{L\left(0, \hat{\hat{\theta}}\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}$



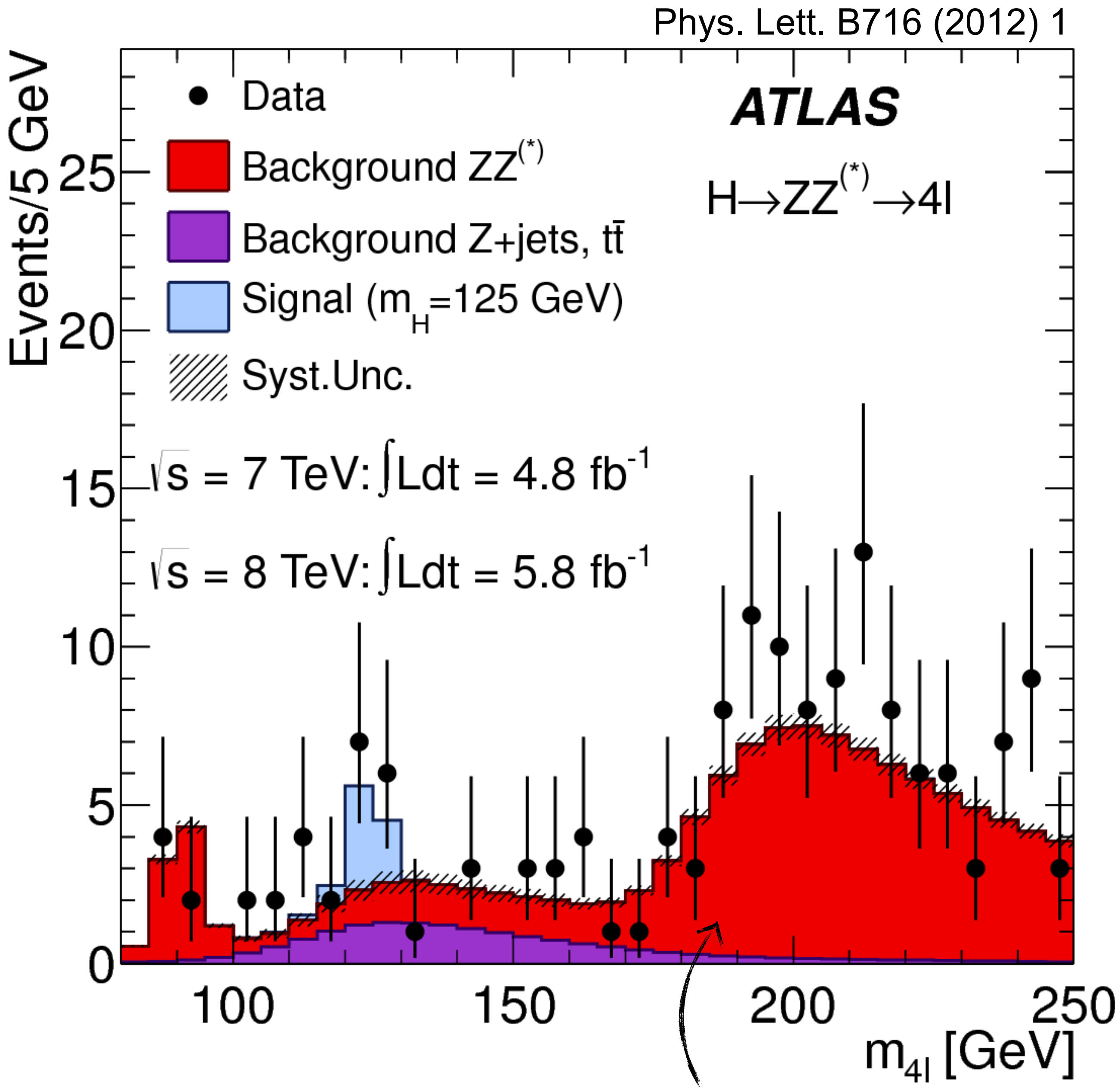
Phys. Lett. B716 (2012) 1

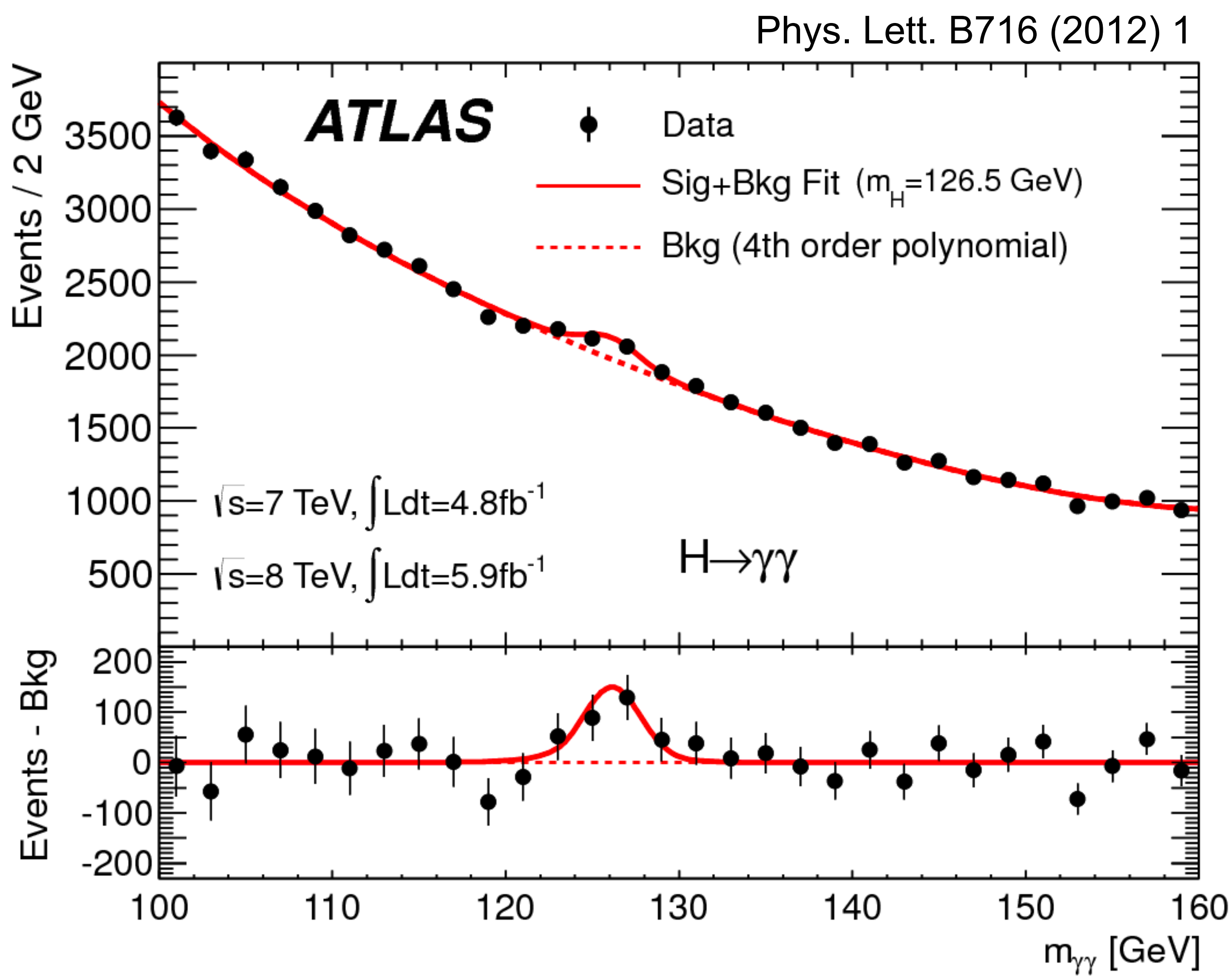
Observing the Higgs boson

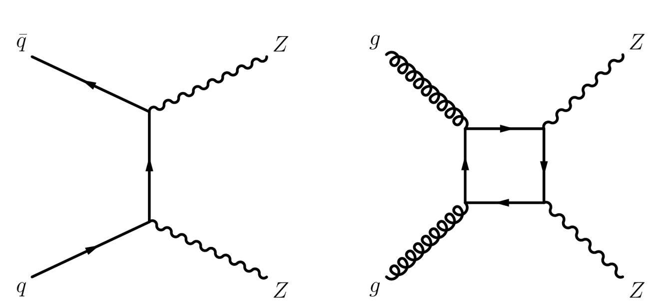




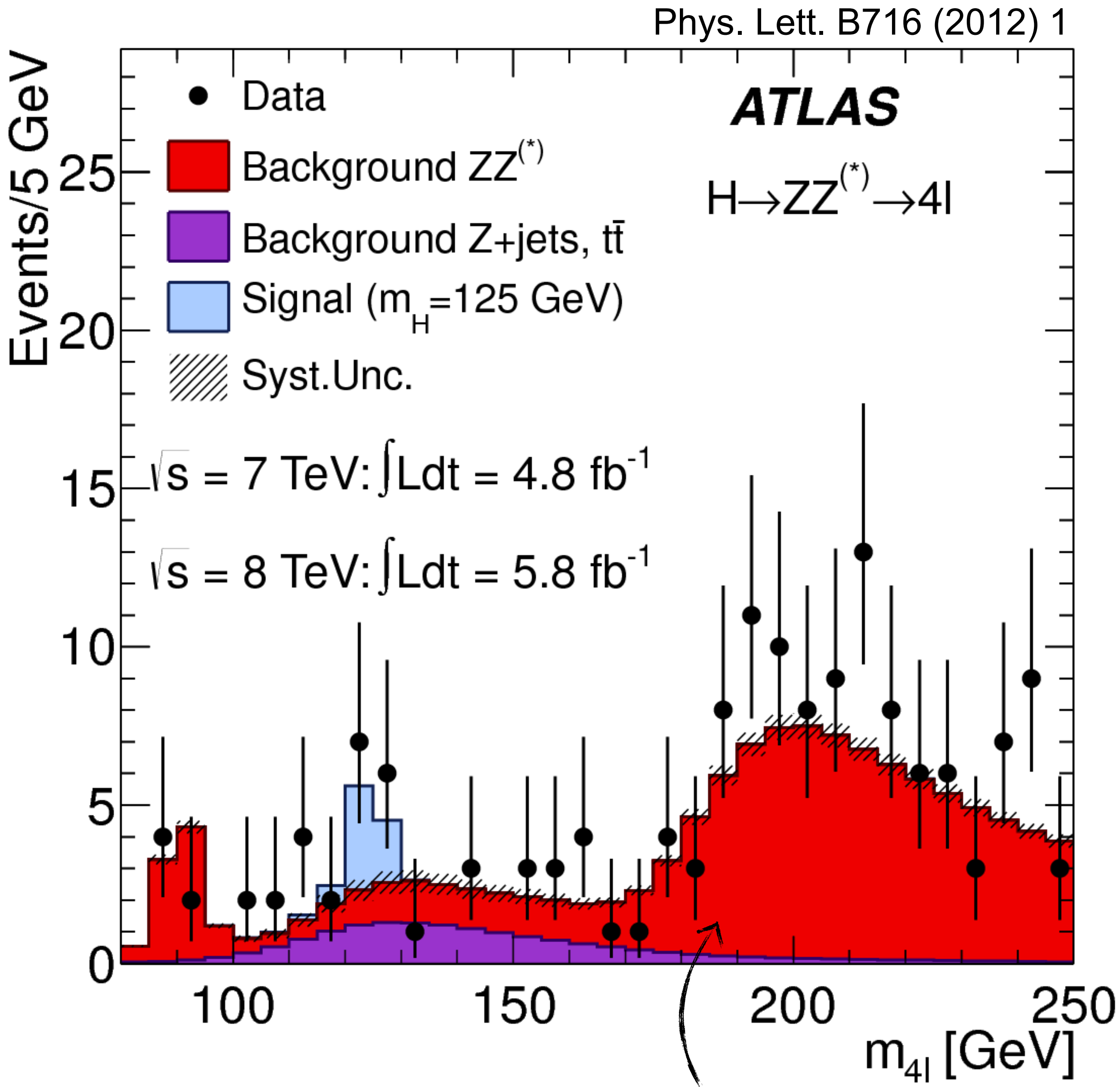
Observing the Higgs boson

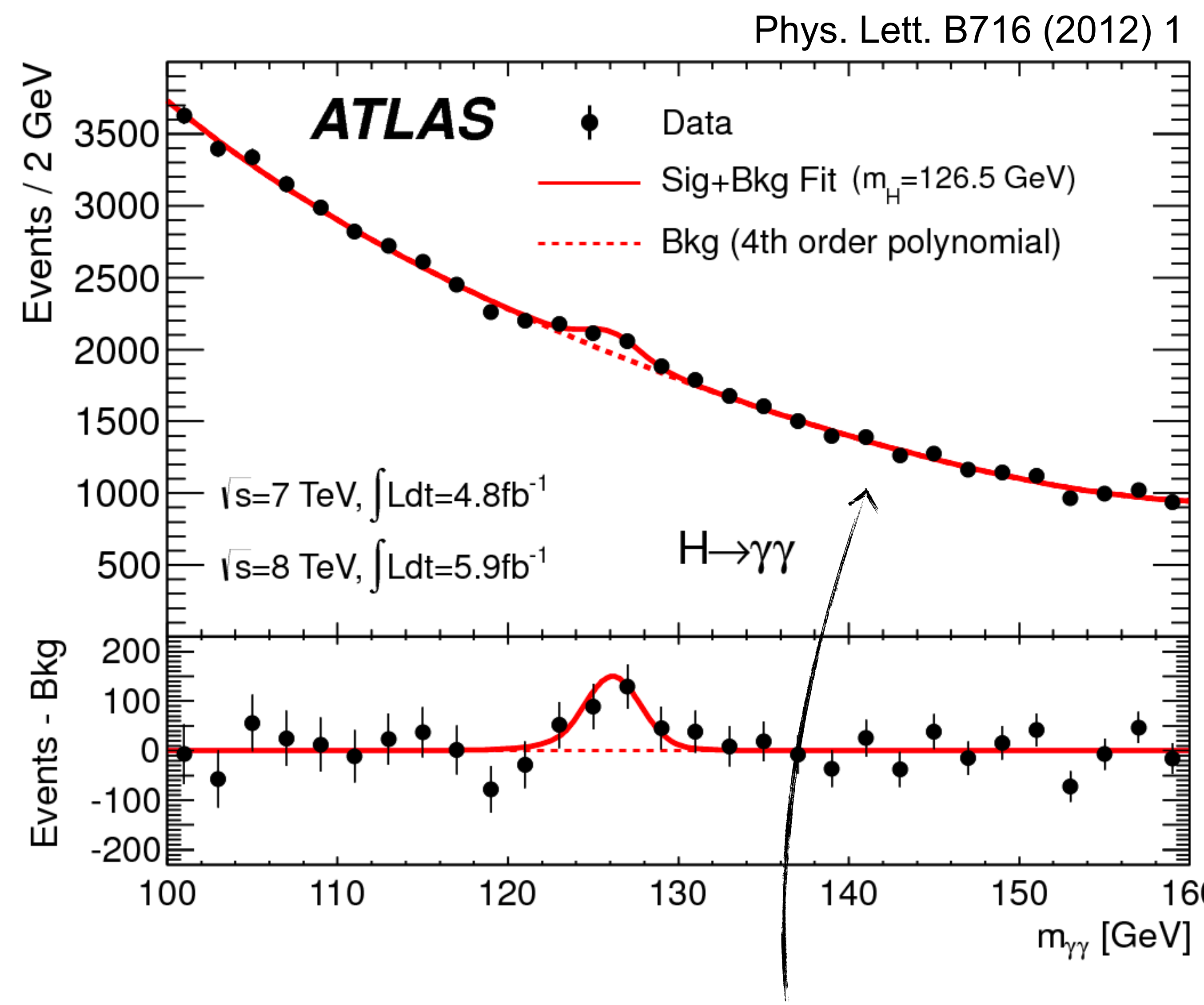


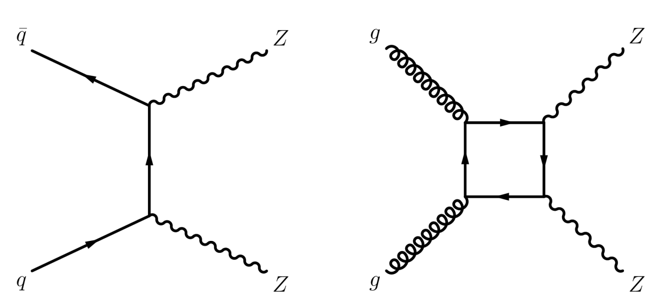




Observing the Higgs boson

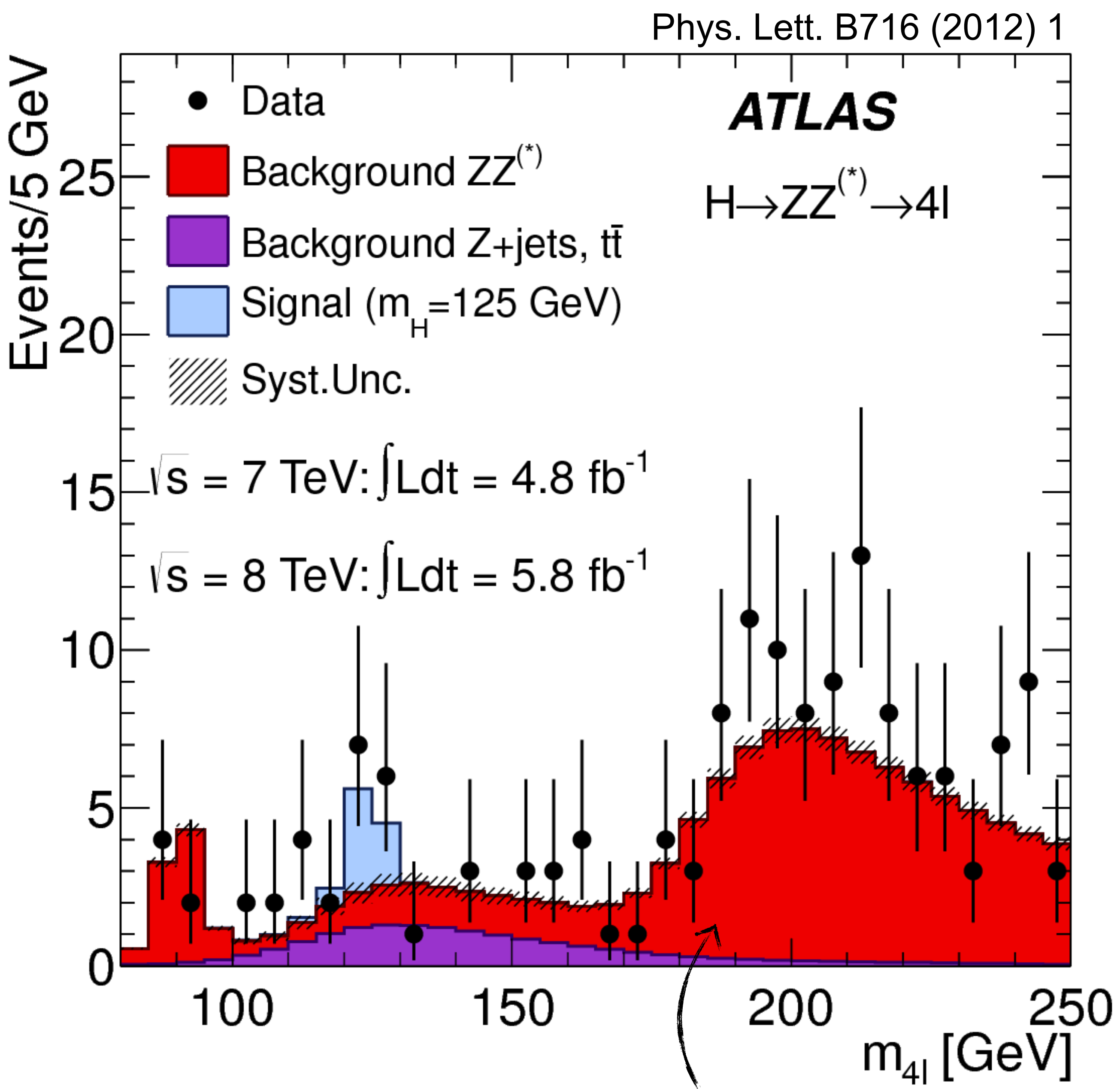


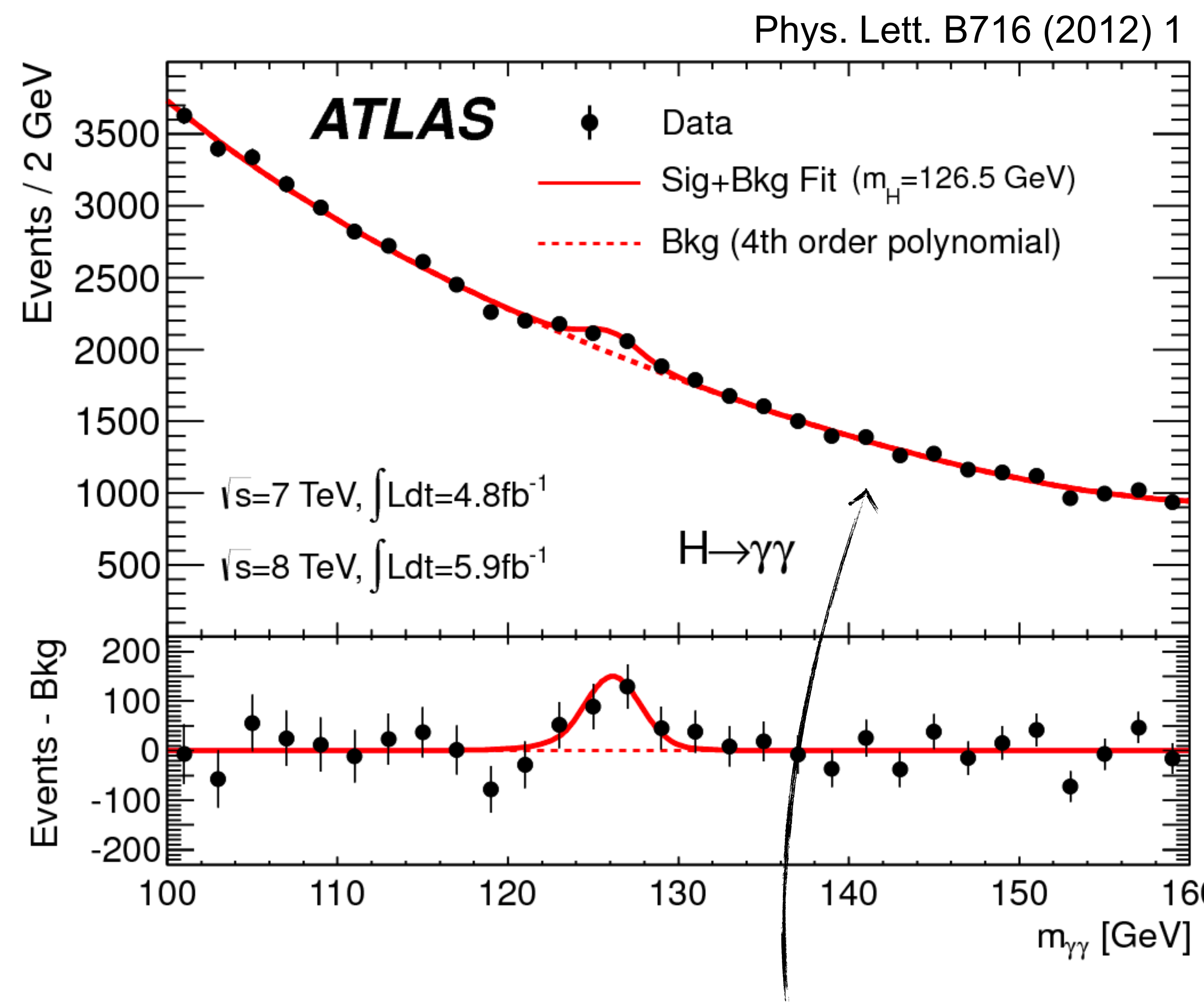


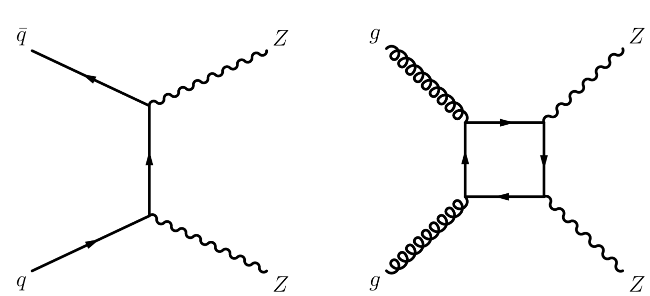


di-photon, photon+jet, jet+jet

Observing the Higgs boson







di-photon, photon+jet, jet+jet

Trying to maximise data-driven input, e.g. signal side-bands

Parametric methods

Both ATLAS and CMS $H \to \gamma\gamma$ use parametric methods

- Also $H \to \mu\mu$, $H \to Z\gamma$, $H \to bb\gamma\gamma$, etc

Parametric methods

Both ATLAS and CMS H→γγ use parametric methods

- Also H→μμ, H→Zγ, H→bbγγ, etc

Choose a function with N_{par} free parameters

- Too **many** parameters: Reduced statistical power
- Too **few** parameters: Not enough flexibility to model the background

Parametric methods

Both ATLAS and CMS H→γγ use parametric methods

- Also H→μμ, H→Zγ, H→bbγγ, etc

Choose a function with N_{par} free parameters

- Too **many** parameters: Reduced statistical power
- Too **few** parameters: Not enough flexibility to model the background

Question: Does the true, but unknown, background shape belong to the family of curves parametrised by the chosen function?

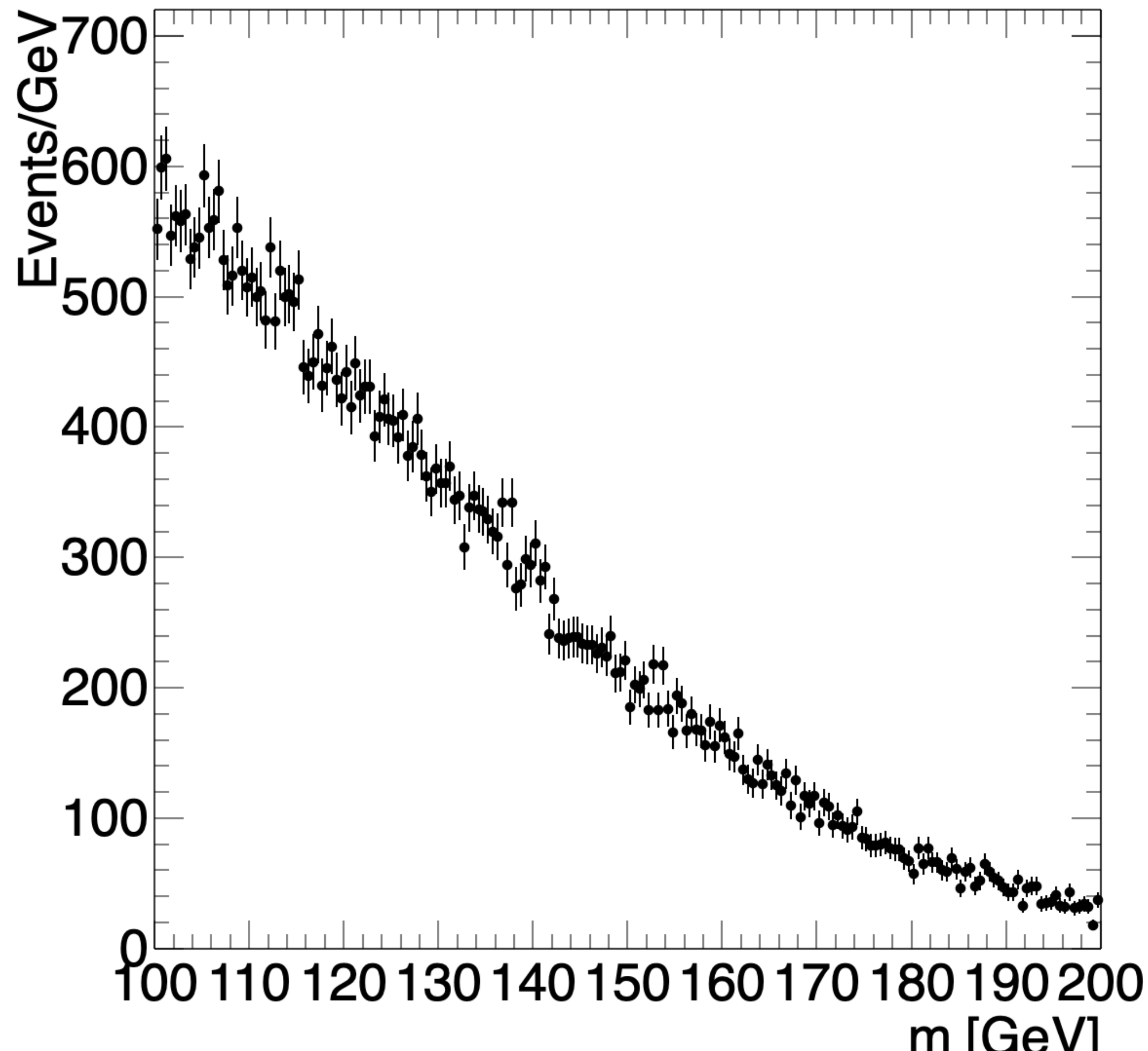
Parametric methods

Both ATLAS and CMS H→γγ use parametric methods

- Also H→μμ, H→Zγ, H→bbγγ, etc

Choose a function with N_{par} free parameters

- Too **many** parameters: Reduced statistical power
- Too **few** parameters: Not enough flexibility to model the background



Question: Does the true, but unknown, background shape belong to the family of curves parametrised by the chosen function?

Parametric methods

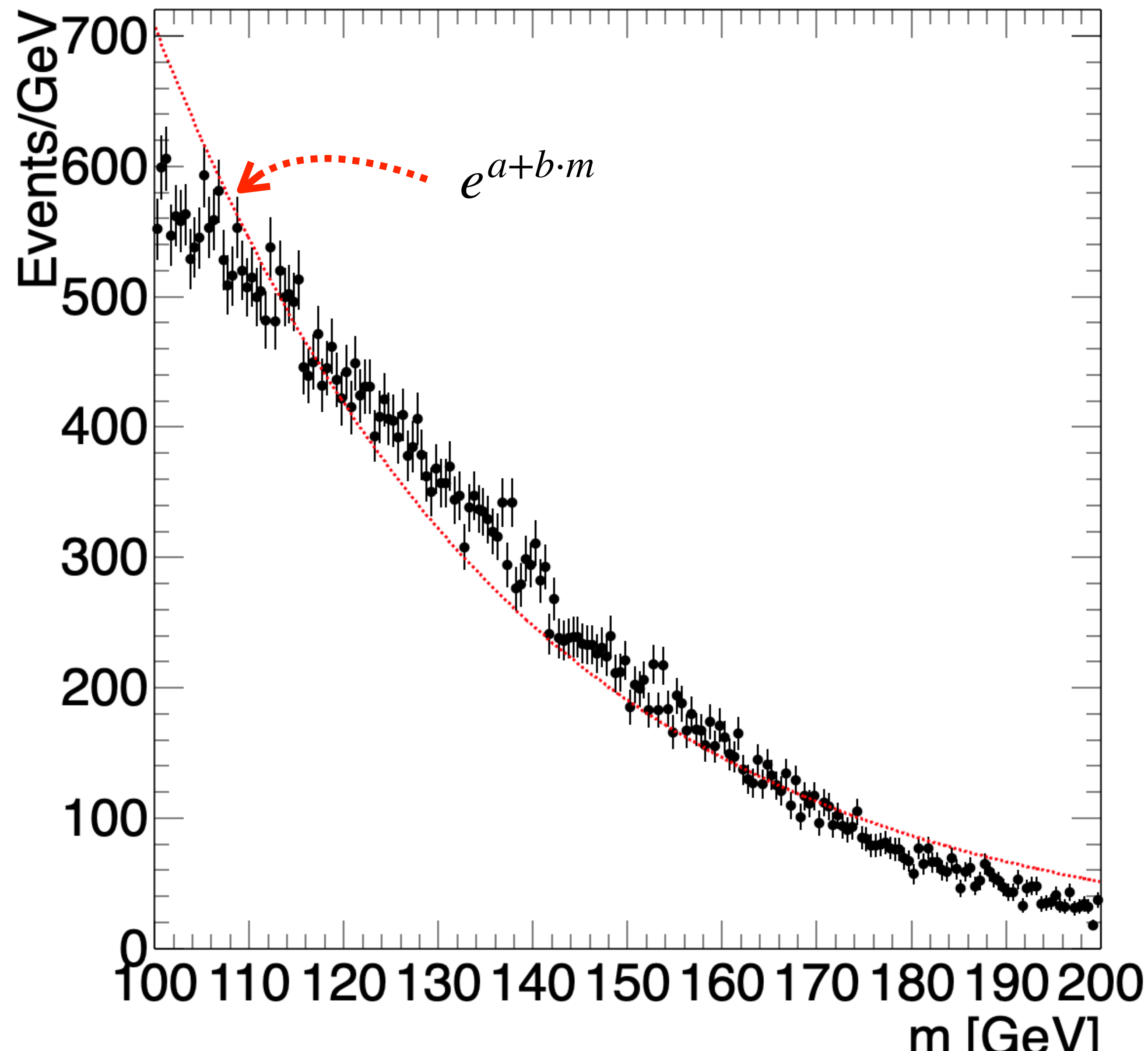
Both ATLAS and CMS H→γγ use parametric methods

- Also H→μμ, H→Zγ, H→bbγγ, etc

Choose a function with N_{par} free parameters

- Too **many** parameters: Reduced statistical power
- Too **few** parameters: Not enough flexibility to model the background

Question: Does the true, but unknown, background shape belong to the family of curves parametrised by the chosen function?



Parametric methods

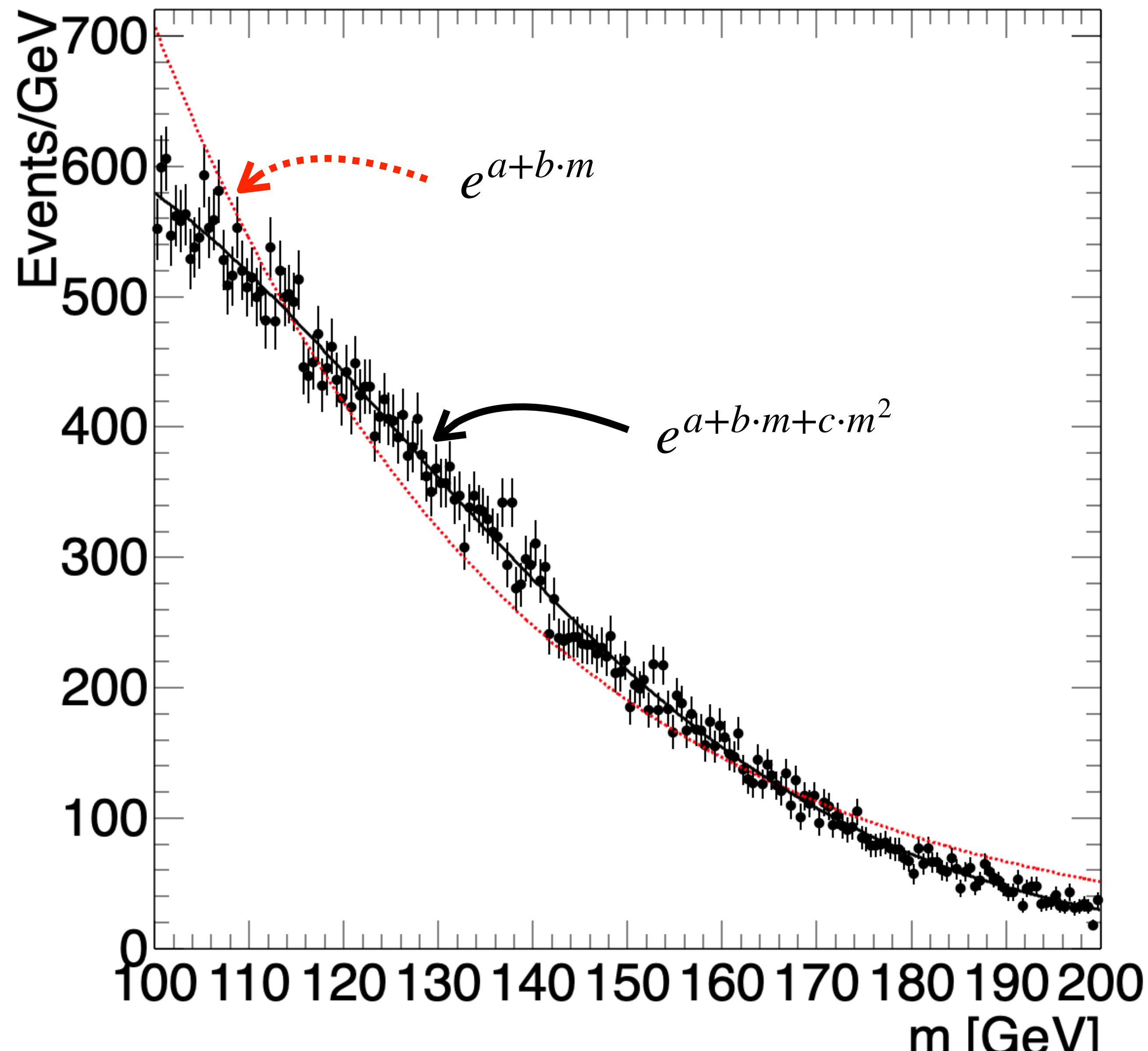
Both ATLAS and CMS H→γγ use parametric methods

- Also H→μμ, H→Zγ, H→bbγγ, etc

Choose a function with N_{par} free parameters

- Too **many** parameters: Reduced statistical power
- Too **few** parameters: Not enough flexibility to model the background

Question: Does the true, but unknown, background shape belong to the family of curves parametrised by the chosen function?



Spurious signal

ATLAS uses the concept of "**spurious signal**"

- Possible systematic mismodelling due to function choice leading to apparent signal

Spurious signal

ATLAS uses the concept of "**spurious signal**"

- Possible systematic mismodelling due to function choice leading to apparent signal

Choices:

- What function to use?
- What systematic uncertainty to assign?

Spurious signal

ATLAS uses the concept of "**spurious signal**"

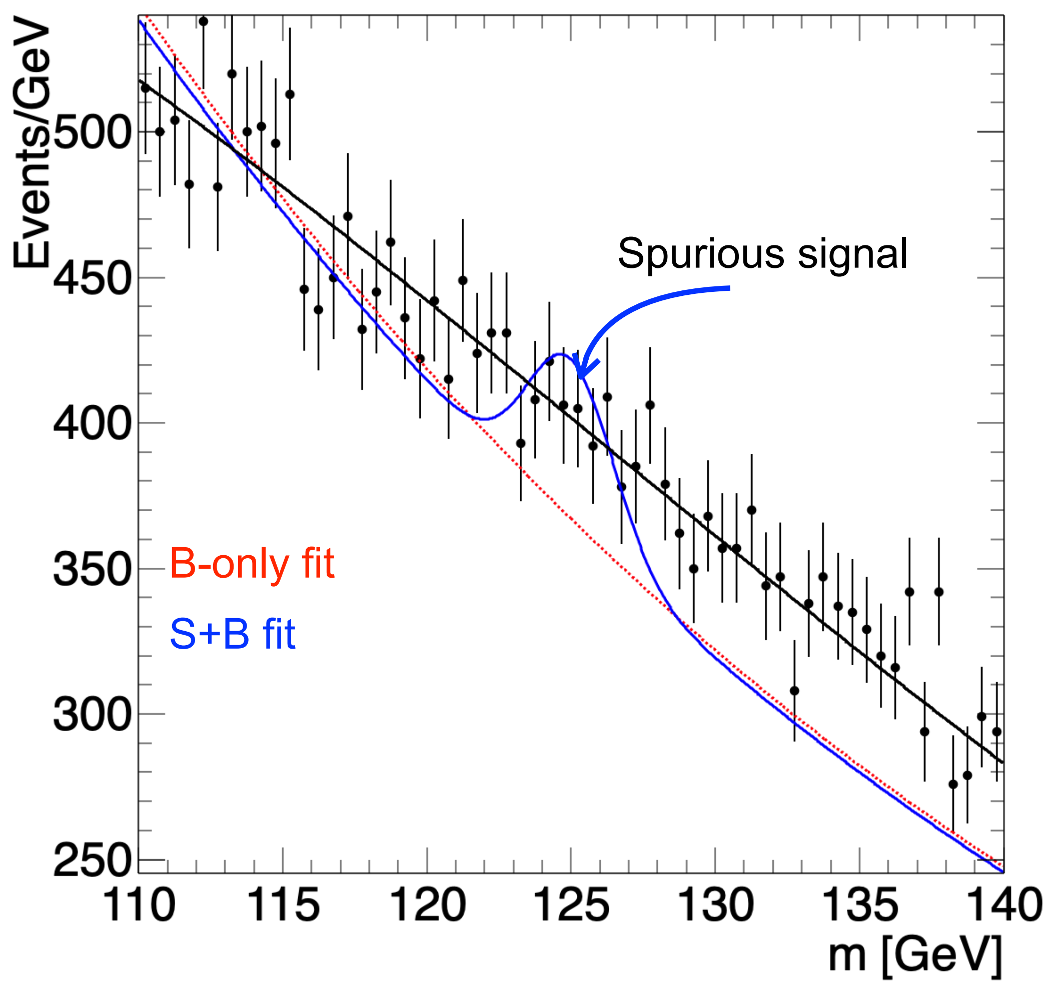
- Possible systematic mismodelling due to function choice leading to apparent signal

Choices:

- What function to use?
- What systematic uncertainty to assign?

Use MC sample of background to perform S+B fits.

- Use function with lowest obtained $S_{spurious}$, and said $S_{spurious}$ as systematic



Spurious signal

ATLAS uses the concept of "**spurious signal**"

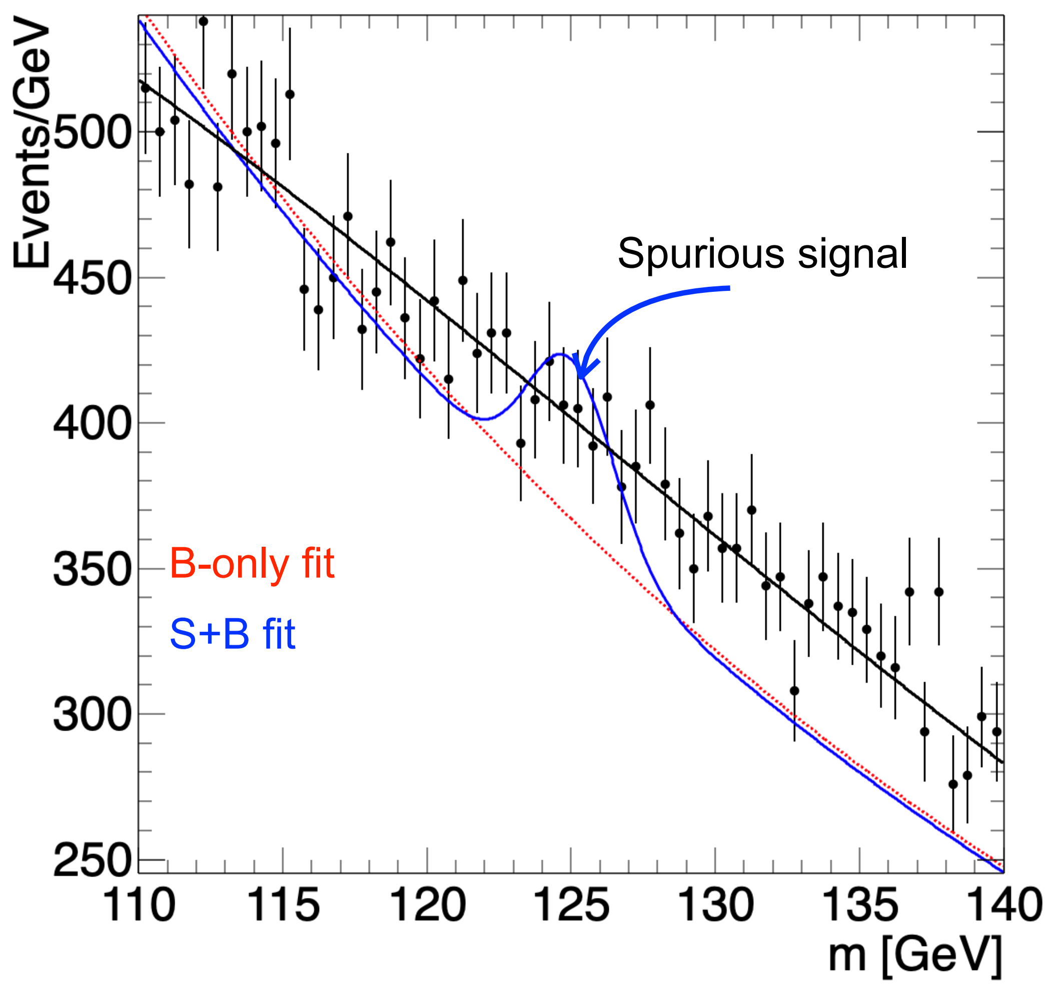
- Possible systematic mismodelling due to function choice leading to apparent signal

Choices:

- What function to use?
- What systematic uncertainty to assign?

Use MC sample of background to perform S+B fits.

- Use function with lowest obtained $S_{spurious}$, and said $S_{spurious}$ as systematic



Challenges:

- **Conceptual:** use MC sample not deemed reliable for modelling the background
- **Practical:** required MC sample orders of magnitude larger than dataset of interest

Spurious signal

H→γγ inclusive fiducial cross section measurement uncertainties

Source	Uncertainty (%)
Fit (stat.)	10
Fit (syst.)	8.3
Photon energy scale & resolution	4.0
Background modeling (spurious signal)	7.3
Correction factor	5.2
Photon isolation efficiency	4.6
Pileup	1.9
Photon ID efficiency	1.3
Trigger efficiency	0.7
Dalitz Decays	0.4
Theoretical modeling	$^{+0.3}_{-0.4}$
Diphoton vertex selection	0.1
Photon energy scale & resolution	0.1
Luminosity	2.0
Total	14

ATLAS-CONF-2018-028

Higgs boson mass measurement with H→ZZ→4l and H→γγ

Source	Systematic uncertainty in m_H [MeV]
EM calorimeter response linearity	60
Non-ID material	55
EM calorimeter layer intercalibration	55
$Z \to ee$ calibration	45
ID material	45
Lateral shower shape	40
Muon momentum scale	20
Conversion reconstruction	20
$H \to \gamma\gamma$ background modelling	20
$H \to \gamma\gamma$ vertex reconstruction	15
e/γ energy resolution	15
All other systematic uncertainties	10

Phys. Lett. B 784 (2018) 345

Discrete Profiling Method

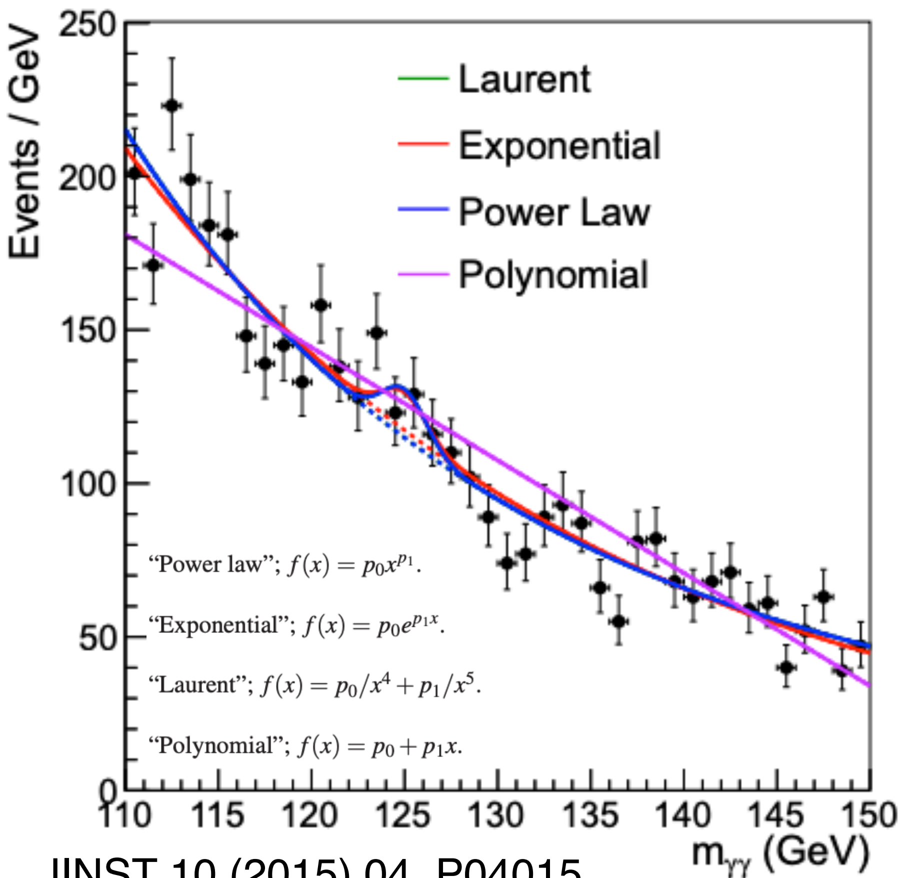
CMS uses the **discrete profiling method**

- Combine different parametric models at the likelihood level
- Treat shape options as discrete nuisance parameter
  - Use envelope of individual likelihood scans to obtain result

Discrete Profiling Method

CMS uses the **discrete profiling method**

- Combine different parametric models at the likelihood level
- Treat shape options as discrete nuisance parameter
  - Use envelope of individual likelihood scans to obtain result

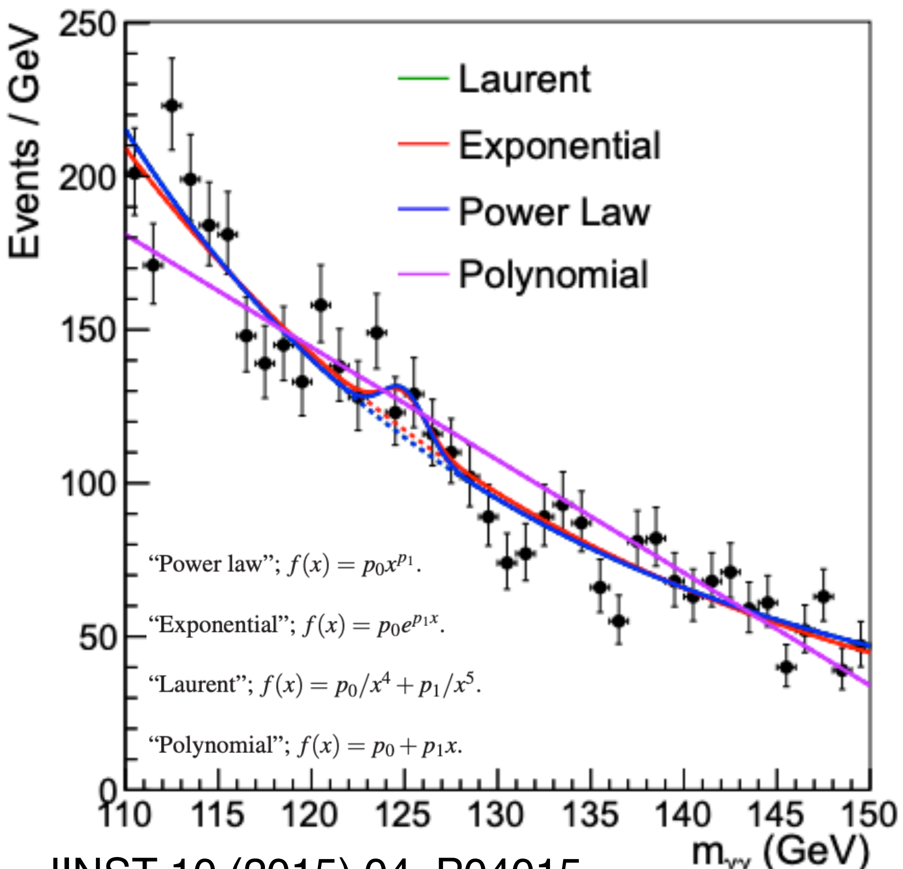


JINST 10 (2015) 04, P04015

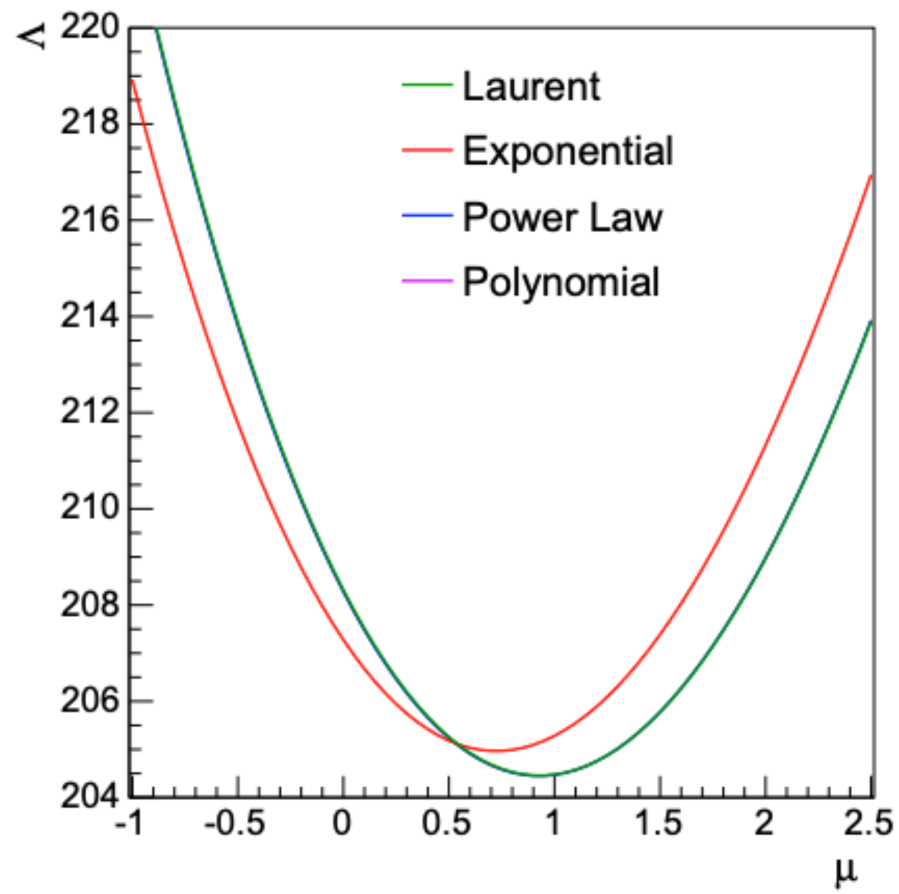
Discrete Profiling Method

CMS uses the **discrete profiling method**

- Combine different parametric models at the likelihood level
- Treat shape options as discrete nuisance parameter
  - Use envelope of individual likelihood scans to obtain result



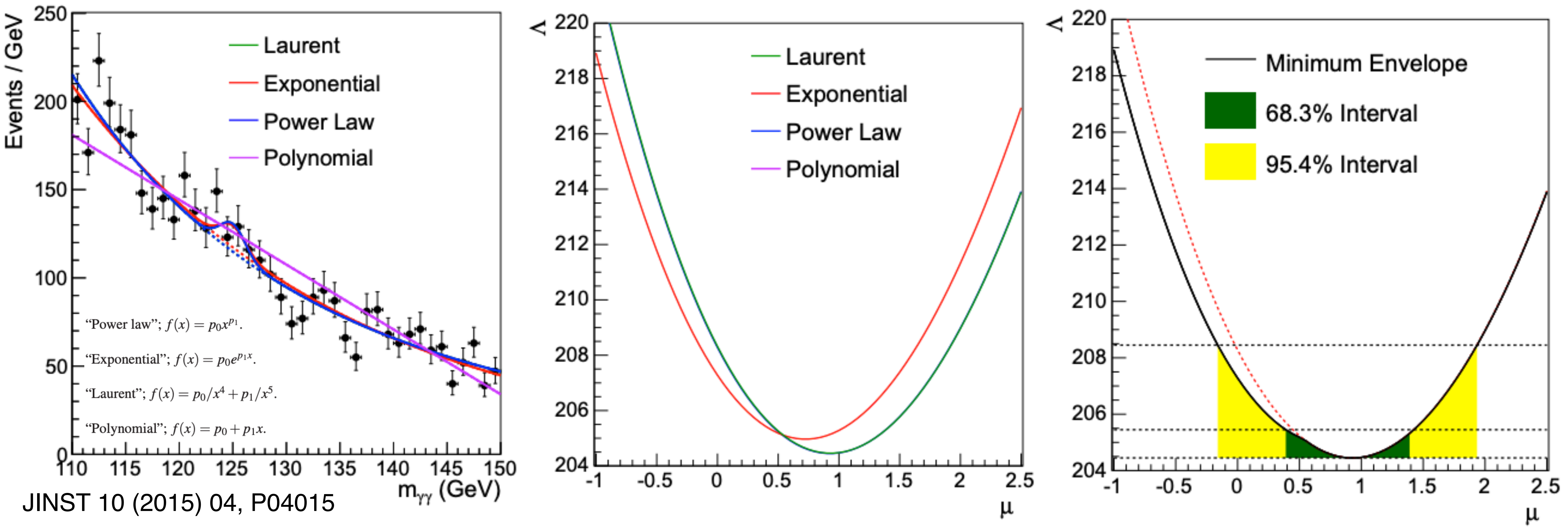
JINST 10 (2015) 04, P04015



Discrete Profiling Method

CMS uses the **discrete profiling method**

- Combine different parametric models at the likelihood level
- Treat shape options as discrete nuisance parameter
  - Use envelope of individual likelihood scans to obtain result



JINST 10 (2015) 04, P04015

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

Correction: penalise functions with more parameters

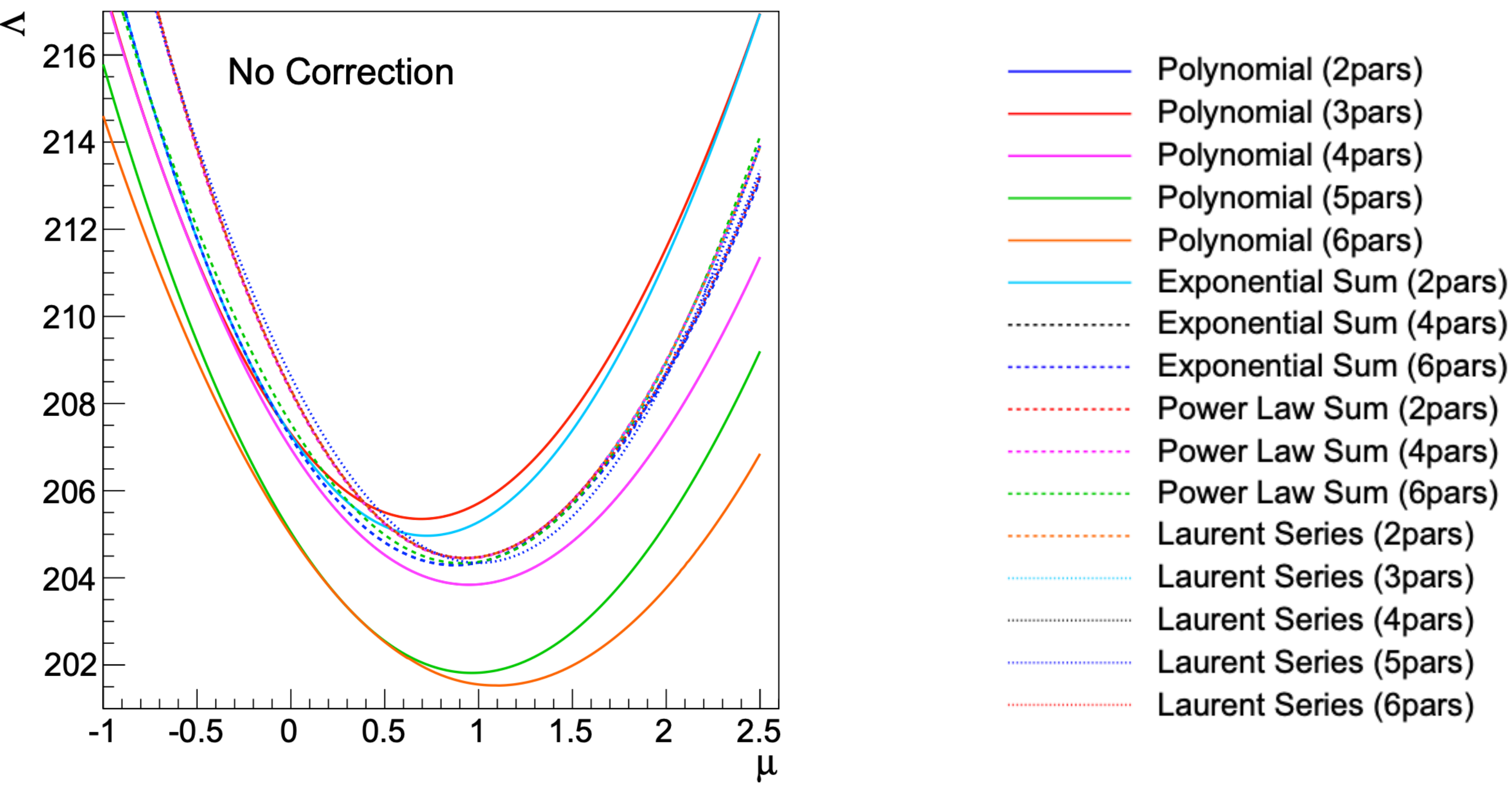
- Inspired by p-value and Akaike information criterion
- Parametrised as $\Lambda_{corr} = \Lambda + cN_{par}$
- Bias vs coverage trade-off versus c studied case-by-case

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

Correction: penalise functions with more parameters

- Inspired by p-value and Akaike information criterion
- Parametrised as $\Lambda_{corr} = \Lambda + cN_{par}$
- Bias vs coverage trade-off versus c studied case-by-case



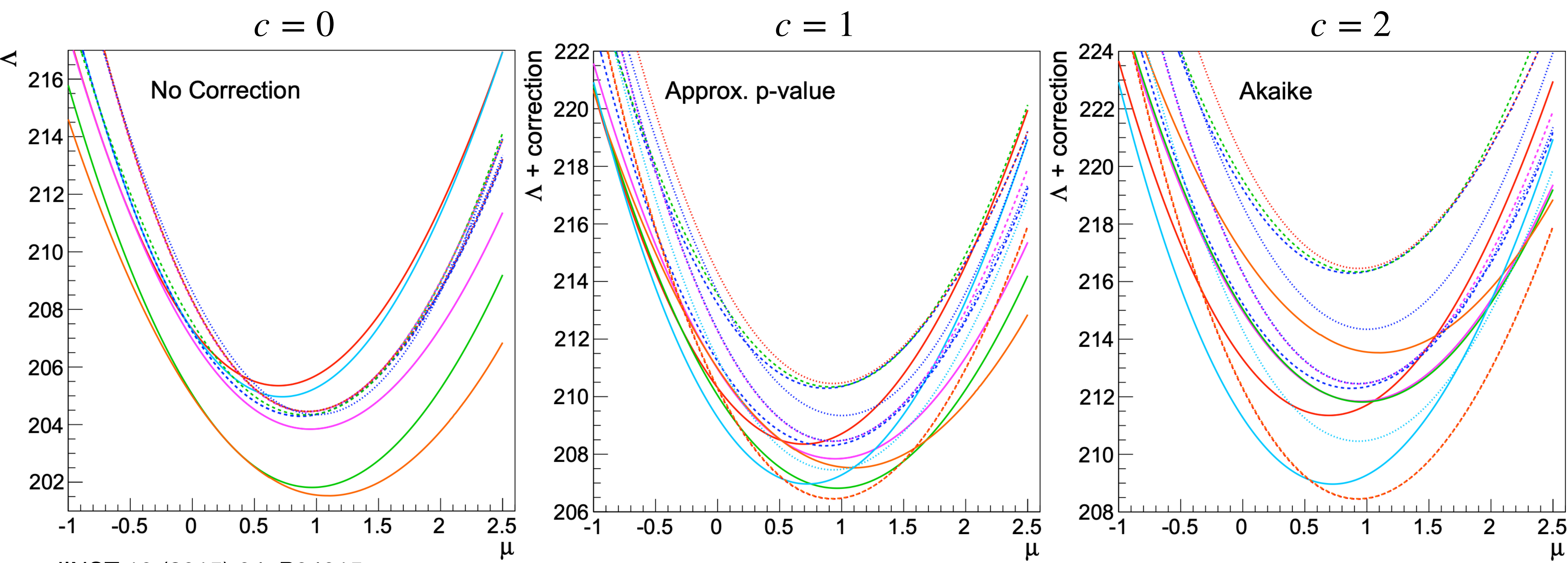
JINST 10 (2015) 04, P04015

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

Correction: penalise functions with more parameters

- Inspired by p-value and Akaike information criterion
- Parametrised as $\Lambda_{corr} = \Lambda + cN_{par}$
- Bias vs coverage trade-off versus c studied case-by-case



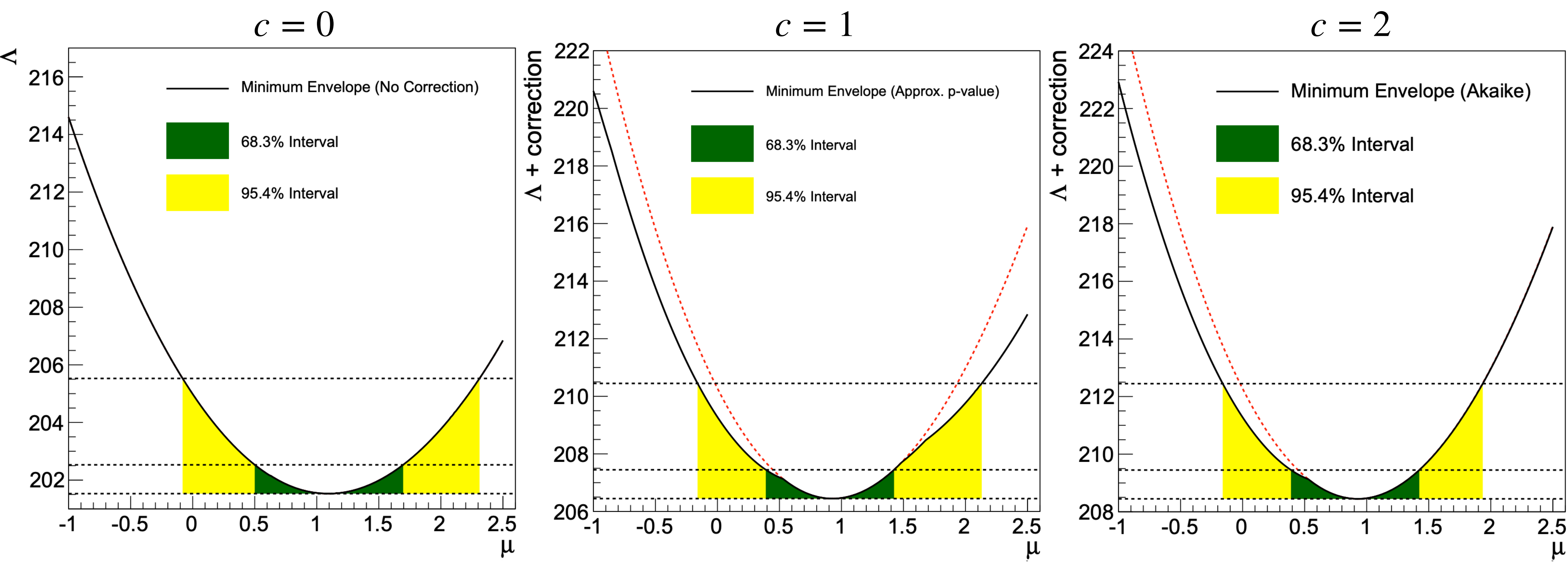
JINST 10 (2015) 04, P04015

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

Correction: penalise functions with more parameters

- Inspired by p-value and Akaike information criterion
- Parametrised as $\Lambda_{corr} = \Lambda + cN_{par}$
- Bias vs coverage trade-off versus c studied case-by-case



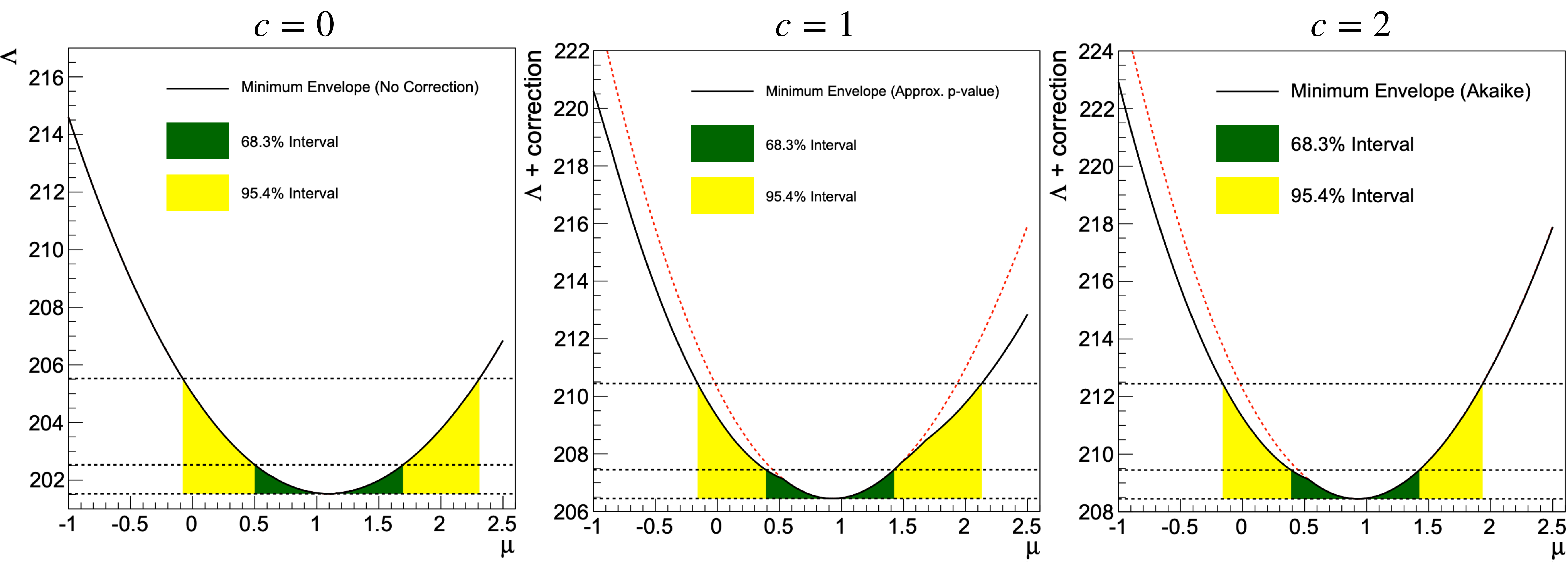
JINST 10 (2015) 04, P04015

The Discrete Profiling Method

Practical and conceptual complications when models have different N_{par}

Correction: penalise functions with more parameters

- Inspired by p-value and Akaike information criterion
- Parametrised as $\Lambda_{corr} = \Lambda + cN_{par}$
- Bias vs coverage trade-off versus c studied case-by-case



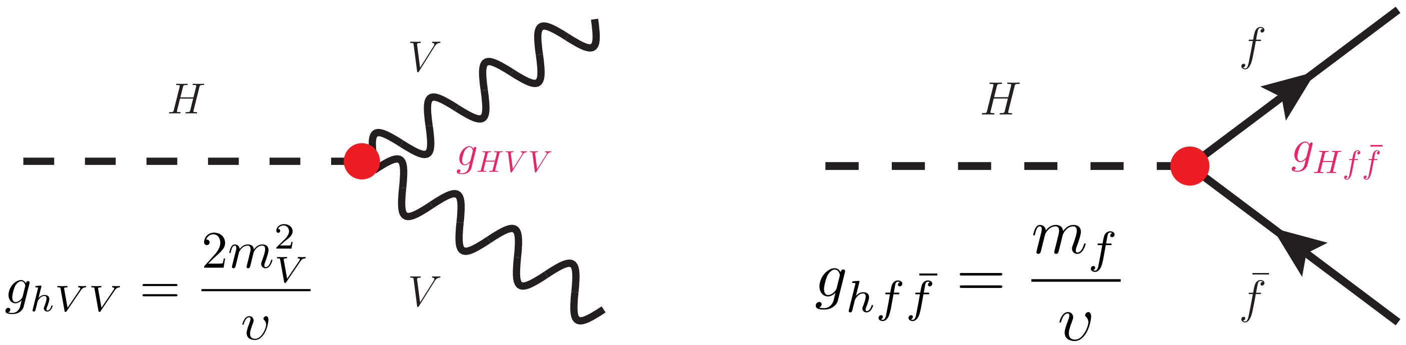
JINST 10 (2015) 04, P04015

Common systematic effects across categories: All combinations of functions and nuisance parameters need to be scanned

→Naive implementation impractical and usually approximations used.

Higgs-fermion interactions

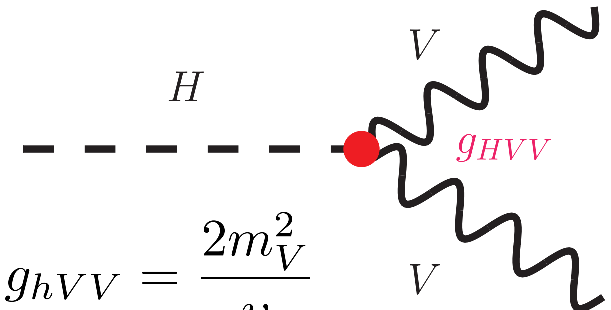
- **Higgs interactions to vector bosons:** defined by symmetry breaking
- **Higgs interactions to fermions:** ad-hoc hierarchical Yukawa couplings $\propto m_f$



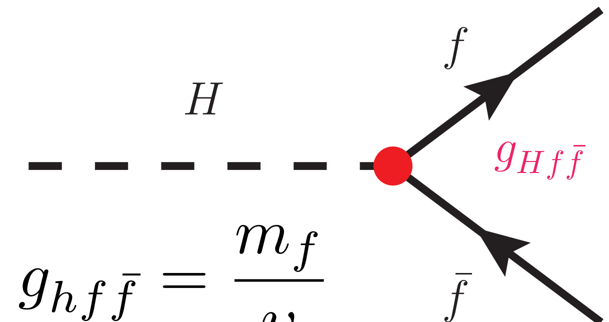
$$g_{hVV} = \frac{2m_V^2}{v} \qquad g_{hf\bar{f}} = \frac{m_f}{v}$$

Higgs-fermion interactions

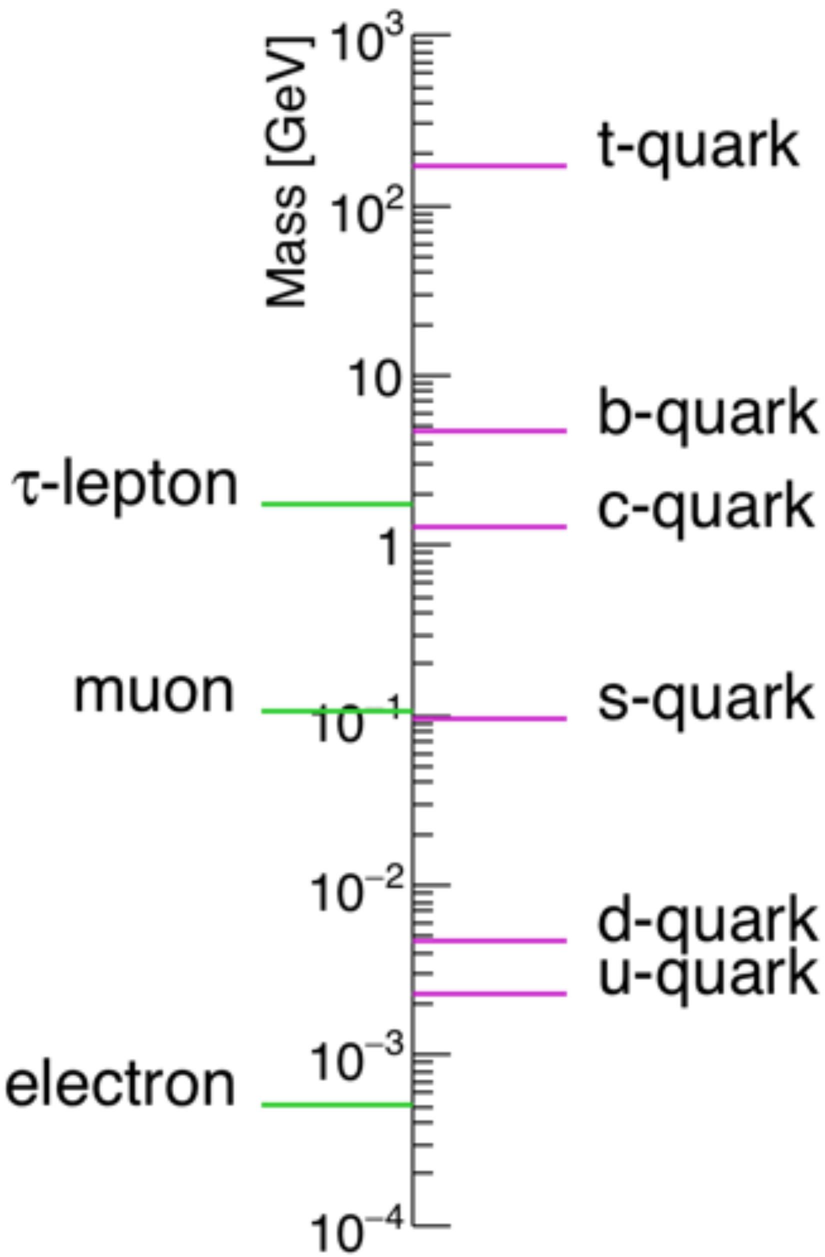
- **Higgs interactions to vector bosons:** defined by symmetry breaking
- **Higgs interactions to fermions:** ad-hoc hierarchical Yukawa couplings $\propto m_f$



$$g_{hVV} = \frac{2m_V^2}{v}$$



$$g_{hf\bar{f}} = \frac{m_f}{v}$$



$$\frac{m_e}{m_t} \approx 3 \times 10^{-6}$$

- Yukawa couplings **not** imposed by fundamental principle
- Modified Higgs-fermion couplings in BSM scenarios
- Probing fermion mass generation scale→independent task

Standard Model successful

but matter particle mass

hierarchy unexplained!

UNIVERSITY OF BIRMINGHAM

Extended Higgs sectors

- The Standard Model **Higgs sector is an $SU(2)_L$ doublet of complex scalar fields:** this is the most economic way to obtain spontaneous symmetry breaking
- **Extended Higgs sectors** are possible, and can potentially provide answers to a number of open questions
- The ρ parameter puts tight constraints on model viability
  - For SM ρ=1 (with small corrections)
  - Constraints naturally fulfilled for appropriate configurations of scalar singlets and doublets

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1.00039 \pm 0.00019$$

UNIVERSITY OF BIRMINGHAM

Extended Higgs sectors

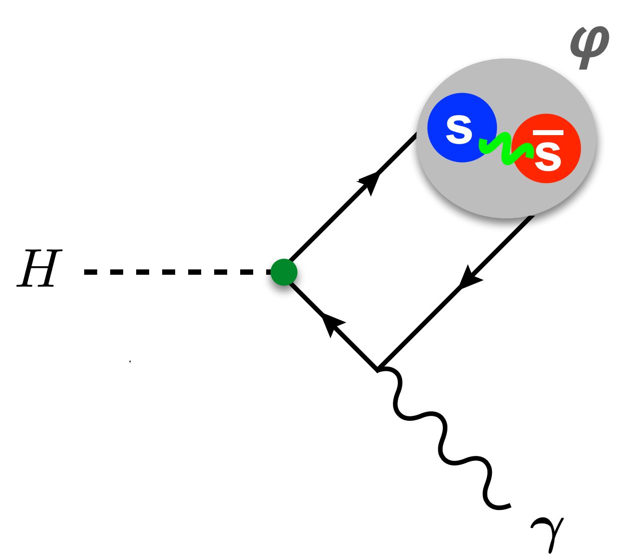
- The Standard Model **Higgs sector is an $SU(2)_L$ doublet of complex scalar fields:** this is the most economic way to obtain spontaneous symmetry breaking
- **Extended Higgs sectors** are possible, and can potentially provide answers to a number of open questions
- The ρ parameter puts tight constraints on model viability
  - For SM ρ=1 (with small corrections)
  - Constraints naturally fulfilled for appropriate configurations of scalar singlets and doublets

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1.00039 \pm 0.00019$$

- A number of possibilities with **rich phenomenology**:
  - Higgs double with one or more scalar singlets,
  - Two Higgs Doublets (2HDM),
  - 2HDM with additional scalar singlet (2HDM+S)
- Particularly interesting: additional scalar lighter than observed Higgs boson.
  - $h \to aa$
  - $h \to Za$

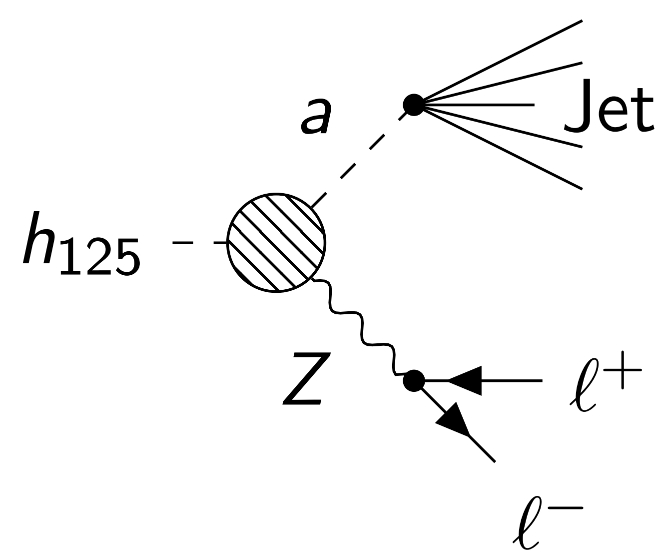
Searches for new physics

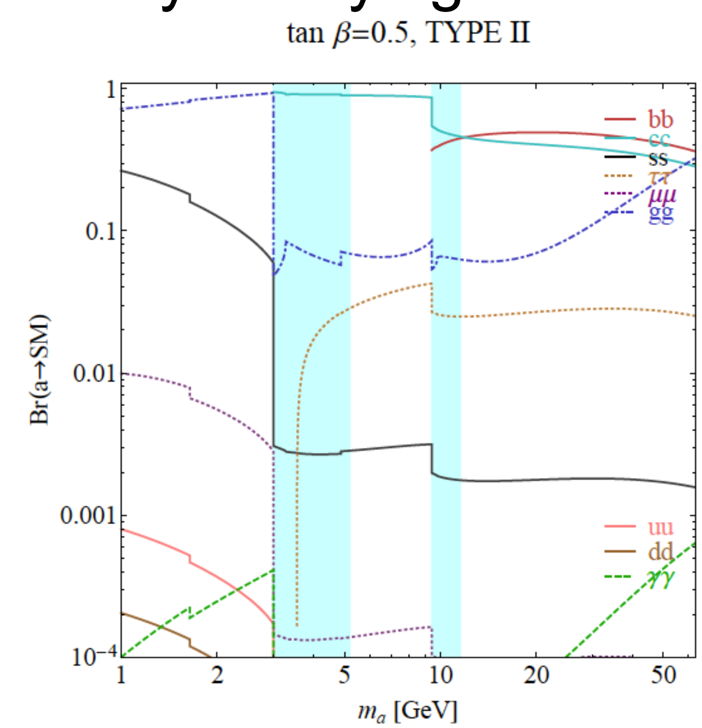
Exclusive Higgs decays



$$BR(h \to \phi\gamma) = (2.31 \pm 0.03_{f_\phi} \pm 0.11_{h\to\gamma\gamma}) \cdot 10^{-6}$$

Higgs decays to light hadronically decaying scalars





PRD 90 (2014) 7, 075004

These analyses share the challenge that the respective backgrounds are not straightforward to model with simulations.

Beyond Parametric Methods

Parametric methods have several advantages but also important issues

In the following: aim to develop **fully data-driven non-parametric background models**

Beyond Parametric Methods

Parametric methods have several advantages but also important issues

In the following: aim to develop **fully data-driven non-parametric background models**

Non-Parametric Data-Driven Background Modelling using Conditional Probabilities

---

A. Chisholm, T. Neep, K. Nikolopoulos, R. Owen,[1] E. Reynolds[2] and J. Silva

School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom

E-mail: andrew.chisholm@cern.ch, tom.neep@cern.ch, konstantinos.nikolopoulos@cern.ch, rhys.owen@cern.ch, elliot.reynolds@cern.ch, julia.manuela.silva@cern.ch

arXiv:2112.00650

Beyond Parametric Methods

Parametric methods have several advantages but also important issues

In the following: aim to develop **fully data-driven non-parametric background models**

Non-Parametric Data-Driven Background Modelling using Conditional Probabilities

A. Chisholm, T. Neep, K. Nikolopoulos, R. Owen,[1] E. Reynolds[2] and J. Silva

School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom

E-mail: andrew.chisholm@cern.ch, tom.neep@cern.ch, konstantinos.nikolopoulos@cern.ch, rhys.owen@cern.ch, elliot.reynolds@cern.ch, julia.manuela.silva@cern.ch

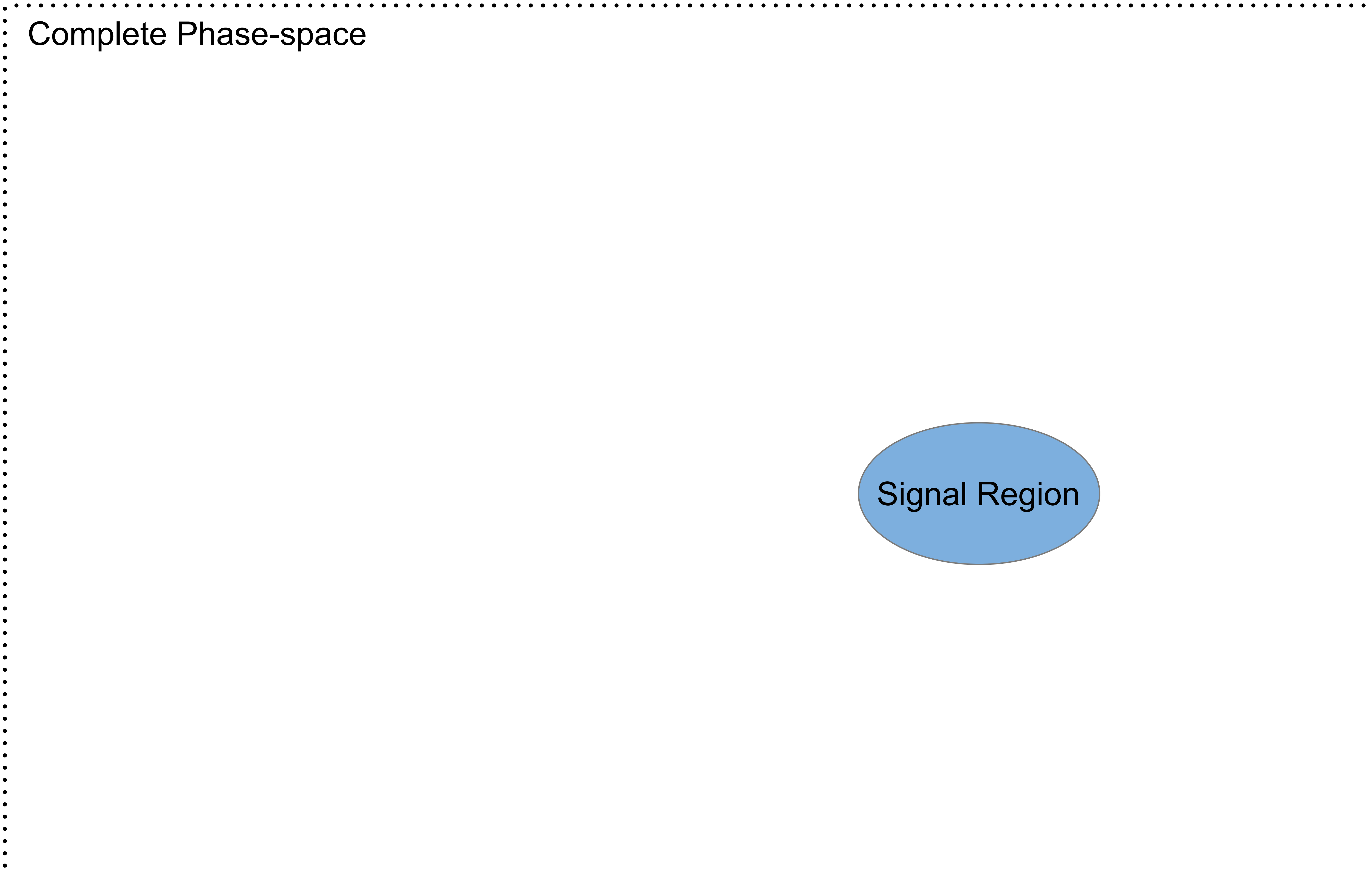
arXiv:2112.00650

Methods motivated by specific analyses, but with wide applicability

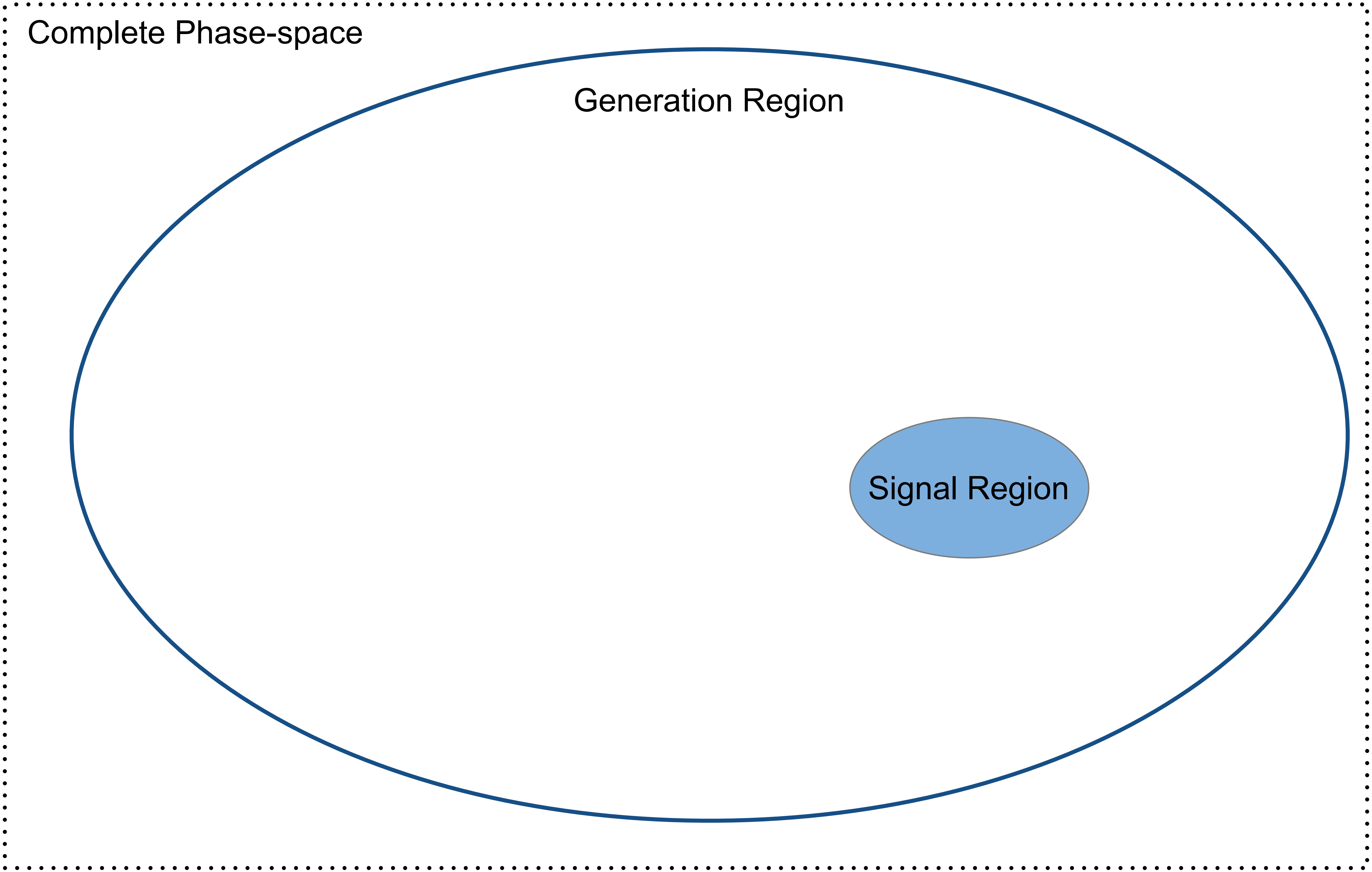
The strategy

Complete Phase-space

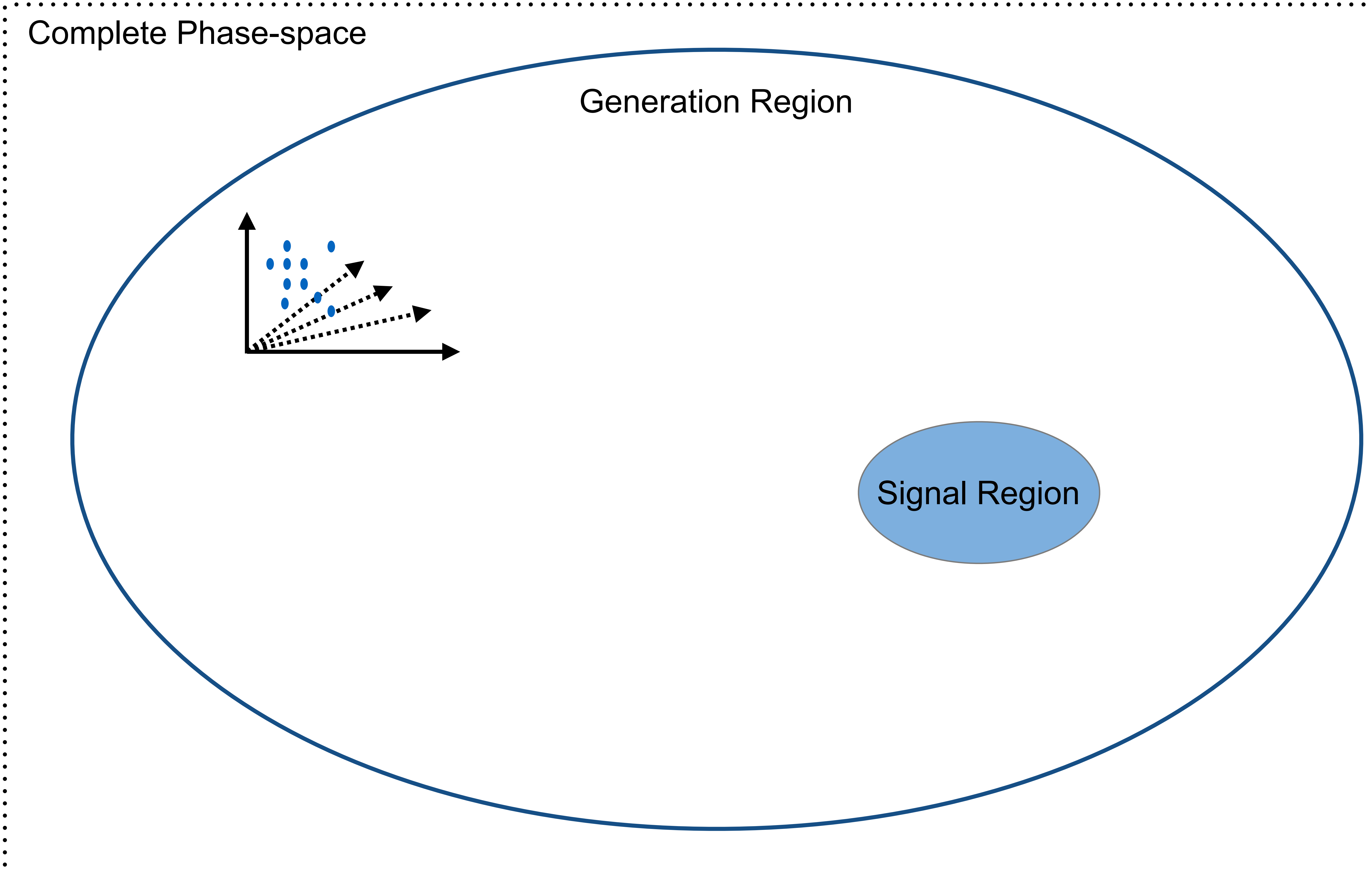
The strategy



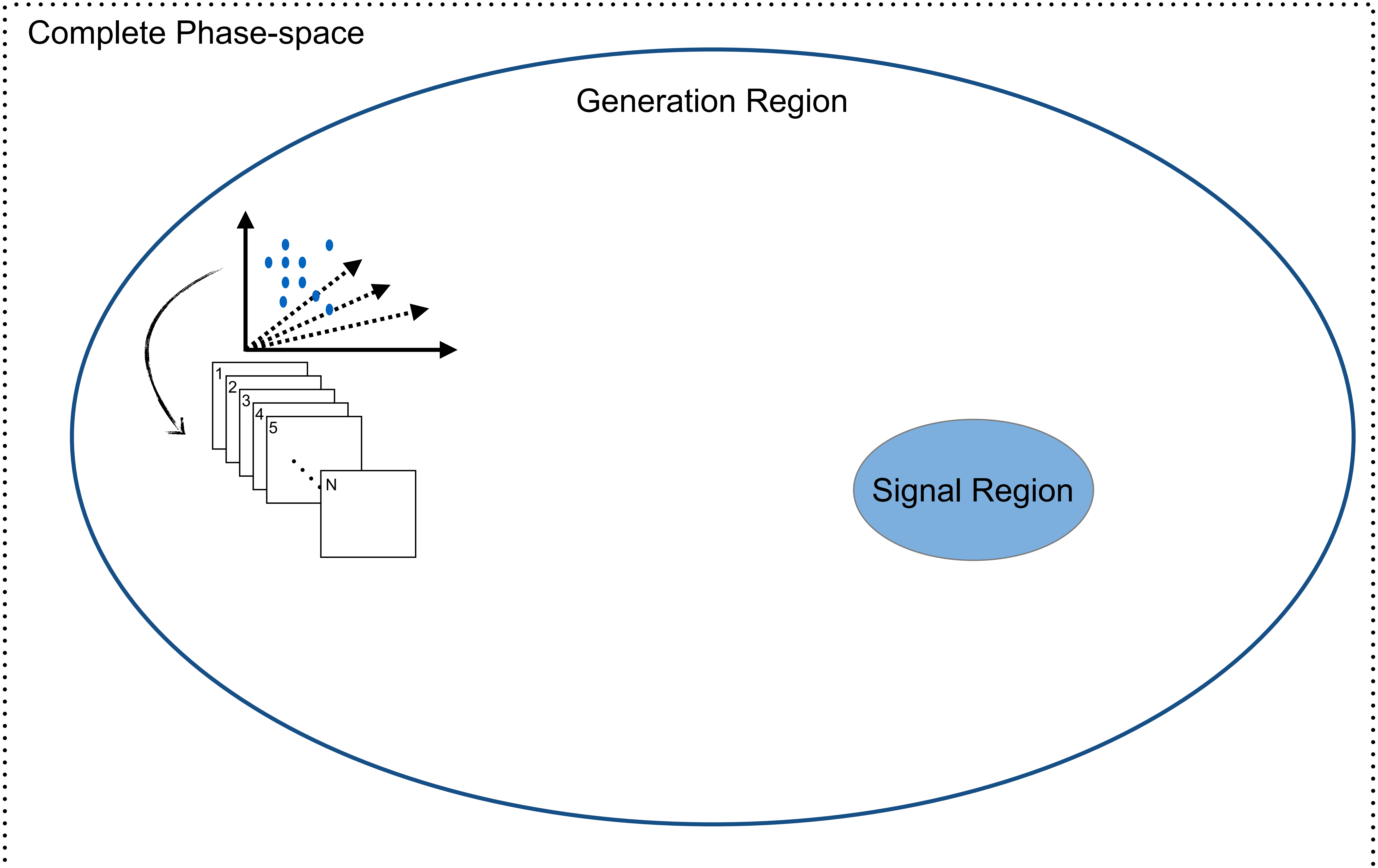
The strategy



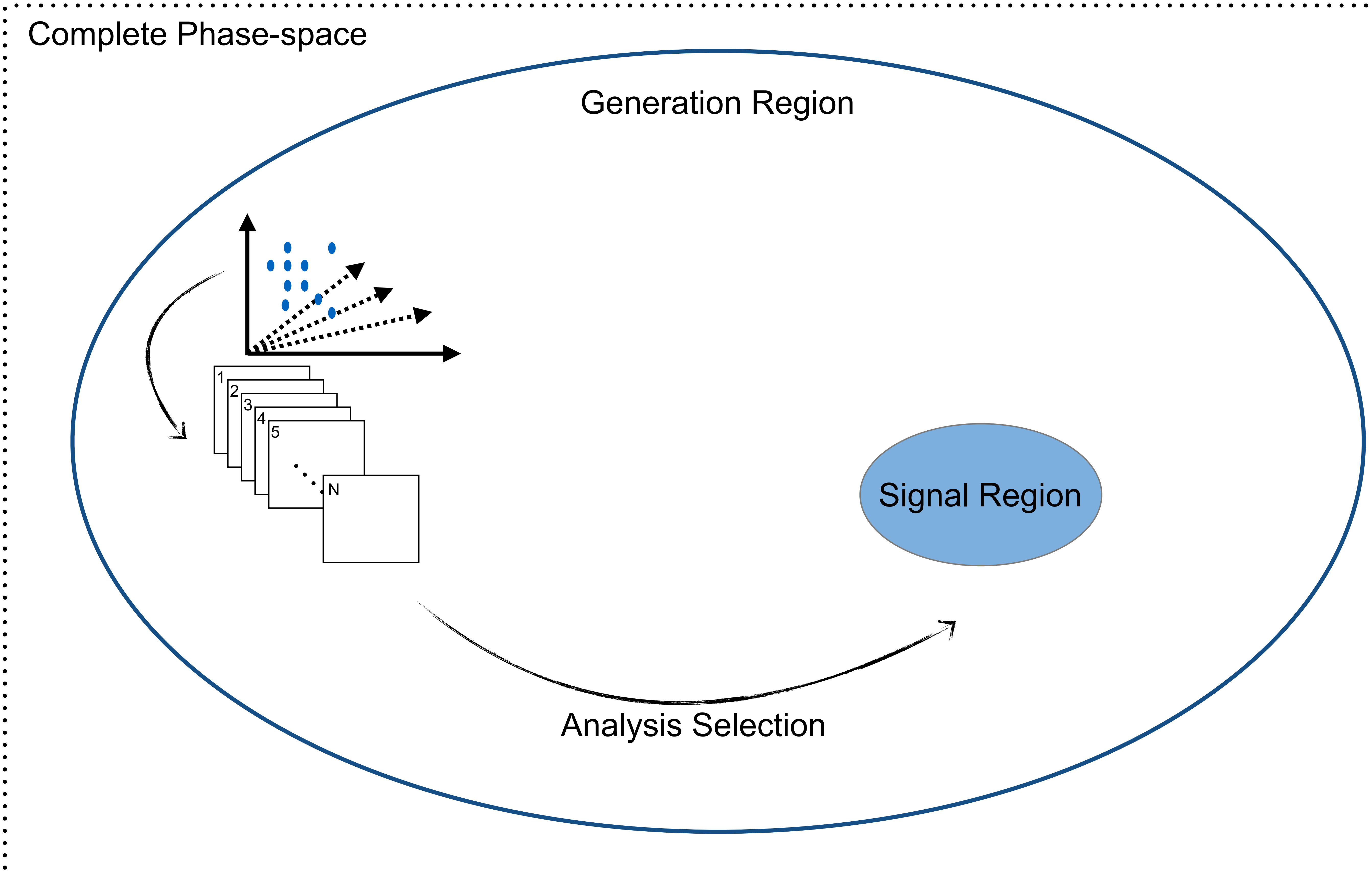
The strategy



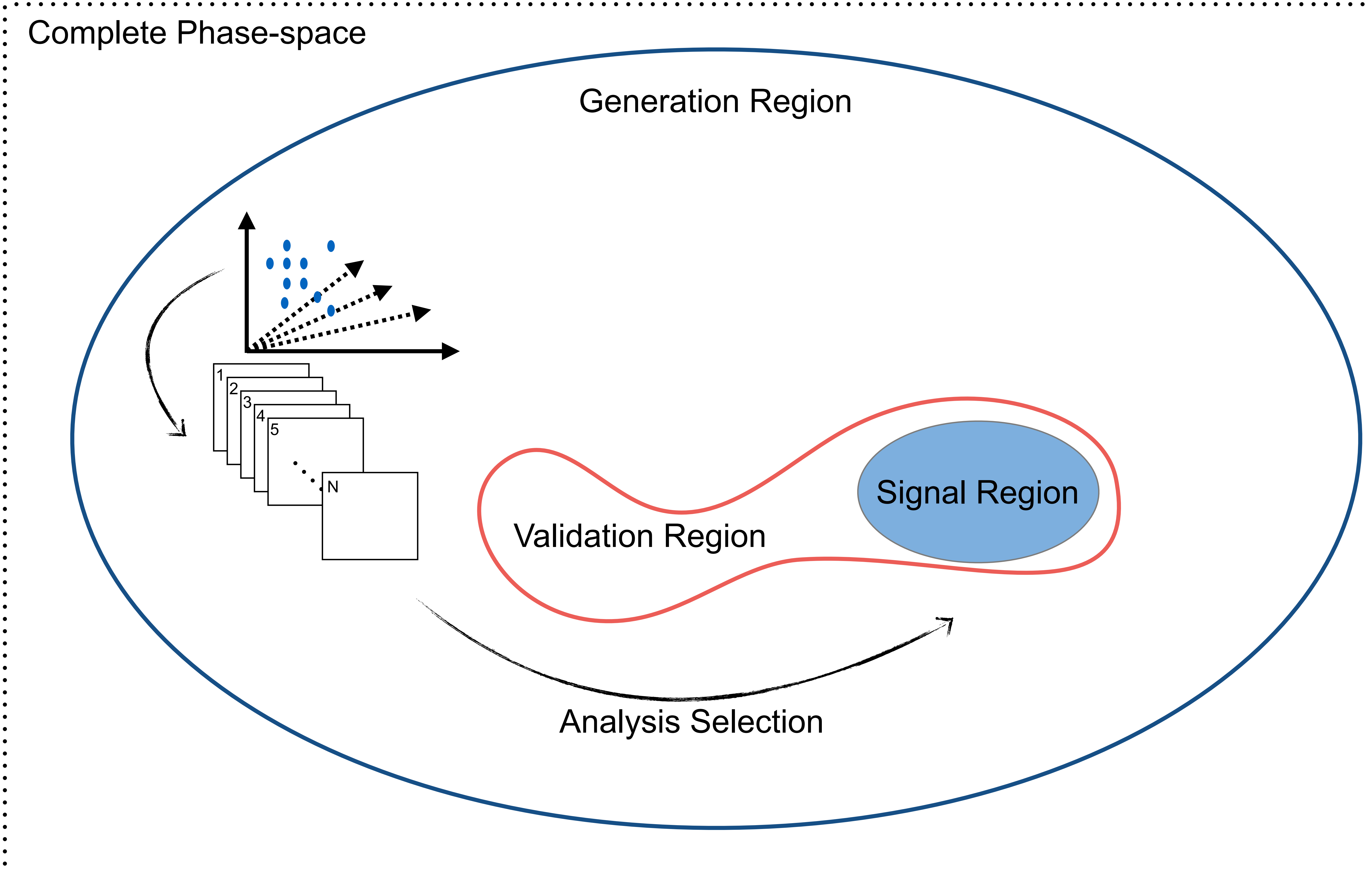
The strategy



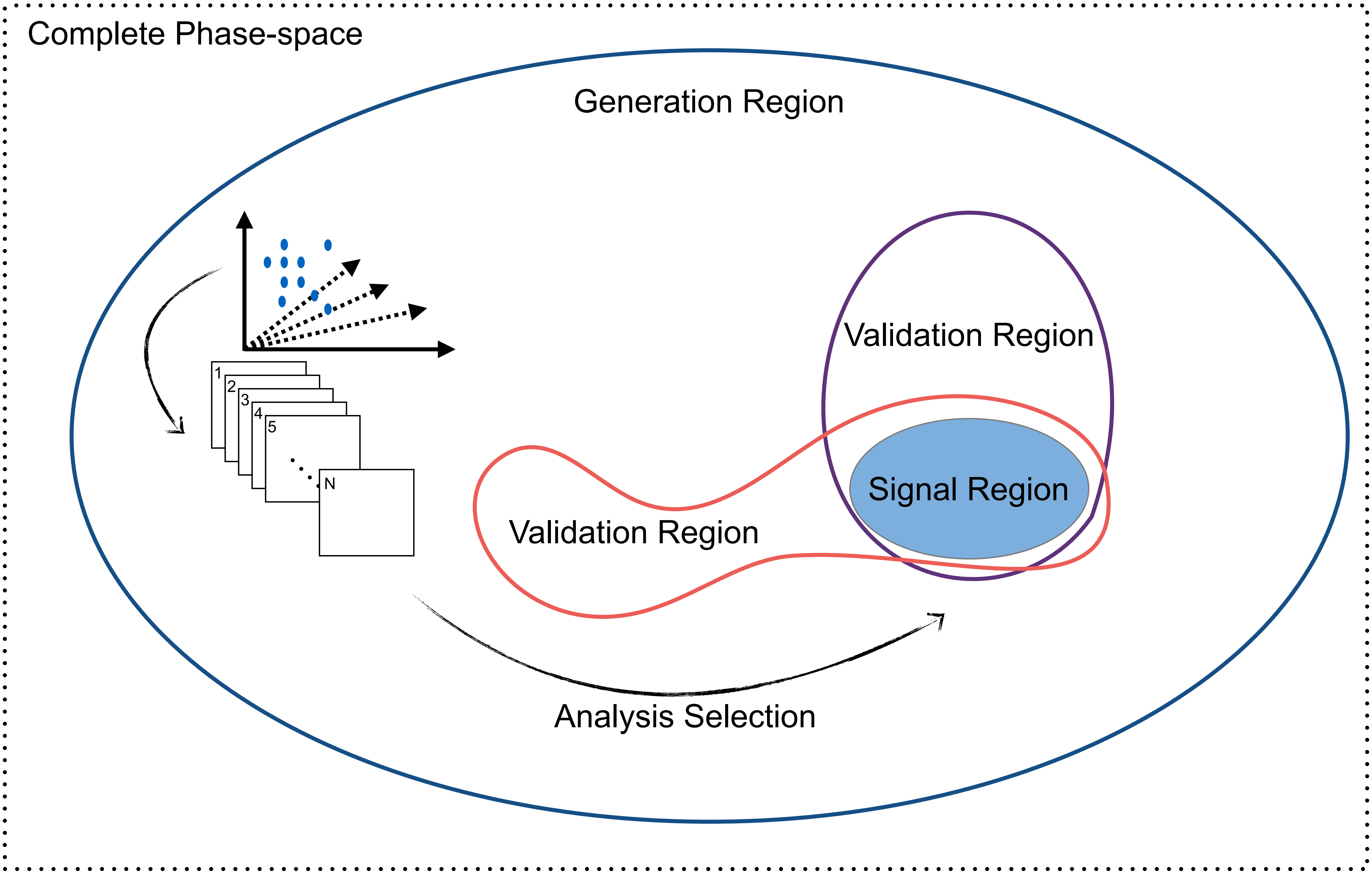
The strategy



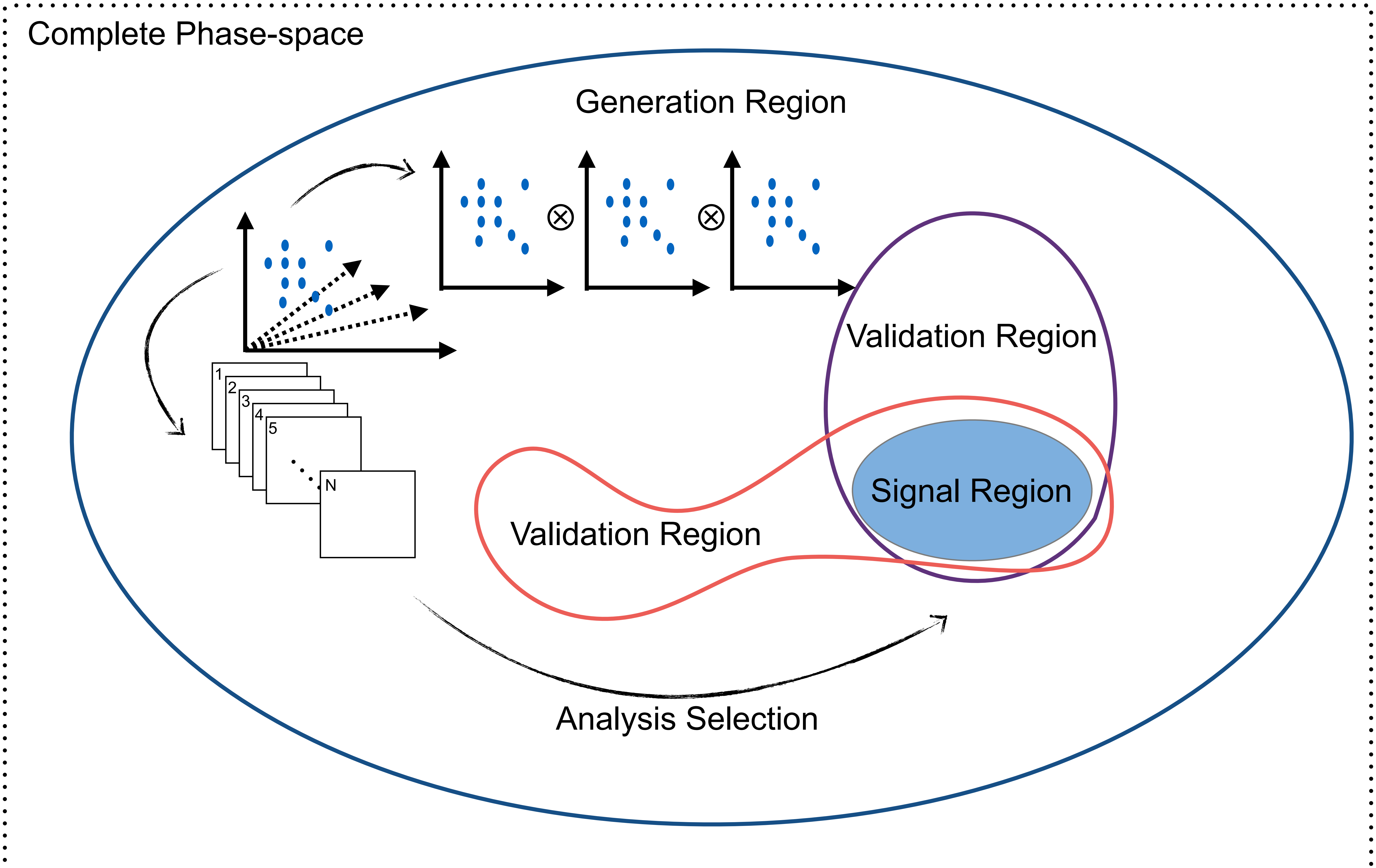
The strategy



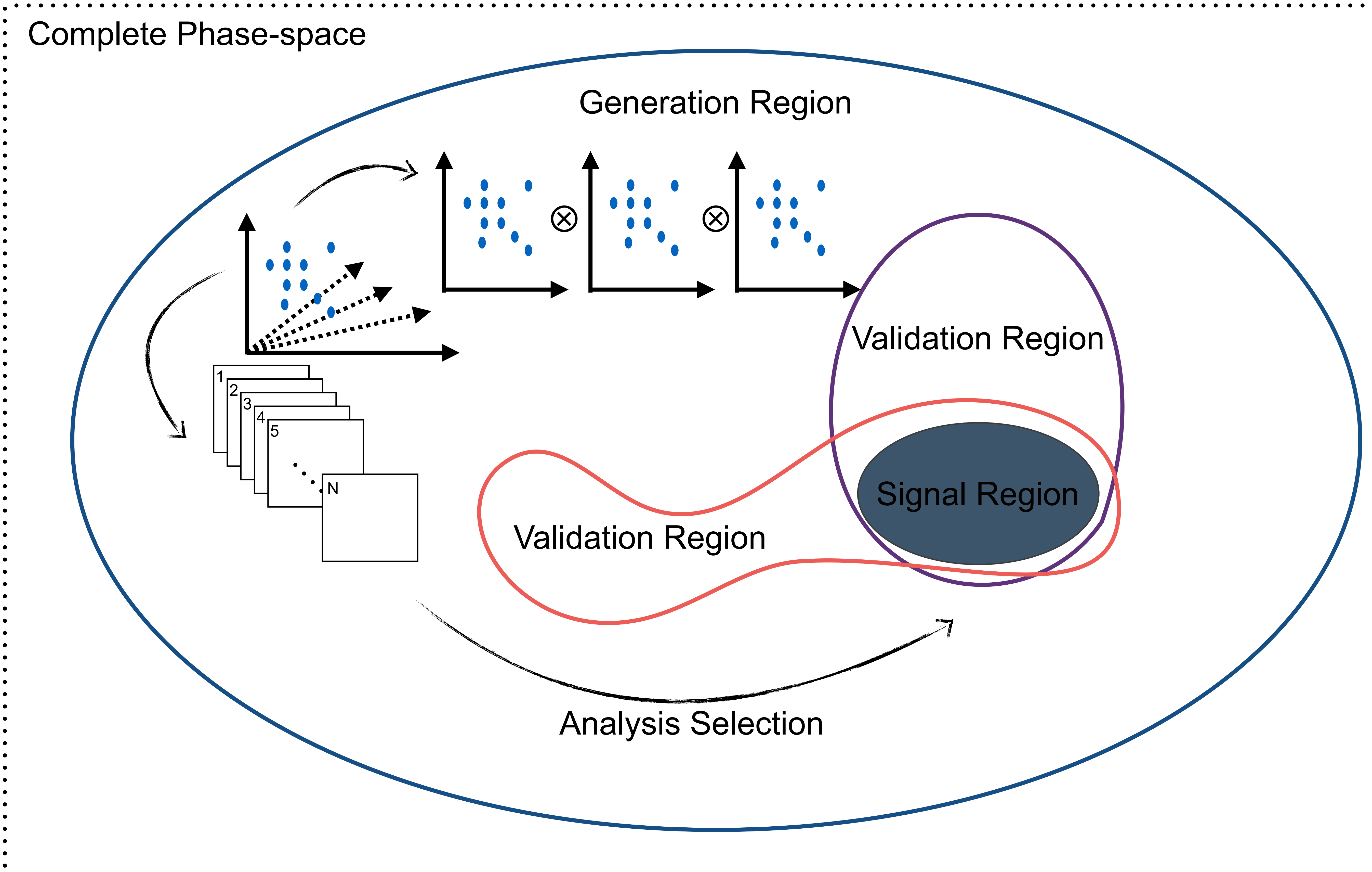
The strategy



The strategy

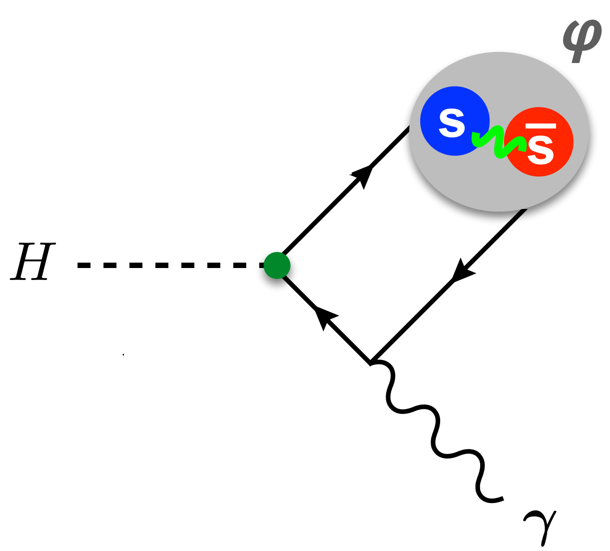


The strategy



h/Z→ϕγ/ργ

Exclusive Higgs decays

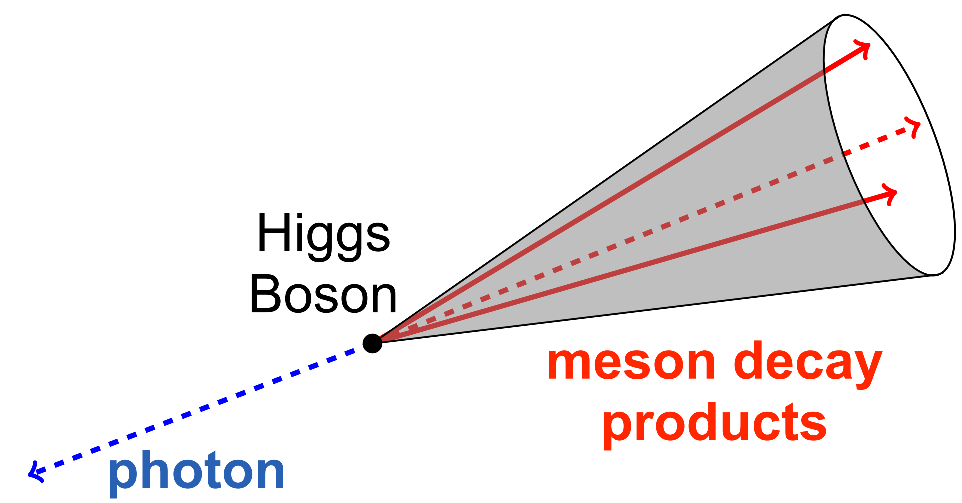


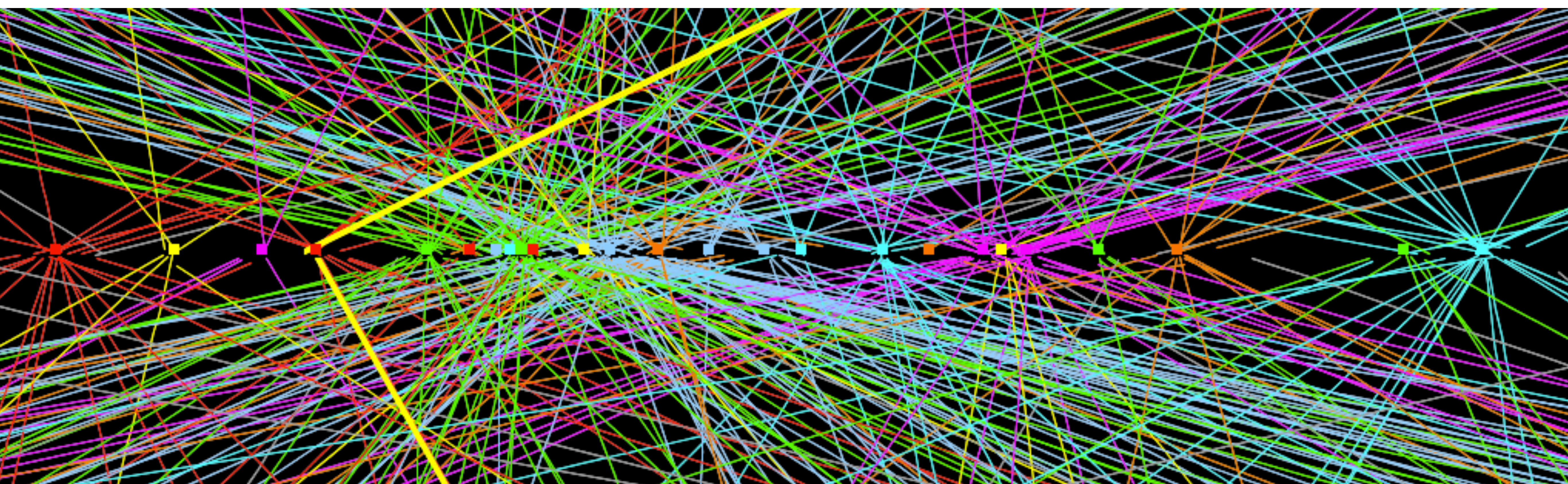
$$BR(h \to \phi\gamma) = (2.31 \pm 0.03_{f_\phi} \pm 0.11_{h\to\gamma\gamma}) \cdot 10^{-6}$$

h/Z→φγ/ργ

- **Exclusive decays → distinct experimental signature**
  - Pair of collimated high-pT isolated tracks

  recoils against high-pT isolated photon
- **Meson decays:**
  - **$\phi \to K^+K^-$**, BR=49%
  - **$\rho \to \pi^+\pi^-$**, BR~100%
- **Small opening angles between decay products**
  - Particularly for $\phi \to K^+K^-$
  - Tracking in dense environments

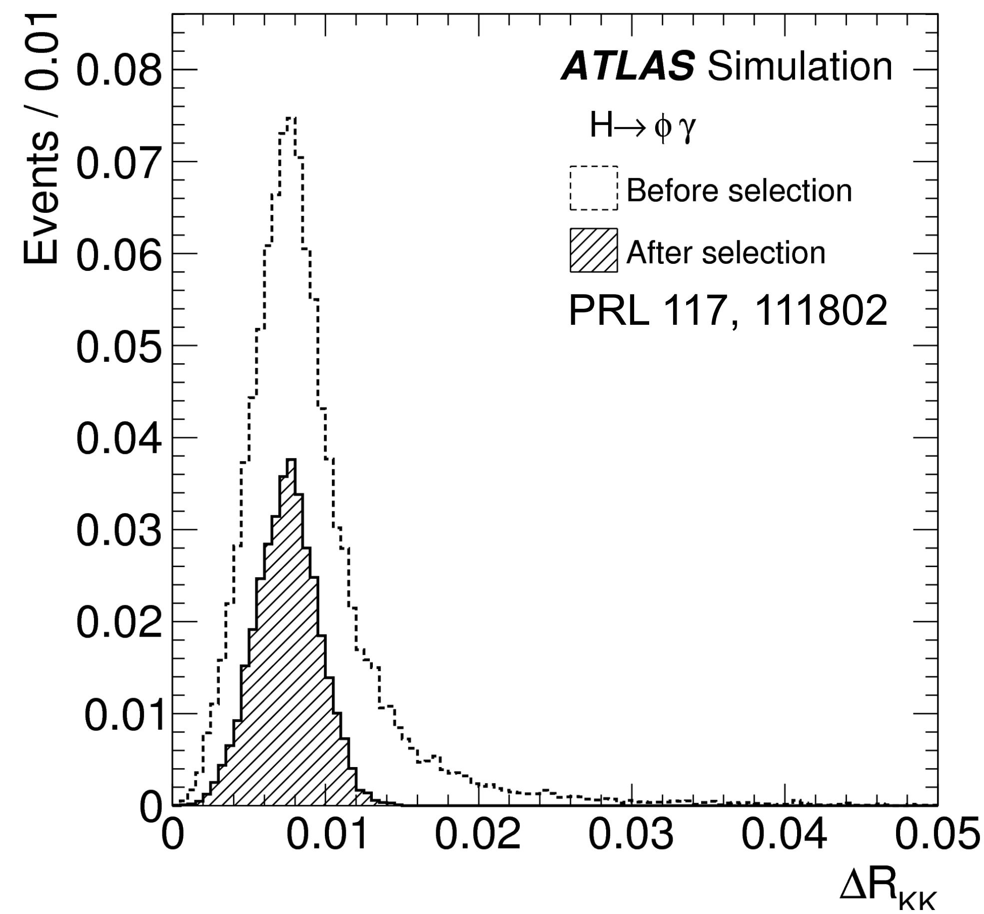


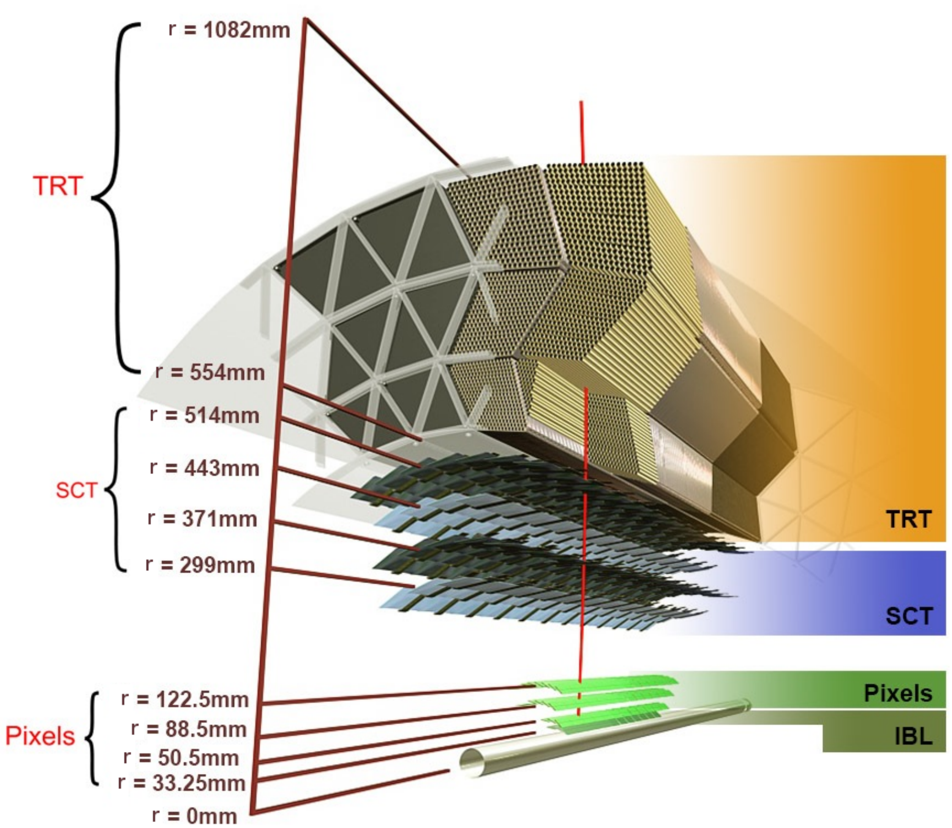


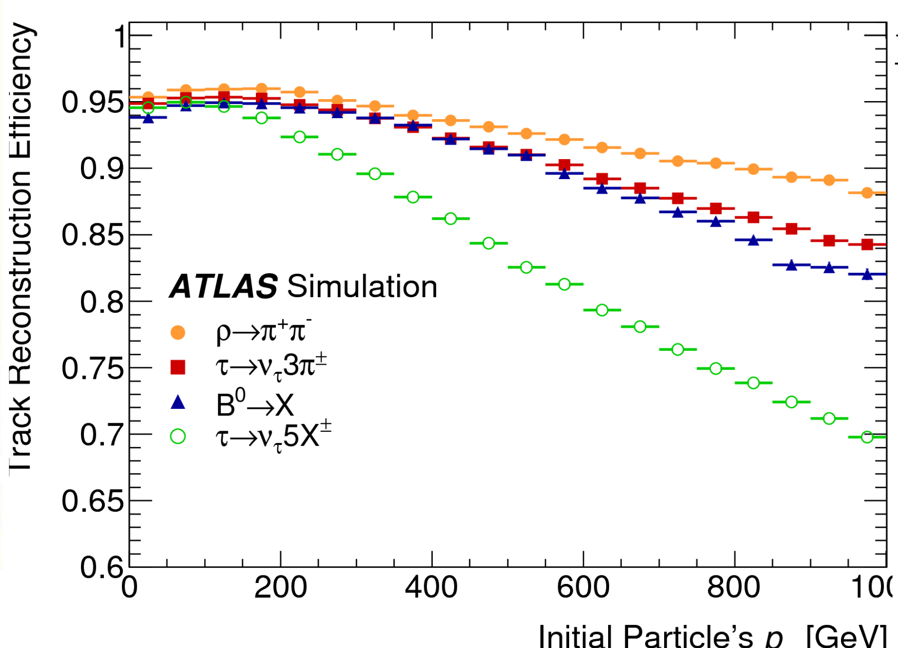
Z→μμ candidate with 25 reconstructed vertices from the 2012 run. Only good quality tracks with pT>0.4GeV are shown

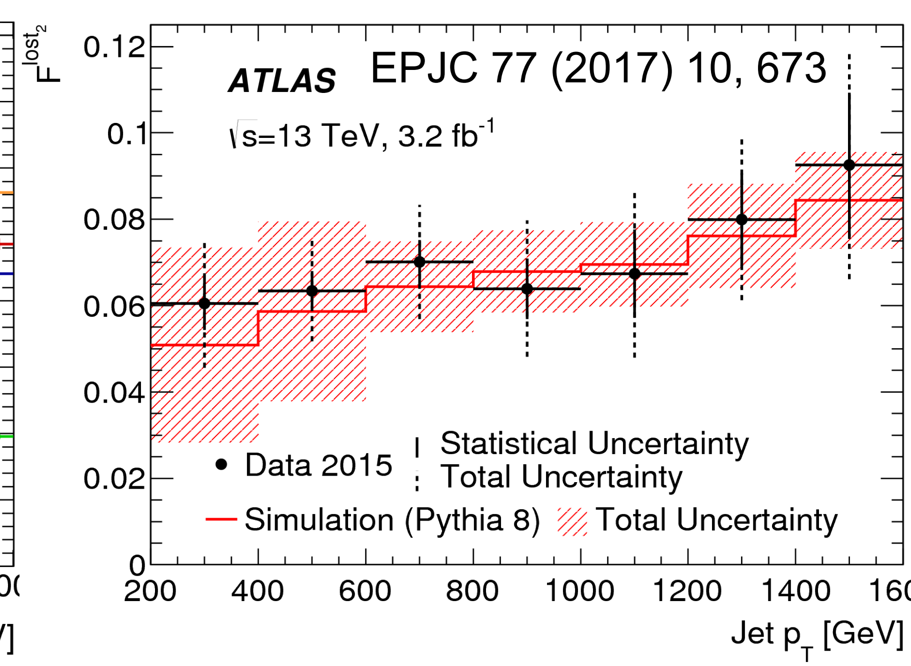
h/Z→ϕγ/ργ

- **Exclusive decays → distinct experimental signature**
  - Pair of collimated high-pT isolated tracks
    recoils against high-pT isolated photon
- **Meson decays:**
  - **$\phi \to K^+K^-$**, BR=49%
  - **$\rho \to \pi^+\pi^-$**, BR~100%
- **Small opening angles between decay products**
  - Particularly for $\phi \to K^+K^-$
  - Tracking in dense environments

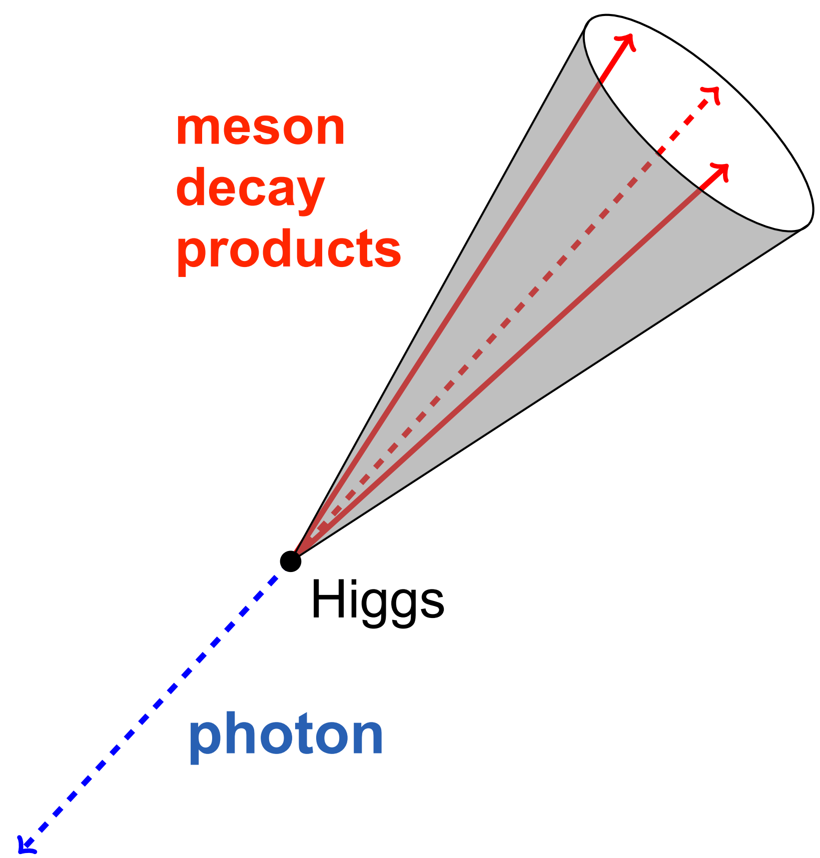


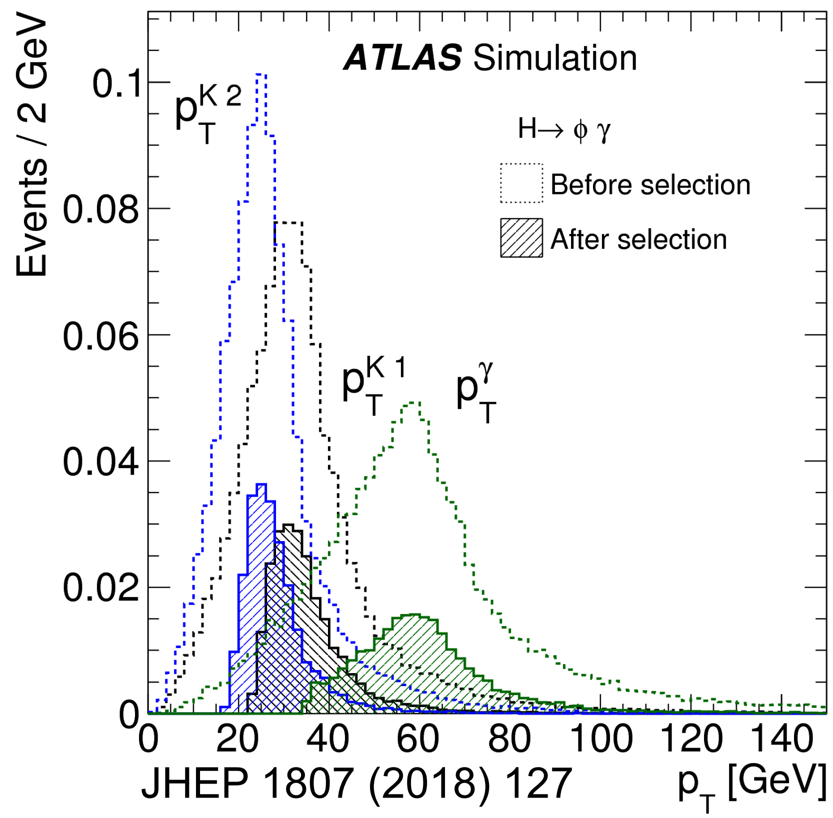






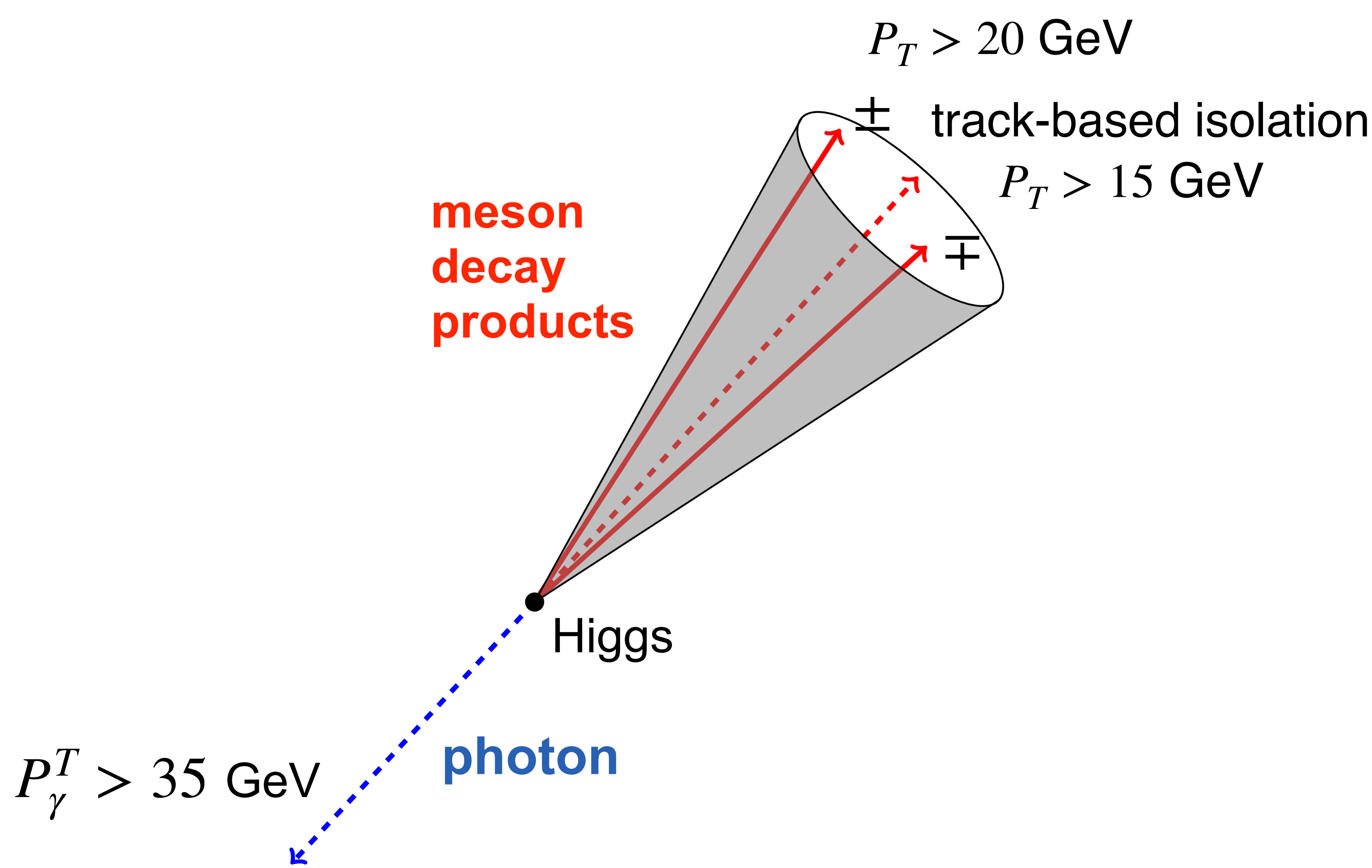
Event Selection



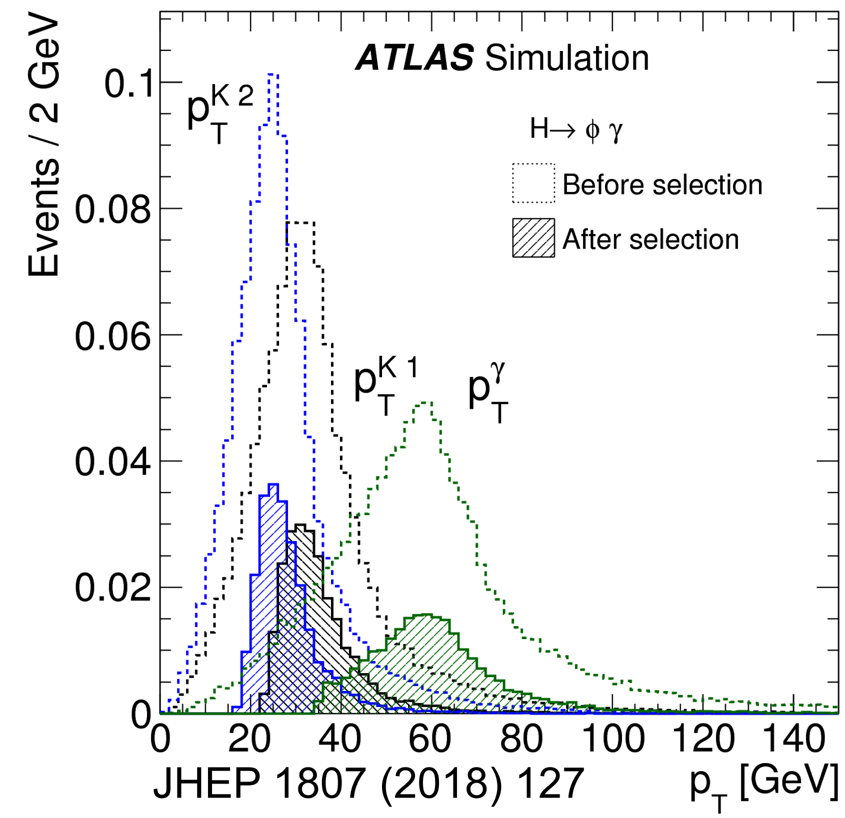


JHEP 1807 (2018) 127

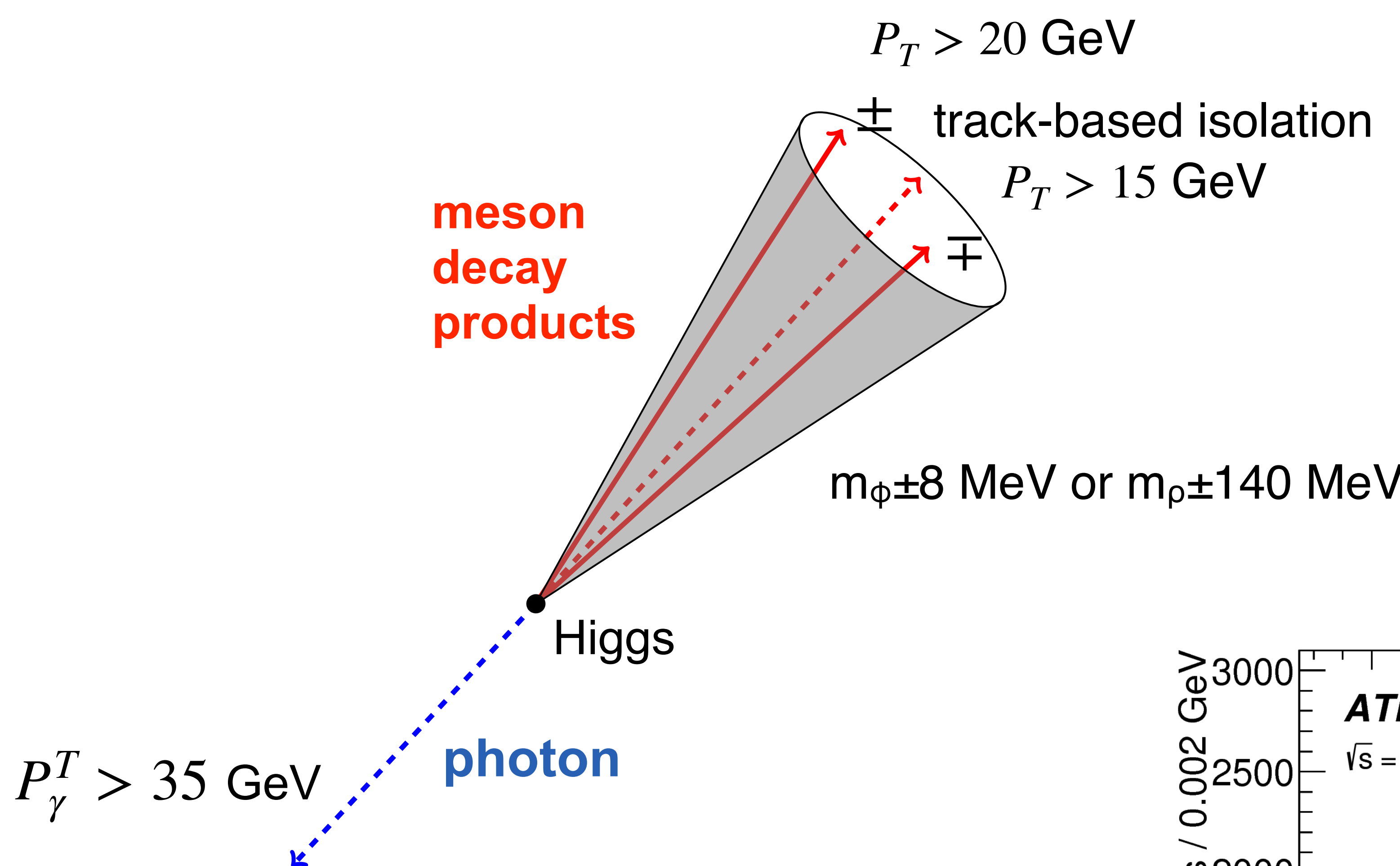
Event Selection



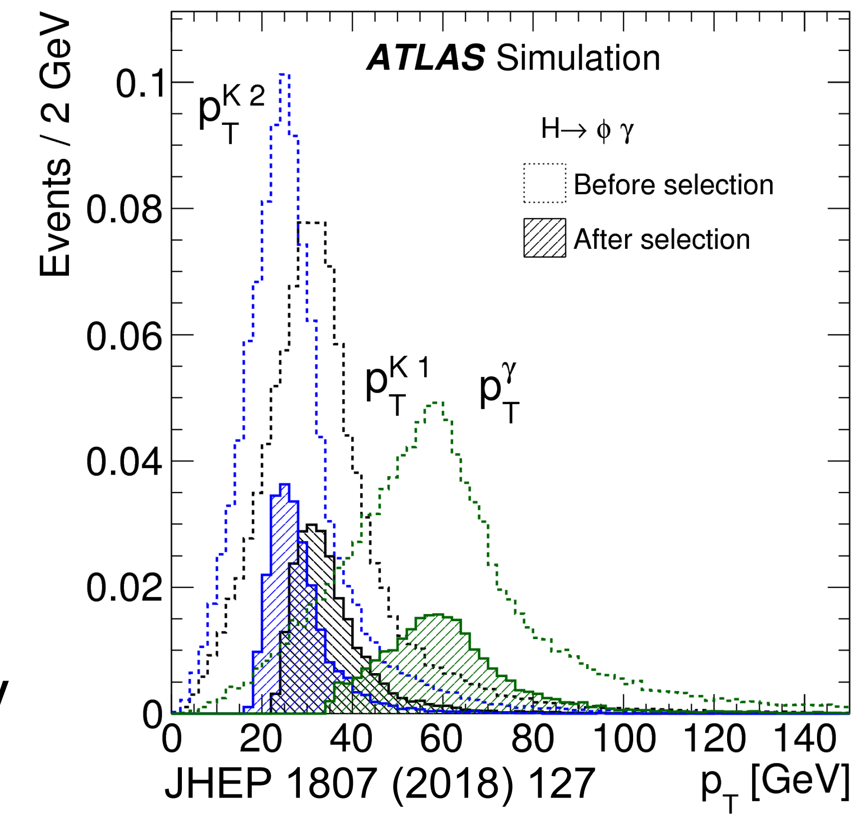
“Tight” identification criteria
Isolated (calorimeter- and track-based)

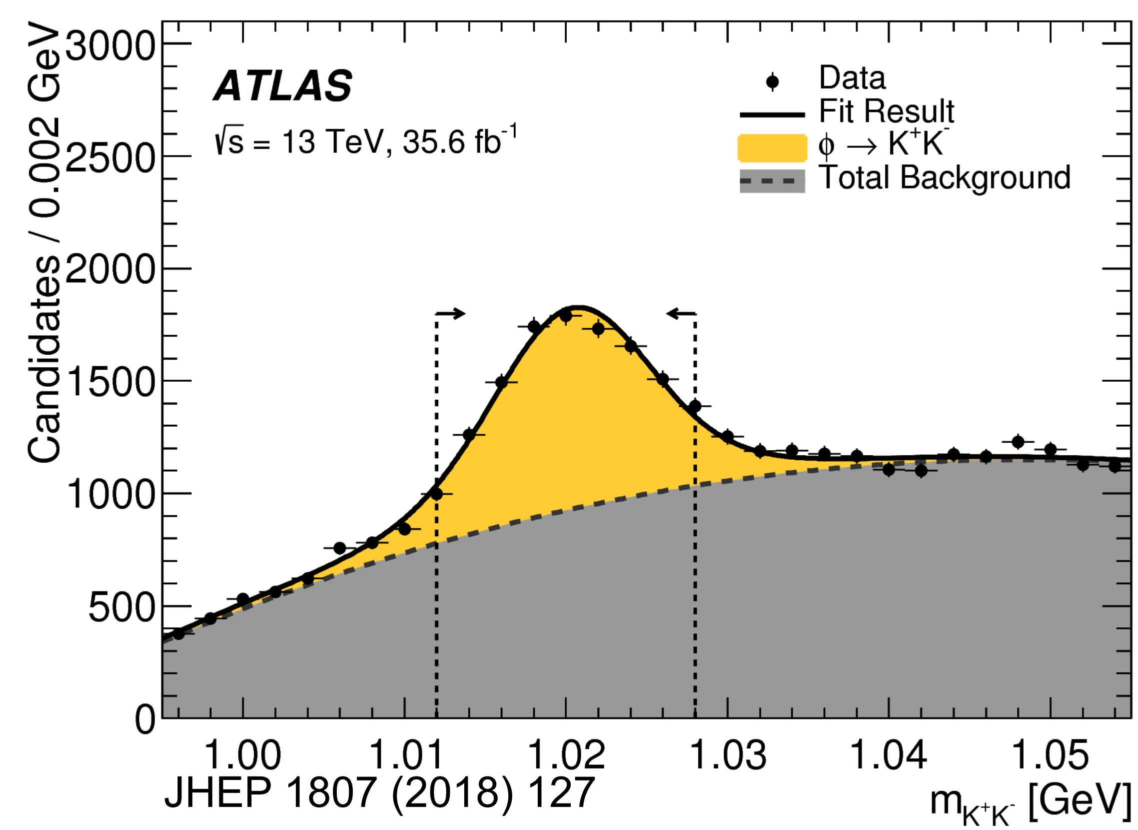


Event Selection



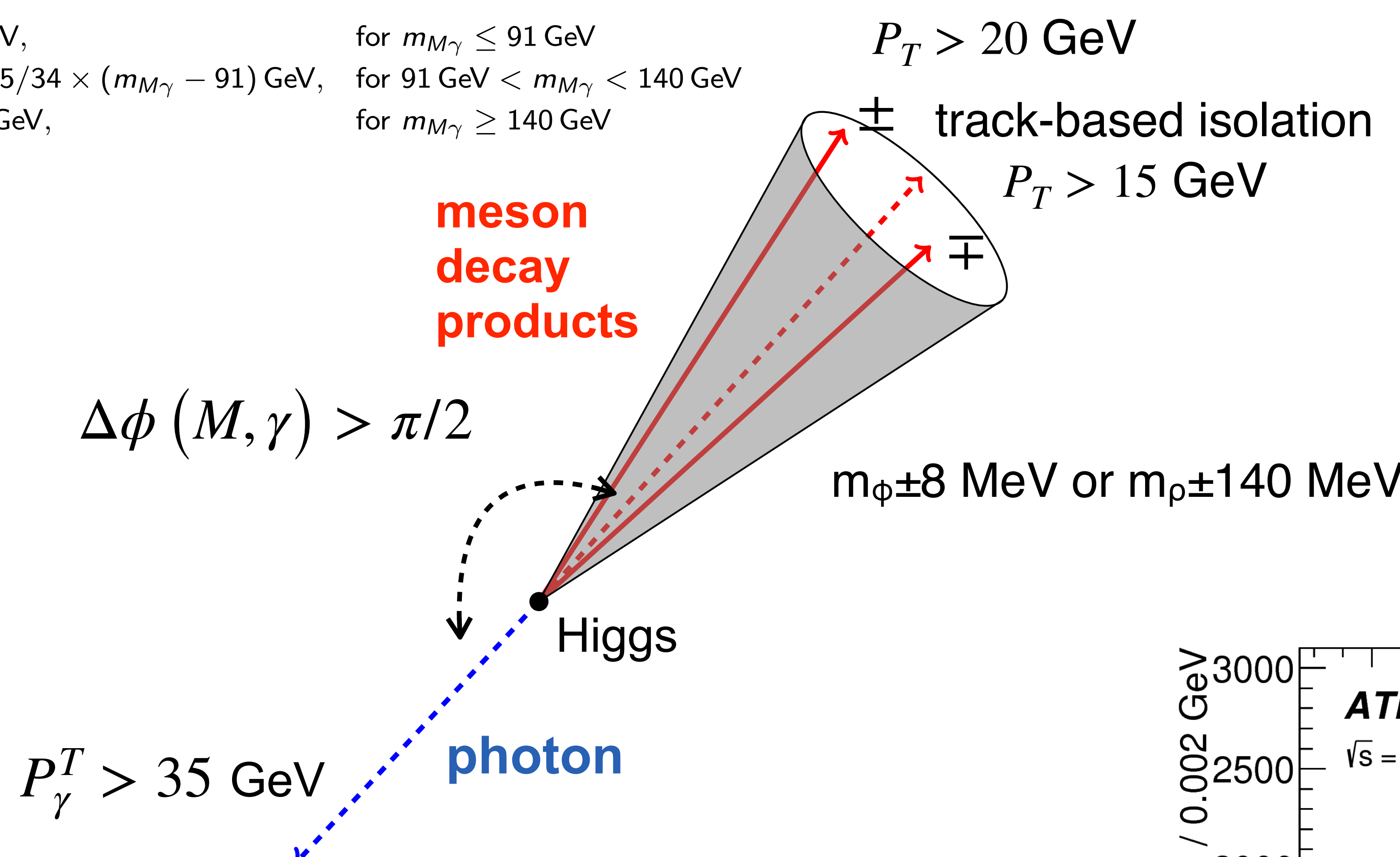
“Tight” identification criteria
Isolated (calorimeter- and track-based)



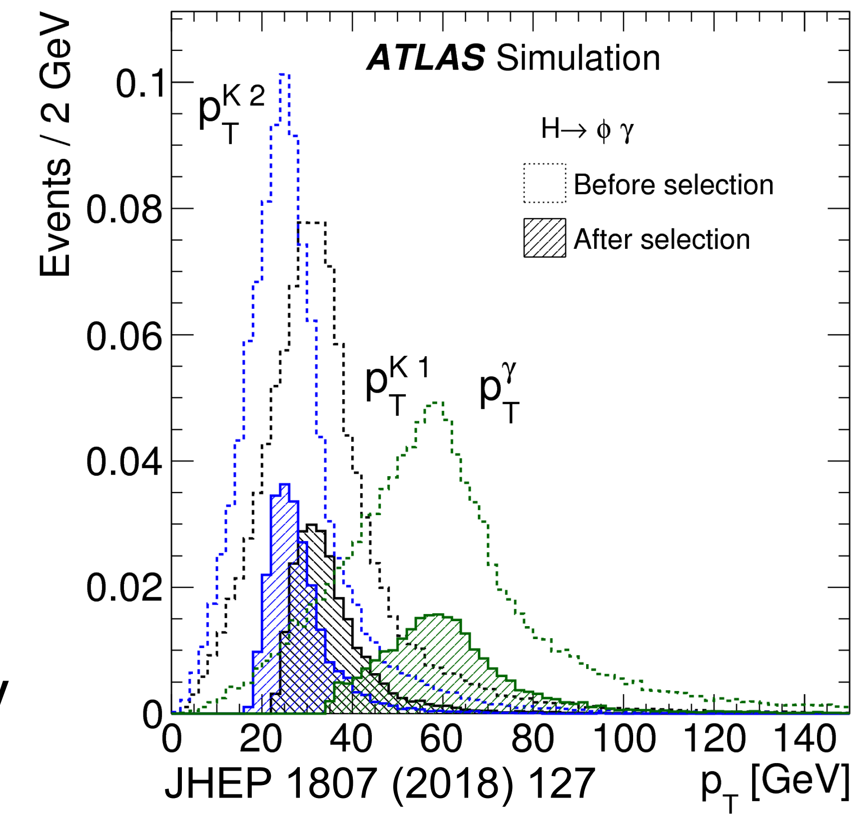


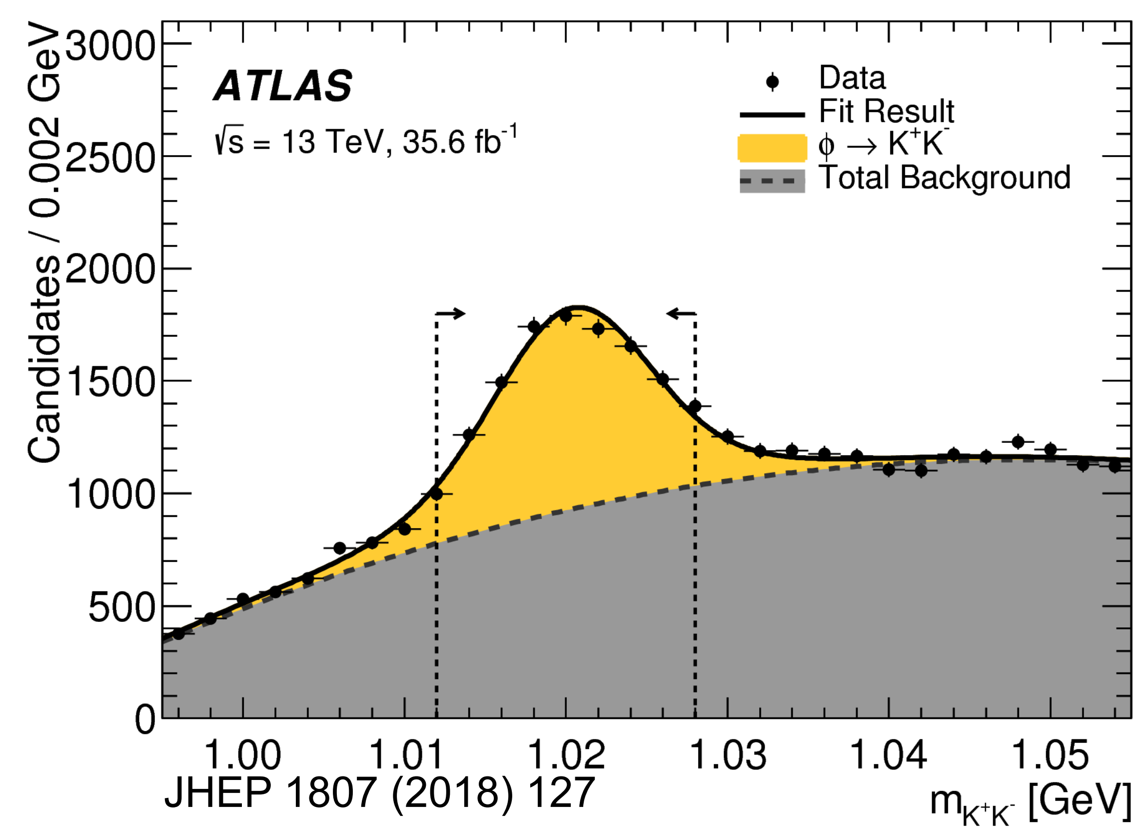
Event Selection

$$p_{\mathrm{T}}^{M} > \begin{cases} 40\ \text{GeV}, & \text{for } m_{M\gamma} \le 91\ \text{GeV} \\ 40 + 5/34 \times (m_{M\gamma} - 91)\ \text{GeV}, & \text{for } 91\ \text{GeV} < m_{M\gamma} < 140\ \text{GeV} \\ 47.2\ \text{GeV}, & \text{for } m_{M\gamma} \ge 140\ \text{GeV} \end{cases}$$



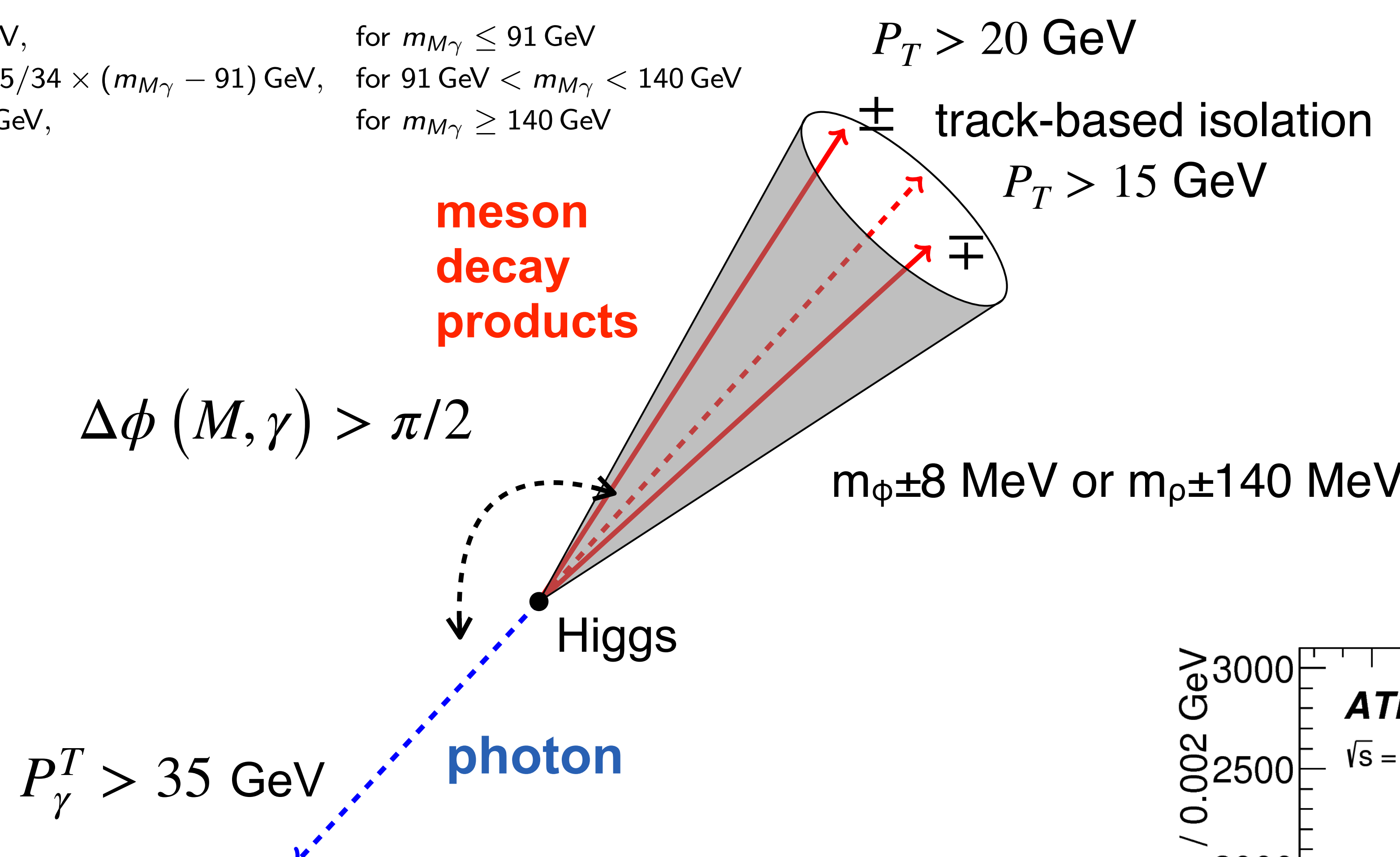
“Tight” identification criteria
Isolated (calorimeter- and track-based)



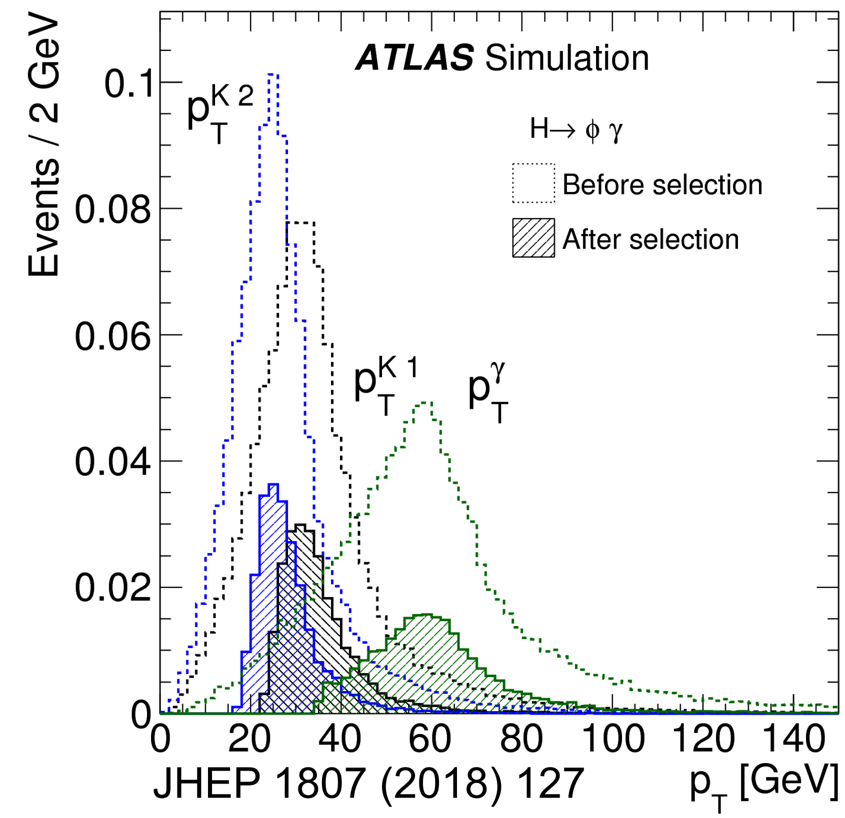


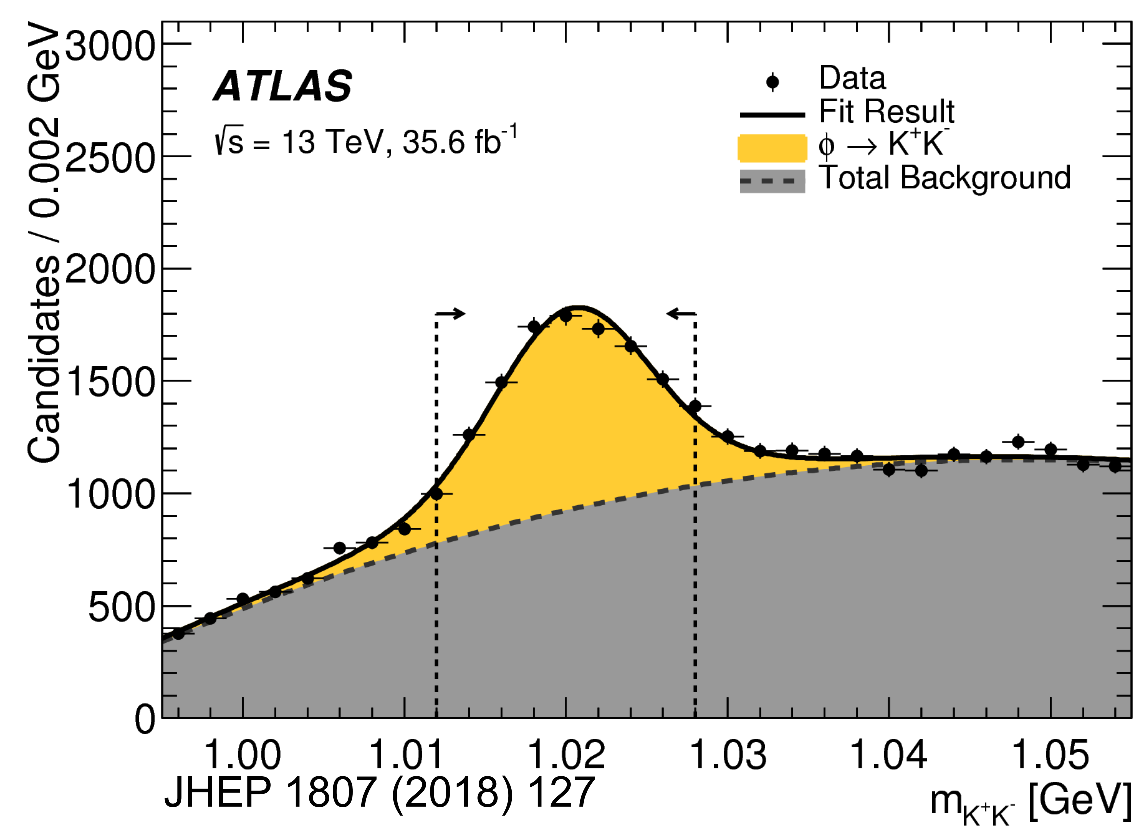
Event Selection

$$p_{\mathrm{T}}^{M} > \begin{cases} 40\ \text{GeV}, & \text{for } m_{M\gamma} \leq 91\ \text{GeV} \\ 40 + 5/34 \times (m_{M\gamma} - 91)\ \text{GeV}, & \text{for } 91\ \text{GeV} < m_{M\gamma} < 140\ \text{GeV} \\ 47.2\ \text{GeV}, & \text{for } m_{M\gamma} \geq 140\ \text{GeV} \end{cases}$$



“Tight” identification criteria
Isolated (calorimeter- and track-based)





- **“Inclusive” backgrounds**
  - γ+jet, di-jet with jet “seen” as γ

Background Model

- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analyses already!**

[Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]

Background Model

- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analyses already!**

  [Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]

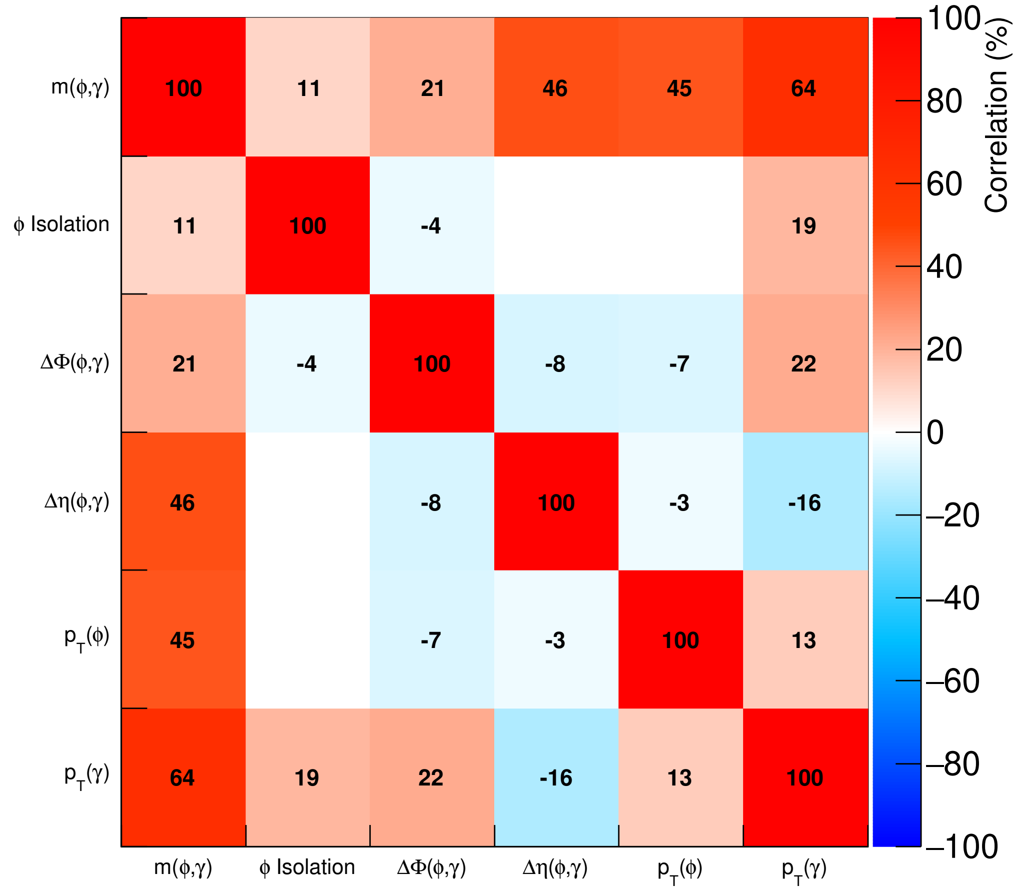
Example application on γ+jet MC sample

Background Model

- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analyses already!**

[Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]

Example application on γ+jet MC sample



γ+jet MC

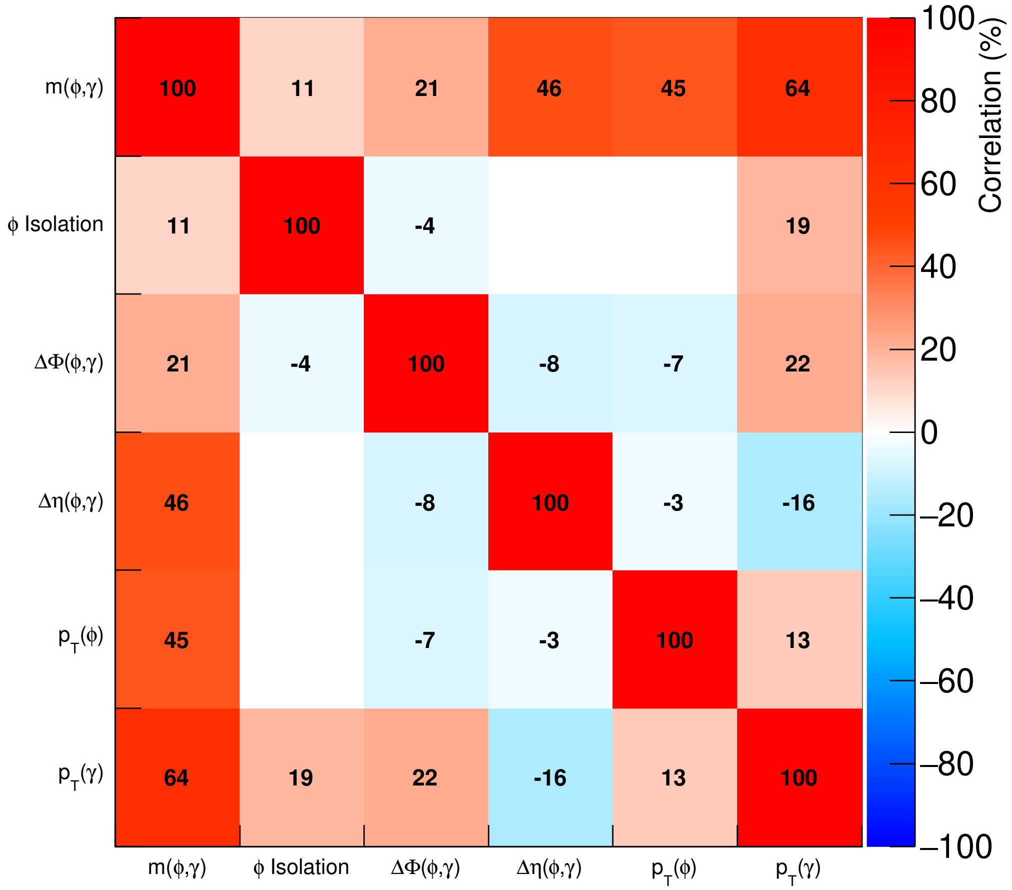
arXiv:2112.00650

Background Model

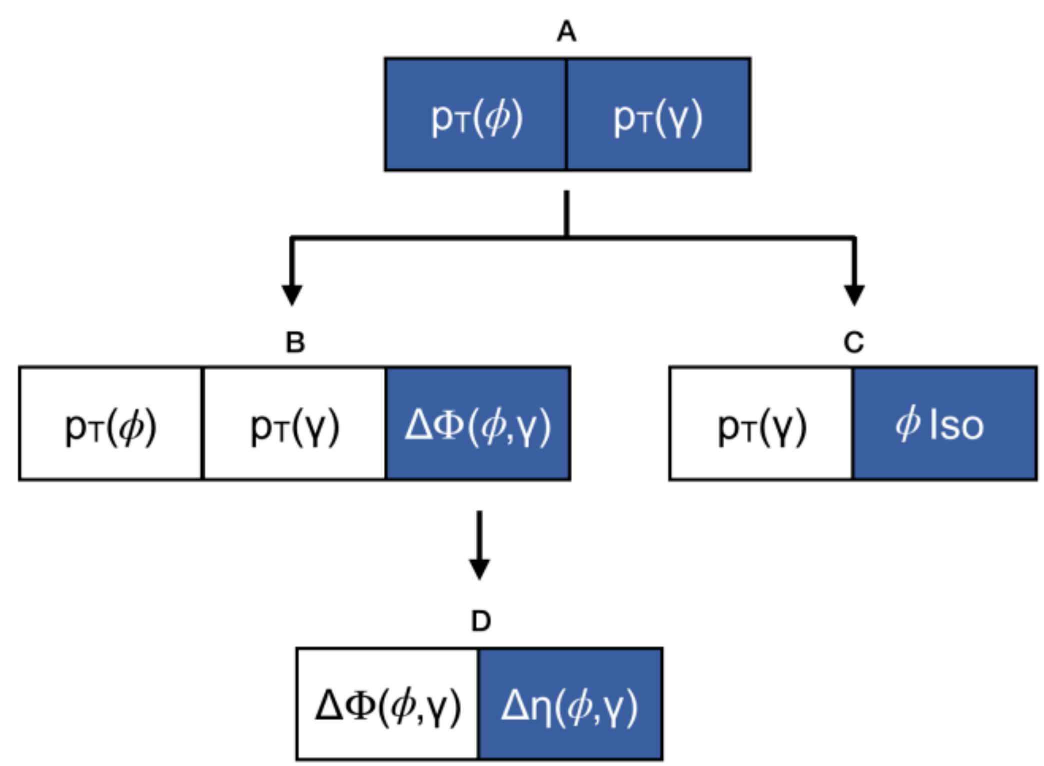
- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analyses already!**

[Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]

Example application on γ+jet MC sample



γ+jet MC



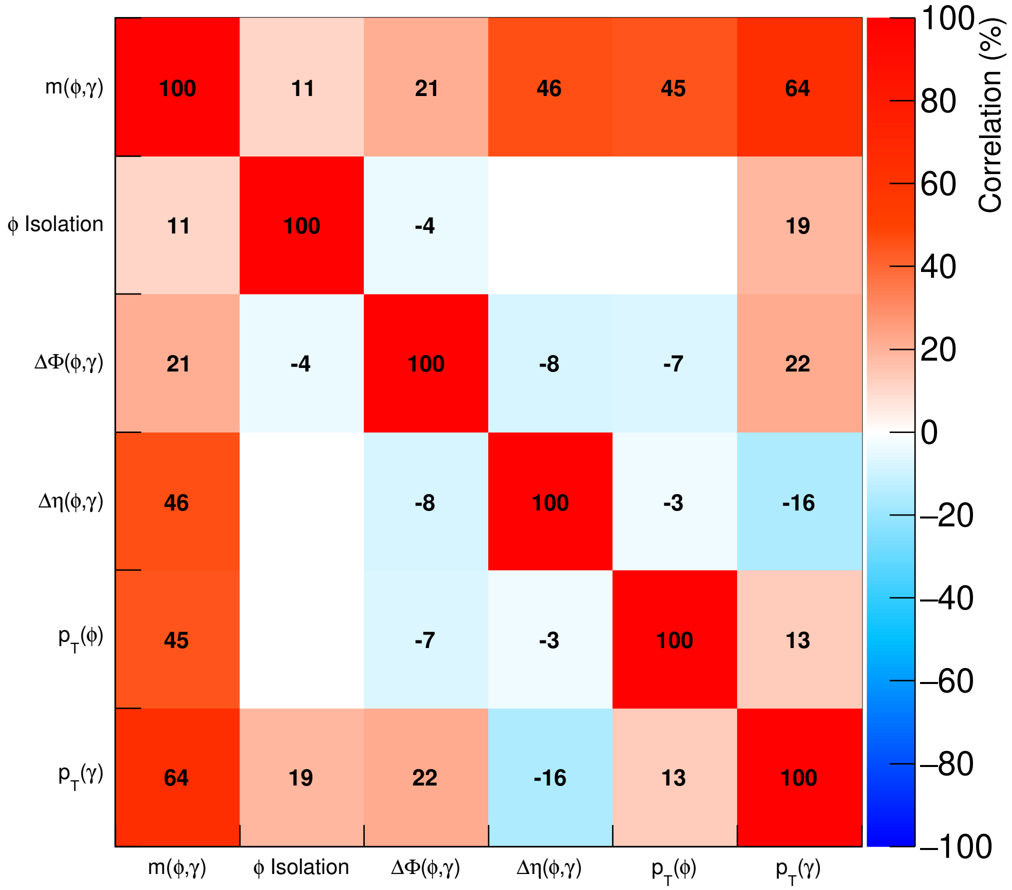
arXiv:2112.00650

Background Model

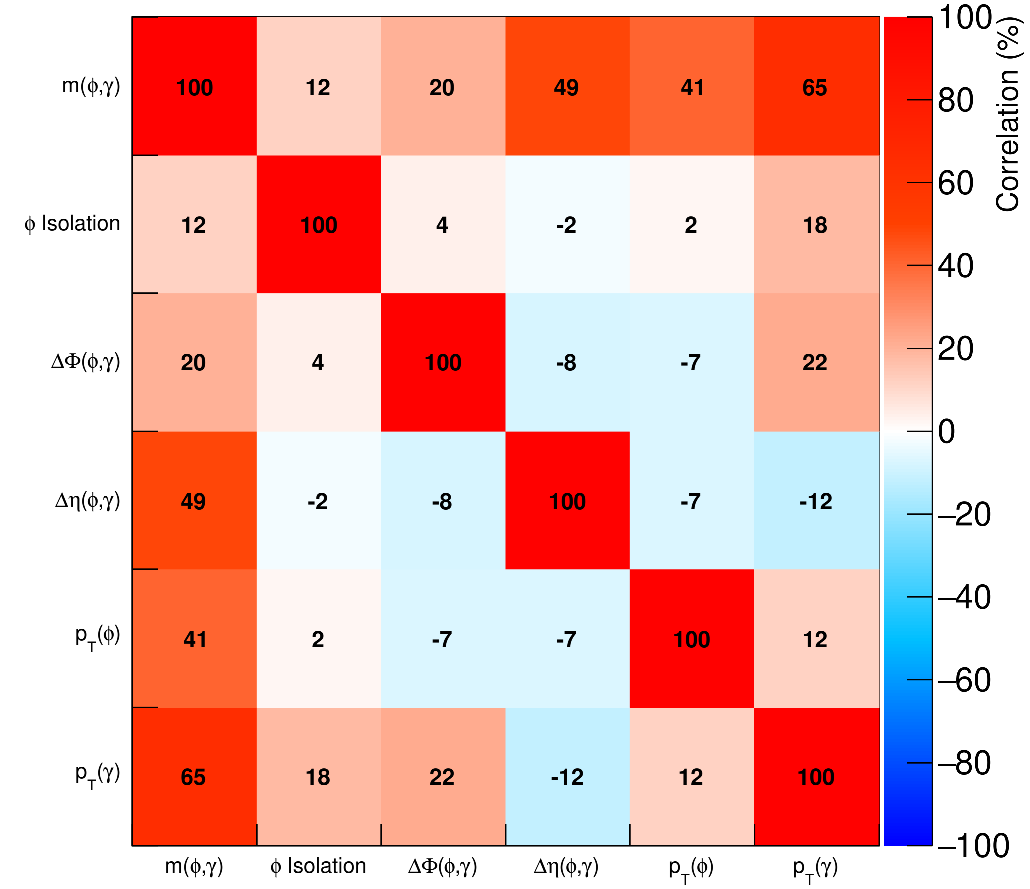
- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analyses already!**

[Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]

Example application on γ+jet MC sample



γ+jet MC

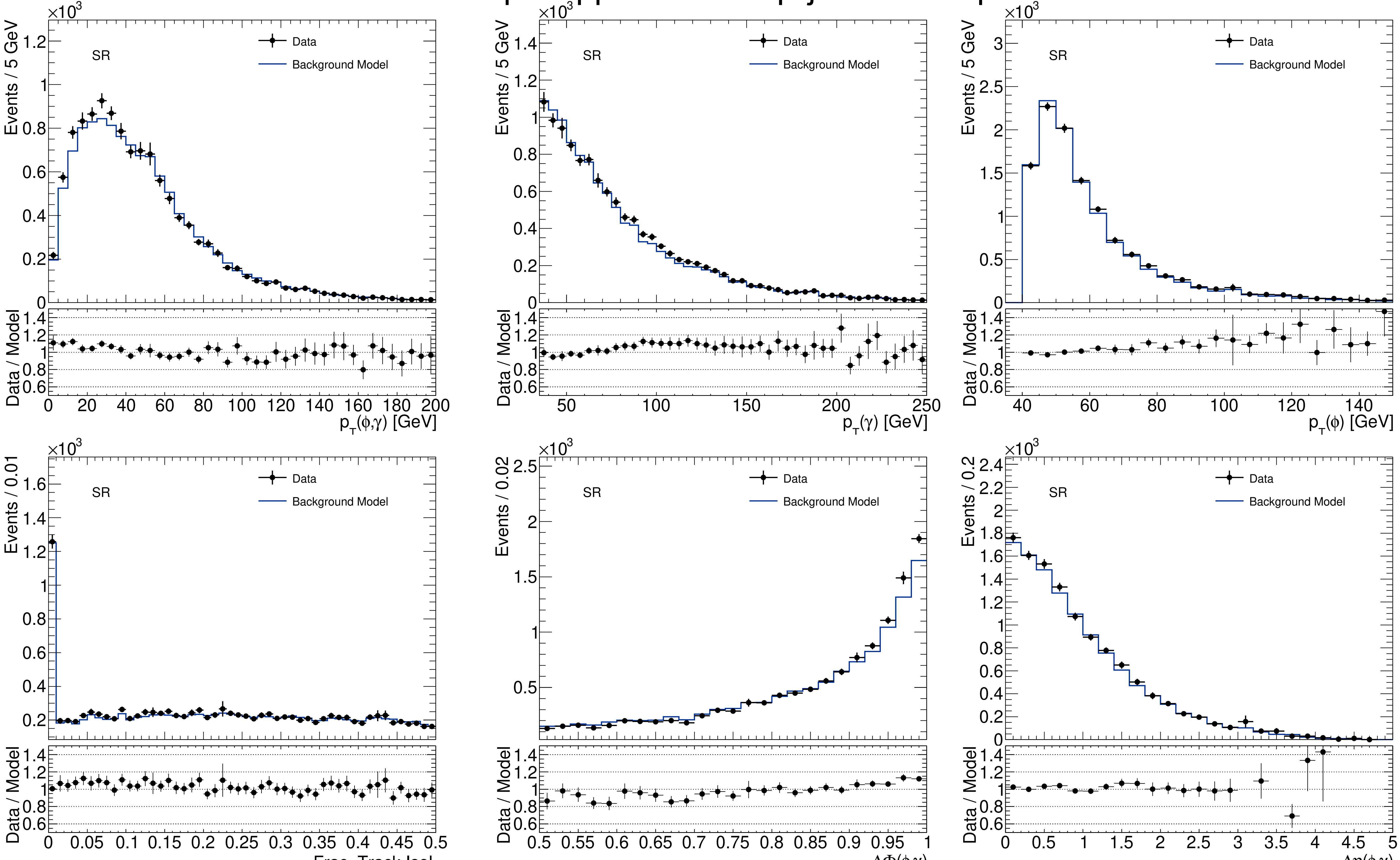


Model

arXiv:2112.00650

Background Model

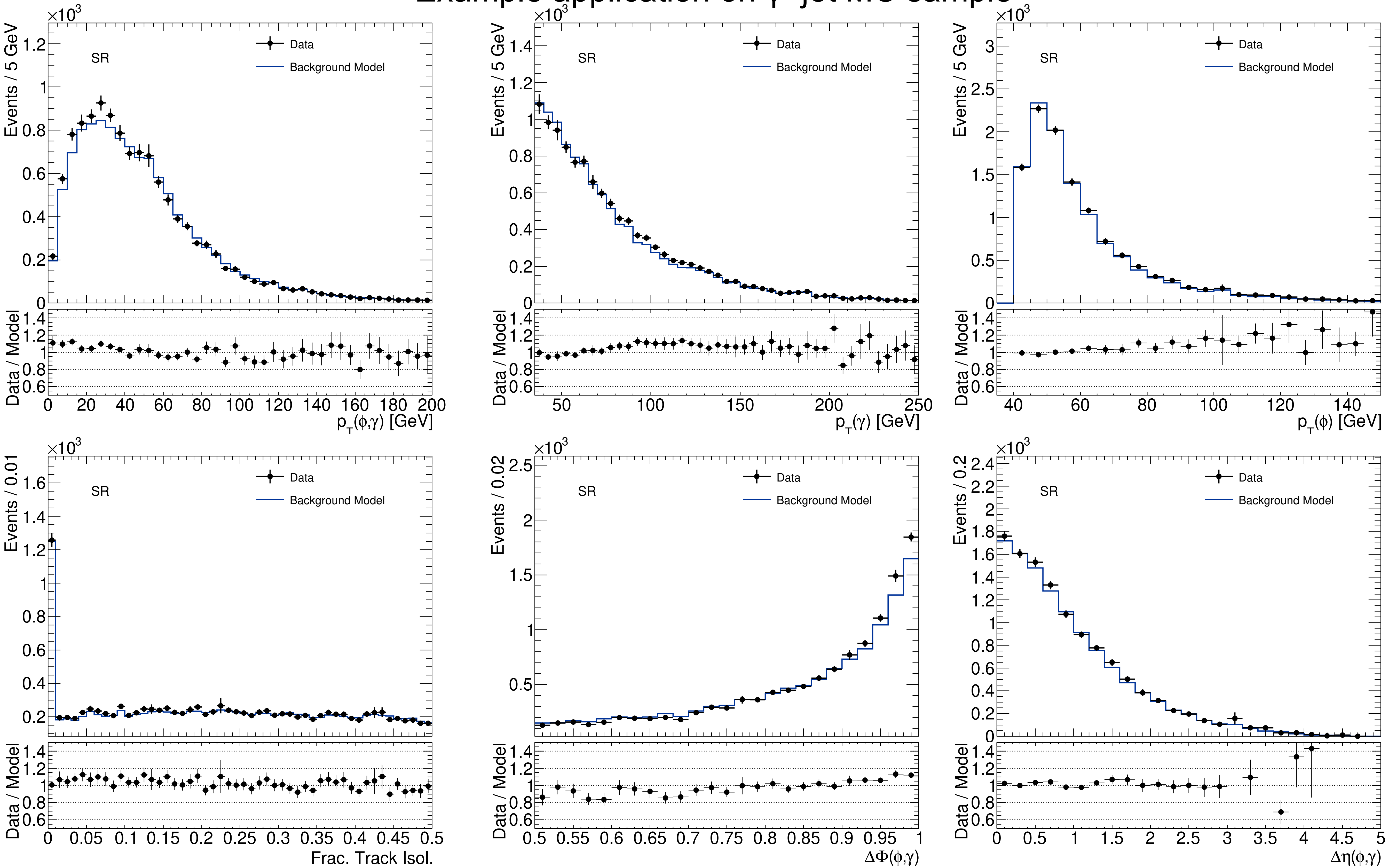
Example application on γ+jet MC sample



Background Model

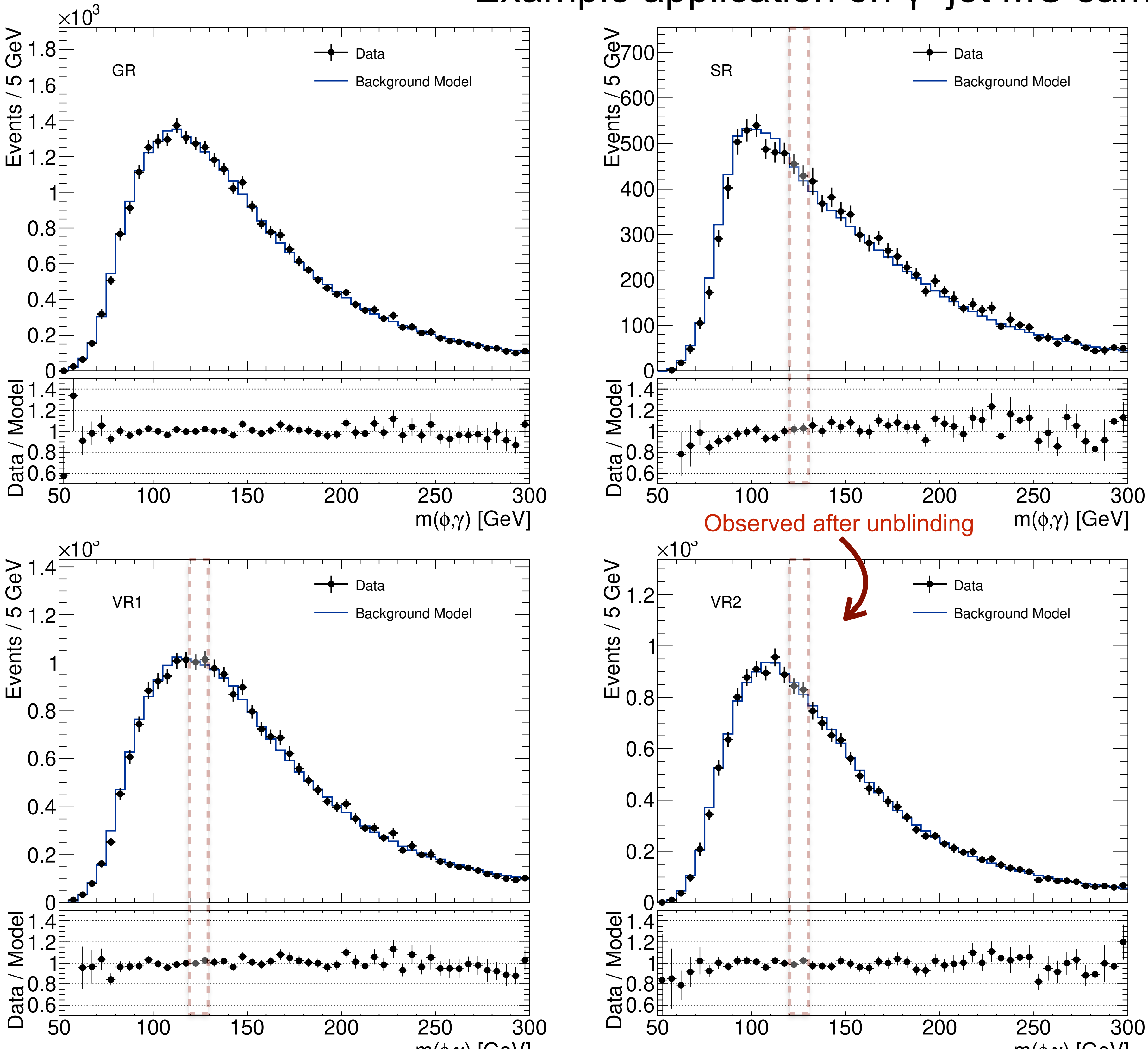
Example application on γ+jet MC sample

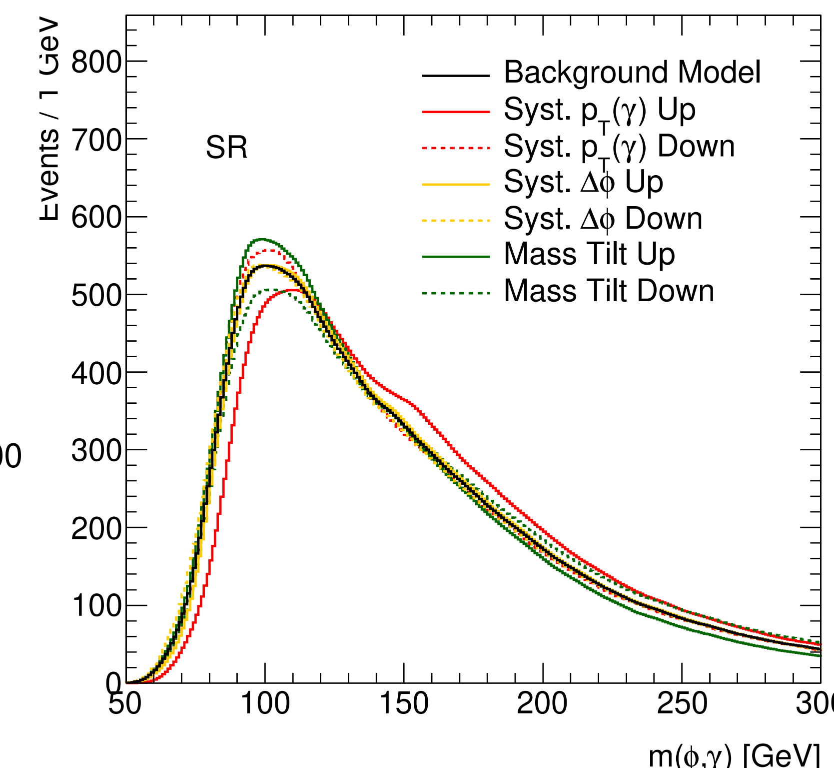
arXiv:2112.00650



Background Model

Example application on γ+jet MC sample





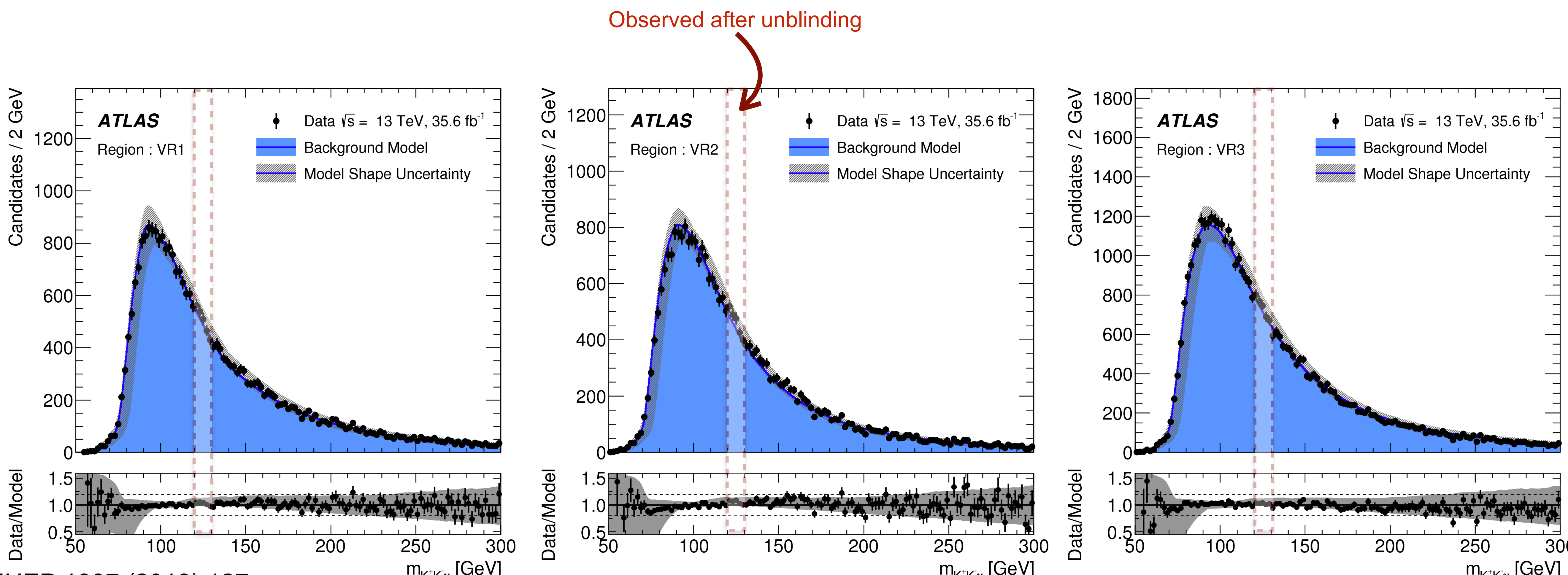
Shape variations

- Modifying sampling distributions
- Overall transformations of signal shape

Background Model

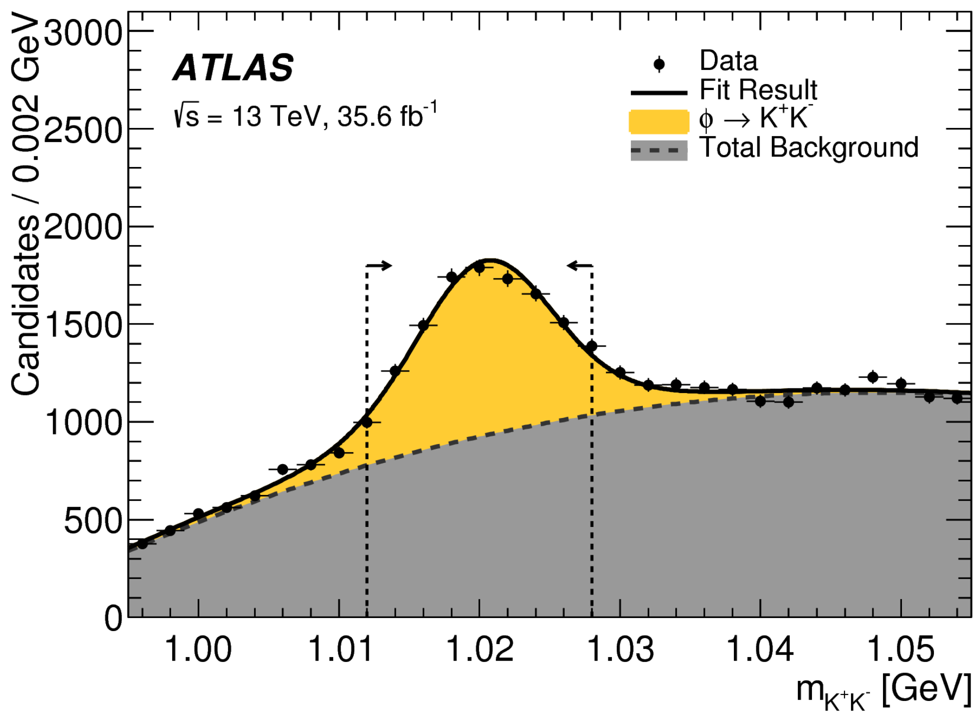
- **Non-parametric data-driven** background model based on **Ancestral Sampling**
  - Obtain loose sample of candidates
  - Model kinematic and isolation distributions
    - Conditional PDFs modelled using histograms
  - Generate “pseudo”-background events and apply event selection
- **Used in several analysis already!**

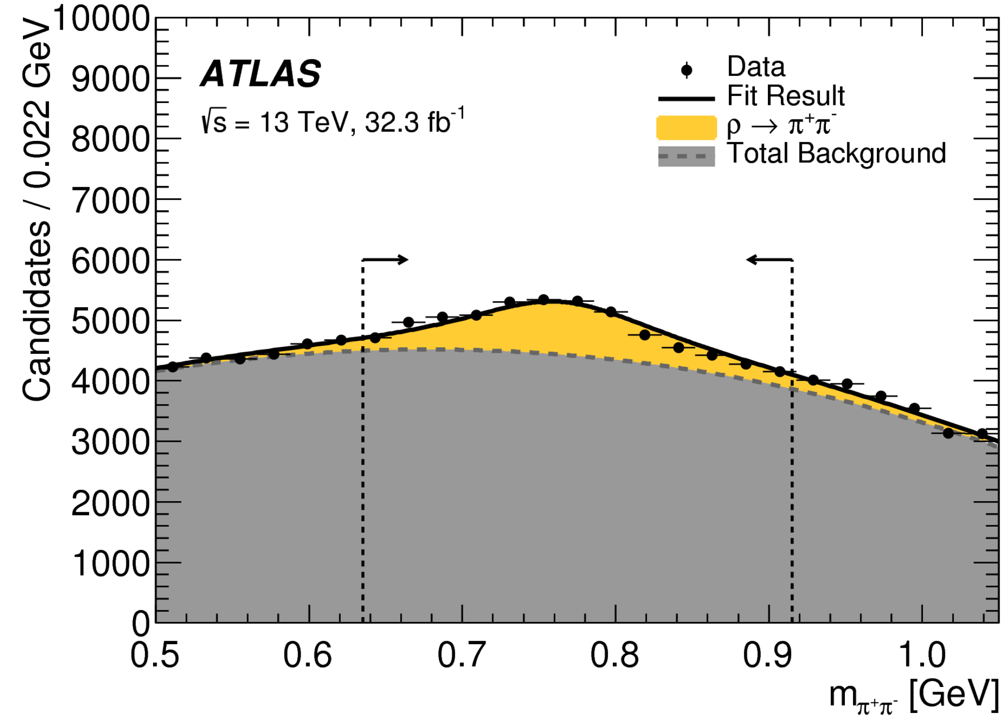
[Phys. Rev. Lett. 114 (2015) 121801, Phys. Rev. Lett. 117, 111802 (2016), JHEP 07 (2018) 127, Phys. Lett. B 786 (2018) 134]



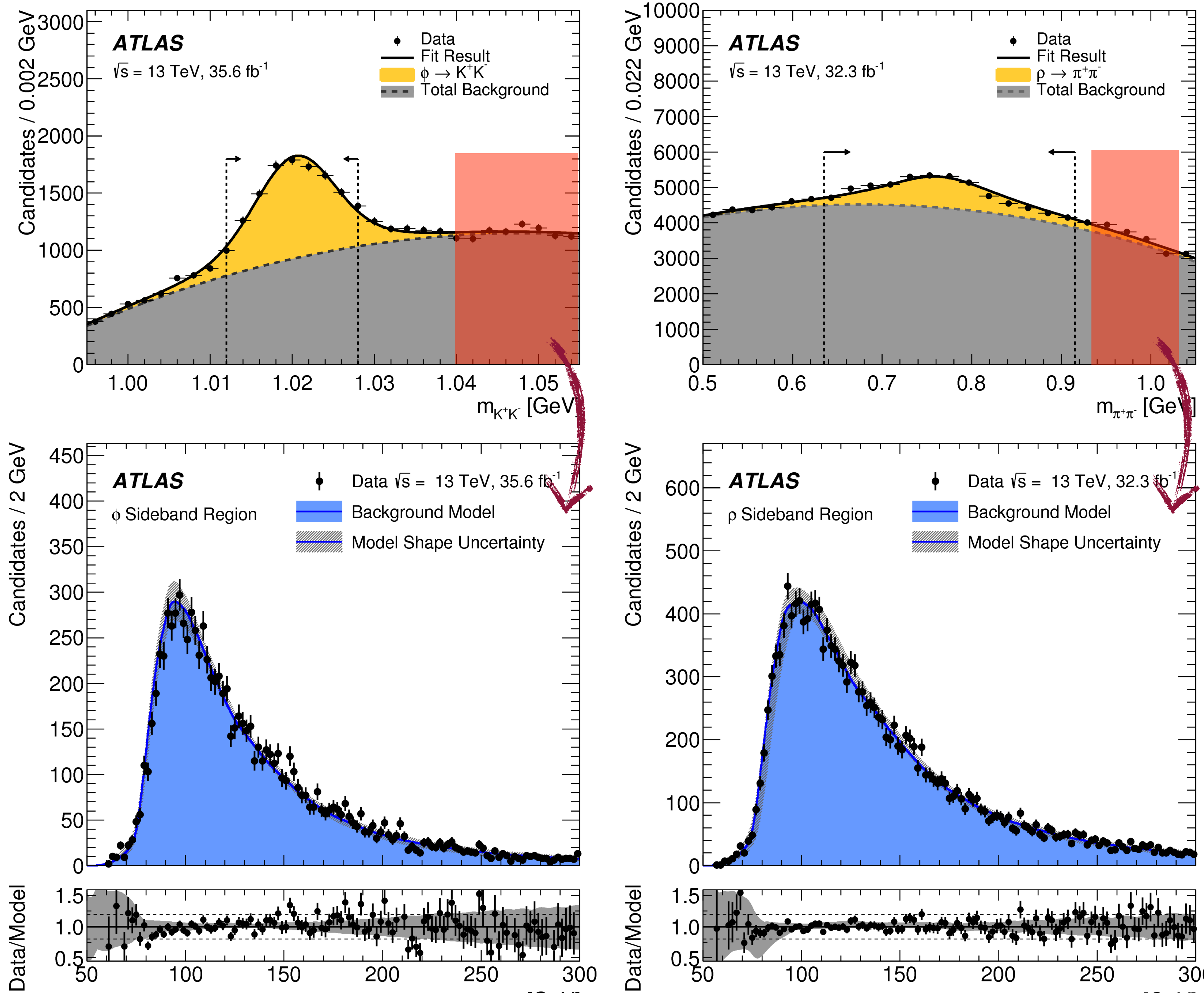
JHEP 1807 (2018) 127

Background Validation



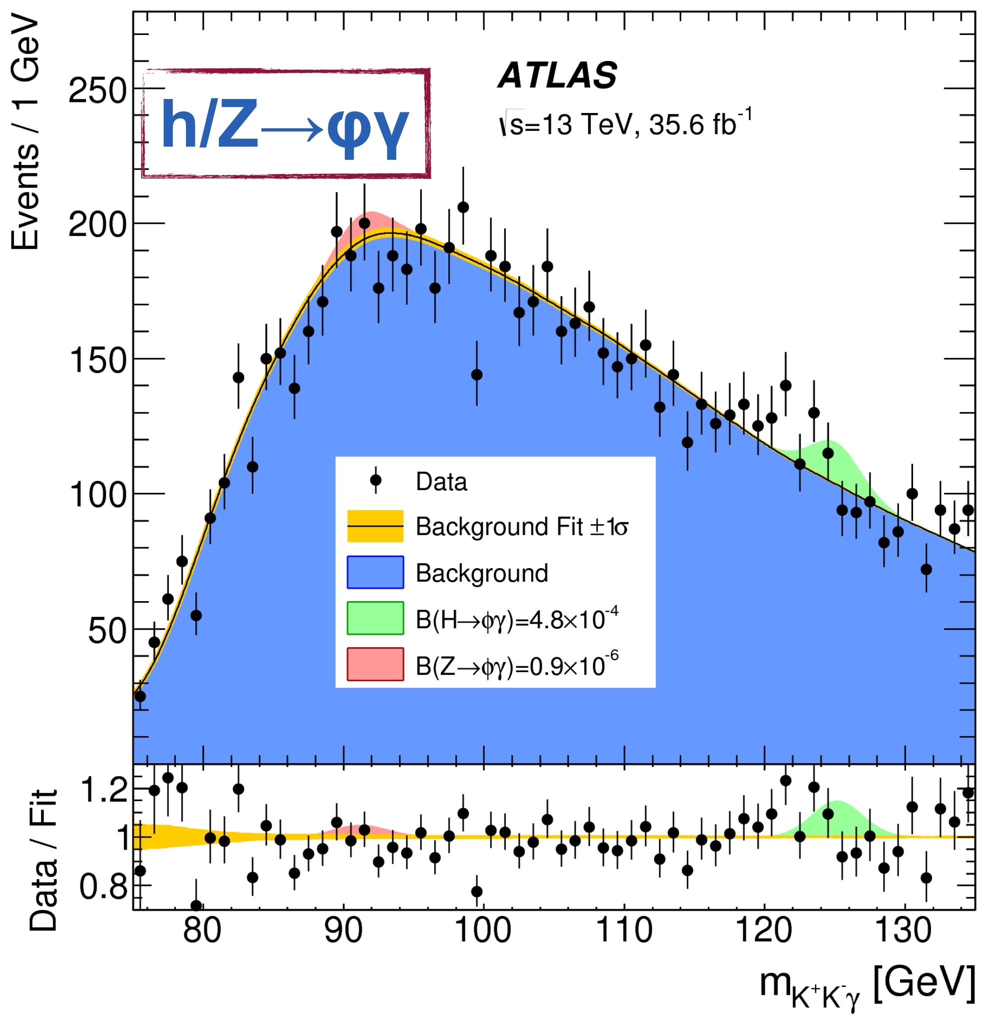


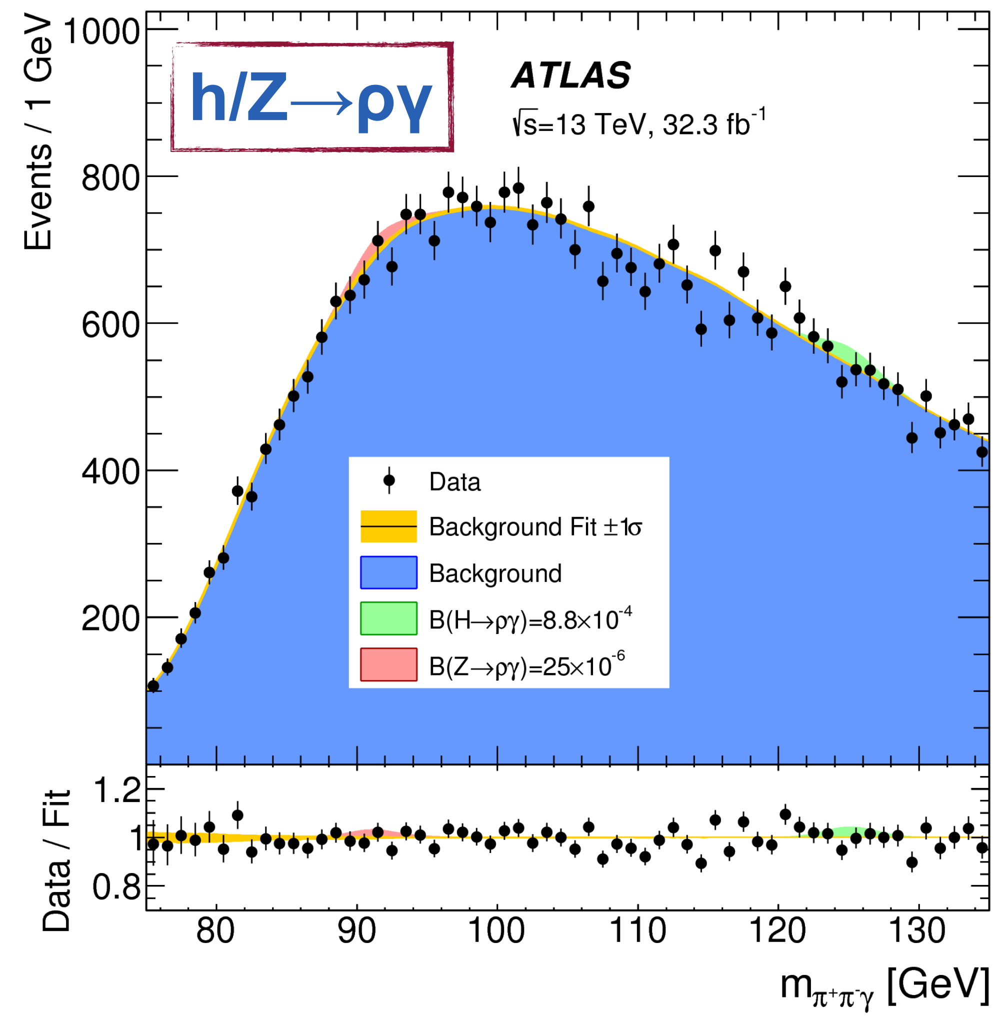
Background Validation



JHEP 1807 (2018) 127 JHEP 1807 (2018) 127

h/Z→φγ/ργ: Results





Final discriminant: $m_{KK\gamma}$ and $m_{\pi\pi\gamma}$

No significant signal observed

Branching Fraction Limit (95% CL)	Expected	Observed
$\mathcal{B}(H \to \phi\gamma)\ [10^{-4}]$	$4.2^{+1.8}_{-1.2}$	4.8
$\mathcal{B}(Z \to \phi\gamma)\ [10^{-6}]$	$1.3^{+0.6}_{-0.4}$	0.9
$\mathcal{B}(H \to \rho\gamma)\ [10^{-4}]$	$8.4^{+4.1}_{-2.4}$	8.8
$\mathcal{B}(Z \to \rho\gamma)\ [10^{-6}]$	33^{+13}_{-9}	25

JHEP 1807 (2018) 127

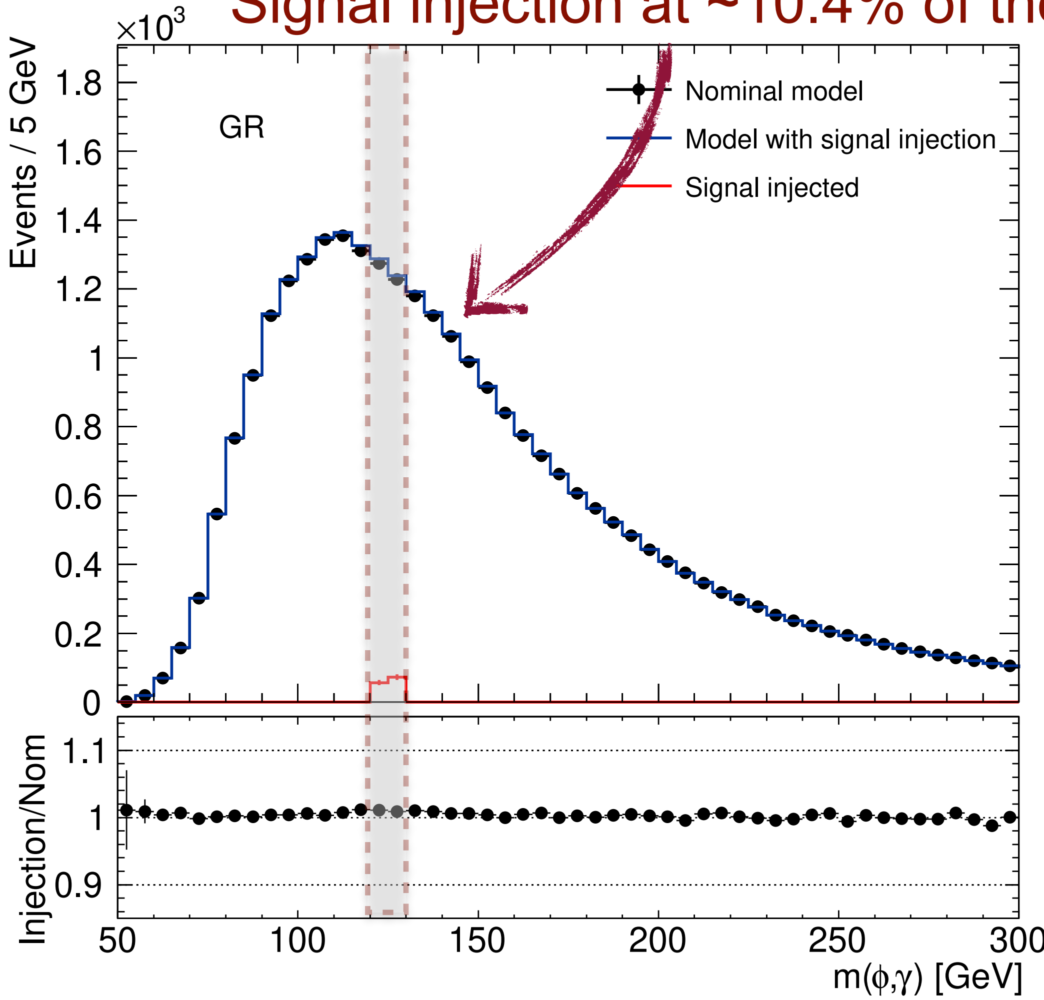
Model Robustness

- **Model describes main features of background**
  - Robust under signal contamination
  - Resonant backgrounds need to be considered separately

Model Robustness

- **Model describes main features of background**
  - Robust under signal contamination
  - Resonant backgrounds need to be considered separately

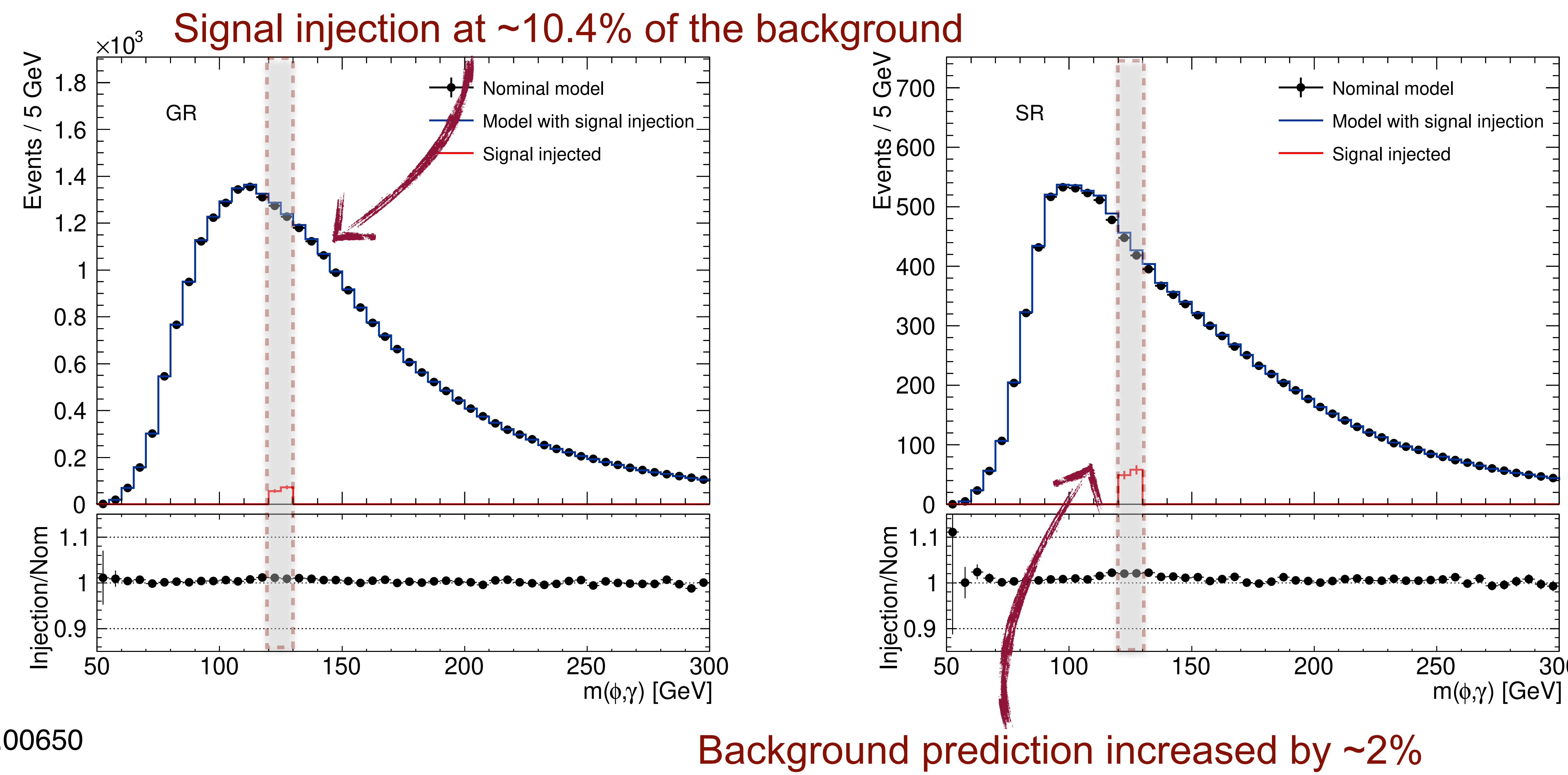
Signal injection at ~10.4% of the background



arXiv:2112.00650

Model Robustness

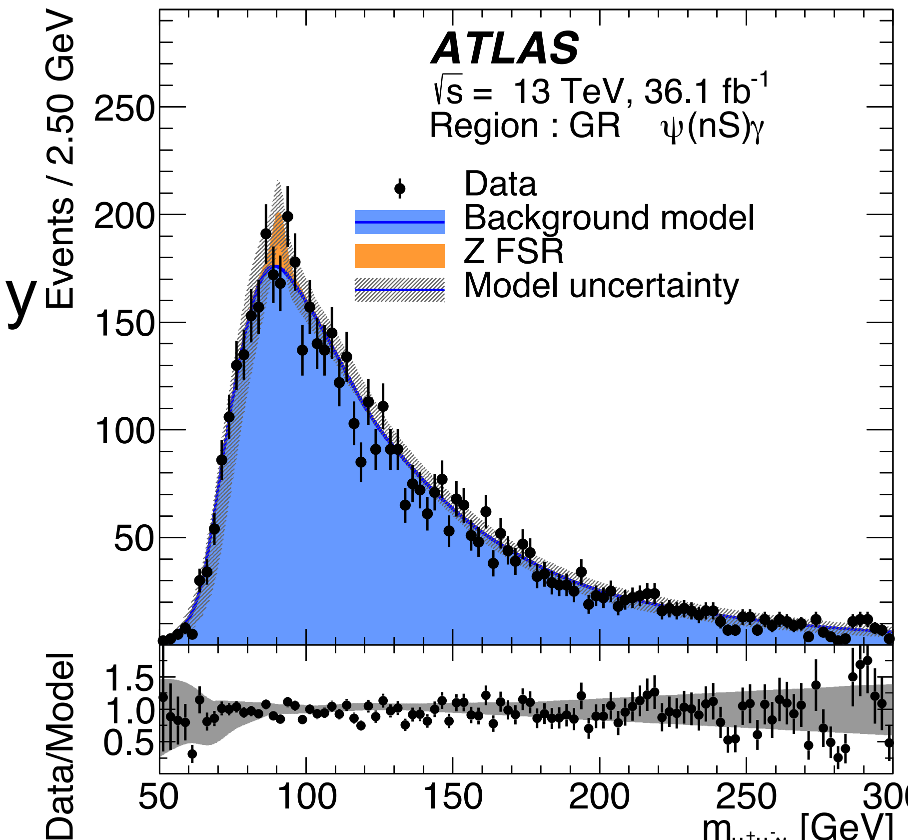
- **Model describes main features of background**
  - Robust under signal contamination
  - Resonant backgrounds need to be considered separately



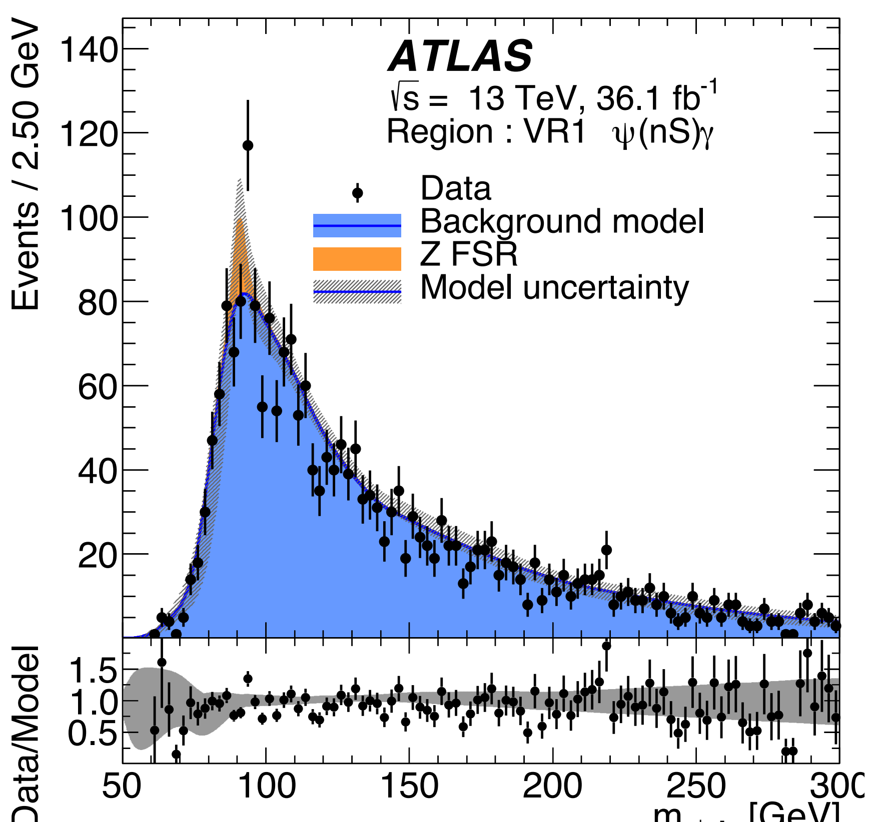
arXiv:2112.00650

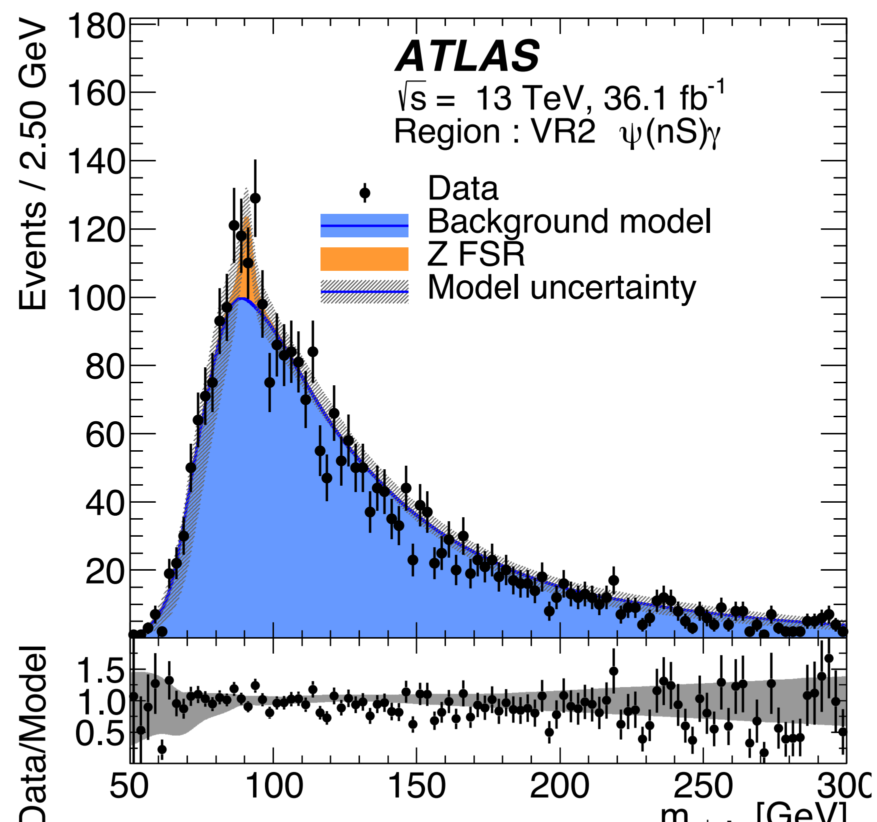
h/Z→Qγ: Resonant Backgrounds

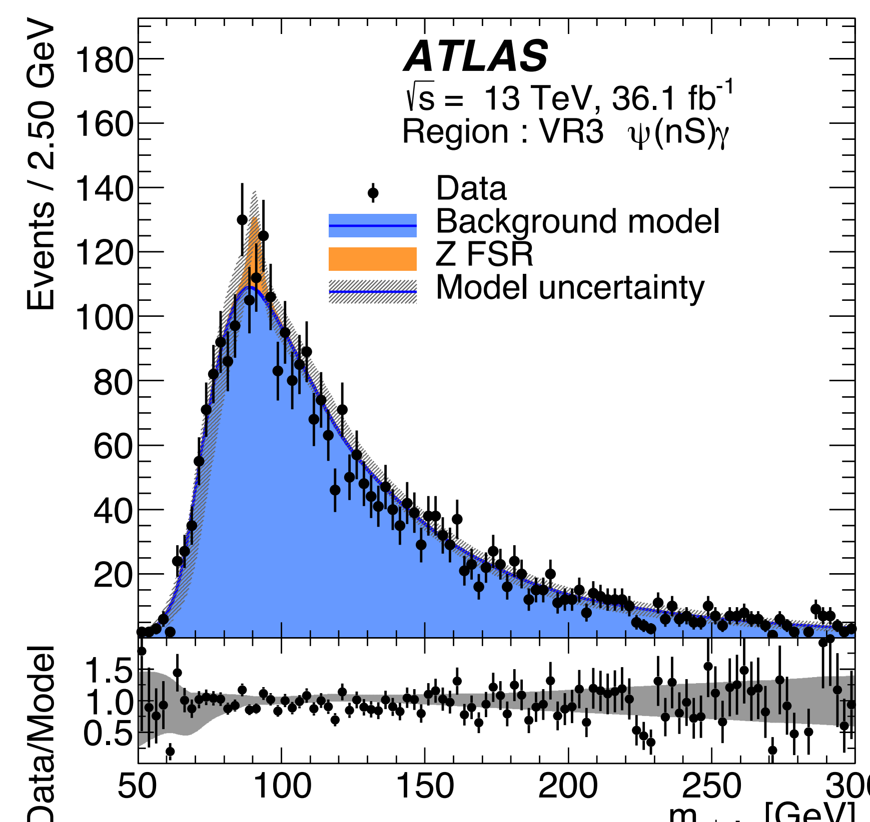
- **Model describes main features of background**
  - Robust under signal contamination
  - Resonant backgrounds need to be considered separately



arXiv:1807.00802

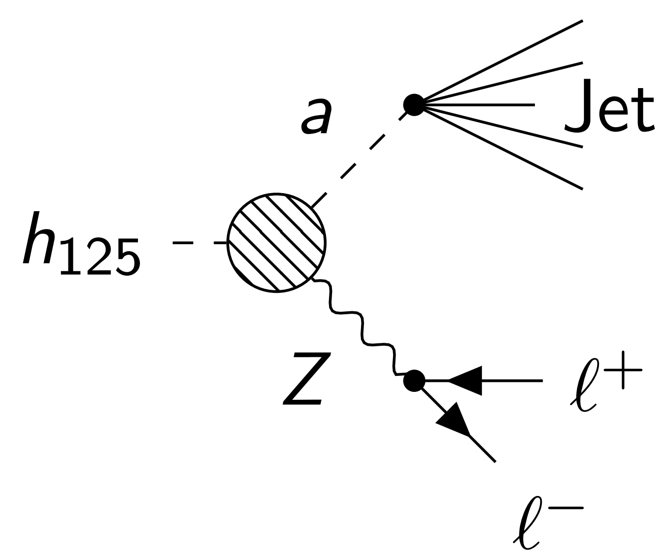


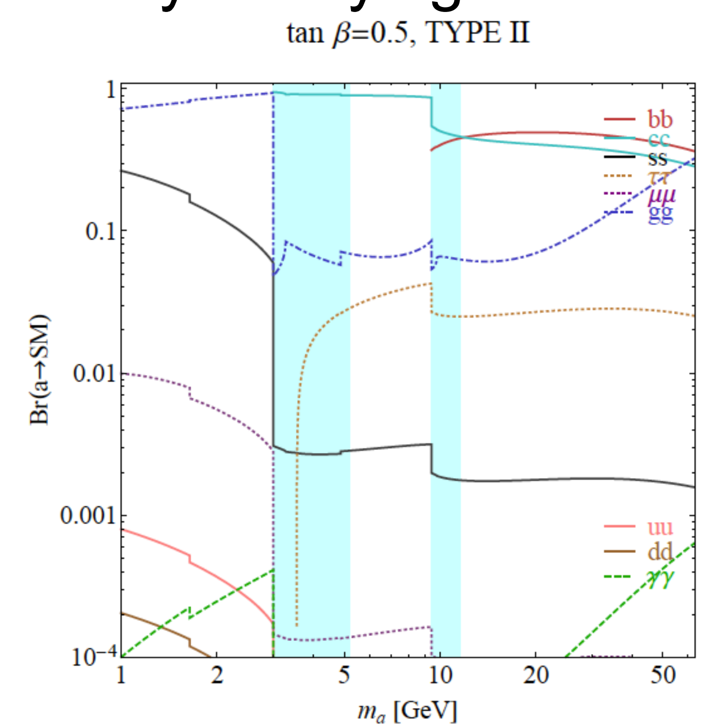




h→Za→ll+jet

Higgs decays to light hadronically decaying scalars





PRD 90 (2014) 7, 075004

h→Za→ll+jet

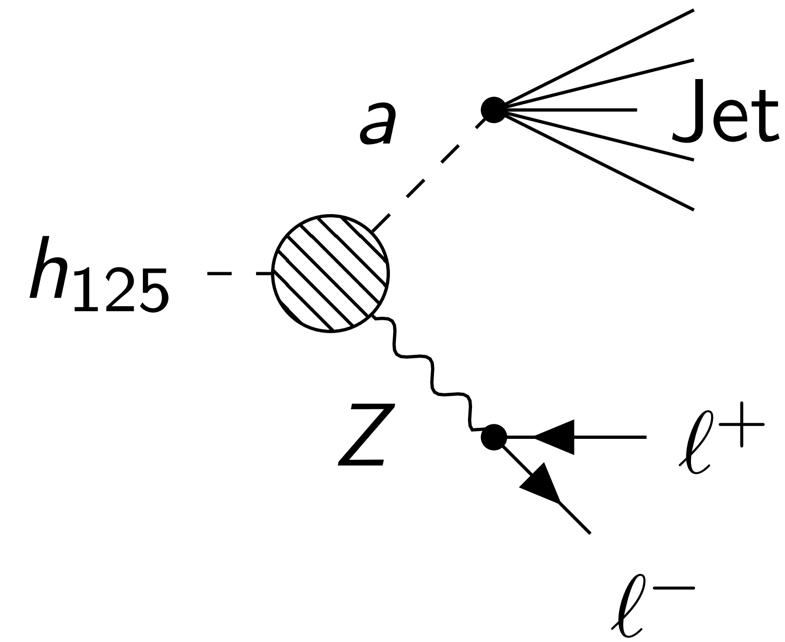
PRL 125 (2020) 22, 221802

Experimental focus mostly on:

- h→aa
- a→down-type fermions

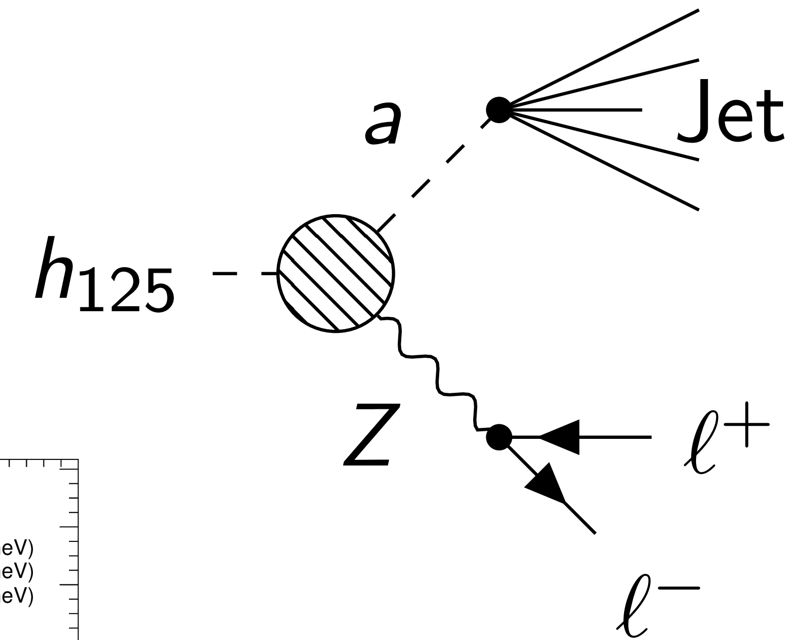
New search: h→Za with a→hadrons

- Overwhelming Z + jets background
- a→hadrons reconstruction using sub-structure techniques



h→Za→ll+jet

PRL 125 (2020) 22, 221802

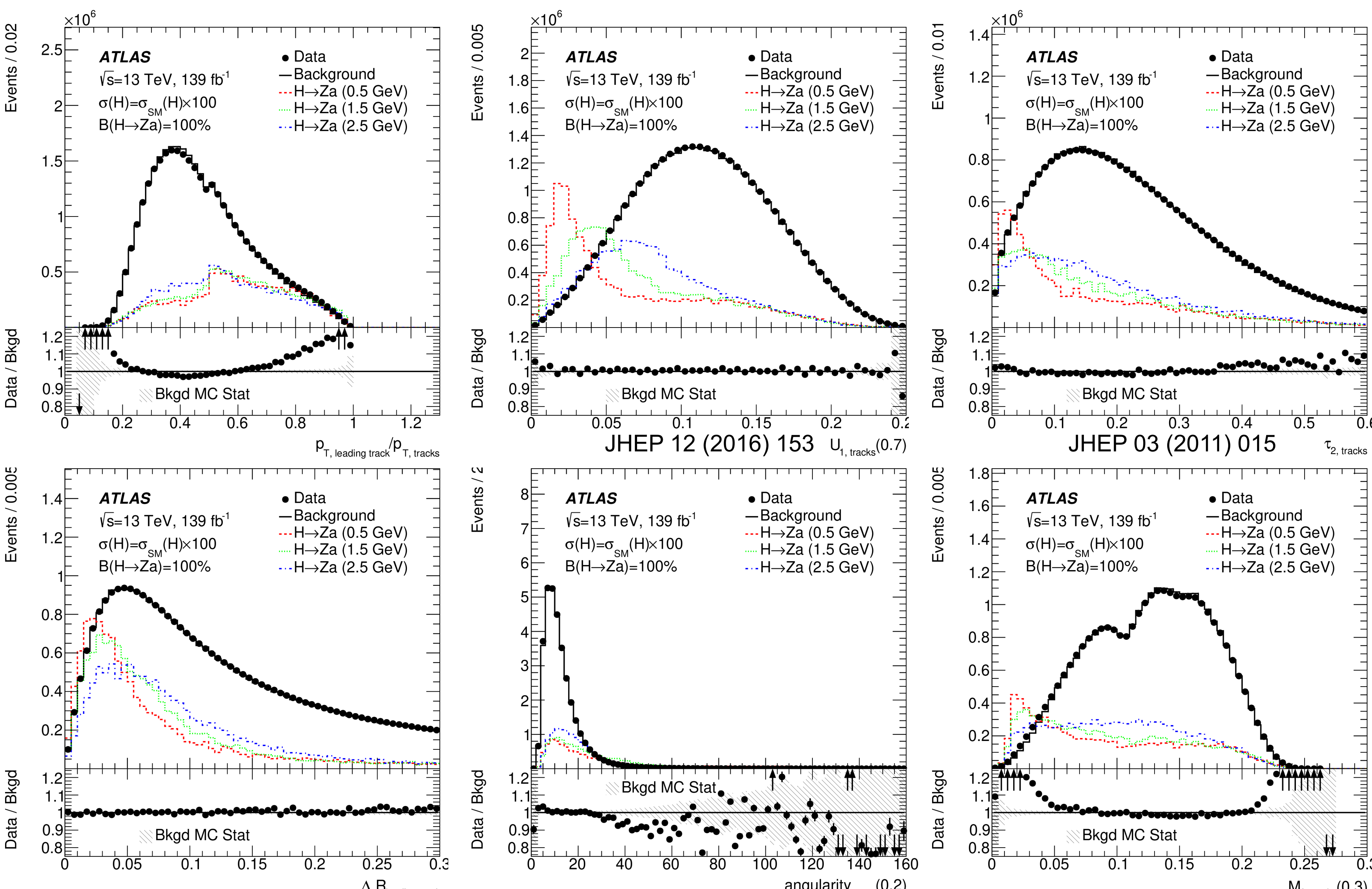


Experimental focus mostly on:

- h→aa
- a→down-type fermions

New search: h→Za with a→hadrons

- Overwhelming Z + jets background
- a→hadrons reconstruction using sub-structure techniques



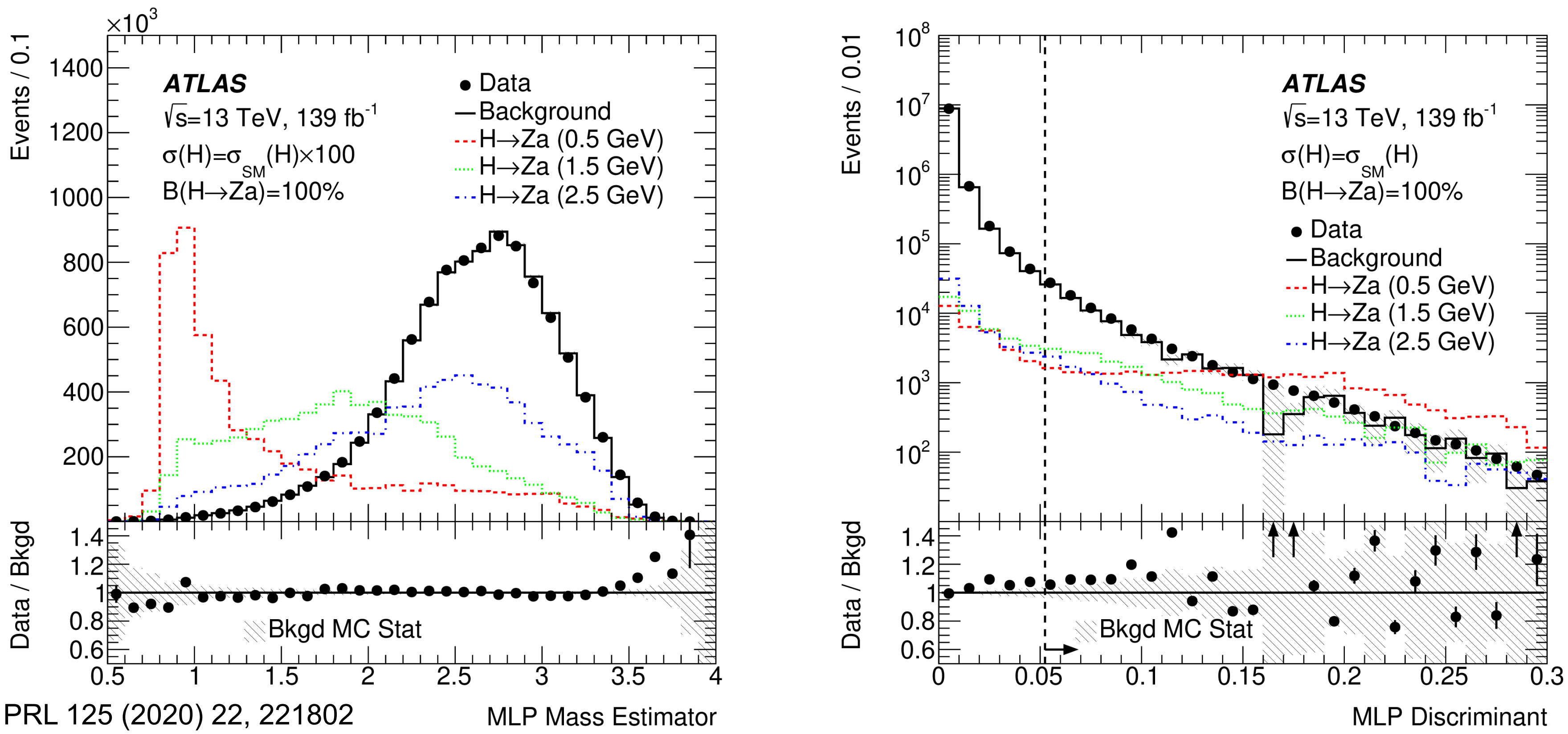
JHEP 12 (2016) 153

JHEP 03 (2011) 015

PRD 79 (2009) 074017

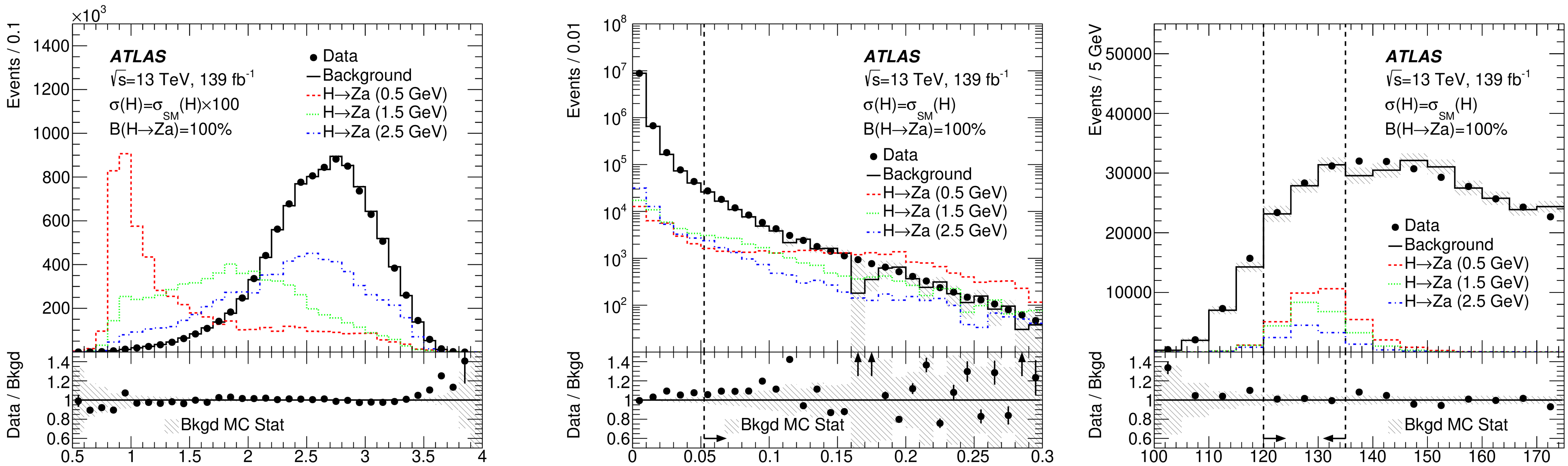
JHEP 12 (2016) 153

h→Za→ll+jet



PRL 125 (2020) 22, 221802

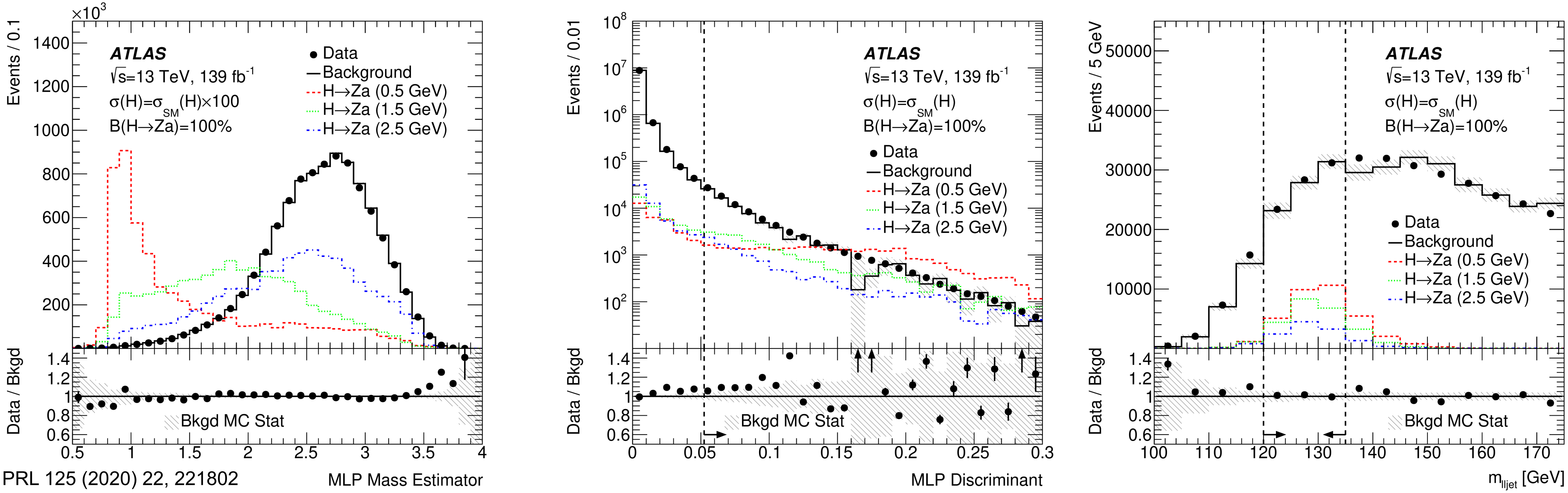
h→Za→ll+jet



PRL 125 (2020) 22, 221802

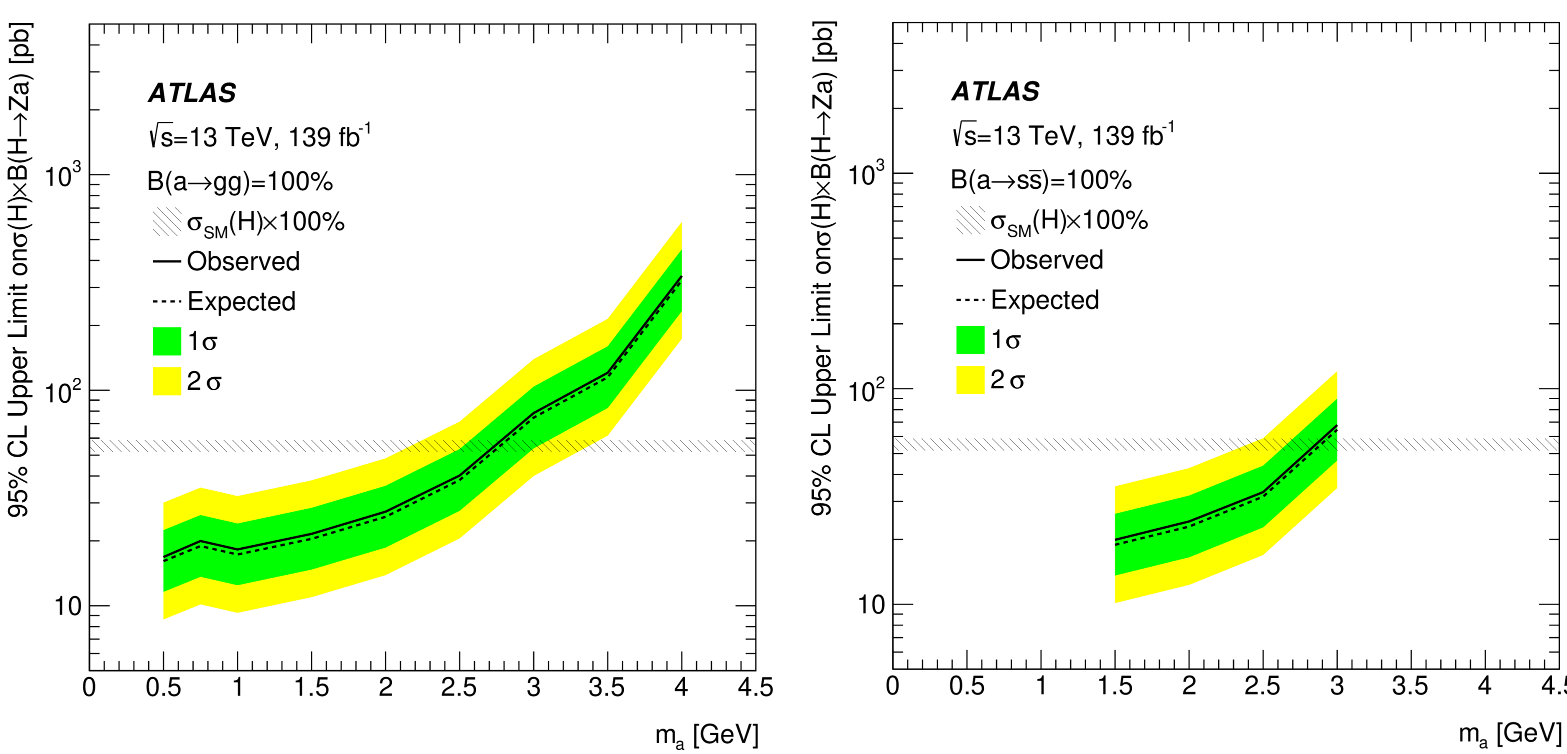
Expected Bkg: 82400±3700
Observed: 82908

h→Za→ll+jet



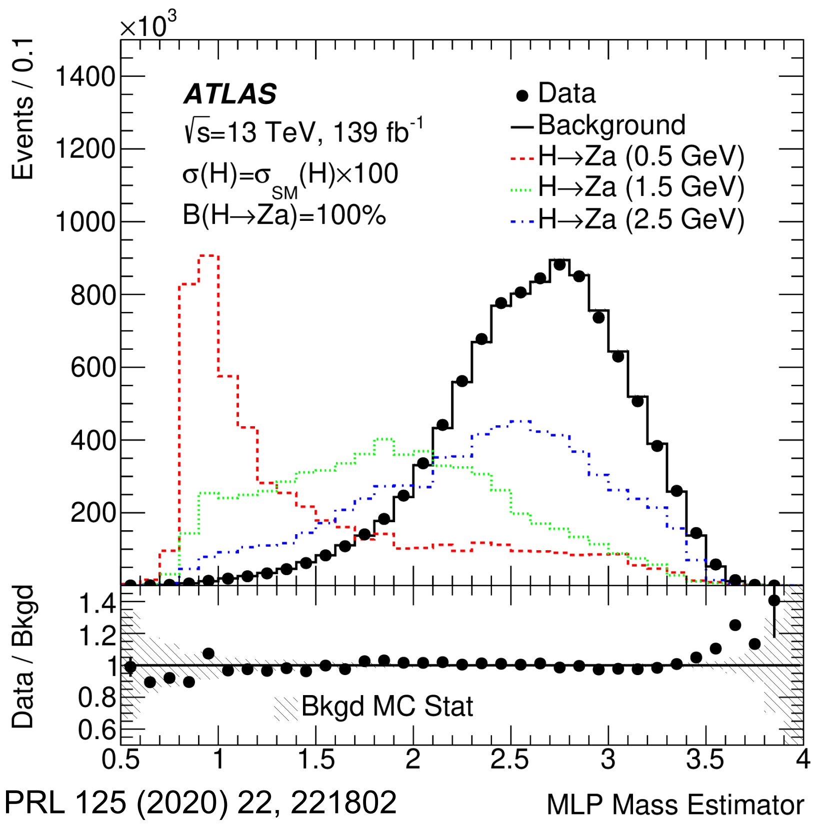
PRL 125 (2020) 22, 221802

Expected Bkg: 82400±3700
Observed: 82908

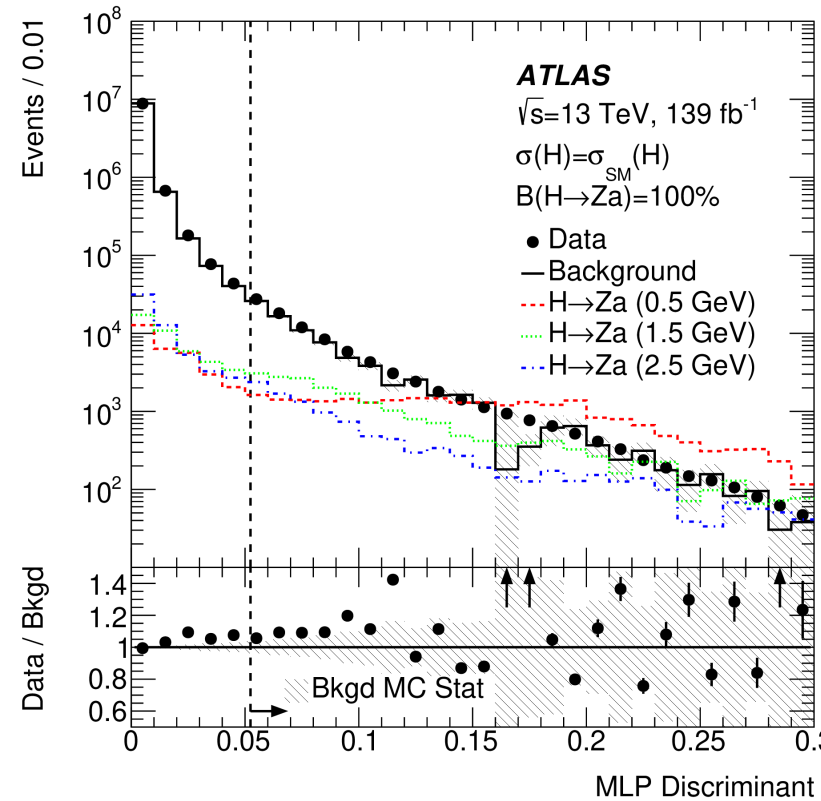


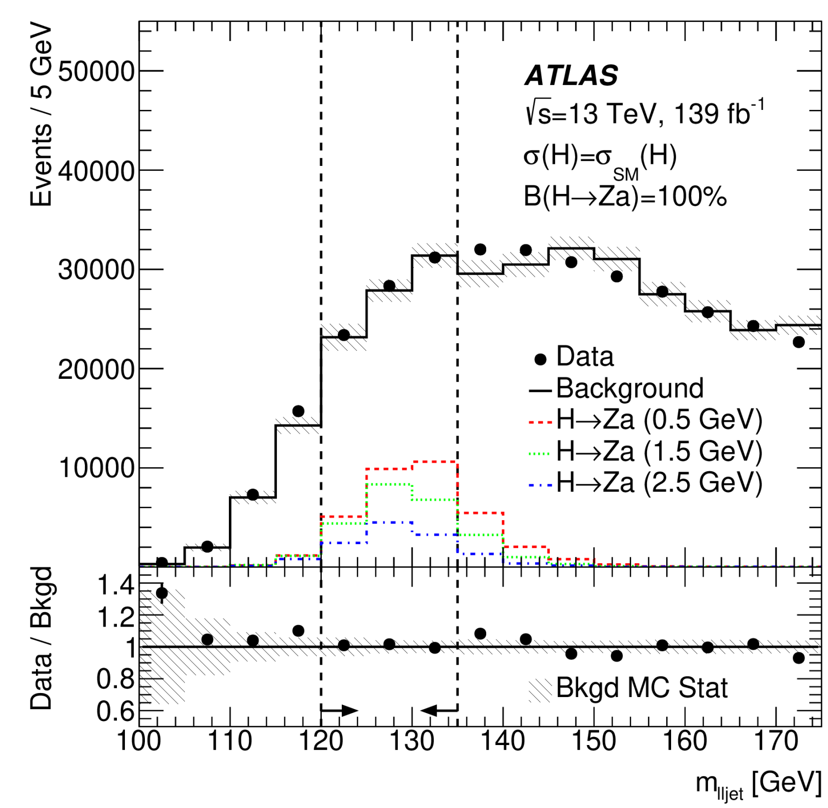
Expressed in $B(H \to Za) \times B(a \to \text{hadrons})$ limits start from BR<31%

h→Za→ll+jet



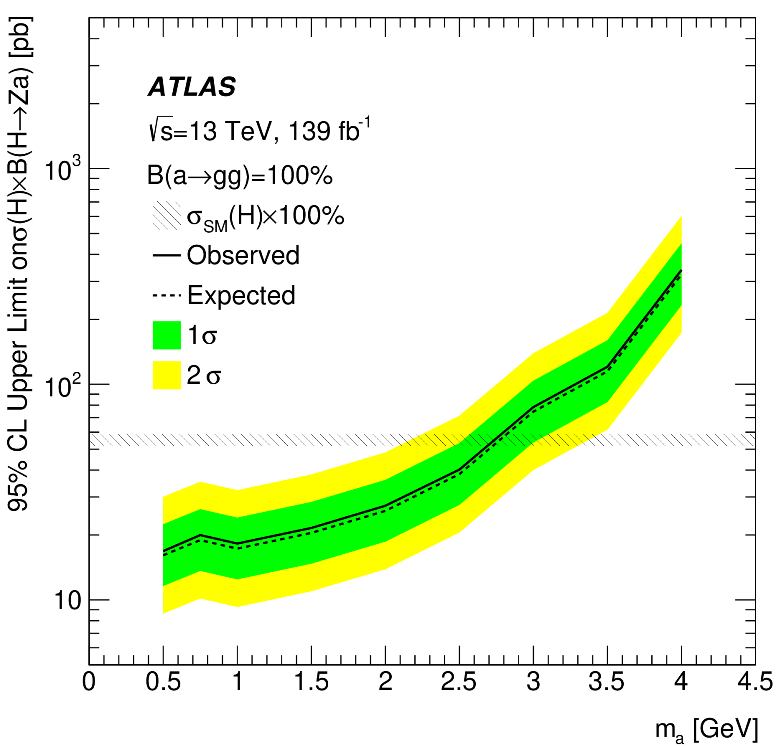
PRL 125 (2020) 22, 221802

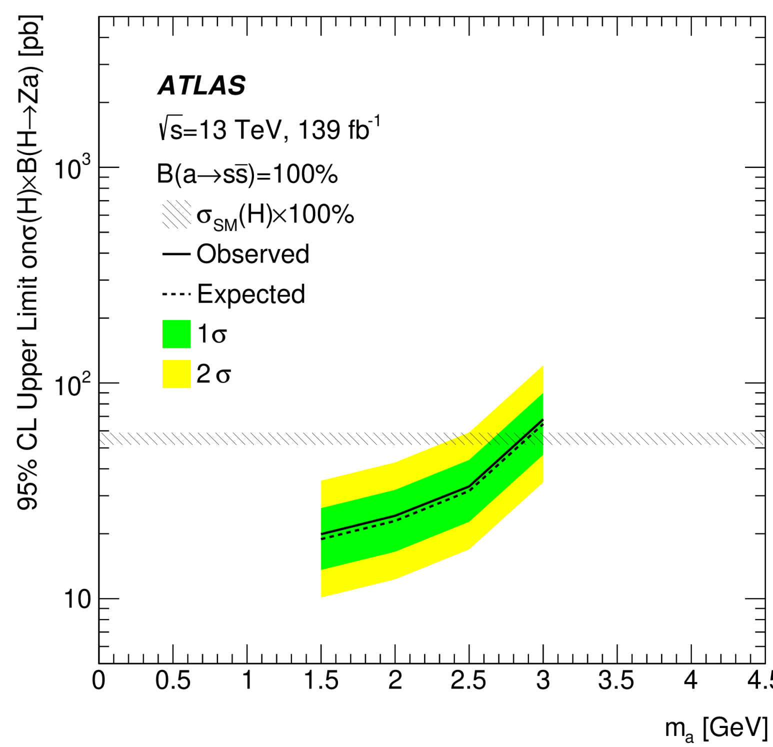


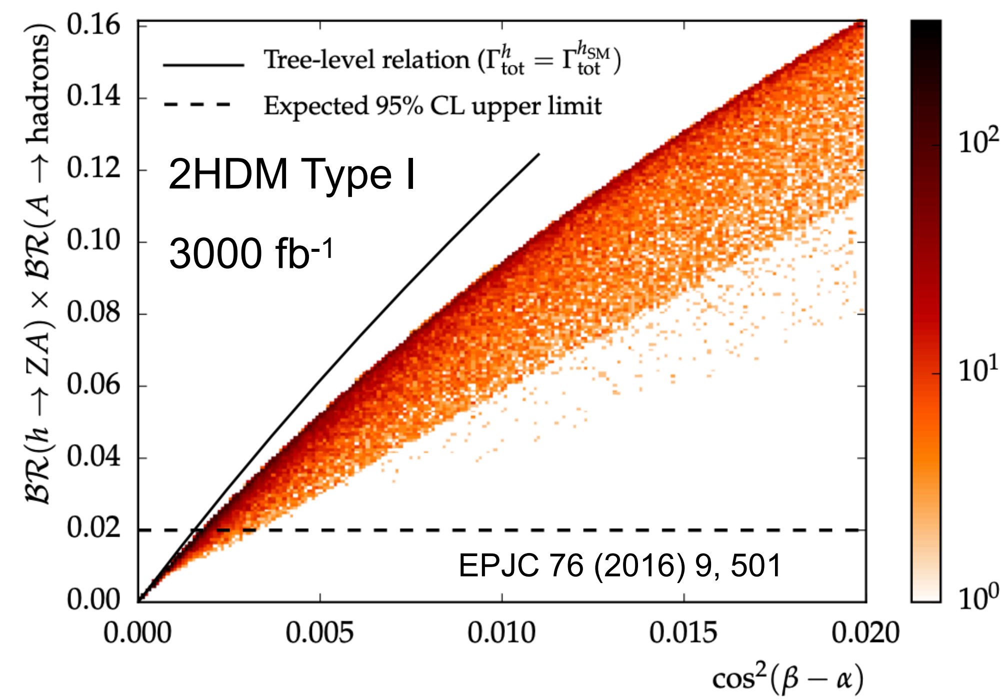


Expected Bkg: 82400±3700

Observed: 82908







Expressed in $B(H \to Za) \times B(a \to \text{hadrons})$ limits start from BR<31%

h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

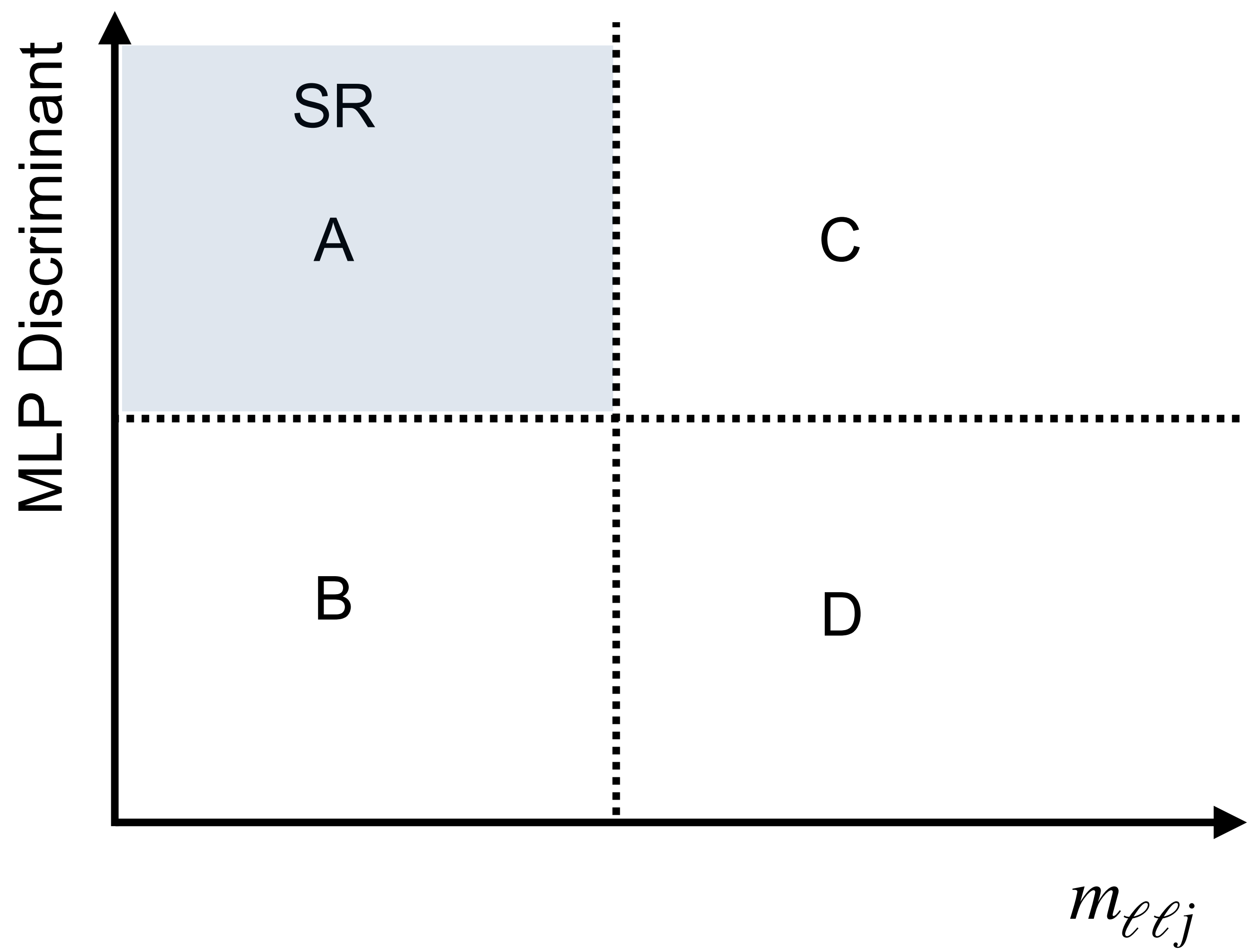
$$A_{\text{SR}}^{\text{ABCD Est.}} = \underbrace{\frac{B_{\text{data}} C_{\text{data}}}{D_{\text{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\text{MC}}}{\frac{B_{\text{MC}} C_{\text{MC}}}{D_{\text{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}}C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}}C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

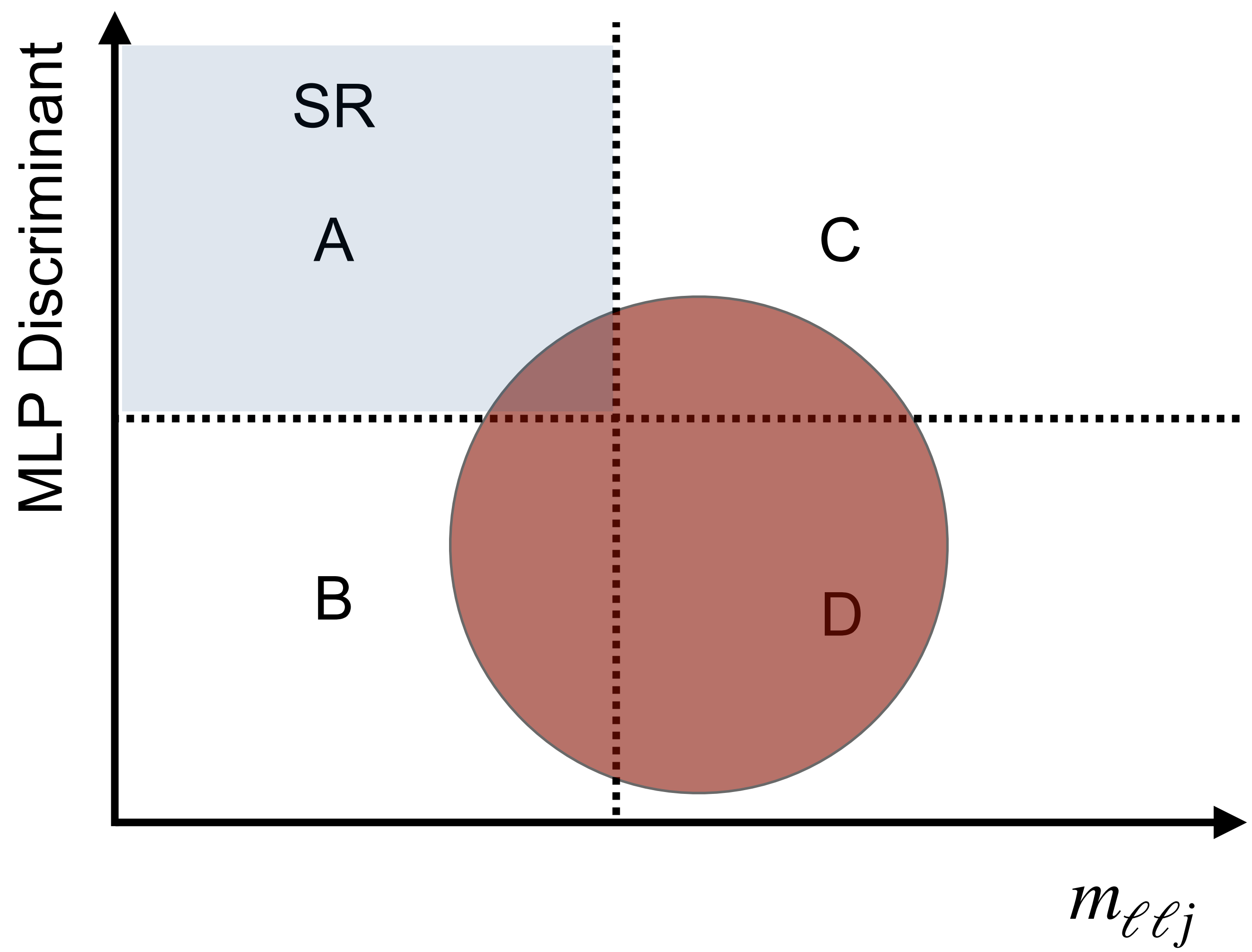


h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}} C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}}$$

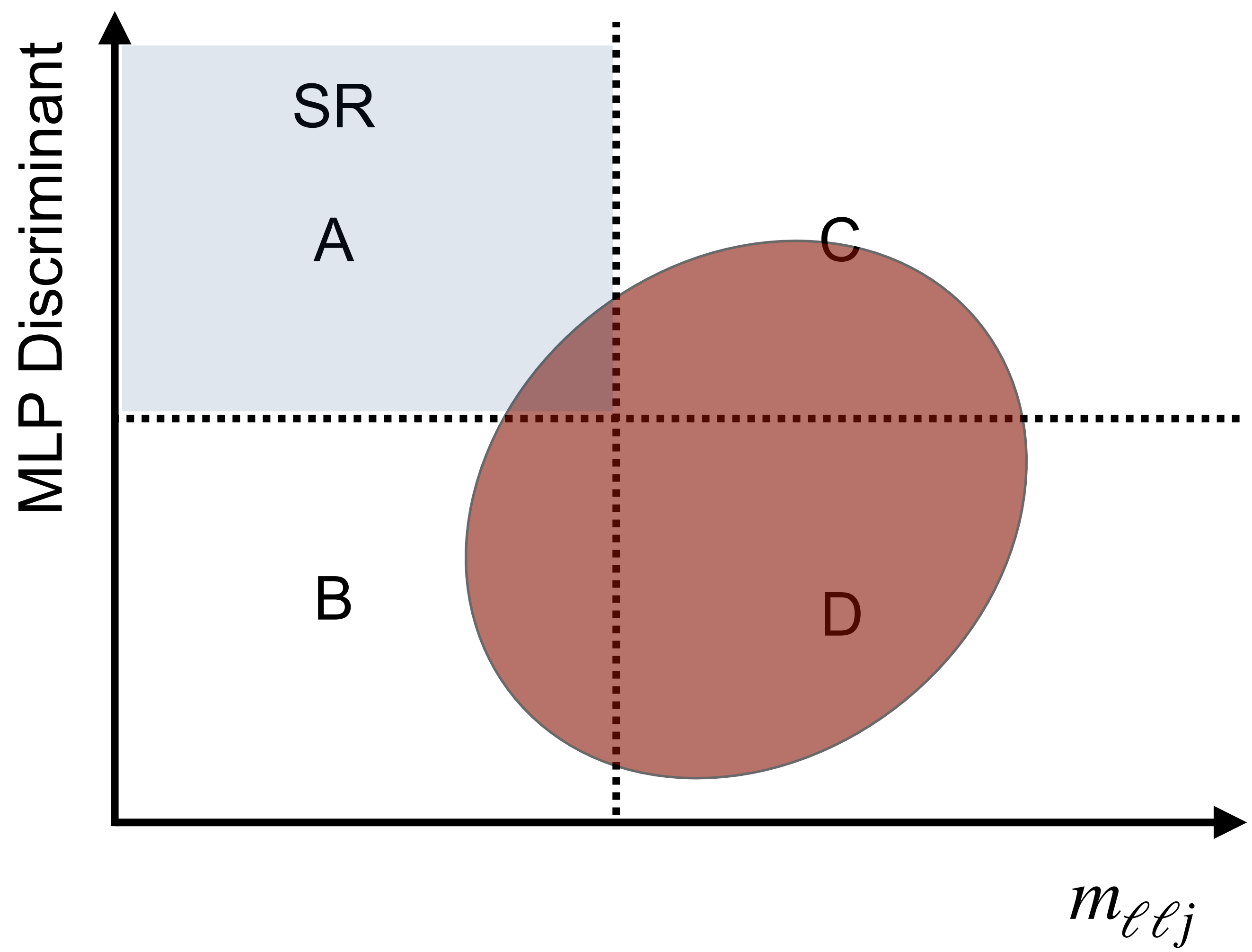


h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}}\,C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}}\,C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$



h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\text{ABCD Est.}} = \underbrace{\frac{B_{\mathrm{data}} C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}} C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

$h \to Za \to ll$+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}} C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}} C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

a mass	0.5 GeV	1.5 GeV	2.5 GeV
Total Uncertainty	8.3	10.7	20.3
Total Statistical Uncertainty	0.6	0.8	1.6
Total Systematic Uncertainty	8.2	10.7	20.2
Signal Systematic Uncertainties			
Jet Energy Scale	1.3	1.5	1.5
Parton Shower	1.4	1.4	1.4
Luminosity, Pileup, Trigger, Leptons, & JVT	0.2	0.3	0.5
MC Statistics	0.2	0.2	0.6
Renormalization Scale	0.1	< 0.1	0.2
Acceptance	0.1	< 0.1	0.2
Background Systematic Uncertainties			
MC Statistics	6.4	8.4	15.8
Parton Shower and ME	3.9	5.1	9.6
Renormalization Scale	3.4	4.4	8.3

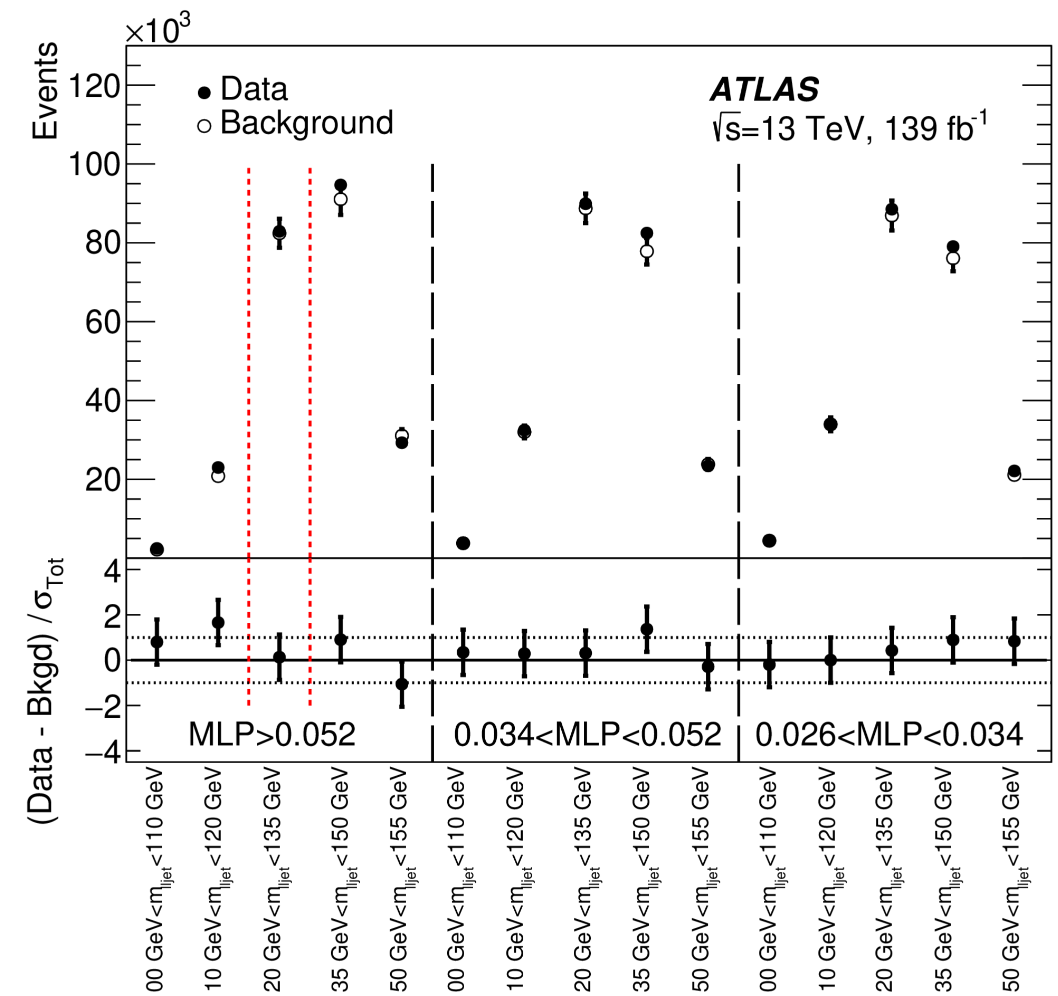
$h \to Za \to ll$+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}} C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}} C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

a mass	0.5 GeV	1.5 GeV	2.5 GeV
Total Uncertainty	8.3	10.7	20.3
Total Statistical Uncertainty	0.6	0.8	1.6
Total Systematic Uncertainty	8.2	10.7	20.2
Signal Systematic Uncertainties			
Jet Energy Scale	1.3	1.5	1.5
Parton Shower	1.4	1.4	1.4
Luminosity, Pileup, Trigger, Leptons, & JVT	0.2	0.3	0.5
MC Statistics	0.2	0.2	0.6
Renormalization Scale	0.1	< 0.1	0.2
Acceptance	0.1	< 0.1	0.2
Background Systematic Uncertainties			
MC Statistics	6.4	8.4	15.8
Parton Shower and ME	3.9	5.1	9.6
Renormalization Scale	3.4	4.4	8.3



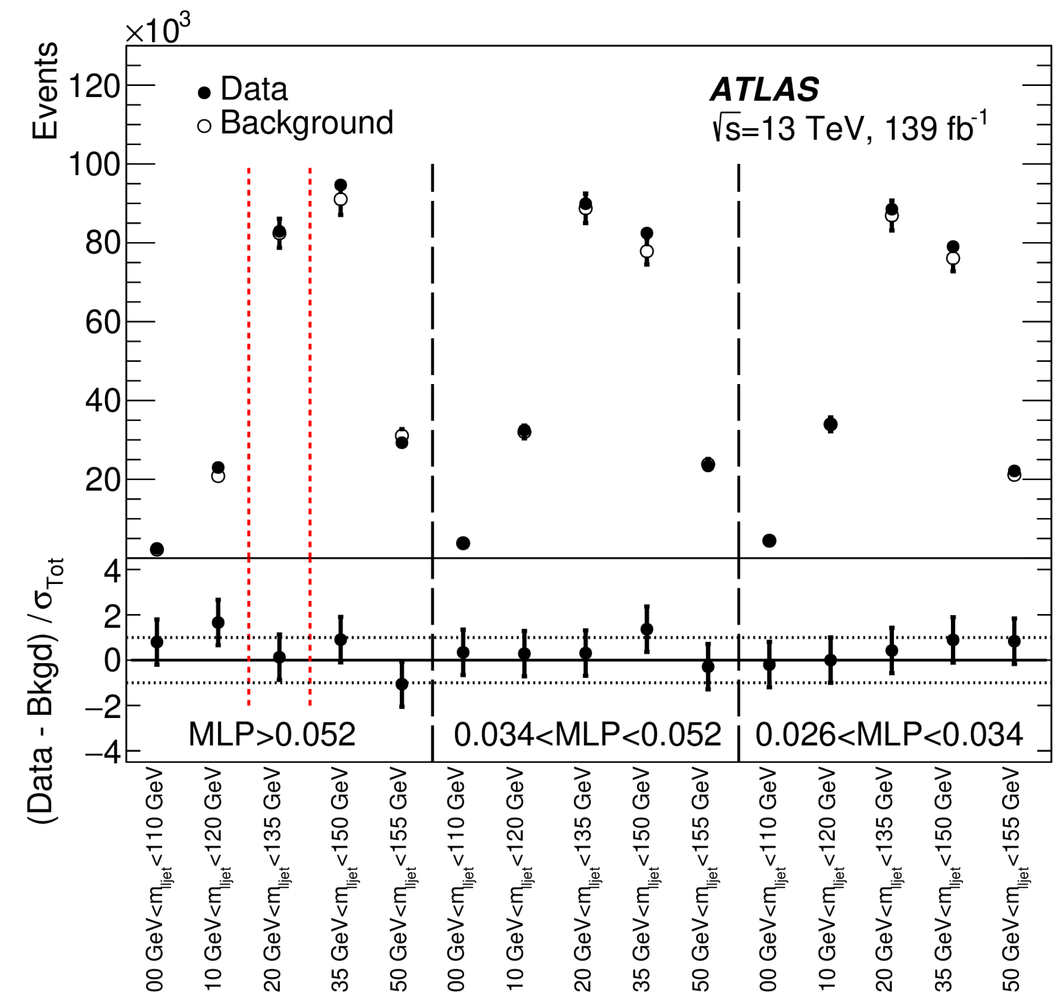
h→Za→ll+jet

Background estimation: MC-corrected ABCD method using $m_{\ell\ell j}$ and MLP discriminant

- Accounts for 13% correlation between $m_{\ell\ell j}$ and MLP discriminant

$$A_{\mathrm{SR}}^{\mathrm{ABCD\ Est.}} = \underbrace{\frac{B_{\mathrm{data}}C_{\mathrm{data}}}{D_{\mathrm{data}}}}_{\text{Data-driven ABCD Estimate}} \times \underbrace{\frac{A_{\mathrm{MC}}}{\frac{B_{\mathrm{MC}}C_{\mathrm{MC}}}{D_{\mathrm{MC}}}}}_{\text{MC-based ABCD Correction Factor}}$$

a mass	0.5 GeV	1.5 GeV	2.5 GeV
Total Uncertainty	8.3	10.7	20.3
Total Statistical Uncertainty	0.6	0.8	1.6
Total Systematic Uncertainty	8.2	10.7	20.2
Signal Systematic Uncertainties			
Jet Energy Scale	1.3	1.5	1.5
Parton Shower	1.4	1.4	1.4
Luminosity, Pileup, Trigger, Leptons, & JVT	0.2	0.3	0.5
MC Statistics	0.2	0.2	0.6
Renormalization Scale	0.1	< 0.1	0.2
Acceptance	0.1	< 0.1	0.2
Background Systematic Uncertainties			
MC Statistics	6.4	8.4	15.8
Parton Shower and ME	3.9	5.1	9.6
Renormalization Scale	3.4	4.4	8.3



Suppressing MC statistical/modelling uncertainties would improve limit from 31% to 7.5%!

Generative Adversarial Network

To improve analysis sensitivity → improve background model

- Increase sample size
- Improve Generator-level modelling uncertainties

Generative Adversarial Network

To improve analysis sensitivity → improve background model

- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Generative Adversarial Network

To improve analysis sensitivity → improve background model

- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Solution to sample size: Use a **Generative Adversarial Network** to generate the background sample

Generative Adversarial Network

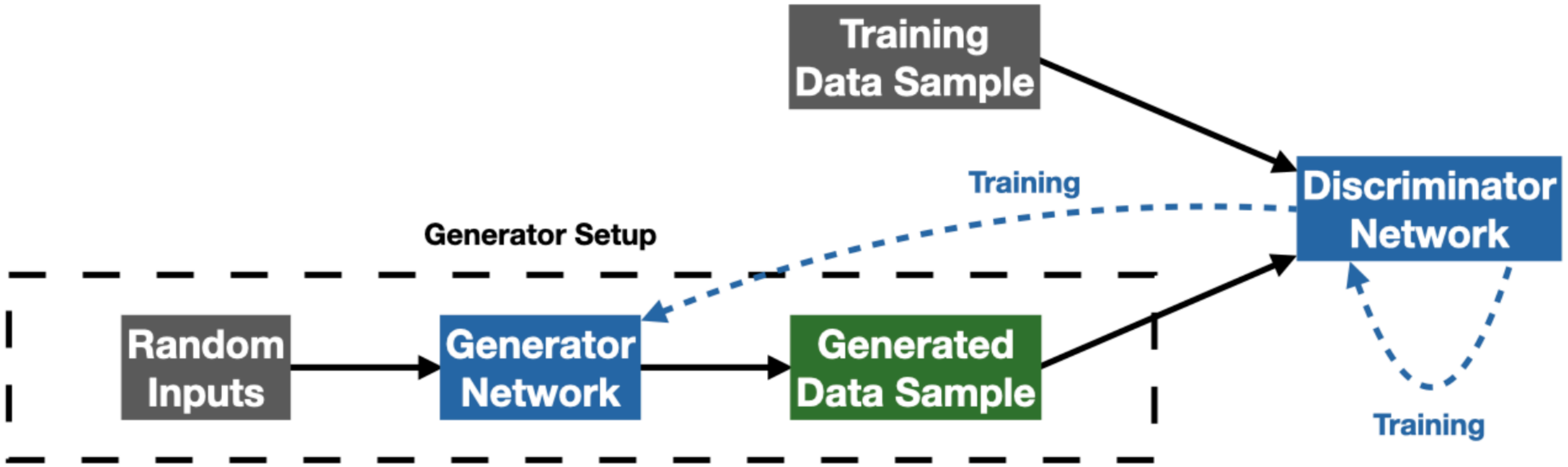
To improve analysis sensitivity → improve background model

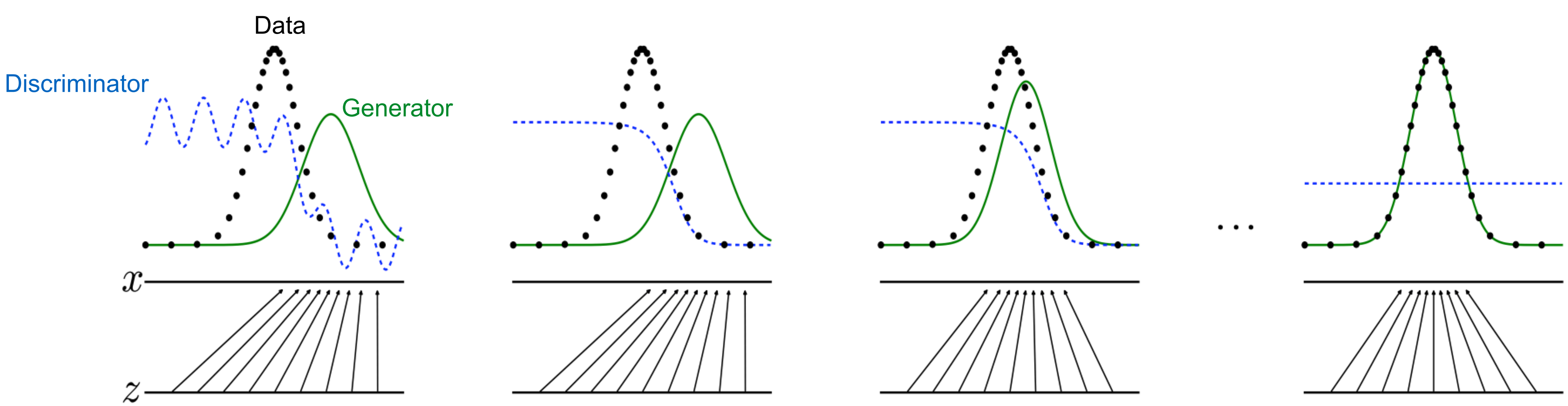
- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Solution to sample size: Use a **Generative Adversarial Network** to generate the background sample





Generative Adversarial Network

To improve analysis sensitivity → improve background model

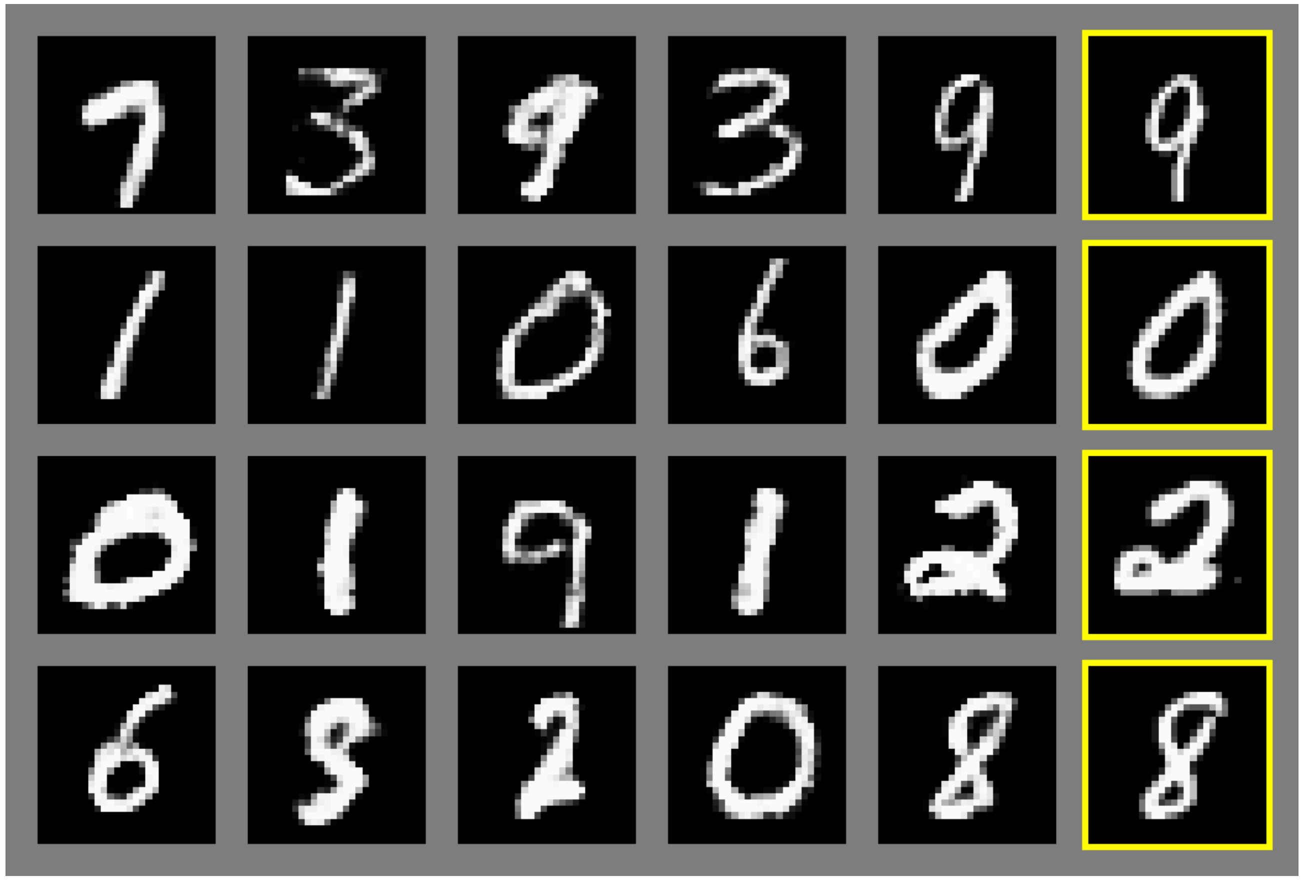
- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Solution to sample size: Use a **Generative Adversarial Network** to generate the background sample

arXiv:1406.2661



Generative Adversarial Network

To improve analysis sensitivity → improve background model

- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Solution to sample size: Use a **Generative Adversarial Network** to generate the background sample

Generative Adversarial Network

To improve analysis sensitivity → improve background model

- Increase sample size
- Improve Generator-level modelling uncertainties

Ancestral sampling procedure presented earlier is impractical

- Culprit: background discrimination uses multivariate techniques on variables

Solution to sample size: Use a **Generative Adversarial Network** to generate the background sample


StyleGAN2 (Dec 2019) - Karras et al. and Nvidia

Novelty: directly use data in superset of signal region for model generation

- Resolves concerns about modelling uncertainties

conditioned-GAN

Complication: dataset used for model generation may be contaminated by signal

- Blind the Signal Region while training the GAN

conditioned-GAN

Complication: dataset used for model generation may be contaminated by signal

- Blind the Signal Region while training the GAN

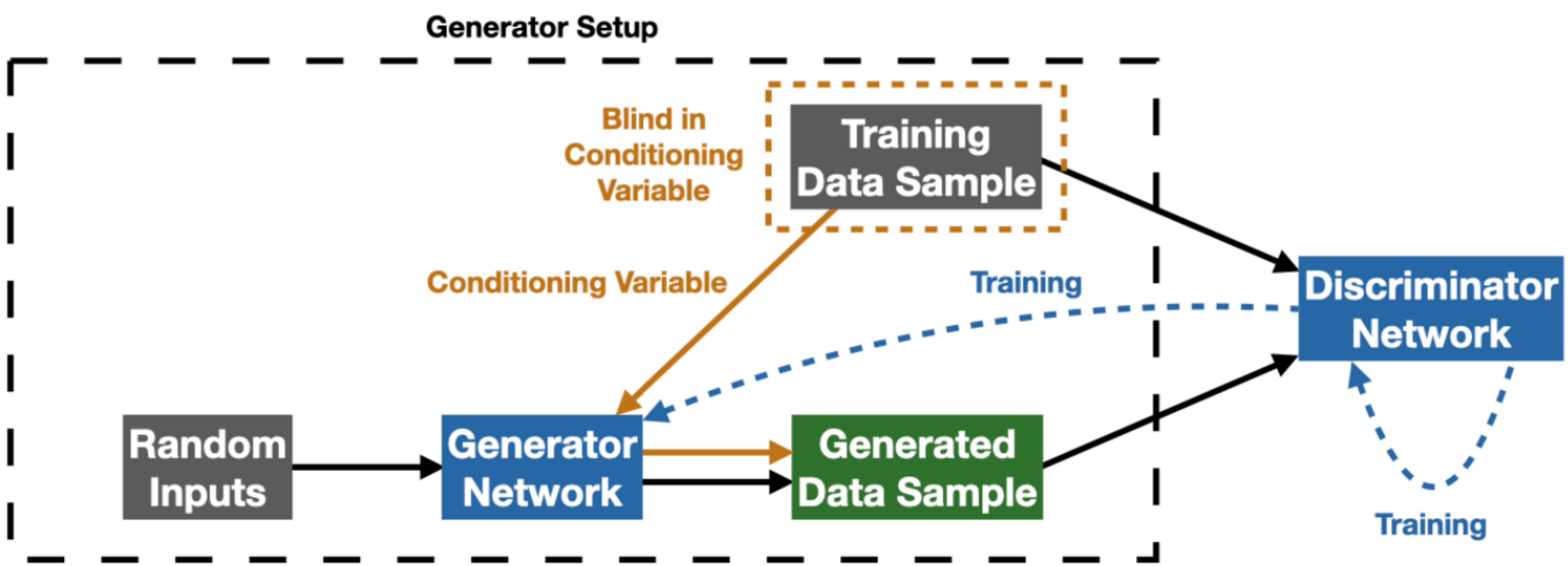
conditioned-GAN (cGAN): generator depends on **conditioning variable** → model can be interpolated

conditioned-GAN

Complication: dataset used for model generation may be contaminated by signal

- Blind the Signal Region while training the GAN

conditioned-GAN (cGAN): generator depends on **conditioning variable** → model can be interpolated



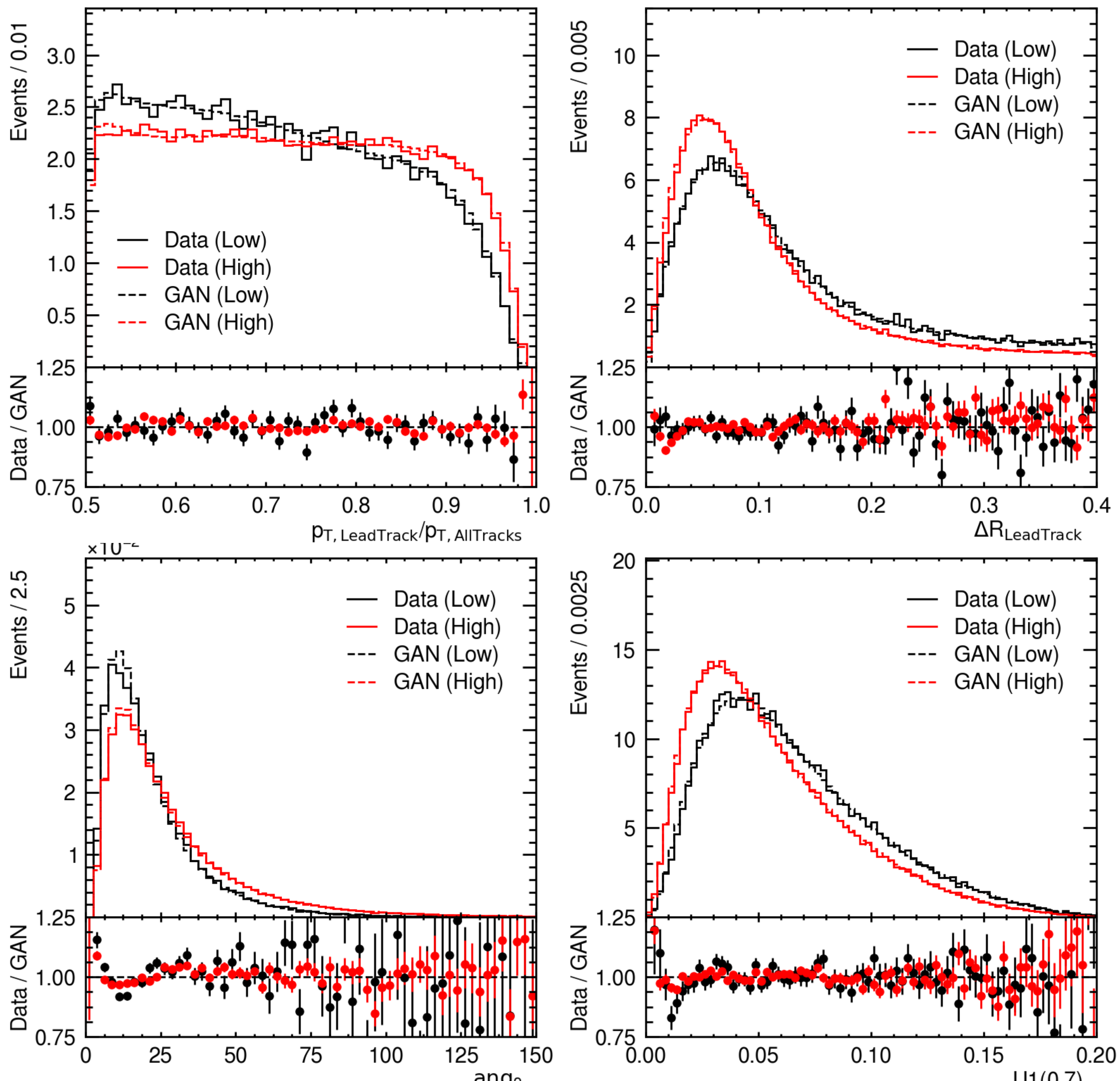
Generator and discriminator:

- 5 layers × 256 hidden nodes with leaky ReLU activation function
- Binary cross entropy loss function and L2 regularisation

cGAN: Modelling of variables

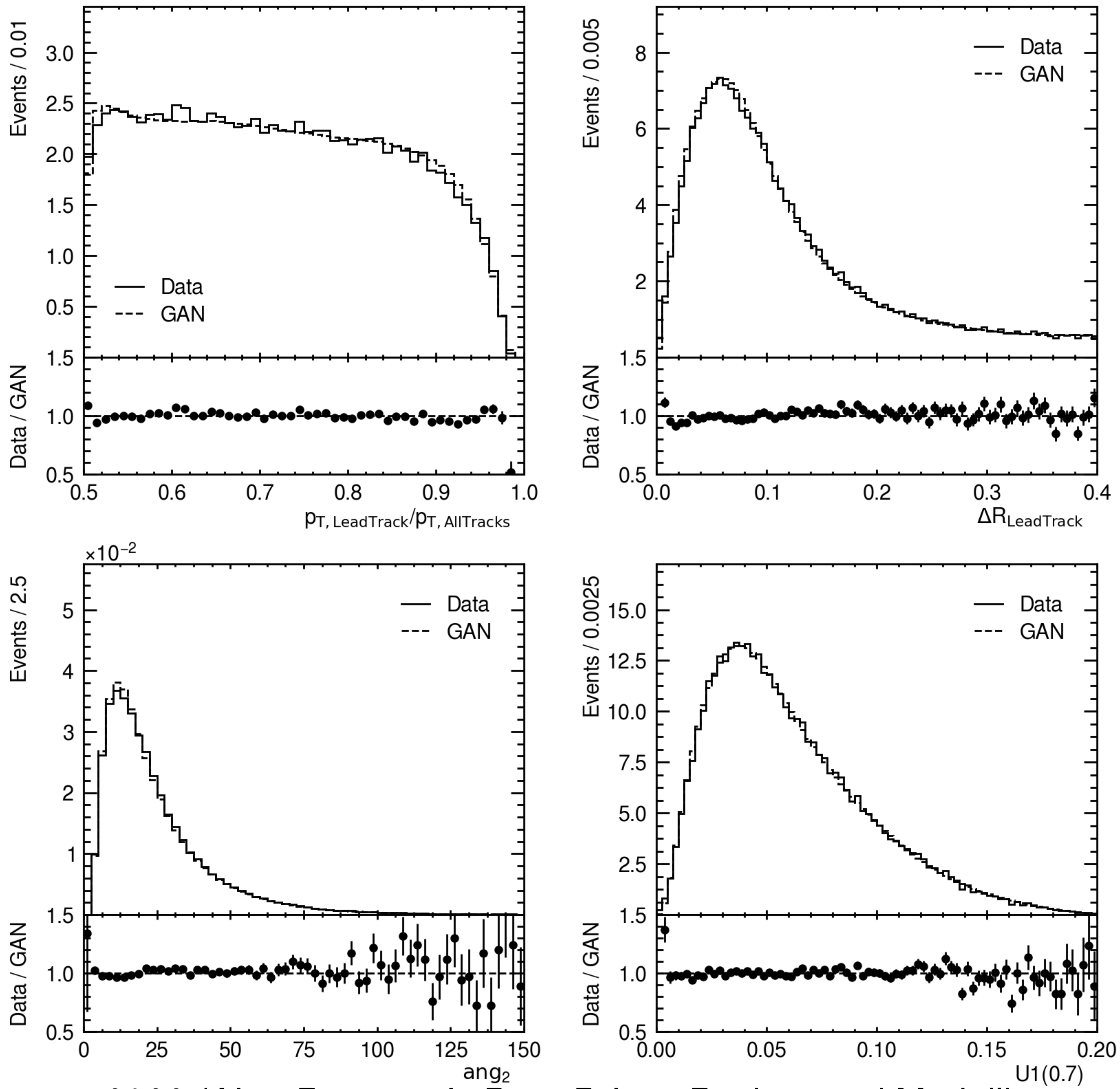
Trained 100 cGANs with random hyper-parameters

- Ensemble of top 5 cGANs, based on χ^2, retained

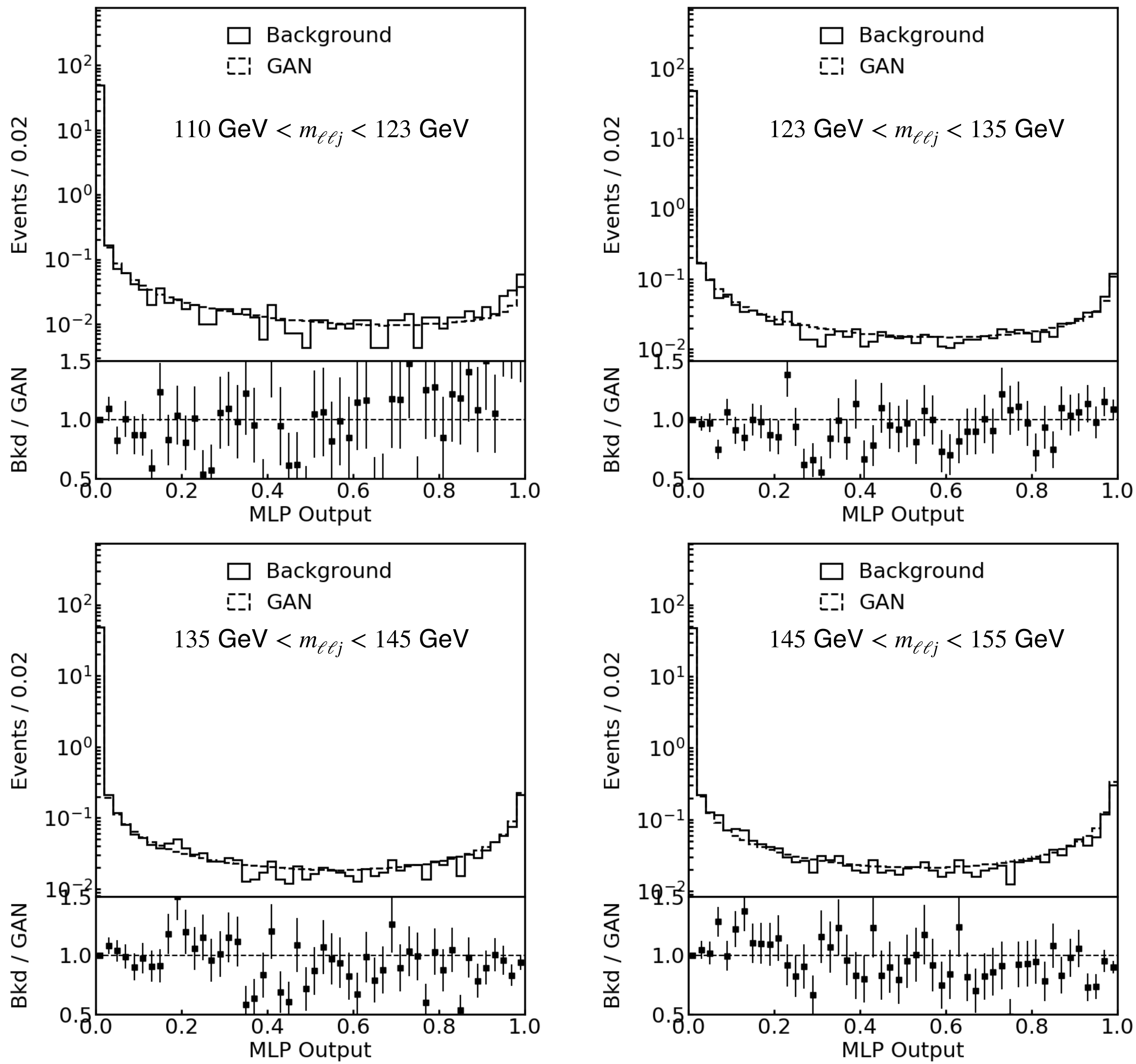


cGAN: Modelling of variables

$$123 \text{ GeV} < m_{\ell\ell j} < 135 \text{ GeV}$$



cGAN: Modelling of variables

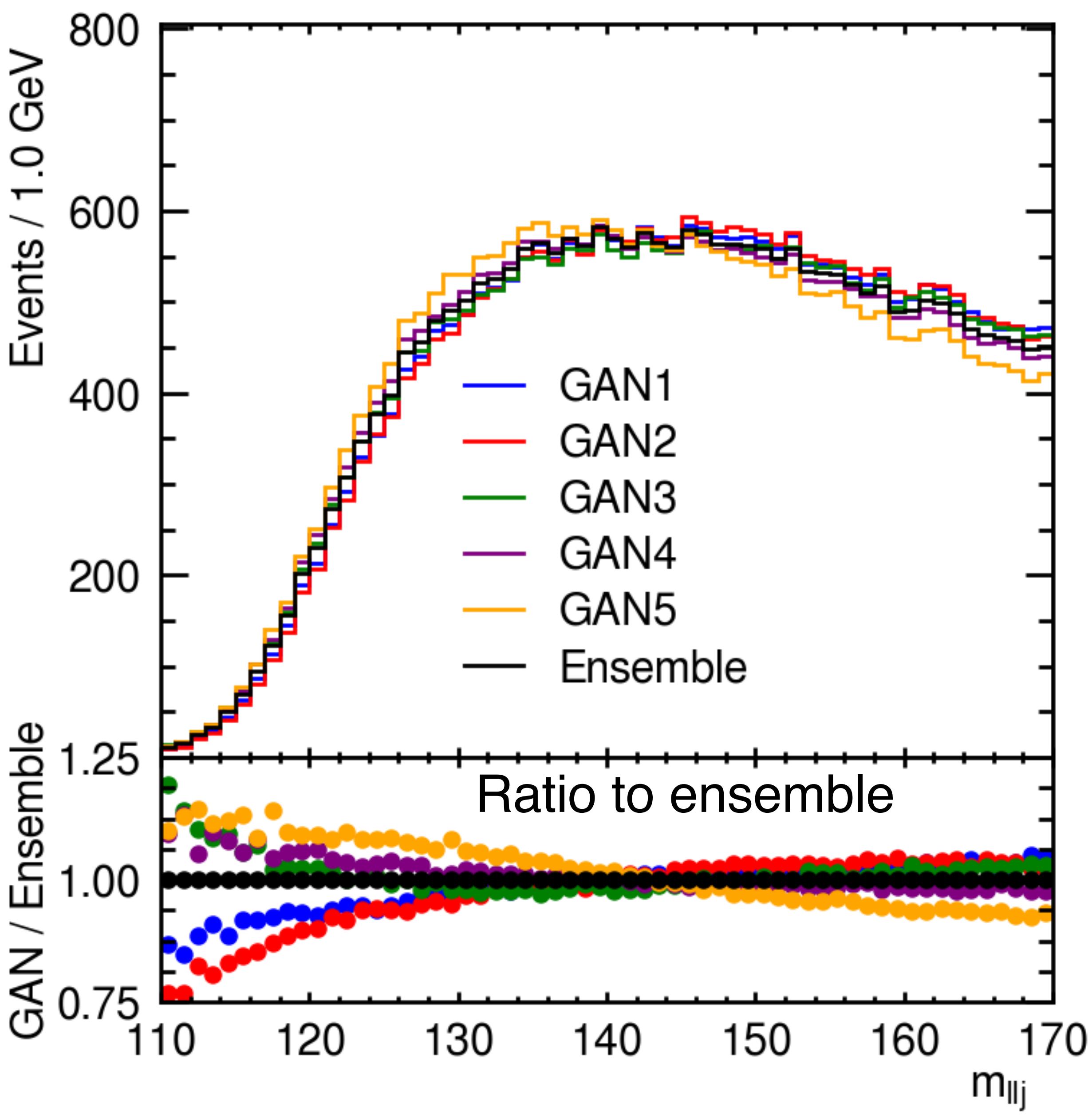


UNIVERSITY OF BIRMINGHAM

cGAN: Ensemble and Shape Variations

Shape variations:

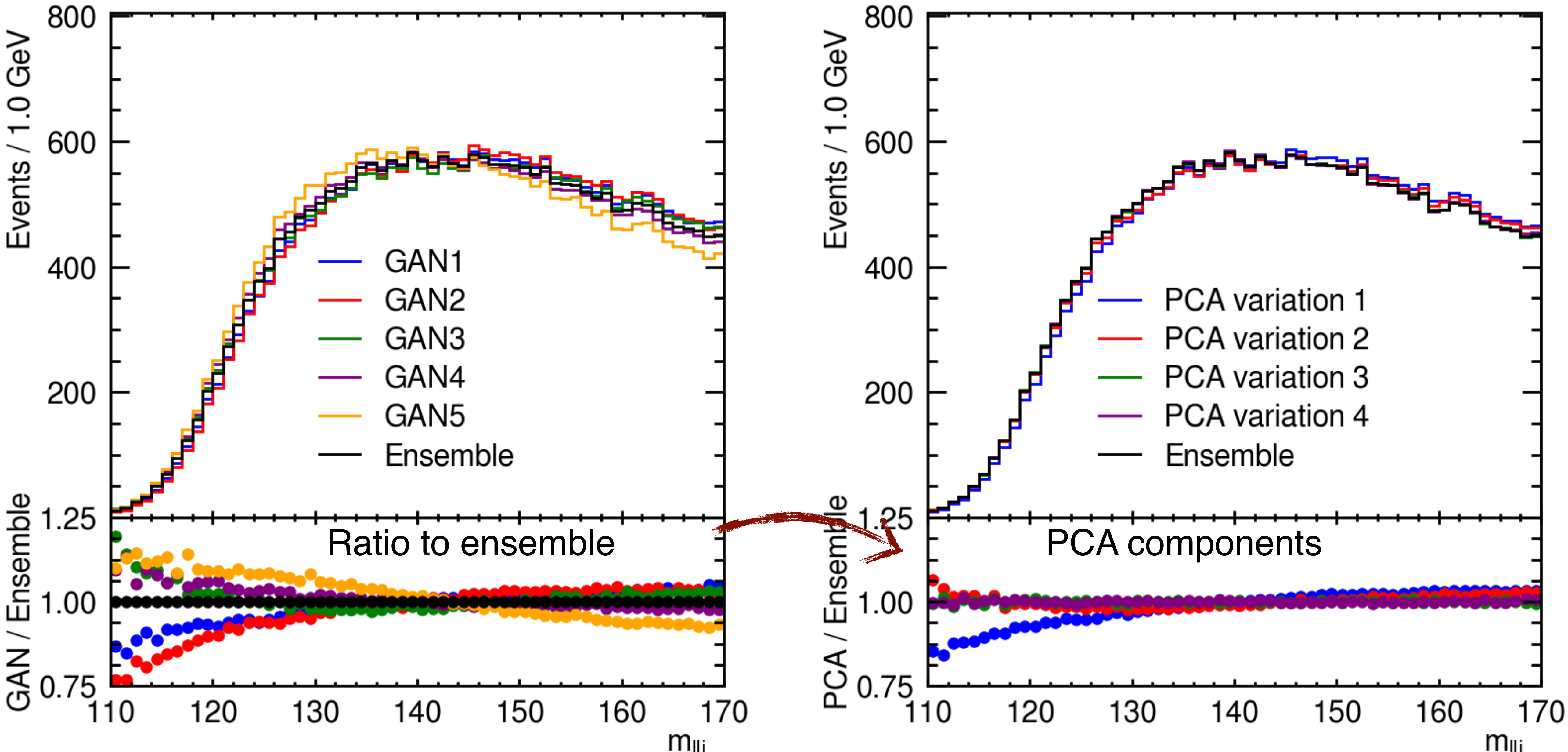
- Perform Principal Component Analysis on differences of individual cGANs to ensamble



cGAN: Ensemble and Shape Variations

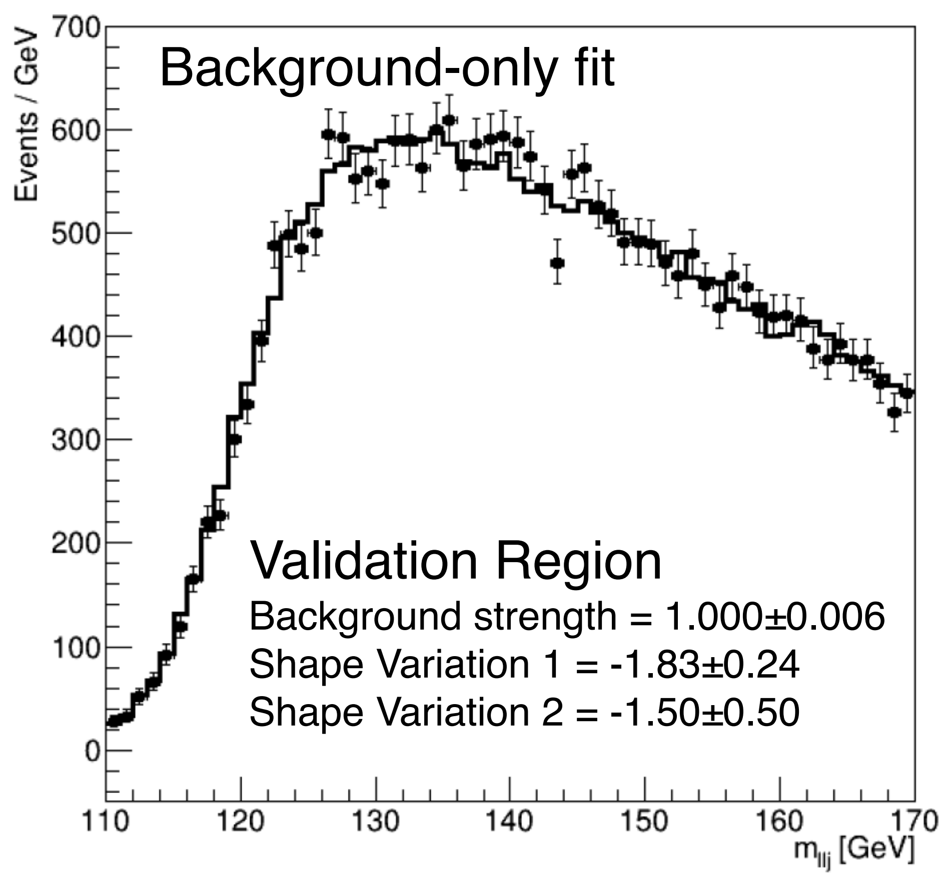
Shape variations:

- Perform Principal Component Analysis on differences of individual cGANs to ensamble

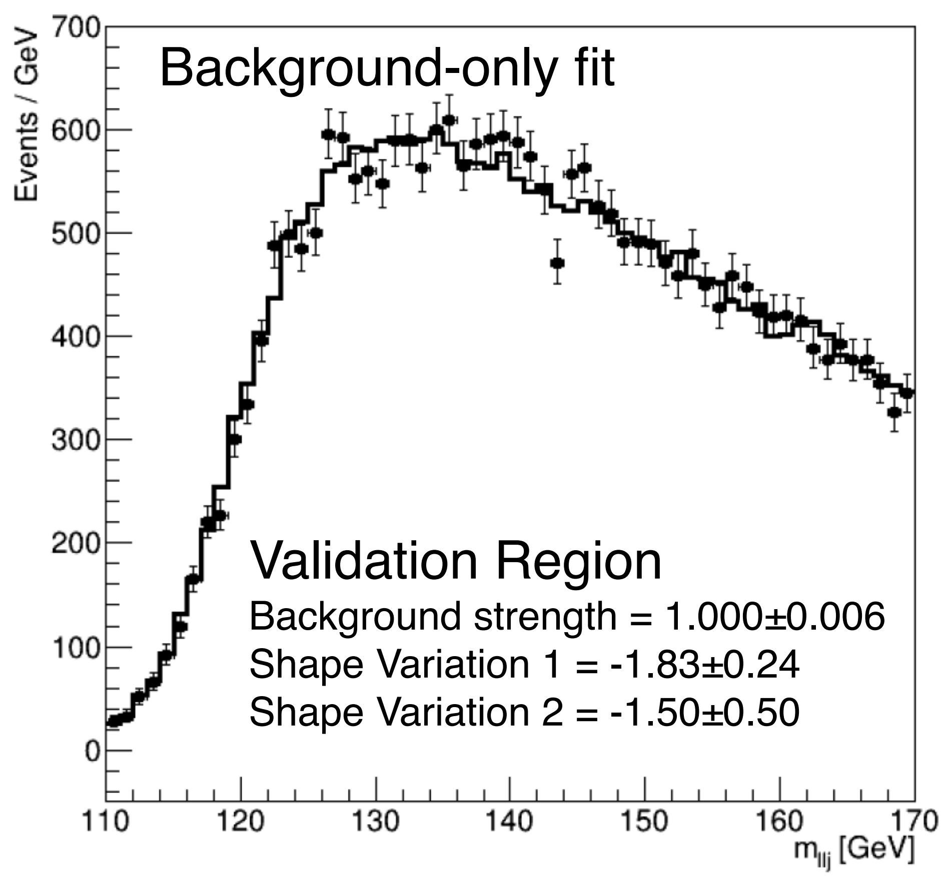


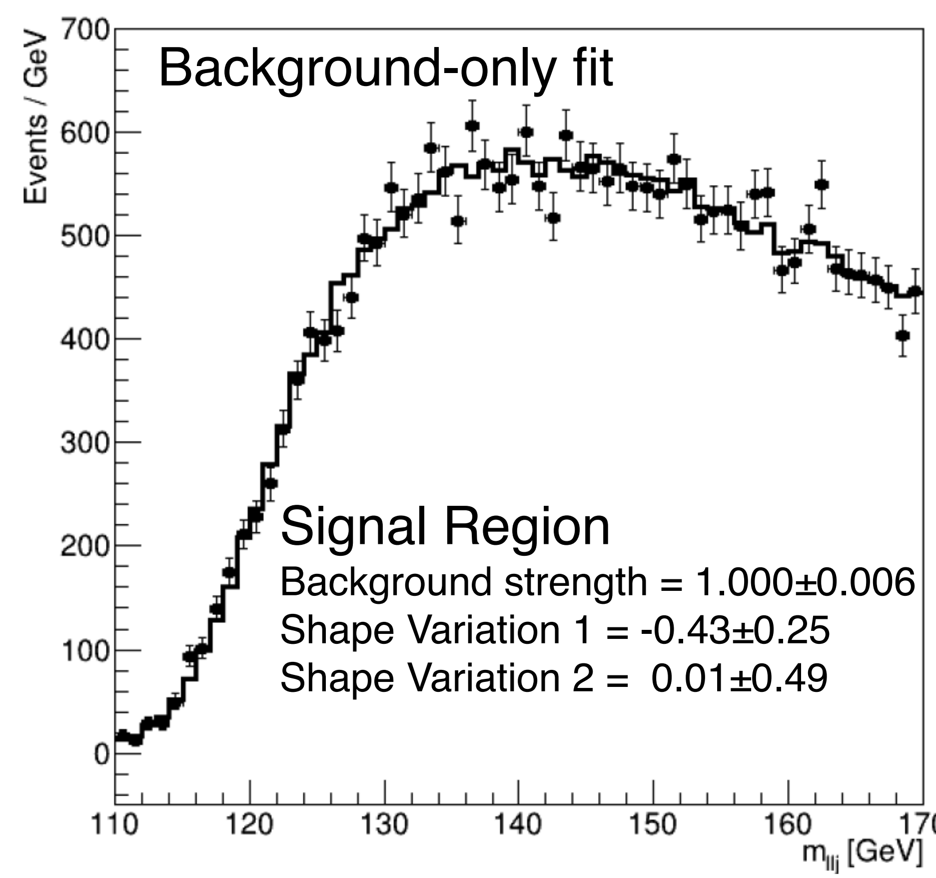
PCA components account for: 89%, 9.6%, 0.55%, and 0.40% of variance

cGAN: Fitting the "data"

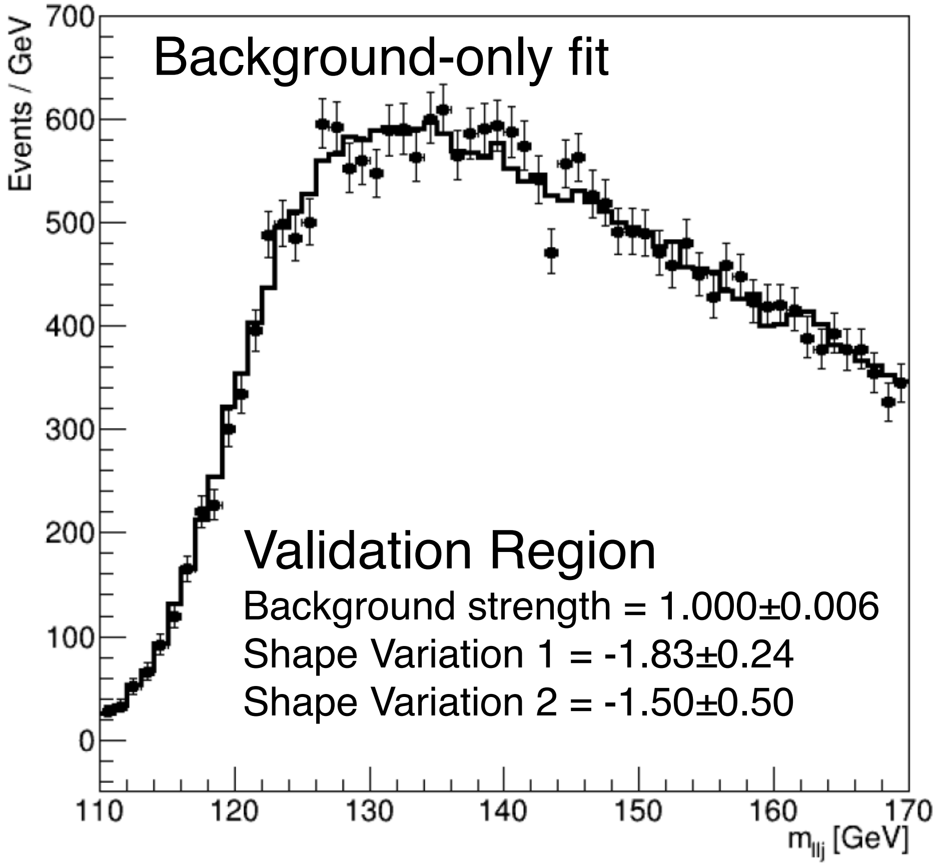


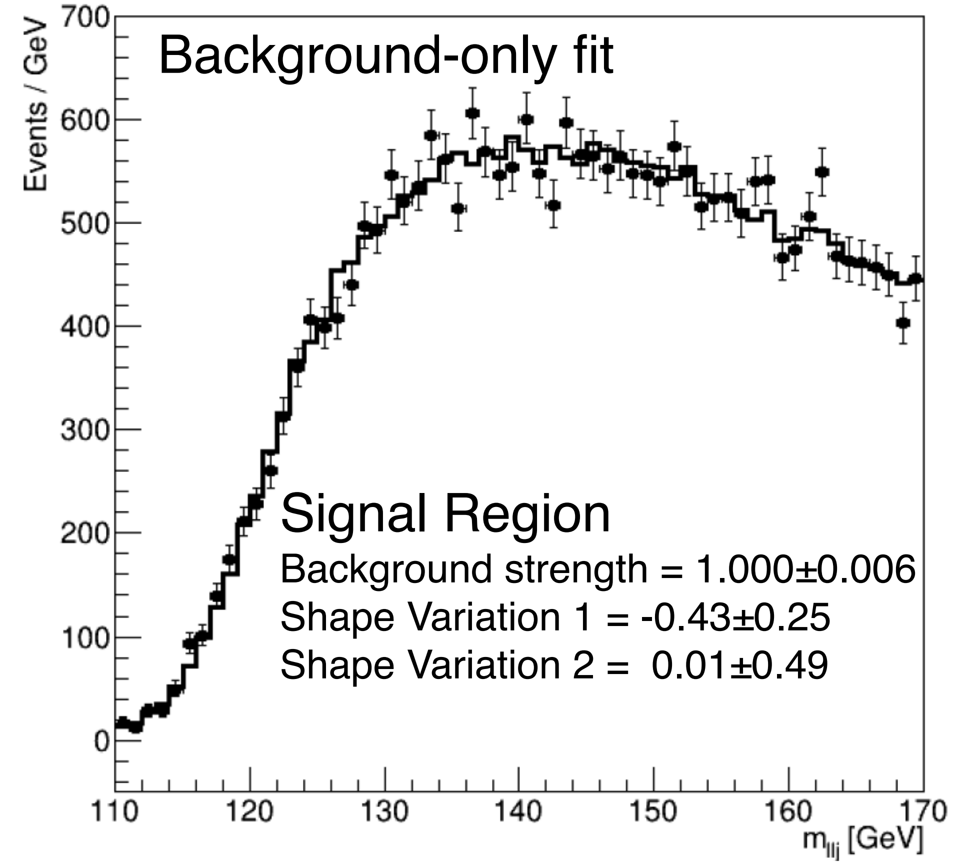
cGAN: Fitting the “data”

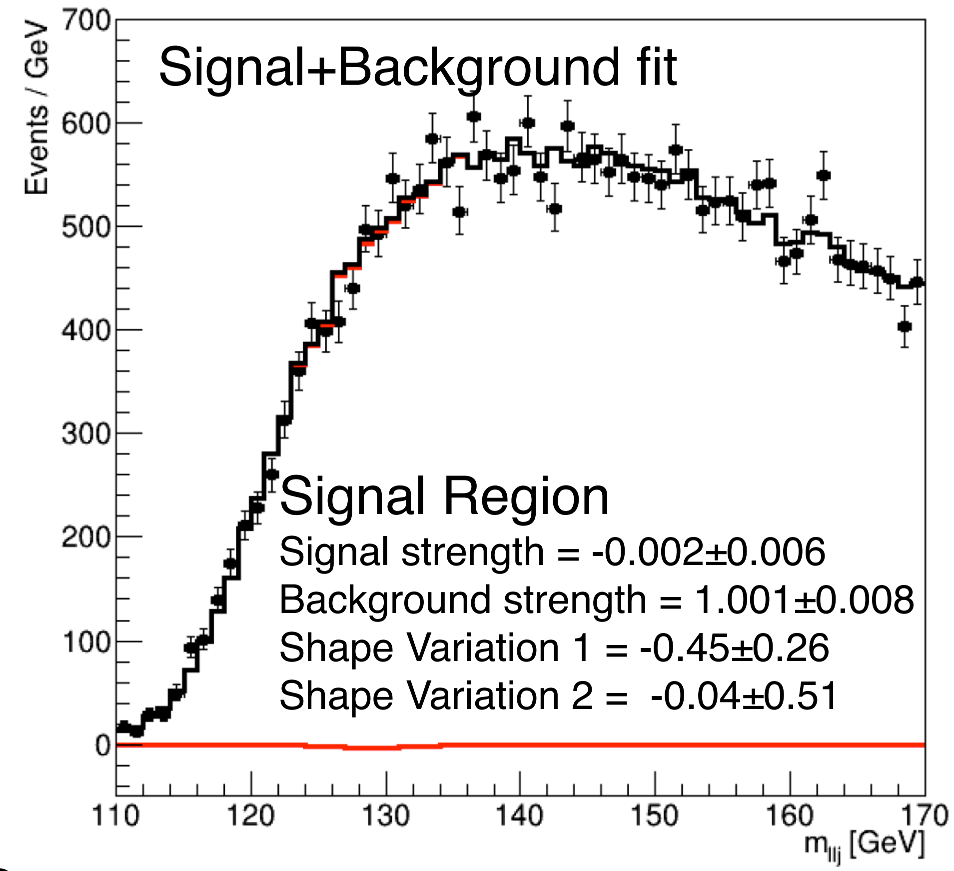




cGAN: Fitting the “data”







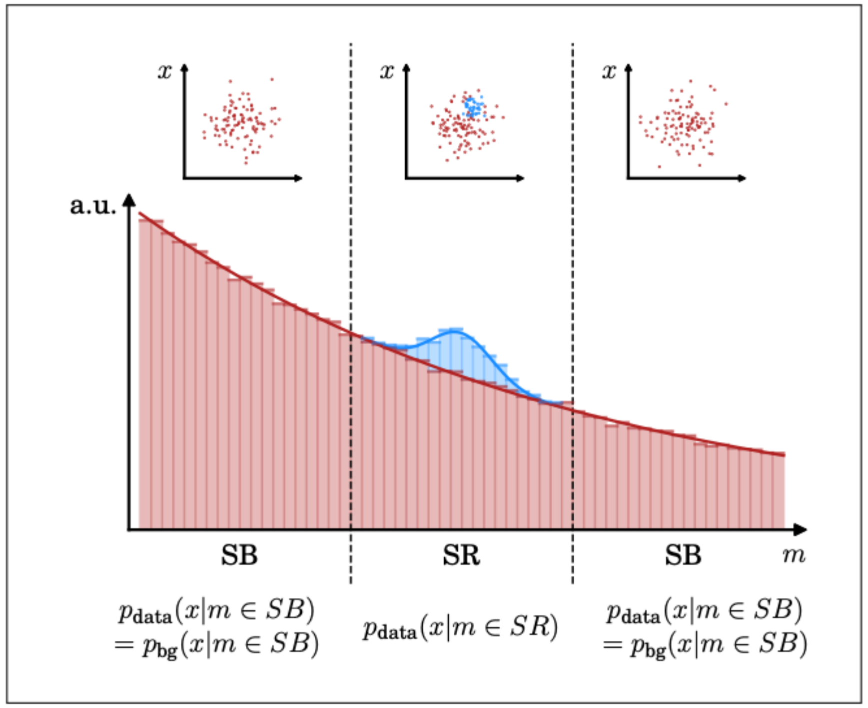
Signal+Background fit behaves as expected

- Obtained signal compatible with 0

CATHODE

- CATHODE: Classifying Anomalies Through Outer Density Estimation
  - Training a conditional density estimator (Masked Autoregressive Flow) on the discriminant variables in the side-band
  - Interpolating it into the signal region and sampling from it
  - Train classifier: separate SR data from produced "background" sample
  - Anomaly detection: Apply the trained classifier to data in SR
- In real life: the CATHODE method would need to be combined with a background estimation procedure

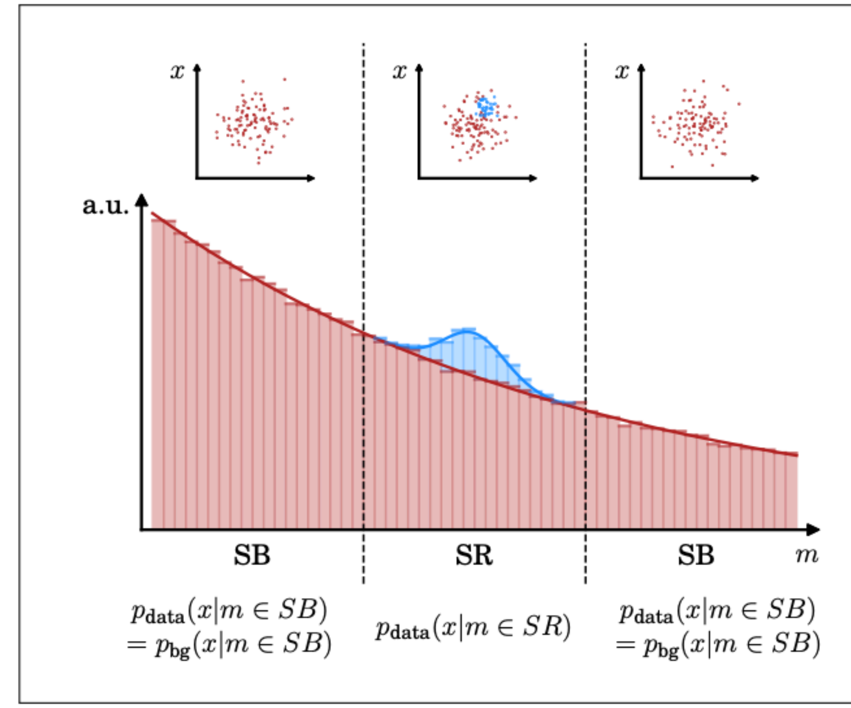
arXiv:2109.00546

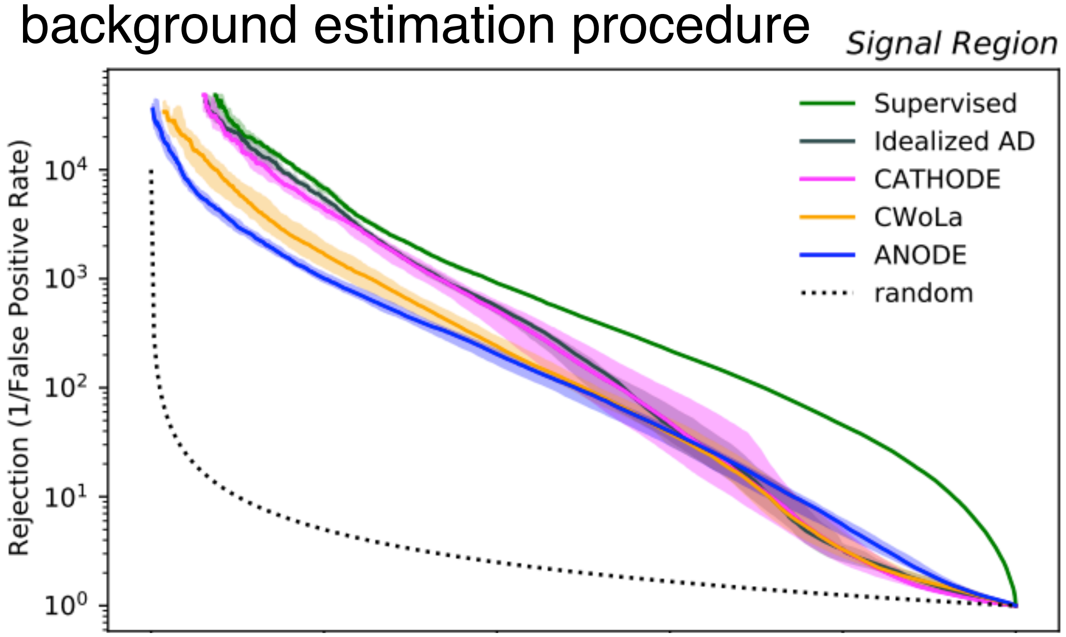


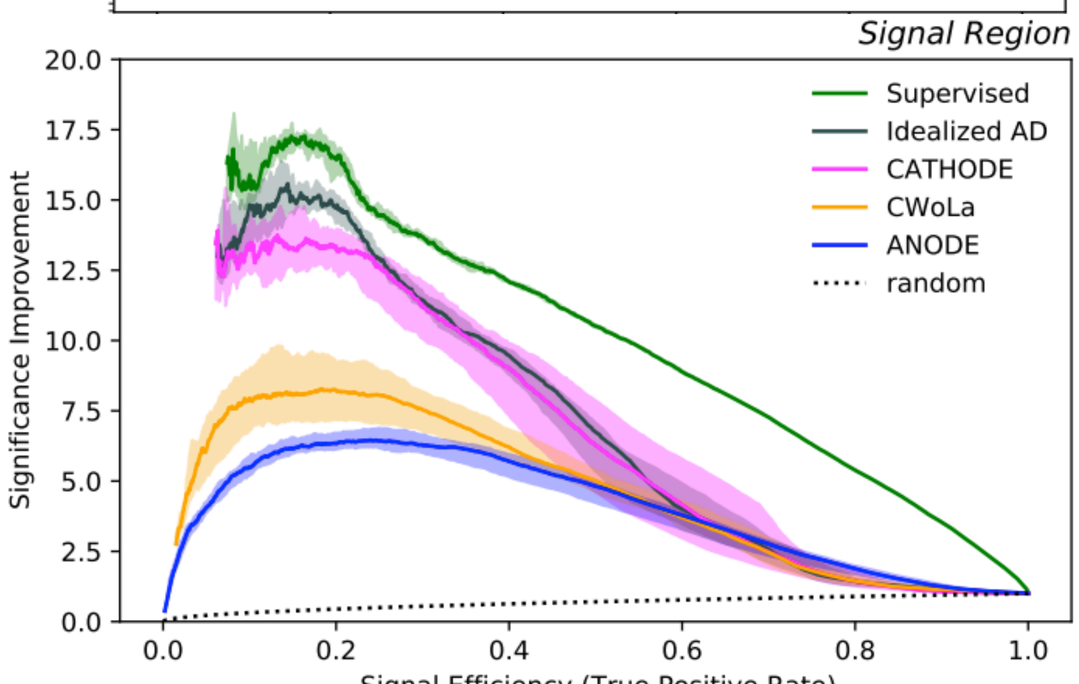
CATHODE

arXiv:2109.00546

- CATHODE: Classifying Anomalies Through Outer Density Estimation
  - Training a conditional density estimator (Masked Autoregressive Flow) on the discriminant variables in the side-band
  - Interpolating it into the signal region and sampling from it
  - Train classifier: separate SR data from produced “background” sample
  - Anomaly detection: Apply the trained classifier to data in SR
- In real life: the CATHODE method would need to be combined with a background estimation procedure



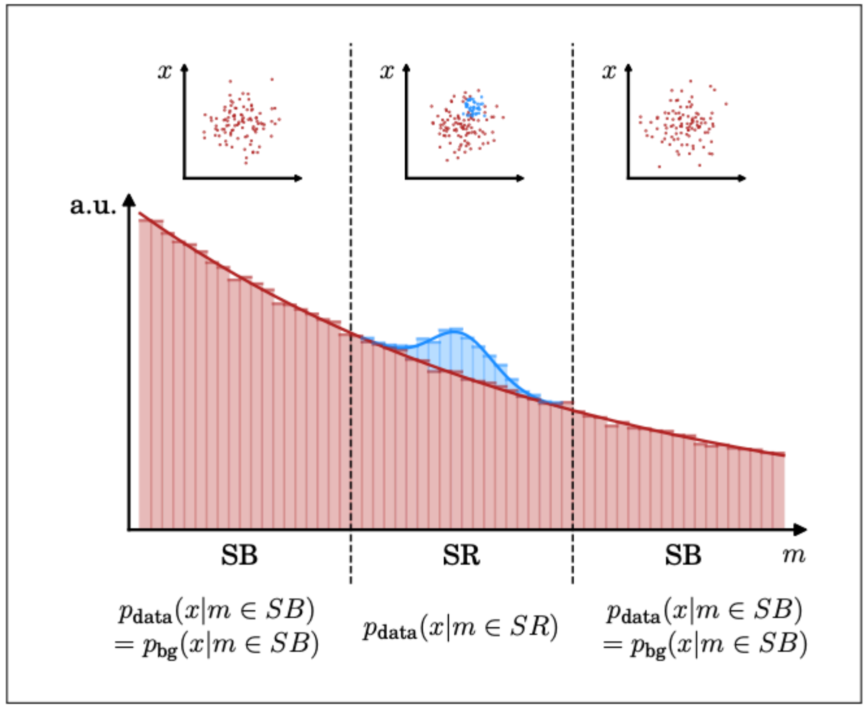


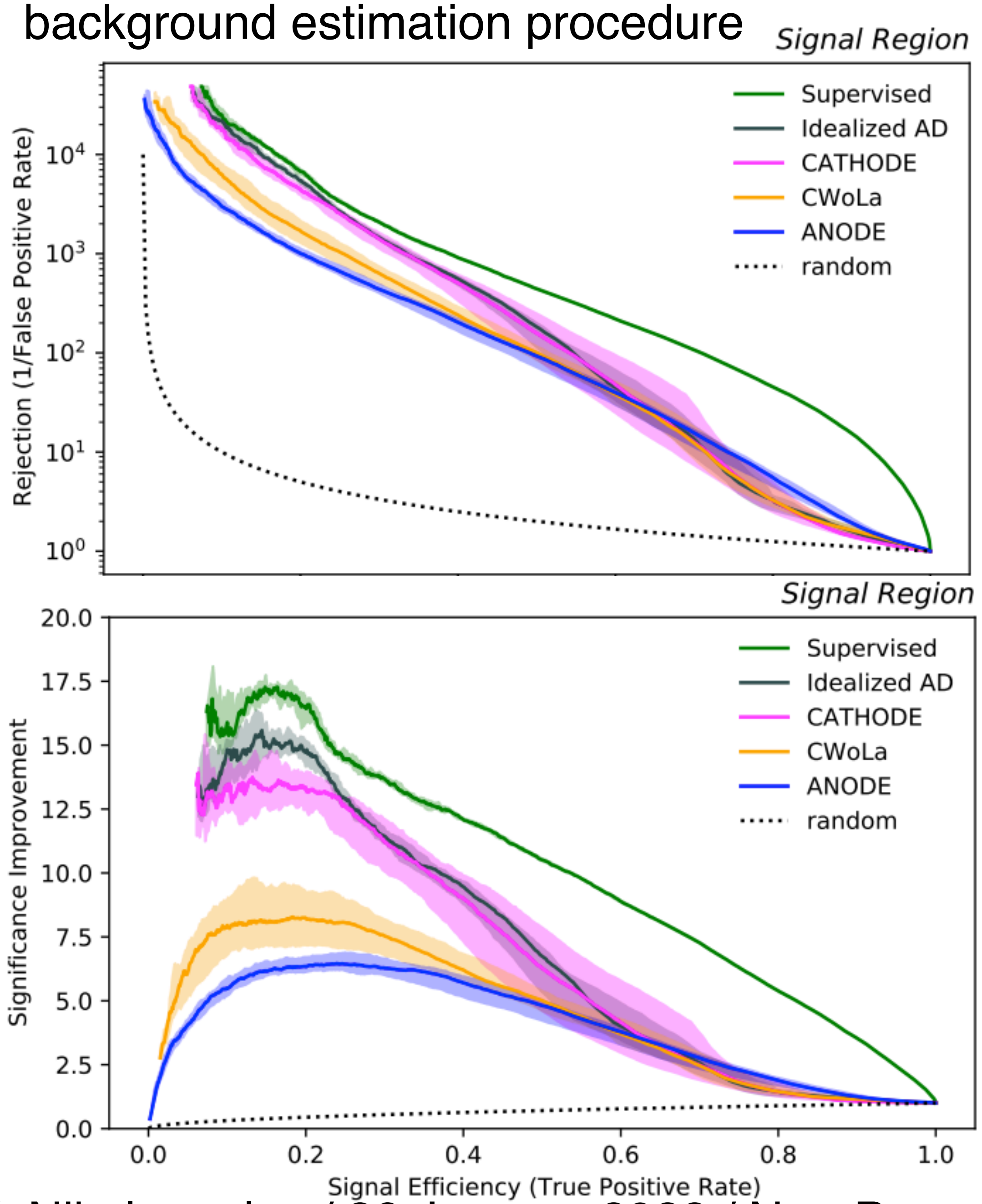


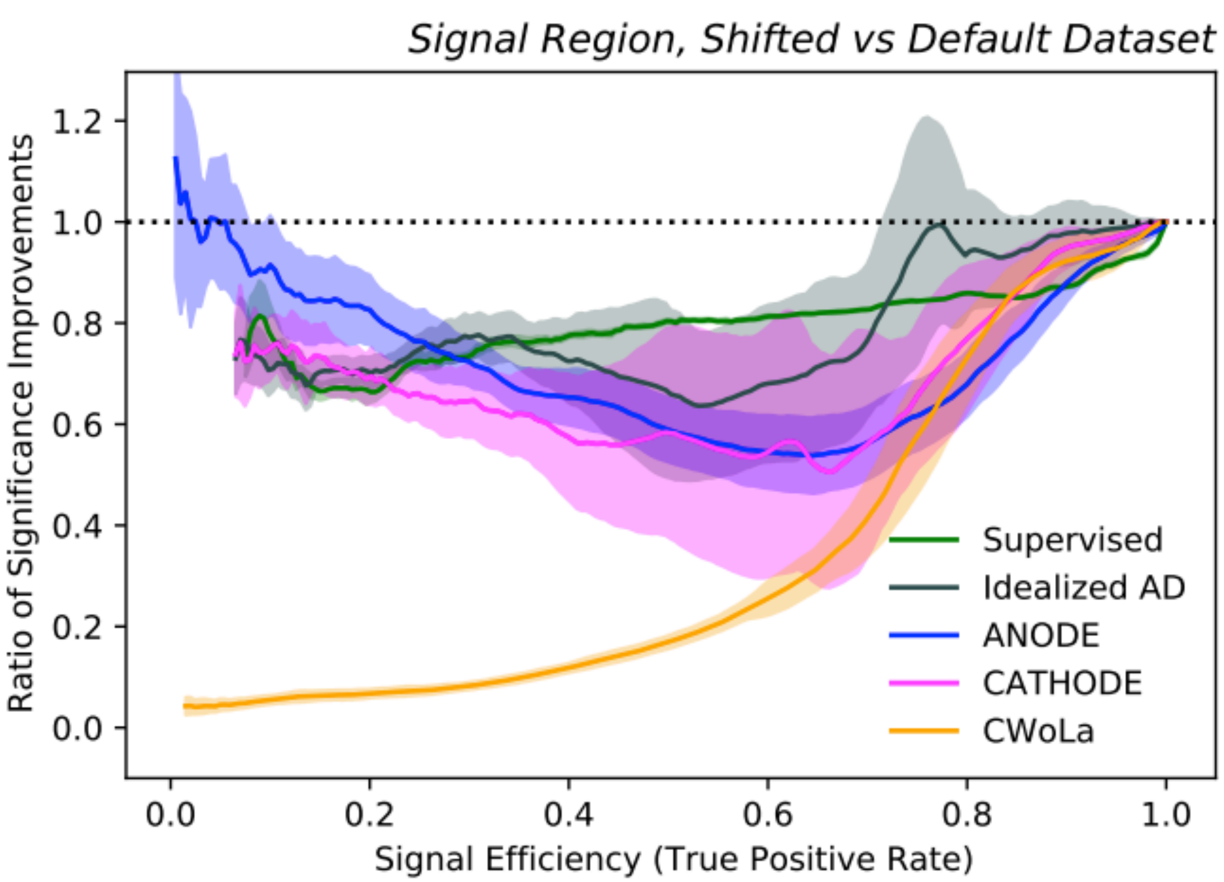
CATHODE

- CATHODE: Classifying Anomalies Through Outer Density Estimation
  - Training a conditional density estimator (Masked Autoregressive Flow) on the discriminant variables in the side-band
  - Interpolating it into the signal region and sampling from it
  - Train classifier: separate SR data from produced "background" sample
  - Anomaly detection: Apply the trained classifier to data in SR
- In real life: the CATHODE method would need to be combined with a background estimation procedure

arXiv:2109.00546





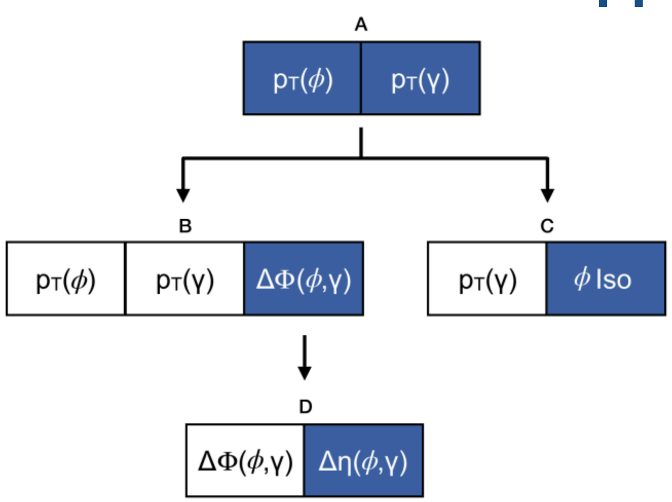


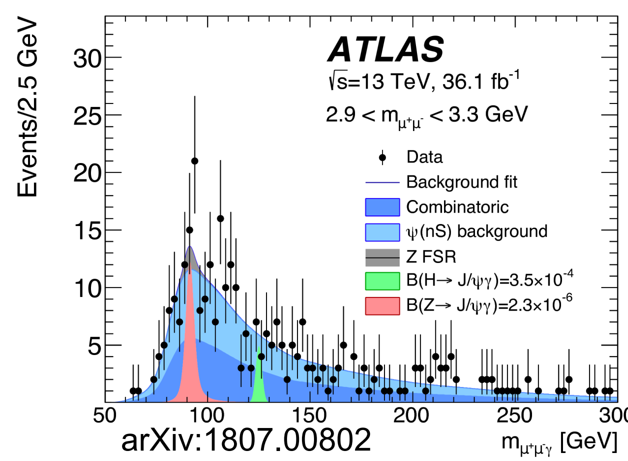
Summary

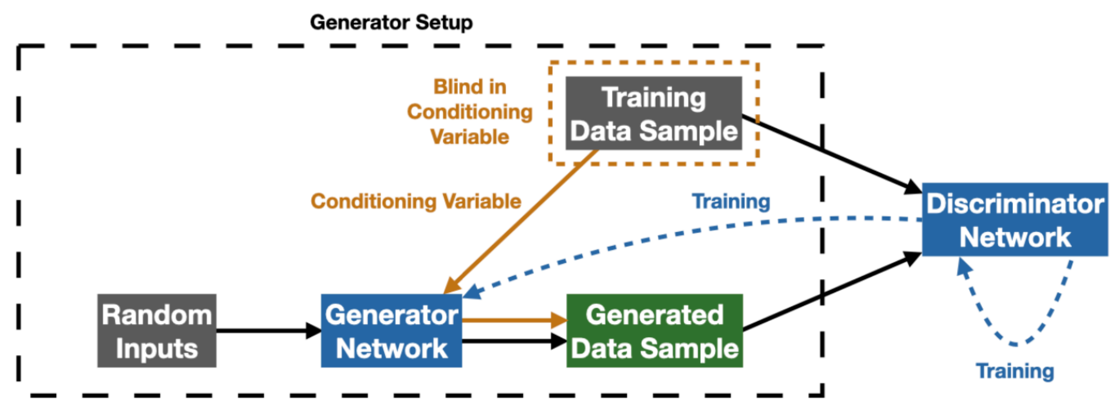
Background modelling crucial in searches for new physics and precision measurements
- Variety of methods has been developed
- Many rely on availability of large, reliable, simulated data samples
- Parametric methods suffer "spurious signal" type of effects

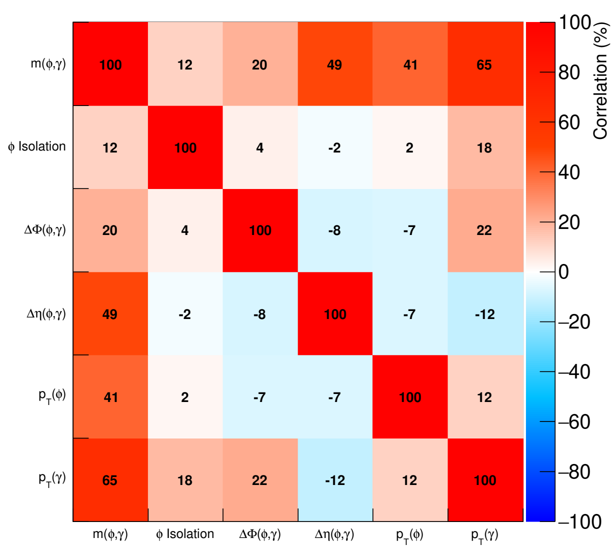
Developed **non-parametric, conditional probability-based, methods for data-driven modelling**:
- Histogram-based ancestral sampling method
- Machine learning technique using conditioned-Generative Adversarial Network

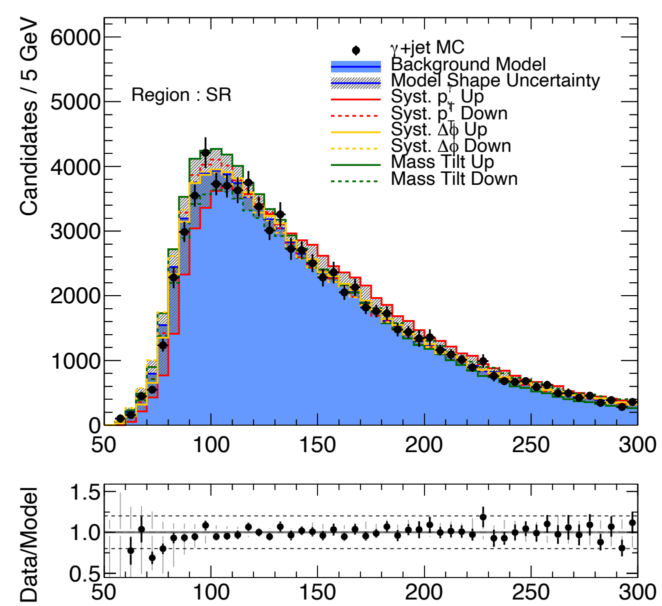
Presented methods applicable to any analysis!

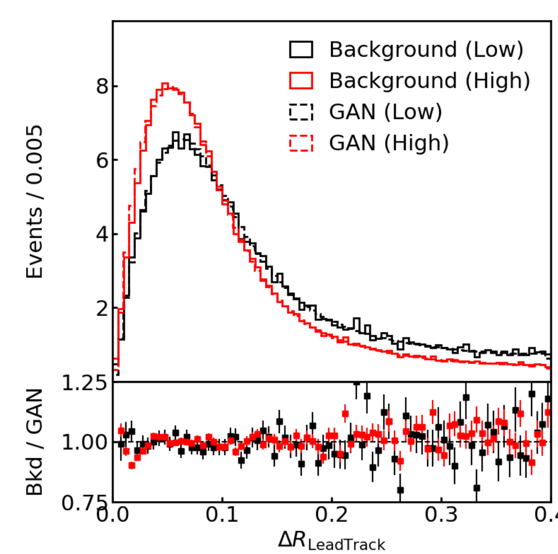


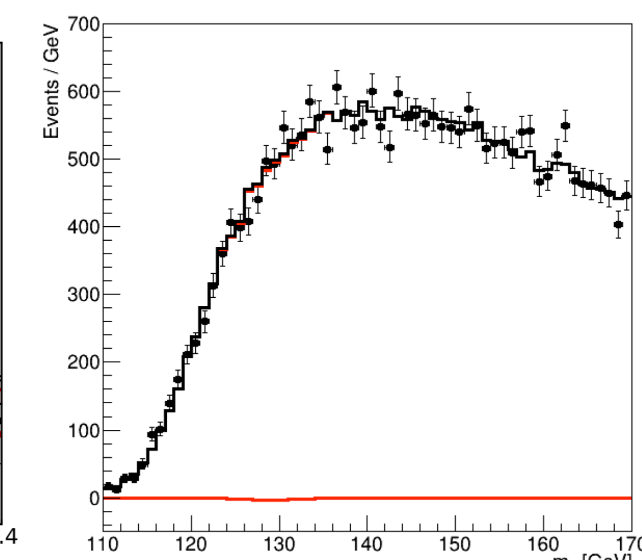












This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement 714893 (ExclusiveHiggs) and under Marie Skłodowska-Curie agreement 844062 (LightBosons)