



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

在 ATLAS 实验上寻找超出标准模型希格斯粒子

作者姓名: 周茂森

指导教师: 方亚泉 研究员 中国科学院高能物理研究所

学位类别: 理学博士

学科专业: 粒子物理与原子核物理

培养单位: 中国科学院高能物理研究所

2019 年 6 月

Search for the BSM Higgs boson with the ATLAS detector at LHC

**A thesis submitted to the
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor of Natural Science
in Particle Physics and Nuclear Physics**

By

Zhou Maosen

Supervisor: Professor Fang Yaquan

Institute of High Energy Physics, Chinese Academy of Sciences

June, 2019

中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期：

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关保存和使用学位论文的规定，即中国科学院大学有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

关键词： ATLAS, 超出标准模型希格斯粒子, ITk

Abstract

Keywords: ATLAS, BSM Higgs, ITk

目 录

| | |
|--------------------------------------|----|
| 第 1 章 数据和蒙特卡罗样本 | 1 |
| 1.1 数据样本 | 1 |
| 1.2 蒙特卡罗样本 | 1 |
| 1.2.1 信号样本 | 1 |
| 1.2.2 背景样本 | 2 |
| 第 2 章 事例筛选和信号优化 | 5 |
| 2.1 粒子鉴别及筛选 | 5 |
| 2.1.1 Object definitions in 4W | 5 |
| 2.1.2 Overlap removal | 5 |
| 2.2 事例筛选 | 6 |
| 2.2.1 初步筛选 | 6 |
| 2.2.2 信号优化 | 8 |
| 第 3 章 背景估计 | 13 |
| 3.1 QmisID 估计 | 15 |
| 3.1.1 似然函数方法 | 15 |
| 3.1.2 系统误差 | 16 |
| 3.2 Fakes 估计 | 18 |
| 3.2.1 Fake factor 方法 | 18 |
| 3.2.2 系统误差 | 21 |
| 3.2.3 总预期本底估计 | 24 |
| 第 4 章 信号优化 | 27 |
| 4.1 优化策略 | 27 |
| 4.2 信号效率 | 27 |
| 4.3 优化结果 | 27 |
| 参考文献 | 29 |
| 作者简历及攻读学位期间发表的学术论文与研究成果 | 31 |
| 致谢 | 33 |

图形列表

| | | |
|-----|--|----|
| 1.1 | ATLAS 数据收集情况: (a) 2015 年, (b) 2016 年。 | 1 |
| 2.1 | The final efficiency of pre-selections for hh signal samples. | 9 |
| 2.2 | The final efficiency of pre-selections for SS signal samples. | 10 |
| 2.3 | Distributions of p_T and number of jet for signal. Figure 2.3a to Figure 2.3e are distributions of p_T of jet before 25 GeV cuts, corresponding for $m_X=260, 300, 400, 500$ GeV and non-resonant signal. Two dashed vertical lines are $p_T=10$ GeV and $p_T=25$ GeV, respectively. Figure 2.3f is number of jet distribution after 25 GeV cuts. | 11 |
| 3.1 | The comparison between data and all MC backgrounds at pre-selection level. Left: ee , middle: $\mu\mu$, right: $e\mu$. The slashed pink bands are corresponding to statistical uncertainties. Each process is normalized to the luminosity of 36.1 fb^{-1} | 14 |
| 3.2 | The electron charge mis-identification rates as a function of (η , p_T) computed in data with the likelihood method. | 16 |
| 3.3 | The comparison between the charge mis-identification rates of electrons measured in simulated $Z \rightarrow e^+e^-$ events with the truth-matching method and the 2D likelihood method. | 17 |
| 3.4 | The systematic uncertainty on the charge mis-identification rate, for different bins in p_T and $ \eta $ | 18 |
| 3.5 | The comparisons between data and backgrounds at pre-selection level, corresponding to $N_{\text{jet}} \geq 2$. Left: ee , middle: $\mu\mu$, right: $e\mu$. The uncertainties, represented by slashed bands, include all systematics on fakes and statistical uncertainties on the other background components. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} | 26 |
| 3.6 | The comparisons between data and backgrounds at pre-selection level, corresponding to $N_{\text{jet}} \geq 3$. Left: ee , middle: $\mu\mu$, right: $e\mu$. The uncertainties, represented by slashed bands, include all systematics on fakes and statistical uncertainties on the other background components. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} | 26 |

表格列表

| | | |
|-----|--|----|
| 1.1 | Summary of the MC hh samples which have been produced for study. | 2 |
| 1.2 | Summary of the MC $X \rightarrow SS$ signal samples used. | 3 |
| 2.1 | 4W 物理分析粒子筛选条件总结 | 5 |
| 2.2 | Overlap removal in 4W analysis. | 6 |
| 2.3 | The event cleaning criteria. | 8 |
| 2.4 | The cutflow of pre-selection for non-resonant hh signal. The cross-section of $pp \rightarrow hh$ is 33.41 fb. The event yields are normalized to the luminosity of 36.1 fb^{-1} , corresponding to the final state of two-signed leptons. The statistical uncertainty is added in the last row. | 9 |
| 3.1 | definitions of tight electrons and anti-tight electrons. In addition to the inverted ID requirement, anti-tight electrons are required to pass the loose selection criteria. | 19 |
| 3.2 | definitions of tight muons and anti-tight muons. In addition to the inverted isolation requirement, anti-tight muons are required to pass the loose selection criteria. | 19 |
| 3.3 | Summary of different regions used to estimate fakes for low mass and high mass searches. | 19 |
| 3.4 | Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of electron in low mass search. Uncertainties are statistical. | 20 |
| 3.5 | Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of muon in low mass search. Uncertainties are statistical. | 20 |
| 3.6 | Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of electron in high mass search. Uncertainties are statistical. | 20 |
| 3.7 | Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of muon in high mass search. Uncertainties are statistical. | 20 |
| 3.8 | Summary of fake factors of electron and muon with different N_{jet} requirements. Uncertainties are statistical. | 21 |
| 3.9 | Observed number of data and expected events yields in high jet multiplicity region, which is used to predict fakes in low mass search. Uncertainties are statistical. | 21 |

| | | |
|------|---|----|
| 3.10 | Observed number of data and expected events yields in high jet multiplicity region, which is used to predict fakes in high mass search. Uncertainties are statistical. | 22 |
| 3.11 | Estimated jet fakes in three channels with different selections. Uncertainties are statistical. | 22 |
| 3.12 | Non-closure uncertainty on θ_e and θ_μ . To reduce the statistical error, SS, $p_T(\ell)$ and $M(\ell\ell) > 15$ GeV requirements are dropped in pre-selections. | 22 |
| 3.13 | The contribution from $t\bar{t}$ becomes bigger as more jets are required. W+jets and $t\bar{t}$ (semi-leptonic) MC samples are produced with the same generator (Sherpa). | 23 |
| 3.14 | The fake factors with and without b -jet. | 23 |
| 3.15 | Summary of systematic uncertainty on θ_e with different N_{jet} selections(in %). | 24 |
| 3.16 | Summary of systematic uncertainty on θ_μ with different N_{jet} selections(in %). | 24 |
| 3.17 | Event yields at pre-selection level, corresponding to $N_{\text{jet}} \geq 2$. The total uncertainties include all systematics on fakes and statistical uncertainties on the others. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} | 25 |
| 3.18 | Event yields at pre-selection level, corresponding to $N_{\text{jet}} \geq 3$. The total uncertainties include all systematics on fakes and statistical uncertainties on the others. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} | 25 |

符号列表

字符

| Symbol | Description | Unit |
|---------------|---|--|
| R | the gas constant | $\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$ |
| C_v | specific heat capacity at constant volume | $\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$ |
| C_p | specific heat capacity at constant pressure | $\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$ |
| E | specific total energy | $\text{m}^2 \cdot \text{s}^{-2}$ |
| e | specific internal energy | $\text{m}^2 \cdot \text{s}^{-2}$ |
| h_T | specific total enthalpy | $\text{m}^2 \cdot \text{s}^{-2}$ |
| h | specific enthalpy | $\text{m}^2 \cdot \text{s}^{-2}$ |
| k | thermal conductivity | $\text{kg} \cdot \text{m} \cdot \text{s}^{-3} \cdot \text{K}^{-1}$ |
| S_{ij} | deviatoric stress tensor | $\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$ |
| τ_{ij} | viscous stress tensor | $\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$ |
| δ_{ij} | Kronecker tensor | 1 |
| I_{ij} | identity tensor | 1 |

算子

| Symbol | Description |
|--------------|------------------------------------|
| Δ | difference |
| ∇ | gradient operator |
| δ^\pm | upwind-biased interpolation scheme |

缩写

| | |
|------|--------------------------------------|
| CFD | Computational Fluid Dynamics |
| CFL | Courant-Friedrichs-Lewy |
| EOS | Equation of State |
| JWL | Jones-Wilkins-Lee |
| WENO | Weighted Essentially Non-oscillatory |
| ZND | Zel'dovich-von Neumann-Doering |

第 1 章 数据和蒙特卡罗样本

1.1 数据样本

本分析利用 2015 年和 2016 年 ATLAS 探测器收集的质心系能量为 13 TeV 的数据，排除掉受损或者探测器未完全运作时的数据，其积分亮度为 36.1 fb^{-1} 。其中，2015 年和 2016 年的数据收集情况如图 1.1 所示。

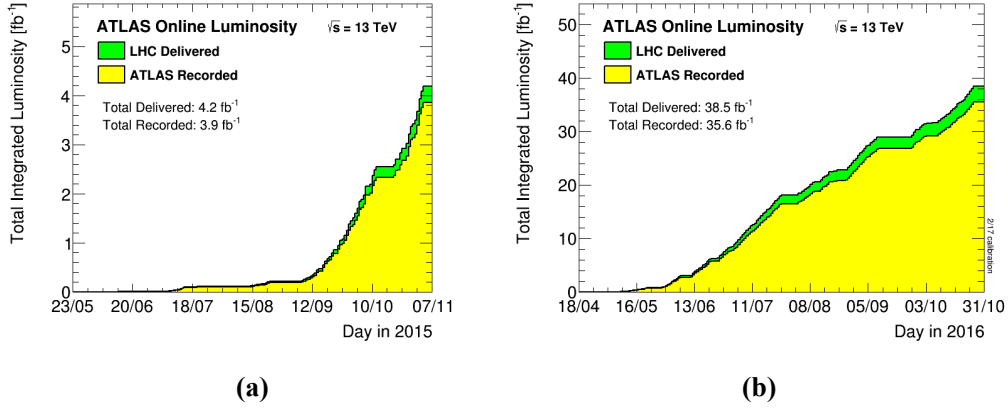


图 1.1 ATLAS 数据收集情况: (a) 2015 年, (b) 2016 年。

1.2 蒙特卡罗样本

1.2.1 信号样本

信号样本包含两种模型，分别为希格斯粒子对 ($gg \rightarrow (X) \rightarrow hh$) 和类希格斯粒子对 ($gg \rightarrow X \rightarrow SS$)。

1.2.1.1 $gg \rightarrow (X) \rightarrow hh$

1. SM 信号（非共振态模式），即 $gg \rightarrow hh$ ，利用包含 NLO 修正的双希格斯粒子模型 [1]，在 MadGraph5_aMC@NLO [2, 3] 中产生。其产生截面为 33.4 fb [4, 5]，考虑了 NNLO QCD 修正和 including resummation of soft-gluon emission at next-to-next-to-leading-logarithmic (NNLL) accuracy for $m_H = 125.09 \text{ GeV}$ 。

2. 共振态模式 ($gg \rightarrow X \rightarrow hh$) 也利用包含 NLO 修正的 2HDMCP_EFT 的信号模型 [6]，在 MadGraph5_aMC@NLO 中产生。其中重标量粒子 X，即共振态粒子，被假设具有远小于实验精度的衰变宽度；在实际模拟中，其宽度设为 10 MeV ，并考虑四个质量点，分别为 260 GeV , 300 GeV , 400 GeV 和 500 GeV 。此过程产生截面假设为 1 pb 。

| DSID | lepton charge | m_X [GeV] | Num. Events | Simulation | e/a/s/r/p-tags |
|--------|---------------|-------------|-------------|------------|---------------------------------|
| 344133 | ++ | Non-res | 500000 | AFII | e5060, a766, a821, r7676, p2949 |
| 344134 | -- | Non-res | 500000 | AFII | e5060, a766, a821, r7676, p2949 |
| 343704 | ++ | 260 | 100000 | AFII | e5234, a766, a821, r7676, p2949 |
| 343712 | -- | 260 | 100000 | AFII | e5234, a766, a821, r7676, p2949 |
| 343706 | ++ | 300 | 100000 | AFII | e5234, a766, a821, r7676, p2949 |
| 343714 | -- | 300 | 100000 | AFII | e5234, a766, a821, r7676, p2949 |
| 343709 | ++ | 400 | 100000 | AFII | e5153, a766, a821, r7676, p2949 |
| 343717 | -- | 400 | 100000 | AFII | e5153, a766, a821, r7676, p2949 |
| 343711 | ++ | 500 | 100000 | AFII | e5153, a766, a821, r7676, p2949 |
| 343719 | -- | 500 | 100000 | AFII | e5234, a766, a821, r7676, p2949 |

表 1.1 Summary of the MC hh samples which have been produced for study.

两个希格斯粒子均被要求衰变到 W 玻色子对，随后，其中两个 $W^+(W^-)$ 被要求衰变到轻子（包括 τ ），而另两个 $W^-(W^+)$ 则到强子对。这一系列衰变通过 Herwig++ [7] 实现，也包括随后的 showering 和强子化过程，其衰变分支比为 $BR(hh \rightarrow 4W \rightarrow \ell^+ \nu \ell^+ qqqq) = 4.4 \times 10^{-3}$ 。两种模式的信号样本的产生情况总结如表 1.1 所示。

1.2.1.2 $gg \rightarrow X \rightarrow SS$

$gg \rightarrow X \rightarrow SS$ 利用 Pythia 8 在 LO 阶产生，PDF 为 A14NNPDF2.3LO，模型为 HiggsBSM:gg2A3， X 和 S 均假设具有远小于实验分辨率的宽度，为各自质量的 1%。与 $gg \rightarrow (X) \rightarrow hh$ 类似， S 被要求衰变到两个 W 玻色子，其中两个 $W^+(W^-)$ 被要求衰变到轻子（包括 τ ），而另两个 $W^-(W^+)$ 则到强子对。随后的 showering 和强子化过程也由 Pythia 8 实现。 m_X 和 m_S 选择使得 $4W$ 末态能够最显著。同样的， $gg \rightarrow X \rightarrow SS$ 截面假设为 1 pb，而 $BR(S \rightarrow WW)$ 则依赖于 m_S ，即希格斯粒子在不同质量点的衰变分支比 [8]。表 1.2 总结了此信号样本产生情况。

1.2.2 背景样本

多玻色子 (VV/VVV) 和 $V\gamma$ 样本通过 Sherpa 2.1 [9] 在 NLO 阶产生； $V + jets$ 则通过 Sherpa 2.2 在 NLO 阶产生，此两种过程均采用 CT10 PDF。 VH 利用 Pythia 8 在 LO 阶产生，采用 NNPDF2.3LO PDF。 $t\bar{t}$ 通过 Powheg-Box 2.0 [10] 在 NLO 阶产生，而后传递到 Pythia 8 进行 parton showering 和强子化模拟，采用 PDF 为

| Charge | m_X | m_S | BR(two SS leptons) | DSID | N_{events} |
|--------|---------|---------|------------------------|--------|---------------------|
| ++ | 280 GeV | 135 GeV | 1.47×10^{-2} | 344927 | 25000 |
| | 300 GeV | 135 GeV | 1.535×10^{-2} | 344928 | 25000 |
| | 320 GeV | 135 GeV | 1.535×10^{-2} | 344930 | 25000 |
| | 340 GeV | 135 GeV | 1.535×10^{-2} | 344933 | 25000 |
| | 340 GeV | 145 GeV | 3.454×10^{-2} | 344934 | 25000 |
| | 340 GeV | 155 GeV | 6.049×10^{-2} | 344935 | 24000 |
| | 340 GeV | 165 GeV | 8.842×10^{-2} | 344936 | 25000 |
| -- | 280 GeV | 135 GeV | 1.47×10^{-2} | 344937 | 25000 |
| | 300 GeV | 135 GeV | 1.535×10^{-2} | 344938 | 25000 |
| | 320 GeV | 135 GeV | 1.535×10^{-2} | 344940 | 25000 |
| | 340 GeV | 135 GeV | 1.535×10^{-2} | 344943 | 25000 |
| | 340 GeV | 145 GeV | 3.454×10^{-2} | 344944 | 24000 |
| | 340 GeV | 155 GeV | 6.049×10^{-2} | 344945 | 25000 |
| | 340 GeV | 165 GeV | 8.842×10^{-2} | 344946 | 25000 |

表 1.2 Summary of the MC $X \rightarrow SS$ signal samples used.

NNPDF2.3LO。单顶夸克过程 ($t + X$) 同样通过 Powheg-Box 2.0 在 NLO 阶产生, 但传递到 Pythia 6.4 [11] 进行后续模拟, 采用 PDF 则为 CT10。 $t\bar{t}V$ 样本则在 NLO 阶通过 MadGraph5_aMC@NLO + Pythia 8 产生, 采用 PDF 为 NNPDF2.3LO。 $t\bar{t}H$ 样本通过 MadGraph5_aMC@NLO + Herwig++ 产生, PDF 为 NNPDF3.0 [12]。更多关于这些背景过程的产生及模拟过程可参考文献 [13–15]。

第 2 章 事例筛选和信号优化

2.1 粒子鉴别及筛选

2.1.1 Object definitions in 4W

粒子鉴别遵循 ATLAS 一般流程，如章节 ?? 所述。对于 4W 分析，进一步的粒子筛选条件总结如表 2.1 所示。

| 粒子 | 选择条件 | |
|-------|---|--|
| 电子 | Baseline | Tight |
| | $E_T > 10 \text{ GeV}$ $ \eta < 2.47$, 排除 $1.37 < \eta < 1.52$ 区间 LooseLH ID, Loose isolation $ z_0 \sin \theta < 0.5 \text{ mm}$, $d_0/\sigma(d_0) < 5$ | TightLH ID FixedCutTight $(E_T^{\text{cone20}}/p_T < 0.06, p_T^{\text{varcone20}}/p_T < 0.06)$ |
| μ | $p_T > 10 \text{ GeV}$ $ \eta < 2.5$ Loose ID, Loose isolation $ z_0 \sin \theta < 0.5 \text{ mm}$, $d_0/\sigma(d_0) < 3$ | Tight ID FixedCutTightTrackOnly $(p_T^{\text{varcone20}}/p_T < 0.06)$ |
| Jet | $p_T > 25 \text{ GeV}$, $ \eta < 2.5$ $ \text{JVT} < 0.59$ if $p_T < 60 \text{ GeV}$ and $ \eta < 2.4$ | |
| MET | $E_T^{\text{miss,TRK}}$ | |

表 2.1 4W 物理分析粒子筛选条件总结

2.1.2 Overlap removal

经过粒子基准（baseline）筛选之后，为了进一步保证没有误重建的重复粒子，专门的 overlap removal 需要完成。该分析中的 overlap removal 总结如表 2.2 所示。

| Keep | Remove | Cone size (ΔR) |
|----------|------------------------|---|
| muon | electron | 0.1 |
| electron | electron(lower p_T) | 0.1 |
| electron | jet | 0.3 |
| jet | muon | $\min(0.4, 0.04+10[\text{GeV}]/p_T(\mu))$ |

表 2.2 Overlap removal in 4W analysis.

2.2 事例筛选

2.2.1 初步筛选

所选事例应当通过如下初步筛选条件：

- **GRL**

2015 data: data15_13TeV.periodAllYear_DetStatus-v79-repro20-02
_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml

2016 data: data16_13TeV.periodAllYear_DetStatus-v88-pro20-21
_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml

- **Event cleaning criteria:** cleaning for Tile corrupted events, LAr noise bursts and corrupted data

- **Vertex criteria:** events are required to contain at least one primary vertex with ≥ 2 associated tracks. The detailed selection on the vertex can be found in [?]

- **Trigger:**

对于 2015 年数据，满足以下任一 trigger：

- Single lepton triggers:
 - * HLT_mu20_loose_L1MU15
 - * HLT_mu50
 - * HLT_e24_lhmedium_L1EM20VH
 - * HLT_e60_lhmedium
 - * HLT_e120_lhloose
- Dilepton triggers:
 - * HLT_2e12_lhloose_L12EM10VH
 - * HLT_e17_lhloose_mu14
 - * HLT_mu18_mu8noL1

对于 2016 年数据，满足以下任一 trigger：

- Single lepton triggers:

- * HLT_mu24_ivarmedium
- * HLT_mu50
- * HLT_e24_lhtight_nod0_ivarloose
- * HLT_e60_lhmedium_nod0
- * HLT_e140_lhloose_nod0
- Dilepton triggers:
- * HLT_2e17_lhvloose_nod0
- * HLT_e17_lhloose_nod0_mu14
- * HLT_mu22_mu8noL1

与数据一样，模拟样本也应当满足以上 trigger 条件，其相应的 trigger 效率修正已添加到每个样本事例中

- 选择通过章节 2.1.1 所述的粒子。
- 轻子数：
 - 两个相同电荷的轻子。
 - 每个 tight 电子应当满足 $\text{ChargeIDBDTTight} > 0.067$ ，此变量是用来压低“假”电子，如附录 ?? 所述。
 - 至少有一个轻子应当能匹配以上任一或多个 trigger，除此之外，大横动量轻子 p_T 应大于 30 GeV，小横动量轻子大于 20 GeV。
- 排除掉任何含有 b -jet 的事例。
- $E_T^{\text{miss,TRK}} > 10 \text{ GeV}$ 。
- 因为 Drell-Yan 过程目前并不能被 MC 很好模拟，所以为了避免此问题，双轻子不变质量应大于 15 GeV。
- 为了压低来自于 Z +jets 过程的本底（电荷误判）， $|M(\ell\ell) - M(Z)| > 10 \text{ GeV}$ 条件须通过。
- jet 数的要求依赖于质量点的选择，低（高）质量点要求至少 2（3）个 jet。此项会在 ?? 深入讨论。

以上的事例筛选过程总结在表 2.3。最后，通过以上筛选条件的事例根据轻子味道分为三个分析道，为 $ee \mu\mu$ 和 $e\mu$ 。表 2.4 展示了标准模型希格斯对信号经过以上一系列条件时的事例数和效率变化，此处对应亮度为 36.1 fb^{-1} ，截面 ($gg \rightarrow hh$) 为 33.4 fb。图 2.1（图 2.2）展示所有 hh (SS) 信号样本的经过初步筛选之后的效率，可以看到：一是随着 m_X 或者 m_S 的增加，效率相应增加（对于 SS ，在 $m_X = 340 \text{ GeV}$, $m_S = 135 \text{ GeV}$ 质量点的效率下降是因为从此点开始要求至少三个

| | |
|----------------|--|
| Pre-selections | GRL |
| | Event clean criteria |
| | Pass any trigger applied |
| | Select objects following object definitions |
| | Overlap removal |
| | Two tight same-signed leptons, with at least one trigger matched |
| | $p_T(\ell_1) > 30 \text{ GeV}, p_T(\ell_2) > 20 \text{ GeV}$ |
| | b veto |
| | $E_T^{miss} > 10 \text{ GeV}$ |
| | $M(\ell\ell) > 15 \text{ GeV}$ |
| | $ M(\ell\ell) - M(Z) > 10 \text{ GeV}$ in ee channel |
| | $N_{\text{jet}} \geq 2(3)$ |

表 2.3 The event cleaning criteria.

jet)；二是 $e\mu$ 道具有最高的效率值， $\mu\mu$ 次之， ee 最低，这是因为理论上 $e\mu$ 的分支比是其他两个道的两倍，以及 μ 比 e 具有更好的鉴别效率。

2.2.2 信号优化

2.2.2.1 hh 信号优化

本分析中的显著信号是两个相同电荷的轻子和 jet 数。两个希格斯粒子倾向于出射到两个相反的半球，随后，两个希格斯粒子均衰变到 W 玻色子，总共四个 W 玻色子中有两个是不在壳的，而不在壳的 W 玻色子会贡献相当部分的低动量的 jet ($p_T < 25 \text{ GeV}$)。在图 2.3a到2.3e中可以看到（此图未通过初步筛选条件），很大一部分的第四条 jet p_T 是低于 25 GeV 的，甚至在高质量信号点。那么加上基本的筛选条件之后，大部分信号事例只有三条 jet，如图 2.3f所示。同时可以发现，对于低质量点，即 $m_X=260 \text{ GeV}$ 和 $m_X=300 \text{ GeV}$ ，其大部分事例最多只有 2 条 jet。所以，对于不同的质量点，应当应用不同的 jet 数条件，对于低质量点，要求 $N_{\text{jet}} \geq 2$ ，而高质量点， $N_{\text{jet}} \geq 3$ 。为了证实该分类能够给出最高的信号显著性，考虑本底后，详细检查可见附录 ?? 最后，为了提高信号显著性，一系列动力学被重建，从而用来优化信号，具体优化方法会在章节 ?? 具体讨论，以下列出一些具有区分度的变量：

- $M(l\bar{l})$, the invariant mass of two same-signed leptons;
- MET , missing transverse energy;
- $M(jj)^W$, the invariant mass of two closest jets among all selected good jets;

| Cut flow | Event yield | | | Efficiency | | |
|------------------------------|------------------|------------------|------------------|------------|----------|--------|
| Evgen | - | | | 100% | | |
| HIGG8D1 | 2.76 | | | 56.34% | | |
| Event cleaning | 2.76 | | | 56.34% | | |
| Trigger | 2.10 | | | 44.84% | | |
| Channel | ee | $\mu\mu$ | $e\mu$ | ee | $\mu\mu$ | $e\mu$ |
| OB, OLR | 0.29 | 0.28 | 0.56 | 5.86% | 6.23% | 11.96% |
| Tight leptons, trigger match | 0.14 | 0.20 | 0.33 | 2.33% | 3.46% | 5.68% |
| $p_T(\ell)$ | 0.11 | 0.15 | 0.24 | 1.93% | 2.70% | 4.53% |
| b veto | 0.10 | 0.14 | 0.23 | 1.79% | 2.49% | 4.18% |
| MET | 0.10 | 0.14 | 0.22 | 1.76% | 2.45% | 4.10% |
| Drell-Yan cut | 0.10 | 0.14 | 0.22 | 1.76% | 2.44% | 4.10% |
| Z veto | 0.08 | 0.14 | 0.22 | 1.58% | 2.44% | 4.10% |
| $N_{\text{jet}} \geq 3$ | 0.05 ± 0.002 | 0.09 ± 0.002 | 0.14 ± 0.003 | 1.03% | 1.92% | 2.99% |

表 2.4 The cutflow of pre-selection for non-resonant hh signal. The cross-section of $pp \rightarrow hh$ is 33.41 fb. The event yields are normalized to the luminosity of 36.1 fb^{-1} , corresponding to the final state of two-signed leptons. The statistical uncertainty is added in the last row.

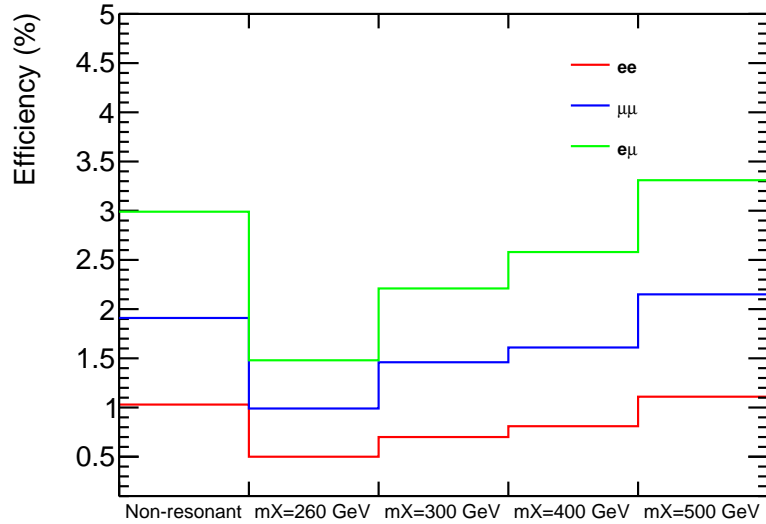


图 2.1 The final efficiency of pre-selections for hh signal samples.

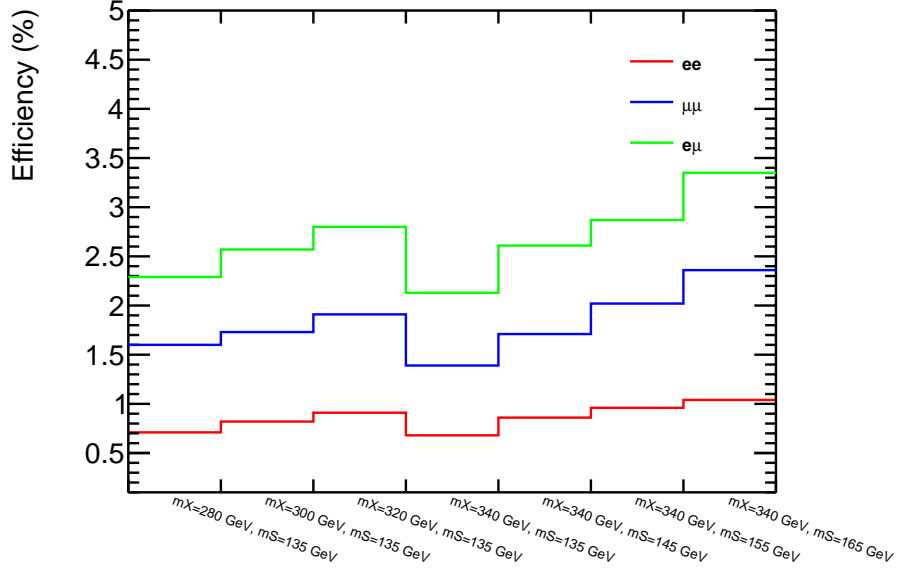


图 2.2 The final efficiency of pre-selections for SS signal samples.

- $M(l_1 jj)$, the invariant mass of leading lepton and two closest jets;
- $M(all)$, the invariant mass of all selected objects;
- M_T , the transverse mass of all selected objects;
- $\Delta R_{min}(\ell_1, j)$, ΔR distance between leading lepton and the closest jet;
- $\Delta R_{min}(\ell_2, j)$, ΔR distance between sub leading lepton and the closest jet;

2.2.2.2 SS 信号优化

S 标量粒子所取质量从 135 GeV 到 165 GeV, X 粒子从 280 GeV 到 340 GeV。 SS 与 hh 具有类似的动力学性质, 为了尽可能增加信号显著性, N_{jet} 分类适用于此, 具体如下:

- 固定 $m_S = 135$ GeV: $m_X = 280$ GeV, $m_X = 300$ GeV and $m_X = 320$ GeV; $N_{jet} \geq 2$ 。
- 固定 $m_X = 340$ GeV: $m_S = 135$ GeV, $m_S = 145$ GeV, $m_S = 155$ GeV and $m_S = 165$ GeV; $N_{jet} \geq 3$ 。

前述章节 2.2.2.1 的动力学变量也可用来进一步优化信号显著性。

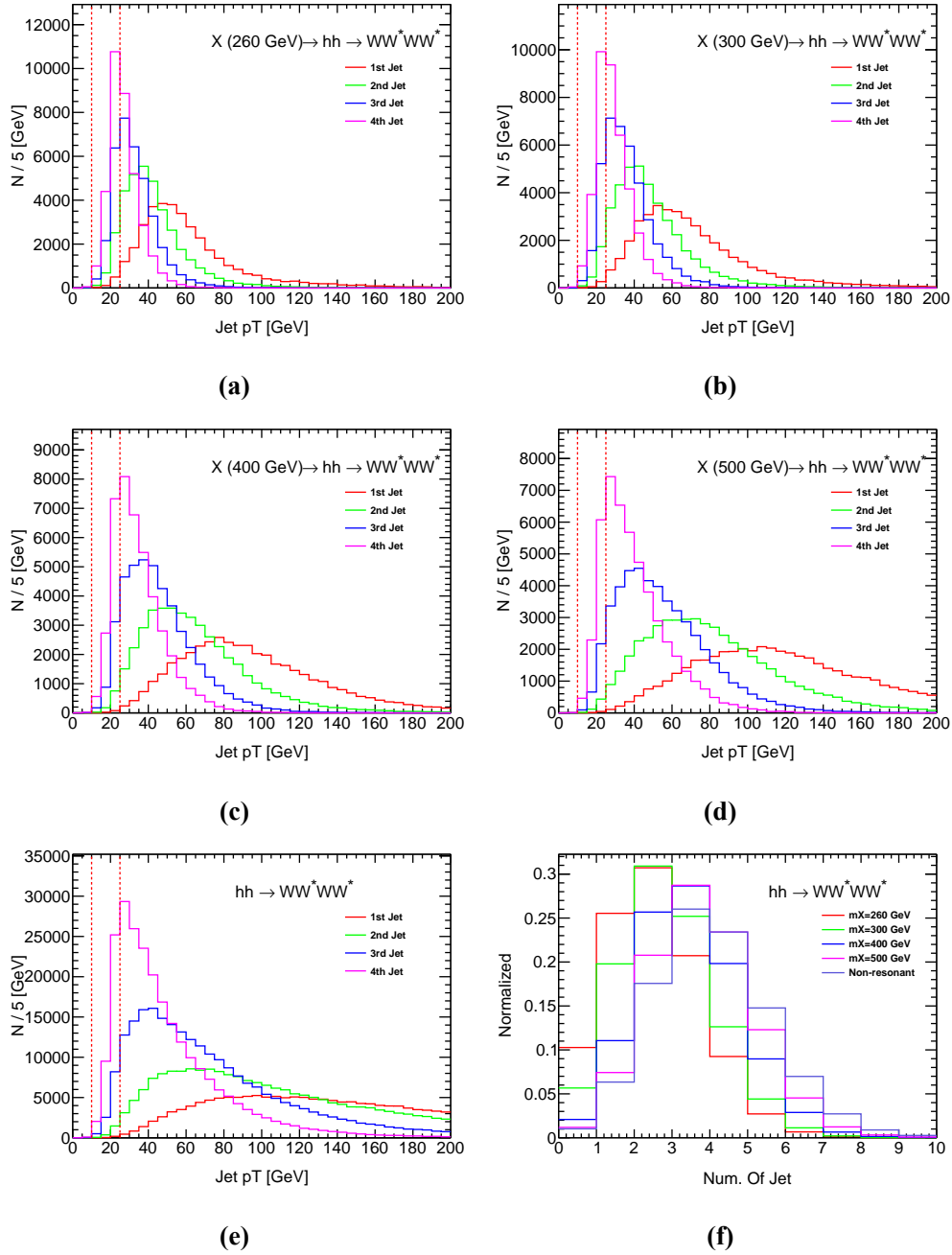


图 2.3 Distributions of p_T and number of jet for signal. Figure 2.3a to Figure 2.3e are distributions of p_T of jet before 25 GeV cuts, corresponding for $m_X=260, 300, 400, 500$ GeV and non-resonant signal. Two dashed vertical lines are $p_T=10$ GeV and $p_T=25$ GeV, respectively. Figure 2.3f is number of jet distribution after 25 GeV cuts.

第 3 章 背景估计

4W 物理分析背景分为可以贡献两个相同电荷的过程 (promptSS), 电荷误鉴别 (QmisID) 和“假”轻子过程 (fakes)。promptSS 主要来自 $t\bar{t}V$, VV , tV 以及 $t\bar{t}H$ 过程, 该背景可用 MC 估计。QmisID 一般来自于 $Z + jets$ 和 $t\bar{t}$ (轻子衰变过程)。fakes 来自于 $W + jets$, $t\bar{t}$ (半轻子衰变) 过程, 其中一个 jet 被误判成轻子或者一个轻子来源于 b -jet (non-prompt)。目前 ATLAS MC 不能很好地描述 QmisID 和 fakes, 如图 3.1 所示, 如果所有背景均用 MC 模拟, 可以看到, 数据跟预期有非常大的偏差。所以, 这表明 data-driven 的方法去估计 QmisID 和 fakes 是有必要的。在以下章节中, 将分别讲述 QmisID 和 fakes 的估计方法。

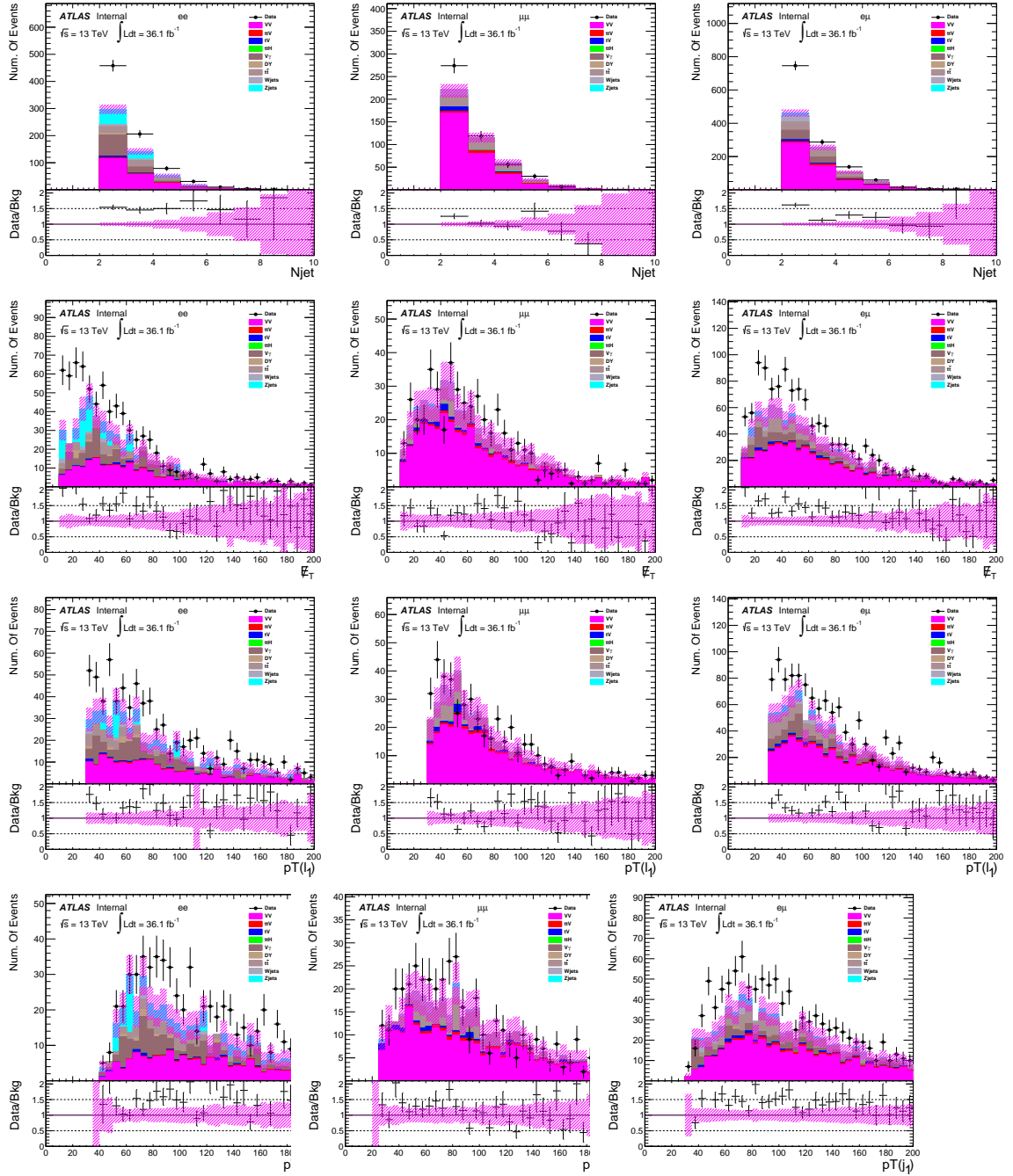


图 3.1 The comparison between data and all MC backgrounds at pre-selection level. Left: ee , middle: $\mu\mu$, right: $e\mu$. The slashed pink bands are corresponding to statistical uncertainties. Each process is normalized to the luminosity of 36.1 fb^{-1} .

3.1 QmisID 估计

双轻子为同电荷和同味道双轻子的 Z veto 的选择条件，会极大地压低 $t\bar{t}$ 和 $Z + jets$ 背景，但如果一个电荷误判，即使很低的误判率，但考虑到这两种过程的极大截面，仍有相当一部分的 QmisID 会贡献到最终背景中。根据 8 TeV 的研究 [16]， μ 电荷误判率非常低（一般低于 10^{-5} ）¹，所以只考虑电子电荷误判背景。

电子电荷误判有两种原因：

- 当电子穿过探测器材料时，出射一个光子（韧致辐射）；而后这个光子转换成一对正负电子，然而在随后的径迹重建中，带相反电荷的电子被利用，从而导致电荷误判。韧致辐射依赖于探测器材料密度，而探测器材料密度随着 $|\eta|$ 变化，所以电子的电荷误判率也依赖于 $|\eta|$ 。
- 第二种贡献相对来讲比较小，主要是因为测量精度不够，当电子径迹的曲率很小或者内部探测器径迹与量能器的簇射匹配错误时，得到完全相反曲率的径迹，从而导致电荷误判。所以，当电子具有高横动量时这个效应比较明显，那么误判率也依赖于 p_T 。

3.1.1 似然函数方法

一般假设电子电荷误判率不依赖于产生模式，因为 Z 玻色子产生截面大，而且其不变质量峰重建比较好，所以 Z 过程可以用来测量电子电荷误判率，采用似然函数技术 [17]，构建的似然函数如下：

$$\ln L(\varepsilon|N_{SS}, N) = \sum_{i,j} \ln[N^{ij}(\varepsilon_i + \varepsilon_j - 2\varepsilon_i\varepsilon_j)]N_{SS}^{ij} - N^{ij}(\varepsilon_i + \varepsilon_j - 2\varepsilon_i\varepsilon_j) \quad (3.1)$$

其中， ε_i 和 ε_j 分别为 $\eta - p_T$ 二维区间中第 i 个和第 j 个电子的误判率， N_{SS} 和 N 是观察到的相同电荷事例数和总事例数。 $\eta - p_T$ 区域总共分为 28 个小区域，其 $|\eta|$ 和 p_T 的分界线分别为 $[0., 0.60, 1.1, 1.37, 1.52, 1.70, 2.00, 2.47]$ 和 $[10, 60, 90, 130, 1000]$ GeV。随后，为了得到电子电荷误判率，需按照以下过程进行：

1. 按照 Z 玻色子过程筛选双轻子数据，其中轻子质量要求应与章节2所述一致。

¹The rate of charge mis-identification for muons is only affected by the track curvature. Because of the long lever arm to the muon system and the fact that the charge is measured in both the inner detector and muon spectrometer the mis-identification rates of the muon charge are very low, making this background negligible compared to the other sources of background

2. 在所选数据中通过高斯函数拟合 Z 不变质量谱，得到 Z 不变质量拟合值 κ 和标准偏差 σ 。而后分为如下三个区间，

| A | B | C |
|--|------------------------|--|
| $(\kappa - 8\sigma, \kappa - 4\sigma)$ | $(\kappa \pm 4\sigma)$ | $(\kappa + 4\sigma, \kappa + 8\sigma)$ |

3. 为了进一步提高 Z 玻色子纯度，剩余本底会被减去，即 $N_Z = n_B - \frac{n_A + n_C}{2}$ 。

4. 所选数据根据电荷分为 SS 和 OS，利用最大似然函数法得到各个 $|\eta| - p_T$ 区间的误判率。

图3.2展示电子电荷误判率随着 $|\eta|$ 增加而增大，因为在高 $|\eta|$ 粒子会穿过更多探测器材料；误判率也随着 p_T 增大而增大，这与前面的讨论一致。最后，利用公式 3.2 计算出 ee 和 $e\mu$ 道经过初步筛选后，在 $N_{\text{jet}} \geq 2$ 时 ($N_{\text{jet}} \geq 3$) 的误判事例数分别为 101.47 ± 0.60 (35.60 ± 0.38) 和 18.21 ± 0.23 (8.38 ± 0.16)。

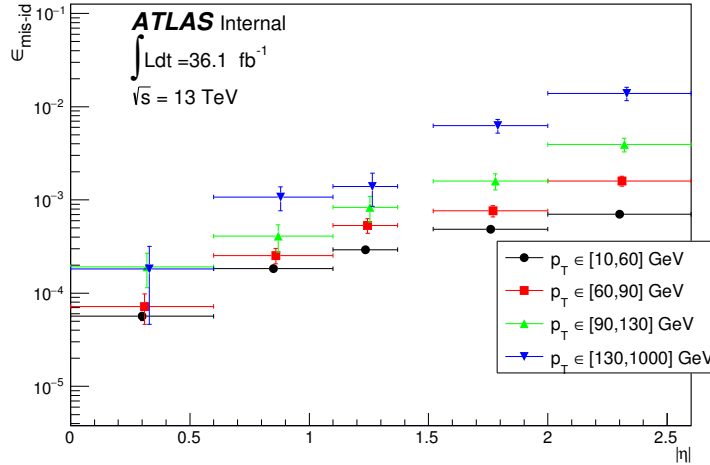


图 3.2 The electron charge mis-identification rates as a function of $(|\eta|, p_T)$ computed in data with the likelihood method.

$$N_{ee}^{\text{QmisID}} = \frac{\varepsilon_i + \varepsilon_j - 2\varepsilon_i\varepsilon_j}{1 - \varepsilon_i - \varepsilon_j + 2\varepsilon_i\varepsilon_j} N^{\text{OS}}, N_{e\mu}^{\text{QmisID}} = \frac{\varepsilon}{1 - \varepsilon} N^{\text{OS}} \quad (3.2)$$

3.1.2 系统误差

在 QmisID 估计中，考虑了三种系统误差：

1. 每个 $|\eta| - p_T$ 小区域中的统计误差；
2. 为了证实似然函数方法的可靠性，可以利用 Z MC 进行以上的误判率估计，因为在 MC 中，电子误判率是已知的，从而比较似然函数估计值与真实值就

可判断该方法的可靠程度（closure test）。图 3.3 比较了真实值与似然函数估计值，总体上是一致的。该项差距将作为 QmisID 的系统误差之一。

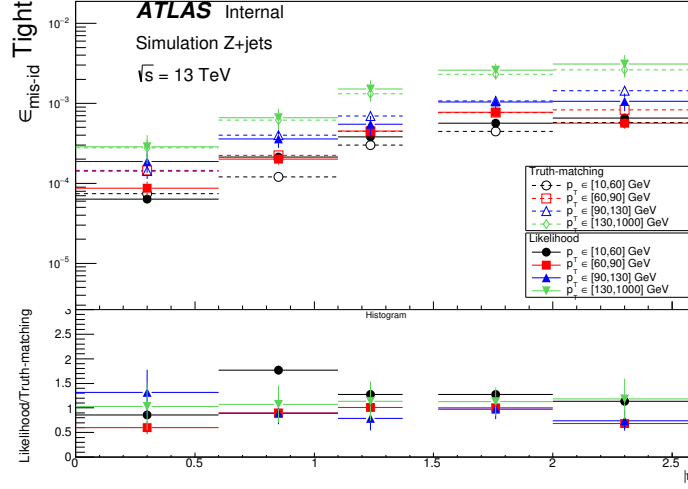


图 3.3 The comparison between the charge mis-identification rates of electrons measured in simulated $Z \rightarrow e^+e^-$ events with the truth-matching method and the 2D likelihood method.

3. Z 峰区间的变动会影响 QmisID 的估计，所以，如果偏移 Z 峰 1σ ，其 QmisID 率的相对变化考虑成系统误差。

图 3.4 总结了几种系统误差在不同 $|\eta| - p_T$ 的相对大小，随着 p_T 的增加统计误差越来越大，因为大部分电子是低动量的，其次是似然函数误差在低动量区更显著。

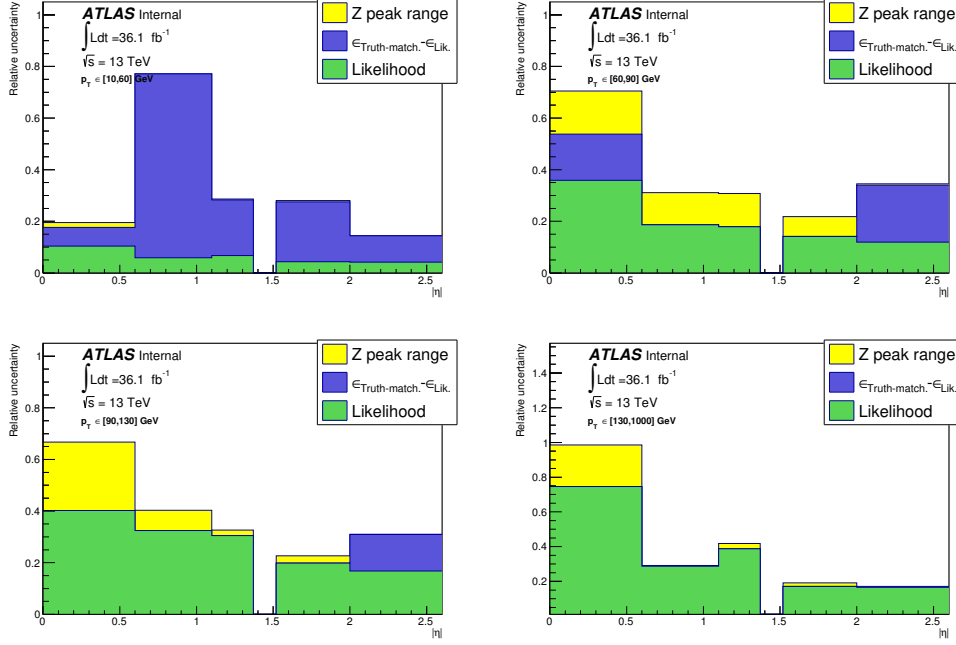


图 3.4 The systematic uncertainty on the charge mis-identification rate, for different bins in p_T and $|\eta|$.

3.2 Fakes 估计

误鉴别粒子是一项非常重要的本底，因为对其形成机制理解不够准确，MC 并不能很好地描述，所以有必要使用 **data-driven** 的方法去估计该项本底。在该分析中，我们使用 **fake factor** 的方法。

3.2.1 Fake factor 方法

在 4W 分析中，Fake factor 方法的假设是 fake factor 不依赖于 jet 数，其定义是具有两个 **tight** SS 轻子的事例数与具有一个 **tight** 和一个 **anti-tight** 的轻子的事例数比例，如等式 3.3 所示。

$$\theta_\ell = \frac{N_{\ell\ell}}{N_{\ell\bar{\ell}}} \quad (3.3)$$

其中， ℓ 为 **tight** 电子或者 μ 子， $\bar{\ell}$ 是 **anti-tight** 的轻子。通常，分母的选择是 fake factor 中最困难的部分：分母的选择应当使得真实轻子极大地压低，而尽量增大误鉴别轻子的比例。如果分母选择条件越严格，与外延相关的系统误差越小，但是另一方面，事例越少，相应的统计误差越大。所以，为了优化整体的系统误差，必须考虑到这些相反的影响。一般来讲，利用 **ID** 和 **isolation** 条件可以很好地压低误鉴别的电子，而 **isolation** 和碰撞参数可以用来压低误鉴别 μ 子 [18]。在本分析中， ℓ 和 $\bar{\ell}$ 的定义如表 3.1 和表 3.2 所示。

| | tight electron | anti-tight electron |
|-----------|------------------------|------------------------|
| ID | TightLH | fail TightLH |
| isolation | isolationFixedCutTight | - |
| QmisID | ChargeIDBDTTight>0.067 | ChargeIDBDTTight>0.067 |

表 3.1 definitions of tight electrons and anti-tight electrons. In addition to the inverted ID requirement, anti-tight electrons are required to pass the loose selection criteria.

| | tight muon | anti-tight muon |
|-----------|---------------------------------|--------------------------------------|
| ID | Tight | - |
| isolation | isolationFixedCutTightTrackOnly | fail isolationFixedCutTightTrackOnly |

表 3.2 definitions of tight muons and anti-tight muons. In addition to the inverted isolation requirement, anti-tight muons are required to pass the loose selection criteria.

值得指出的是，因为在本分析中，有两个 N_{jet} 类别，所以对于低质量点，计算 fake factor 时要求一个 jet；对于高质量点，要求 1 到 2 个 jet。总结如表 3.3 所示。接下来，作为例子，我们将只展示高质量点的 fakes 的计算方式。之前提过，

| | Region | Fake factor CR (low jet multiplicity region) | SR (high jet multiplicity region) |
|-----------|---|--|-----------------------------------|
| Low mass | $hh: m_X=260, 300 \text{ GeV}$ $SS: \text{Fixing } m_S=135 \text{ GeV}, m_X=280, 300, 320 \text{ GeV}$ | $N_{\text{jet}}=1$ | $N_{\text{jet}} \geq 2$ |
| High mass | $hh: m_X=400, 500 \text{ GeV, no-resonant}$ $SS: \text{Fixing } m_X=340 \text{ GeV}, m_S=135, 145, 155 \text{ and } 165 \text{ GeV}$ | $1 \leq N_{\text{jet}} \leq 2$ | $N_{\text{jet}} \geq 3$ |

表 3.3 Summary of different regions used to estimate fakes for low mass and high mass searches.

本底包括 fakes, QmisID, $V\gamma$ 和 promptSS，为了不重复考虑这些本底，实际计算 fake factor 时应当减去这些本底，如公式 3.5 所示。

$$\theta_e(1 \leq N_{\text{jet}} \leq 2) = \frac{N_{ee}^{\text{data}} - N_{ee}^{\text{promptSS}} - N_{ee}^{V\gamma} - N_{ee}^{\text{QmisID}}}{N_{e\cancel{e}}^{\text{data}} - N_{e\cancel{e}}^{\text{promptSS}} - N_{e\cancel{e}}^{V\gamma} - N_{e\cancel{e}}^{\text{QmisID MC}}}(1 \leq N_{\text{jet}} \leq 2) \quad (3.4)$$

$$\theta_\mu(1 \leq N_{\text{jet}} \leq 2) = \frac{N_{\mu\mu}^{\text{data}} - N_{\mu\mu}^{\text{promptSS}} - N_{\mu\mu}^{V\gamma}}{N_{\mu\cancel{\mu}}^{\text{data}} - N_{\mu\cancel{\mu}}^{\text{promptSS}} - N_{\mu\cancel{\mu}}^{V\gamma}}(1 \leq N_{\text{jet}} \leq 2) \quad (3.5)$$

可以看到，promptSS, QmisID 和 $V\gamma$ 在分子分母中均被减去。promptSS 用 MC 估计，并且要求其中一个轻子能够匹配到真实轻子 (truth-matching)。 $V\gamma$ 也利用

MC 估计, 但是并不要求 truth-matching, 因为其中一个轻子很可能来自于 γ 。对于 QmisID, 分子中的 N_{ee}^{QmisID} 计算如章节 3.1 所述, 而 $N_{e\cancel{e}}^{\text{QmisID MC}}$ 直接利用 MC 估计, 其中要求电子 truth-matching。表 ?? 到表 ?? 总结在不同 N_{jet} 类别下, 用来计算 fake factor 各个成分的值。

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | QmisID | Data |
|----------------------|---------------|--------------------|-----------------|-----------------|-----------------|--------------------|--------------------|------|
| $N_{\text{jet}} = 1$ | ee | 204.64 \pm 19.13 | 1.09 \pm 0.08 | 5.08 \pm 0.93 | 0.03 \pm 0.01 | 135.94 \pm 12.84 | 164.46 \pm 0.65 | 976 |
| | $e\cancel{e}$ | 44.26 \pm 3.51 | 0.13 \pm 0.03 | 8.25 \pm 1.32 | 0.00 \pm 0.00 | 67.33 \pm 10.49 | 135.54 \pm 71.62 | 1116 |

表 3.4 Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of electron in low mass search. Uncertainties are statistical.

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | Data |
|----------------------|-------------------|-------------------|-----------------|------------------|-----------------|-----------------|------|
| $N_{\text{jet}} = 1$ | $\mu\mu$ | 296.37 \pm 9.72 | 1.92 \pm 0.11 | 5.91 \pm 1.01 | 0.02 \pm 0.02 | 0.00 \pm 0.00 | 455 |
| | $\mu\cancel{\mu}$ | 56.84 \pm 5.00 | 0.13 \pm 0.03 | 20.80 \pm 2.34 | 0.00 \pm 0.00 | 0.63 \pm 0.45 | 378 |

表 3.5 Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of muon in low mass search. Uncertainties are statistical.

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | QmisID | Data |
|--------------------------------|---------------|--------------------|-----------------|------------------|-----------------|--------------------|--------------------|------|
| $1 \leq N_{\text{jet}} \leq 2$ | ee | 309.38 \pm 19.75 | 3.67 \pm 0.16 | 11.27 \pm 1.47 | 0.10 \pm 0.02 | 213.30 \pm 17.29 | 230.40 \pm 0.81 | 1434 |
| | $e\cancel{e}$ | 66.58 \pm 5.19 | 0.39 \pm 0.06 | 15.85 \pm 1.89 | 0.02 \pm 0.01 | 104.00 \pm 12.71 | 187.16 \pm 78.65 | 1591 |

表 3.6 Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of electron in high mass search. Uncertainties are statistical.

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | Data |
|--------------------------------|-------------------|--------------------|-----------------|------------------|-----------------|-----------------|------|
| $1 \leq N_{\text{jet}} \leq 2$ | $\mu\mu$ | 463.01 \pm 11.61 | 6.14 \pm 0.21 | 15.20 \pm 2.26 | 0.17 \pm 0.03 | 0.01 \pm 0.01 | 729 |
| | $\mu\cancel{\mu}$ | 74.30 \pm 5.40 | 0.45 \pm 0.06 | 43.59 \pm 3.37 | 0.02 \pm 0.01 | 1.62 \pm 0.74 | 658 |

表 3.7 Observed number of data and expected events yields in low jet multiplicity region, which is used for fake factor calculation of muon in high mass search. Uncertainties are statistical.

最终, 根据公式 3.5, 得到各种类别的 fake factor, 总结在表 3.8。接下来, 就

| Selections | Fake factor | Value |
|--------------------------------|--------------|---------------------|
| $N_{\text{jet}} == 1$ | θ_e | 0.5401 ± 0.0311 |
| | θ_μ | 0.5033 ± 0.0503 |
| $1 \leq N_{\text{jet}} \leq 2$ | θ_e | 0.5472 ± 0.0264 |
| | θ_μ | 0.4544 ± 0.0350 |

表 3.8 Summary of fake factors of electron and muon with different N_{jet} requirements. Uncertainties are statistical.

可以计算在信号区，即高 N_{textjet} 区，的 fakes 估计值。计算方法如下：

$$N_{ee}^{\text{fakes}}(N_{\text{jet}} \geq 3) = (N_{e\cancel{e}}^{\text{data}} - N_{e\cancel{e}}^{\text{promptSS}} - N_{e\cancel{e}}^{V\gamma} - N_{e\cancel{e}}^{\text{QmisID MC}})(N_{\text{jet}} \geq 3) \times \theta_e \quad (3.6)$$

$$N_{\mu\mu}^{\text{fakes}}(N_{\text{jet}} \geq 3) = (N_{\mu\cancel{\mu}}^{\text{data}} - N_{\mu\cancel{\mu}}^{\text{promptSS}} - N_{\mu\cancel{\mu}}^{V\gamma})(N_{\text{jet}} \geq 3) \times \theta_\mu \quad (3.7)$$

$$N_{e\mu}^{\text{fakes}}(N_{\text{jet}} \geq 3) = (N_{e\cancel{\mu}} - N_{e\cancel{\mu}}^{\text{promptSS}} - N_{e\cancel{\mu}}^{V\gamma} - N_{e\cancel{\mu}}^{\text{QmisID}})(N_{\text{jet}} \geq 3) \times \theta_\mu \quad (3.8)$$

$$+ (N_{\cancel{e}\mu} - N_{\cancel{e}\mu}^{\text{promptSS}} - N_{\cancel{e}\mu}^{V\gamma} - N_{\cancel{e}\mu}^{\text{QmisID MC}})(N_{\text{jet}} \geq 3) \times \theta_e$$

各个成分的选择同样遵循计算 fake factor 时要求，表 3.9（表 3.10）总结各种成分在 $N_{\text{textjet}} \geq 2$ ($N_{\text{textjet}} \geq 3$) 时的数值。最后，计算得到信号区的 fakes 结果

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | QmisID | Data |
|-------------------------|-------------------|------------------|-----------------|------------------|-----------------|------------------|--------------------|------|
| $N_{\text{jet}} \geq 2$ | $e\cancel{e}$ | 37.39 \pm 4.24 | 1.67 \pm 0.12 | 11.55 \pm 1.62 | 0.19 \pm 0.04 | 51.74 \pm 8.67 | 137.17 \pm 33.00 | 829 |
| | $\mu\cancel{\mu}$ | 32.41 \pm 2.83 | 1.44 \pm 0.15 | 38.97 \pm 3.15 | 0.12 \pm 0.03 | 1.01 \pm 0.59 | - | 583 |
| | $\cancel{e}\mu$ | 39.71 \pm 3.06 | 2.02 \pm 0.17 | 15.46 \pm 2.13 | 0.19 \pm 0.04 | 53.50 \pm 9.21 | 195.94 \pm 19.80 | 708 |
| | $e\cancel{\mu}$ | 17.89 \pm 2.50 | 0.42 \pm 0.10 | 17.00 \pm 1.99 | 0.03 \pm 0.02 | 0.75 \pm 0.39 | 0.43 \pm 0.03 | 267 |

表 3.9 Observed number of data and expected events yields in high jet multiplicity region, which is used to predict fakes in low mass search. Uncertainties are statistical.

如表 3.11所示。其中，只考虑了统计误差，其计算方式为 $\theta_\ell \times \sqrt{N_{\ell\cancel{\ell}}^{\geq 2\text{jet}(3\text{jet})}}$ [18]。

3.2.2 系统误差

计算 fake factor 时有如下系统误差：

1. **统计误差**，低 N_{jet} 区的统计误差会传递到 fake factor；
2. **QmisID**，QmisID 的贡献在 ee 道是比较大的，其系统误差也会传递到 θ_ℓ 的计算。

| Selections | | VV | $t\bar{t}V$ | tV | $t\bar{t}H$ | $V\gamma$ | QmisID | Data |
|-------------------------|-----------|------------------|-----------------|------------------|-----------------|------------------|-------------------|------|
| $N_{\text{jet}} \geq 3$ | $e\ell$ | 15.07 \pm 1.83 | 1.41 \pm 0.11 | 3.96 \pm 0.90 | 0.17 \pm 0.03 | 15.07 \pm 4.85 | 85.54 \pm 6.45 | 354 |
| | $\mu\mu$ | 14.95 \pm 1.94 | 1.12 \pm 0.13 | 16.18 \pm 2.01 | 0.10 \pm 0.03 | 0.03 \pm 0.03 | - | 303 |
| | $\ell\mu$ | 17.84 \pm 2.04 | 1.60 \pm 0.16 | 6.71 \pm 1.62 | 0.18 \pm 0.04 | 17.98 \pm 5.18 | 102.56 \pm 5.64 | 287 |
| | $e\mu$ | 4.78 \pm 1.06 | 0.36 \pm 0.09 | 7.68 \pm 1.24 | 0.02 \pm 0.02 | 0.44 \pm 0.27 | 0.21 \pm 0.03 | 149 |

表 3.10 Observed number of data and expected events yields in high jet multiplicity region, which is used to predict fakes in high mass search. Uncertainties are statistical.

| Selections | $N_{\text{jet}} \geq 2$ | | | $N_{\text{jet}} \geq 3$ | | |
|-------------|-------------------------|-------------------|-------------------|-------------------------|-------------------|-------------------|
| | ee | $\mu\mu$ | $e\mu$ | ee | $\mu\mu$ | $e\mu$ |
| Event yield | 318.27 \pm 9.64 | 256.20 \pm 8.06 | 332.69 \pm 9.62 | 127.38 \pm 6.10 | 122.97 \pm 5.58 | 138.25 \pm 6.16 |

表 3.11 Estimated jet fakes in three channels with different selections. Uncertainties are statistical.

3. **Closure test**, 此分析中 fake factor 的假设是其值不依赖于 jet 数, 但此假设本身是有误差的, 所以为了考虑此项误差, 可以利用 MC (semi-leptonic $t\bar{t}$) 重复一遍 fake factor 方法, 将真实的 fakes 与预测值作为系统误差。具体流程如下:

- 要求 $1 \leq N_{\text{jet}} \leq 2$, 其中为了增大统计量, 去除 b -veto, Z veto 和轻子 p_T 选择条件。

- 选择 ee ($e\ell$) 和 $\mu\mu$ ($\mu\mu$) 事例, 计算 fake factor ($\frac{N_{\ell\ell}}{N_{\ell\ell}}$):
 $\theta_e = 0.32 \pm 0.12$, $\theta_\mu = 0.12 \pm 0.04$;

- 预测高 N_{extjet} 区的 fakes 数 ($\theta \times N_{\ell\ell}$)。

表 3.12 的总结了在 $t\bar{t}$ MC 中真实 fakes, 预测值以及它们之间的相对差别, 其中 $e\mu$ 道最大的相对差别会作为 fake factor closure 系统误差。

| | Predicted | Real | Relative difference |
|----------|-------------------|------------------|---------------------|
| ee | 24.69 \pm 9.47 | 26.92 \pm 2.06 | 9.03% |
| $\mu\mu$ | 30.44 \pm 10.00 | 34.88 \pm 2.35 | 14.59% |
| $e\mu$ | 40.80 \pm 11.31 | 56.63 \pm 3.01 | 38.80% |

表 3.12 Non-closure uncertainty on θ_e and θ_μ . To reduce the statistical error, SS, $p_T(\ell)$ and $M(\ell\ell) > 15$ GeV requirements are dropped in pre-selections.

4. **样本成分**, 低 N_{jet} 与高 N_{jet} 区的一个显著区别是背景成分, 以表 3.13 为例证, 可以看到, 随着 jet 数的增加, $t\bar{t}$ 的比例相应增大。不同味道的 jet 具有不同的误鉴别率, 从而不同本底会有不同的 fake factor。在此估计中, fake factor 是在低 N_{jet} 区估计的, 而后应用在高 N_{jet} , 那么从表 3.13 推论出, $t\bar{t}$ 本底被低估了。为了补偿此项偏差, 可以重复以上 fake factor 方法, 加上至少一个 b -jet 的条件。最后, 把他们之间的差别作为系统误差, 结果如表 3.14 所示。

| Pre-selections | $N_{\text{jet}}=1$ | | | $N_{\text{jet}}=2$ | | | $N_{\text{jet}} \geq 3$ | | |
|----------------------------------|--------------------|----------|--------|--------------------|----------|--------|-------------------------|----------|--------|
| | ee | $\mu\mu$ | $e\mu$ | ee | $\mu\mu$ | $e\mu$ | ee | $\mu\mu$ | $e\mu$ |
| Sherpa W+jets | 38.84 | 30.74 | 152.01 | 4.49 | 13.98 | 49.85 | 5.20 | 3.96 | 19.88 |
| Sherpa $t\bar{t}$ | 7.20 | -0.34 | 10.32 | 9.62 | 37.35 | 62.95 | 15.66 | 28.04 | 59.79 |
| $N_{t\bar{t}}/N_{\text{W+jets}}$ | 0.19 | -0.011 | 0.068 | 2.14 | 2.67 | 1.26 | 3.01 | 7.08 | 3.00 |

表 3.13 The contribution from $t\bar{t}$ becomes bigger as more jets are required. W+jets and $t\bar{t}$ (semi-leptonic) MC samples are produced with the same generator (Sherpa).

| $N_{\text{jet}}=1$ | with b veto | with b -jet | uncer. |
|--------------------------------|---------------------|---------------------|--------|
| θ_e | 0.5401 ± 0.0311 | 0.7228 ± 0.1919 | 33.83% |
| θ_μ | 0.5033 ± 0.0503 | 0.3438 ± 0.0856 | 31.69% |
| $1 \leq N_{\text{jet}} \leq 2$ | with b veto | with b -jet | uncer. |
| θ_e | 0.5472 ± 0.0264 | 0.8000 ± 0.1171 | 46.20% |
| θ_μ | 0.4544 ± 0.0350 | 0.3060 ± 0.0413 | 48.50% |

表 3.14 The fake factors with and without b -jet.

5. **prompt SS 产生截面**, 在 fake factor 计算中, prompt SS 作为减去项, 那么它们的截面理论值也会影响 fake factor 的结果, 它们的理论误差会传递到 fake factor 的误差中, 其中低于 1% 影响的过程被忽略。

Fake factor 的所有系统误差总结在表 3.15 和表 Table 3.16 中。可以发现, 最显著的误差是 Non-closure 和样本成分; 对于 μ 而言, WZ 的产生截面也有 30% 到 40% 的影响, 而对于 electron fake factor, QmisID 误差大小约为 30%; 其次是统计误差最小, 只有不到 10% 的影响, 说明轻子选择条件是比较合理的, 没有引入较大统计误差; 虽然在各个 N_{jet} 类别, 电子 fake factor 误差略高于 μ 子的, 但它们

总误差都在 60% 到 72% 之间。

| | $N_{\text{jet}} == 1$ | $1 \leq N_{\text{jet}} \leq 2$ |
|-------------------|-----------------------|--------------------------------|
| Statistics | 5.76 | 4.82 |
| QmisID | 33.0 | 30.0 |
| θ_e syst. | 38.80 | 38.80 |
| Sample dependence | 33.83 | 46.20 |
| $W^\pm W^\pm$ | 1.22 | 2.08 |
| WZ | 8.93 | 7.94 |
| $V\gamma$ | 11.15 | 12.28 |
| QmisID MC | 1.50 | 2.00 |
| Total | 63.09 | 69.18 |

表 3.15 Summary of systematic uncertainty on θ_e with different N_{jet} selections(in %).

| | $N_{\text{jet}} == 1$ | $1 \leq N_{\text{jet}} \leq 2$ |
|--------------------|-----------------------|--------------------------------|
| Statistics | 9.99 | 7.70 |
| θ_μ syst. | 38.80 | 38.80 |
| Sample dependence | 31.69 | 48.50 |
| $W^\pm W^\pm$ | 6.06 | 10.39 |
| WZ | 39.0 | 33.6 |
| Total | 64.55 | 71.79 |

表 3.16 Summary of systematic uncertainty on θ_μ with different N_{jet} selections(in %).

3.2.3 总预期本底估计

表 3.17和表 3.18分别总结在 $N_{\text{jet}} \geq 2$ 和 $N_{\text{jet}} \geq 3$ 时, 经过初步筛选之后, 的各项本底的估计值与观测数据数; 表内的误差考虑了 fakes 的统计误差和 fake factor 的系统误差, 假设它们相互独立, 总的误差为 $\sqrt{(\theta_\ell^{\text{sys.}} \times N_{\text{jet fakes}}^{\text{median}})^2 + \theta_\ell \times N_{\text{jet fakes}}^{\text{median}}}$, 其中 $\theta_\ell^{\text{sys.}}$ 是 fake factor 系统误差值, $N_{\text{jet fakes}}^{\text{median}}$ 是 fakes 预期值。在三个轻子道中, fakes 都占有比较大的比例, 都高于 30%, 尤其在 ee 中, fakes 作为最大的本底成分, 高达 44%。图 3.5和图 3.6是 N_{jet} 的分布, 分别对应 $N_{\text{jet}} \geq 2$ 和 $N_{\text{jet}} \geq 3$ 。

相比图 3.1，数据与预期本底吻合度得到极大地提升，其偏差基本控制在 2 个标准偏差之内。

| | ee | $\mu\mu$ | $e\mu$ |
|-------------------|---------------|---------------|----------------|
| Jet fakes | 318.27±201.23 | 256.20±165.77 | 332.69±156.43 |
| PromptSS | 208.92±6.64 | 334.71±8.74 | 560.18±10.63 |
| $V + \gamma$ | 105.39±12.43 | 0.01±0.01 | 107.99±15.17 |
| QmisID | 101.47±0.60 | 0.00±0.00 | 18.21±0.23 |
| Total backgrounds | 734.07±201.72 | 590.93±166.00 | 1019.06±157.52 |
| Observed | 790 | 487 | 1257 |

表 3.17 Event yields at pre-selection level, corresponding to $N_{\text{jet}} \geq 2$. The total uncertainties include all systematics on fakes and statistical uncertainties on the others. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} .

| | ee | $\mu\mu$ | $e\mu$ |
|-------------------|--------------|--------------|--------------|
| Jet fakes | 127.38±88.52 | 122.97±88.60 | 138.25±69.55 |
| PromptSS | 95.34±4.30 | 154.40±5.64 | 262.03±7.04 |
| $V + \gamma$ | 28.03±4.52 | 0.01±0.01 | 51.62±13.75 |
| QmisID | 35.60±0.38 | 0.00±0.00 | 8.38±0.16 |
| Total backgrounds | 286.35±88.74 | 277.38±88.78 | 460.27±71.25 |
| Observed | 332 | 213 | 511 |

表 3.18 Event yields at pre-selection level, corresponding to $N_{\text{jet}} \geq 3$. The total uncertainties include all systematics on fakes and statistical uncertainties on the others. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} .

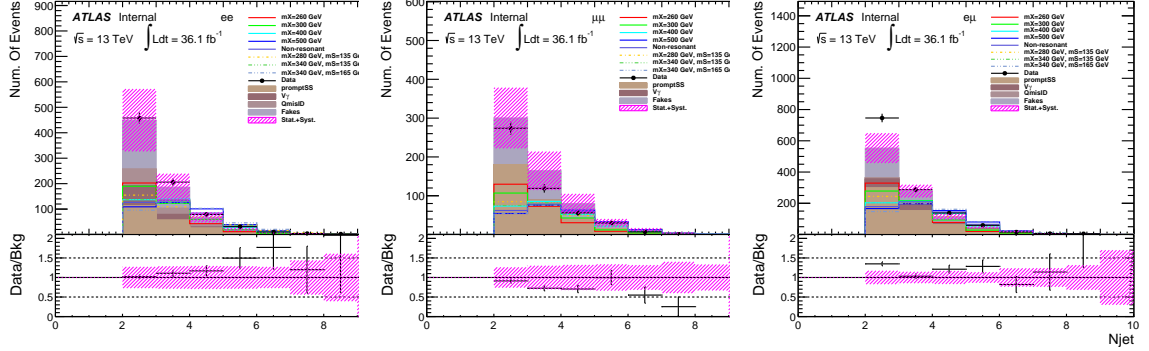


图 3.5 The comparisons between data and backgrounds at pre-selection level, corresponding to $N_{\text{jet}} \geq 2$. Left: ee , middle: $\mu\mu$, right: $e\mu$. The uncertainties, represented by slashed bands, include all systematics on fakes and statistical uncertainties on the other background components. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} .

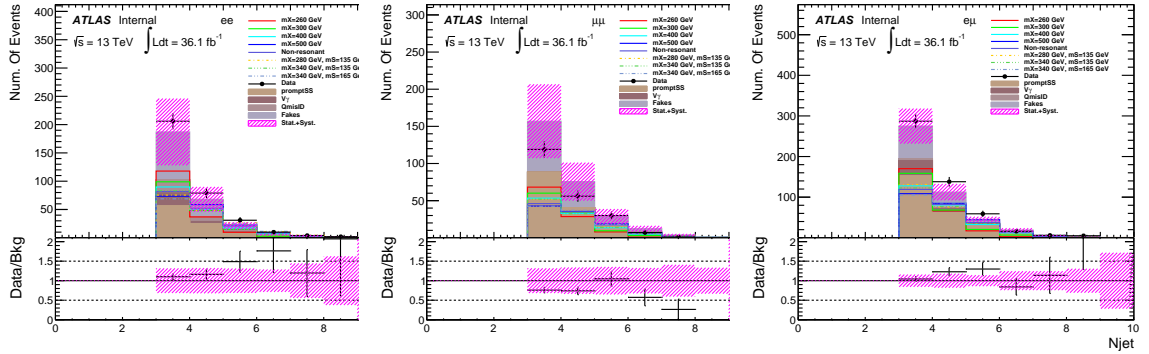


图 3.6 The comparisons between data and backgrounds at pre-selection level, corresponding to $N_{\text{jet}} \geq 3$. Left: ee , middle: $\mu\mu$, right: $e\mu$. The uncertainties, represented by slashed bands, include all systematics on fakes and statistical uncertainties on the other background components. PromptSS and $V + \gamma$ are normalized to the luminosity of 36.1 fb^{-1} .

第4章 信号优化

在初步筛选之后，为了进一步加强信号显著性，利用动力学性质进行信号优化是有必要的。因为每个信号质量点的动力学性质差别较大，所以，每个质量点都会进行信号优化过程，而后通过各自动力学选择条件之后的区域才作为每个质量点的最终信号区。

4.1 优化策略

MVA 方法用于确定不同运动学变量的分离能力，并考虑所有变量之间的相关性。最终，前五个运动学变量用于形成优化选择，分别是 $M(\ell\ell)$, $\Delta R_{min}(\ell_2, j)$, $\Delta R_{min}(\ell_1, j)$, $M_{\ell_1 jj}$ 和 $M(all)$ ，它们具有很强的分离能力，而且相互之间的相关性很低 (图 ??)。通常， $M(\ell\ell)$ 和 $M_{\ell_1 jj}$ 对低质量点敏感，而其余对高质量和非共振信号敏感。基于这些知识， $\Delta R_{min}(\ell_1, j)$, $M(\ell\ell)$, $M_{\ell_1 jj}$ 和 $M(all)$ 用于在低质量搜索中形成优化削减，而 $\Delta R_{min}(\ell_2, j)$, $\Delta R_{min}(\ell_1, j)$, $M(\ell\ell)$ 和 $M_{\ell_1 jj}$ 用于高质量搜索中。它们的相应分布分别见图 ??和图 ??。

4.2 信号效率

4.3 优化结果

参考文献

- [1] FREDERIX R, FRIXIONE S, HIRSCHI V, et al. Higgs pair production at the LHC with NLO and parton-shower effects[J/OL]. Phys. Lett., 2014, B732:142-149. DOI: [10.1016/j.physletb.2014.03.026](https://doi.org/10.1016/j.physletb.2014.03.026).
- [2] ALWALL J, FREDERIX R, FRIXIONE S, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations[J/OL]. JHEP, 2014, 07:079. DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [3] KALOGEROPOULOS A, ALWALL J. The SysCalc code: A tool to derive theoretical systematic uncertainties[J]. 2018.
- [4] BOROWKA S, GREINER N, HEINRICH G, et al. Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence[J/OL]. Phys. Rev. Lett., 2016, 117(1):012001. DOI: [10.1103/PhysRevLett.117.012001](https://doi.org/10.1103/PhysRevLett.117.012001).
- [5] DE FLORIAN D, et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector[J/OL]. 2016. DOI: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002).
- [6] MALTONI F. Higgs pair production[EB/OL]. <https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/HiggsPairProduction>.
- [7] BELLM J, et al. Herwig++ 2.7 Release Note[J]. 2013.
- [8] DENNER A, HEINEMEYER S, PULJAK I, et al. Standard Model Higgs-boson branching ratios with uncertainties[J/OL]. Eur. Phys. J., 2011, C71:1753. DOI: [10.1140/epjc/s10052-011-1753-8](https://doi.org/10.1140/epjc/s10052-011-1753-8).
- [9] GLEISBERG T, HOECHE S, KRAUSS F, et al. Event generation with SHERPA 1.1[J/OL]. JHEP, 2009, 02:007. DOI: [10.1088/1126-6708/2009/02/007](https://doi.org/10.1088/1126-6708/2009/02/007).
- [10] ALIOLI S, NASON P, OLEARI C, et al. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX[J/OL]. JHEP, 2010, 06:043. DOI: [10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043).
- [11] SJÖSTRAND T, STEPHEN M, PETER S. PYTHIA 6.4 physics and manual[J/OL]. JHEP, 2006, 05:026. DOI: [10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026).
- [12] BALL R D, et al. Parton distributions for the LHC Run II[J/OL]. JHEP, 2015, 04:040. DOI: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040).
- [13] ATLAS Collaboration. Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV[M/OL]. 2016. <https://cds.cern.ch/record/2120417>.
- [14] ATLAS Collaboration. Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses[M/OL]. 2016. <https://cds.cern.ch/record/2120826>.
- [15] ATLAS Collaboration. Multi-boson simulation for 13 TeV ATLAS analyses[M/OL]. 2016. <https://cds.cern.ch/record/2119986>.

- [16] The ATLAS Collaboration. Search for heavy neutrino, W_R and Z_R gauge bosons in events with two high- P_T leptons and jets with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV[J]. ATL-COM-PHYS-2013-810, 2013.
- [17] A. Alonso and B. Meirose. New data-driven methods for lepton charge mis-identification[J]. ATL-COM-PHYS-2012-164, 2012.
- [18] ALISON J. The fake factor method[M/OL]. Cham: Springer International Publishing, 2015: 151-209. http://dx.doi.org/10.1007/978-3-319-10344-0_9.

作者简历及攻读学位期间发表的学术论文与研究成果

本科生无需此部分。

作者简历

casthesis 作者

吴凌云，福建省屏南县人，中国科学院数学与系统科学研究院博士研究生。

ucasthesis 作者

莫晃锐，湖南省湘潭县人，中国科学院力学研究所硕士研究生。

已发表 (或正式接受) 的学术论文:

[1] ucasthesis: A LaTeX Thesis Template for the University of Chinese Academy of Sciences, 2014.

申请或已获得的专利:

(无专利时此项不必列出)

参加的研究项目及获奖情况:

可以随意添加新的条目或是结构。

致 谢

