



北京大学

博士研究生学位论文

题目：多夸克喷注探针的建立及
三规范玻色子共振态的寻找

姓 名：	吕旭东
学 号：	1601110102
院 系：	物理学院
专 业：	粒子物理与原子核物理
研究方向：	实验高能物理方向
导师姓名：	冒亚军 教授
	李强 长聘副教授

二〇二一年 六 月



版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。



摘要

粒子物理中的标准模型是一项非常成功的理论，它可以描述自然界中组成物质的基本粒子和四种基本相互作用力中的三种（包括电磁、弱、强相互作用，引力并未统一）。物质世界是由自旋为 $\frac{1}{2}$ 的费米子组成，它们之间的相互作用通过传递自旋为 1 的规范玻色子实现。标准模型的最后一块拼图 Higgs 粒子在 2012 年由大型强子对撞机 LHC 上的 CMS 和 ATLAS 实验发现。然而，标准模型也有一些缺陷。最显著的缺陷是并不能统一引力。另外，电弱相互作用和普朗克尺度能级存在巨大的差异，后者是前者的 10^{16} 倍。标准模型存在的这些问题推动了在 TeV 尺度上超出标准模型 (BSM) 理论的研究，从这些 BSM 理论出发可以得到丰富且新颖的物理过程。寻找 BSM 引入的新粒子是 LHC 上的重大课题。近些年，两体衰变双玻色子共振态的寻找并没有发现存在新物理的证据。最近几年内，BSM 中一些模型比如拓展额外维模型预测了三玻色子共振态过程并且在 LHC 的能量尺度上有很大的发现潜力。然而目前 LHC 上并没有寻找末态为三玻色子的共振态的分析。

本文是第一次对于重共振态衰变到三个重 W 玻色子 $W_{KK} \rightarrow RW \rightarrow WWW$ 过程的分析，其中 W_{KK} 是大质量规范玻色子 Kaluza-Klein 激发态，R 是标量粒子 radion。本文要求末态包括一个孤立轻子、丢失横动量 (p_T^{miss}) 和一个或两个大圆锥角的胖喷注。对于大多数情形，信号中一个 W 衰变到轻子加中微子，另外两个 W 经强子衰变的末态结构会根据 W_{KK} 和 radion 的质量比值不同而被重建为两种情形，包括 2 个分别合并的 AK8 W 玻色子喷注或者单个合并的 WW 喷注。另外，两个 W 中其中一个轻子衰变但轻子非孤立的情形也被考虑在内。所以强子化的末态有多重情形，可能包含分立的 R 到两个合并的 W 玻色子喷注、单个合并 Radion 喷注或有夸克在圆锥外的部分合并的 radion 喷注。信号喷注的质量谱会很复杂，因此分析中在不同的质量区间内使用不同的标记加以区分信号。

本文的分析是基于 LHC 上 CMS 探测器在 2016-2018 三年间收集到的对应积分亮度为 137fb^{-1} 的数据。两个强子衰变 W 合并成单个 WW 喷注（合并 radion 情形）和重建为两个 W 喷注（分立情形）都被同时探查。在分析中采用了基于深度神经网络 (DNN) 的 DeepAK8 标记来提高信号和本底的区分度。其中使用 deep-W 来标记分立情形，对于合并情形，定义了 deep-WH 作为四夸克喷注的鉴别量。标记值校准时一个难题是在标准模型中并不存在有四夸克结构且统计量足够的粒子，因此本文通过在全合并顶夸克的基础上额外要求一个夸克或胶子从而形成四夸克喷注作为信号过程的代理，并在标记校准时开发了全新的“矩阵方法”。这项多夸克探针的开发与校准为很多分析

的开展提供了可能性, 比如 $\text{Higgs} \rightarrow \text{WW} \rightarrow 4q$ 的测量。最后采用数据驱动的方法探测了 $\text{W}_{\text{KK}}\text{-R}$ 质量平面上新物理存在的可能性 ($M_{\text{W}_{\text{KK}}}: 1.5\text{-}5 \text{ TeV}$, $M_{\text{R}}: 0.18 \text{ TeV}-M_{\text{W}_{\text{KK}}}$)。拟合后数据与估计的本底相近, 并没有新物理存在的迹象。对应合并和分立情形, 可以排除 W_{KK} 质量分别为 3.5TeV 和 3.3TeV 。

本文另外一项工作是从 2019 年起担任 CMS 两个组 JME 和 B2G 组蒙特卡洛联系人 (MC contact) 的工作。CMS 实验可以提供用于分析的蒙特卡洛模拟的官方样本。作为 MC contact 需要为分析者提供的物理过程创建用于申请样本的 request 并检验是否正常工作, 并在线上 MCM 网站准备 ticket 用于提交样本。为了减少重复性工作以提高效率, 采用多个 gridpack 用单样本产生技术和全自动化脚本并得到了组内推广。

关键词: 三玻色子共振态寻找, DeepAK8, 新物理, 标记校准技术

Development for multi-quarks tagger and search for resonances decaying to triple W-boson

Xudong Lyu (Particle Physics and Nuclear Physics)

Directed by Prof. Yajun Mao

Prof. Qiang Li

ABSTRACT

The Standard Model (SM) is an extremely successful theory describing the elementary particles in the nature and three of the four well-known fundamental forces (the electromagnetic, weak, and strong force and not including the gravitational force). In the theory, fundamental particles are spin- $\frac{1}{2}$ fermions with interact between them through exchange of spin-1 gauge bosons. The last piece of SM, Higgs, was found by both CMS and ATLAS experiments at Large Hadron Collider (LHC) in 2012. However, the SM has shortcomings. The obvious one is that it fails to include the gravitational force at high energies. In addition, there is a hierarchy problem between the electroweak force and Planck scale, the latter one is 10^{16} times larger than the former one. All these problems motivate the presence of Beyond the Standard Model (BSM) at TeV scale, which leads to rich and novel phenomenology. Searching for new particles, which are introduced by the BSM, is an important topic at LHC. In recent years, probing these heavy resonances through two-body decays into diboson final states have so far found no evidence of new physics. In recent years, many new BSM, i.e., warped extra dimensional model, motivate a triboson resonance indicating discovery potential within LHC reach, while this has not been performed before.

This thesis presents the first search for heavy resonances decaying in triple massive W bosons, through $W_{KK} \rightarrow RW \rightarrow WWW$, where W_{KK} is a Kaluza-Klein excited massive gauge boson and R is a scalar radion. The final state topology analyzed consists of one isolated lepton, missing transverse momentum (p_T^{miss}), and one or two massive large radius jets. In most cases, it originates from a W decaying to isolated leptons while the other two Ws decaying hardonically into either two merged AK8 W-jet or even merged to a single merged WW-jet depending on the relative masses of W_{KK} and R bosons. I also consider the case where one of the two merged W bosons decay leptonically with a nonisolated lepton inside the jet. So the

hadronic final states can be quite different, including resolved R into two merged W-jets, one merged radion R-jet, or partly merged R-jet with particles leaked outside the AK8-jet. Thus the jet mass spectrum can be very complicated, and the tagger to be used can be different in separate mass regions.

The analysis is based on proton-proton collision data at $\sqrt{s} = 13$ TeV collected by the CMS experiment at the CERN LHC during 2016, 2017 and 2018, corresponding to an integrated luminosity of 137fb^{-1} . Both regimes with two W-bosons merged in a single WW-jet (merged R) and with two separated W-jets (resolved R) are simultaneously probed. In this search DeepAK8 tagger implementing Deep Neural Networks (DNN) are used to further increase signal-background discrimination, deep-W are used to tag the resolved R, while for the merged R reconstructed as 3- or 4- prong jets, deep-WH is defined as the discriminator. Since there is no 4-prong object with large statistics in the SM, fully merged top together with an extra quark or gluon forms the proxy as a mimic. A novel method, which is referred as "Matix Method", is developed to calibrate the discriminator. The development and calibration of multi-quarks tagger offer possibilities for many analysis, like the measurement of $\text{Higgs} \rightarrow \text{WW} \rightarrow 4\text{q}$. And the background contributions after signal selection are estimated using a data-driven technique in dedicated control regions. Finally, a wide area of the $\text{W}_{\text{KK}}\text{-R}$ mass parameter space (ranging in $\text{M}_{\text{W}_{\text{KK}}}: 1.5\text{-}5\text{TeV}$, $\text{M}_{\text{R}}: 0.18\text{TeV}-\text{M}_{\text{W}_{\text{KK}}}$) is probed for the first time. The observed data are in agreement with the estimated background simulation and W_{KK} scenarios are excluded up to masses of 3.5 TeV and 3.3 TeV for merged and resolved R cases respectively.

Another work mentioned in this thesis is the Monte Carlo (MC) contact work for 2 of CMS object or analysis groups (JME and B2G) since 2019. The CMS experiment provides the analyzers with official simulation of Monte Carlo samples, which can be used for analyses. MC contact should create and validate requests for dedicated processes required by the analyzers and prepare tickets for submission in the online system. Multi-gridpack technique and smarter automatic scripts were implemented to reduce repeating work and help to improve the efficiency.

KEYWORDS: triboson resonance search, DeepAK8, new physics, tagger calibration

目录

第一章 引言	1
1.1 标准模型	1
1.1.1 标准模型的粒子组成	1
1.1.2 标准模型的拉氏量	2
1.1.3 强相互作用	2
1.1.4 电弱相互作用	3
1.1.5 Higgs 机制	4
1.2 超出标准模型的新物理	6
1.2.1 标准模型的缺陷	6
1.2.2 超对称模型	9
1.2.3 额外维模型	9
1.2.4 拓展额外维模型	10
1.2.5 玻色子共振态研究现状	12
1.3 论文结构	14
第二章 LHC 对撞机与 CMS 探测器	15
2.1 LHC 对撞机	15
2.2 CMS 探测器	15
2.2.1 赝快度	16
2.2.2 径迹探测器	17
2.2.3 电磁量能器	18
2.2.4 强子量能器	19
2.2.5 μ 子探测器	21
2.2.6 触发系统	22
第三章 CMS 探测器上事例的重建	23
3.1 径迹与初级顶点的重建	23
3.2 粒子流算法	24
3.2.1 μ 子	24
3.2.2 电子与光子	26
3.2.3 强子	26

3.2.4	丢失横能量	26
3.3	Pile-up 消除	27
3.3.1	Charged Hadron Subtraction (CHS)	27
3.3.2	Pile up per particle identification (PUPPI)	27
3.4	喷注的重建	28
3.4.1	喷注聚类算法 (jet clustering)	29
3.4.2	PF 喷注	30
3.4.3	喷注能量修正	31
3.5	喷注结构的重建	32
3.5.1	喷注质量修饰 (Grooming)	33
3.5.2	子喷注数量鉴别技术 (N-subjettiness)	35
3.5.3	DeepAK8 标记技术	36
3.5.4	官方修正因子	38
第四章	末态为轻子和喷注的三玻色子共振态的寻找	41
4.1	半轻衰变三玻色子共振态的寻找	41
4.2	信号与本底的蒙卡样本	42
4.3	事例的挑选	43
4.3.1	噪声过滤与顶点选取	43
4.3.2	触发	43
4.3.3	μ 子挑选	45
4.3.4	电子挑选	45
4.3.5	喷注挑选	46
4.4	事例选择条件	47
4.4.1	初选条件	47
4.4.2	玻色子标记	48
4.4.3	信号区选择条件	48
4.4.4	信号区喷注质量分类	49
4.5	6 个信号区观测量	49
4.6	数据与蒙卡模拟样本对比	61
4.6.1	2018 年电子道 C 与 D 时期异常	61
4.6.2	不同轻子道数据蒙卡对比	61
4.6.3	不同采集年份数据蒙卡对比	61

第五章 三玻色子共振态寻找本底估计方法	69
5.1 本底过程喷注的匹配分解	70
5.2 修正 deepAK8 标记差异的比例因子	72
5.2.1 矩阵方法的说明	74
5.2.2 SFs 对横动量的依赖	77
5.3 对标准模型喷注使用 SFs 后的检验	80
5.4 分立信号 SFs 的使用	87
5.5 信号合并 radion 喷注对应的 SFs	89
5.6 关于 SFs 的系统误差	92
5.6.1 PS 和产生子误差	92
5.6.2 子样本挑选条件带来的偏差	92
5.6.3 代理对象误差	93
5.6.4 信号高横动量外推误差	93
5.6.5 信号关于 SFs 误差总结	94
5.7 top 本底的控制区的定义与估计	99
5.8 W+jets 本底控制区的定义与估计	101
5.9 本底估计方法的总结	101
第六章 系统误差及结论	105
6.1 本底和信号共用的系统误差	106
6.2 本底系统误差	108
6.2.1 本底归一化误差	108
6.2.2 本底形状误差	108
6.3 信号系统误差	108
6.3.1 信号归一化误差	108
6.3.2 信号形状误差	109
6.4 拟合与排除限设置	113
6.4.1 预拟合与后拟合分布	113
6.4.2 拟合结果评价	113
6.4.3 信号注入偏离检验	115
6.4.4 排除限	115
6.4.5 一维排除限	119

第七章 蒙特卡洛 contact 工作	121
7.1 蒙卡样本的产生过程	121
7.2 MC contact 的工作任务	123
7.3 contact 工作中的创新	126
7.3.1 multi-gridpack	126
7.3.2 自动化产生 request 脚本	126
第八章 总结与展望	129
参考文献	131
附录 A 博士期间的研究成果	139
A.1 CMS 公开文档	139
A.2 国际会议报告	139
附录 B 附录	141
B.1 不同轻子道其他物理量数据与模拟对比	141
B.2 矩阵方法中由于子样本挑选条件带来的偏差	143
B.3 t^3 和 t^4 关于 deep-W (WH) 的差异	145
B.4 信号 R^{3q} 、 R^{4q} 和 $t^{3,4}$ 在添加代理误差后的对比	146
B.5 top 与 W+jets 控制区传递系数线性拟合	148
B.6 二维质量平面的插值方法	149
B.7 信号系统误差的研究	150
致谢	155
北京大学学位论文原创性声明和使用授权说明	157

第一章 引言

1.1 标准模型

标准模型 (Standard Model) 是现代基本粒子物理学中的核心理论 [55、83、101], 它结合了多种量子场论的概念, 在解释高能物理世界方面非常成功。自然界的四种相互作用中, 它包含其中三个, 包括电磁相互作用, 弱相互作用和强相互作用, 引力除外。此外, 它还对所有已知的基本粒子进行了分类, 最新的基本粒子是希格斯玻色子 (Higgs), 该粒子于 2012 年在欧洲核子研究组织的大型强子对撞机上被观测到。

1.1.1 标准模型的粒子组成

根据标准模型, 基本粒子基于自旋为半整数或整数被分为费米子和玻色子, 如图 1.1 所示。构成物质的费米子包括各有六种的夸克和轻子, 自旋为 $\frac{1}{2}$, 分别有三代, 每代有两种。除中微子外, 所有第二代和第三代粒子均不稳定。另外, 所有的夸克和轻子都有对应的反粒子, 它们具有相同的质量, 相反的电荷。轻子中电子、 μ 子及 τ 子可以参与电磁相互作用与弱相互作用, 而中微子不带电, 所以只参与弱相互作用, 通常会给实验探测带来困难。夸克可以参与所有的相互作用。规范矢量玻色子共有 12 种, 用于传播相互作用。另外, 还引入了 Higgs 标量玻色子用于解释规范矢量玻色子及费米子质量的来源。



图 1.1 组成标准模型的基本粒子。

1.1.2 标准模型的拉氏量

标准模型是一种量子规范场论，其中每个粒子都被描述为在时空坐标处的动力学场。这些场由标准模型拉格朗日密度函数控制。比如一个自由费米子不与任何其他场相互作用的拉氏量可写为

$$\mathcal{L} = \bar{\Psi}(x^\mu)(i\lambda^\mu\partial_\mu - m)\Psi(x^\mu) \quad (1.1)$$

其中 $\Psi(x^\mu)$ 表示任何自旋为 $\frac{1}{2}$ 的狄拉克场，是关于时间和空间的函数。基本粒子之间的任何相互作用都可以用上述等式并添加附加项描述。作为规范场论，标准模型具有规范对称的性质，要求在转换场后观察量保持不变。费米场可以被表述为粒子多重态，属于某一多重态的粒子经过对称变换中仍在其中。标准模型的对称群可表示为

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \quad (1.2)$$

1.1.3 强相互作用

$SU(3)_C$ 群描述了胶子传导的强相互作用，并由量子规理论量子色动力学 (QCD) 描述。该群是由 8 个线性独立矩阵 $T^a = \frac{\lambda^a}{2}$ 产生，其中 λ^a 是 Gell-Mann 矩阵 [52]。这些矩阵满足 $[\lambda^i/2, \lambda^j/2] = if^{ijk}\lambda_k/2$ ，因此 $SU(3)_C$ 群为非阿贝尔群。任何群元素可写作 $e^{-i\theta^a g_a}$ 的形式， a 表示 8 个颜色自由度， g_a 为 8 个线性独立矩阵之一。夸克场中规范变换表示为

$$\Psi(x^\mu) \rightarrow e^{-ig_s\theta^a(x^\mu)T^a}\Psi(x^\mu), \quad (1.3)$$

为了保持方程1.1在这样变换时保持拉氏量不变，需添加一个附加项 $\partial_\mu + ig_s\mathcal{A}_\mu^a T^a$ 来替换偏导数 ∂_μ 。这时 QCD 拉氏量可写为

$$\mathcal{L}_{QCD} = -\frac{1}{4}\mathcal{F}_a^{\mu\nu}\mathcal{F}_{\mu\nu}^a + \bar{\Psi}(x^\mu)(i\lambda^\mu D_\mu - m)\Psi(x^\mu), \quad (1.4)$$

其中规范场张量 $F_{\mu\nu}^a$ 可表示为

$$F_{\mu\nu}^a = \partial_\mu\mathcal{A}_\nu^a - \partial_\nu\mathcal{A}_\mu^a + g_sf^{abc}\mathcal{A}_\mu^b\mathcal{A}_\nu^c. \quad (1.5)$$

方程1.4的第一项表示夸克和胶子的相互作用，第二项导致了胶子的自相互作用。这些自相互作用会导致夸克会被有相同颜色的虚胶子和夸克包围。当以越来越高的能量探测夸克颜色时，对应于越来越短的距离，色荷会减少，直到只有裸荷可见，那里的夸

克基本上是自由的，可以观察到为可分辨的粒子，夸克之间的耦合常数随着能量的增加而减小，此属性称为渐近自由。出于同样的原因，当远离裸色荷时，围绕它的色荷海会使观察到的色荷增加，这导致长距离时强大吸引力，其中两者之间的势能随它们之间的距离线性增长，可以表示为

$$V(r) = -\frac{4\alpha_s}{3r} + kr, \quad (1.6)$$

其中 r 为夸克之间的距离， α_s 为 QCD 耦合常数。当距离增加时，势能也会增加，增加到足够大时，足够在真空中产生夸克-反夸克对来降低势能。因此夸克不会单独存在，这种现象被称为夸克禁闭。

1.1.4 电弱相互作用

电磁相互作用和弱相互作用是由 $SU(2)_L \otimes U(1)_Y$ 对称性的破坏引起的。属于 $SU(2)_L \otimes U(1)_Y$ 自由费米场的拉氏量如下

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi \quad (1.7)$$

在如下变化中

$$\Psi \rightarrow e^{\frac{i}{2}g_W T^k \Lambda^k(x)} e^{\frac{i}{2}g'_W Y_\alpha(x)} \Psi, \quad (1.8)$$

其中 g_W 为 $SU(2)_L$ 耦合常数， g'_W 为 $SU(1)_Y$ 耦合常数。为了使得变换满足规范对称，引入了协变微商 D_μ 来替换偏导数 ∂_μ ：

$$(D_\mu) = \delta_{ij}\Psi_\mu + ig_W(T^k W_\mu^k)_{ij} + ig'_W Y \delta_{ij} B_\mu \quad (1.9)$$

式中引入了玻色子场与费米场的相互作用。最后电弱相互作用的拉氏量如下：

$$\mathcal{L}_{EW} = i\bar{\Psi}D^\mu\Psi - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}\mathcal{B}^{\mu\nu}\mathcal{B}_{\mu\nu} \quad (1.10)$$

到现在为止，我们考虑了自发对称破坏之前的拉氏量，其中有三个带电的无质量玻色子 (W^1, W^2, W^3) 和一个无质量的中性玻色子，这与我们的所知相违背。我们将在下节1.1.5中证明，由于自发的对称性破坏，带电的玻色子场 $W^\pm$ 实际上是 W^1 与 W^2 的线性组合：

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2). \quad (1.11)$$

Z^0 和 γ 玻色子可表示为关于电弱混合角 θ_W 的矩阵：

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \quad (1.12)$$

1.1.5 Higgs 机制

自发对称性破缺

为了实现自发的对称破坏, Goldstone[54] 引入具有真空量子数的标量场 $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, 并赋予该场真空期望值。场的拉氏量可写作：

$$\mathcal{L} = \partial^\mu \bar{\phi} \partial_\mu \phi - \mu_0^2 \bar{\phi} \phi - \frac{\lambda_0}{6} (\bar{\phi} \phi)^2, \quad (1.13)$$

其中 λ_0 为耦合常数, μ_0 为质量, 该拉氏量满足 $U(1)$ 的对称变换。如果 μ_0^2 为负值, 势能会在半径为 v 的圆上最小, 满足 $v^2 = \frac{\mu_0^2}{\lambda_0}$, v 也被称为真空期望值点。哈密顿量的最低值在 $\phi = v$ 处, 而不是 $\phi = 0$ 。通过将场转换到最小能量位置 $\phi' = \phi + v$, 可以有效地打破了两个场分量 η 、 ϵ 之间的对称性。新的拉氏量可写作：

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu_0^2 \eta^2 + \dots \quad (1.14)$$

第三项是标量场 η 质量项形式。Goldstone 理论并没有提供矢量玻色子的质量项, 却得到了一个大质量标量和一个无质量标量玻色子。

Higgs 机制

对于 $U(1)$ 对称性, 要求拉氏量在 $\phi \rightarrow e^{i\theta(x)}\phi$ 旋转变换下不变, 协变微商需写作 $D_\mu = \partial_\mu - ieA_\mu$ 来替换 ∂_μ 。将场 ϕ 转换到基态可以得到拉氏量：

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 - v^2 \lambda \eta^2 + \frac{1}{2} v^2 e^2 A_\mu A^\mu - ev, A_\mu \partial^\mu \epsilon - \frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a \quad (1.15)$$

可以发现包含一个无质量的 Goldstone 玻色子 ϵ 、一个有质量的标量 η 和有质量 $m_A = ev$ 的矢量场 A_μ 。此时无质量标量仍存在, 这是由于给矢量场 A_μ 赋予质量, 新增了纵向极化, 极化自由度从 2 增加到 3。这时通过用简单的规范变换：

$$A_\mu \rightarrow A_\mu + \frac{1}{ev} \partial_\mu \theta, \quad (1.16)$$

Goldstone 玻色子消失并转化为有质量的纵向极化玻色子。最后 Goldstone 理论只剩下一个带质量的矢量规范玻色子 A_μ 和带质量的标量玻色子 h 。

1968 年, S. Glashow、A. Salam 和 S. Weinberg 在此基础上得到可重整的弱电统一规范理论, 也被称为 GWS 理论 [77]。该理论在电弱作用拉氏量添加了规范不变项, 即:

$$\mathcal{L} = \left| (i\partial_\mu - g_2 T_a W_\mu^a - v g_1 \frac{Y}{2} B_\mu^a) \phi \right|^2 - V(\phi). \quad (1.17)$$

$SU(2) \times U(1)$ 对应 4 个生成元, 有 4 个规范玻色子, 对于 $SU(2)$ 有 3 个, 对应式中 W_μ , 耦合常数为 g_2 ; 而 $U(1)$ 有 1 个, 对应式中 B_μ , 耦合常数为 g_1 。引入 $SU(2)$ 的标量二重态, 表示为:

$$\phi = \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 + i\phi_4) \end{pmatrix} \quad (1.18)$$

拉氏量的质量项可得出 $m_W = v \cdot g_1/2$ 、 $m_Z = v/2 \cdot \sqrt{g_1^2 + g_2^2}$, 二者的比值即为电弱混合角 θ_W 。费米子质量是源于与 Higgs 场的耦合, 以电子的质量为例, 需引入电子与标量场 ϕ 的 Yukawa 耦合:

$$L_{Yukawa} = -G_e \left[L \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} e^R + \bar{e}_R \begin{pmatrix} \phi^{(-)} & \bar{\phi}^{(0)} \end{pmatrix} \bar{L} \right], \quad (1.19)$$

其中 G_e 是电子耦合常数。通过规范变换去除无质量的 Goldstone 玻色子:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \quad (1.20)$$

h 即 Higgs 粒子, v 为真空预期值点。经变换后拉氏量 Yukawa 项可写为:

$$\mathcal{L}_{Yukawa} = -\frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) \quad (1.21)$$

电子因此获得质量 $m_e = G_e \cdot v/\sqrt{2}$ 。上述讨论同样适用于轻子 μ 和 τ 。Higgs 场自发对称性破缺赋予规范玻色子质量, 与轻子的 Yukawa 耦合使轻子获得质量。

1.2 超出标准模型的新物理

1.2.1 标准模型的缺陷

尽管标准模型非常成功且成功预测了 Higgs 玻色子的存在，但仍有其不足之处。

引力问题

最显著的一个问题是它无法与引力相互作用结合。广义相对论对引力进行了细致的研究，在存在物质和能量的情况下，由时空弯曲产生引力。该理论没有利用量子场，并且能量没有被量子化。标准模型量子场论和广义相对论之间的能量标度完全不同，时空尺度在天文尺度上是弯曲的，而量子场论通常在尽可能小的尺度解决问题，时空的变化在这样尺度下非常小。因此，对于标准模型，时空大约是平坦的，因此不存在引力。

等级问题

除了将引力纳入量子场论框架的困难外，标准模型在普朗克尺度上也会出现问题。该尺度对应的是普朗克质量，表示可能的最小黑洞的质量。普朗克质量约为 1.22×10^{19} GeV，比参与电弱相互作用的 W 和 Z 玻色子大 10^{16} 倍，因此弱核相互作用力和引力在能量尺度存在巨大的差异。这种差异与 Higgs 真空期望值有关，Higgs 场具有 246 GeV 的实测真空期望值，这正是 W 和 Z 玻色子质量的来源。但是，当实际计算 Higgs 真空期望值并考虑包括普朗克能量的修正时，得到 Higgs 玻色子质量比观察到的大 10^{16} 倍，这被称为等级问题。

如此大的能量级差违背了自然性原理，为了描述精细调节 (fine-tuning) [80] 引入无量纲的质量参数 $\mu = m_H^2(\Lambda)/\Lambda^2$ ，其中 $m_H(\Lambda)$ 表示在截至能量尺度 Λ 下 Higgs 的重整化质量，可以简单表示为：

$$m_{H0}^2 - m_H^2 = g^2(\Lambda_0^2 - \Lambda^2) \quad (1.22)$$

其中 Λ_0 是新物理预测的大尺度， g 近似是常数，当非零 Higgs 场真空预期值 v 满足 $m_H^2 = \lambda v_{EW}^2$ 时，无量纲参数 $\mu_{EW} = \lambda$ ，定义 $h = \Lambda_0^2/v_{EW}^2$ ，可以得到：

$$\mu_0 = g^2 + (\lambda - g^2)h^{-1} \quad (1.23)$$

对于等级问题，在普朗克尺度时，公式中系数 $h \sim 10^{34}$ ，因此调节越精细会导致越严重的等级问题。因此标准模型可能只能在小能量尺度下才能做到精确的描述。

μ 子反常磁矩

μ 子反常磁矩与标准模型预言有一定偏差，对新物理有较高敏感度。首先，对于质量 m_ℓ 、电荷 q 、自旋 $\vec{s}$ 的带电粒子，其磁矩满足

$$\vec{\mu}_\ell = g_\ell \left(\frac{q}{2m_\ell} \right) \vec{s}, \quad (1.24)$$

其中 g_ℓ 为比例系数，即该带电粒子的 g 因子。对于电子，由狄拉克方程预言的 g 因子为 2。然而实验测量值被非严格等于 2，因此将 g 因子与 2 的偏差的一半作为反常磁矩， $a_\ell = \frac{g_\ell - 2}{2}$ 。1947 年，Schwinger 提出了对电子磁矩的领头阶修正，即电子被非孤立，其自身产生的电磁场对其本身也会产生作用，通过量子电动力学计算得电子反常磁矩为 $\alpha/2\pi \simeq 0.00116$ [85]， α 为精细结构常数，与同年利用微波方法测量得到的结果一致[66]。

对于 μ 子反常磁矩，标准模型的贡献如图 1.2 所示。其中， μ 子参与的最强相互作用是电磁相互作用，因此来自量子电动力学的贡献是 μ 子反常磁矩最主要的来源。第一个单圈图的贡献与上述电子的反常磁矩相同，同样为 $\alpha/2\pi \simeq 0.00116$ ，与轻子类型无关。单圈图的贡献占反常磁矩的 99.6%，多圈图情形要复杂许多，其结果正比于 $(\frac{\alpha}{\pi})^N$ ，其中 N 为圈图数。目前理论计算到十阶项，共一万多张圈图，得到标准模型计算的理论值 $a_\mu^{QED} = 116591840(59) \times 10^{-11}$ [18]。弱相互作用通过 W 、 Z 和 Higgs 粒子传递，这些粒子对 μ 子反常磁矩也有一定的贡献，在 2012 年 Higgs 粒子发现后，这一项得到重新计算，总的电弱贡献 $a_\mu^{EW} = (153.6 \pm 1.0) \times 10^{-11}$ [53]。

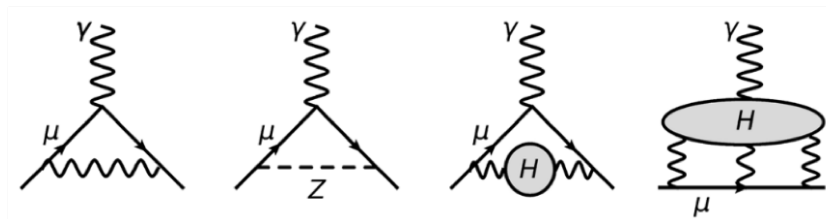


图 1.2 标准模型对 μ 子反常磁矩贡献的费曼图，从左到右分别为量子电动力学、弱相互作用单圈图，量子色动力学真空极化图与光子-光子散射图。

上述是电弱统一理论对 μ 子反常磁矩贡献的讨论，尽管 μ 子不直接参与强相互作用，但可以通过弱相互作用间接参与，从而影响反常磁矩。其中最简单的如图 1.2 第三个费曼图所示，是来源于强相互作用引起的真空极化。图中“H”的圆圈表示由夸克和胶子组成的各种圈图，在低能区微扰方法不再适用。理论学家提出非微扰方法如色散关系（dispersion relation）将光子自能谱中虚数部分与正负电子对湮灭成强子的反应截

面联系 [58]。强相互作用真空极化领头阶可表示为：

$$a_{\mu}^{LO,NP} = \frac{1}{2} \left(\frac{\alpha^2}{\pi} \right) \int_{m_{\pi}^2}^{\infty} \frac{R(s)}{s} K^{(1)}(s) ds, \quad (1.25)$$

其中 $R(s)$ 是正负电子对撞中强子衰变的总的截面经归一化后得到：

$$R(s) = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma_{pt}}, \sigma_{pt} = 4\pi\alpha^2/(3s), \quad (1.26)$$

$K^{(1)}(s)$ 由如下定积分计算：

$$K^{(1)}(s) = \int_0^1 \frac{x^2(1-x)}{x^2 + (1-x)\frac{s}{M_{\mu}^2}} dx, \quad (1.27)$$

更高阶精度下的修正可参考文献 [65]。最后一种费曼图，即光子-光子散射，主要来源于质量较轻的强子，比如 π 、 η 等介子，贡献为 $a_{\mu}^{bl} = 136(25) \times 10^{-11}$ 。最终将所有贡献叠加得理论计算的 g 因子为 $a_{\mu}^{SM} = 116591810(43) \times 10^{-11}$ 。

Brookhaven[15] 和 FNAL 实验组 (Fermilab National Accelerator Laboratory) 通过测量 μ 子在磁场中的进动频率测量其反常磁矩，最新的测量结果如图1.3所示 [14]，表明测量值与标准模型预言有 4.2σ 的偏差，虽然没有达到粒子物理 5σ 的发现标准，但也十分显著。目前发表的结果只使用了第一期采集的数据，在分析全部五期数据得到更精确的实验结果后，有可能会发现新物理。

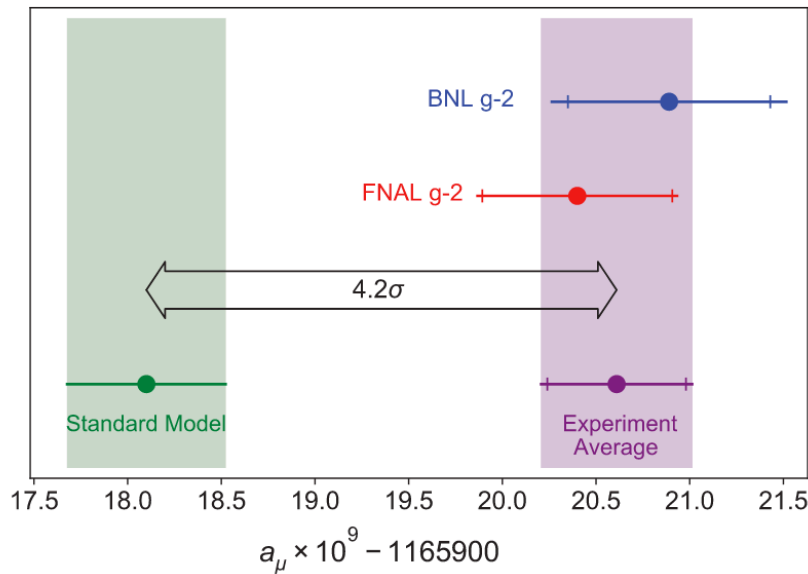


图 1.3 标准模型预言的 μ 子反常磁矩以及 Brookhaven、FNAL 实验组测量值和合并后的结果。

除此之外，还存在解释占宇宙绝大多数的暗物质和暗能量困难、中微子质量等难题。为了解决这些问题，理论学家提出了一些超出标准模型 (BSM) 的理论，比如引入新粒子和新作用力的超对称、小 Higgs 等理论。也有些理论学家认为目前对引力的理解是错误的，存在新的未知维度并且可以在 LHC 的实验中观测到，比如额外维模型。下文将介绍其中几种理论：超对称模型、额外维模型和拓展额外维模型。

1.2.2 超对称模型

超对称模型 [72] 通过要求玻色子与费米子间的额外对称性，可以使得粒子质量远小于普朗克尺度从而解决等级问题。其中一种简单的模型是最小超对称标准模型 (Minimal Supersymmetric Standard Model, MSSM)，要求每个标准模型粒子都对应一个超对称粒子，图1.4为模型要求的新粒子的多重态与对应的规范群量子数。与标准模型相同，最小超对称标准模型的规范群也是 $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ ，对于轻子，包含左、右螺旋度的手征费米场及超对称伙伴，与标准模型只有不同，最小超对称标准模型需 2 个 Higgs 二重态才能使粒子获得质量。

Names		spin 0	spin 1/2	$SU(3)_C, SU(2)_L, U(1)_Y$
squarks, quarks ($\times 3$ families)	Q	$(\tilde{u}_L \ \tilde{d}_L)$	$(u_L \ d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
	$\bar{u}$	$\tilde{u}_R^*$	$u_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	$\bar{d}$	$\tilde{d}_R^*$	$d_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
sleptons, leptons ($\times 3$ families)	L	$(\tilde{\nu} \ \tilde{e}_L)$	$(\nu \ e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
	$\bar{e}$	$\tilde{e}_R^*$	$e_R^\dagger$	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs, higgsinos	H_u	$(H_u^+ \ H_u^0)$	$(\tilde{H}_u^+ \ \tilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, +\frac{1}{2})$
	H_d	$(H_d^0 \ H_d^-)$	$(\tilde{H}_d^0 \ \tilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

图 1.4 最小超对称标准模型的基本粒子及对应规范群量子数。

尽管超对称模型可以在电磁力、弱核力与强核力的耦合系数达到统一，并且能解决标准模型的部分问题，但是在 TeV 量级并没有发现超对称伙伴粒子，实验也并未发现由超对称粒子构成的暗物质，经过近 60 年的探索，预期目的并没有实现。

1.2.3 额外维模型

额外维理论为等级问题提供了解决方案。在 Randall-Sundrum[82] 额外维模型中，提出了新的弯曲空间维度 [6]，如图1.5所述，引入了普朗克能量尺度的 UV “膜” 和 TeV 能量尺度的 IR “膜”，两个膜之间为额外的空间维度 (bulk)。新的度量取决于新额外

维度的半径 r 、曲率因子 k 及反应额外维度尺寸的 r_c ：

$$ds^2 = e^{-2kr_c\phi} \eta_{\mu\nu} dx^\mu dx^\nu + (r_c d\phi)^2, \quad (1.28)$$

引力集中在 UV 膜处，标准模型粒子位于 IR 膜处，该模型预测了 Kaluza-Klein 引力子共振态，自旋为 2 且能量在 TeV 量级，因此适合在高能实验中观测。另外，标准模型粒子包括 Higgs 也允许在额外维中传播。轻的第一代和第二代费米子位于 UV 膜附近，与 Higgs 耦合较小，而 Higgs 玻色子仍然驻留在 IR 膜中，从而解释了它较小的质量。顶夸克同样位于 IR 膜附近，从而导致与 Higgs 玻色子较强的耦合。此外，由于引力子位于 IR 膜附近，而费米子位于 UV 膜附近，引力子与费米子的耦合受到了强烈抑制。胶子在整个空间均匀分布，使得胶子-胶子的产生成为引力子的主要产生渠道。由于弱的矢量玻色子在自发对称破缺中吸收了 Higgs 自由度，因此它们的波函数在 IR 膜附近急剧下降，因此与 Higgs 的耦合与顶夸克和 Higgs 耦合类似。该理论的唯一自由参数是最轻的 KK 引力子的质量和空间因子与普朗克质量的比值 $\tilde{k} = \frac{k}{M_{Pl}}$ ，该参数控制共振态的宽度和截面。由于维度弯曲，在 UV 膜上的普朗克质量在 IR 膜上衰减为了 $e^{-\pi k R}$ ，从而解决了等级问题。基于额外维模型有一些双玻色子共振态的寻找分析，但是目前并没有发现新物理粒子。

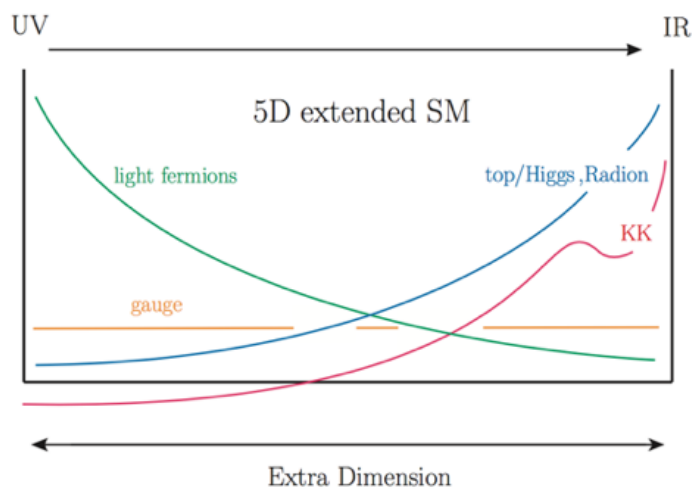


图 1.5 额外维模型的构成及标准模型粒子与 Kaluza-Klein 引力子在其传播。

1.2.4 拓展额外维模型

根据最新的额外维模型中 KK 规范玻色子味道及 CP 测试，它们的质量在大于 10TeV 的量级 [48、57]，因此 IR 膜的能量尺度应该在 $O(10)\text{TeV}$ 。但是 LHC 质子对撞的中心能量为 13TeV，可探测的信号能量在 TeV 量级。综上考虑，一个自然的想法就

是在现有的 UV 膜和 IR 膜之间增加一个中间膜，Higgs 膜，能量量级在 10 TeV，如图1.6所示 [3、9]。引力子可以全空间传播，新模型中所有的 $SU(2)_L \times U(1)_Y$ 规范玻色子场同样也可以传播到最终的 IR 膜，SU(3) 玻色子场、费米子只能在 UV 膜和 Higgs 膜之间传播。因此，模型预言的自旋为 0 的新粒子 radion 与规范玻色子有很强的耦合，而胶子与顶夸克或 Higgs 之间的耦合被压制。KK 模式下电弱规范玻色子、标准模型电弱规范玻色子及 radion 之间的拉氏量可写作：

$$\mathcal{L}^{EW} \supset \frac{g_V^2}{g_{V_{KK}}} V_{KK}^\mu J_{V\mu} + \left(-\frac{1}{4} \frac{g_{grav}}{g_{V_{KK}}^2} g_V^2 V_{\mu\nu} V^{\mu\nu} + \epsilon g_{V_{KK}} g_V V_{\mu\nu} V_{KK}^{\mu\nu} \right) \frac{\phi}{m_{KK}} \quad (1.29)$$

其中 V_{KK} 代表 KK 电弱规范玻色子，包括 W_{KK} 、 Z_{KK} 和 γ_{KK} ，而 V 表示标准模型电弱规范玻色子，包括 W 、 Z 及 γ 。 $V^{\mu\nu}$ 和 $V_{KK}^{\mu\nu}$ 是标准模型和 KK 模式规范玻色子强度张量， g_V 和 $g_{V_{KK}}$ 为对应的耦合常数， g_{grav} 是 KK 模式引力子耦合常数， ϕ 为标量 radion 场， ϵ 为约束参数。式中第一项表示 KK 激发态规范玻色子与标准模型费米子的相互作用，这一项与标准的额外维模型一致。第二项为 radion 粒子与标准模型电弱规范玻色子的耦合，这一项控制了 radion 到一对标准模型玻色子的衰变模式。最后一项是 radion、标准模型玻色子和 KK 激发态玻色子之间的耦合，描述 V_{KK} 到 V 和 radion 是最主要的衰变模式。允许的参数范围为： $1 \leq g_{grav} \leq 6$ ， $3 \leq g_{W_{KK}} \leq 6$ ， $\sqrt{g_{W_{KK}}^2 + 3^2} \leq g_{Z_{KK}} \leq \sqrt{g_{W_{KK}}^2 + 6^2}$ ， $0.3 \leq \epsilon \leq 0.5$ 。

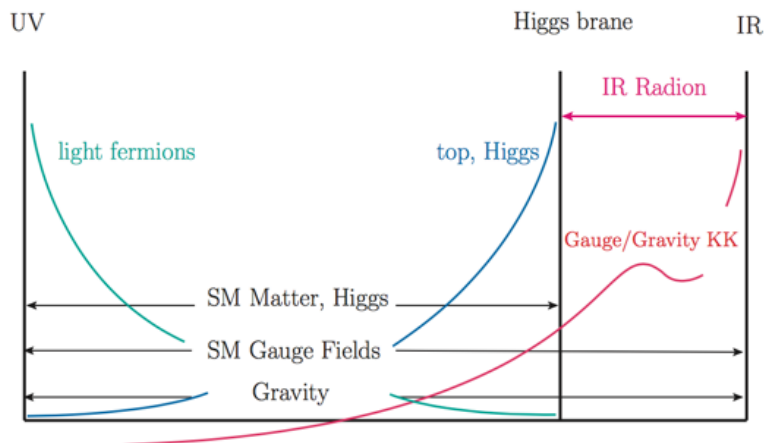


图 1.6 拓展额外维模型的构成及其间粒子的传播。

1.2.5 玻色子共振态研究现状

双玻色子共振态

大质量共振态的寻找是 LHC 上一个重要的课题。在过去的半个世纪中，在粒子对撞机上有很多通过两体衰变过程寻找更重的共振态的研究。LHC 上的 ATLAS 和 CMS 实验开展了末态为 2 个标准模型规范玻色子的分析用于研究大质量的规范玻色子 (W' 和 Z' ，统称为 V') 与标准模型粒子通过电弱作用的耦合，并且设置了质量和与共振态耦合参数的上限。新粒子状态可能主要与标准模型中费米子耦合，如最小 W' 和 Z' 模型 [20、56、84]，或与标准模型规范玻色子耦合，如强耦合复合 Higgs 和小 Higgs 模型 [67、73、76]。前文介绍的弯曲额外维模型也提供了大规模共振态的候选粒子，比如自旋为 2 的引力子 Kaluza-Klein (KK) 激发态 [7]，有较大的分支比可以衰变到一对 W 、 Z 和 H 玻色子。

表 1.1 总结了 CMS 基于 BSM 理论的分析，使用了 2016 年在 CMS 上采集的积分亮度为 35.9fb^{-1} 的数据。由于共振态的质量超过 1TeV ，因此在衰变中产生的标准模型玻色子具有较大的洛伦兹增强 (boost)，玻色子的衰变产物会很集中，非常准直，需要专用的重建技术来进行识别和重建。在强子衰变的情况下，衰变得到的—对夸克是用单个大锥角喷注重建，该喷注具有源自两个强子化夸克的结构。在双玻色子共振态的寻找中开发了很多应用于 boost 喷注的技术，将会在 3.4 章中介绍。将表 1.1 中所有分析合并可以得到更高的统计量，从而进一步提高排除限，合并后的结果如图 1.7 [88]，在 95% 置信度水平下 W' 、 Z' 质量分别低于 4.3、3.7 TeV 被排除。ATLAS 实验也有使用 2016 年的数据集开展的类似的分析 [1]，同样没有发现新物理的存在。

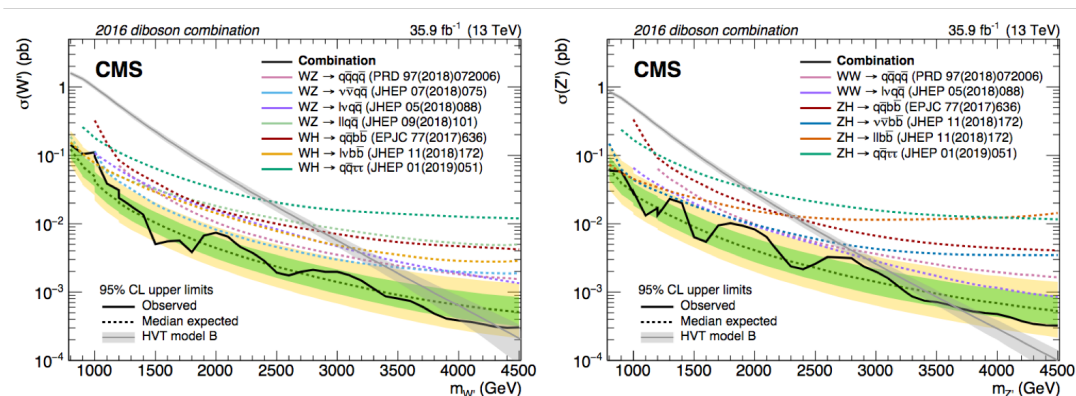


图 1.7 2016 年双玻色子共振态的寻找和合并后给出的质量排除限。

表 1.1 CMS 上对双玻色子共振态的寻找总结, 其中横纵向分别为两个玻色子的衰变模式, 表格中是相应分析的参考文章。

	$V \rightarrow q\bar{q}$	$W \rightarrow \ell\nu$	$Z \rightarrow \ell\ell$	$Z \rightarrow \nu\nu$	$H \rightarrow b\bar{b}$	$H \rightarrow \tau\tau$
$V \rightarrow q\bar{q}$	[98]	[94]	[93]	[92]	[97]	[96]
$W \rightarrow \ell\nu$	[94]				[95]	
$Z \rightarrow \ell\ell$	[93]			[99]	[95]	
$Z \rightarrow \nu\nu$	[92]		[99]		[95]	
$H \rightarrow b\bar{b}$	[97]	[95]	[95]	[95]	[90]	[96]
$H \rightarrow \tau\tau$	[96]				[96]	

三玻色子共振态

三玻色子共振态在 LHC 上并没有被直接测量过, 近些年的很多 BSM 中的物理情形涉及三个玻色子共振态 [8、12、64、69], 如图1.8所示。本分析是从拓展额外维模型着手, 该模型的引入带来了新颖的分析课题, 模型中引入了自旋为 1 的 KK 激发态对应的新的 W 和 Z 规范玻色子, 以及自旋为 0 的标量粒子 radion。图1.9、1.10为 W_{KK} 和 R 的衰变及分支比情况。本文的题目对应 $W_{KK} \rightarrow WR \rightarrow WWW$ 过程, 衰变分支比较高, 是 LHC 上第一次对于三个重玻色子的共振态的寻找。公式1.29中的参数采用推荐值, $g_{grav} = 6$, $g_{W_{KK}} = 3$, $g_{Z_{KK}} = 6.708$, $\epsilon = 0.5$ 。CMS 中采用同样模型的分析有 $g_{KK} \rightarrow gR \rightarrow ggg$, 是末态为三胶子共振态的寻找。而三玻色子共振态寻找是一个全新的课题, 其中四夸克喷注的标记与校准是一个难点, 为此开发了矩阵方法用于校准, 可以用于类似的多夸克喷注信号的校准, 因此本分析对新物理的寻找具有较为重要的意义。

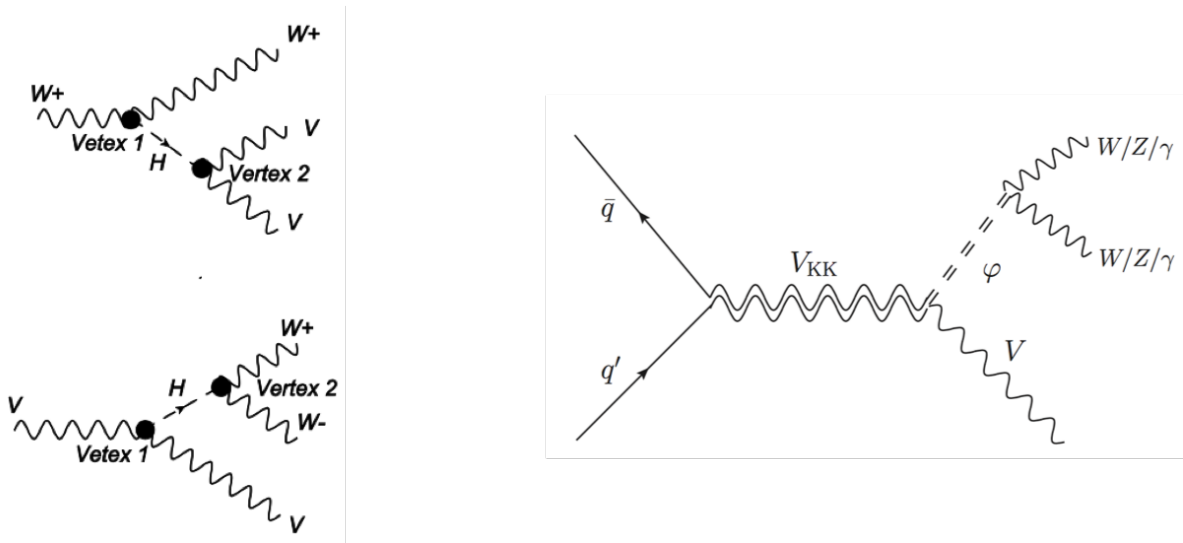


图 1.8 左图为邝宇平院士提出的 VH^* 到三玻色子的过程, 右图为拓展额外维模型预测的 V_{KK} 到三玻色子的费曼图。

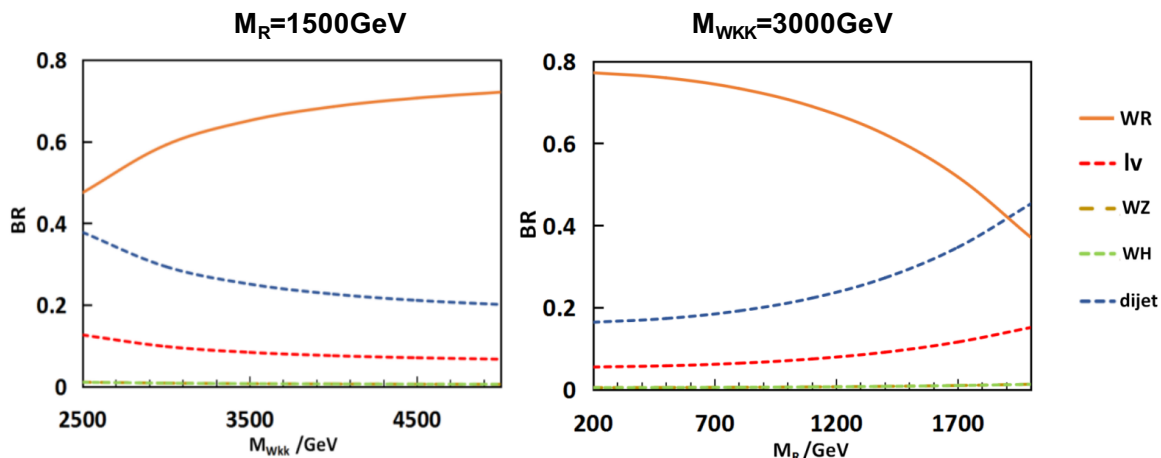


图 1.9 耦合系数取模型推荐值 $g_{W_{kk}}=3$, $g_{Z_{kk}}=6.708$, $g_{grav}=3$ 时固定 radion 质量为 1500GeV (左)、固定 W_{KK} 质量为 3000GeV 时其的衰变模式与分支比, 衰变到 WW 是最主要的过程。

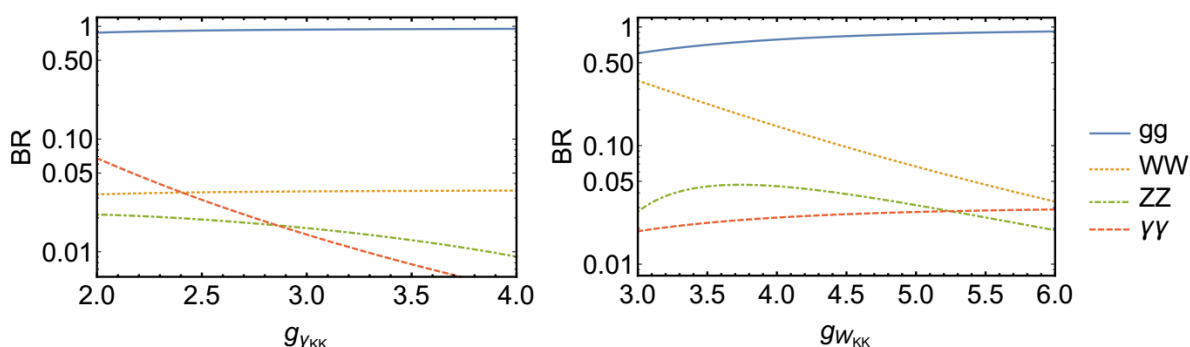


图 1.10 固定 $g_{W_{kk}}=6$, radion 的衰变分支比关于 $g_{Y_{kk}}$ 的分布 (左); 固定 $g_{Y_{kk}}=2.5$, radion 的衰变分支比关于 $g_{W_{kk}}$ 的分布。

1.3 论文结构

本文包括三个部分。第一部分包括前三章, 主要介绍 LHC、CMS 探测器及 CMS 实验中事例的重建。第二部分分为第四到第六章, 是实验分析的全部流程, 包括事例挑选、数据驱动的本底估计方法、为校准多夸克喷柱探针开发的矩阵修正技术及设定截面上限。第七章简单总结在分析之外担任 JME 和 B2G 两个组蒙特卡洛模拟样本联系人的服务性工作。第八章是分析工作的总结和展望。

第二章 LHC 对撞机与 CMS 探测器

2.1 LHC 对撞机

大型强子对撞机 (Large Hadron Collider) [74] 是世界上最大的、能量最高的大型强子对撞机, 于 2008 年 9 月首次通质子流测试, 半径为 27 公里, 能够将质子加速到质心能量约 13TeV。加速器由两个反向循环的质子束环组成, 它们彼此独立并且处于超高真空下。质子通过射频腔被加速到接近光速, 然后集中在环的四个不同的对撞点发生碰撞。这四个碰撞点对应于四个 LHC 粒子探测器, ATLAS、CMS、LHCb 和 ALICE。尽管 ATLAS 和 CMS 是通用检测器, 但其构建目的是为了研究各种不同的物理过程, LHCb 和 ALICE 建造是为了特殊用途, 其中 LHCb 适用于探测 b 物理过程, ALICE 适用于重离子碰撞。质子从气体到 LHC 对撞点的过程如图2.1所示, 首先压缩氢气经过电磁场电子被剥离剩下氢核, 即质子, 然后质子注入到线性加速器 LINAC2, 将质子能量增加到约 50 MeV。LINAC2 还通过仔细调整开关的频率, 将质子流分成等间距的束团。然后将被加速的质子注入质子同步加速器 (Proton Synchrotron Booster), 在其中将其能量增加 30 倍, 达到约 1.4GeV。质子到达 LHC 环之前的两个最终加速装置是质子同步加速器和超级质子同步加速器, 产生的质子的总能量为 450GeV。之后质子已准备好进入到最终阶段, 并沿相反方向注入 LHC 的两个束管中, 直到质子束均匀分布在两个环上。这时质子经过八个射频腔被加速到目前的最大能量 6.5 TeV, 这个过程大约需要 20 分钟。在质子束达到其最大能量并在 LHC 环中稳定循环之后, 束团开始对撞。每束大约 10^{11} 个质子, 束团大约相距 7.5m, 相隔 25ns 对撞一次, 涉及亮度大约为 $10^{34} \text{cm}^{-2} \text{s}^{-1}$ 。

RunI 数据采集阶段质心能量为 8TeV, CMS 实验组于 2012 年发现了 Higgs 粒子, 对标准模型的完善具有重要意义。RunII 阶段从 2015 年 6 月开始, 质心能量提升到 13TeV, 并且峰值亮度在设计亮度的 1-2 倍之间运行, 本文中使用的就是 2016-2018 三年采集到的 137fb^{-1} 的数据。

2.2 CMS 探测器

紧凑型 μ 子螺线管探测器 (Compact Muon Solenoid detector, CMS) [44] 是 LHC 上两个通用型探测器之一, 它的直径为 15m, 重量为 14000 吨, 体积比 ATLAS 探测器小 60%, 重量却是其两倍。CMS 有有史以来最大的超导磁铁, 能够产生 3.8T 的磁场, 连同相应的磁轭返回, 一共占探测器总重量的 90%。CMS 检测器是圆柱对称的, 结构

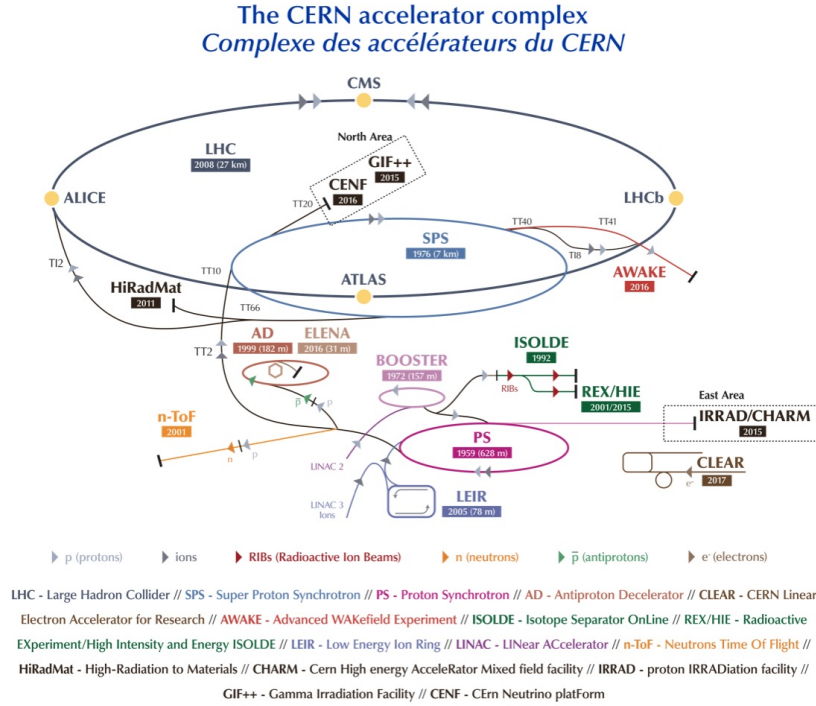


图 2.1 LHC 上质子经多级加速到达对撞点示意图。

如图2.2所示。内层径迹探测器分布于距束流管约 3cm 半径到大约 1.2m 的半径的圆筒环内，由内层硅像素检测器和外部硅微条探测器组成。径迹探测器外面是两个量能器：由钨酸铅闪烁晶体组成的电磁量能器（ECAL），负责测量电磁相互作用粒子的沉积能量，然后是测量强子能量的强子量能器（HCAL）。电磁量能器位于超导螺线管内部，这使得整个探测器相当紧凑，这也是探测器名字中 compact 的来源。最外侧为测量 μ 子动量的 μ 子探测器。接下来会详细介绍各个子探测器。

2.2.1 赝快度

如果用空间直角坐标系表示 CMS 探测器内的位置， z 轴的正方向沿束流管道指向西， x 轴的正方向指向 LHC 环的中心， y 轴的正方向指向地球的表面。由于 CMS 探测器是圆柱对称性的，使用极坐标会更方便。方位角 ϕ 是在 xy 平面中测量，其中 $\phi = 0$ 对应于正轴正向， $\phi = \pi/2$ 对应于 y 轴正向。相对于 z 轴测量极角 θ ， $\theta = 0$ 与指向 z 轴正向。为了定义粒子相对于束流的角度，定义赝快度 $\eta = -\ln(\theta/2)$ 比直接用 θ 更方便。这是由于以下事实：粒子的产生量对赝快度是常数，而且更方便进行洛伦兹变换。

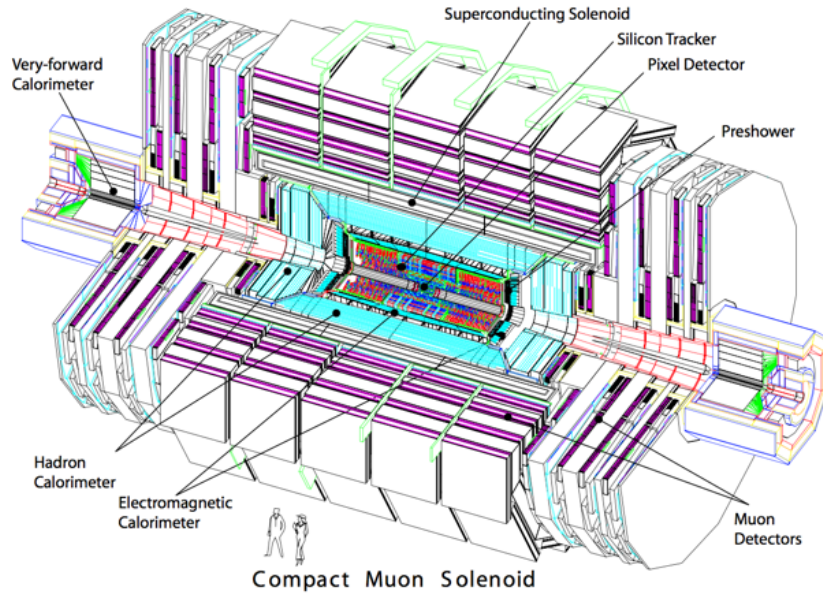


图 2.2 CMS 探测器结构示意图。

2.2.2 径迹探测器

径迹探测器由内层硅像素检测器和外部硅微条探测器组成，负责精确重建带电粒子的动量。硅像素探测器位于最靠近相互作用点且粒子强度最高的位置。它于 2017 年进行了升级，从所谓的 Phase-0 升级到 Phase-1，其结构为四个圆柱筒形层，半径分别为 2.9、6.8、10.9 和 16.0cm。最靠近束流管处撞击强度非常高，因此需要严格限制探测器像素大小。传感器单元像素为 $100\mu\text{m} \times 150\mu\text{m}$ ，厚度为 $285\mu\text{m}$ ，总计为 1.24 亿个通道。传感器安装在每个带有 16 个读出条的探测器模块上，内层读出条的工作频率为 600 MHz/cm^2 ，高精细的粒度单元及高速的读出能力保障了径迹的探测效率。

当远离束流管时，粒子束流强度降低，因此像素探测器被更大尺寸的硅微条探测器所取代，总共有十个条状层，延伸到大约 1.3m 的半径位置。这些探测器被分为四部分：具有四个条层的内筒（TIB），两个内端盖（TID），包含六个圆柱层的外筒（TOB）以及两个端盖（TEC），具体结构如图 2.3 所示。TIB 和 TID 中的条带长 10cm，宽度为 $80\mu\text{m}$ ，厚度为 $320\mu\text{m}$ 。TOB 和 TEC 部分由长宽 $25\text{cm} \times 180\mu\text{m}$ ，厚 $500\mu\text{m}$ 的稍大的条带组成。微条探测器总共有 1000 万个微条，覆盖面积 $\sim 200\text{m}^2$ 。为了延长径迹探测器的使用寿命，整个探测器通过专用的冷却系统保持在 -20°C 。径迹探测器的覆盖范 $|\eta| < 2.6$ 的区域，分辨率约为 $\sigma/p_T \approx 1.5 \times 10^{-5} p_T \oplus 0.005$ 。

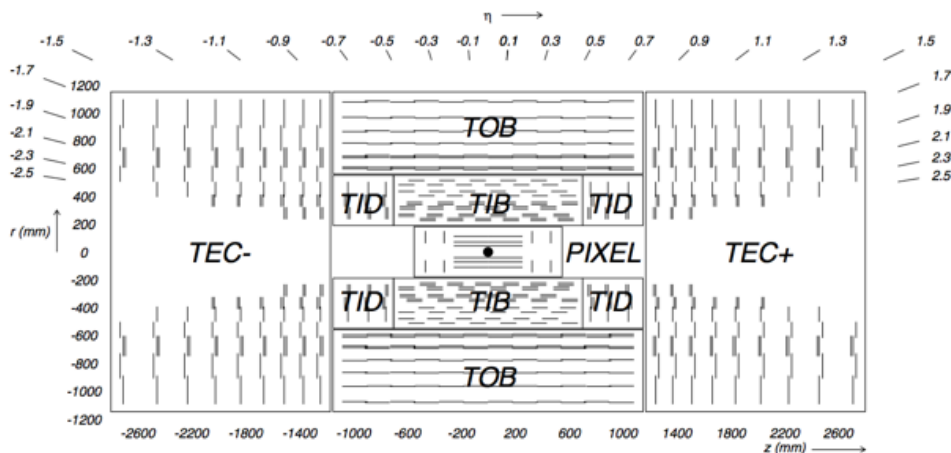


图 2.3 径迹探测器的横截面示意图，每条线代表一个检测器模块。

2.2.3 电磁量能器

径迹探测器外侧是电磁量能器（ECAL），它由 75848 个钨酸铅（ PbWO_4 ）晶体组成，具有很好的能量和位置分辨率。对能量高于 100GeV 的光子和电子设计的能量分辨率为 0.5%，因此电磁量能器的材料选择一直是其最关键的设计特征。为了承受探测器中高剂量辐射和强磁场，同时需要在两次粒子碰撞之间的 25 纳秒内产生清晰的响应信号，选择了重金属钨酸铅晶体。其密度为 8.28g/cm^3 ，晶体足够致密，又不占用过多的体积。钨酸铅的辐射长度为 0.85cm，莫里埃半径为 2.19cm，因此可以很好的区分强子和光子的能量沉积。钨酸铅晶体的最大缺点是光产率低，需要专用的雪崩光电二极管来增加增益。电磁量能器是完全封闭的，分为覆盖 $|\eta| < 1.48$ 的筒部（barrel）和两个将覆盖范围扩大至 $|\eta| < 3$ 的端盖部分（endcap）。为了提高对 γ 和 π_0 的分辨能力，前端 $1.65 < |\eta| < 2.6$ 的区域安装了使用铅吸收剂和硅传感器的顶簇射探测器（pre-shower detector）。筒部的晶体被划分为每个含大约 1700 个晶体的超级模块，端部则分为两个半盘，每个半盘由 3662 个晶体组成。每个晶体单元重约 1.5 千克，呈略带锥形的形状。电磁量能器的布局如图2.4所示。

电磁量能器晶体全部是横向排布，没有纵向排布，因此依赖径迹探测器对事例初级顶点的准确重建，以便正确重建光子角度。其能量分辨率包括随机，噪声和常数项三部分 [2]：

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%,$$

0.3% 的常数项源于纵向光收集的不均匀性，能量分辨率与电子能量的关系如图2.5所示。

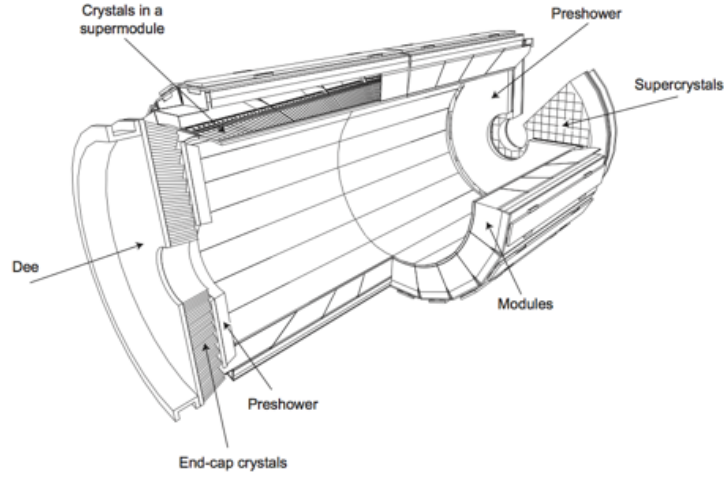


图 2.4 电磁量能器的布局示意图，包括筒部、端部及前端的顶簇射探测器。

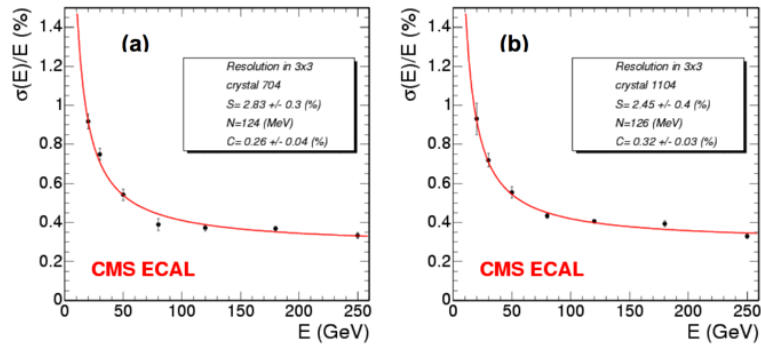


图 2.5 电磁量能器两块不同的晶体能量分辨率随能量变化的分布函数。

2.2.4 强子量能器

电磁量能器外测是强子量能仪（HCAL），由致密的黄铜吸收剂材料和塑料闪烁体的交替层组成。当粒子撞击吸收板时，它将与吸收材料相互作用，并以次级粒子为中心产生簇射。这些粒子然后在闪烁材料中产生与它们的能量成比例的光，并且将连续层上的产生光的总量相加代表初始粒子的能量。综合电磁量能器与强子量能器的响应重建喷注的能量和缺失横能量来测量夸克、胶子和中微子的能量。强子量能器分为四个区域：内筒（HB）、外筒（HO）、端盖（HE）和前向区域（HF），如图2.6所示。内筒位于超导螺线管内，覆盖 $|\eta| < 1.3$ 的区域，由 36 个相同的楔形块组成，每个楔形块重 26 吨，分成两个半筒（HB+ 和 HB-）。楔形块由平行于束流轴的扁平黄铜吸收板组成。这些板由一个 4cm 厚的前钢板，然后是八个 5cm 厚的黄铜板，六个 5.6cm 厚的黄铜板，最后是一个 7.5cm 厚的钢背板组成。吸收板与 4mm 厚的塑料闪烁体砖交替排列，这些闪烁体砖是探测器的活性介质，能量沉积由塑料纤维读出。筒部强子量能仪在与相互作用长度相关的有效厚度随着极角 θ 的增加而增加，从 90 度角处的 $5.8\lambda_I$ 开

始，并在 $|\eta| < 1.3$ 时增加到 $10.6\lambda_I$ 。厚度越厚，可沉积的能量越多，能量分辨率越好。

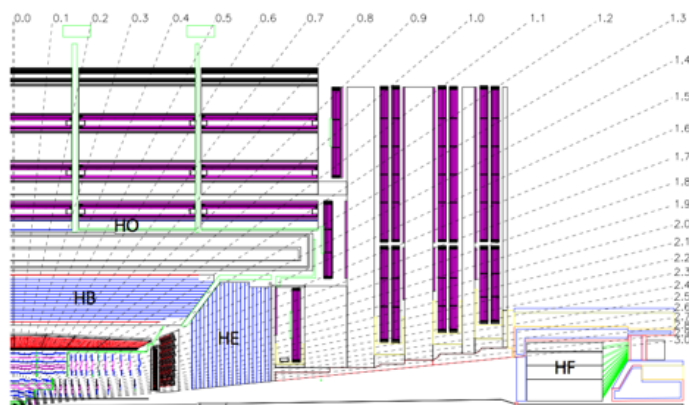


图 2.6 强子量能器结构示意图，包括内筒、外筒、端盖和前向四个区域。

在 $|\eta|$ 中心区域，由于内筒被限制在超导线圈内，电磁量能器和强子量能器的组合相互作用长度太小，不足以完全吸收强子簇射。为了确保足够的采样，因此在电磁线圈的外面增加了一层闪烁体作为外筒（HO）。它使用线圈本身作为吸收材料，从而将筒形量能器的总相互作用长度增加到 $11.8\lambda_I$ 。强子量能器端盖（HE）位于靠近束流管的前部区域，并覆盖 $1.3 < |\eta| < 3.0$ 的区域，该区域包含约 35% 的对撞产生的粒子。由于端盖非常靠近束流管，因此端盖需要极高的处理信号速率并具有很强的辐射耐受性。由于端盖区域中的分辨率会受到堆积效应（pileup）和磁场效应的影响而降低，强子量能器端盖设计时优先保障最大程度地减少与筒部之间空隙，而不是优先考虑单粒子分辨率。端盖中的吸收板交错安装排列，而不是像在筒部一个正好覆盖另一个的形式安装，从而提供封闭的自支撑结构。强子量能器是以塔的组合形式工作，单个塔大小为 $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ ，对于较大的赝快度区域为 0.17×0.17 。

为了获得完全封闭的强子量能器，在前向区域添加了一个额外的前向量能器（HF）。延伸到赝快度满足 $|\eta| = 5.2$ 的区域，此时探测器的位置非常靠近束流管，以至于粒子采集速率超过 $10^{11}/\text{cm}^2$ ，每次质子-质子对撞接收约 760 GeV 的能量，而其余探测器平均为 100 GeV。它由两个圆柱形钢结构组成，每个钢结构的外半径为 130cm，内半径为 12.5cm，位于相互作用点两侧的 11.2m 处，由带凹槽的 5mm 厚的钢制吸收板组成，并将石英纤维活性介质插入这些凹槽中。电磁量能器与强子量能器对 π 介子的能量分辨率作为能量的函数如图 2.7 所示 [87]。强子量能器典型的电子噪声为每座塔 200MeV。对于能量为 300GeV 的介子，前向量能器的加入可将分辨率提高 10%。与电子量能器一样，使用内筒和外筒时的最终能量分辨率参数化由随机和常数项给出，可表示为

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\%$$

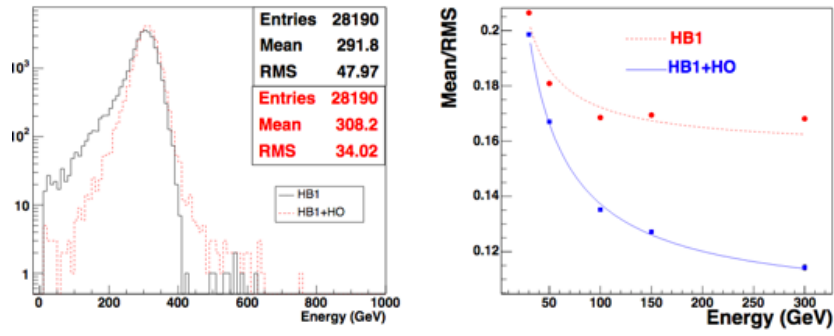


图 2.7 强子量能器仅使用内筒与同时使用内筒和外筒时能量的分布（左）及能量分辨率的对比（右）。

2.2.5 μ 子探测器

CMS 探测器最外测是专用于 μ 子鉴别，动量测量和触发的 μ 子探测器，由三种不同类型的气体探测器组成：漂移管室 (drift tube, DT)、阴极条室 (cathode strip chambers, CSC) 和阻抗板室 (resistive plate chambers, RPC)，如图 2.8 所示 [61]，它们均集成在磁轭返回结构中。在筒部区域中通过的粒子较少且磁场均匀，因此使用漂移管室并覆盖 $|\eta| < 1.2$ 区域。但是在端盖区域， μ 子计数率和背景粒子要高得多，磁场本身很强且不均匀，因此在此区域使用具有更精细、具有更高抗辐射性和收集信号更快的阴极条室，覆盖 $0.9 < |\eta| < 2.4$ 的区域。阻抗板室被用作专用的 μ 子触发系统，在筒部和端盖中都有安装。它们提供了出色的时间分辨率，并覆盖了 $|\eta| < 1.6$ 区域。

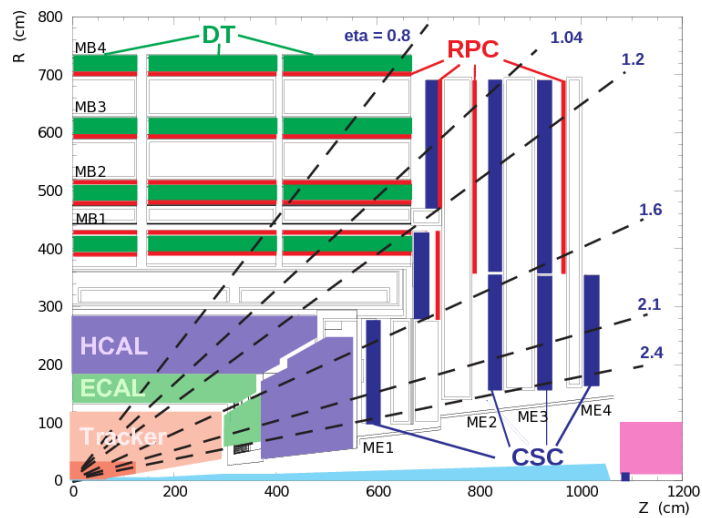


图 2.8 μ 子探测器结构示意图。

2.2.6 触发系统

LHC 中质子-质子束团以 40MHz 的频率交汇，两次交汇之间只有 25 纳秒的时间可用于处理数据，在设计亮度下，每秒大约有 10 亿次对撞，而每个事例大小大约为 1MB，因此不可能将所有事例都读出并存储到磁盘中。CMS 触发系统旨需要对事例是否有用做出快速且高质量的判断。触发的第一阶段 L1，旨在通过定制设计的硬件将事件输出率降低到最大 100kHz。它使用来自 μ 子系统和量能器的粗略数据来决定是否应该记录该事例，每次决定需要在 $3.2\mu s$ 之内完成。在这段时间内，来自每个子探测器的完整事件数据存储在探测器前端电子设备中，等待 L1 决策。L1 分为三个步骤，如图 2.9 所示，对于 μ 子系统，记录由三种类型的 μ 子探测器中的探测到的轨道片段。对于量能器，通过计算触发塔 $\Delta\eta - \Delta\phi$ 为 0.087×0.087 中的横向能量，并将其与正确的束团交汇点匹配。然后将来自量能器的信息传递到区域触发器并定义电子、介子和喷注等候选对象。这些信息中的一部分传递给区域 μ 子触发器，用于提供有关粒子是否最小化电离的信息。然后全局 μ 子触发器将径迹信息与量能器信息结合起来，最多选择四个 μ 子候选者，并计算它们的动量、位置、电荷和质量。区域量能器触发的输出也将传递到全局量能器触发，该触发器将提供有关喷注总横向能量和丢失能量的信息。结合来自全局 μ 子触发器和全局量能器触发的信息，L1 触发器通过简单的逻辑运算来决定是否保留事例。

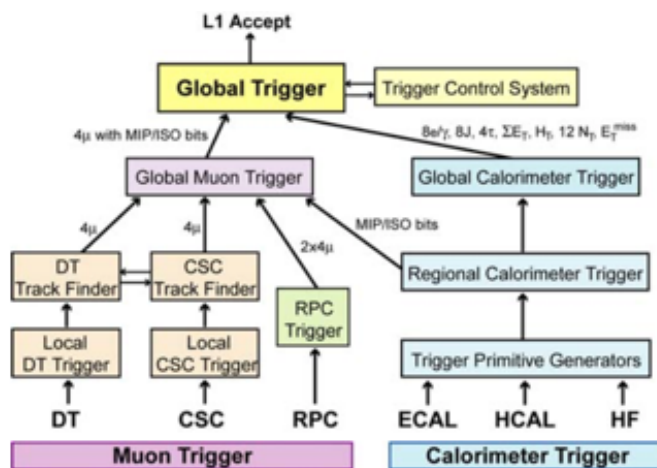


图 2.9 CMS 中触发流程。

如果事例通过触发，那么该事例的全部信息将以 100kHz 的速率被读出，并传递给高水平触发器 (HLT)。HLT 会使用探测器高精度的数据，基于离线重建算法进行决策。HLT 的目标是最终将事件输出率降低到平均 400 Hz。

第三章 CMS 探测器上事例的重建

3.1 径迹与初级顶点的重建

CMS 径迹探测器在每次质子-质子束流对撞时会同时被对撞产生的大约 1000 个带电粒子穿过，因此径迹重建极具挑战性。在整个 Run II 数据采集期间，每次对撞的反应次数如图3.1所示，平均为 34 次。径迹重建包括以下过程：从内层硅像素探测器获

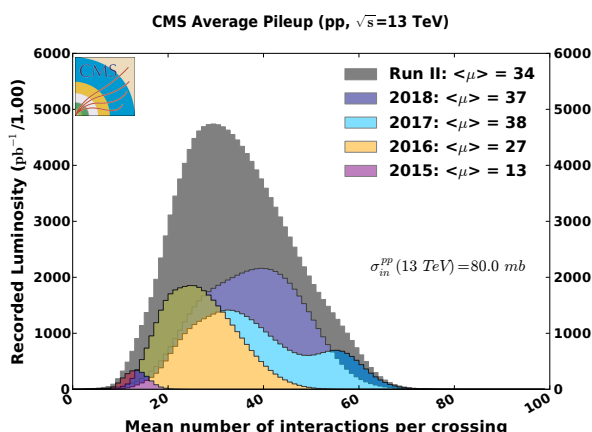


图 3.1 CMS 探测器在整个 Run II 数据采集期间平均每个事例中反应次数 [27]。

取打火点组合为粒子径迹，然后估计产生该打火点的带电粒子的动量和飞行方向。在 CMS 上，计算是基于所谓的组合卡尔曼过滤器 (Kalman filter)[22]。卡尔曼过滤器是一种使用时间相关的观测值来估计未知变量的算法从而确定径迹的方向。CMS 中的径迹重建软件被称为组合径迹寻找器 (Combinatorial Track Finder, CTF)，通过迭代循环打火点来重建径迹，然后删除那些已经用作前一径迹输入的打火点来构造径迹集合。它从最里面的径迹探测器层中两个或三个打火点作为种子开始，寻找与该候选者关联的其他打火点以推断种子径迹。它会同时考虑到打火点和径迹的不确定性以及丢失打火点的数量，基于 χ^2 的计算结果，忽略拟合失败的径迹，径迹重建算法在整个径迹探测器覆盖范围 $|\eta| < 2.5$ 内有效，并且可以重建动量低至 0.1 GeV 及距离束流 60 cm 以内的粒子。在中心区域，动量为 100 GeV 的粒子的 p_T 分辨率约为 2.8%，横向位置分辨率为 $10\mu\text{m}$ ，纵向约为 $30\mu\text{m}$ 。

重建算法将通过特定选择标准的上一步重建径迹作为输入，将共享共同起点的径迹聚类，将所有径迹的共同起点定义为初级顶点，通常位于距束流轴数 mm 的半径内。要求每条径迹在硅像素探测器层中至少有 2 个打火点，在硅像素和微条探测器中不少于 5 个打火点，并且径迹拟合满足 $\chi^2 < 20$ 才能被用于作为顶点寻找器的输入。主

顶点要求有至少 50 条径迹，分辨率在 x 方向约为 $12\mu\text{m}$ ，在 z 方向约为 $10\mu\text{m}$ 。

离线软件要求所有事例都必须在沿束流轴的 24cm 窗口内重建至少一个主顶点，并且与反应区域中心的横向距离应小于 2cm 。选择所有粒子求和的 p_T^2 最大时的重建顶点作为发生硬散射过程的初级顶点。

3.2 粒子流算法

质子束对撞后首先经过径迹探测器重建径迹，然后经过电磁量能器簇射，最后在 μ 子探测器打火。为了关联并整合各个子探测器得到的信息，CMS 使用了粒子流算法 (Particle Flow Algorithm, PF) 来判断事例中真实产生的粒子信息。这些重建物理对象包括：

- μ 子，通过径迹探测器和 μ 子探测器中的打火点重建；
- 带电强子，通过径迹探测器中的打火点和电磁量能器中的电磁能量沉积重建；
- 中性强子，会在电磁量能器中沉积能量但不会在径迹探测器中留下打火点；
- 光子，只在电磁量能器中沉积能量而非强子量能器，并且不会在径迹探测器中打火；
- 电子，会在径迹探测器打火并在电磁量能器沉积能量。

不同粒子在 CMS 探测器中留下的轨迹如图3.2所示。图3.3描述了 PF 算法来重建物理对象的例子，左边是包含五个粒子的喷注在强子量能器留下的簇团，右图为电磁量能器。在强子量能器中，观测到两个簇团 H_1 、 H_2 ，并且两个带电粒子 π^+ 、 π^- 被重建为 2 条径迹 T_1 、 T_2 ，且指向 2 个簇团。在电磁量能器中， K_L^0 、 π^- 和 π^0 衰变产生的两个光子共产生 4 个簇团。

3.2.1 μ 子

μ 子重建可以通过可高效鉴别 μ 子的 μ 子探测器及精确测量动量的内层径迹探测器实现。在 CMS 中，定义了以下 3 种：

- 独立 μ 子 (standalone muon)：只基于 μ 子探测器中的径迹信息重建；
- 全局 μ 子 (global muon)：如果独立 μ 子径迹与在内层径迹探测器中的相互匹配；
- 径迹 μ 子 (tracker muon)： μ 子径迹横向动量大于 0.5 GeV ，总动量大于 2.5 GeV ，并且至少有一个 μ 子探测器中的打火点与径迹探测器中相匹配。

对于 $|\eta| < 2.4$ 的 μ 子，99% 都被重建为全局 μ 子或者径迹 μ 子或者两种都是。

在 PF 算法中，孤立 μ 子是最先被挑选的，要求径迹横向动量和及在 $\eta - \phi$ 平面上距离 $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ 的圆锥张角内的沉积能量小于 μ 子动量的 10%。非孤立的 μ 子需要通过严格的挑选条件 [46]，在 μ 子探测器中至少有三段匹配的径迹段或者

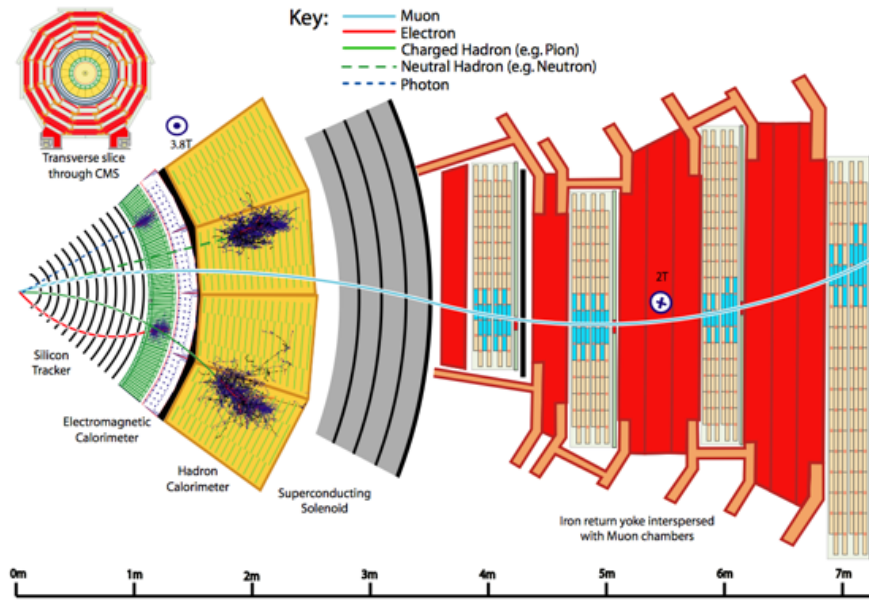
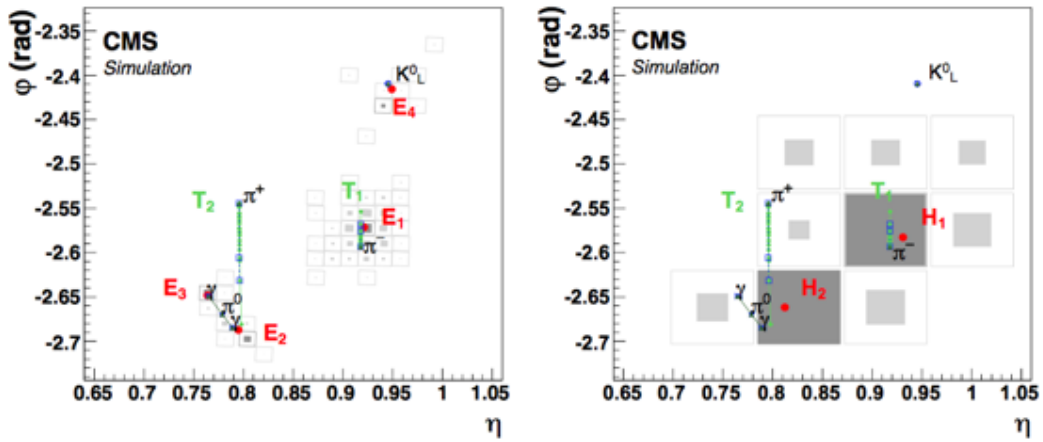


图 3.2 CMS 探测器中不同粒子反应示意图 [13].


 图 3.3 K_L^0 、 π^+ 、 π^- 和 π^0 衰变产生的两个光子，共五粒子组成的喷注在强子量能器（左）和电磁量能器（右）中的能量沉积。方格对应量能器单元模块，方格灰色面积与沉积能量的对数成正比，绿色线条表示在径迹探测器中重建的径迹。[13]

在量能器中的能量沉积与假设是 μ 子时的结果匹配。如果 μ 子不满足以上条件，除非独立 μ 子的径迹有很高的质量并在 μ 子探测器中留下大量的打火点，都不会被保留。如果 μ 子在径迹探测器中测得的横向动量小于 200 GeV 时，选取这个测量值作为最终的横向动量。否则，将根据不同拟合方式选取 χ^2 最小的方式 [13]。

3.2.2 电子与光子

电子重建是有基于电磁量能器能量沉积和基于径迹两种不同的模式。在前一种模式中，由于大量的电子能量在到达电磁量能器之前被韧致辐射产生的光子带走，因此选取电磁量能器超簇团时在 η 方向窗口较小，在 ϕ 方向窗口较大从而完全包含电子和产生的韧致辐射光子。为了减小这种超簇团选取方式带来的污染，需要采用较严格的孤立条件，会导致重建效率下降。因此引入基于径迹重建方式，所有横向动量大于 2 GeV 的径迹都被用于与电磁量能器簇团匹配，要求沉积能量大约与径迹动量相等。之后候选电子被用于 GSF (Gaussian-sum filter) 拟合，径迹探测器中打火数目、GSF 径迹的 χ^2 ，沿着径迹的能量沉积以及 GSF 径迹延长线与最近电磁量能器簇团的距离都用于 boosted-decision-tree (BDT) 输入来得到候选电子的判决条件。为了区分电子与带电强子，要求在 $\Delta R < 0.15$ 内强子量能器中的能量沉积小于在电磁量能器中的沉积。候选电子需要通过专用的基于 BDT 的条件。在这一步中，孤立光子也被从横向能量大于 10 GeV 的电磁量能器超簇团重建，要求与 GSF 径迹无关联以与电子区分。之后所有用于重建电子和孤立光子的径迹和能量沉积都从 PF 块中去除。

3.2.3 强子

在移除 μ 子、电子和孤立光子之后，PF 算法会鉴别剩余的强子和非孤立光子。分别在强子量能器、电磁量能器中产生簇团并且无法与任何径迹匹配的粒子被鉴别为中性强子、光子，后者是基于中性强子在电磁量能器中的沉积能量小于 3% 的判断。剩余的强子量能器簇团都会和至少一条径迹相关联。此时会通过比较径迹动量之和与量能器能量来确定粒子的组成。如果两者大致相等，则可以推断出每个粒子对应一条径迹并且相应的能量等于径迹动量。如果沉积能量大于径迹动量之和，每条径迹除了对应一个带电强子外还需要加一个光子或者中性强子。

3.2.4 丢失横能量

由于中微子不与探测器反应，导致出现探测器横向能量不对称的情况，因此丢失横动量被定义为一个事例中所有重建 PF 粒子横向动量矢量和的负值：

$$\vec{p}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}, \quad (3.1)$$

它的模被定义为丢失横能量 E_T^{miss} (MET)，用于近似代替中微子的横向动量。

3.3 Pile-up 消除

非产生于主顶点的质子-质子碰撞的相互作用的粒子被称为 **pileup** 事例，与电子线路中的堆积信号类似，这些事例会污染我们感兴趣的源于硬散射过程的观测量，因此必须通过专用的 **pileup** 去除技术加以消除。

3.3.1 Charged Hadron Subtraction (CHS)

如前面小节所述，CMS 中使用带电强子的径迹来重建反应顶点。如果事例的硬散射顶点与主顶点不一致，则将与该顶点关联的经粒子流算法重建的带电强子从粒子流的列表集合中移除，并且不会后续的任何物理对象重建，该方法称为 **CHS**。

3.3.2 Pile up per particle identification (PUPPI)

CHS 只能消除带电 **pileup** 粒子的影响。2014 年，提出了一种新的性能更强的 **pileup** 消除算法，即 **PUPPI**[21、62] 算法。**PUPPI** 使用局部形状信息，事例 **pileup** 属性和径迹信息的相结合来计算权重，该权重描述了给定粒子可能源于 **pileup** 的可能性。首先，基于来自 **pileup** 的软辐射与较硬的共线 **QCD** 模式之间的差异，计算表示为 α 的形状变量。然后将带电粒子的 α 形状用作所有 **pileup** 粒子的替代，并在逐个事件的基础上用于计算每个粒子的权重。该权重描述了哪些事例像 **pileup**，并用于重新按比例缩放粒子的四动量。给定粒子 i 的形状变量定义为

$$\alpha_i = \log \sum_{\substack{j \in \text{Ch, PV} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right) \times \Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij}), \quad (3.2)$$

其中 Θ 是 Heaviside 阶跃函数， ΔR_{ij} 是粒子 i 和 j 在 $\eta - \phi$ 平面上的距离， j 是指与初级顶点相邻半径为 R_0 的圆锥体内所有来自初级顶点的带电粒子。如果带电粒子与事件的初级顶点相关联，或者与初级顶点的距离 $d_z < 0.3$ cm 之内，则认为该带电粒子来自初级顶点。判断粒子是否来自 **pileup** 的概率定义为

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}, \quad (3.3)$$

其中 $\bar{\alpha}_{PU}$ 是给定事例中 **pileup** 粒子的 α_i 分布的中心值，而 RMS_{PU} 是其均方根值。如果一个粒子来自硬的物理过程，它和来自同一个父代粒子的其他粒子靠的很近，因此 α_i 较大，而来自 **pileup** 的粒子并没有类似的团簇结构，与来自初级顶点的粒子没有相关性，因此与粒子 i 的距离是随机的， α_i 通常较小。对应来自初级顶点的中性粒子，近

似认为与带电粒子的分布相同，可以用带电粒子的 α 形状作为代理。然后为每个粒子（中性粒子和带电粒子）分配权重 $w_i = F_{\chi^2, NDF=1} \chi_i^2$ ，其中 $F_{\chi^2, NDF=1}$ 是自由度为 1 的 χ^2 分布的累积分布函数。 $w_i < 0.01$ 的粒子将被排除掉。图3.4展示了 PUPPI 算法、CHS 及粒子流 PF 得到的 p_T 分辨率的对比，PUPPI 得到的分辨率明显优于 CHS 及 PF 喷注，并且更加对称。图3.5为质量分辨率作为 p_T 和 N_{PV} 的函数，可以看到 PUPPI 有显著的提升效果。

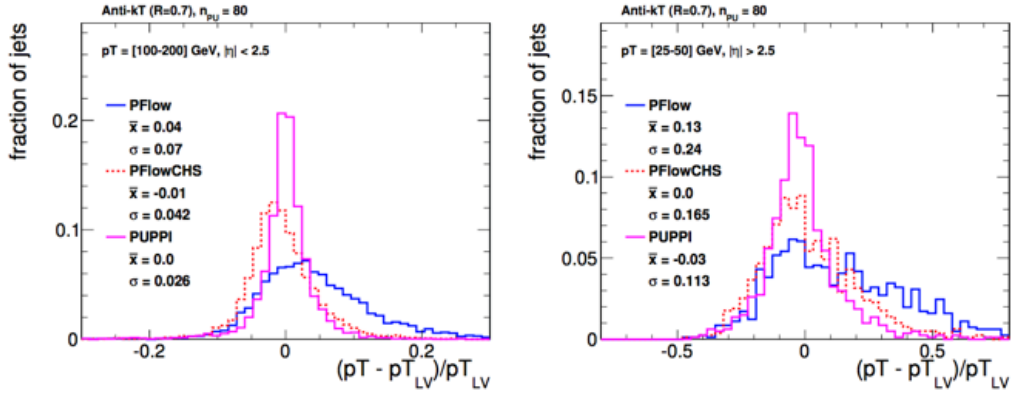


图 3.4 喷注 p_T 在 $N_{PV} = 80$ 时的分辨率：100 < p_T < 200 GeV（左），25 < p_T < 50 GeV（右），其中 $p_{T_{LV}}$ 为初级顶点（Leading Vertex, LV）的横动量。

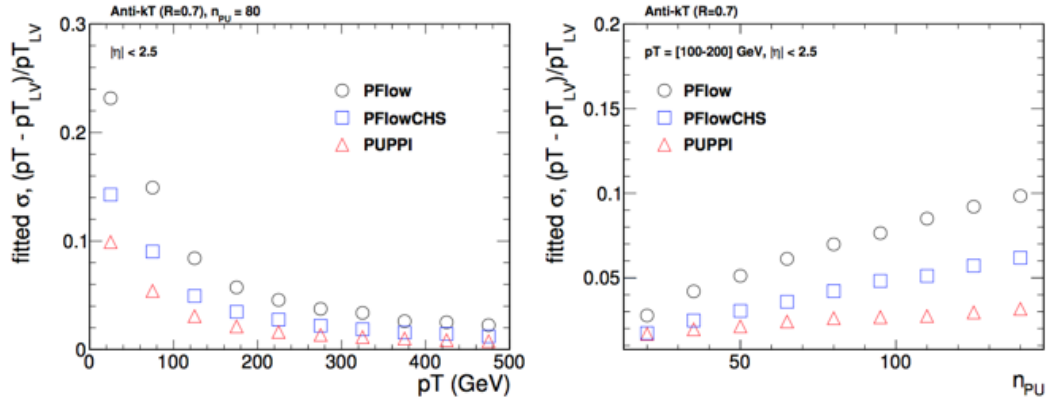


图 3.5 PUPPI 算法、CHS 及粒子流 PF 得到的喷注 p_T 分辨率按 p_T 分布（左）和按 N_{PV} 分布（右）的对比。

3.4 喷注的重建

在探测器中并不能直接探测带有颜色的夸克和胶子，只能观测到经过强子化后形成的强子喷注。通过研究喷注的特性可以推断出产生喷注的原始部分子的特性。目前

存在几种算法喷射聚类算法提供了将粒子分组为喷注的规则，分组通常基于粒子之间的特定距离要求以及粒子动量的要求。在粒子流算法中已经定义了带电强子、中性强子和光子等对象，以及它们的能量和方向，因此 CMS 中的喷注聚类就是将这些粒子与一个共同的起源相关联。

3.4.1 喷注聚类算法 (jet clustering)

在强子对撞实验中最常用的喷注聚类算法是 k_T 算法 [51]、Cambridge/Aachen 算法 [50] 和 anti- k_T 算法 [25]。这些都是顺序重组算法，这意味着它们会系统地遍历每个粒子对，并在组合满足特定条件时将它们重组为一个粒子。这三种算法共享的规则如下：

1. 对于每对粒子 i 和 j ，按如下公式计算距离

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.4)$$

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \quad d_{iB} = p_{Ti}^{2p}, \quad (3.5)$$

其中 d_{ij} 是粒子 i 和 j 之间的 k_T 距离， ΔR_{ij}^2 是两粒子在 $\eta - \phi$ 平面上的距离， R 对应于喷注的圆锥张角。 p 是用来将三种算法彼此分开的参数。对于 anti- k_T 算法， $p=1$ ；对于 k_T 算法， $p=0$ ；对于 C/A 算法， $p=0$ 。如何选取三种算法将会在下面进行详细说明。

2. 找到 d_{ij} 和 d_{iB} 中最小的距离。
3. 如果 d_{ij} 最小，那么将粒子 i 和 j 组合起来后返回步骤 1。
4. 如果 d_{iB} 最小，那么认为粒子 i 是末态喷注，并且将其从粒子列表中去除，然后继续返回步骤 1。
5. 重复上述操作直到没有粒子剩余。

红外和共线安全 (infrared and collinear safety)

定义喷注算法时，必须满足红外和共线安全的要求。红外安全性对应于以下要求：由于 QCD 事件（微扰和非微扰）中始终存在软发射，如果最终状态粒子由于存在软粒子发射而被修改，应保证硬粒子喷注保持不变。如果算法是共线不安全的，那么意味着如果硬部分子经共线分裂导致喷注的配置将发生变化。比如高能部分子经历共线分裂辐射出来一个小角度的胶子，需设计阈值或其他条件使得重建得到的喷注保持不变。因此喷注算法设计时要尽量满足这两点要求。

k_T 算法

k_T 算法是顺序重组算法中最古老的算法之一，由于其在公式 3.4 和 3.5 中的 $p=1$ 的定义，因此在 p_T 和角度方面均遵循 QCD 分支结构。首先将软粒子聚类在一起，最后才将两个最硬的粒子聚类在一起。按该定义，很有可能一个附近没有其他粒子的软粒子被定义为喷注，因此应该需要引入能聚类为喷注的最小 p_T 。尽管这种算法有一些优势，但由于大多数对撞机实验产生的喷注不规则， k_T 算法会优先聚类软粒子导致效果不佳。

Cambridge/Aachen 算法

Cambridge/Aachen(C/A) 算法要求公式 3.4 和 3.5 中 $p=0$ ，聚类顺序仅基于空间角度。这是最简单的算法，它会重新组合距离 ΔR 较小的粒子直到 $\Delta R_{ij} > R$ 。这样做的好处是，聚类历史记录中包含任何与喷注的几何子结构有关的信息，这一特征对研究喷注的子结构十分重要。

anti- k_T 算法

CMS 中的默认喷注聚类算法是 anti- k_T 算法，该算法同样遵循公式 3.4 和 3.5，其中 $p=-1$ 。该算法优先考虑大横动量粒子，这意味着该算法围绕硬粒子生长，主要倾向于高横动量粒子之间的聚类，而非软粒子之间的聚集，因此会产生具有明确定义的圆锥形区域的喷注。由于该算法满足红外和共线安全性，因此被选为 CMS 中的主要喷注重建算法。图 3.6 显示了 k_T 、C/A 和 anti- k_T 聚类后在 $\phi - \eta$ 平面中所得喷注区域的比较。 z 轴对应于部分子横动量 p_T 。可以清楚地看到，当使用 anti- k_T 算法进行聚类时，生成的喷注对应一块以硬粒子为中心的对称圆形结构，其半径由圆锥张角 ΔR 决定。

3.4.2 PF 喷注

CMS 中的喷注算法主要使用 PF 候选者的四向量作为输入，通常先经过 pileup 消除算法去除软粒子，然后再进行喷注聚类。如果使用 CHS 算法，则在聚类之前会丢弃与主顶点无关的带电强子。如果使用 PUPPI，所有 PF 候选对象都要根据其来自于 pileup 的可能性来重新加权。对于 anti- k_T 算法，CMS 常用的两个喷注半径为 $R=0.4$ 和 $R=0.8$ 。 $R=0.4$ 的喷注称为 PFAK4，用于研究只有一个子喷注的情况，而较大的 $R=0.8$ 对应的是 PFAK8 喷注，在寻找包含多个硬夸克、胶子的喷注时更常用，可以容纳全部的强子化末态。

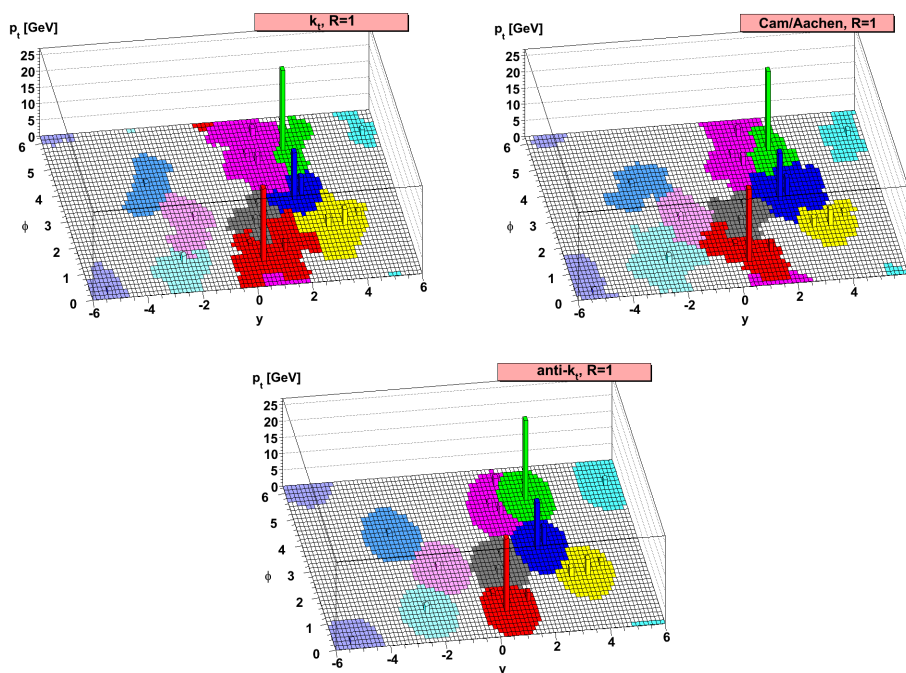


图 3.6 三种不同喷注重建算法 k_T 算法、C/A 算法和 $\text{anti-}k_T$ 算法在 $\phi - \eta - p_T$ 平面上喷注截面。

3.4.3 喷注能量修正

由于喷注的能量与 p_T 和赝快度非线性相关，因此 CMS 有官方的喷注能量修正 (JEC) [45]。这些旨在通过修正喷注能量标度 (JES) 和喷注能量分辨率 (JER) 使测得的喷注能量更接近真实的喷注能量。能量修正分为三个步骤：

- L1：对真实数据及蒙特卡洛模拟样本的能量偏移修正，旨在消除 pileup 和电子噪声；
- L2L3：对粒子水平喷注响应的相对 (L2) 和绝对 (L3) 修正，对真实数据及蒙特卡洛模拟样本进行修正；
- Residual：仅对真实数据的修正，用于修正数据和模拟样本之间的残留差异。

一下分三小节简单总结这三个步骤，更多的详细技术细节可以参考 [45]。

L1 能量偏移修正

L1 pileup 偏移修正是此步最重要的修正，用于减去由于 pileup 而导致的喷注中的额外能量。该过程使用了有效喷注面积方法，该方法使用喷注的有效面积乘以平均事例能量密度来计算要从每个喷注中应该减去的偏移能量的大小。为了解决探测器不同部分、不同喷注能量不同时 pileup 密度不同的问题，增加了一个与 p_T 和 η 相关的附加项。对于探测器收集到的数据，计算出一个额外的比例来修正数据和模拟样本差异。这是通过构造一个以给定 $\eta - \phi$ 为中心的随机圆锥体，计算没有硬相互作用的数据集

(Zero Bias) 在该圆锥体的能量密度除以模拟样本中的真实能量偏移。对于 PUPPI 算法, 由于已经消除了由于 pileup 引起的额外能量, 因此不需要加这项修正。

L2 相对与 L3 绝对修正

在添加 L1 修正后, 基于 QCD 蒙特卡洛模拟样本中的真实探测器的响应来得到探测器对强子响应的修正值。粒子响应定义为重建与产生子水平横动量的比值

$$R = \frac{p_{T, \text{reco}}}{p_{T, \text{gen}}} \quad (3.6)$$

这些系数是对粒子级 p_T 和重建的 η 分区得到的, L2 相对修正旨在使探测器器响应均匀, 是 η 的函数, 而 L3 绝对修正是关于 p_T 的函数, 对真实数据及蒙特卡洛模拟样本进行修正。

剩余修正

在应用 L1, L2 和 L3 修正后, 仍需对数据进行附加修正, 以解决数据与模拟样本之间的残余差异。这是通过查看要校准的喷注与某个参考对象比如另一个喷注, Z 玻色子或光子之间的横向动量平衡来完成的。如果喷注能量标度不等于 1, 则表示 p_T 不平衡。L2 相对残余修正用双喷注数据样本来测量的, 通过比较 $\eta < 1.3$ 的基准喷注测得的 p_T 与 η 不受约束的待修正的喷注 p_T , 测量时根据喷注平均 p_T 分区, 修正是关于喷注 η 的函数。L3 相对残余误差修正用 Z 或光子加喷注事例来测量, 通过对比待测度的喷注 p_T 与 Z 玻色子或光子可以精确测得的 p_T 。相应差异被定义为

$$R_{\text{jet}, p_T} = \frac{p_{T, \text{jet}}}{p_{T, \text{base}}}$$

可以用数据和模拟样本得到两个值, 两者的比例被定义为残余修正。

3.5 喷注结构的重建

在寻找高能 (boosted) 矢量玻色子的分析中, 矢量玻色子夸克衰变产物之间距离会很小, 被重建为单个胖喷注而不是两个分离良好的小喷注。在强子衰变的情况下, 两个夸克之间的距离取决于矢量玻色子的质量及其 p_T , 可以表示为

$$\Delta R = \frac{2M}{p_T} \quad (3.7)$$

因此，如果的玻色子的 $p_T > 200 \text{ GeV}$ ，两个夸克合并为一个大小为 $R=0.8$ 的单个大锥形喷注。如果 W 的 p_T 远低于 200 GeV ，则其末态可被重建为两个明确分离的喷注。为了区分由强子衰变的矢量玻色子（ W 或 Z 玻色子）与由 QCD 产生的夸克或胶子的喷注，从理论上讲，喷注质量是一个很好的判别量， W 玻色子的质量约为 80 GeV ，而夸克或胶子质量接近零。然而，在非常高的横向动量下，QCD 喷注的宽度（以及质量）可能会变得同样大。此外，由潜在事件和 pileup 引起的辐射会在事例中引起大量其他粒子，从而增加喷注的总质量。因此，要能够准确有效地将 boosted QCD 喷注与矢量玻色子分开，就需要其他方法。为了消除 pileup，可以使用 PUPPI 之类的算法。如果想进一步提高质量分辨率，必须采用专用的修饰算法。

3.5.1 喷注质量修饰 (Grooming)

修饰技术旨在在不显著改变背景和信号事例数的情况下提高大半径喷注的质量分辨率。它包括通过重新聚类并确定喷注中可以去除的软粒子，从而使测量值更接近其真实质量。图3.7与3.8描述了 Trimming、pruning 和 soft drop mass 三种不同的修饰技术取不同参数值时对信号及本底喷注的表现，下文会分三小节来介绍。

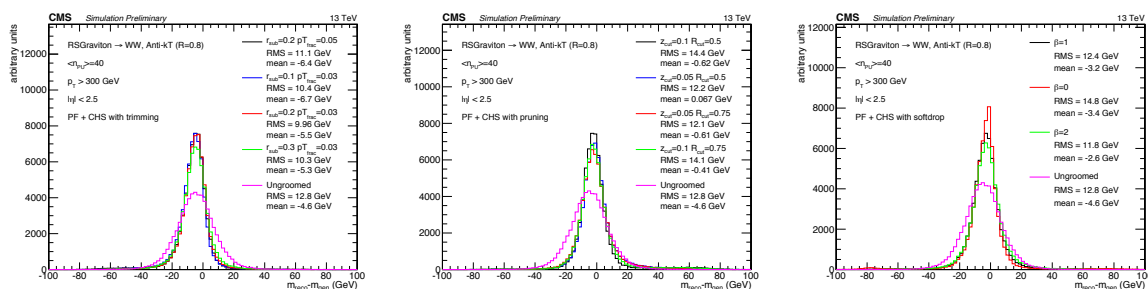


图 3.7 Trimming、pruning 和 soft drop mass 三种不同的修饰技术取不同参数值时对信号喷注的表现 [79]。

Trimming 算法

Trimming 算法 [63] 是一种在 CMS 的触发级别大量使用的修饰技术。从一个 anti- k_T 或 C/A 算法得到的喷注开始，使用 k_T 算法对喷注进行再次聚类，以创建子喷注。然后检查每个子喷注的动量相对整个喷注的占比是否都超过某个阈值。如果子喷注不满足阈值条件，那么会被移除。然后将其余的子喷注组成新的 Trimming 喷注。如图3.7左所示，在子喷注半径及阈值取不同值的情况下，Trimming 对实际 W 玻色子喷注和 QCD 夸克或胶子喷注的影响。当子喷注半径取 0.2，占比阈值取 0.03 时获得最佳信号质量分辨率，这也是 CMS 中 Trimming 算法中参数的默认值，从图3.8左可以看到看到经过 Trimming 后 QCD 喷注的质量趋于 0。

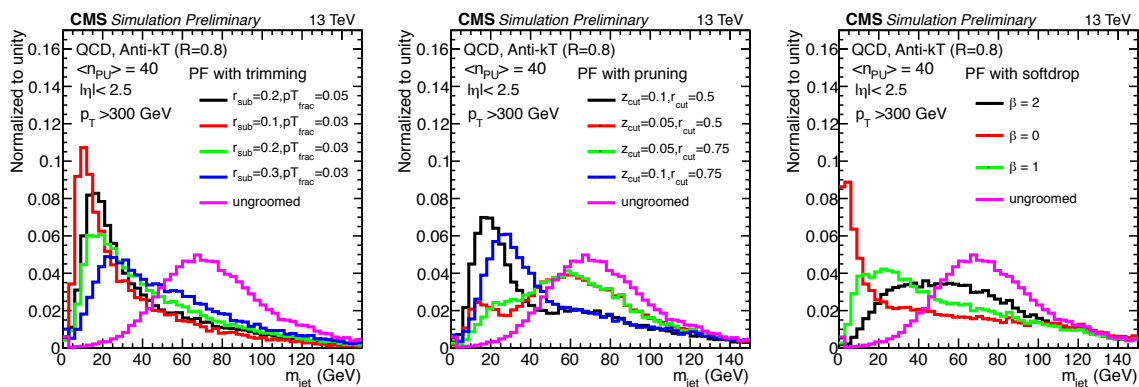


图 3.8 Trimming、pruning 和 soft drop mass 三种不同的修饰技术取不同参数值时对本底喷注的表现 [79]。

pruning 算法

除了除去软粒子外，pruning 算法还对组合之间的距离有额外的要求。对 C/A 算法喷注进行再次聚类，这在每个步骤中要求

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,com}} > z_{cut} \quad , \quad \Delta R_{i,j} < D_{cut} = \frac{2 \times r_{cut} m_{jet}}{p_T} \quad (3.8)$$

第一个条件是对组合是否为硬过程的要求。变量 p_{Ti} 和 p_{Tj} 对应于每个原喷注的横向动量和 $p_{T,com}$ 是两者组合的横动量。如果横向动量小的喷注不满足第一个条件，或者组合的角度大于 D_{cut} ，则将其移除。在 CMS 中，调整后的参数设置为 $r_{cut} = 0.5$ ， $z_{cut} = 0.1$ ，与图3.7和3.8中间图所示一致。

Soft Drop 算法

对于常见的 W 或 Z 喷注，是由两个夸克子喷注形成的，因此，每个子喷注的质量都远小于它们本身的质量。另外，QCD 喷注是由连续的软辐射形成的，这意味着 QCD 的最重子喷注应接近于喷注本身的质量。基于这些判断，质量下降标记器 (MDT) [24] 从使用 C/A 算法聚类的喷注 i 出发，然后再次对其再次进行聚类，定义子喷注 s_1 和 s_2 ，其中 $m_{s_1} > m_{s_2}$ 。如果在找子喷注时发生了明显的质量下降，说明分裂不是太不对称，如果满足 $\min(p_{T,s_1}^2, p_{T,s_2}^2) \Delta R(s_1, s_2) / m_i^2 > y_{cut}$ ， m_i 为喷注 i 的质量，则将喷注 i 打上标记。否则，将 i 设置为 s_1 ，然后继续循环。修改后的质量下降标记器 (mMDT) [49]，用 $m^2 + p_T^2$ 替代了质量 m 。这修复了 MDT 标记器中的一个缺陷，因此，在不满足质量下降和不对称条件的情况下，将选择能量更高的分支而非更重的分支。推荐的默认 mMDT 版本不再使用质量下降条件，因为它对 QCD 喷注质量分布影响可以忽略不计。另外，默认情况下，mMDT 算法使用 z_{cut} 标准进行修剪，而不是使用 y_{cut} 标准。修改

后的质量下降被概括为 soft drop[68]，该方法允许使用不同类型的角度要求来作为条件。Soft drop 要求如下：

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta, \quad (3.9)$$

现在不对称条件是直接通过子喷注的横动量占比来定义的，而不是通过到 k_T 定义的距离。如果分裂是相对对称的，则满足条件，并且将全喷注视为 soft drop 喷注。否则，仅保留最高横动量的子喷注，并且继续进行子结构分解。如果无法进一步分解喷注，可以将其从考虑中删除，即所谓的“标记”（tagging）模式，或者将其视为最终的 soft drop 喷注，将其视为“整理”（grooming）模式。 $\beta = 0$ 对应于 mMDT 算法，并删除分布比主要的两子夸克（2-prong）结构更宽的所有软发射。对于 $\beta > 0$ ，去除了软辐射，但是保留了部分软共线辐射。 $\beta < 0$ 的情况下，Soft Drop 可以消除软辐射和共线辐射。 $z_{cut} = 0.1$ ， $\beta=0$ 为 CMS 中的默认 soft drop 算法的参数设置，此时提供了最佳的信号/背景辨别力，同时保持了出色的信号质量分辨率，如图3.7与3.8右所示。

目前 CMS 用于 boost 喷注的标准是 puppi+soft drop，这是由于前者可以去除包括带电和中性的 pileup 粒子，后者可以去除喷注中来自其他软辐射的粒子，使用这两项技术可以有效使得喷注的质量分辨率更佳。

3.5.2 子喷注数量鉴别技术 (N-subjettiness)

使用上述算法后，喷射结构本身中仍然存在可以区分 W 或 Z 喷注和夸克/胶子喷注的信息。W 或 Z 喷注由两个定义明确的高横动量的子喷注组成。夸克或胶子喷注由单个部分子组成，因此利用子喷注数量判别可以进行有效区分。N-subjettiness 算法 [100] 充分利用了这一特点，对 n-subjettiness 进行量化，该变量定义为 τ_N

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}) \quad (3.10)$$

其中 k 循环所有的喷注子元素， $p_{T,k}$ 该组分的横向动量， $\Delta R_{i,k}$ 为该组分到候选子喷注轴的距离。前面的归一化因子为

$$d_0 = \sum_k p_{T,k} R_0 \quad (3.11)$$

其中 R_0 对应于初始喷注的圆锥尺寸。在此定义下， $\tau_N = 0$ 的喷注中大部分成分沿子喷注轴线方向。如果 $\tau_N \gg 0$ ，则大部分能量从子喷注方向外辐射出去，并且该喷注更有可能具有 N 个以上的子喷注。在 CMS 中，通常用 $\tau_{21} = \frac{\tau_2}{\tau_1}$ 用于区分 W 和 QCD 喷注。尽管预计信号射流的 τ_1 较大，但由于存在辐射，夸克或胶子也可能具有较大的 τ_1 。但

是，QCD 喷注往往也会有 τ_2 ，而信号喷注由于只有两个子喷注，因此 τ_2 较小，两者的比值提供了更大的区分能力。这可以从3.9中看到。

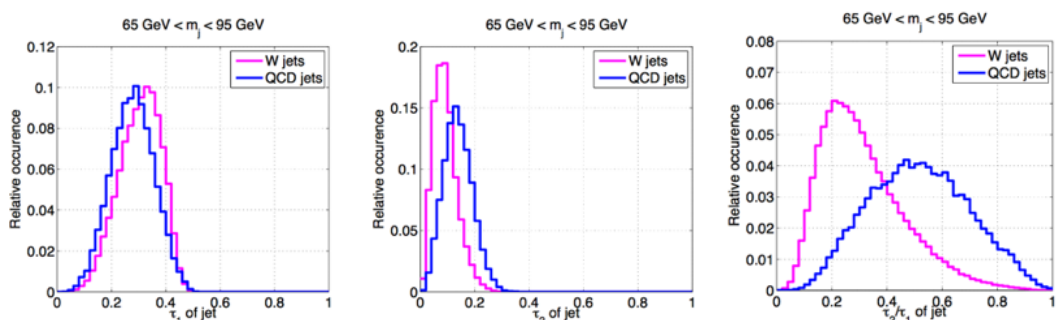


图 3.9 boosted W 和 QCD 喷注 τ_1 , τ_2 及两者的比值的分布，可以看到 $\frac{\tau_2}{\tau_1}$ 有很好的区分度。

3.5.3 DeepAK8 标记技术

原则上，利用粒子流算法重建喷注的完整信息和其他粒子级测量值的算法可以对喷注进行高效识别。为此，开发了一种新算法 DeepAKX[89]，可通过定制的机器学习方法直接利用此类粒子级信息进行机器学习。DeepAKX 算法是用于识别强子衰变粒子的分类器，具有五个主要类别：将喷注识别为源自顶夸克 (top)、W 玻色子、Z 玻色子、Higgs 和 QCD 背景。为了提高算法的通用性，主要类别进一步细分为次要类别，对应于每个粒子的衰减模式，例如 $Z \rightarrow b\bar{b}$, $Z \rightarrow c\bar{c}$ 和 $Z \rightarrow q\bar{q}$ 等，如图3.10右所示。目前，默认将 DeepAKX 算法与距离参数 $R=0.8$ 聚类的喷注一起使用，即 DeepAK8。

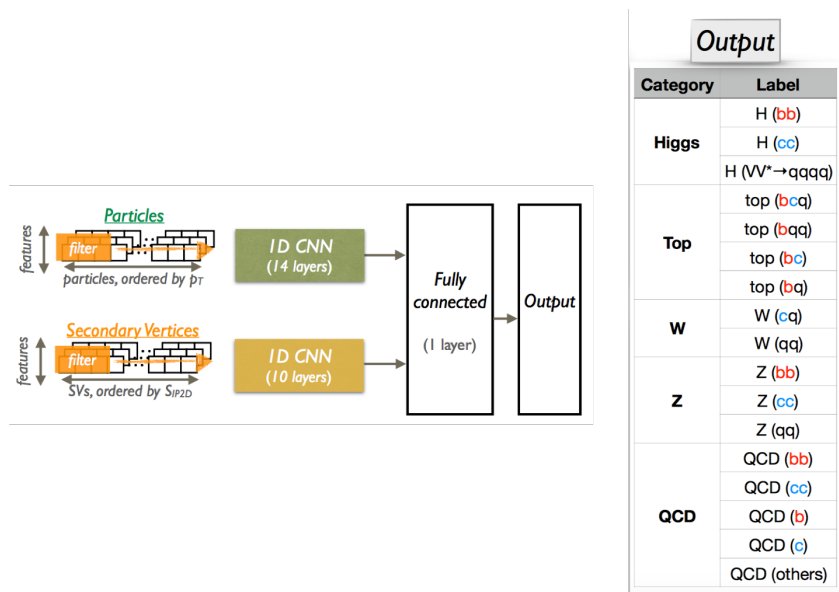


图 3.10 DeepAK8 算法的网络结构（左）与可识别的粒子衰变分类（右）。

在 DeepAK8 算法中, 对于每个喷注定义了 2 个列表作为输入 (图3.10左)。第一个列表是粒子列表, 包括超过 100 个组成喷注的粒子, 按横动量降序排序。粒子可观测的性质比如横动量、能量沉积、带电量、粒子与喷注轴或子喷注轴的夹角等, 都被用作输入来帮助算法训练得到能反应喷注结构的特征量。对于带电粒子, 额外的信息比如径迹的位置质量等也会作为输入。在粒子列表中总共有 42 个变量。第二个列表是次级顶点 (secondary vertex, SV) 相关量, 包括 7 个次级顶点, 每个顶点包含 15 个特征量, 比如次级顶点运动量、质量等。次级顶点列表可以提供用于推断喷注中夸克的味道。算法直接使用粒子级信息带来的重大挑战是输入数量的大幅增加。同时, 这些输入之间的相关性至关重要。为了有效地处理输入和之间的关联, DeepAK8 开发了定制的 DNN 架构。该体系结构包括两个步骤, 在第一步, 对粒子列表和 SV 列表采用两个一维 CNN 用于提取有用的功能。在第二步中, 将 CNN 得到的输出结合起来连入简单的全连接网络以执行喷注分类。第一步中的 CNN 结构基于 ResNet 模型, 粒子列表有 14 层, 而 SV 列表有 10 层。该算法使用了五千万事例, 80% 用于训练, 剩余 20% 用于验证。

尽管喷注质量没有被直接被用于输入, 但是 CNN 能够提取与质量相关的特征用于提高分辨能力。通常实验分析会使用喷注质量窗来区分信号和本底, 这会导致使用标记量做区分条件时会影响最终观测量。因此有喷注质量去相关的版本 (DeepAK8-MD) 被开发出来。如图3.11所示, 添加了预测质量的网络, 当进行训练时, 质量预测网络可以判断特征量与喷注质量的相关程度, 如果强烈相关, 那么会阻止 CNN 使用该特征量从而实现质量的去相关, 最终达到分辨能力和质量去相关的平衡。图3.12展示了 DeepAK8 普通版本、DeepAK8-MD 与 soft drop+ τ_{21} 的 ROC 曲线对比, 可以发现在相同信号效率时, DeepAK8 对 W 的鉴别有着更强的本底压低能力, 这对于提高敏感度有极大的作用。其中质量去相关版分辨效率会有下降, 但相对于其他方法仍有优势。

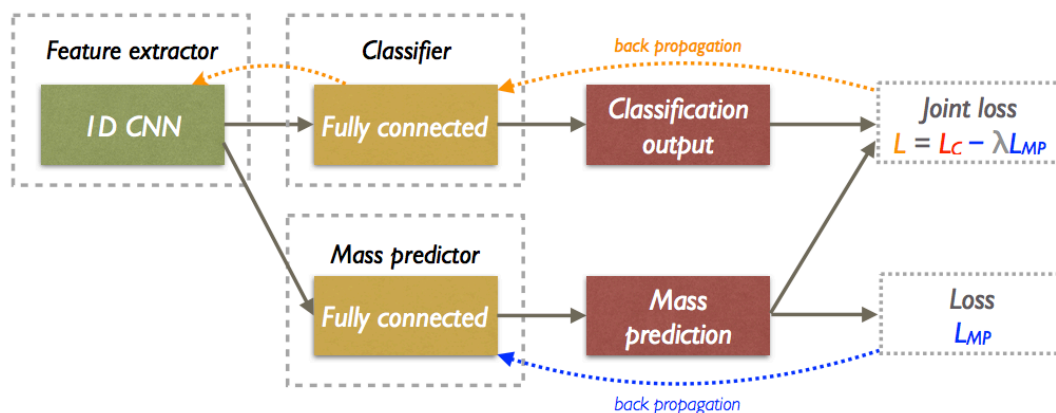


图 3.11 DeepAK8 算法喷注质量去相关算法的网络结构。

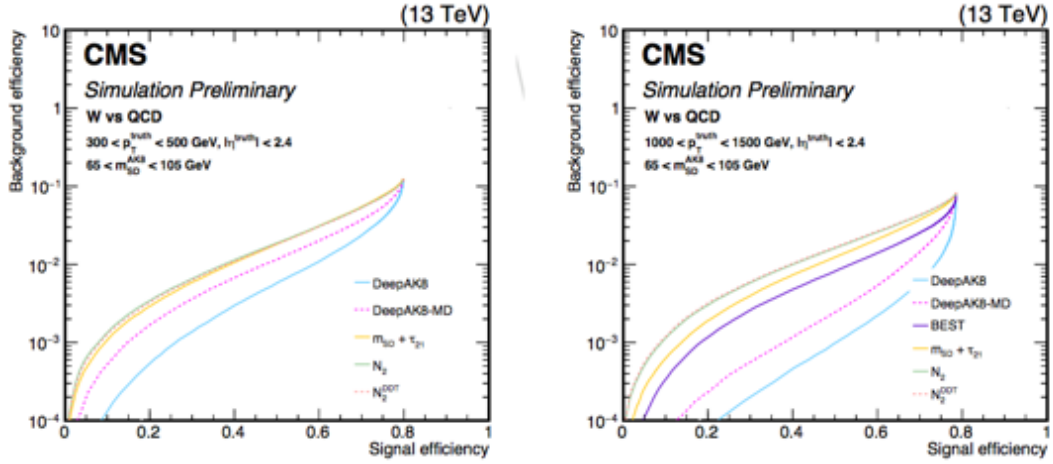


图 3.12 DeepAK8 算法与 N-subjettiness 的 ROC 曲线的对比。

3.5.4 官方修正因子

JME 官方提供了 DeepAK8 标记对 top 夸克和 W 玻色子的标记效率在数据中的表现 [89]，在计算时对单 μ 样本采用标记-探针方法 (tag and probe method) [60]， μ 与 b 夸克喷注组合形成标记 (tag)，而事例另一边的喷注作为探针 (probe)。总的标准模型本底样本按 top 衰变得到的部分子 (b、q、 $\bar{q}$) 与 AK8 喷注的空间距离被分为 3 类。其中，如果 3 个部分子与喷注距离都满足 $\Delta R < 0.6$ ，则被归为合并 top 类；如果只有 W 衰变得到的 2 个部分子与喷注距离都满足 $\Delta R < 0.6$ 且 b 夸克在喷注锥角外，被划分为合并 W 玻色子类；其他剩余的喷注都是非合并类 (Nonmerged)。蒙卡模拟样本得到的 3 类喷注的质量分布都作为模板来与数据中的喷注质量分布拟合。对于每一个给定的工作点，同时对通过和未通过工作点的事例得到的 3 个模板形状同时拟合，分别计算得到数据和蒙卡模拟样本的效率，二者的比值定义为比例因子 SFs。对应 W 玻色子标记的 SFs，按横动量 200-300、300-400、400-800 GeV 分为三段，图 3.13 为喷注横动量取 400-800 GeV 时喷注 soft drop 质量在通过和未通过工作点情形下的分布，图 3.14 为 W 标记的 SFs。

官方提供的 SFs 是在特定工作点下的修正，然而并不能直接用于本文的分析，有以下几点原因：

- 对应本文中不同的信号区，分别优化选取最佳的标记截断条件，而官方提供的工作点并非最佳优化点；
- 图 3.14 只提供了标记大于工作点截断值时的修正因子，本文使用全喷注不变质量作为信号区，因此无法使用常规的选喷注质量两侧做本底控制区，事实上分析采用标记小于信号区截断值作为控制区；
- 对于合并 radion 情形，采用 deep-W 与 deep-H 标记值组合作为新的鉴别量，而后

者并没有官方的 SFs。

因此本文开发了新的矩阵方法对全标记值范围的修正，具体方法在5.2章。

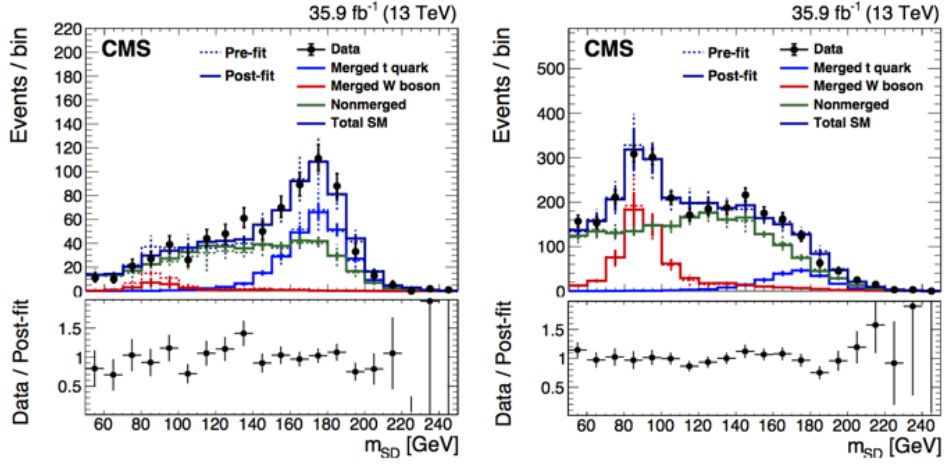


图 3.13 DeepAK8 去质量相关版本数据和模拟样本在通过和未通过工作点时喷注 soft drop 质量分布图，喷注横动量为 400-800GeV，虚线是预拟合时 3 类喷注模板的质量分布，实线是经极大似然拟合后的分布。

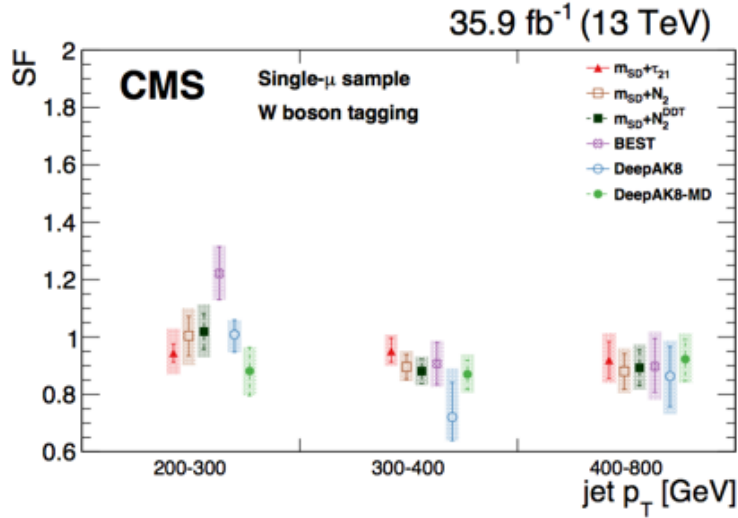


图 3.14 DeepAK8 算法得到的 W 标记 SFs。

第四章 末态为轻子和喷注的三玻色子共振态的寻找

4.1 半轻衰变三玻色子共振态的寻找

本分析是对一个大质量共振态经过 $W_{KK} \rightarrow W\phi$ 及 $\phi \rightarrow WW$ 衰变为变成三个 W 玻色子过程的首次寻找，其中 W_{KK} 是自旋为 1 的 Kaluza-Klein 规范玻色子， ϕ 是自旋为 0 的 radion，在本文中用 R 标记，该过程是由 1.2.5 中的拓展额外维模型预言的。分析要求三个 W 中其中一个轻子衰变、另外两个强子衰变。 W 轻子衰变产生孤立的轻子 ℓ 和中微子 ν ，而强子衰变情况更加复杂，取决于 R 和 W_{KK} 的相对质量，如果相对质量较大时，两个 W 会被重建为 2 个分立的 AK8 喷注，如果相对质量较小，根据公式 3.7，当 R 的横动量很高时， R 的衰变产物张角会很小，因此会得到 WW 合并的大型喷注。在后一种情况下，其中一个 W 也可以轻子衰变，但该轻子并非孤立的。因此在最终状态要求有一个孤立的轻子、丢失横能量 MET 和一个或两个喷注。图 4.1 描述了本分析的两种情形：分立和合并情形以及对应的最终状态。

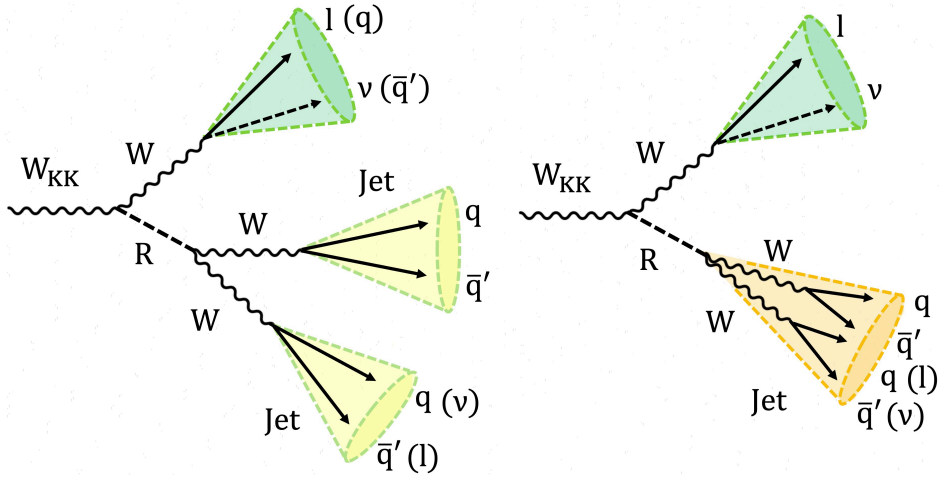


图 4.1 $W_{KK} \rightarrow W\phi$ 及 $\phi \rightarrow WW \rightarrow WWW$ 过程半轻衰变的分立（左）和合并（右）情形的费曼图。

这是一个全新的物理研究方向，之前从未直接研究过，该过程是由标准模型在 [4, 5, 8, 10] 中预测的，并且在大部分质量参数空间中都有很大的发现潜力。本分析基于 CMS 探测器在 2016-2018 年整个 RunII 阶段收集到的对应 137fb^{-1} 的数据。该分析要求轻子为 $\ell = e$ 或 μ ，另外也包含了 $W \rightarrow \tau\nu$ 其中 $\tau \rightarrow \ell\nu\nu$ 的情况，根据 W 衰变分支比的计算，一轻子道是 WWW 衰变的主要过程，因此作为本次分析的课题。该分析的难点在于信号如图 4.1 所示有 2 个共振态，对于第二种情况，即合并情形，喷注可以是含非孤立轻子的 W 喷注、含 4 个夸克的全合并 R 喷注或有部分夸克泄露的部分合并 R

喷注，导致最后的喷注质量是一个很宽的分布，需要寻找合适的标记方法来区分不同类型的喷注。对于轻子衰变的 W ，可以用 MET 来近似计算中微子的动量。利用 W 的质量 80.385GeV 、轻子四动量及 MET 信息，得到关于中微子 p_z 的一元二次方程，如果方程有 2 个实数解，选取最接近轻子 p_z 的解，如果方程无解，则选取实数根部分。最后选用轻子衰变的 W 加喷注的不变质量作为观测量。本文设计了专用的本底控制区域并采用数据驱动（data-driven）技术来估计本底。

4.2 信号与本底的蒙卡样本

所用蒙卡样本都是使用 CMS 官方产生的 MINIAOD 样本，其中 2016 年为 RunI-ISummer16MiniAODv3, 2017 年为 RunIIFall17MiniAODv2, 2018 年为 RunIIAutumn18MiniAOD。其中信号蒙卡样本是用 Madgraph V2.4.2[17] 产生子以领头阶（Leading Order, LO）精度产生，覆盖了 $M_{W_{KK}}$ 从 1.5 到 5TeV， M_R 从 0.6% 到 90% 倍 $M_{W_{KK}}$ 的 2 维质量参数空间，如图4.2所示。另外基于模型 [4, 5, 8, 10] 推荐，参数值选取如下，KK 引力耦合参数 $g_{\text{grav}} = 6$ ，规范耦合参数 $g_{W_{KK}} = 3$ 和 $g = 6.708$ 以及约束参数 $\epsilon = 0.5$ 。

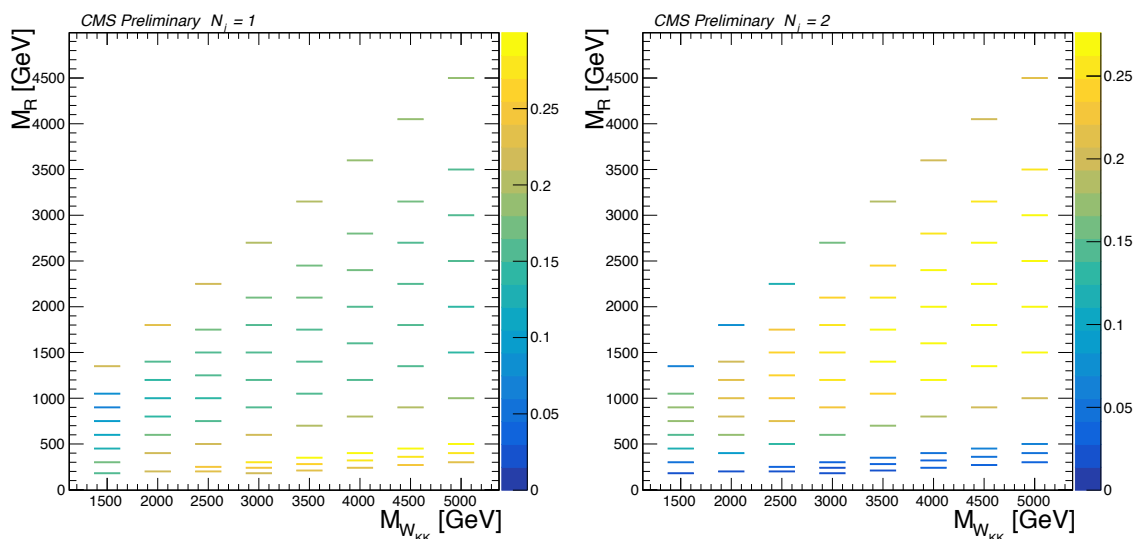


图 4.2 本文使用的信号样本在 2D 质量参数空间上的分布，一共 75 个点。颜色对应于只有 1 个（左）和 2 个（右）AK8 喷注时的信号效率（将会之后介绍）。

本底蒙卡样本主要包括顶夸克-反顶夸克对 ($t\bar{t}$) 和 W +jets 样本。 $t\bar{t}$ 是用 Powheg V2[16] 产生子在次领头阶 (NLO) 产生。 W +jets 样本使用了 Madgraph 产生子在 LO 精度进行模拟，其截面是归一化到更高阶 NNLO 进行修正。其余本底样本包括双玻色子样本、三玻色子、单顶夸克 (single top)、QCD 样本是用 Madgraph 或 Powheg 来模拟。

上述信号和本底样本只是由产生子在费曼图水平（所谓 matrix element level, ME）产生事例，之后还需要经过 PYTHIA 来进行部分子簇射过程（Parton Shower, PS）。此

外, PYTHIA 还使用了 tune 来模拟 UE 的影响, 其中 2016 年使用 CUETP8M1 tune, 2017 和 2018 年使用 CP5 tune[59、91], 部分子函数 (PDFs) 分别使用 NNPDF 3.0、3.1[71、19]。所有粒子在 CMS 探测器中的响应都是用 GEANT4 模拟 [11]。在产生样本时 PU 的影响也考虑在内, 对于每一个蒙卡样本中的事例, 都会使用重建的初级顶点数量信息来得到所谓的 PU 权重来与数据匹配。

4.3 事例的挑选

事例的挑选是基于第三章介绍的 PF 粒子流重建算法, 基于分析所需要的信号特征来对重建物理量加一些选择条件来加以挑选。

4.3.1 噪声过滤与顶点选取

考虑到强子量能器中信号异常、电磁量能器异常的能量沉积等因素的影响, 会导致 MET 的错误重建, 因此需要丢弃掉这部分事例。基于 MET 组的官方推荐 [39], 添加了如下的噪声过滤:

对应 2016 年样本:

- 对好的顶点过滤 (goodVertices)
- 对束晕过滤 (globalSuperTightHalo2016Filter)
- 对强子量能器端盖与筒部过滤 (HBHENoiseFilter)
- 对强子量能器孤立度过滤 (HBHENoiseIsoFilter)
- 对电磁量能器坏死单元触发过滤 (EcalDeadCellTriggerPrimitiveFilter)
- 对坏的 PF 流 μ 子过滤 (BadPFMuonFilter)
- 对坏的电磁量能器单元过滤, 仅对数据使用 (eeBadSCFilter)

对应 2017 年及 2018 年样本除了要求以上 6 个过滤条件外, 还需要通过电磁量能器减少失败校准的过滤 (ecalBadCalibReduced), 此项需使用官方推荐的方法对每个事例重新计算。另外, 要求事例中至少包含一个好的初级顶点和至少有 4 个自由度, 且距离质子对撞点中心满足 $x - y$ 平面上小于 2cm, z 方向小于 24cm。

为了修正 2016 和 2017 年电磁量能器 L1 预打火 (L1 prefiring) 的影响, 使用了官方推荐的权重来调整事例 [37]。

4.3.2 触发

本文使用的数据主要是单电子和单 μ 子数据集, 分别使用 $p_T > 27/32/35$ 或 115 GeV、 $p_T > 50/100$ 的高级触发 (High Level Trigger, HLT)。另外由于这两种触发对大

横动量电子或 μ 子效率较低, 因此, 对于电子道, 额外使用了光子触发; 对应 μ 子道, 额外使用 MET 触发来弥补降低的效率。本文使用的全部触发总结如表4.1。

表 4.1 本分析对数据使用的 HLT 详表。

年份	分析道	数据集	HLT 触发	具体描述
2016	电子	单电子	HLT_Ele27_WPTight_Gsf_v* 或 HLT_Ele115_CaloIdVT_GsfTrkIdT_v* HLT_Photon175_v*	$p_T > 27$ GeV, 电子 ID 紧工作点, $p_T > 115$ GeV, $p_T > 175$ GeV,
	μ 子	单 μ 子 MET	HLT_Mu50_v* OR HLT_TkMu50_v* HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_v*	$p_T > 50$ GeV MET > 120 GeV
2017	电子	Single 电子	HLT_Ele35_WPTight_Gsf_v* 或 HLT_Ele115_CaloIdVT_GsfTrkIdT_v* HLT_Photon200_v*	$p_T > 35$ GeV, 电子 ID 紧工作点, $p_T > 115$ GeV, $p_T > 200$ GeV,
	μ 子	单 μ 子 MET	HLT_Mu50_v* OR HLT_OldMu100_v* OR HLT_TkMu100_v* HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_v*	$p_T > 50$ GeV or $p_T > 100$ GeV MET > 120 GeV
2018	电子	电子光子	HLT_Ele32_WPTight_Gsf_v* 或 HLT_Ele115_CaloIdVT_GsfTrkIdT_v* 或 HLT_Photon200_v*	$p_T > 32$ GeV, 电子 ID 紧工作点, $p_T > 115$ GeV, $p_T > 200$ GeV,
	μ 子	单 μ 子 MET	HLT_Mu50_v* OR HLT_OldMu100_v* OR HLT_TkMu100_v* HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_v*	$p_T > 50$ GeV or $p_T > 100$ GeV MET > 120 GeV

由于 HLT 对于数据和蒙卡样本有不同的效应, 因此需要考虑比例修正因子 (scale factor) 来修正这种差异, 对于电子道触发效率, 使用官方推荐的比例修正因子如图4.3, 4.4, 4.5所示, 分别对应 2016、2017、2018 年的影响因子。

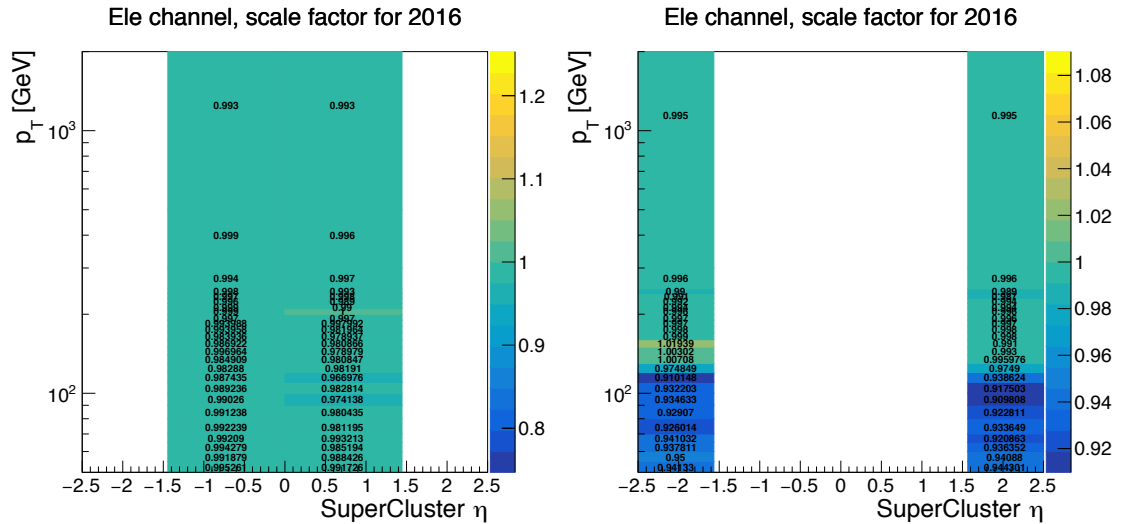


图 4.3 2016 年单电子触发在筒部区域 (左) 和端盖 (右) 的数据与蒙卡样本的比例修正因子。

μ 子触发效率用富集 boost W+jets 数据测量的, 图4.6, 4.7, 4.8是计算得到的三年的 μ 子和 MET 触发效率。我们用 μ 子或 MET 触发, 因此总的触发效率为: $\epsilon_{total} = \epsilon(\ell) + \epsilon(\text{MET}) - \epsilon(\ell) * \epsilon(\text{MET})$ 。新的比例修正因子定义为: $SF(total) = \frac{\epsilon_{total}(data)}{\epsilon_{total}(MC)}$ 。

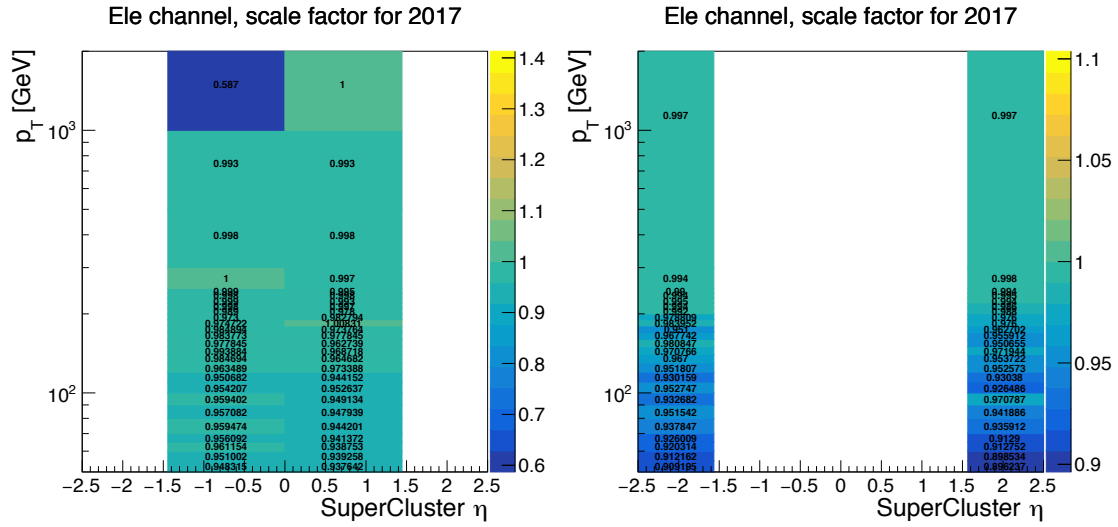


图 4.4 2017 年单电子触发在筒部区域（左）和端盖（右）的数据与蒙卡样本的比例修正因子。

4.3.3 μ 子挑选

分析在选用的 μ 子要求通过高动量 μ 子鉴别 (high-pt muon ID) [34]:

- μ 子被重建为全局 μ 子;
- 至少有一个 μ 子探测器中的打火点在全局 μ 子拟合径迹中;
- μ 子径迹至少在 2 个探测器子单元留下片段;
- 最好的 μ 子横动量的相对偏差 $\sigma(p_T)/p_T < 0.3$;
- 以防混入来自宇宙线和其他粒子衰变得到的 μ 子, 要求相对于初级顶点的横向影响参数小于 2mm, 径迹的纵向距离小于 5mm;
- μ 子径迹至少在内层硅像素探测器中留有一个打火点;
- 重建中至少有 6 层径迹探测器刘希打火点。

除了 high-pt ID 的要求之外, 所选 μ 子应满足 $p_T > 55 \text{ GeV}$ 且 $|\eta| < 2.4$ 。为了进一步减少喷注被误判为 μ 子的情况, 额外要求基于粒子流算法的孤立条件 (ISO), 要求 $\Delta\beta < 0.05$, 相应的数据与蒙卡样本差异的比例修正因子如图 4.9 所示, 第 6.1 章会讨论由 ID、ISO 条件带来的系统误差。

4.3.4 电子挑选

粒子流电子候选者是基于与径迹相连的电磁量能器簇射重建得到。重建的电子要求通过 HEEP v7.0[32] 鉴别条件, 且 $p_T > 55 \text{ GeV}$, $|\eta| < 2.5$, 但是要排除电磁量能器筒部与端盖的交界区 [1.4442, 1.566]。HEEP v7.0 鉴别要求重建电子包含与孤立的电磁量能器能量沉积相连的高质量径迹, 并且电磁量能器的能量沉积必须符合真正的簇射

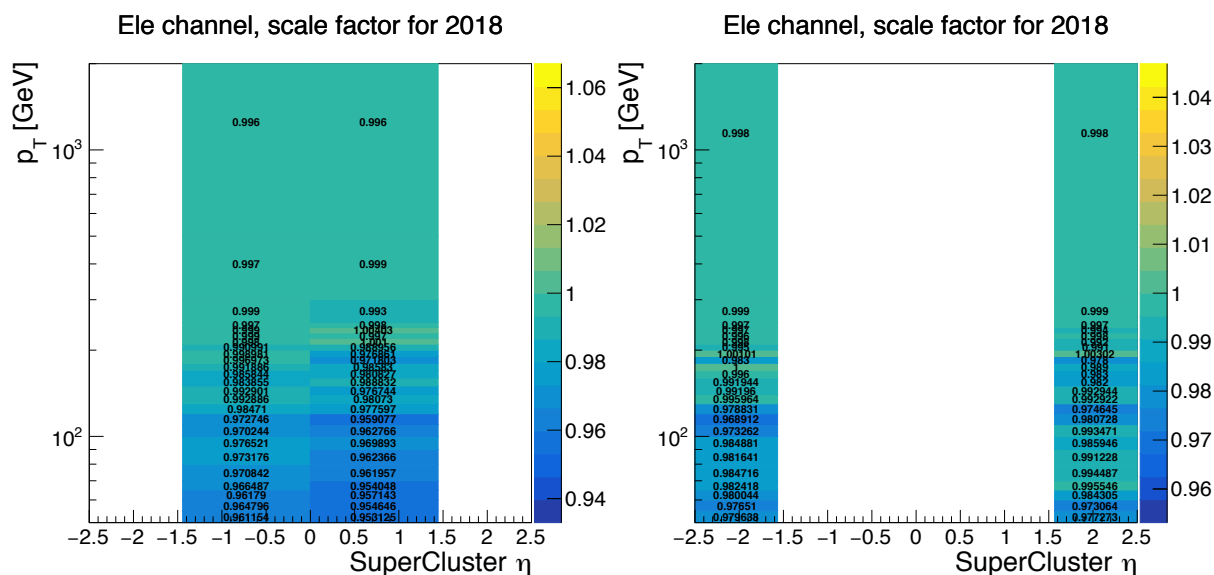


图 4.5 2018 年单电子触发在筒部区域（左）和端盖（右）的数据与蒙卡样本的比例修正因子。

形状。同样，ID 带来的差异及重建的比例修正因子都被添加到分析中，均使用了官方推荐值 [30]。

4.3.5 喷注挑选

喷注重建是采用章节3.4描述的粒子流重建方法，本文使用了 2 个喷注集合：

- AK8 PUPPI 喷注: anti- k_T 算法重建的距离参数半径为 0.8 的喷注集合，用于挑选 boost W 玻色子或者 R 衰变的喷注；
- AK4 CHS 喷注: anti- k_T 算法重建的距离参数半径为 0.4 的喷注集合，用于要求或者排除标记为底夸克 (b-tagged) 喷注。

对应两种集合，都采用了紧的鉴别条件 [36]，所有的喷注都采用了 L1、L2 和 L3 修正来校准喷注的能量，对于数据需要额外考虑 L2+L3 剩余修正 [42]。另外还需要考虑数据与蒙卡喷注能量分辨率 (JER) 的差异，对于蒙卡样本重建的喷注，按照横动量分辨率来弥散 (smear) 从而与数据匹配 [43]，相应的误差是从 global tag 读取并计算，将会在6.1章详细讨论。对于 AK8 喷注，采用 PUPPI+soft drop 技术来降低 pileup 的影响，具体细节可参考第三章内容。

对强子衰变的 W 或者 R 候选者，需挑选 AK8 喷注横动量大于 200GeV 且 $|\eta| < 2.4$ ，另外为了防止轻子与喷注的重叠，要求在喷注附近 $\Delta R < 1.0$ 范围内不存在好的电子或 μ 子，否则不将该喷注计数。对于事例中其他的喷注，挑选粒子流重建的横动量大于 30GeV 的 CHS AK4 喷注，同样要求喷注附近 ($\Delta R < 0.4$) 与挑选的电子或 μ 子不

重叠。AK4 喷注主要用于排除 b-tagged 喷注从而减少 $t\bar{t}$ 和 single top 本底。b-tagged 喷注要求其满足 $|\eta| < 2.4$ 并且通过 DeepCSV 标记推荐的中等工作点 [40]。

b-tagged 标记修正因子

由于 b-tagged 效率在数据中的测量值与蒙卡模拟预测不一致，因此需要考虑比例修正因子来修正蒙卡模拟样本。利用官方推荐的修正因子及蒙卡底夸克标记效率来缩放每个事例的权重 [31]，其中对于底夸克、粲夸克和其他轻夸克型喷注，采用 DeepCSV 底夸克标记技术，标记效率如图4.10、4.11和4.12所示，分别为 2016、2017 和 2018 年在 p_T 和 η 分区内的分布。

4.4 事例选择条件

首先根据信号的基本特征给出很松的条件作为初选条件，然后基于初选条件来优化确定最终的选择条件。

4.4.1 初选条件

基于信号费曼图4.1的基本特征可以确定一下动力学挑选条件来区分信号与本底。初选条件要求如下：

- $N_\ell=1$ ，即有且仅有一个轻子 ℓ ，为电子 (e) 或 μ ， $p_T > 55\text{GeV}$ ， $|\eta_{e(\mu)}| < 2.5(2.4)$ 且相对孤立程度满足 $Iso < 0.05$ ；
- 没有第二个 $e(\mu)$ 满足 $p_T > 35(20)\text{GeV}$ ；
- 对于 $e(\mu)$ 道满足 $E_T^{\text{miss}} > 80(40)\text{GeV}$ ；
- 轻子衰变的 W 横动量满足 $P_T(W_{\ell\nu}) > 200\text{GeV}$ ；
- AK8 喷注数量 $N_j=1, 2$ 或 3（下表 j 表示喷注），且 $|\eta_j| < 2.4$ ， $p_T(j) > 200\text{GeV}$ ；
- 最大的喷注质量满足 $M_j^{\text{max}} > 40\text{GeV}$ ，式中上标“max, mid, min”表示 AK8 喷注按质量排序（如果事例只有一个喷注，则 $M_j^{\text{max}} = M_j > 40\text{GeV}$ ）；
- 当 $N_j=1$ （2、3）时三玻色子不变质量 $M_{j\ell\nu}(M_{jj\ell\nu}) > 500\text{GeV}$ ，是一个或两个喷注加轻子衰变的 $W_{\ell\nu}$ 系统的不变质量。

在初选条件下，一些运动学变量的分布如图4.13、4.14和4.15。它们被用来研究哪些变量可以显著区分信号与本底。每张图的下部小图展示了定义为“ $S/\sqrt{1.2B+1}$ ”的信号显著度，其中 S 和 B 分别表示信号和本底的事例数。1.2 用来表示本底 20% 的误差。图中所有的蒙卡事例被归一到 RunII 数据对应的积分亮度 137fb^{-1} 。

基于这些图，可以得出结论：合并情形主要对应于 $N_j = 1$ 的情况；分立情形对应 $N_j = 2$ 。 S_T ，是轻子衰变的 W 的横动量与所有喷注横动量标量和，它与 W_{KK} 大小是相

关的，因此对于信号有很好的区分度，可以作为后续信号区的选择条件。 W 和 R 共振态的质量中 R 喷注质量更大，为了更好地区分不同喷注类型，因此选用质量排序而非按横动量排序，喷注质量对应合并情形如图4.14所示。

4.4.2 玻色子标记

信号中的喷注对应 boost radion($R \rightarrow WW \rightarrow qq\bar{q}\bar{q}$, $qq\ell\nu$) 或者 2 个 boost $W(W \rightarrow qq)$ ，分析选取的 AK8 喷注距离参数 $R=0.8$ ，有时第一种情况下的 4 个夸克可能有部分在重建的喷注尺寸外，因此重建的喷注可能包含 4 个、3 个或 2 个子喷注。CMS 实验有多种技术来标记这种差异，本分析采用其中高效的 DeepAK8 质量去相关版 (DeepAK8-MD) 的标记技术，基于 CNN 神经网络训练，有着很强的信号区分能力，参考3.5.3章。这种技术使用提供了多种模块来区分多种标准模型对象，比如 W 、顶夸克、Higgs 等，算法基于事例的喷注及顶点信息给出某对象的原始分辨值 (raw score) 来判断是该对象的可能性，使用时通常需要归一化处理。在本文中使用如下的标记：

- $\text{deep-}W \equiv [\text{raw score } W] / [\text{raw score } W + \text{raw score QCD}]$ ，用来鉴别 W ，用于质量在 60-100GeV 的喷注鉴别；
- $\text{deep-WH} \equiv [\text{raw score } W + \text{raw score H4q}] / [\text{raw score } W + \text{raw score H4q} + \text{raw score QCD}]$ ，用于鉴别合并的 radion 标记，对质量大于 100GeV 的喷注使用；
- $\text{deep-t} \equiv [\text{raw score } t] / [\text{raw score } t + \text{raw score QCD}]$ ，用于排除或提高顶夸克喷注；
- $\text{deep-t}/W \equiv [\text{raw score } t] / [\text{raw score } t + \text{raw score } W]$ ，在特殊的控制区域区分 W 和顶夸克；
- $\text{deep-t}/Wg \equiv [\text{raw score } t] / [\text{raw score } t + \text{raw score } W + \text{raw score QCD}]$ ，在特殊的控制区域区分顶夸克和其他类型的喷注。

图4.16展示了在初选条件下信号和本底的表现差异，可以发现对于信号喷注， $\text{deep-}W$ 和 deep-WH 更集中在值较大的右侧。对于 $60 < M_j < 100 \text{ GeV}$ 、 $M_j > 100 \text{ GeV}$ 的喷注分别要求 $\text{deep-}W$ 、 $\text{deep-WH} > 0.7$ ，如果有 2 个喷注都满足 $60 < M_j < 100 \text{ GeV}$ ，则将条件放松到 > 0.5 。如果 $M_j < 60 \text{ GeV}$ ，那么喷注可能是单夸克或者胶子产生的，此时不需要加任何的标记挑选条件。

4.4.3 信号区选择条件

最终的信号区域 (signal region, SR) 选择条件是在初选条件的基础上 (参考4.4.1)，新添加的条件如下：

- $N_j=1,2$;
- b-tagged 喷注数量 $N_b^M=0$ ，与 AK8 喷注不重叠的 AK4 喷注数目满足 $N_{j\text{-AK4-exc}} \leq 2$;

- $N_j=1$ 时 $M_j^{\max} > 60\text{GeV}$, $N_j=2$ 时 $60 < M_j^{\max} < 100\text{GeV}$;
- 如果 $60 < M_j^{\max} < 100\text{GeV}$, 要求 $\text{deep-W} > 0.7$, 如果 $M_j^{\max} > 100\text{GeV}$, 要求 $\text{deep-WH} > 0.7$, 当 $N_j=2$ 且满足 $M_j^{\min} > 60\text{GeV}$ 时将条件放松至 $\text{deep-W} > 0.5$;
- $S_T > 1$;
- $M_{j\ell\nu}, M_{jj\ell\nu} > 1.1\text{ TeV}$ 分别对应于 $N_j=1, 2$ 。

其中 deep-W 标记用来在 W 质量窗 $60\text{-}100\text{GeV}$ 内区分 W 型喷注与 QCD 喷注, deep-WH 标记用于在 $M_j > 100\text{GeV}$ 时区分多子喷注结构的合并 radion 与 QCD 喷注。

4.4.4 信号区喷注质量分类

由于分析涉及两个共振态, W_{KK} 与 radion, 所以末态会有很多种情形。比如轻子衰变的 W 是直接来自 W_{KK} 而非 radion 衰变, radion 的末态可能被重建为一个胖喷注, 或一对 $W(qq)$ 喷注, 甚至更多, 来自 W 的 2 个夸克可能相距较远因此被重建为 2 个单独的夸克喷注。对于 boost radion, 末态也不一定会全部在 AK8 喷注张角内, 有可能有部分夸克没有被包含或者其中一个 W 轻子衰变含有非孤立的轻子, 这些情形下, 喷注质量都会低于 M_R 。综上所述, 喷注质量分布非常复杂, 在不同的质量区间需要使用不同的标记, 这些可以从图4.14和4.15的 M_j 与 $M_j^{\min}$ 分布中观察到。更细致的研究将会在图5.16和5.17中提到。

为了保留这些不同的末态, 同时预测合并和分立 radion 情形, 决定对喷注质量进行分区, 信号区域被分为 6 个正交的部分:

- $N_j=1$ 的事例按喷注质量分为 3 个信号区: $M_j : 60 - 100, 100\text{-}200, > 200\text{GeV}$ 分别对应 SR1、SR2 和 SR3, 是对合并情形 radion 质量的分区;
- $N_j=2$ 的事例先按小质量喷注分为两部分 $M_j^{\min} : 60 - 100, 0\text{-}60\text{GeV}$, 其中第一个条件下含有 2 个大质量的喷注, 要求全部满足 $\text{deep-W} > 0.5$ 定义为 SR4, 如果只有一个满足 $\text{deep-W} > 0.5$ 定义为 SR6; 后一个条件满足 $M_j^{\min} < 60\text{ GeV}$ 时只有一个大质量喷注, 被定义为 SR6。

6 个信号区如图4.17所示, 表4.2总结了它们之间的差异。

4.5 6 个信号区观测量

SR2 和 SR3 中只有 1 个质量大于 100GeV 的喷注, 因此只用来研究合并 radion 情形, 包含 $R \rightarrow WW \rightarrow qq\bar{q}\bar{q}, qq\ell\nu, qq\tau_h\nu$ 及存在夸克泄露的情况。

SR1 也是只有一个喷注, 质量在 W 质量窗口内, 对于合并 radion 情形, 喷注质量有可能在 $60\text{-}100\text{GeV}$, 对于孤立情形, 非 boost W 强子衰变 W 可能会得到 2 个分立的

表 4.2 6 个信号区不同的挑选条件

信号区	$M_j^{\max}[\text{GeV}]$	deepAK8 标记	$M_j^{\min}[\text{GeV}]$	tagger	N_j^{AK8}	N_j^{AK4}	N_b^M
SR1	60-100	deep-W > 0.7	-	-	1	≤ 2	0
SR2	100-200	deep-WH > 0.7	-	-	1	≤ 2	0
SR3	≥ 200	deep-WH > 0.7	-	-	1	≤ 2	0
SR4	60-100	deep-W > 0.5	60-100	deep-W > 0.5	2	≤ 2	0
SR5	60-100	deep-W > (<) 0.5	60-100	deep-W(<) > 0.5	2	≤ 2	0
SR6	60-100	deep-W > 0.7	0-60	-	2	≤ 2	0

夸克喷注，由于 $p_T < 200 \text{ GeV}$ ，因此不会计数为 AK8 喷注，同样在 SR1 的接收范围内，这种情形对于 W_{KK} 和 M_R 接近时比较常见，因此 SR1 对合并和孤立情形都敏感。

SR4 中要求 2 个满足 W 标记的喷注，因此它是显著度最高区域；SR5 与 SR4 类似，但是只有 1 个 W 标记的喷注，显著度相对 SR4 要低；最后 SR6 包含 1 个 W 标记的喷注和另一个分立的 W，同样是由于研究分立 radion 情形。

最终分析的观测量是三玻色子不变质量，对于 $N_j=1$ (SR1、SR2、SR3) 使用 $M_{j\ell\nu}$ ；而对于 $N_j=2$ (SR4、SR5、SR6) 使用 $M_{jj\ell\nu}$ 。六个观测量如图 4.18 所示，分析正是基于这六个观测量的形状。

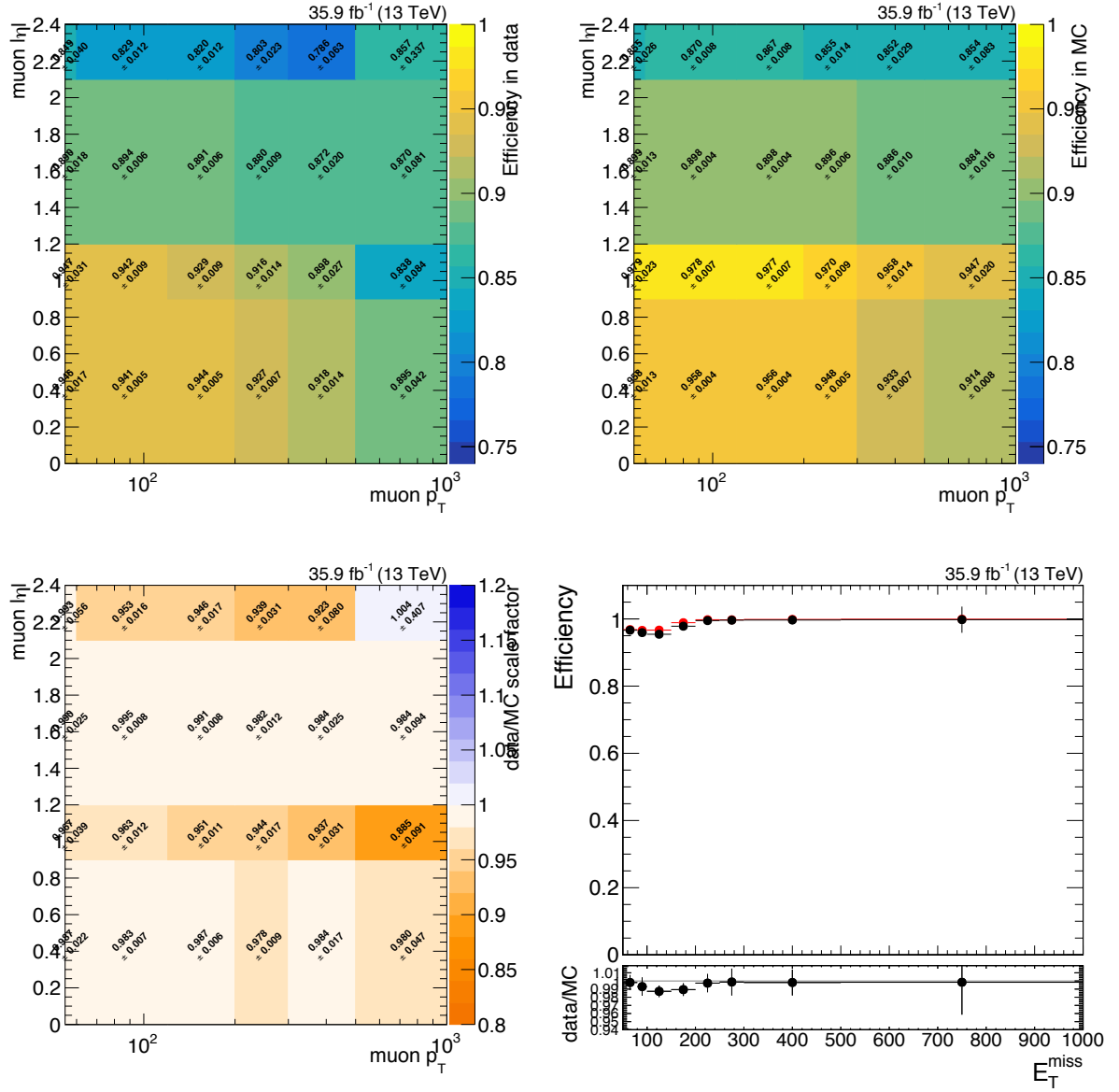


图 4.6 2016 年单 μ 子在数据（上左）和蒙卡样本（上右）的触发效率及两者的比例作为比例修正因子（下左），下右图是 MET 触发效率和数据与蒙卡样本的比例修正因子。

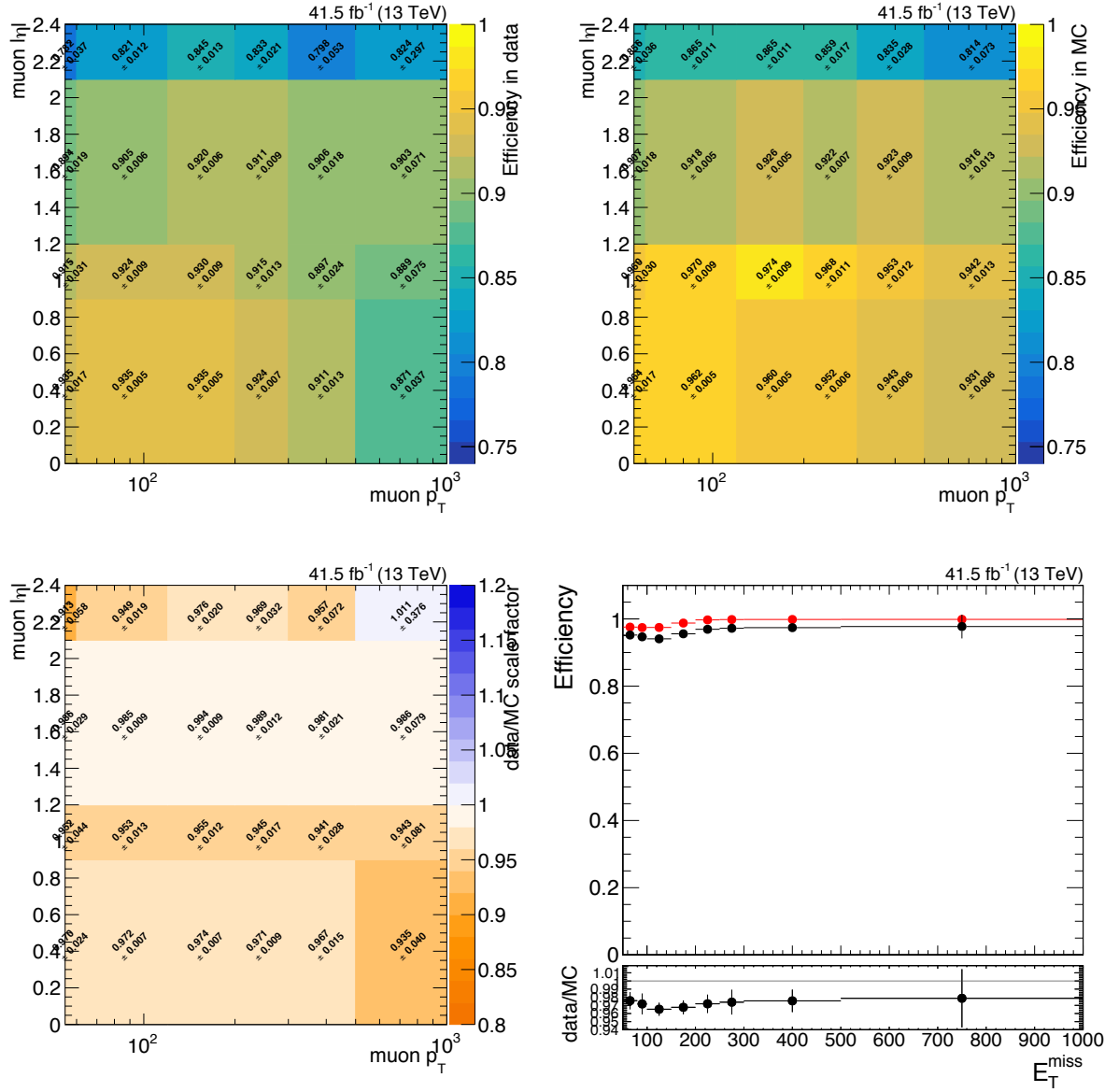


图 4.7 2017 年单 μ 子在数据 (上左) 和蒙卡样本 (上右) 的触发效率及两者的比例作为比例修正因子 (下左), 下右图是 MET 触发效率和数据与蒙卡样本的比例修正因子。

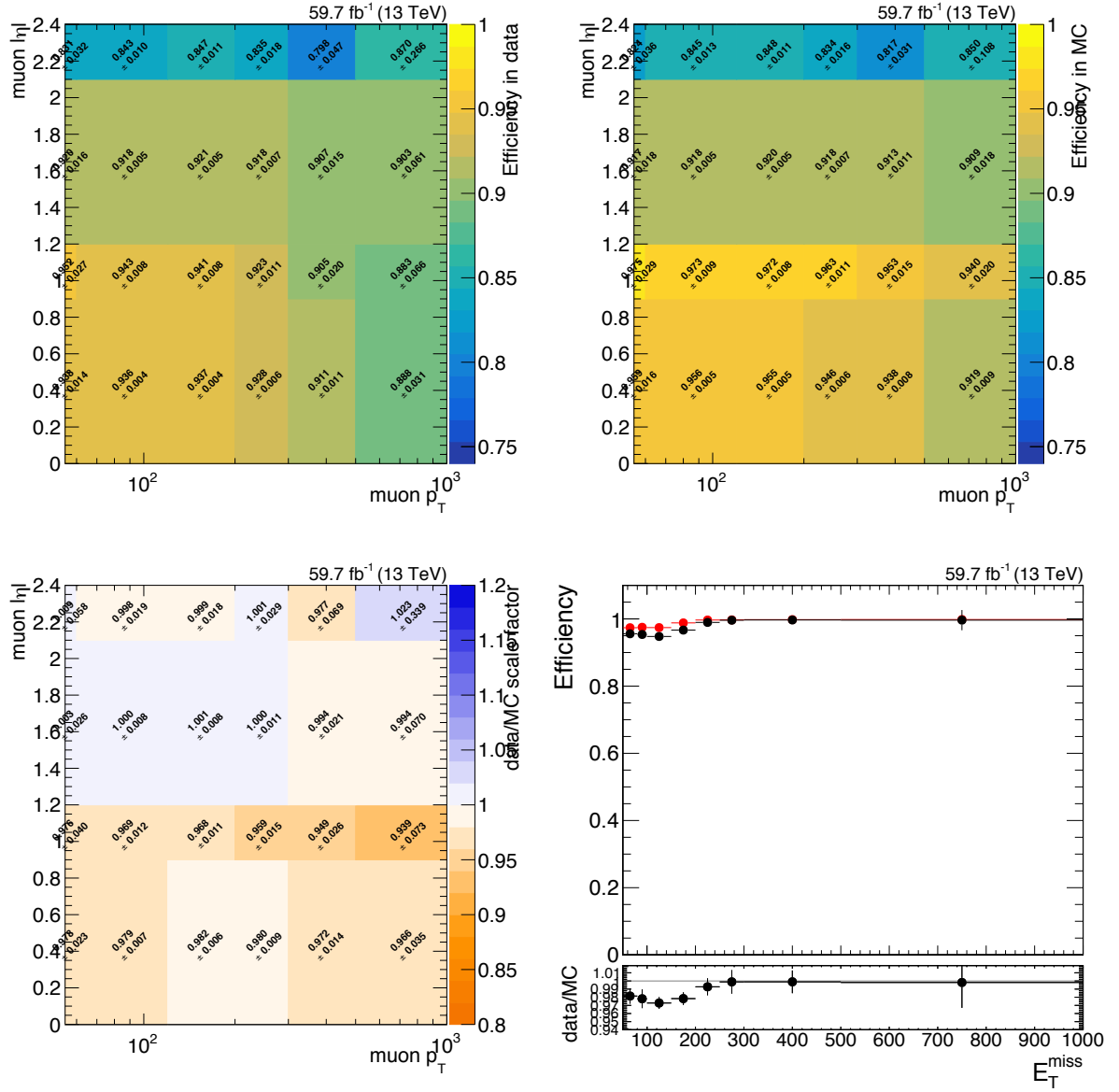


图 4.8 2018 年单 μ 子在数据（上左）和蒙卡样本（上右）的触发效率及两者的比例作为比例修正因子（下左），下右图是 MET 触发效率和数据与蒙卡样本的比例修正因子。

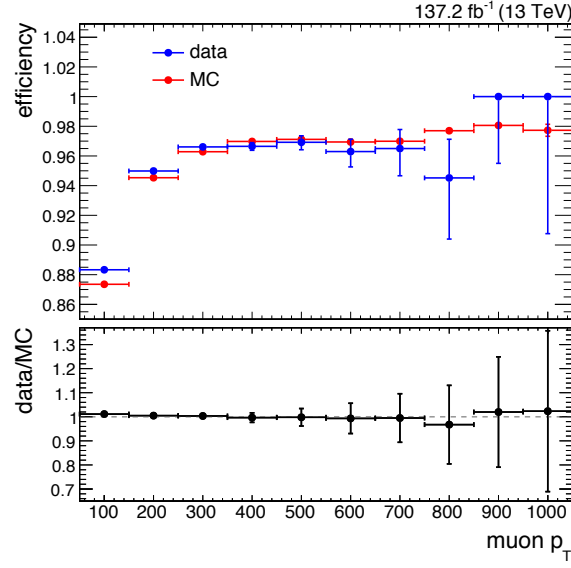


图 4.9 使用 boost Z 和 DY 事例计算得到的 μ 子孤立条件在数据和蒙卡中的效率以及二者的比例修正因子。

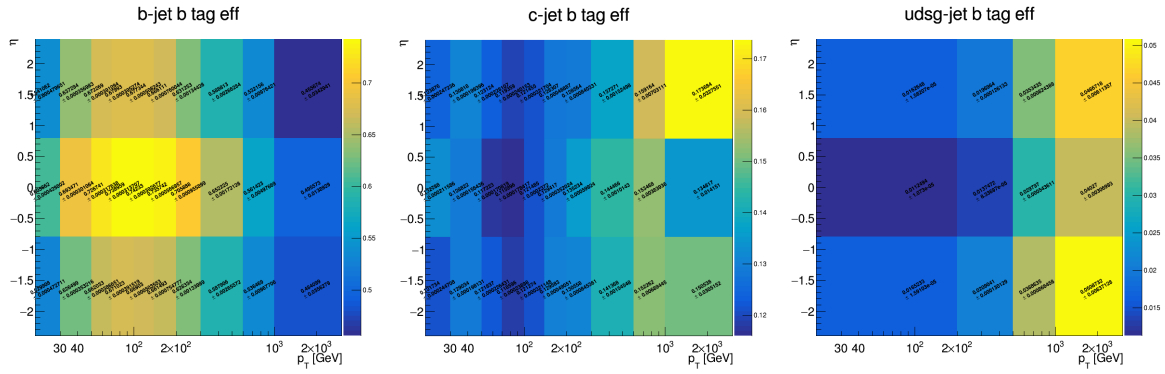


图 4.10 采用 2016 年 POWHEG $t\bar{t}$ 样本计算的对于底夸克、粲夸克、轻夸克型喷注的 DeepCSV 底夸克标记效率，标记时使用了中间工作点。

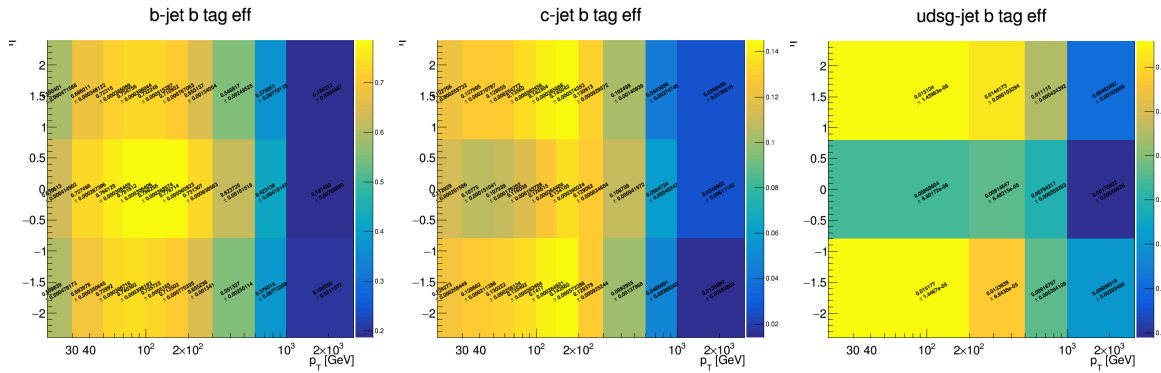


图 4.11 采用 2017 年 POWHEG $t\bar{t}$ 半轻衰变样本计算的对于底夸克、粲夸克、轻夸克型喷注的 DeepCSV 底夸克标记效率，标记时使用了中间工作点。

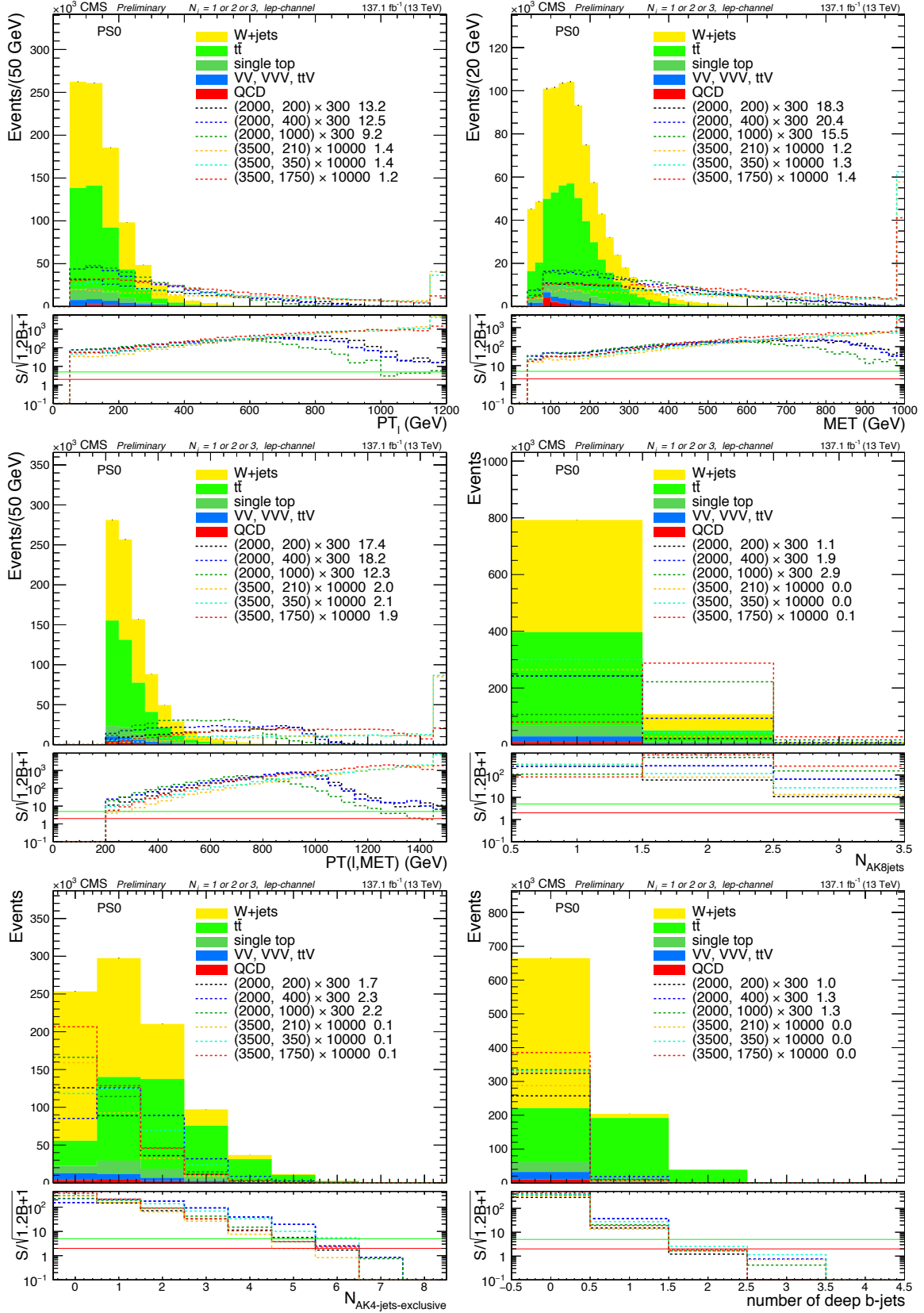


图 4.13 初选条件下的 RunII 三年样本电子 + μ 子道的一些运动学变量分布，分别为轻子 p_T 、MET, $p_T(W) = p_T(\ell, \nu)$, $N_j = N_{jAK8}$ 、 N_{jAK4} 和 N_b^M ，图例右上角说明中信号后面第一个数为信号的放大倍数使信号可见，第二个数为下方显著度的积分用于比较不同条件下显著度的比较。

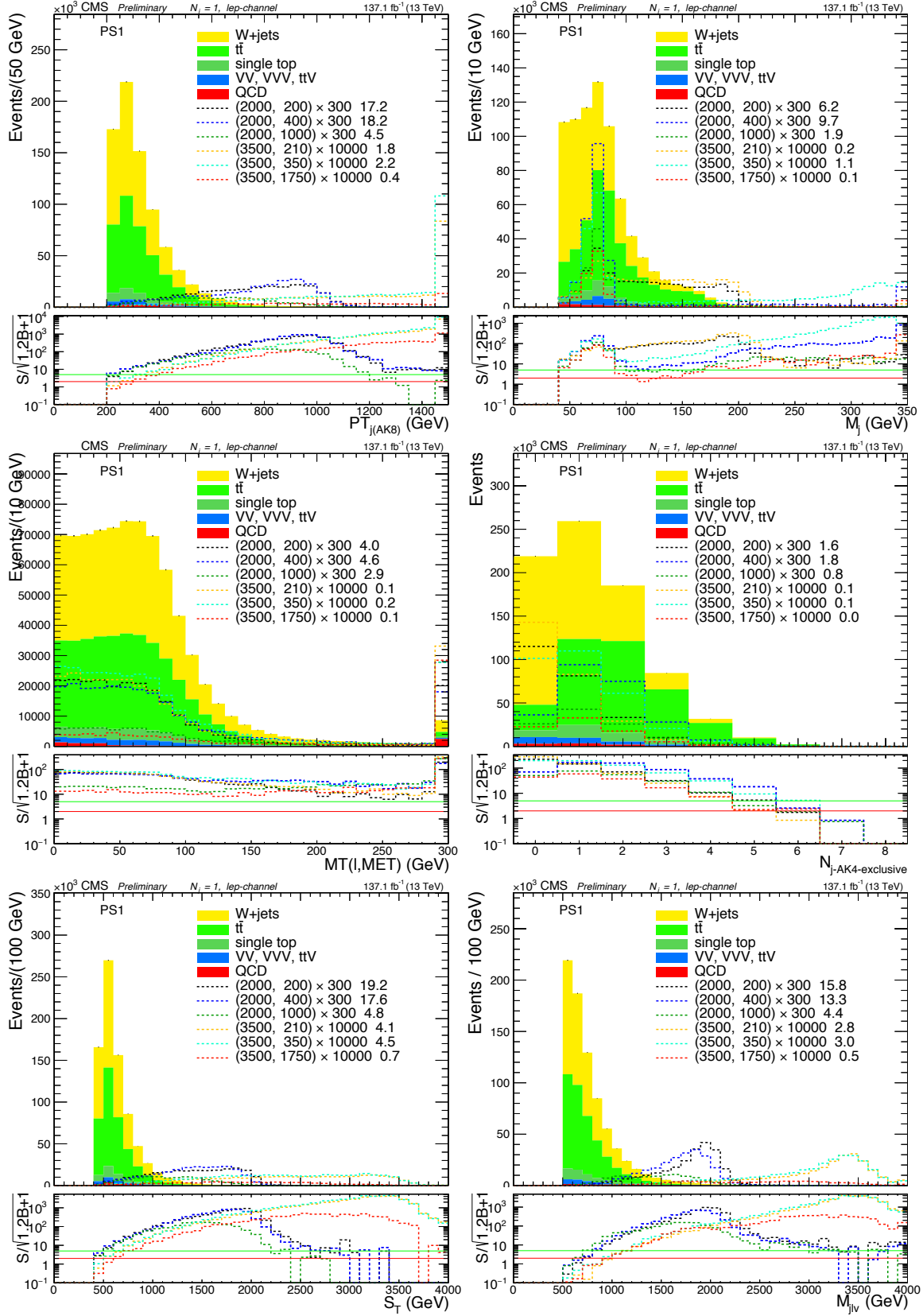


图 4.14 初选条件下且只有 1 个 AK8 喷注对应 RunII 三年样本电子 + μ 子道的一些运动学变量分布, 分别为 AK8 喷注 p_T 、soft drop (SD) 质量、横向质量 $M_T(\ell, MET)$ 、 N_{j-AK4} 、 S_T 和 M_{jlv} , 最后一张图反应了信号的合并情形。

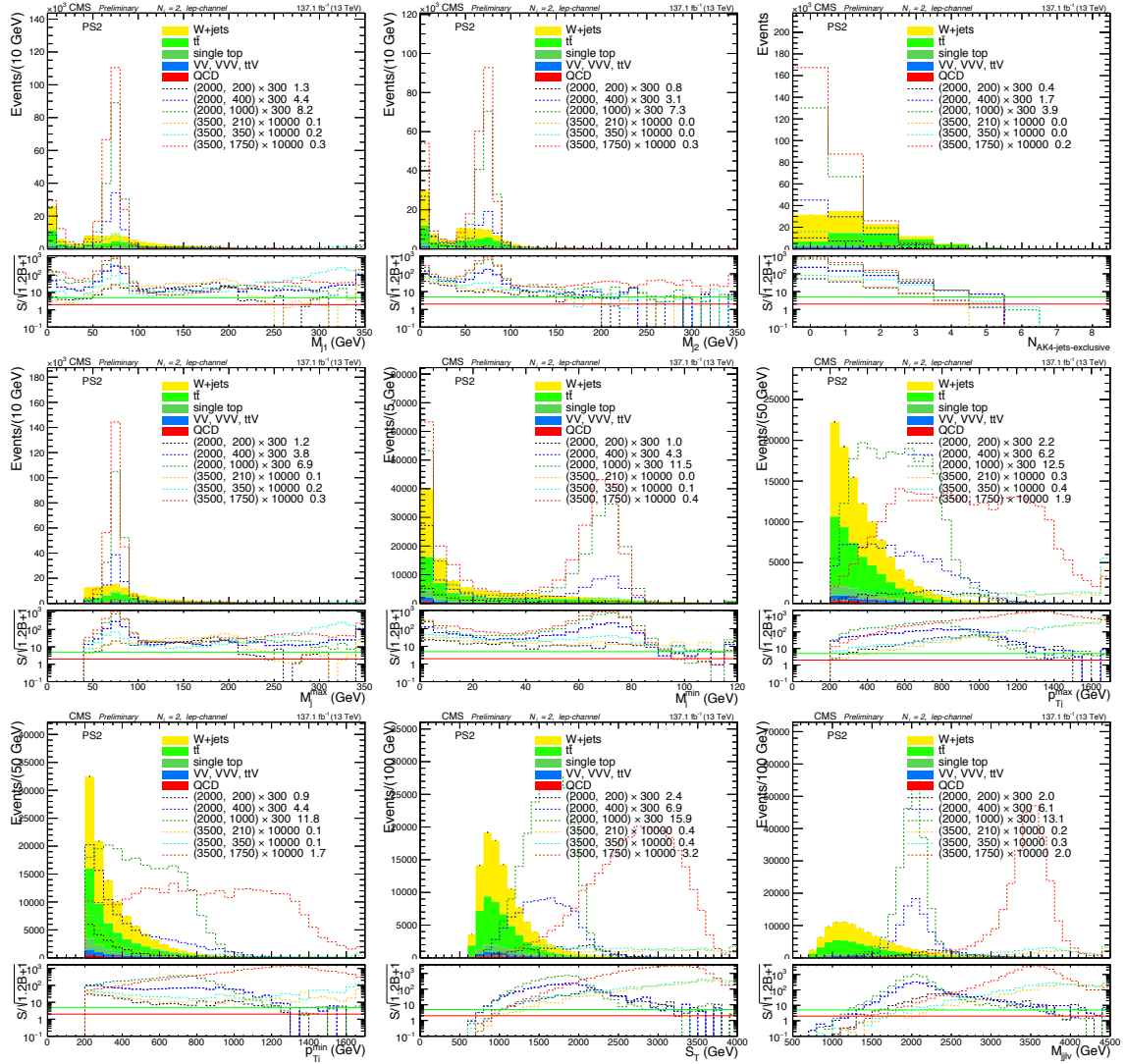


图 4.15 初选条件下且只有 2 个 AK8 喷注对应 RunII 三年样本电子 + μ 子道的一些运动学变量分布，分别为喷注 p_T 和 soft drop 质量。喷注 j 的数字下标表示按其横动量降序排序，上标 $M^{max, mix}$ 表示按喷注 soft drop 质量排序。另外包括 AK8 喷注外的 AK4 喷注数目 $N_{AK4-jets-exclusive}$ 、AK8 喷注数目 N_j 、 S_T 、 $M_{j\ell\nu}$ ，其中 S_T 为轻子衰变的 W 的横动量与所有喷注横动量之和，最后一个变量反应了信号的分立情形。

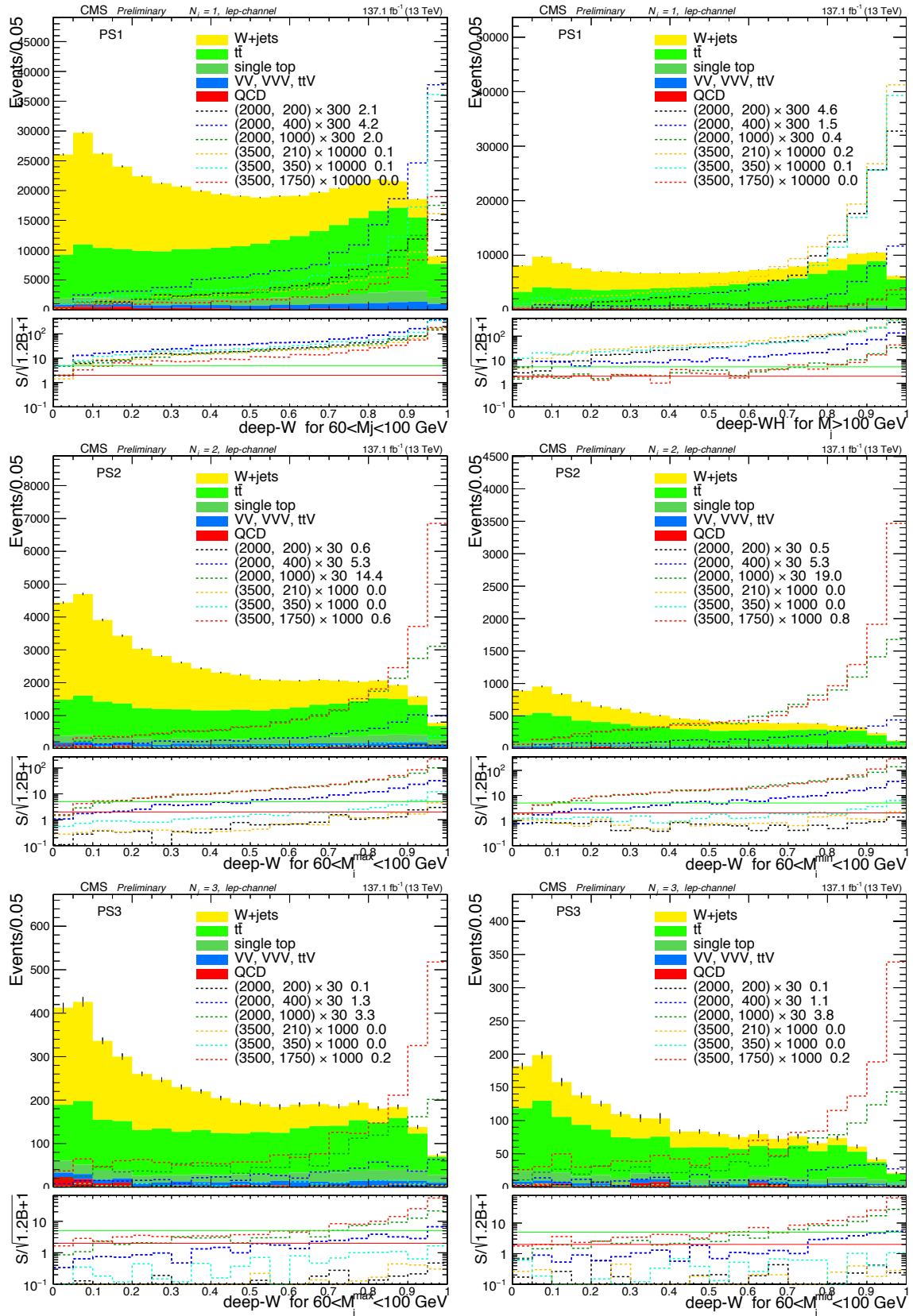


图 4.16 信号和本底 Deep-AK8 标记值在初选条件下的分布，第一、二、三行分别表示 $N_j=1$ 、2、3。右上图是 radion 标记，所以使用 $M_j > 100$ GeV，其他图喷注质量都满足 $60 < M_j < 100$ GeV。

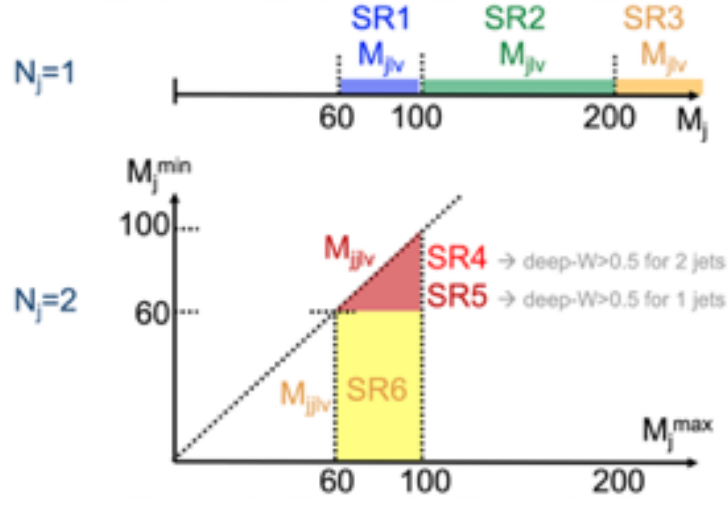


图 4.17 信号区根据喷注质量和 deepAK8 标记分为 6 个正交的区域，定义为 SR1-6。

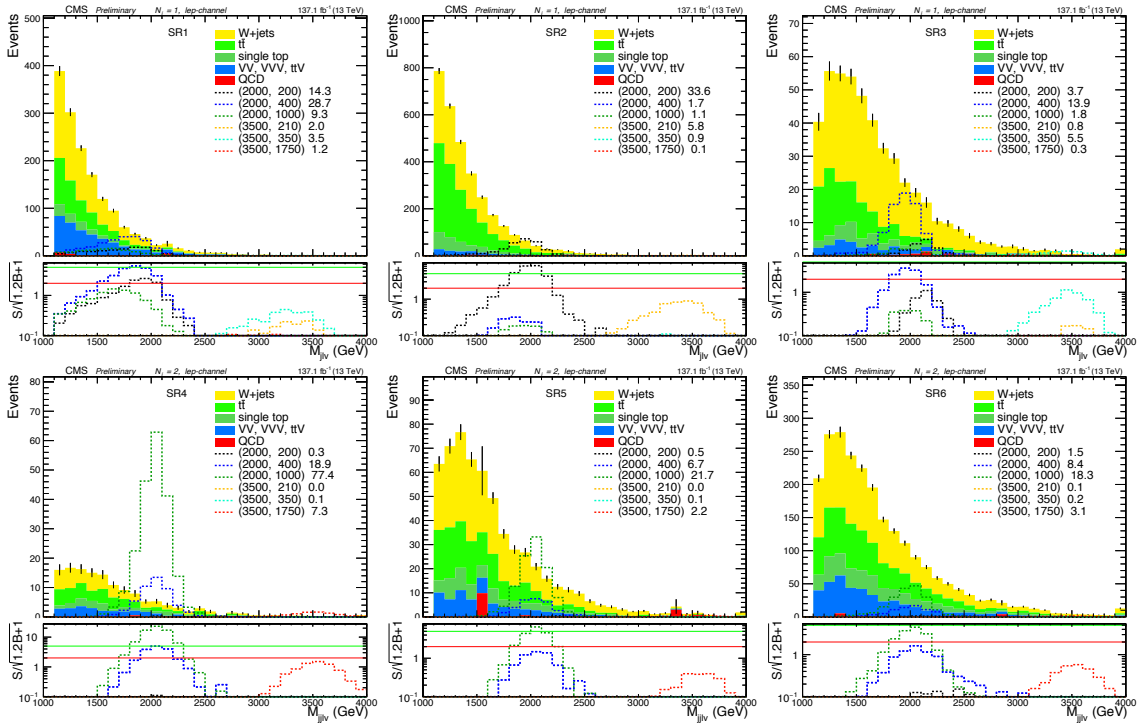


图 4.18 6 个信号区的观测：不变质量对应于 $M_{j\ell\nu}$ SR1、SR2、SR3（上）与 $M_{jj\ell\nu}$ 对应于 SR4、SR5、SR6（下）。图中蒙卡形状对应于原始模拟和截面归一到数据亮度。

4.6 数据与蒙卡模拟样本对比

在本节会使用数据与蒙卡模拟样本比较一些重要物理量，分为 μ 子、电子和 μ 子 + 电子道。另外也按收集年份分解来观察是否在不同年份有相同的表现。2018 年数据采集时，由于强子量能器问题导致在特殊 $\eta - \phi$ 窗口内探测到假的电子信号，因此分析中这部分事例会被丢弃掉。

在作比较时，除了使用之前章节4.4.1中定义的初选条件，还需要排除掉信号区域的事例来保证不直接观测敏感数据，要求 $M_{j\ell\nu}(M_{jj\ell\nu}) < 1.1 \parallel S_T < 1 \parallel N_b^M \geq 1 \parallel N_{j-AK4} \geq 3$ ，新的条件图中用“PS_”表示。每张图 PS_ 之后的数字是数据与蒙卡事例数之比，图中蒙卡样本被归一到数据积分亮度 137fb^{-1} 。

4.6.1 2018 年电子道 C 与 D 时期异常

强子量能器收集塔 HEM15/16 在 2018 年 C 与 D 数据采集期间出现异常 [33]，导致在特殊区间电子严重超出： $-1.57 < \phi(e) < -0.87$ ， $-2.5 < \eta(e) < -1.3$ 。HEM15/16 在 Run 319077 之后停止运行，为了解决这个问题，如果 Run 319077 之后事例中含有电子或喷注落在这个 $\eta - \phi$ 窗内，这个事例将会被丢掉。这项操作对蒙卡样本在信号区影响可以忽略不计，因此只对数据进行如上处理。图4.19展示了使用 2018 年电子道数据丢弃事例前后一些物理量的对比，做丢弃事例处理后问题得到了有效的解决。

4.6.2 不同轻子道数据蒙卡对比

图4.20、4.21和4.22展示了 RunII 3 年内的数据一些物理量按 μ 子 + 电子道、 μ 子和电子道从左到右的分布。所有的图中除了 deepAK8 相关量数据和蒙卡都合理吻合，因此本文选择合并 μ 子与电子道。deepAK8 标记修正会在第五章中介绍。

4.6.3 不同采集年份数据蒙卡对比

图4.23、4.24和4.25 中合并 μ 子 + 电子道，按数据采集年份分解，从左到右分别为 2016、2017 和 2018 年，物理量有上一节相同。三年的积分亮度分别为： 35.86fb^{-1} 、 41.53fb^{-1} 和 59.72fb^{-1} 。同样除了 deepAK8 相关的量，其他吻合的都很好。

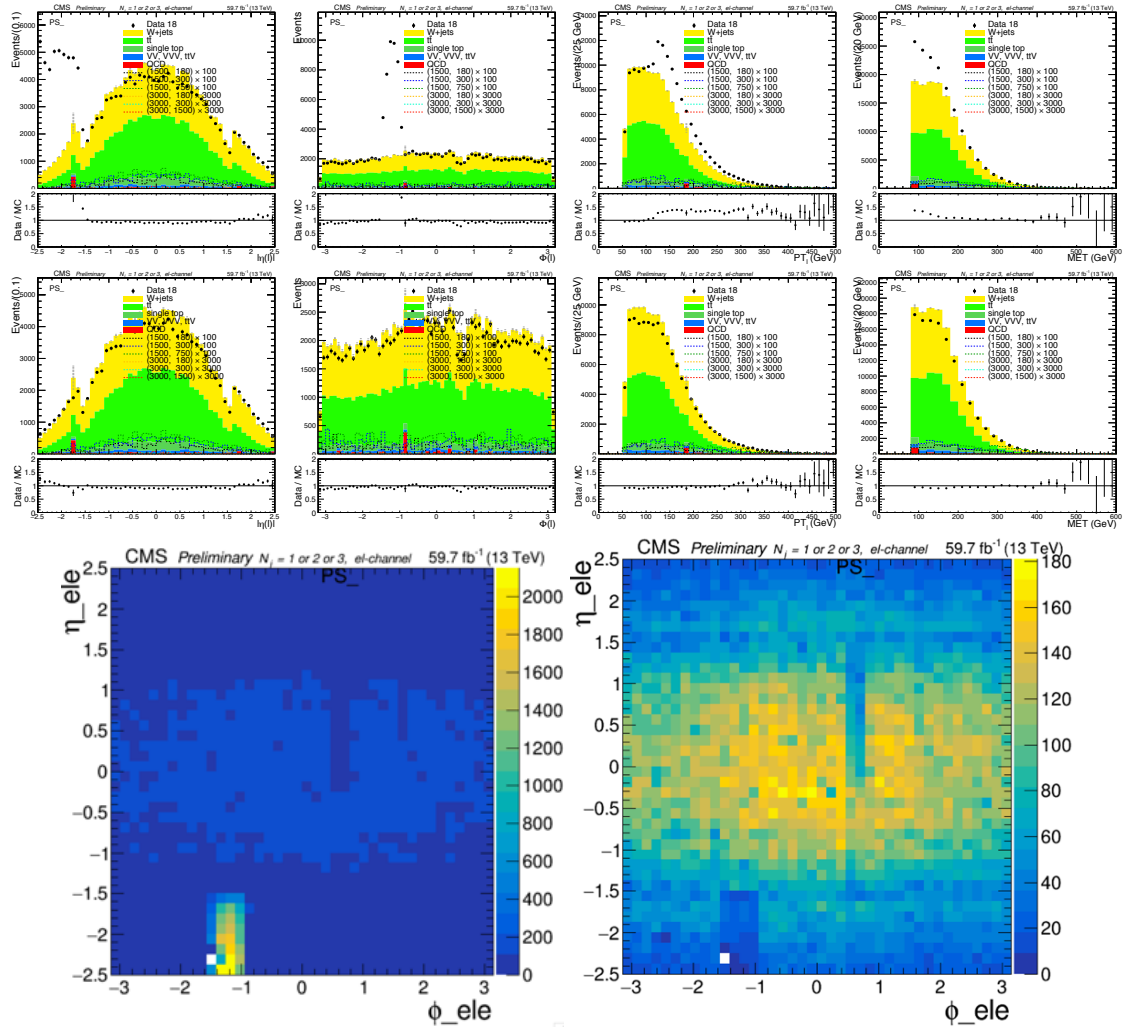


图 4.19 2018 年电子道数据与蒙卡样本的对比，第一行是丢弃事例 $\eta(ele)$ 、 $\eta(ele)$ 、 $p_T(ele)$ 及 MET 的分布对比，可以发现在 $-1.57 < \phi(e) < -0.87$ 且 $-2.5 < \eta(e) < -1.3$ 窗口内数据的显著超出，第二行丢弃在窗口内存在电子或喷注的事例后的分布，结果得到很大改善，第三行是电子 $\eta-\phi$ 在处理前（左）后（右）的 2D 分布图。

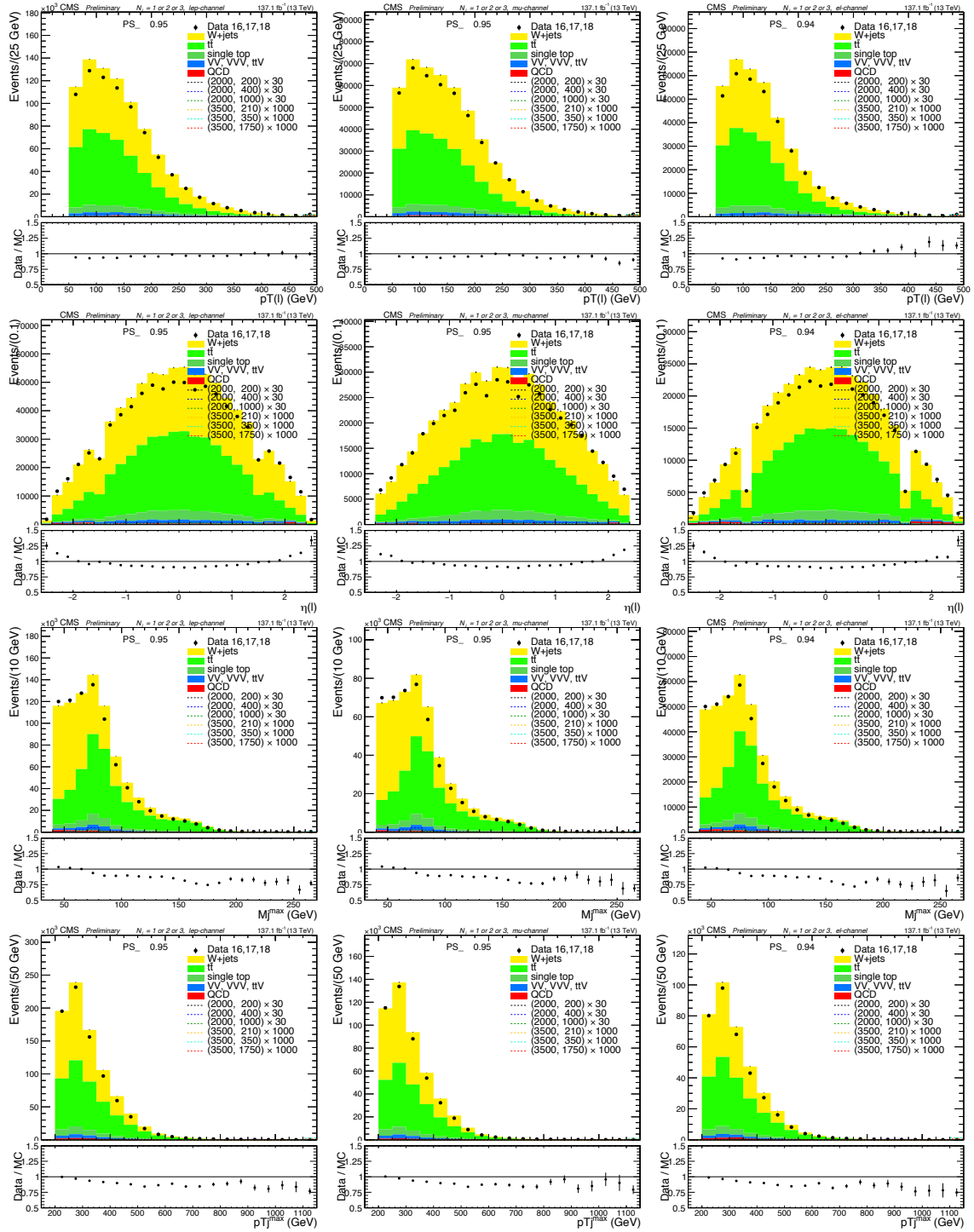


图 4.20 RunII 3 年内的数据得到的部分物理量的分布图，按 μ 子 + 电子道、 μ 子和电子道从左到右的顺序。

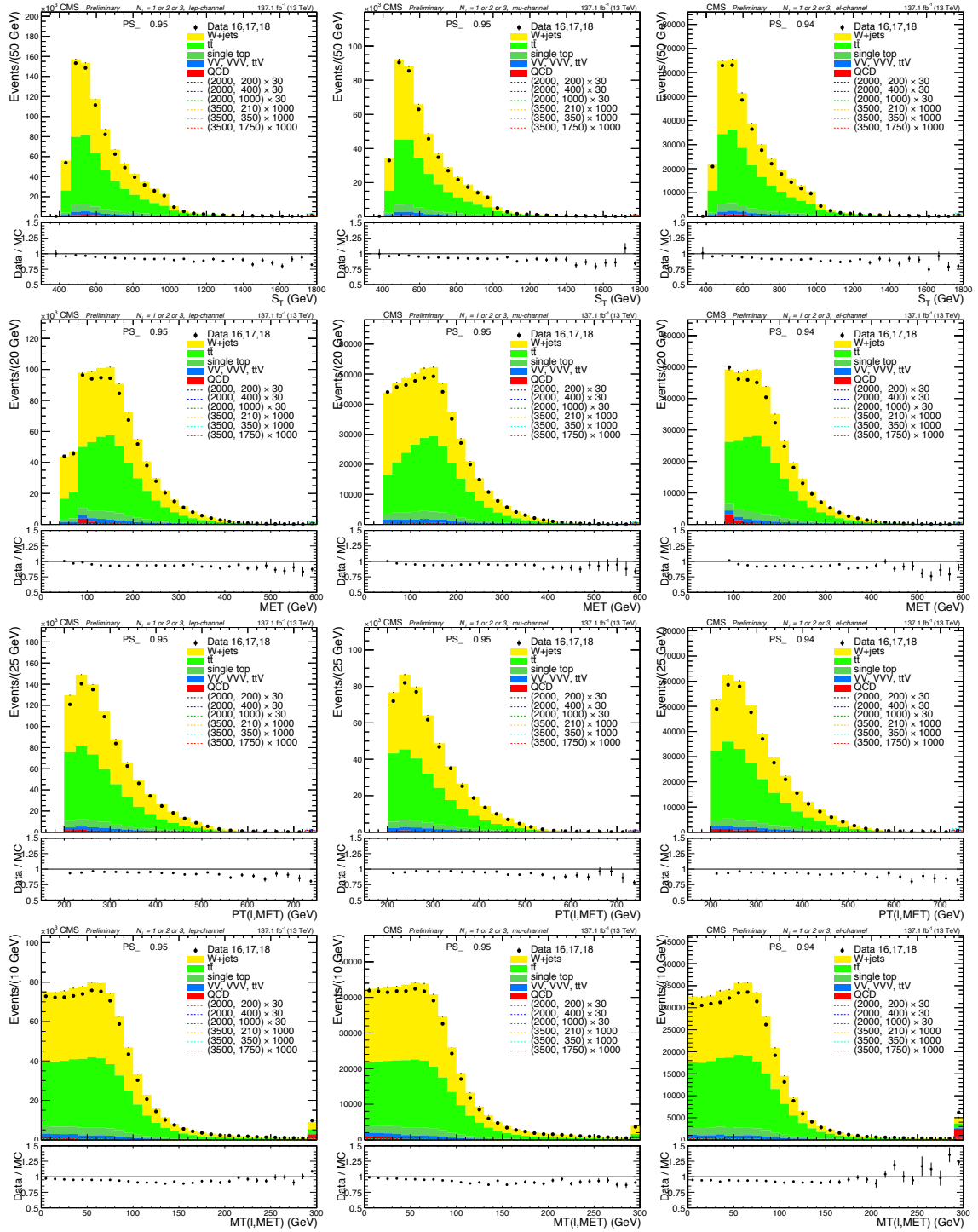


图 4.21 RunII 3 年内的数据得到的部分物理量的分布图，按 μ 子 + 电子道、 μ 子和电子道从左到右的顺序。

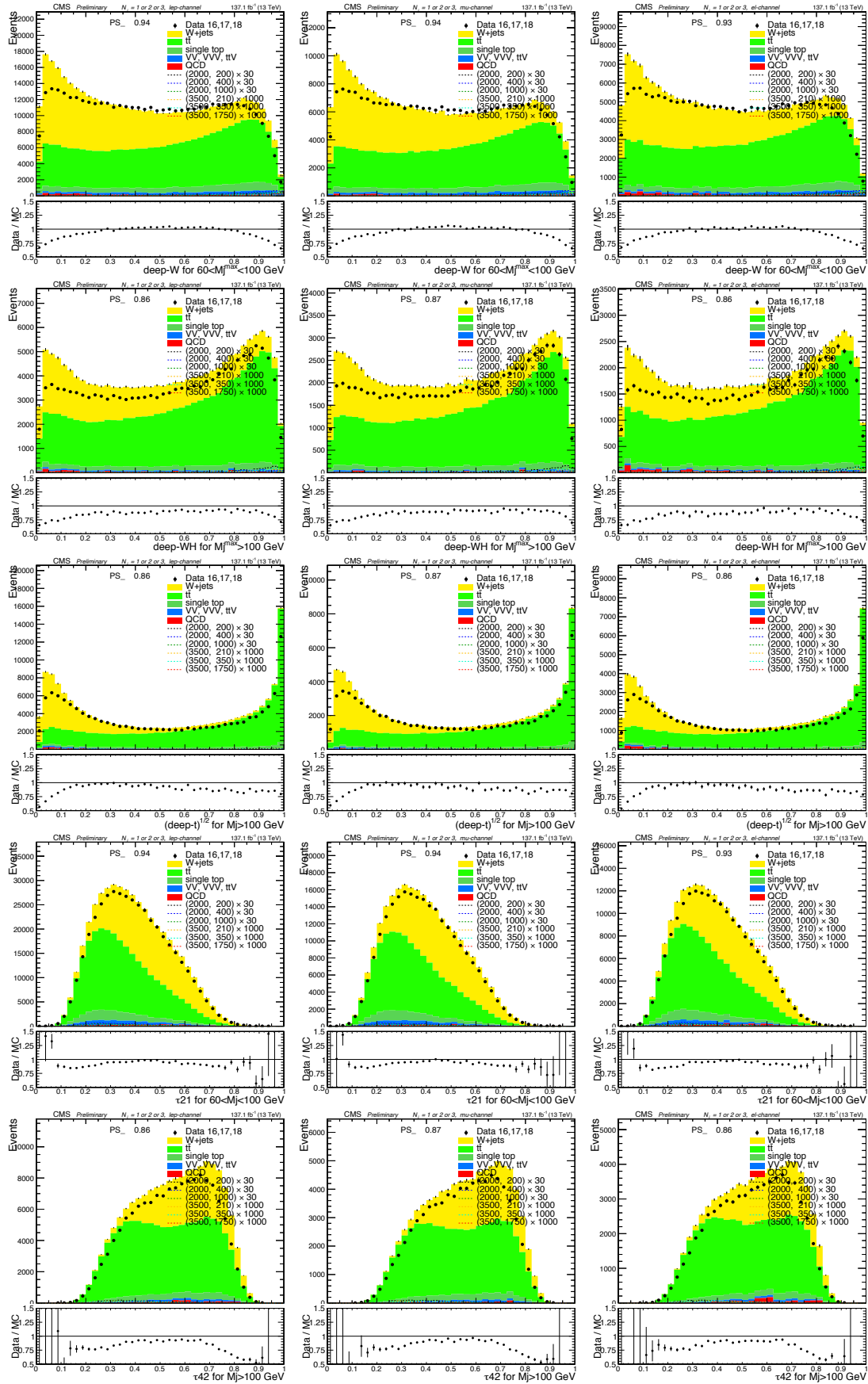
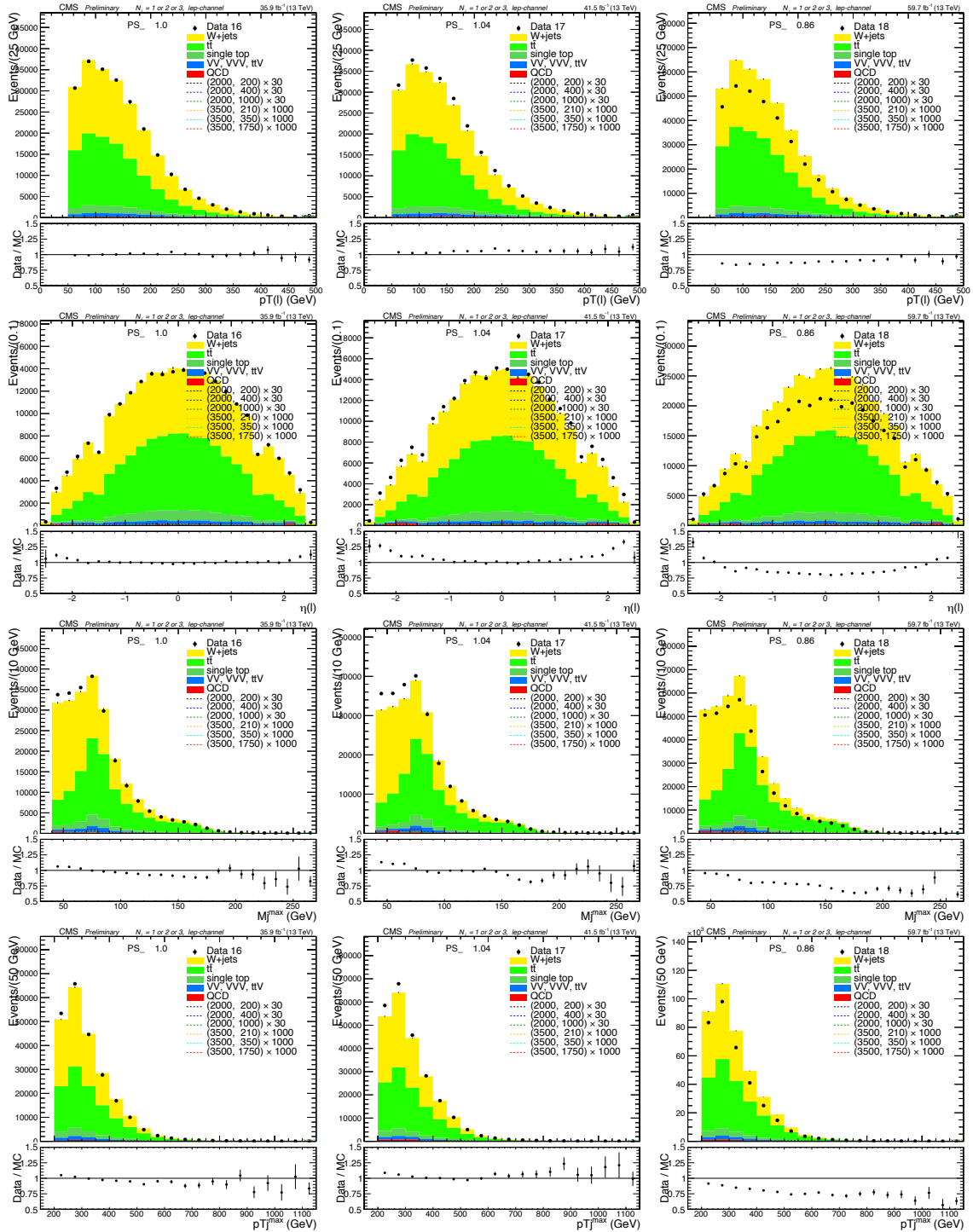
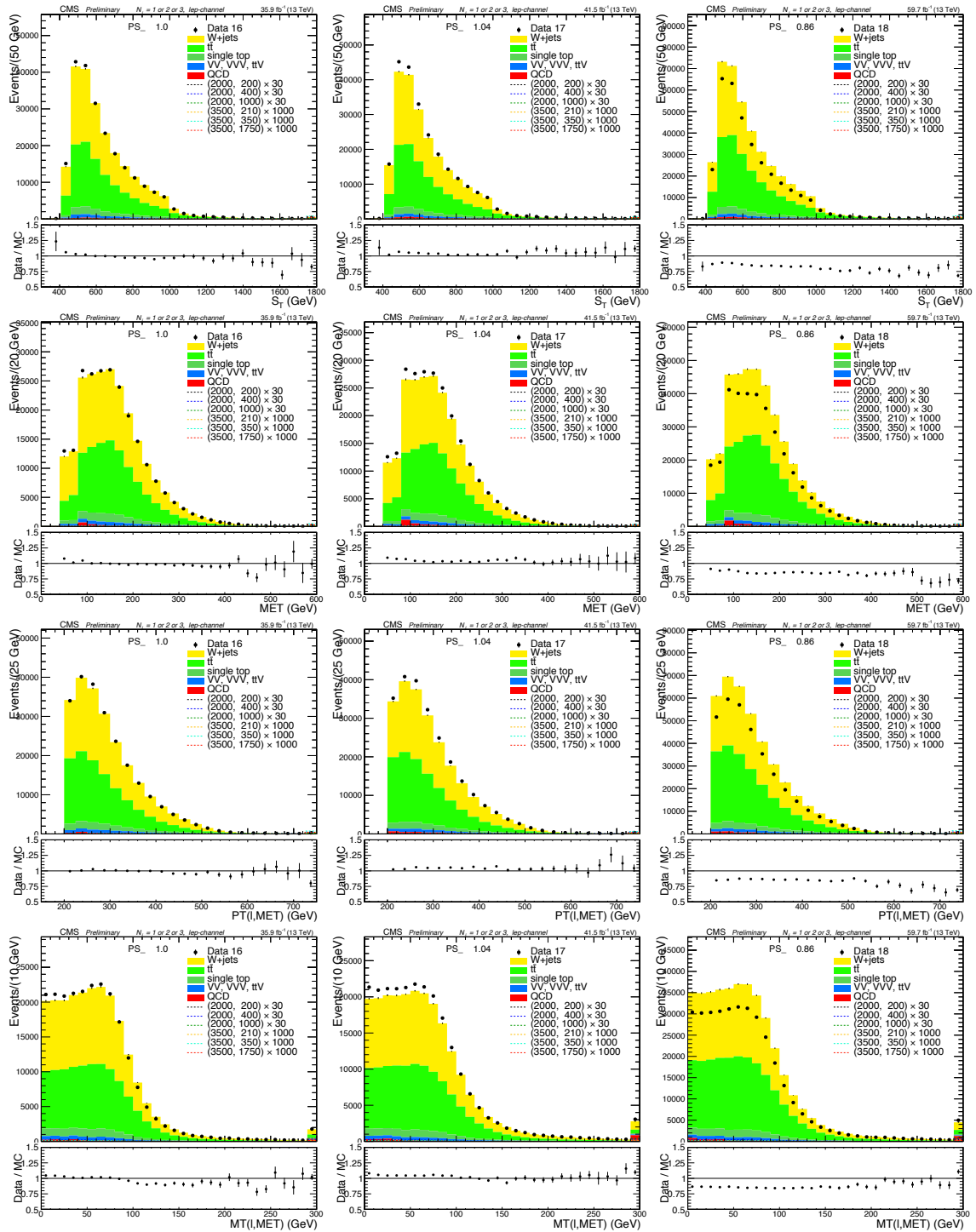
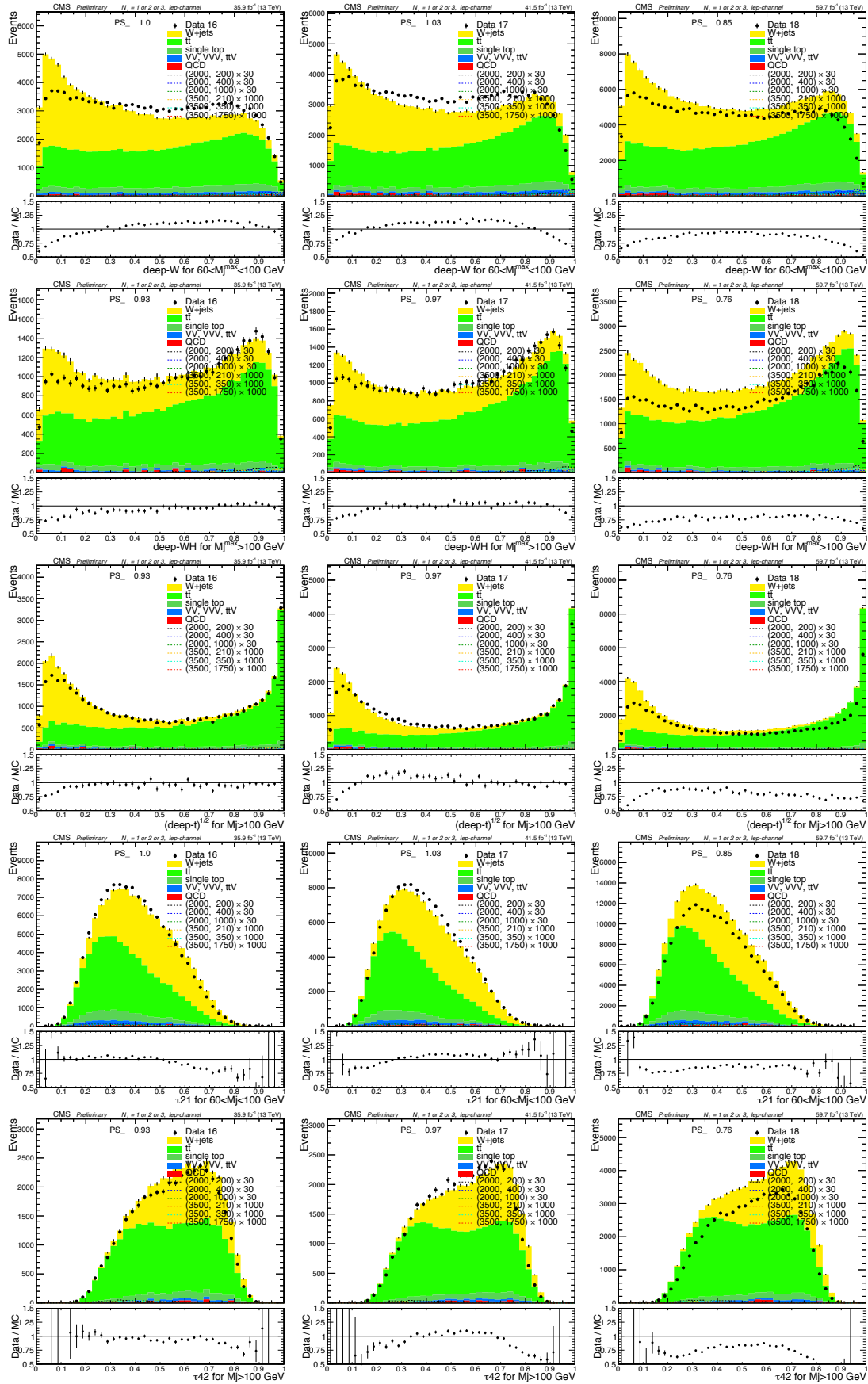


图 4.22 RunII 3 年内的数据得到的部分物理量的分布图，按 μ^+ 子 + 电子道、 μ 子和电子道从左到右的顺序。


 图 4.23 μ 子 + 电子道合并后部分物理量的分布图，按 2016、2017、2018 年从左到右的顺序。


 图 4.24 μ 子 + 电子道合并后部分物理量的分布图，按 2016、2017、2018 年从左到右的顺序。


 图 4.25 μ 子 + 电子道合并后部分物理量的分布图, 按 2016、2017、2018 年从左到右的顺序。

第五章 三玻色子共振态寻找本底估计方法

在4.5章介绍了分析使用的6个信号区以及对应的观测量 $M_{j\ell\nu}$, $M_{jj\ell\nu}$ 。图4.18为6个信号区观测量的分布，为了更直观展示各本底中标准模型物理过程的占比，对应的饼图如图5.1所示，图中给出了每个信号区的事例数与相对占比。其中：

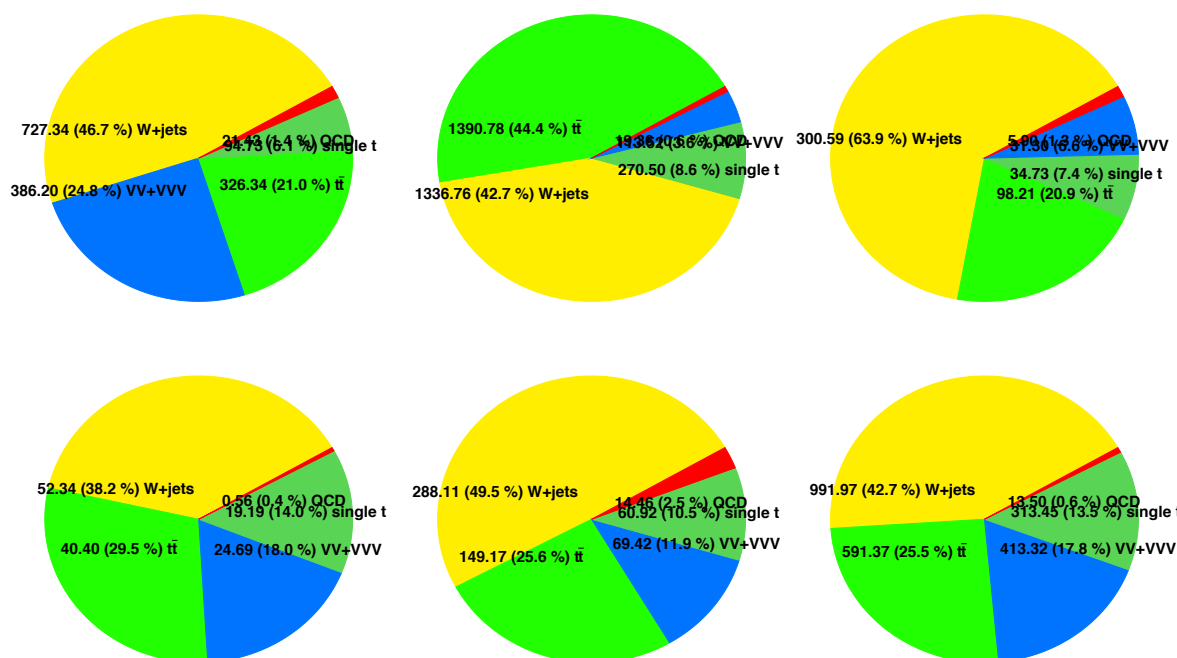


图 5.1 6 个信号区中本底不同物理过程对应的事例数与占比饼图，上面为信号区 1-3，下方为信号区 4-6。

- W+jets 是最主要的本底，占比约为 40-65%；
- top (single top 与 $t\bar{t}$) 是占比第二多的本底过程，约为 20-45%；
- 其他稀有过程包括 WW、WZ、ZZ、WWW、WWZ、WZZ、ZZZ、 $t\bar{t}W$ 、 $t\bar{t}Z$ 、QCD，总共约占 5-25%。

所以分析中本底按这三类划分：W+jets、top 与剩余过程。

正如第四章讨论的，图4.22左下方关于 deepAK8 标记，数据与蒙卡样本并不是很好的吻合，特别是标记值左测和右侧即接近 0 或 1 的部分。因此需要研究这种差异的原因，并且从数据中拿到相应的比例缩放因子来修正蒙卡事例来修正差异。

在 deepAK8 缩放因子的研究中使用了4.6中提到的“PS_”，即使用去掉信号区的敏感事例的初选条件来获得大量的低横动量和高横动量的喷注样本，并且将喷注与产生

子水平的不同物理对象匹配，开发了全新的“矩阵方法”来计算缩放因子。

5.1 本底过程喷注的匹配分解

图5.2为 PS_ 条件下数据与蒙卡样本的对比，如信号区条件所述，在喷注质量 60-100、>100 时分别使用 deep-W 和 deep-WH，图中第二行展示了两种标记在对应喷注质量下的分布，另外喷注性质在高、低横动量时也会有所不同，因此分别使用 400、500GeV 作为两种情况下高低横动量的判别区分。

这图中可以看到这两种 deep 标记并不是很好的吻合，在标记值两侧这种差异更加明显。从这些图中并不能看出是哪种过程导致了差异，因为每种过程都包含了不同物理对象的喷注。为了研究模拟差异的起源，需要将喷注与产生子水平的物理对象匹配，比如 W、Z 玻色子，轻夸克/胶子（用“q/g”标记），顶夸克（后续用 top 或 t 表示）等。重建水平的喷注与产生子水平的粒子匹配条件在表5.1中，top 喷注强子衰变包含部分合并（ $t \rightarrow bq$ ，用 t^2 标记）和全合并（ $t \rightarrow bqq$ ）情况，其中后者按其是否有额外的夸克或胶子组成喷注分为两种情形，分别标记为 t^3 、 t^4 ，其中额外的夸克或胶子的母粒子不能是 W 或者 t 且横动量大于 50GeV，上标中的数目表示了喷注内夸克的个数。而轻子衰变的 t（ $t\nu$ ）由于中微子能量丢失，所以用 $t^{\ell b}$ 表示。匹配总体的原则是喷注与对应的产生粒子距离满足 $\Delta R < 0.6$ ，与产生粒子的子代距离满足 $\Delta R < 0.8$ 。在这种匹配规则下，只有很少的喷注没有成功匹配。

物理对象:	W $\rightarrow$ qq	g/q	t^4	t^3	t^2	$t^{\ell b}$	Z $\rightarrow$ qq
ΔR	(W, j) < 0.6	(g/q, j) < 0.6	(t, j) < 0.6	(t, j) < 0.6	(t, j) < 0.6	(t, j) < 0.6	(Z, j) < 0.6
ΔR	(q_{1W}, j) < 0.8	-	(b_i, j) < 0.8	(b_i, j) < 0.8	(b_i, j) < 0.8	(b_i, j) < 0.8	-
ΔR	(q_{2W}, j) < 0.8	-	(q_{1W}, j) < 0.8	(q_{1W}, j) < 0.8	(q_{1W}, j) < 0.8	(ℓ_W, j) < 0.8	-
ΔR	(b_i, j) > 0.8	-	(q_{2W}, j) < 0.8	(q_{2W}, j) < 0.8	(q_{2W}, j) > 0.8	-	-
ΔR	-	-	(g/q, j) < 0.8	! [(g/q, j) < 0.8]	-	-	-

表 5.1 重建的 AK8 喷注与产生子水平的粒子匹配规则，每一种物理对象对应的几种距离条件要全部满足，匹配优先级为 t、W、Z、q/g。

图5.3展示了喷注质量 M_j 及 deep-W (WH) 将喷注匹配分解后的分布，图中 W+jets 和 top 是按 b-tagged 喷注数量为 0 或大于 0 计算得到缩放因子归一的，如果 b 喷注数量为 0 则主要为 W+Jets 本底，如果非 0，则本底过程主要为 top，这一步的目的是为了将蒙卡样本归一到数据上，从而只研究不同物理对象对 deep-W (WH) 形状的影响。从第一行可以看出，g/q 在整个质量谱都有分布， t^2 分布在 $< \sim 145\text{GeV}$ ，W 集中在 60-100GeV，全合并的 top（ t^3 、 t^4 ）分布在 $> \sim 120\text{GeV}$ 。基于质量分布，把 M_j 分为 60-120、>120GeV 两部分以分离 W 和 $t^{3,4}$ ，这样低质量区由 g/q、 t^2 和 W 主导，高质量区间由 t^2 、g/q 和 $t^{3,4}$ 主导。

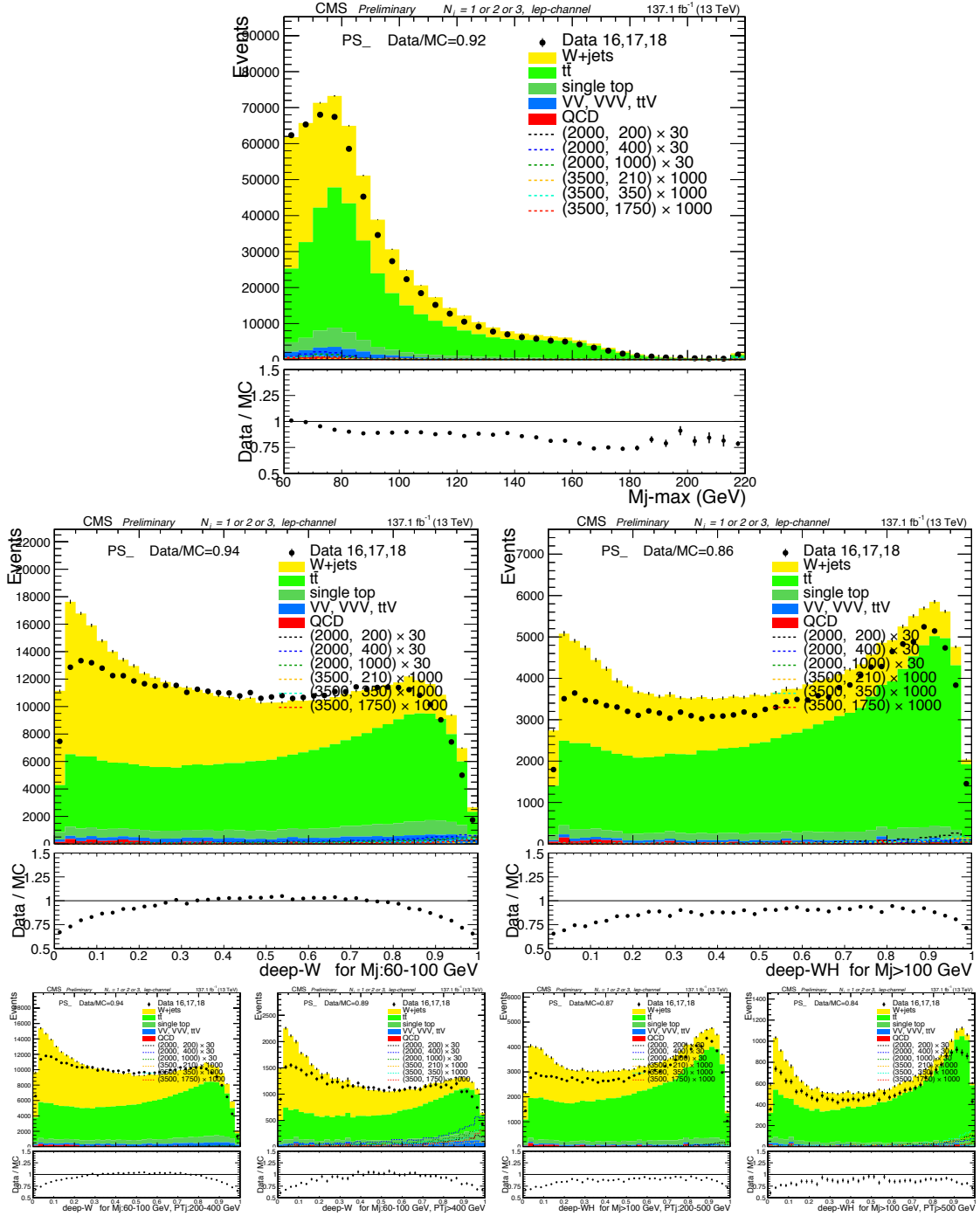
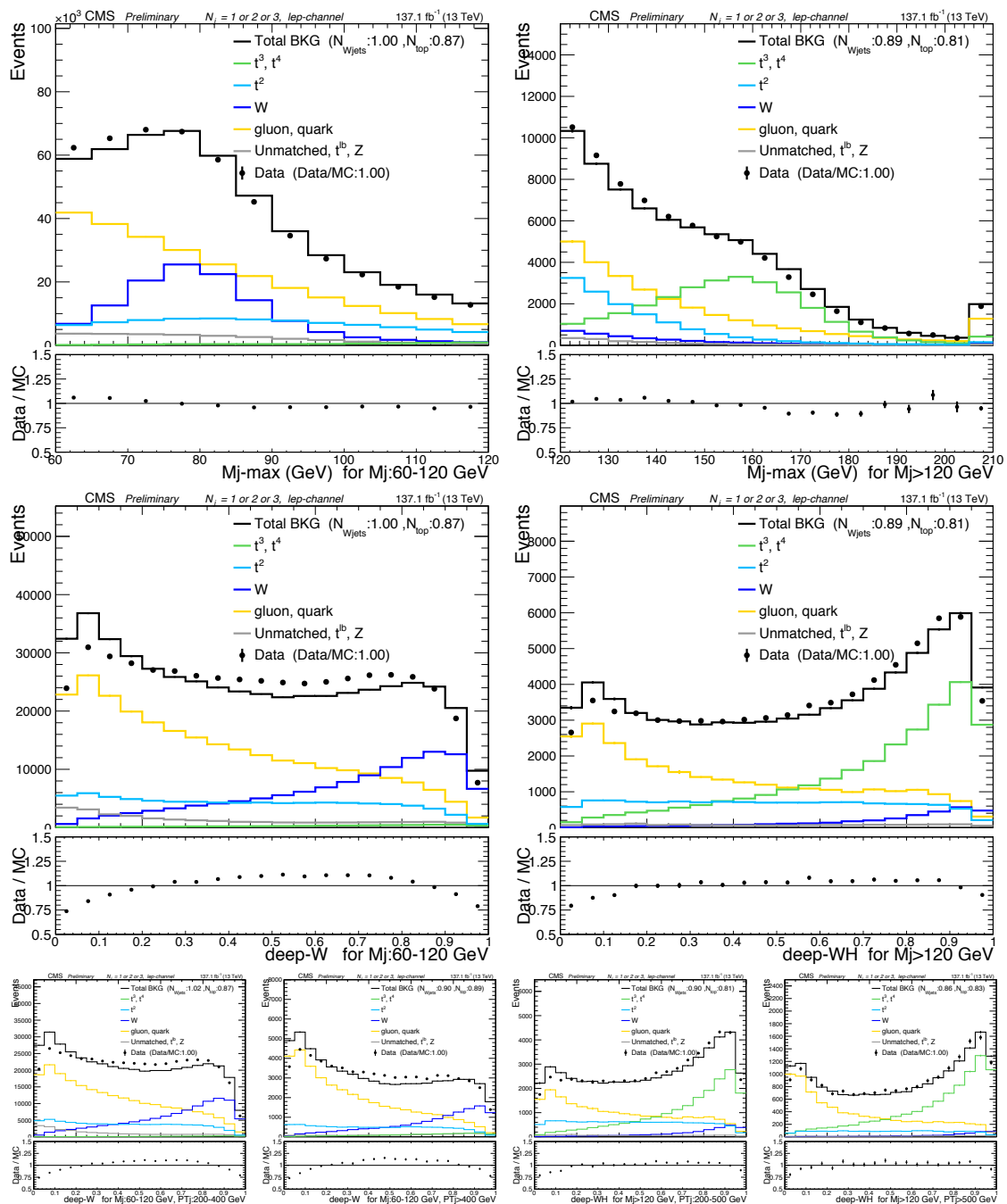


图 5.2 最上方是事例中质量最大的喷注质量的分布，第二行是 deep-W 与 deep-WH 在喷注质量 60-100、 >100 GeV 时的分布，第三行是 deep-W 与 deep-WH 分别在低、高横动量时的分布。

由于每一种物理对象都可能对 deep-W (WH) 有不同的影响，图5.3第二、三行可以看到每张图在全范围内占主导的三种喷注都有分布。按常规的方法，需要对每种物理对象找到纯度很高的样本，然后用数据与蒙卡样本的比例作为修正因子。然而这种思路对于本分析是不可行的，需要考虑的物理对象太多，不可能找到对应于每一种的


 图 5.3 M_j 及 deep-W (WH) 将喷注与产生子水平粒子匹配分解后的分布图。

高纯度样本。为了解决这一问题，我们开发了全新的方法。

5.2 修正 deepAK8 标记差异的比例因子

相对于常规的对于每种喷注使用一个高纯度样本的处理方法，我们定义了多个样本，只需要某一种喷注在部分区域有较高的纯度，然后通过求解多个线性方程的方式

来计算比例缩放因子 (scale factor, SFs)。具体方法如下, 首先将"PS_"事例按喷注质量 M_j 分为 60-120、 $>120\text{GeV}$ 两部分, 然后分别找出三种主要物理对象占主导的条件将样本分为三部分。

对于 $M_i \in [60, 120] \text{ GeV}$ 范围内, 样本分为 g/q、 t^2 和 W 主导的三个区域:

1. t^2 占主导: $\text{deep-}t/\text{Wg} > 0.25 \ \& \ \tau_{21} < 0.5$;
2. W 占主导: $! [\text{deep-}t/\text{Wg} > 0.25 \ \& \ \tau_{21} < 0.5] \ \& \ 65 < M_j < 100 \ \& \ \tau_{21} < 0.5 \ \& \ [(N_b^L = 0 \ \& \ \text{deep-}t > 0.2) \ || \ (N_b^L \geq 1 \ \& \ \text{deep-}t > 0.02)]$;
3. g/q 占主导: 剩余事例。

对于 $M_j > 120\text{GeV}$, 分为 t^2 、 g/q 和 $t^{3,4}$ 主导的三个区域:

1. g/q 占主导: $\text{deep-}t < 0.2$;
2. t^2 占主导: $\text{deep-}t > 0.2$ & $120 < M_j < 145$ & $\tau_{32} > 0.6$;
3. $t^{3,4}$ 占主导: 剩余事例。

上述条件可以有效地增强对应物理对象的纯度。图5.4展示了用于 M_j : 60-120GeV 区间做阶段条件的物理量分布。

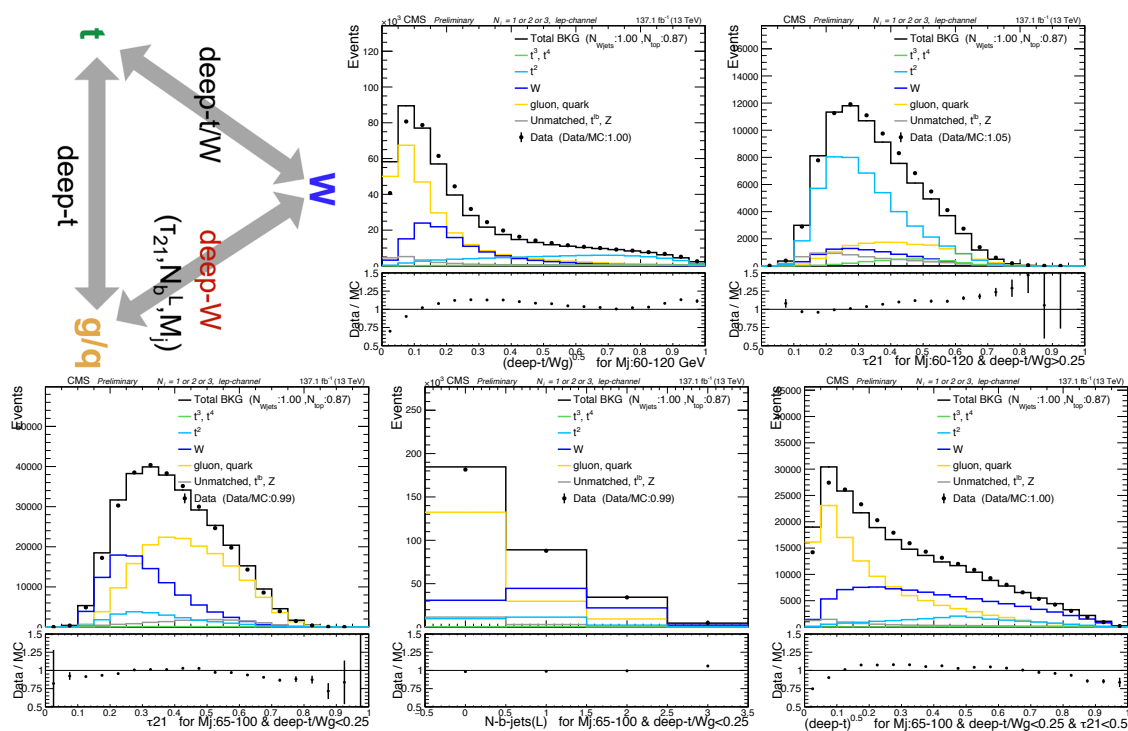


图 5.4 左上方为区分三种对象用到的物理量, 上中为 $(\text{deep-t}/Wg)^{1/2}$ 的分布, 要求 >0.25 来提高 top 纯度得到上右 τ_{21} 分布图, 要求 <0.5 得到 t^2 占主导的样本; 第二行是剩余样本分布图, 按 deep-t 、 N_b^L 、 τ_{21} 和 M_1 被分为 W 占主导和 q/g 占主导的样本。

对于标记值第 i 个分区 (bin) 可以读出每种喷注对应的事例数, 被记为 "g、W、t" (为了描述方便, 将 g/q 记为 g, top 省略了上标), 数据事例数记为 "D", 剩余事例比如

Z 、 t^{lb} 和无法匹配的成分记为" d "，剩余部分相对 D 非常小，通常 $<5\%$ ，因此认为不会对数据蒙卡差异产生贡献，所有的事例数都是已知量。然后引入未知的 SFs " a 、 b 、 c " 对应于 " g 、 W 、 t " 使得蒙卡事例数按相应比例缩放后与数据减剩余事例 ($D-d$) 匹配，即要求满足如下三个方程：

$$D_1 = ag_1 + bW_1 + ct_1 + d_1$$

$$D_2 = ag_2 + bW_2 + ct_2 + d_2$$

$$D_3 = ag_3 + bW_3 + ct_3 + d_3$$

其中下标 1、2、3 表示 3 个主导样本，图5.5第二行从左到依次与公式中下标 1、2、3 对应。上述方程可以写成矩阵的形式：

$$\begin{pmatrix} D_1 - d_1 \\ D_2 - d_2 \\ D_3 - d_3 \end{pmatrix} = \begin{pmatrix} g_1 & W_1 & t_1 \\ g_2 & W_2 & t_2 \\ g_3 & W_3 & t_3 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}。$$

解决这个 3×3 的矩阵就可以得到对应于 deep- W 标记值分区 i 对应的三种主导成分的 SFs，如图5.5下方所示。

在计算时，如果使用相同的 0.05 作为分区宽度，在某些区域内会导致很低的统计量，而矩阵方法计算中使用分区事例数，统计量太低会带来较大的误差，因此整合了数据与蒙卡相对趋势不变且喷注成分变化较小的分区，比如图5.5中左，0-0.35 使用了一个很宽的分区来增加统计量，第三行 W 的 SF 曲线很好的反应出了 W 成分对标记值差异的贡献。

对于 $M_j > 120\text{GeV}$ ，同样有 3 中占主导的成分，可以得到类似的 3×3 矩阵，从而解出对于 deep-WH 的 SFs，如图 5.6下方所示。

得到 SFs 之后需要对所有的蒙卡样本每个事例基于喷注类型和质量重新分配权重。基于图5.5与5.6得到的结果，在信号区要求 deep- $W(WH) > 0.7$ ，因此 W 和全合并 top 在高喷注质量区分别平均缩减 $\sim 15\%$ 和 $\sim 10\%$ ，而 q/g 取决于它们的质量。在信号区 4 和 5 (SR4、SR5)，分立信号有 2 个 W 喷注，因此每个事例权重需要乘 SFs(W) 两次，会在5.4中讨论。

5.2.1 矩阵方法的说明

关于 SFs 计算使用的矩阵方法，有如下的细节与说明：

1. 在使用该方法时，并没有用数据拟合，而是将蒙卡样本按比例缩放与数据匹配。
2. 方程中系数 a 、 b 、 c 在不同的子样本中被认为是相同的，实际由于不同样本物理的差异，真实的修正因子是不相同的，这样的处理会带来一定的偏离，在讨论系统误差时会考虑偏离带来的影响。

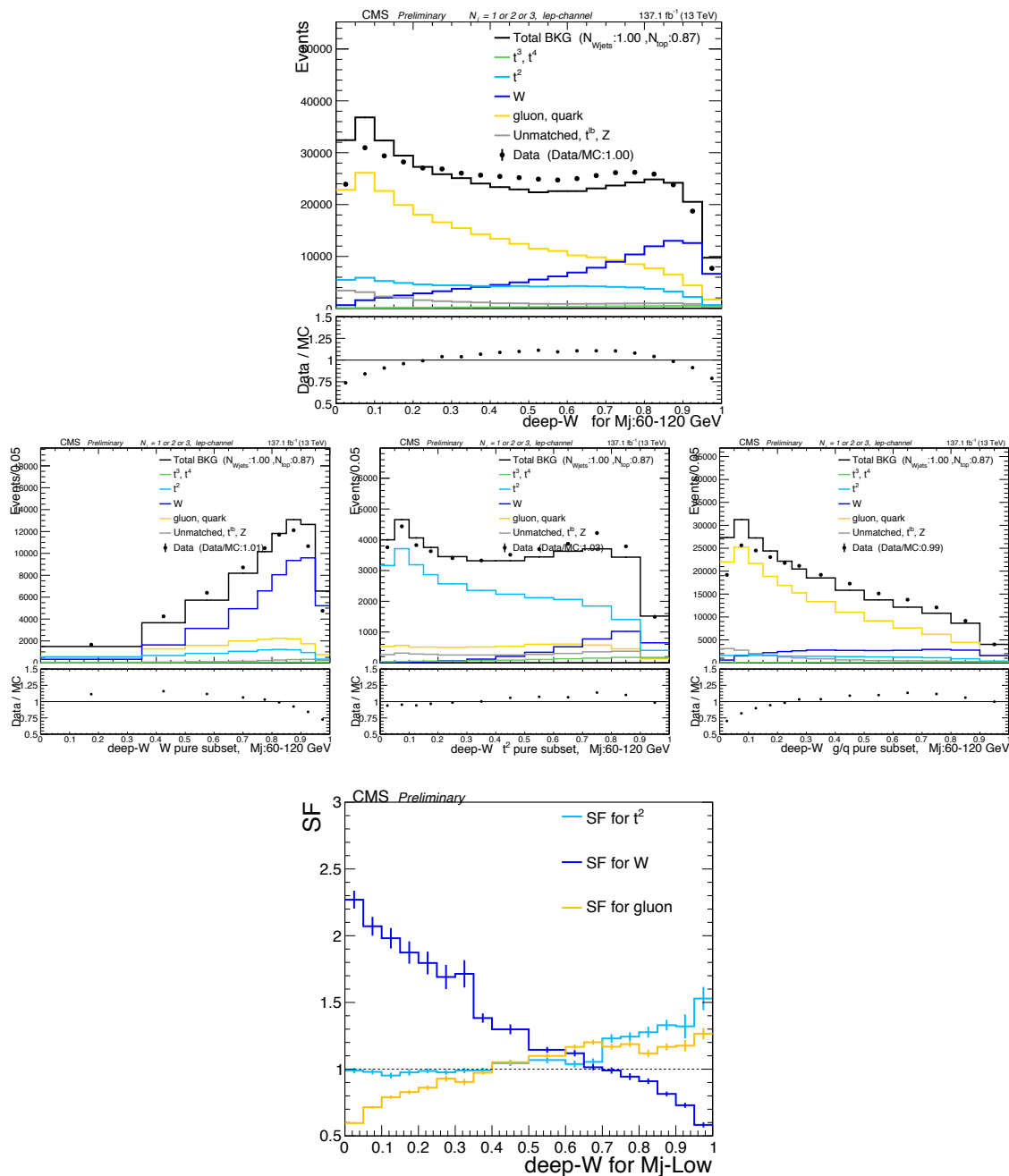


图 5.5 deep-W 在 M_j 取 60-120GeV 时的分布 (上图), 分解为 3 个样本后的分布 (中) 和最终的 SFs 分布 (下)。

- 在矩阵方法中, 比如低质量区, 定义了 g/q 、 t^2 和 W 分别占主导的 3 个子样本, 如果其中 2 个子样本中的三种类型喷注的事例数比例相近, 即矩阵中 2 行相除近似为一个相同的常数, 那么会导致矩阵行列式为 0 无解, 实际不可能为 0, 但是如果行列式很小也会带来很大的误差。因此在定义子样本时要满足不同成分的纯度要求, 这也是设计 3 种成分分别占主导的 3 个子样本的原因。

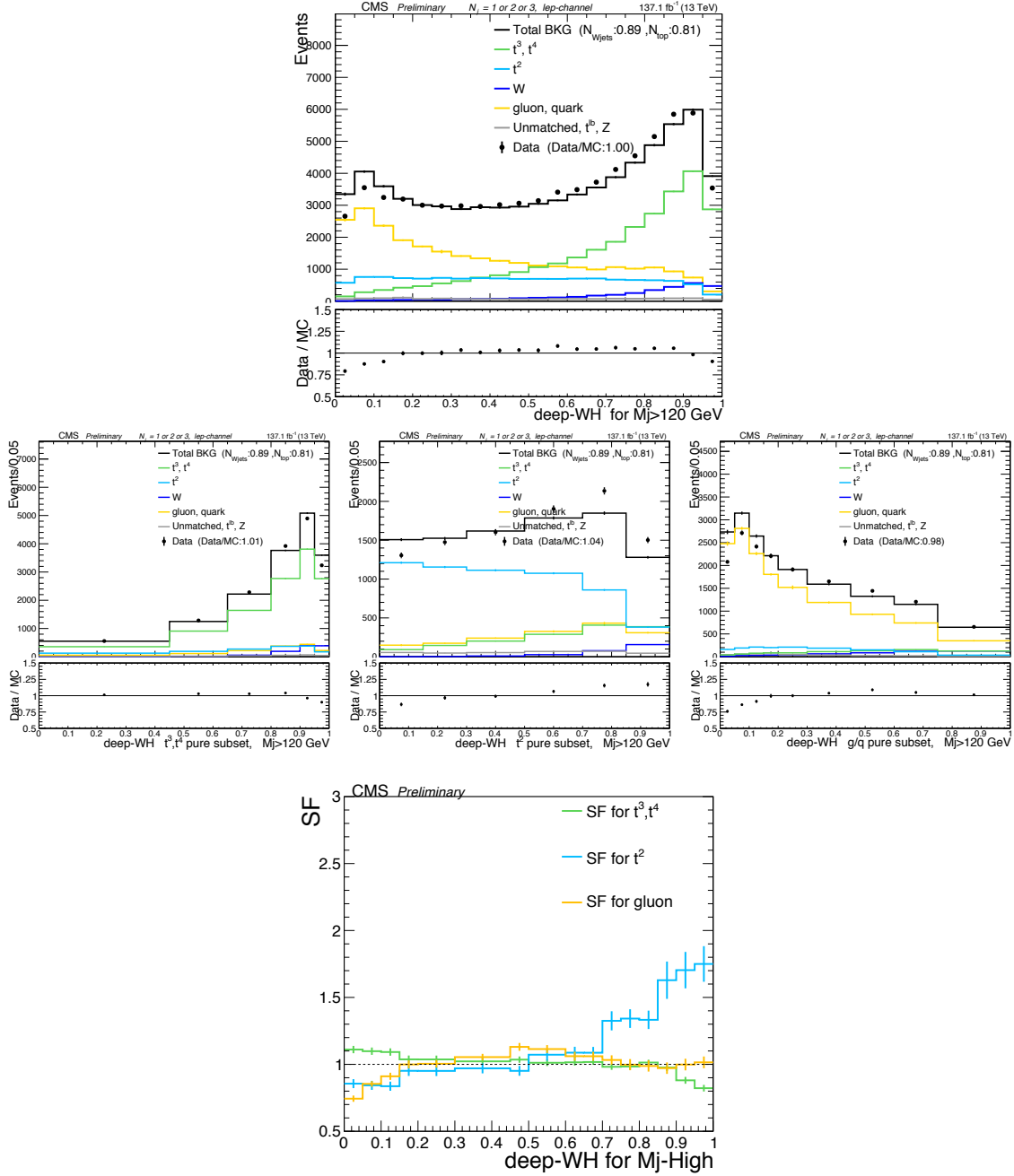


图 5.6 deep-WH 在 $M_j > 120 \text{ GeV}$ 时的分布 (上图), 分解为 3 个样本后的分布 (中) 和最终的 SFs 分布 (下)。

4. 方程中三个方程对应 3 个子样本, 尽管它们的统计量不同, 在处理时同等对待, 因此采用前文提到的合理的再分区从而较小误差。
5. 如前文提到的, 首先利用 b-tagged 喷注数量将 WJets、top 分布乘缩放因子修正到与数据匹配, 即利用 0 个和 ≥ 1 b 喷注时的两种本底事例数得到 2×2 矩阵从而求解修正因子, 高、低质量区的因子是分布计算的, 这样最终得到的关于 deepAK8 标记的 SFs 不会影响总本底的事例数, 只会修正形状。

6. 应用 SFs 后其他变量如 deep-t、deep-t/W、 τ_{ij} 也得到了一定的改善，因此该求解修正因子的矩阵方法对其他分析也有很强的参考意义。
7. 3 个子样本的定义方式是正交的，如果改变其中一个样本的定义条件，那么对应的事例会从一个子样本进入另一个子样本，同样会参与矩阵运算影响 SFs 的计算。
8. 与官方提供 SFs[75] 中等工作点的 SFs（见 3.5.4 章），即 deep-W 大于某个值时的修正因子，与我们的结果是一致的。出于数据驱动方法设计控制区的需求，分析需使用全范围的 deep-W（WH）的 SFs，因此使用了全新的方法来得到 SFs 关于 deep-W（WH）的全范围分布。

5.2.2 SFs 对横动量的依赖

由于横动量不同时，喷注的子结构距离等性质会有差异。在上文计算 SFs 时使用的 PS_ 主要由低横动量的喷注构成，在信号区 $S_T > 1\text{TeV}$ 的条件下包含了很多大横动量的喷注，因此决定考虑 SFs 对喷注横动量的相关性。最终将上文提到低、高 M_j 样本分为了低、高喷注横动量区。出于矩阵方法需要大统计量样本的要求，需要考虑两个分区分界的平衡，因此定义如下：对 M_j 在 60-120GeV，低（高）横动量区域为 200-400 (>400) GeV；对于 $M_j > 120\text{GeV}$ ，低（高）横动量区域为 200-500 (>500) GeV。

矩阵方法需要的子样本（W、 t^2 、g/q 主导区域）如图 5.7 前三列所示，最后一列为计算得到的 SFs，第一行到第四行依次为 (M_j, p_T)：低低、低高、高低、高高，可以发现相同对象的 SFs 曲线趋势相同但略有差异。

目前得到了 deep-W 在 M_j 取 60-120GeV 及 deep-WH 取 $M_j > 120\text{GeV}$ 时的 SFs。对于高质量区，W 的统计量很少，无法使用矩阵方法计算，因此用低质量区的结果近似代替。对于 deep-WH 在 60-120GeV 时的 SFs，使用同样的子样本定义及相同的矩阵方法进行计算。图 5.8 总结了分析使用的全部 SFs。表 5.2 为每个事例中不同类型喷注关于 SFs 的使用情况总结。

$M_j \square$	60-100	100-120	>120
W	$SF_L^{\text{deep-W}}(\text{W})$	$SF_L^{\text{deep-WH}}(\text{W})$	
$t^{3,4}$	$SF_H^{\text{deep-W}}(t^{3,4})$	$SF_H^{\text{deep-WH}}(t^{3,4})$	
g/q	$SF_L^{\text{deep-W}}(g/q)$	$SF_L^{\text{deep-WH}}(g/q)$	$SF_H^{\text{deep-WH}}(g/q)$
t^2	$SF_L^{\text{deep-W}}(t^2)$	$SF_L^{\text{deep-WH}}(t^2)$	$SF_H^{\text{deep-WH}}(t^2)$

表 5.2 S 应用于修正标准模型过程中每种类型的喷注的 SFs 总结。下标 L 或 H 指代低、高质量区，即 M_j 在 60-120GeV 或 $> 120\text{GeV}$ ，上标 deep-W 或 deep-WH 表示修正哪种 deepAK8 标记，在 M_j 小于（大于）100GeV 时使用 deep-W（WH），括号中代指不同的喷注类型。

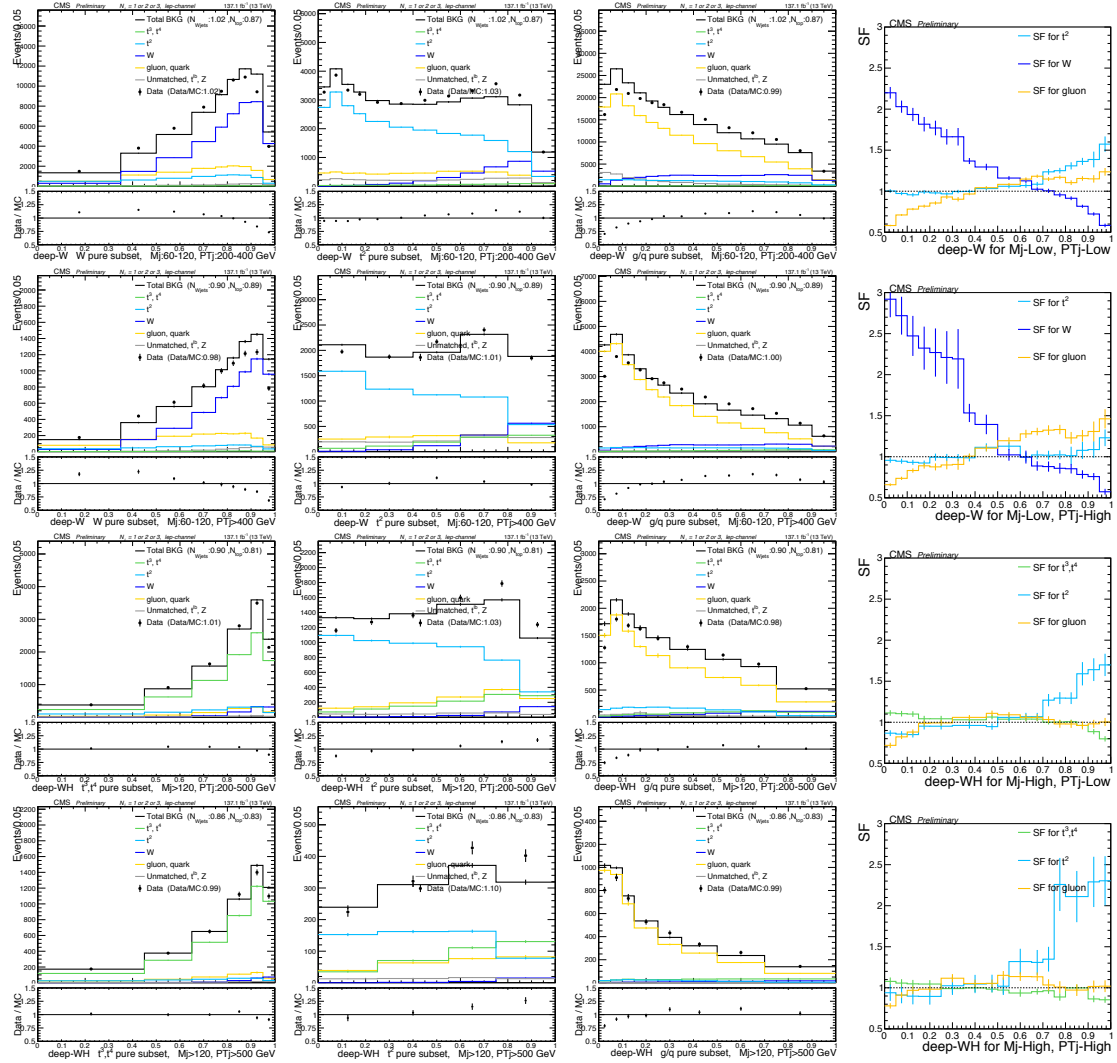


图 5.7 从左到右依次为 W、 t^2 、g/q 主导的子样本 deep-W (WH) 分布及最终的 SFs。从上到下行依次为 (M_j, p_T) : 低低、低高、高低、高高。

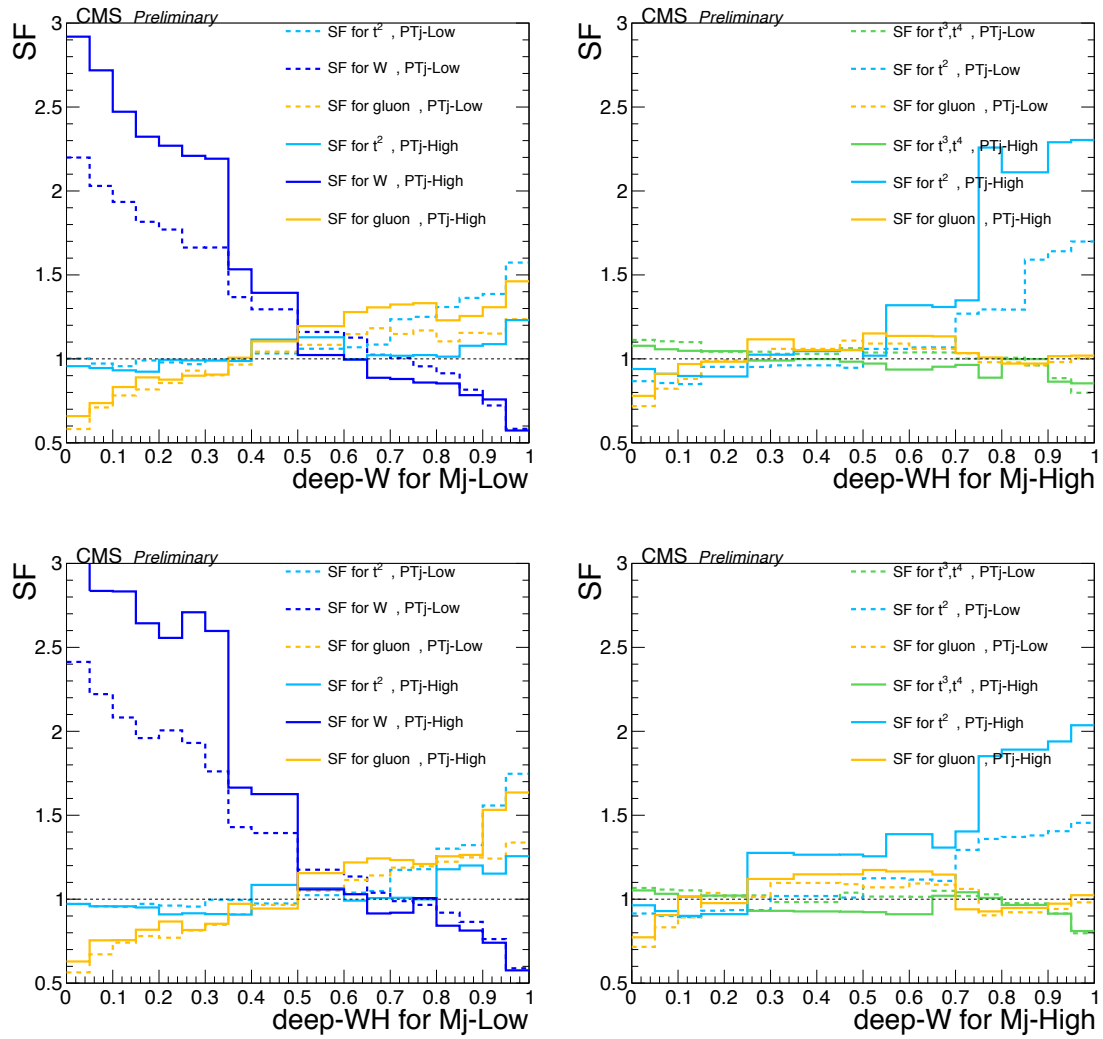


图 5.8 分析使用的全部 SFs, 左侧为低质量区 deep-W (WH)、右侧为高质量区 deep-W (WH) 的 SFs 分布。

5.3 对标准模型喷注使用 SFs 后的检验

在得到了关于 deepAK8 标记的 SFs，需要对蒙卡样本使用来验证矩阵方法是否可靠。

第一项测试是使用 SFs 后重新产生图5.3，即用相同的样本验证修正后的表现，结果如图5.9所示，可以发现蒙卡与数据吻合得极好，说明矩阵方法是有效的。

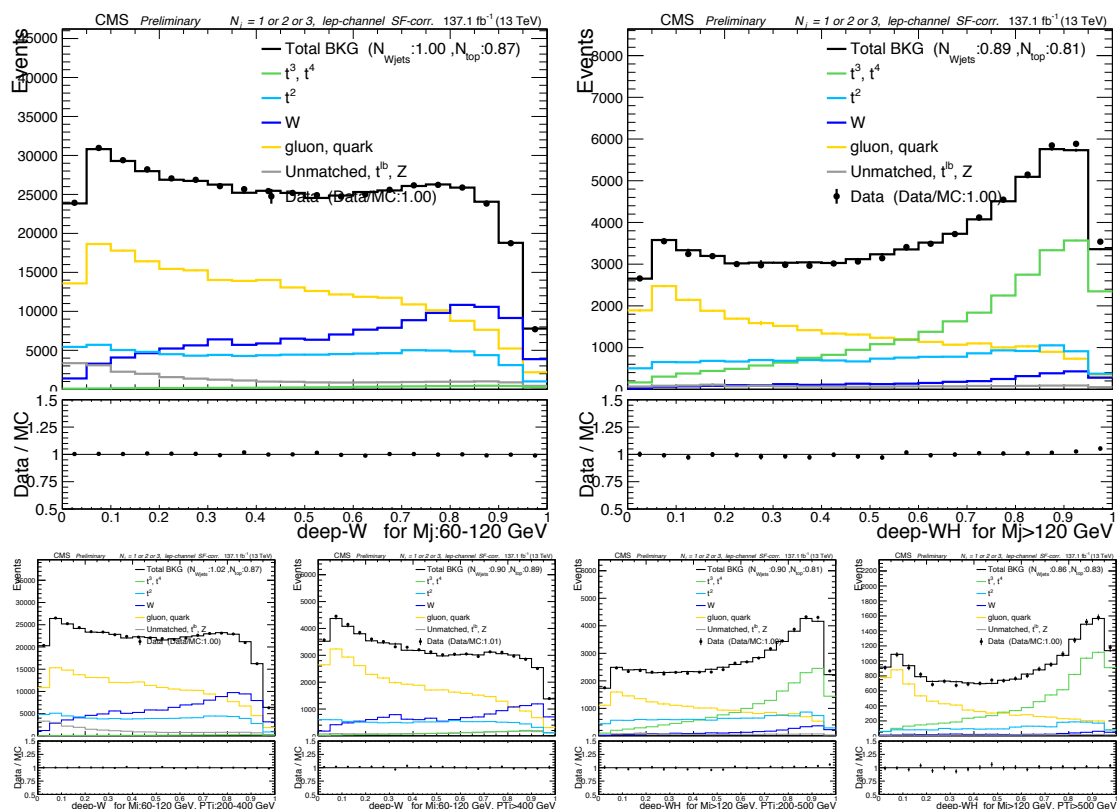


图 5.9 使用与图5.3相同的样本验证添加 SFs 修正之后的表现。

第二项测试是使用在低、高质量区定义的 6 个子样本在添加 SFs 后 p_T 取 200-300-400-500-600-...GeV 时的表现，结果如图 5.10和5.11所示，整体上数据与蒙卡相符合，在大横动量时会有较大的涨落出现，在可接受范围内。

第三项测试是用来检验最终分析相关的区域，即使用与分析区类似的条件，而非初选条件，这时候主要挑选与分析区类似的含大横动量喷注的事例，因此是至关重要的。检验包括以下几个部分：

- a. 将信号区 1-6 的样本中关于 b-tagged 喷注数目 0 个替换为 $N_b^M \geq 1$ 并且去除 deepAK8 标记条件，此时原有的分析区 4 和 5 将合并为 1 个样本，因此共有 5 个样本，这些样本将会被用作 top 的控制区，加 SFs 之前与之后的 deep-W (WH) 的表现如图5.12所示。

- b. 同样基于信号区 1-6 但是要求低能量 $650 < S_T < 950 \text{ GeV}$ 并且去除 deepAK8 标记条件得到 5 个样本，添加 SFs 前后对比如图5.13。
- c. 最后将信号区 1-6 的 deepAK8 标记条件反转，即只使用 deep-W(WH) 0-0.7(0.5) 区间，新的样本 4 和 5 同样会合并，最后 5 个样本修正前后对比如图5.14。

可以发现添加 SFs 修正后蒙卡样本与数据的差异会变小。所有的检测结果都很合理，证明了测得的 SFs 在所有区域内使用都是可靠的。

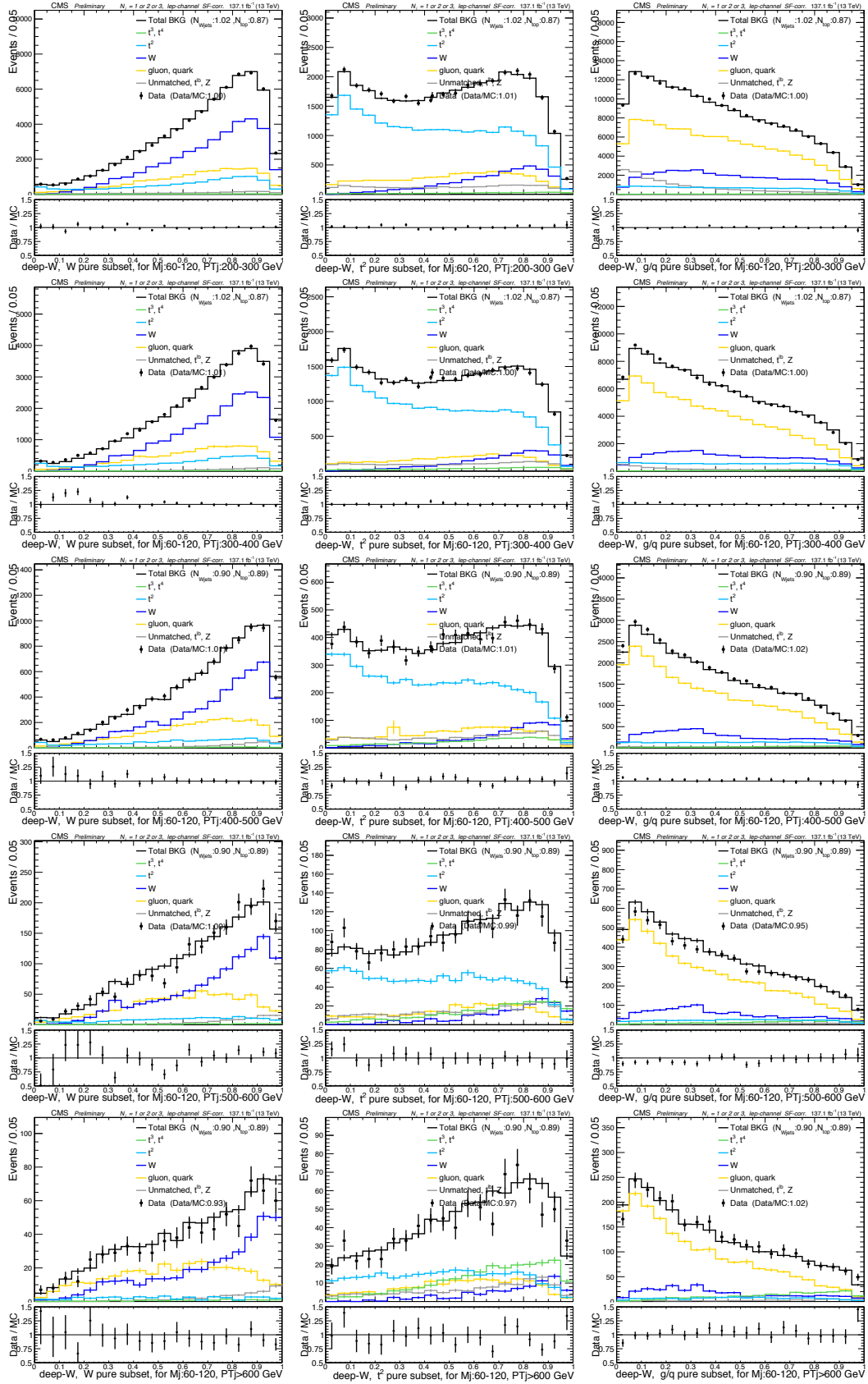


图 5.10 添加 SFs 修正后 deep-W 的表现, 从左到右依次为 W、 t^2 、g/q 主导子样本, 从上到下 p_T 依次取 200-300-400-500-600-...GeV。

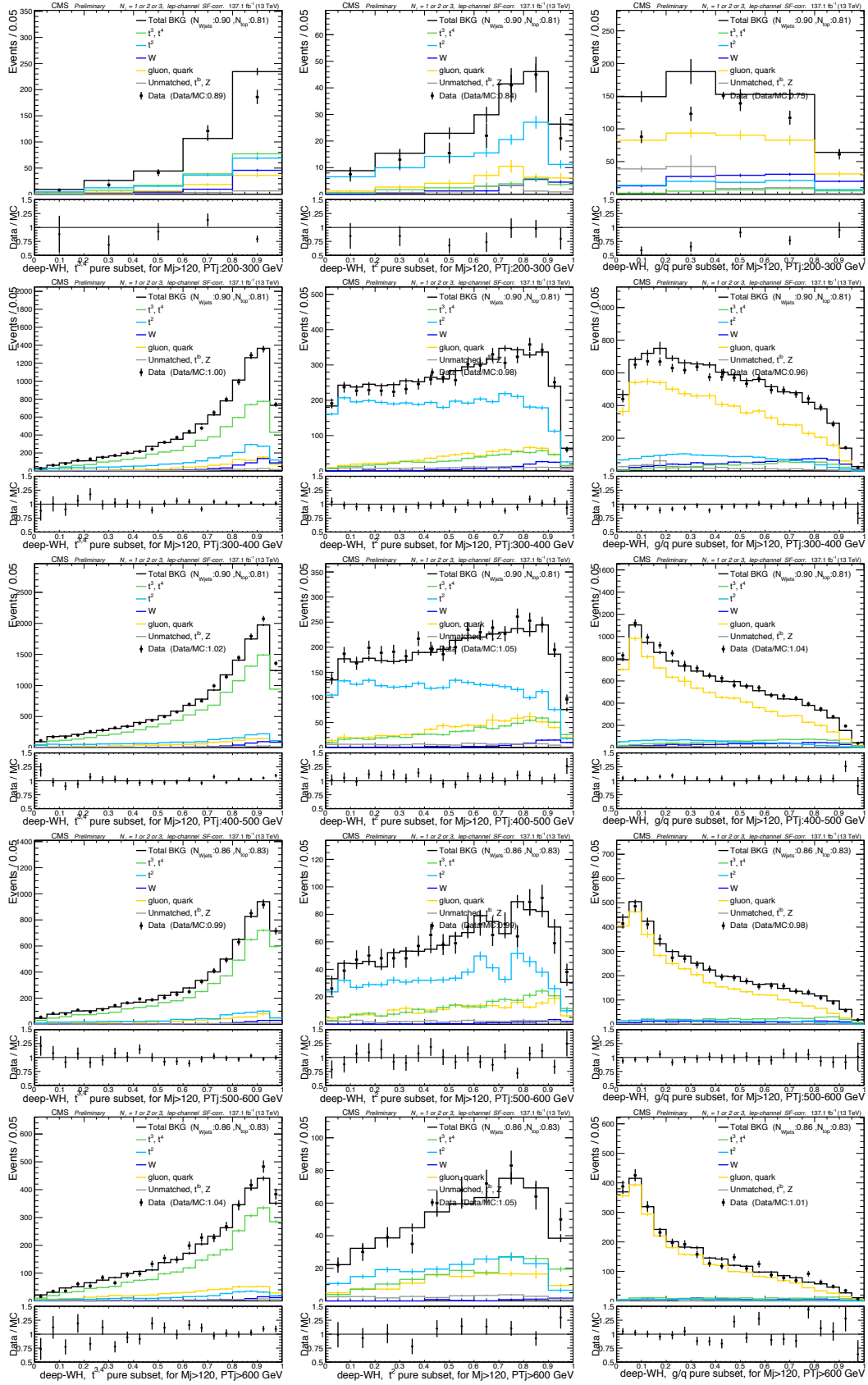


图 5.11 添加 SFs 修正后 deep-WH 的表现, 从左到右依次为 $t^{3,4}$ 、 t^2 、g/q 主导子样本, 从上到下 p_T 依次取 200-300-400-500-600-...GeV。

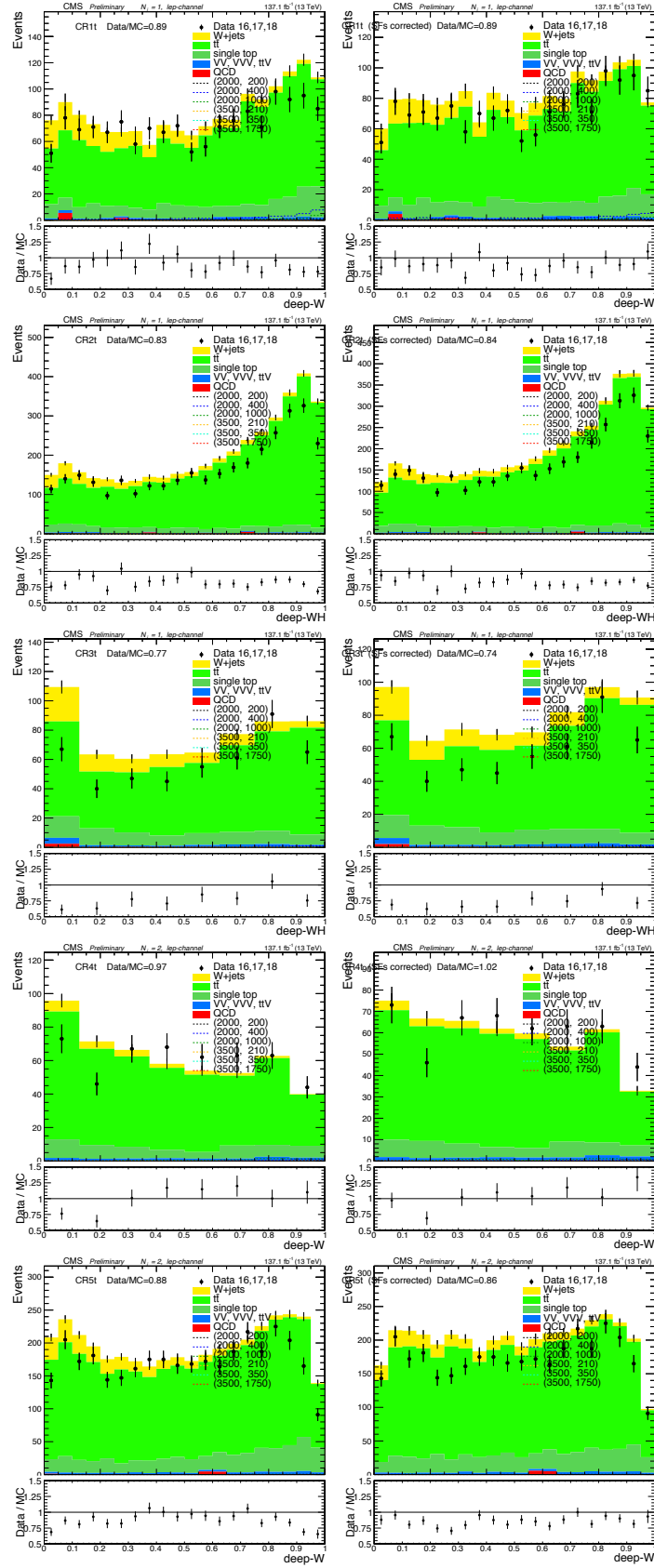


图 5.12 信号区 1-6 的样本中 $N_b^M = 0$ 替换为 $N_b^M \geq 1$ 并且去除 deepAK8 标记条件得到的 5 个样本 SFs 修正前后对比, 5 行对应信号区 1-6, 由于去除标记条件, 所以分区 4 与 5 合并。

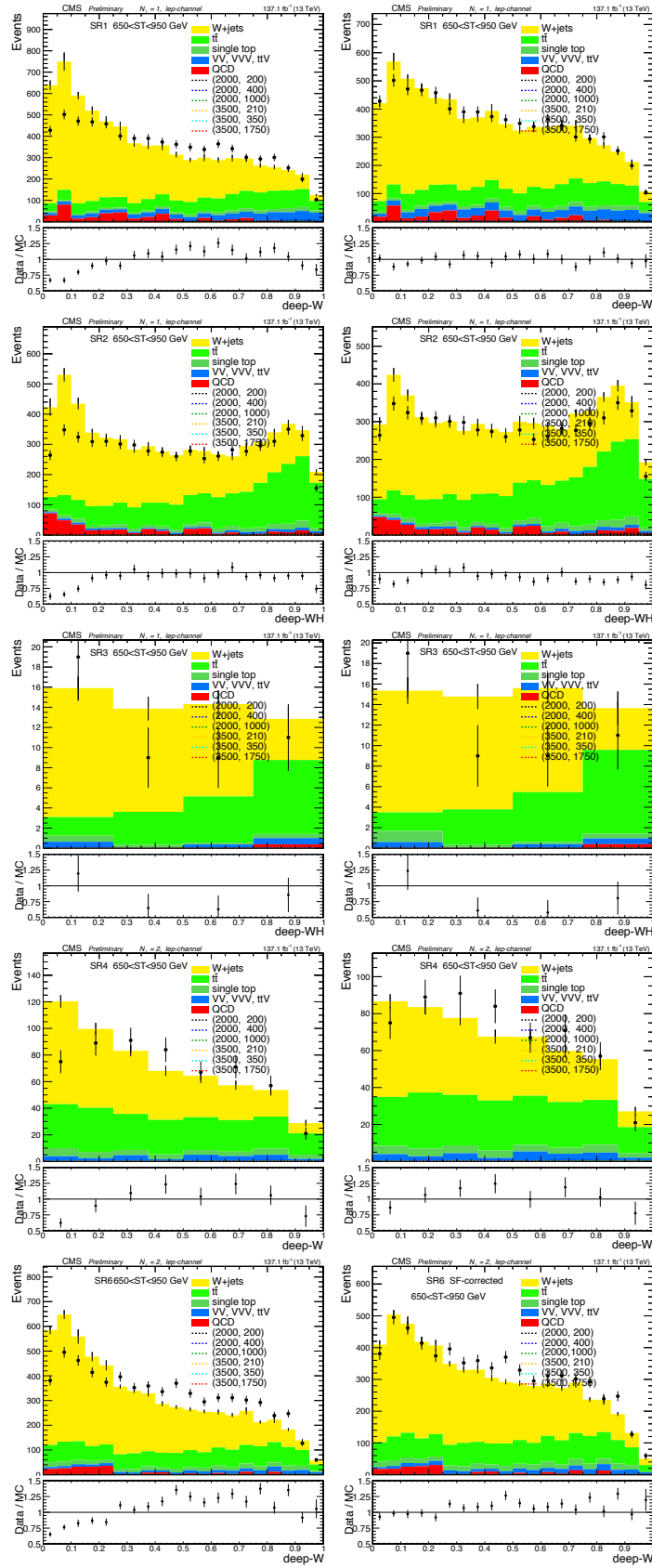


图 5.13 信号区 1-6 的样本中 $S_T > 1000\text{GeV}$ 替换为 $650 < S_T < 950\text{GeV}$ 并去除 deepAK8 标记条件得到的 5 个样本 SFs 修正前后对比, 5 行对应信号区 1-6, 由于去除标记条件, 所以分区 4 与 5 合并。

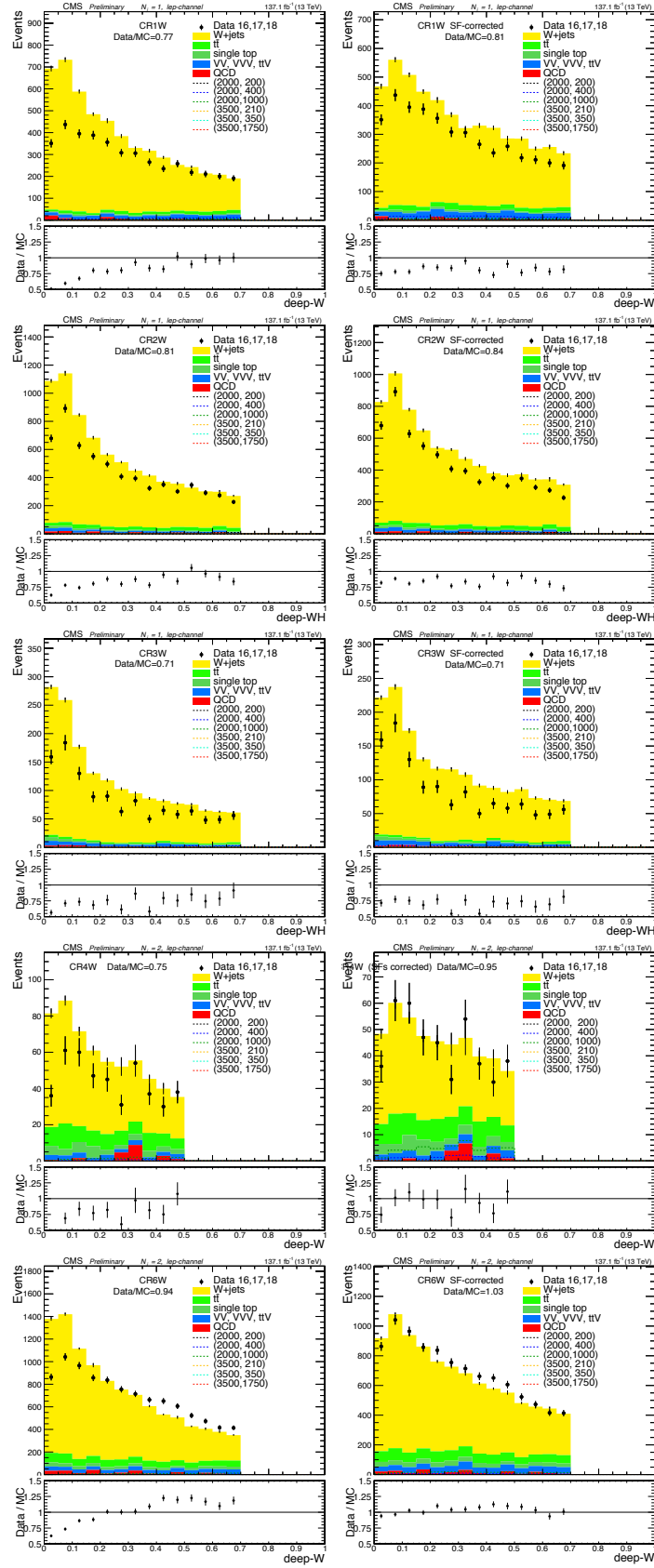


图 5.14 信号区 1-6 的样本中反转 deepAK8 标记条件得到的 5 个样本 SFs 修正前后对比, 5 行对应信号区 1-6, 分区 4 与 5 合并。

5.4 分立信号 SFs 的使用

对于分立 radion 情形的信号，末态重建的 W 喷注原则上和标准模型下的 W 喷注是一样的，二者仅在横动量上有区别，信号的横动量更高，意味着喷注的子结构更紧凑。我们已经得到了标准模型 W 喷注的 SFs，如图5.8上左所示，被记为 $SF_L(W)$ 。并且前文图5.10验证了所得的 SFs 在至少 $p_T \sim 650\text{GeV}$ 是可靠的。所以对于信号的 W 喷注可以使用与标准模型相同的 $SF_L(W)$ 来修正信号 deep-W 的表现并估计事例数。该选择有以下两项检测作为支撑：

1. 挑选富含 W 喷注的高横动量标准模型样本来作为信号 W 喷注的代理，比较与信号类似的 W 喷注与数据的 deep-W 表现。图5.15前两行中，将信号的 W 喷注归一到高横动量的标准模型本底的含 W 喷注的事例数，其中信号使用的去掉 deepAK8 标记的信号区 1 或 4，对于本底事例，采用 PS_ 并且要求喷注质量在 W 的质量窗内。第一行是不加 SFs 时本底喷注在不同横动量要求下与信号中 W 分布的对比，第二行是对应的加 SFs 后的对比情况，可以发现考虑 SFs 后数据与蒙卡形状相似。
2. 使用其他的 PS 比如 Herwig++ 来替代 Pythia8 信号样本来验证二者 deep-W 表现是否一致或者将差异作为系统误差。分析中使用的信号是经过 Pythia8 进行 PS，为了对比不同 PS 对信号过程的影响，额外产生了相同质量点的 Herwig++ 样本。图5.15最后两行中对于本底喷注横动量的要求外，还要求满足前文提到的 W 主导的子样本的条件来提高 W 喷注的纯度，图中 W++ 用来表示 Herwig++ 信号的 W 喷注，其分布与 Pythia8 信号 W 喷注对应的 deep-W 表现近乎一致。

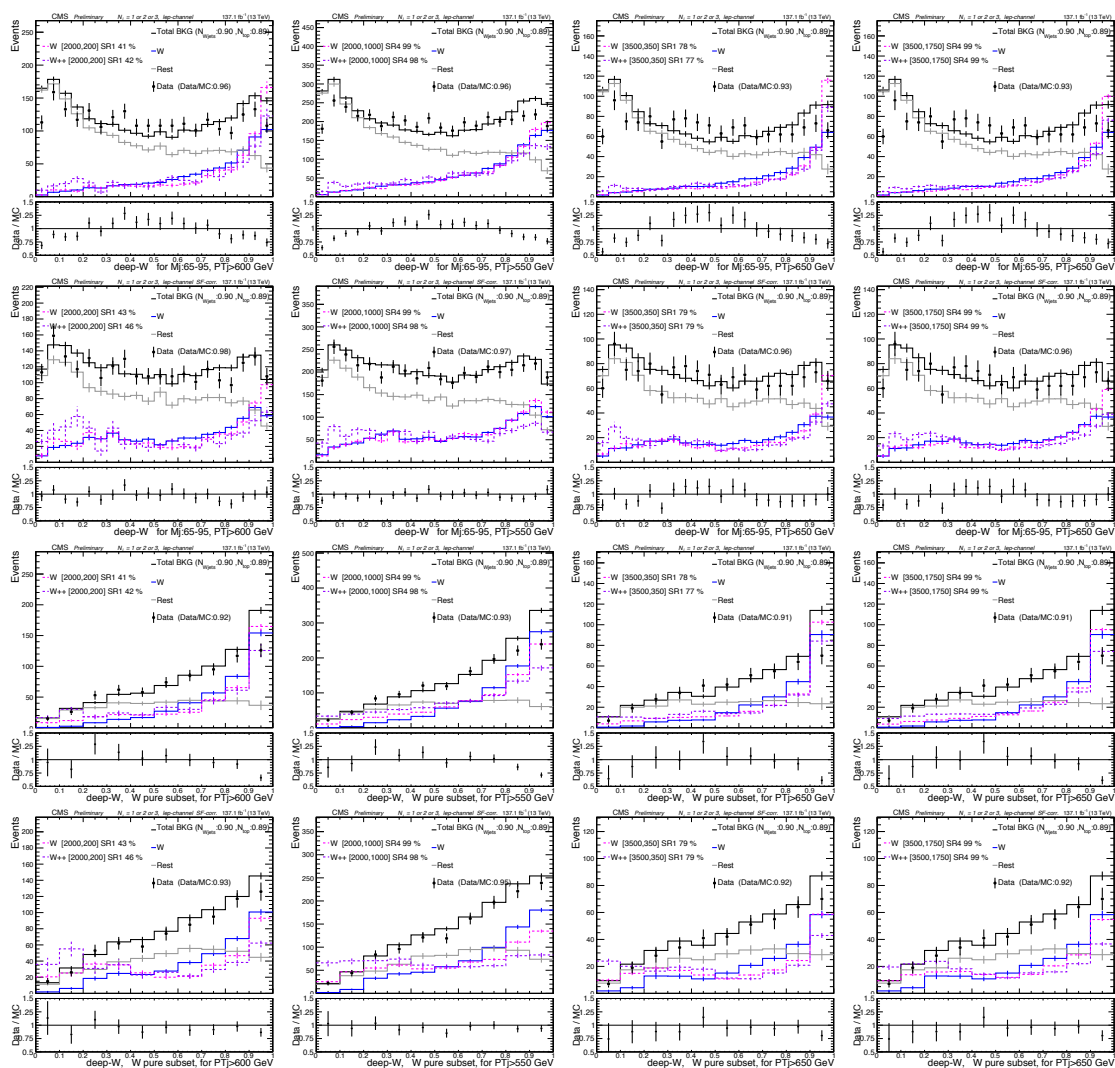


图 5.15 第一行是不加 SFs 时将信号的 W 归一到标准模型的 W 类喷注后二者 deep-W 形状的对比, 其中对于信号使用的去掉 deepAK8 标记的信号区 1 或 4 的条件, 本底要求通过初选条件 PS₋ 及各图中横轴名字要求的喷注 p_T 和 M_j 条件; 第二行与第一行相同, 但添加了 SFs。第三行是类似的 deep-W 分布, 但要求通过 W 主导子样本的选择条件且不加 SFs; 在第四行是添加 SFs 后的对比。图例中比如第一张图 [2000,200]GeV 表示两个共振态的质量, 即 $[W_{KK}, M_R]$ 的格式。

5.5 信号合并 radion 喷注对应的 SFs

如果两个共振态的质量比 $M_R/W_{KK} \lesssim 0.2$, radion 衰变的 2 个 W 会形成一个合并的喷注。真实的喷注结果是复杂的, 包括以下几种情况: $R \rightarrow qq\bar{q}\bar{q}$ 、 $\rightarrow qq\ell\nu$ 、 $\rightarrow qq\tau_h\nu$, 另外在每种情况下还会有一个或多个粒子在重建喷注的张角之外。图5.16展示了文章对不同 radion 喷注结构及命名。radion 衰变的 2 个 W 可能全强衰变, 约占 60%, 在图片第一列, 剩余半轻过程在第二列, 每列前直接从 W_{KK} 衰变得到的 W (prompt W) 与 radion 是背对背结构, 每个图例的左侧是 W, 右侧为 radion 结构, 每列的顺序大致是按占比排序, 对于大部分信号点, 主要的成分是 R^{4q} 、 R^{3q} 和 R^{lqq} 。图中标有红色叉号表示该情况很稀有或者不在挑选条件的接收范围内, 可以在分析时忽略其影响。

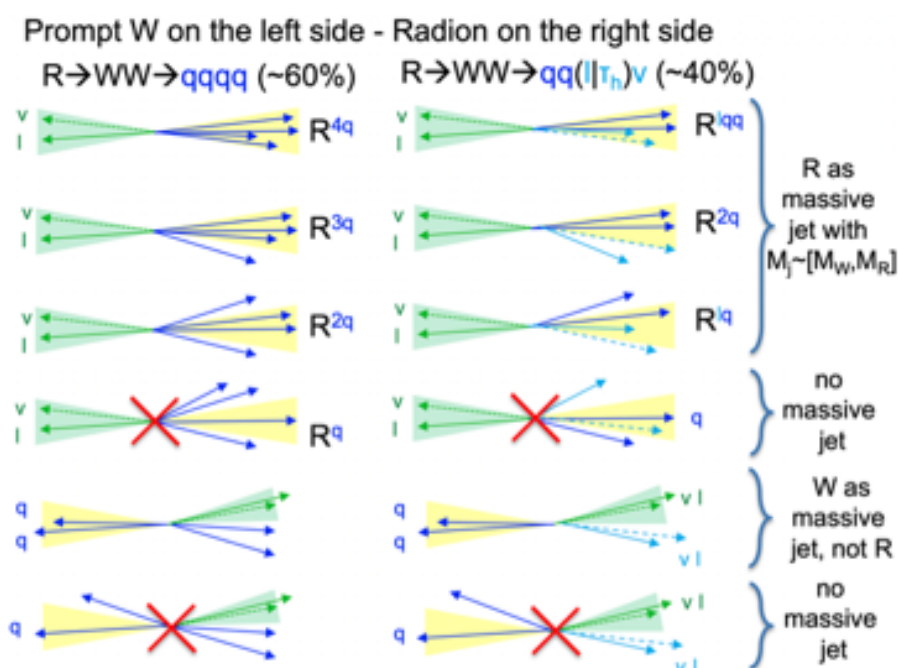


图 5.16 合并 radion 情形下可能存在的末态, 要求只是有一个 W 强子衰变。黄色三角形表示 AK8 喷注, 绿色三角形表示轻子衰变的 W, 黑色、绿色实、绿色虚箭头分别表示夸克、轻子、中微子。每种情形的标记中 R 表示 radion, 上标为 AK8 喷注含有的粒子数量及类型, 比如 R^{4q} 表示 AK8 喷注由 4 个夸克组成。

由于这些多样的 radion 结构的贡献, 最终得到的喷注质量谱是一个展宽的并且复杂的分布。不同的成分会在信号区 1-3 有不同的分布和贡献。为了研究不同的信号成分, 利用产生子水平的信息制定了相应的匹配条件来加以鉴别, 归纳为表5.3。

之后挑选 6 个有代表性的信号点, 将喷注 M_J 、观测量 $M_{J\ell\nu}$ 和 deep-W(WH) 按不同的成分分解, 如图5.17所示, 使用的选择条件为去掉标记条件后的信号区 1-3。图中所有信号点都被归一到数据亮度 137fb^{-1} , 括号内为每种成分的占比, 可以发现 R^{4q} 、 R^{3q} 、 R^{lqq} 与 W 是信号的主要过程, 约占总事例数的 95%。W 和 R^{lqq} 主要分布于 M_J

信号成分	R^{4q}	R^{3q}	R^{2q}	$R^{\ell qq}$	$R^{\ell q}$
ΔR	$(R, j) < 0.6$	$(R, j) < 0.6$	$(R, J) < 0.6$	$(R, j) < 0.6$	$(R, j) < 0.6$
ΔR	$(q_1, j) < 0.8$	$(q_1, j) < 0.8$	$(q_1, J) < 0.8$	$(q_1, j) < 0.8$	$(q_1, j) < 0.8$
ΔR	$(q_2, j) < 0.8$	$(q_2, j) < 0.8$	$(q_2, J) < 0.8$	$(q_2, j) < 0.8$	$(q_2, j) > 0.8$
ΔR	$(q_3, j) < 0.8$	$(q_3, j) < 0.8$	$(q_3, J) > 0.8$	$(\ell, j) < 0.8$	$(\ell, j) < 0.8$
ΔR	$(q_4, j) < 0.8$	$(q_4, j) > 0.8$	$(q_4, J) > 0.8$	-	-

表 5.3 图5.16中介绍的几种合并喷注： R^{4q} 、 R^{3q} 、 R^{2q} 、 $R^{\ell qq}$ 、 $R^{\ell q}$ 与产生子水平粒子的匹配条件，同一列中的条件要满足“与”的关系，基本原则与标准模型过程喷注的匹配要求类似。除这几种成分之外，其他部分占比非常少，后续用“剩余“(或 rest) 来代替。

在 [60,100]GeV 区域，即信号区 1，而且二者的 deep-W 形状接近。 R^{4q} 和 R^{3q} 主要集中在 $M_j > 100\text{GeV}$ ，即信号区 2 和 3，并且它们的 deep-WH 分布相似。另外可以发现尽管不同信号点 radion 的质量变化很大，从 180GeV 到 350GeV，deep-W (WH) 的形状都很接近，证明了分析使用的标记确实满足质量去相关的要求。

同样，对于每一个重要的成分都需要用数据驱动的方法来计算相应的修正因子。有一个难题是标准模型过程中并没有类似信号的 4 子结构 (4-prong) 的对象 (类似 R^{4q}) 或者由 2 个夸克和 1 个轻子 (类似 $R^{\ell qq}$) 组成的喷注来作为信号的替代。可能的过程 $H \rightarrow qq\bar{q}\bar{q}$ 或 $H \rightarrow qq\ell\bar{\nu}$ 过程中 Higgs 喷注与 radion 信号类似，但是并没有足够的统计量来推导出相应的 SFs。还有一种方法是将 2 个常规的 W 喷注通过旋转的方式使它们靠近作为含 4 个夸克喷注的代理，但是对于信号来说，每个 W 的能量很高，为 TeV 量级，此时末态 2 个夸克的距离极小，此时使用探测器得到的粒子流物理量旋转重合的方法不一定可靠。最终分析选用的最接近信号的全合并的 top 喷注来作为 R^{3q} 和 R^{4q} 代理，做为替代要求能用标准模型的物理量模仿信号中的结构，用到的代理对象如下：

- 可以形象成为 3 子喷注 (3-prong) 的强子衰变的全合并 t^3 ，即 $t \rightarrow qq\bar{b}$;
- 在 t^3 的基础上额外加一个夸克或者胶子组成四子喷注结构的 t^4 ，即 $t \rightarrow qq\bar{b} + g/q$ 。

图5.18左侧为合并 radion 的主要成分结构示意，右侧为接近的标准模型过程，所有标准模型对象比如 t^3 和 t^4 都与前文表5.1的匹配条件一致。

图5.19，同样使用相同的 6 个信号的，将不同的成分归一到 1 从而更方便比较形状。可以发现 $R^{\ell qq}$ 与 W 的表现相似，这是因为 deepAK8 在训练时并不会特殊处理喷注中含有轻子的情况，这导致视 $R^{\ell qq}$ 与 W 为相同的物理对象，因此对这两种成分可以使用相同的 SFs (W); 另外， R^{4q} 和 R^{3q} 也非常接近，因此可以同一个代理对象，即 t^3 和 t^4 ，使用 SFs ($t^{3,4}$) 来修正。图5.20与5.21为信号 $R^{3q}R^{4q}$ 与本底 $t^{3,4}$ 的 deep-WH 与 deep-W 对比，三者形状非常相似，因此使用 $t^{3,4}$ 作为代理对象是合理的。

表5.4总结了合并 radion 信号不同成分在不同质量区间 SFs 的应用说明。

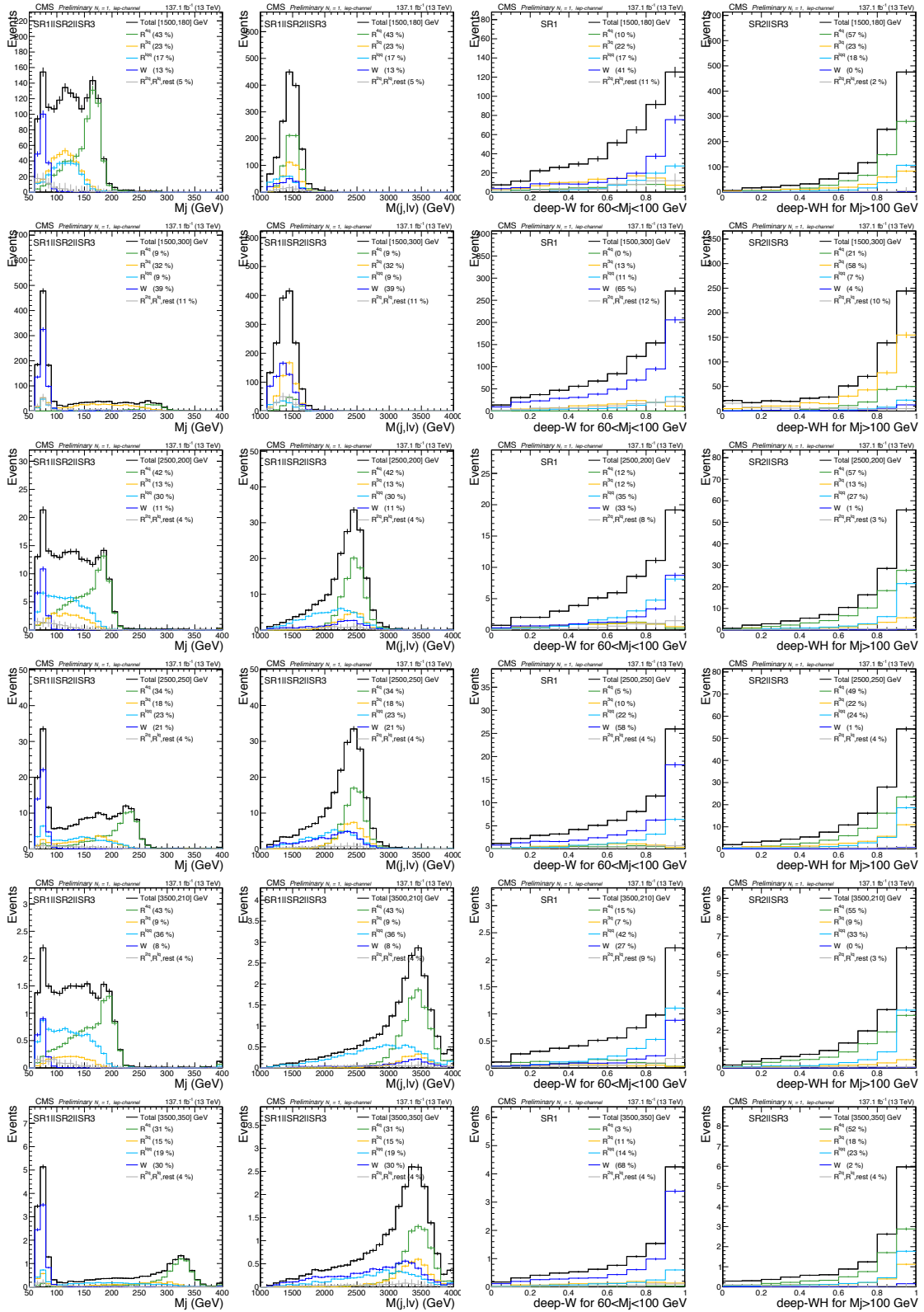


图 5.17 6 个代表性的合并 radion 信号点 (对应 6 行) 按成分分解图, 从左到右分别为 M_j 、 $M_{j,l\nu}$ 、 $60 < M_j < 100$ GeV 时 deep-W 和 $M_j > 100$ GeV 时 deep-WH 在去掉标记条件后的信号区 1、2 和 3 下的分布。

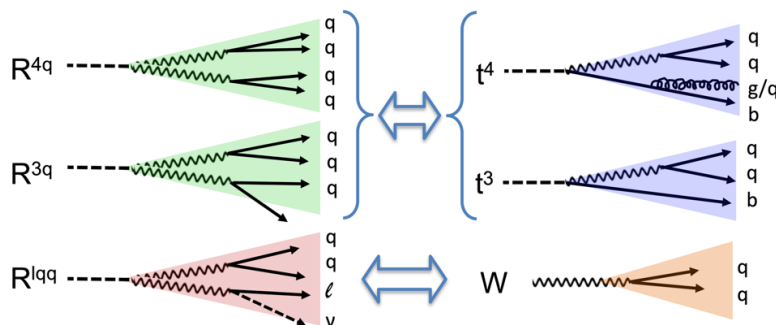


图 5.18 信号中合并 radion 主要成分（左）与对应的标准模型对象（右）的喷注结构的对比。

M_j [GeV]	60-100	100-120 > 120
$W \leftrightarrow R^{\ell qq}$	$SF_{\text{H}}^{\text{deep-W}}(W)$	$SF_{\text{H}}^{\text{deep-WH}}(W)$
$t^{3,4} \leftrightarrow R^{3q}, R^{4q}$	$SF_{\text{H}}^{\text{deep-W}}(t^{3,4})$	$SF_{\text{H}}^{\text{deep-WH}}(t^{3,4})$

表 5.4 信号合并 radion 喷注与标准模型过程对应情况及 SFs 应用。

5.6 关于 SFs 的系统误差

对于信号和本底中不同类型的喷注考虑了以下系统误差：

5.6.1 PS 和产生子误差

分析以及计算 SFs 时使用产生子 Powheg 和部分子簇射（PS）为 Pythia8 的 $\bar{t}t$ 样本，考虑到不同产生子及 PS 在蒙卡模拟时的差异，使用了 2 个替换的 $\bar{t}t$ 样本，分别为 "Madgraph+Pythia8" 和 "Powheg+Herwig7"。通过替换默认的 $\bar{t}t$ 样本可以改变 W 和 top 喷注关于 deep-W（WH）形状，重复 5.2 中的矩阵方法，可以得到新的 2 套 SFs。在一共三套 SFs，选用最大差值作为对称的双向误差范围，图 5.22 中误差带是它们关于 SFs 的 PS 系统误差。

5.6.2 子样本挑选条件带来的偏差

在矩阵方法中定义了三个子样本并使用其中的事例数组成矩阵，在使用不同子样本挑选条件定义时结果会有不同，而且对于同一物理对象认为在不同子样本中修正因子相同，而实际可能不同，这种处理方式会带来一定的偏差，在附录 B.2 中有详细的推导过程。在分析中，首先将 Powheg+Pythia8 的 $\bar{t}t$ 样本替换为 Powheg+Herwig7 的 $\bar{t}t$ ，其余本底过程保持不变，然后用此时的总本底作为假数据。这样可以将每个子样本中假数据与蒙卡的比例作为每个子样本中真实的 SFs，即附录中的 a_i 、 b_i 、 c_i 。用假数据替换真实数据后，重复同样的矩阵方法可以得到真实的 SFs，与测量值的比较如图 5.23。从图中第二行的对比可以观测到最大为 10% 的偏差，除了 $t^2 \sim 100\%$ ，后者是由于其在

对应主导子样本中占比较低导致的。因此对于表5.2中所有的 SFs 除了 t^2 都赋予 10% 的不确定度作为偏离误差，其中 t^2 为 100%。

5.6.3 代理对象误差

前文5.5中使用标准模型 W 作为信号 W 和 R^{lqq} 的代理对象， $t^{3,4}$ 作为 R^{4q} 和 R^{3q} 的代理对象。信号与对应代理对象喷注结构相似，且 deep- W (WH) 有相似的形状，不过也有一些不同之处，比如夸克味道的差异、内部夸克动力学量的差异、硬夸克与软胶子的差异等，因此用信号和代理对象 deep- W (WH) 在 0.7 以上区间的差别作为代理误差来估计这些难以直接处理的差异。如前文图5.20和5.20所示，分别将 R^{4q} 、 R^{3q} 归一到 $t^{3,4}$ 的事例数，计算 deepAK8 标记值大于 0.7 (即最后 6 个分区) 事例数的百分比差异作为代理误差 (proxy uncertainty)。具体公式如下：

$$Proxy - unc. = \sqrt{\left(\frac{\sum_i |t_i^3 - t_i^4|}{\sum_i t_i^4}\right)^2 + \left(\frac{\sum_i |R_i^{3q|4q} - t_i^{3,4}|}{\sum_i t_i^{3,4}}\right)^2}$$

其中 t_i^3 、 t_i^4 、 $t_i^{3,4}$ 、 $R_i^{3q|4q}$ 表示相应的喷注在第 i 个分区内的事例数， i 循环最后 6 个分区。式中第一项表示 t^3 和 t^4 的差异 (10% 或 5%，见附录B.3)，第二项为 R^{4q} 、 $R^{3q}t^{3,4}$ 的差异，这一项的误差在 10%-40%，是代理误差的主要部分。

信号 R^{lqq} 使用 W 代理的，图5.24和5.25为信号 R^{lqq} 成分与来自信号、本底的 W 类喷注的对比。这种情况是代理误差是选用 R^{lqq} 和来自信号的 W 的差异，因为它们有着类似的横动量，对于 R^{lqq} 的代理误差可表示为： $\frac{\sum_i |R_i^{lqq} - W_i|}{\sum_i W_i}$ ， i 同样循环最后 6 个分区。对于其他类型的信号成分 (主要是未匹配部分)，赋予 100% 的不确定度作为代理误差。

因此对于每个信号点及每种 radion 喷注类型都需要独立计算其代理误差，表5.5总结了 6 个信号点的来自信号区 2 和 3 的事例数、等效的 SFs 及各项系统误差，本节中讨论的代理误差为表中最后一列中“代理”项。在计算代理误差的图中 deep- W (WH) >0.7 的区间加入新考虑的代理误差带，可以发现误差带可以有效地覆盖信号喷注与代理对象的差异，见附录B.4。

5.6.4 信号高横动量外推误差

本底样本中用以计算的 SFs 的喷注横动量绝大部分小于 650GeV，而大部分信号喷注的横动量大于这个值，因此采用不同 PS 作为估计这项影响的误差。通过对比不同 PS 如 Herwig++ 和 Pythia8 产生的信号样本，将不同喷注类型在不同信号区两种样本中事例数的相对差值作为高横动量外推误差。对于 W 、 R^{lqq} 、 R^{3q} 和 R^{4q} 外推误差分别为

~15-25%、~0-40%、~0-15% 和 ~0-15%。

5.6.5 信号关于 SFs 误差总结

表5.5是 6 个信号点在信号区 2 和 3 每项误差来源的贡献总结。6 个信号点添加 SFs 后在信号区 2 和 3 的总事例数缩减到 ~0.85，相对误差在 $\pm 25\text{-}40\%$ 之间。

$M_{W_{KK}}, M_R$ (GeV),(GeV)	喷注成分 比例	事例数: SFs 修正后 [未修正]	$SFs \pm$ 偏移 $\pm$ PS $\pm$ 外推 $\pm$ 替代 [% Total]
1500, 180	R^{4q} 61%	449.3 [494.4]	$0.91 \pm 0.09 \pm 0.07 \pm 0.00 \pm 0.41$ [47%]
	R^{3q} 21%	159.8 [173.9]	$0.92 \pm 0.09 \pm 0.08 \pm 0.04 \pm 0.18$ [24%]
	R^{lqq} 15%	113.2 [159.7]	$0.71 \pm 0.07 \pm 0.15 \pm 0.04 \pm 0.25$ [43%]
	W 0%	2.3 [3.3]	$0.69 \pm 0.07 \pm 0.15 \pm 0.44$ - [67%]
	Rest 1%	9.1 [9.1]	1.00 - - $\pm 0.17 \pm 1.00$ [101%]
	合计	733.8 [840.5]	$0.87 \pm 0.09 \pm 0.09 \pm 0.01 \pm 0.34$ [42%]
1500, 300	R^{4q} 25%	103.9 [112.4]	$0.92 \pm 0.09 \pm 0.07 \pm 0.16 \pm 0.37$ [45%]
	R^{3q} 61%	250.2 [275.6]	$0.91 \pm 0.09 \pm 0.08 \pm 0.04 \pm 0.26$ [32%]
	R^{lqq} 6%	24.5 [34.3]	$0.71 \pm 0.07 \pm 0.15 \pm 0.20 \pm 0.22$ [48%]
	W 3%	13.8 [19.4]	$0.71 \pm 0.07 \pm 0.15 \pm 0.32$ - [50%]
	Rest 2%	11.8 [11.8]	1.00 - - $\pm 0.11 \pm 1.00$ [101%]
	合计	404.2 [453.5]	$0.89 \pm 0.09 \pm 0.08 \pm 0.06 \pm 0.30$ [37%]
2500, 200	R^{4q} 60%	51.5 [56.2]	$0.92 \pm 0.09 \pm 0.07 \pm 0.10 \pm 0.21$ [29%]
	R^{3q} 12%	10.8 [11.9]	$0.91 \pm 0.09 \pm 0.08 \pm 0.02 \pm 0.15$ [21%]
	R^{lqq} 25%	21.4 [30.7]	$0.70 \pm 0.07 \pm 0.15 \pm 0.04 \pm 0.19$ [37%]
	W 0%	0.3 [0.4]	$0.72 \pm 0.07 \pm 0.15 \pm 0.37$ - [57%]
	Rest 1%	1.5 [1.5]	1.00 - - $\pm 0.12 \pm 1.00$ [101%]
	合计	85.5 [100.6]	$0.85 \pm 0.09 \pm 0.10 \pm 0.07 \pm 0.21$ [30%]
2500, 250	R^{4q} 53%	45.0 [49.0]	$0.92 \pm 0.09 \pm 0.08 \pm 0.07 \pm 0.20$ [27%]
	R^{3q} 21%	18.3 [20.1]	$0.91 \pm 0.09 \pm 0.08 \pm 0.09 \pm 0.20$ [28%]
	R^{lqq} 21%	18.5 [26.6]	$0.69 \pm 0.07 \pm 0.15 \pm 0.11 \pm 0.19$ [39%]
	W 0%	0.7 [1.0]	$0.68 \pm 0.07 \pm 0.15 \pm 0.16$ - [34%]
	Rest 1%	1.6 [1.6]	1.00 - - $\pm 0.04 \pm 1.00$ [100%]
	合计	84.0 [98.4]	$0.85 \pm 0.09 \pm 0.09 \pm 0.09 \pm 0.21$ [31%]
3500, 210	R^{4q} 57%	5.5 [6.0]	$0.92 \pm 0.09 \pm 0.08 \pm 0.00 \pm 0.12$ [19%]
	R^{3q} 8%	0.8 [0.9]	$0.91 \pm 0.09 \pm 0.08 \pm 0.03 \pm 0.14$ [21%]
	R^{lqq} 31%	3.0 [4.4]	$0.69 \pm 0.07 \pm 0.15 \pm 0.05 \pm 0.12$ [31%]
	W 0%	0.0 [0.0]	$0.68 \pm 0.07 \pm 0.14 \pm 0.20$ - [38%]
	Rest 1%	0.2 [0.2]	1.00 - - $\pm 0.11 \pm 1.00$ [101%]
	合计	9.5 [11.4]	$0.83 \pm 0.08 \pm 0.10 \pm 0.02 \pm 0.14$ [23%]
3500, 350	R^{4q} 58%	4.9 [5.3]	$0.91 \pm 0.09 \pm 0.07 \pm 0.05 \pm 0.33$ [39%]
	R^{3q} 18%	1.6 [1.7]	$0.90 \pm 0.09 \pm 0.07 \pm 0.04 \pm 0.31$ [37%]
	R^{lqq} 20%	1.7 [2.5]	$0.68 \pm 0.07 \pm 0.15 \pm 0.01 \pm 0.15$ [32%]
	W 1%	0.1 [0.2]	$0.67 \pm 0.07 \pm 0.15 \pm 0.47$ - [75%]
	Rest 1%	0.1 [0.1]	1.00 - - $\pm 0.04 \pm 1.00$ [100%]
	合计	8.4 [9.8]	$0.85 \pm 0.08 \pm 0.09 \pm 0.04 \pm 0.29$ [38%]

表 5.5 6 个信号点在信号区 2 和 3 的总事例数按不同喷注类型分解后的事例数，最后一列给出了有效 SFs 和各误差项。

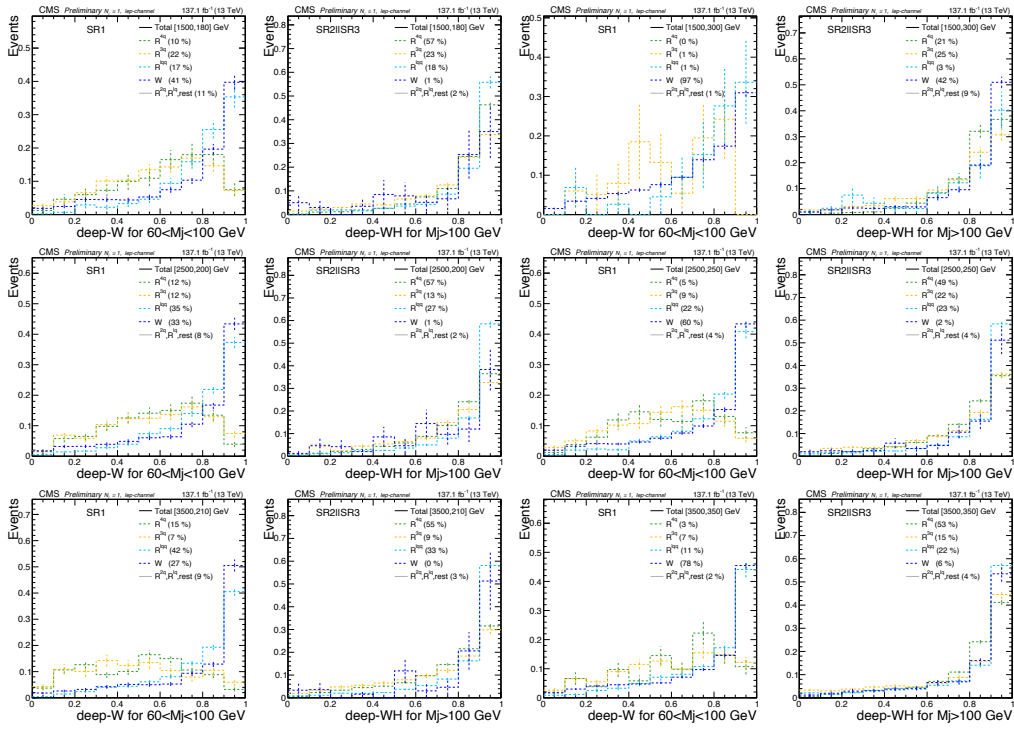


图 5.19 6 个代表性的合并 radion 信号点按成分分解后 M_j 在 $[60, 100]$ GeV 时 deep-W 与 $M_j > 100$ GeV 时 deep-WH 的分布，为了方便对比，所有直方图都被归一到 1。

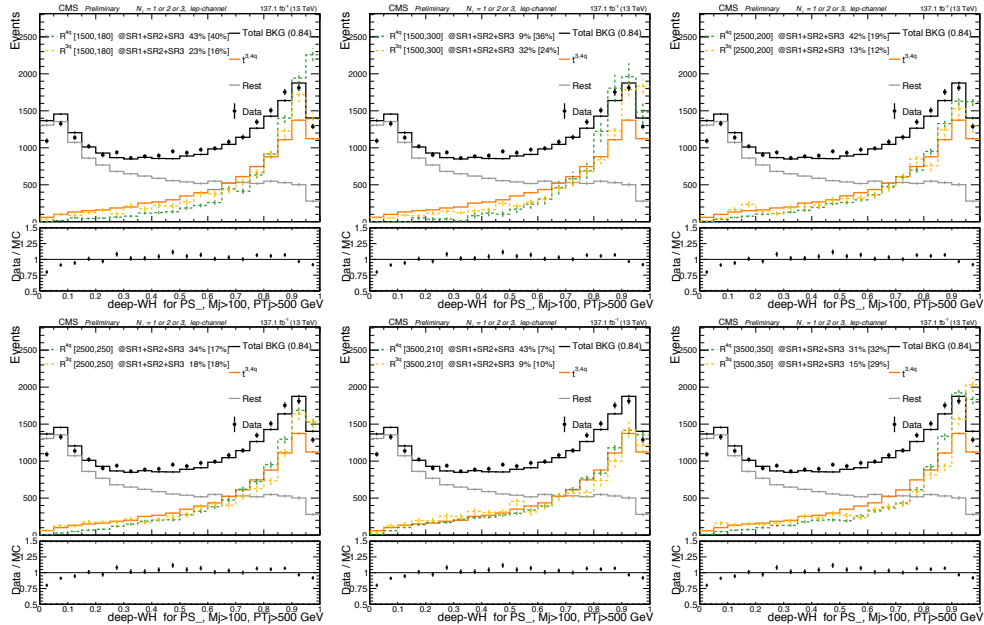


图 5.20 6 个信号点的 R^{3q} 、 R^{4q} 与本底 $t^{3,4}$ 的 deep-WH 分布对比， R^{3q} 、 R^{4q} 分别归一到 $t^{3,4}$ 上，本底喷注的条件如 x 轴标题所示，信号调节为去除 deep-W (WH) 条件的信号区 1、2、3，图例中第一个数字是该成分在信号中的比例，第二个数字为代理对象形状的差异（即 $\Sigma |R_i^{3,4} - t_i^{3,4}| / \Sigma t_i^{3,4}$ ，其中 i 为最后 6 个分区编号，对应 deep-WH > 0.7）。

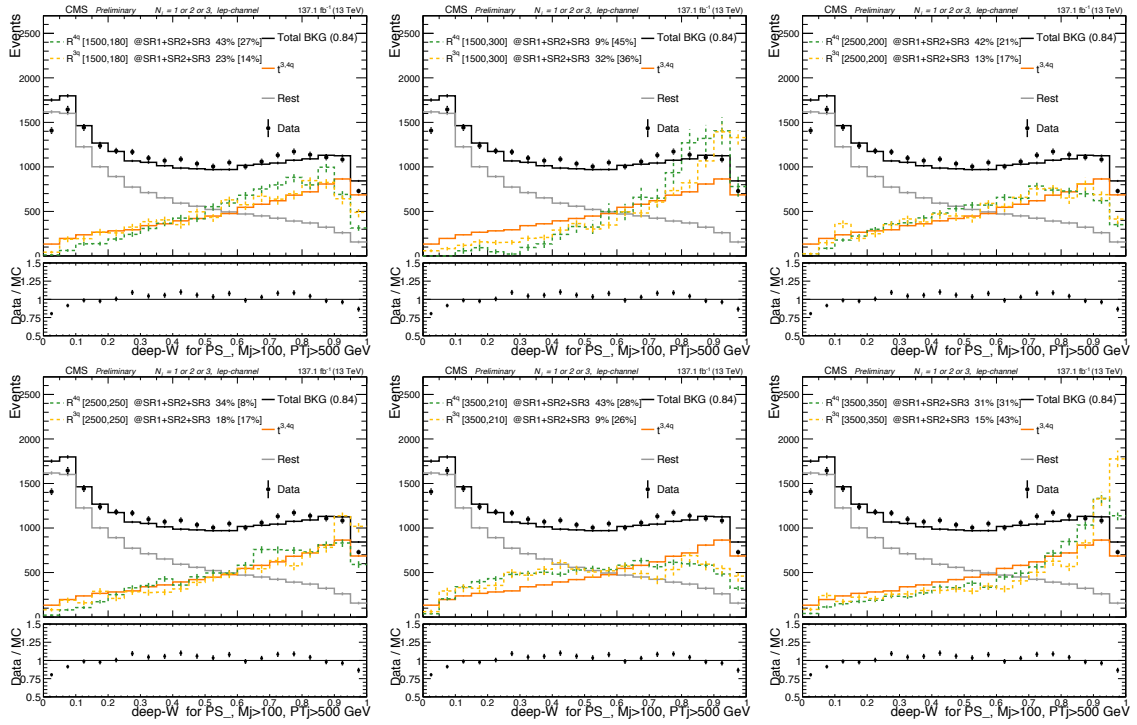


图 5.21 6 个信号点的 R^{3q} 、 R^{4q} 与本底 $t^{3,4}$ 的 deep-W 分布对比, R^{3q} 、 R^{4q} 分别归一到 $t^{3,4}$ 上, 本底喷注的条件如 x 轴标题所示信号调节为去除 deep-W (WH) 条件的信号区 1、2、3, 图例中第一个数字是该成分在信号中的比例, 第二个数字为代理对象形状的差异 (即 $\Sigma |R_i^{3,4} - t_i^{3,4}| / \Sigma t_i^{3,4}$, 其中 i 最后 6 个分区编号, 对应 deep-WH>0.7)。

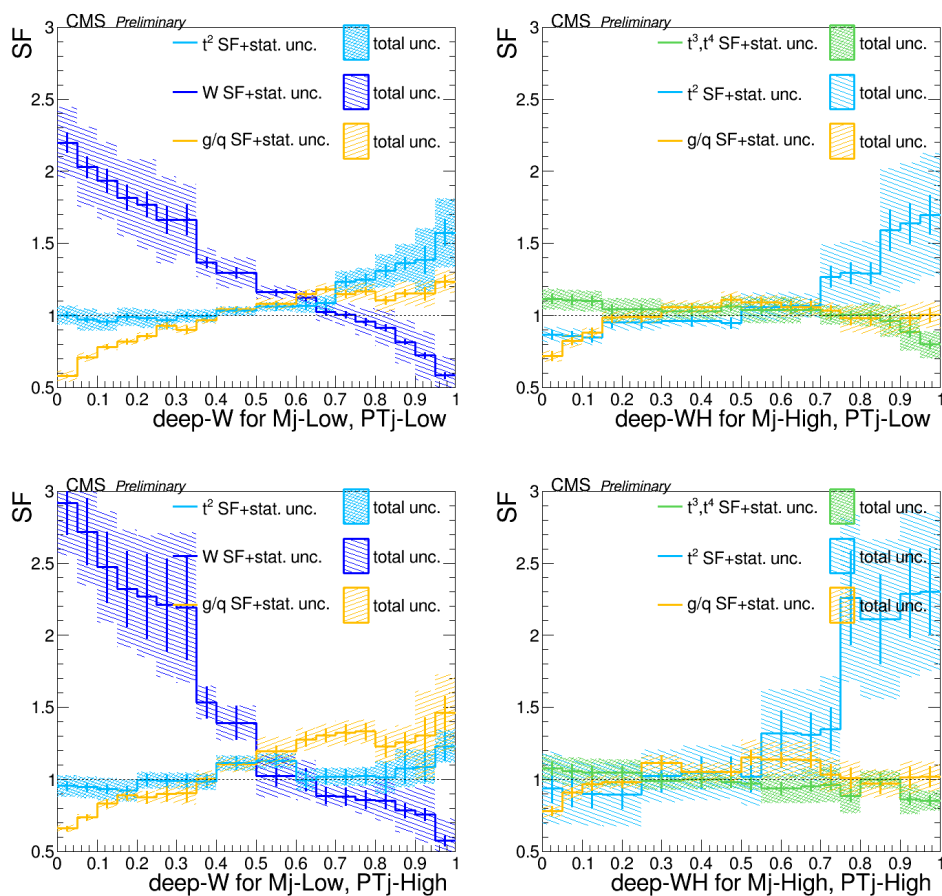


图 5.22 分析使用到的 SFs 及相应的 PS 系统误差。

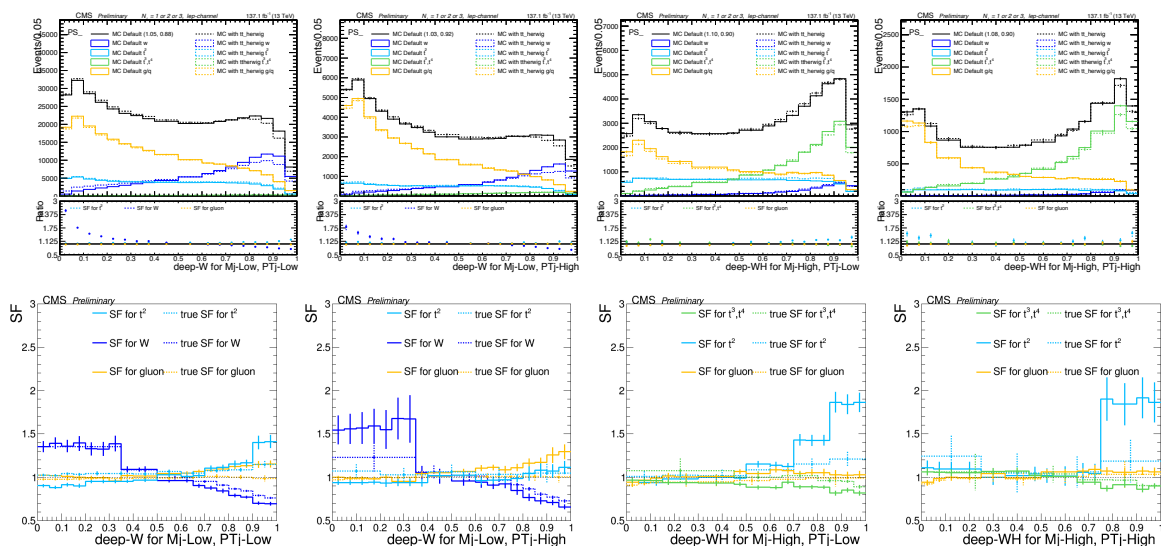


图 5.23 第一行是喷注质量-横动量为低低、低高、高低、高高时真实 SFs 的分布，第二行为相应的四个区域中 SFs 的真实值与测量值的比较。

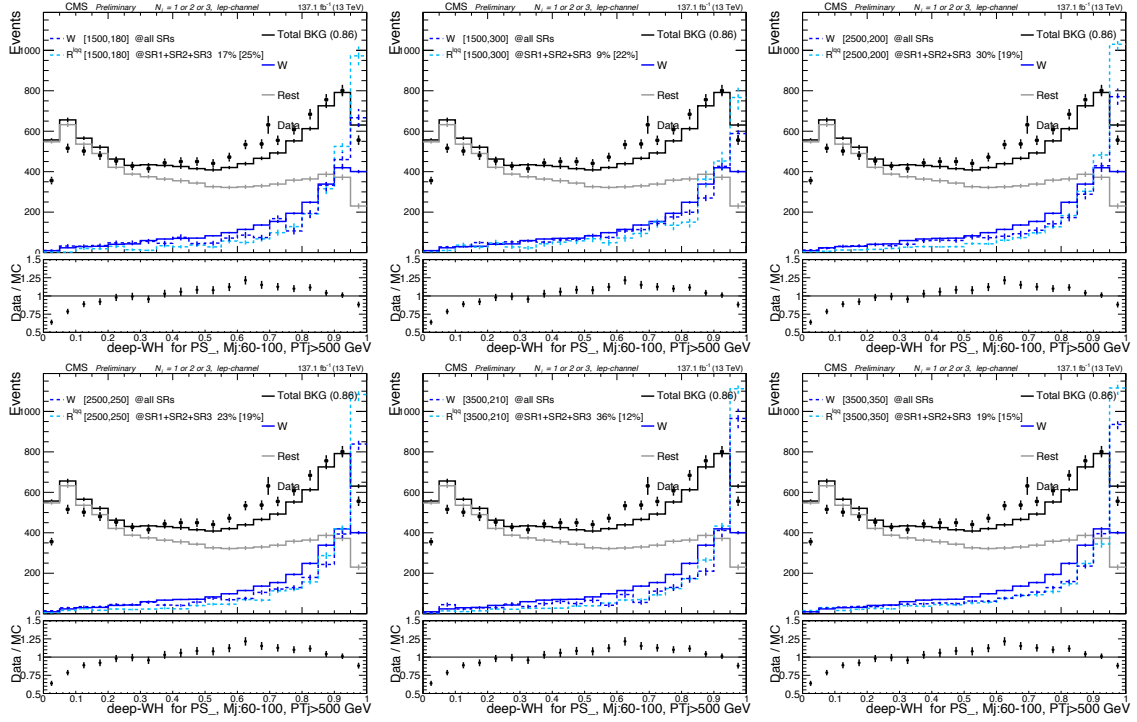


图 5.24 6 个信号点 R^{lqq} 与 W 关于 deep-WH 的对比，信号和本底不同的挑选条件都在图例中给出，图例中 R^{lqq} 后的 [...] 中的数字为其代理误差。

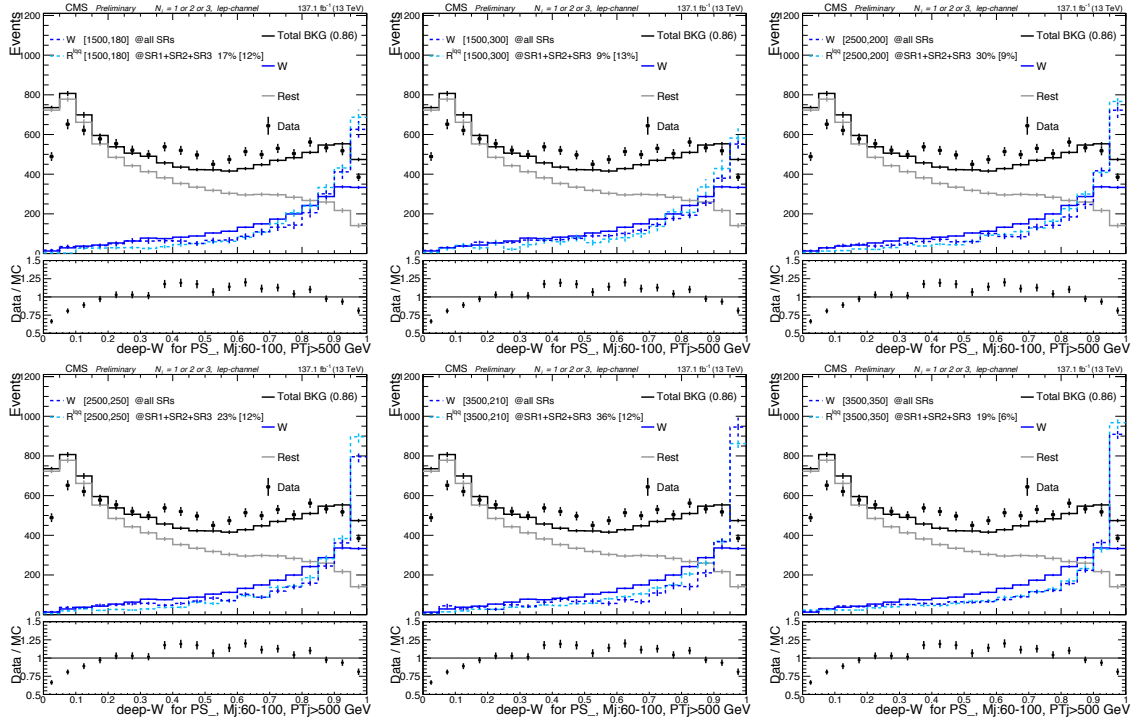


图 5.25 6 个信号点 R^{lqq} 与 W 关于 deep-W 的对比，信号和本底不同的挑选条件都在图例中给出，图例中 R^{lqq} 后的 [...] 中的数字为其代理误差。

5.7 top 本底的控制区的定义与估计

使用本底估计方法时需要定义某本底占主导的控制区来估计该本底在信号区的缩放比例因子和形状。对于本文，top 和 W+jets 是两项最重要的本底过程。首先对于 top，将 $\bar{t}\bar{t}$ 与 single top 整合来估计，为了得到无信号、top 占主导且与信号区运动学特征相似的控制区样本，定义如下：

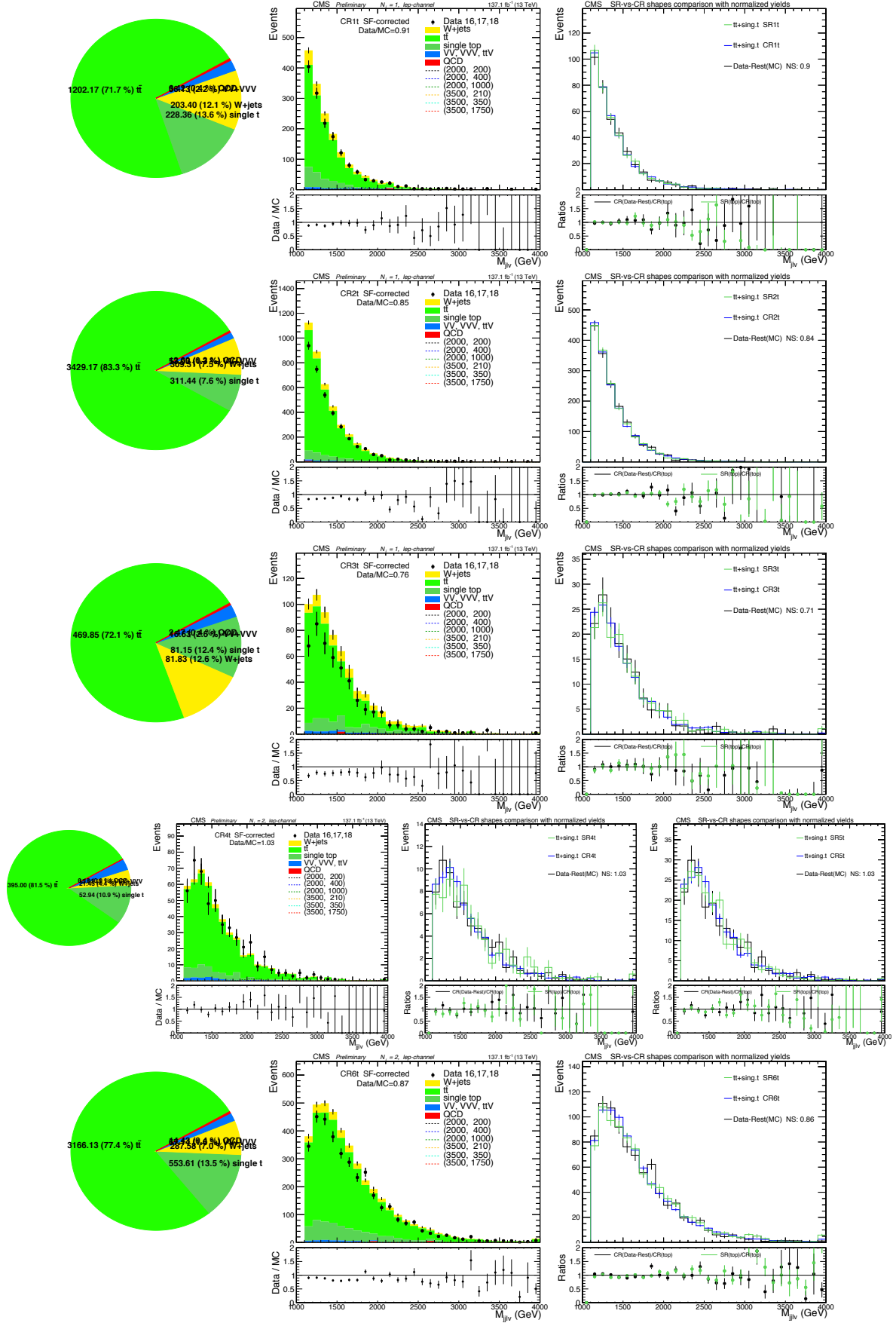
- 首先使用信号区的挑选条件（4.4.3节中定义）来最大限度保证相似的运动学特征；
- 要求标记为 b 的喷注数量满足 $N_b^M \geq 1$ ，即替换信号区 $N_b^M = 0$ 的条件来提高 top 过程的占比，同时压低信号，防止信号污染控制区样本；
- 允许最多有 4 个与 AK8 喷注没有重叠的 AK4 喷注数量， $N_{j-AK4-exc} \leq 4$ ，来容纳更多的 b 喷注；
- 去除 deep-W（WH）截断条件来提高统计量。

与 6 个信号区一一对应定义了 top 的控制区（control region for top, CRt），但是由于去掉了 deep-W（WH）条件，所以区域 4 和 5 合并，最终得到 5 个 top 控制区，命名为 CR1t、...、CR45t、CR6t。

图5.26中第一列是 5 个 top 控制区各过程占比饼图，top 始终是主导过程，占 ~85-95%。第二列为观测量 $M_{j\ell\nu}$ 或 $M_{jj\ell\nu}$ 的分布，数据与蒙卡样本符合情况较好，而且基本没有信号分布。第三列 top 过程在信号区、top 的控制区及数据减除 top 外其他过程（Data-Rest）在控制区的观测量形状对比，后两个直方图都被归一到第一个的总事例数，在 CR45t 的对比图中，一张是与来自信号区 4 的 top 对比，另一张是与来自信号区 5 的 top 对比。在控制区 Data-Rest 与 top 事例数的差异被用于缩放 top 在该控制区与对应信号区的 top 本底的事例数，在对比图中用“NS”表示。另外对比图中三个直方图形状基本相同，这说明可以利用这些控制区来估计信号区内的 top 本底。估计信号区事例数和形状的公式为：

$$PRED_{SRi}^t = [DATA-rest]_{CRti} \frac{MC_{SRi}^t}{MC_{CRti}^t} = N_{CRi}^t MC_{SRi}^t \left(\frac{[DATA-rest]_{CRti}}{N_{CRi}^t MC_{CRti}^t} \right) = N_{CRi}^t MC_{SRi}^t TF_{ti},$$

其中 i 表示控制区的数字编号， N_{CRi}^t 为 top 的比例缩放因子，rest 指非 $\bar{t}\bar{t}$ 或者 single top 过程，该公式中所有蒙卡样本都已添加 SFs。在最式中最后一项定义了传递系数（transfer factor, TF），是图5.26第三列中黑色的比例点，它用来修正信号区尾部高能端的形状。分析中选用了线性函数来拟合传递系数，如附录B.8所示。拟合得到的 $\pm 1\sigma$ 误差带会被用于作为传递系数的形状误差允许观测量在高端上下浮动。


 图 5.26 5 个 top 观测区各过程占比, 对应的观测量 M_{jlv} 的分布和形状的对比。

5.8 W+jets 本底控制区的定义与估计

类似的，对 W+jets 本底的控制区定义如下：

- 使用 6 个信号区的定义来保证类似的运动学特征；
- 将 deepAK8 条件反转为 deep-W(WH)<0.7 或 0.5 来压低信号和 top 过程；
- 要求最大质量喷注满足 deep-t<0.4 来进一步压低 top 本底。

W+jets 控制区的定义与 top 控制区类似，将信号区 4 和 5 中 deep-W 条件反转后得到的样本统计量太低，因为区域 456 相似，因此选择合并作为控制区来估计信号区 4-6，最终得到了 4 个 W+jets 控制区，记为 CR1W、CR2W、CR3W 和 CR456W。

图5.27与上一节的图格式一致，4 行对于 4 个控制区，3 类从左到右依次为过程事例数饼图、数据蒙卡样本观测量分布。控制区 Data-Rest 与 W+jets 事例数的差异被用于缩放 W+jets 在该控制区与对应信号区的 W+jets 本底的事例数，在对比图中用” NS “表示。另外第三列三个直方图形状基本相同，这说明定义的控制区可以有效估计信号区的 W+jets 分布。与 top 控制区估计方法类似，估计信号区事例数和形状的公式为：

$$PRED_{SRi}^W = [DATA-rest]_{CRWi} \frac{MC_{SRi}^W}{MC_{CRWi}^W} = N_{CRi}^W MC_{SRi}^W \left(\frac{[DATA-rest]_{CRWi}}{N_{CRi}^W MC_{CRWi}^W} \right) = N_{CRi}^W MC_{SRi}^W TF_{Wi},$$

最后一项定义了传递系数，为比值 $[DATA-rest]_{CRWi}/MC_{CRWi}^W$ ，采用线性函数来拟合传递系数，见附录B.9。传递系数可以有效地修正观测量高能量即尾部区域， $\pm 1\sigma$ 拟合带来提供关于传递系数的系统误差。

5.9 本底估计方法的总结

首先将本底分为 3 类：W+jets、top 和剩余部分。对于前两种主要本底，定义了对应的控制区，总结如表5.6，使用 SFs 修正后的观测量分布，按控制区得到的数据与蒙卡事例数相对差值来按比例缩放信号区对应的本底过程，并且用传递系数来修正信号区的形状。对于剩余本底，由于占比极少，因此直接使用其在信号区的蒙卡样本形状。图5.28是考虑 SFs 修正后控制区中观测量谱，这些直方图将会用于作为拟合前分布参与统计计算并设置截面排除限。

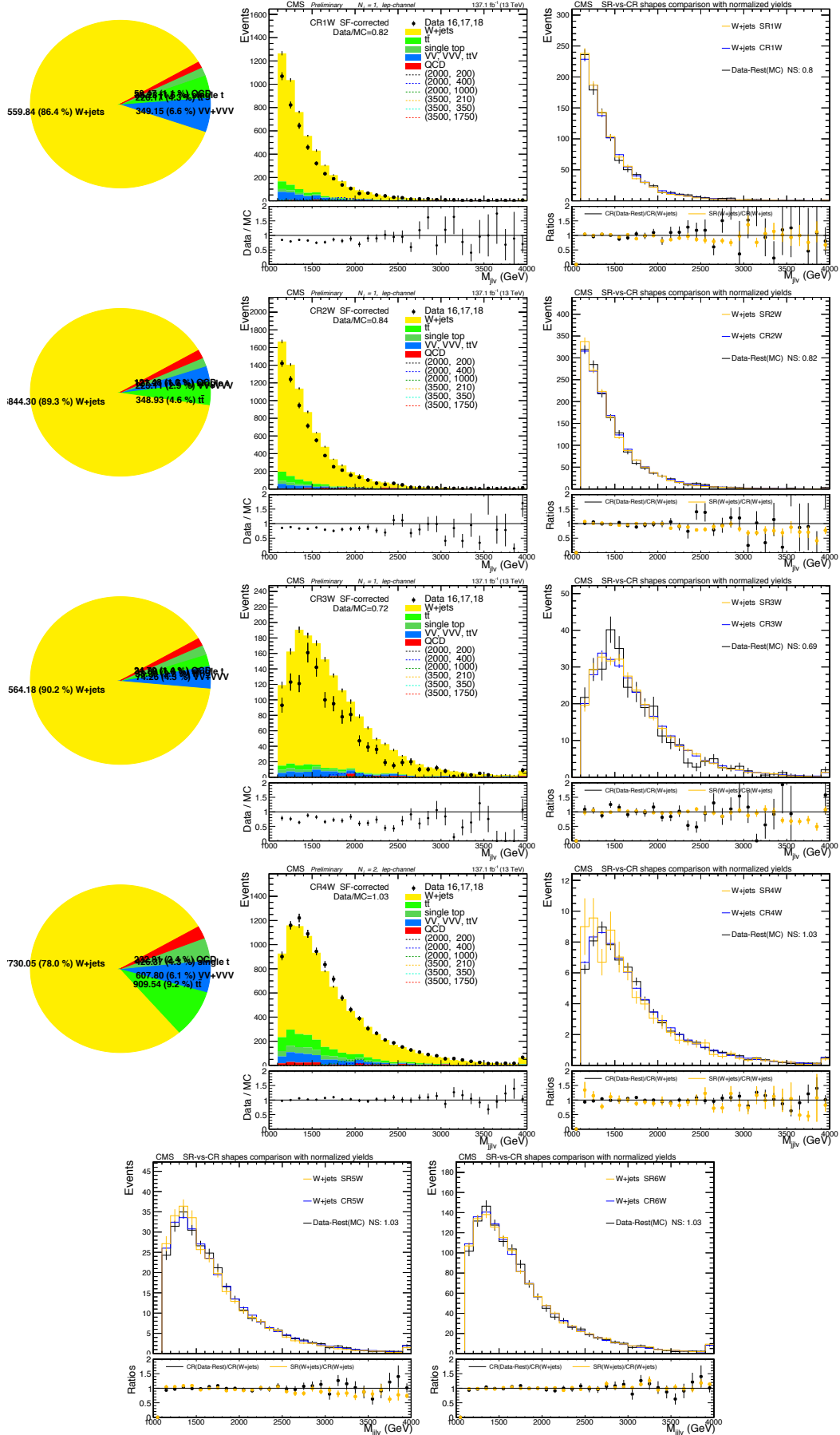


图 5.27 4 个 W+jets 观测区各过程占比, 对应的观测量 M_{jlv} 的分布和形状的对比如图。其中 CR456W 对比图共有 3 张 (最后两行), 分别为与信号区 4、5 和 6 的对比。

表 5.6 信号区、top 和 W+jets 控制区挑选条件的差异总结。共同条件为 $S_T > 1$, $M_{jj\ell\nu} > 1.1$, $p_T(W) > 200 \text{ GeV}$, $p_T(j) > 200$, 横线” - “表示该条件没有要求。

区域	$M_j^{\max} []$	deepAK8 标记	$M_j^{\min} []$	deepAK8 标记	N_j^{AK8}	N_j^{AK4}	N_b^{M}
SR1	60-100	deep-W > 0.7	-	-	1	≤ 2	0
SR2	100-200	deep-WH > 0.7	-	-	1	≤ 2	0
SR3	≥ 200	deep-WH > 0.7	-	-	1	≤ 2	0
SR4	60-100	deep-W > 0.5	60-100	deep-W > 0.5	2	≤ 2	0
SR5	60-100	deep-W < (>) 0.5	60-100	deep-W < (>) 0.5	2	≤ 2	0
SR6	60-100	deep-W > 0.7	0-60	-	2	≤ 2	0
CRt1	60-100	deep-W > 0	-	-	1	≤ 4	≥ 1
CRt2	100-200	deep-WH > 0	-	-	1	≤ 4	≥ 1
CRt3	≥ 200	deep-WH > 0	-	-	1	≤ 4	≥ 1
CRt45	60-100	deep-W > 0	60-100	deep-W > 0	2	≤ 4	≥ 1
CRt6	60-100	deep-W > 0	0-60	-	2	≤ 4	≥ 1
CRW1	60-100	deep-W(t) < 0.7(0.4)	-	-	1	≤ 2	0
CRW2	100-200	deep-WH(t) < 0.7(0.4)	-	-	1	≤ 2	0
CRW3	≥ 200	deep-WH(t) < 0.7(0.4)	-	-	1	≤ 2	0
CRW456	60-100	deep-W(t) < 0.7/0.5(0.4)	0-60/60-100	-/deep-W(t) < 0.5	2	≤ 2	0

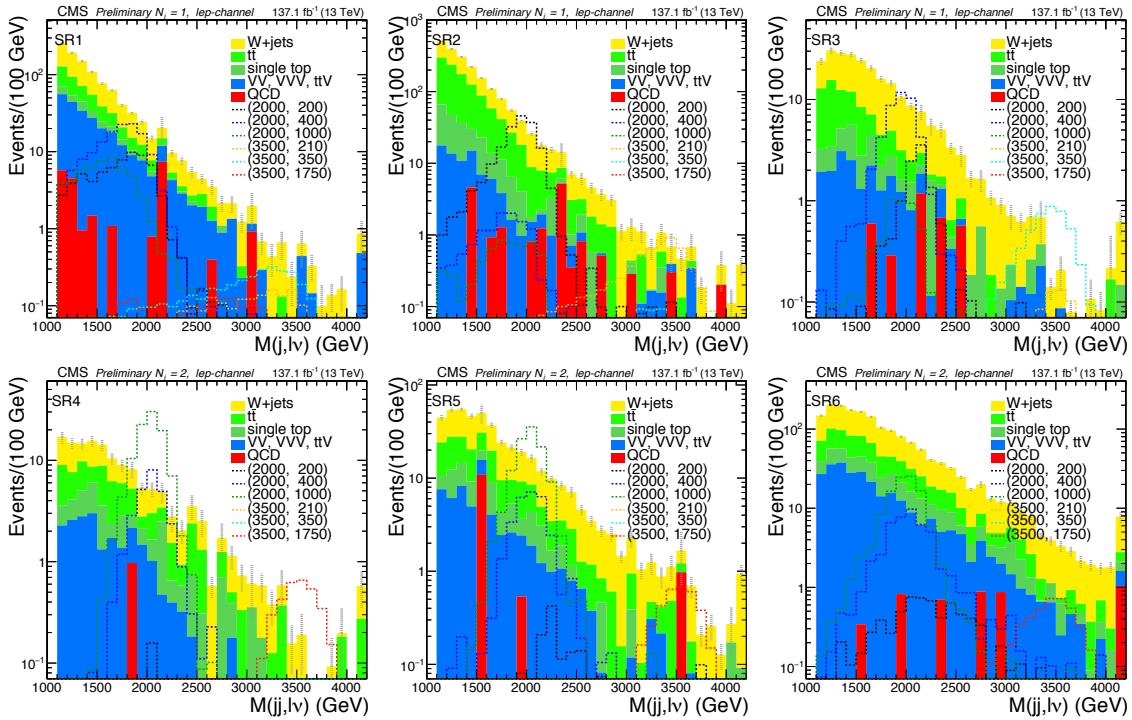


图 5.28 信号区 1-6 考虑 SFs、控制区缩放因子和传递因子修正后的观测量分布。

第六章 系统误差及结论

在上一章介绍中介绍了分析使用的 deepAK8 标记量的 SFs 并验证了其可靠性，在对所有蒙卡样本的每个事例中的喷注加以修正，可以得到各本底的形状。另外也讨论了关于 SFs 的误差，本章将介绍除此之外的其它系统误差。表6.1总结了分析中考虑的全部系统误差，从左到右分别为误差来源、作用于信号 (S) 或本底 (B)、对最终观测量是影响事例数的归一化 (rate) 或形状 (shape)、影响那个信号区和对应冗余参数的数量。在之后会详细介绍各个系统误差。

表 6.1 系统误差和对应的冗余参数表。最后一列表示 SFs 计算中喷注质量和横动量分区，用 L、H 表示低、高。

i	系统误差来源	信号或本底 (S/B)	影响	作用于信号区	冗余参数数量
1	SFs_deepW_PS+Bias(W, R^{lqq})	B+S	Rate	1,4,5,6	2(LL, LH)
2	SFs_deepWH_PS+Bias(W, R^{lqq})	B+S	Rate	2,3	2(HL, HH)
3	SFs_deepW_PS+Bias(t^2)	B	Rate	1,4,5,6	2(LL, LH)
4	SFs_deepWH_PS+Bias(t^2)	B	Rate	2	4(LL, LH, HL, HH)
5	SFs_deepWH_PS+Bias(t^2)	B	Rate	3	2(HL, HH)
6	SFs_deepW_PS+Bias($t^{3,4}, R^{3q,4q}$)	B+S	Rate	1,4,5,6	2(LL, LH)
7	SFs_deepWH_PS+Bias($t^{3,4}, R^{3q,4q}$)	B+S	Rate	2,3	2(HL, HH)
8	SFs_deepW_PS+Bias(g/q)	B	Shape+Rate	1,4,5,6	2(LL, LH)
9	SFs_deepWH_PS+Bias(g/q)	B	Rate	2	4(LL, LH, HL, HH)
10	SFs_deepWH_PS+Bias(g/q)	B	Rate	3	2(HL, HH)
11	SFs_deepW_Proxy(R^{lqq})	S	Rate	1,	1
12	SFs_deepWH_Proxy(R^{lqq})	S	Rate	2,3	1
13	SFs_deepW_Proxy(R^{3q})	S	Rate	1	1
14	SFs_deepWH_Proxy(R^{3q})	S	Rate	2,3	1
15	SFs_deepW_Proxy(R^{4q})	S	Rate	1	1
16	SFs_deepWH_Proxy(R^{4q})	S	Rate	2,3	1
17	SFs_deepW_Proxy(unmatched)	S	Rate	1,4,5,6	1
18	SFs_deepWH_Proxy(unmatched)	S	Rate	2,3	1
19	SFs_deepW_HighPT(W)	S	Rate	1,4,5,6	1
20	SFs_deepWH_HighPT(W)	S	Rate	2	1
21	SFs_deepW_HighPT(R^{lqq})	S	Rate	1	1
22	SFs_deepWH_HighPT(R^{lqq})	S	Rate	2,3	1
23	SFs_deepW_HighPT(R^{3q})	S	Rate	1	1
24	SFs_deepWH_HighPT(R^{3q})	S	Rate	2,3	1
25	SFs_deepW_HighPT(R^{4q})	S	Rate	1	1
26	SFs_deepWH_HighPT(R^{4q})	S	Rate	2,3	1
27	Normalization W+jets	B	Rate	1,2,3,4,5,6	5
28	Normalization top	B	Rate	1,2,3,4,5,6	5
29	Normalization rest	B	Rate	1,2,3,4,5,6	5
30	TF_W from CRWs	B	Shape	1,2,3,4,5,6	6
31	TF_t from CRts	B	Shape	1,2,3,4,5,6	6
32	PU reweight Trigger, ID, Iso_eff. Luminosity	S	Rate	1,2,3,4,5,6	1
33	lepton energy scale	S	Shape	1,2,3,4,5,6	1
34	PDFs	S	Rate	1,2,3,4,5,6	1
35	QCD scales μ_F, μ_R	S	Rate	1,2,3,4,5,6	1
36	JEC	S	Shape	1,2,3,4,5,6	1
37	JER	S	Shape	1,2,3,4,5,6	1
38	JMS	S	Shape	1,2,3,4,5,6	1
39	Signal Strength (β)	S	Rate	1,2,3,4,5,6	1

6.1 本底和信号共用的系统误差

SFs 的误差对应于表6.1的前 26 项。本底过程的喷注按产生子匹配分为 W、 t^2 、 t^3 、 g/q 和未匹配 (unmatched)，信号过程分为 W、 R^{lqq} 、 R^{3q} 、 R^{4q} 和 unmatched，并按每个匹配的喷注计算误差。如5.6章中介绍，包括 PS+ 偏离误差（对应表6.1中 1-10 项）、代理误差（11-18 项）和高动量外推误差（19-26 项），对应一共 36 个冗余参数。对应第一项误差，图6.1展示了本底过程在 6 个信号区考虑误差后观测量分布相对初始分布的变化，可以看到信号区 4 和 5， g/q 分布在整体上分布不均匀，因此考虑这一项对于观测量的形状影响，对应表6.1第 8 行，其余大致为水平分布，对形状影响较小，因此只考虑对事例数的归一化影响。图6.2为 PS+ 偏移误差对 2 个信号的影响，该误差对信号观测量影响很小，因此只考虑整体的归一化影响。每个信号点的误差都要根据5.6章中的方法独立计算，最终三项误差对于所有信号点的事例数的影响如图6.3所示。

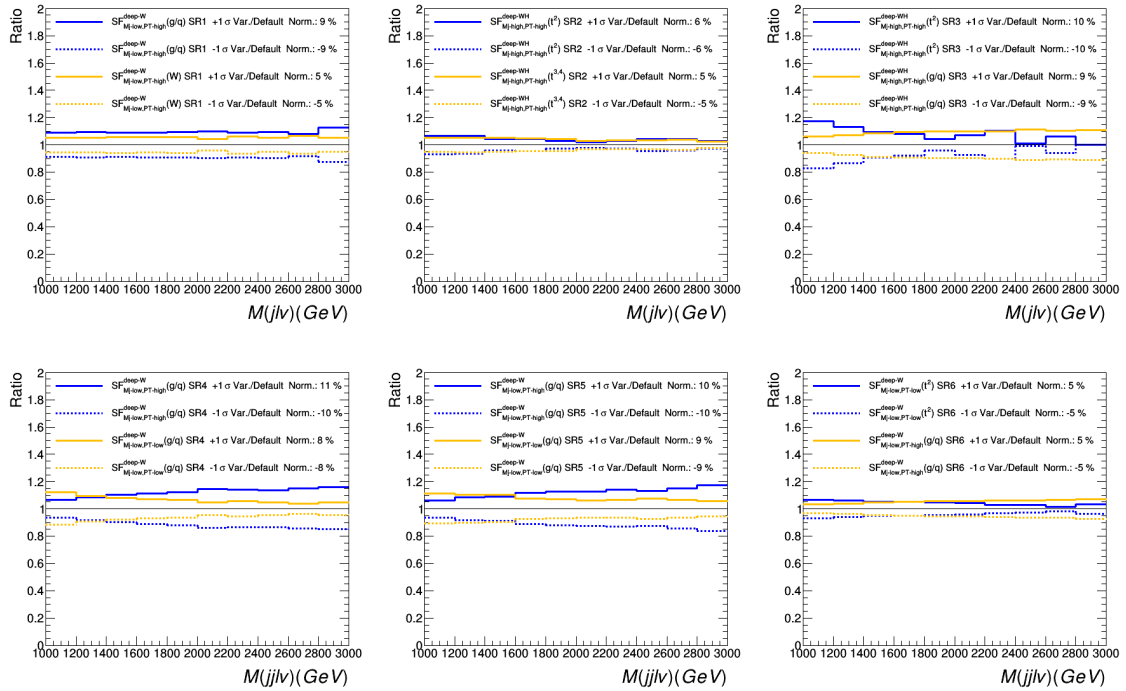


图 6.1 考虑 PS+ 偏移误差后，本底中影响最大的 2 种 SFs 误差相对于初始形状的比例，上方为信号区 1-3，下方为信号区 4-6。

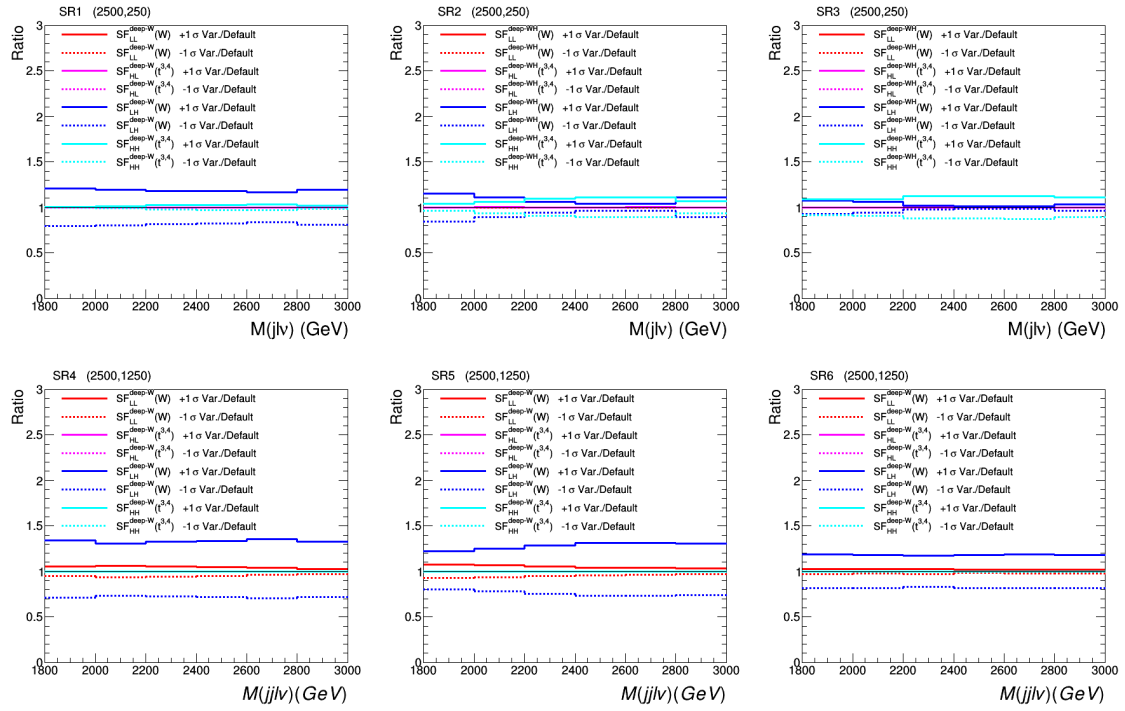


图 6.2 PS+ 偏移误差对于信号在每个信号区的影响，上面 3 张图对应合并信号 (2500,250)GeV 在信号区 1-3，下方为分立信号 (2500,1250)GeV 在信号区 4-6。

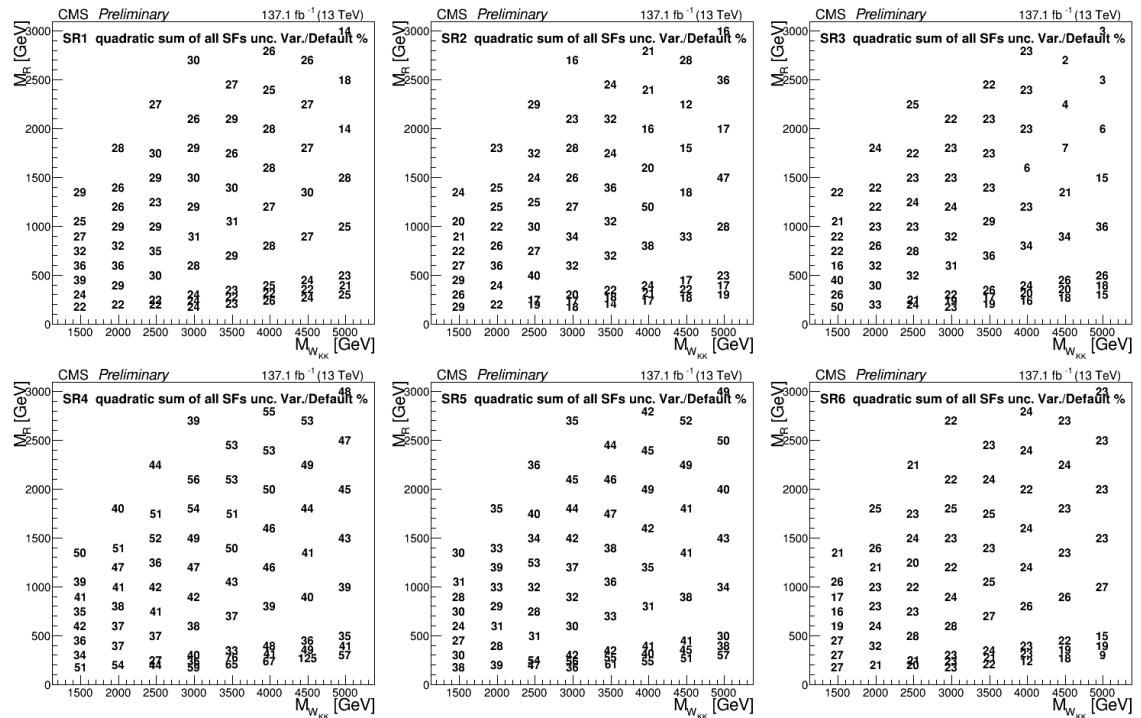


图 6.3 对应全部信号点，PS+ 偏离误差、代理误差和高动量外推误差相对初始事例数归一化的影响，对于大部分信号，事例数在考虑误差后变化 20-50%。

6.2 本底系统误差

6.2.1 本底归一化误差

分析中在信号区有三种本底：W+jets、top 和其他。对于 W+jets 和 top，使用控制区得到的归一化因子，即图5.26和5.27右侧的“NS”，来缩放信号区对应过程的事例数。同时，考虑到信号区和控制区归一化因子的差异，将 $1-NS$ 作为对应信号区的误差。不同本底在 6 个信号区的误差分别为：

Wjets: 20%、18%、31%、3%、3%、3%；

top: 10%、16%、29%、3%、3%、14%；

剩余本底: 50%、50%、50%、50%、50%、50%。

由于在信号区 4 和 5 除了 deep-W 选择条件之外其他条件都相同，因此这两个区域的误差是关联的，其他信号区每个都分别对应不同的冗余参数，因此一共有 15 个。

6.2.2 本底形状误差

在本底估计方法中对控制区采用线性拟合传递系数从而修正信号区的形状，传递系数如图B.8和B.9左侧所示，图中展示了拟合得到的 $\pm 1\sigma$ 误差带。可以将拟合得到的 $\pm 1\sigma$ 误差形状传递到观测量从而得到传递系数误差，图6.4展示了 W+jets 和 top 控制区传递系数误差对观测量初始形状的影响，这项误差允许观测量尾部上下浮动。

6.3 信号系统误差

由于对本底采用了数据驱动的预测方法，因此不需要考虑如喷注能量修正、能量比例缩放刻度、能量分辨、部分子分布函数误差等的影响。但是对于信号，需要额外考虑。

6.3.1 信号归一化误差

亮度、pileup、轻子鉴别分辨及触发误差

对应于表6.132 行，其中累计亮度误差为 1.8%[29]。对于 pileup，通过增加或减小 5% 的最小偏离截面来估算误差 [41]，相关研究见附录图B.12和 fig. B.13分别为 μ 子和电子道 pileup 误差对观测量的影响，可以发现没有显著的形状差异，因此只考虑归一化误差。另外两个道的形状走势相反，是反相关的，该项误差为 0.5%。

轻子鉴别与分辨效率误差对 μ 子和电子道的影响如图B.14和B.15所示，选用图中的最大偏差值 1.8% 和 2.7% 来分别作为 μ 子道和电子道的归一化误差。触发效率测量

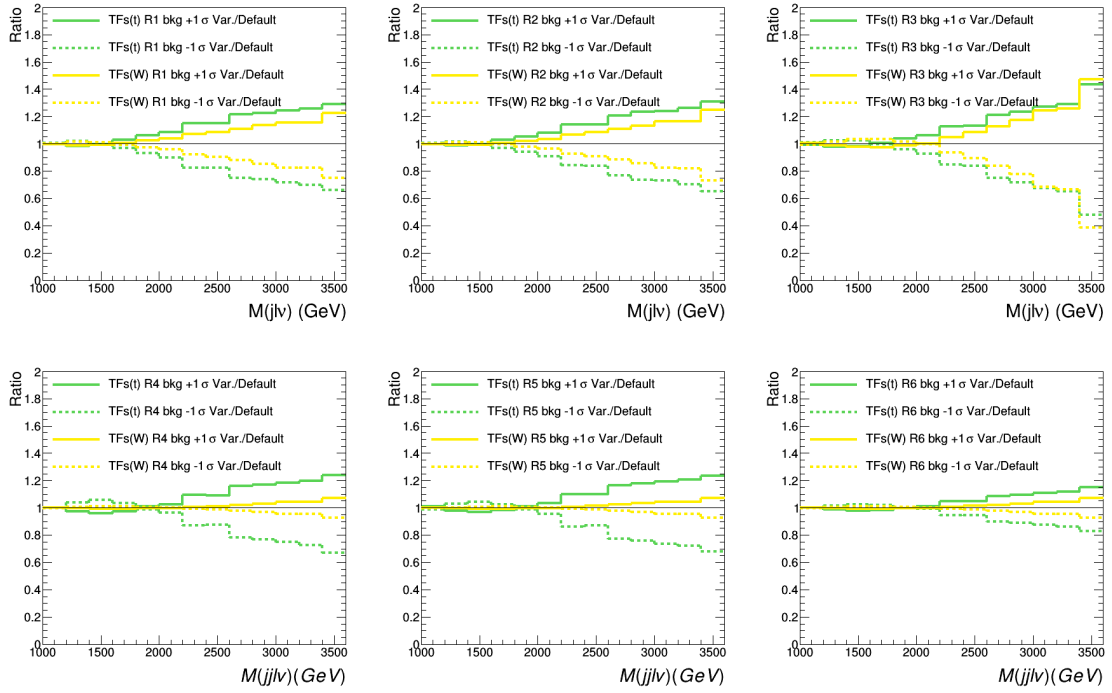


图 6.4 W+jets 和 top 控制区传递系数 $\pm 1\sigma$ 误差带，第一行对应信号区 1、2 和 3，第二行为信号区 4、5 和 6。

误差的影响如图 B.16 和 B.17 所示，分别取 3.1% 和 0.5% 作为 μ 子道和电子道的误差。这几项误差是相互关联的，最终合并之后的误差为 2.9%。

QCD 缩放和部分子分布函数误差

QCD 缩放影响包括两个因子：再归一因子 μ_R 和重整化因子 μ_F ，缩放误差是将两个因子分布变为原来的 1/2 或 2 倍，得到最大的差值作为误差，图 6.5 展示了 6 个信号区的缩放误差，对应大部分信号点为 1%。

信号部分子分布函数函数 (pdf) 采用 NNPDF3.0，包含 100 个权重成员，通过计算不同权重成员与默认权重的差异作为 pdf 误差 [23]，如图 6.6，对应大部分信号点也为 1%。

6.3.2 信号形状误差

轻子能量缩放刻度误差

由于探测器分辨率等的影响，轻子能量刻度有一定误差，会影响轻子衰变的 W 的重建，进而影响最终的观测量。分析中通过将轻子总动量增大或减小 2% 来估计这一误差，图 6.7 反应了轻子动量改变后对观测量的影响。

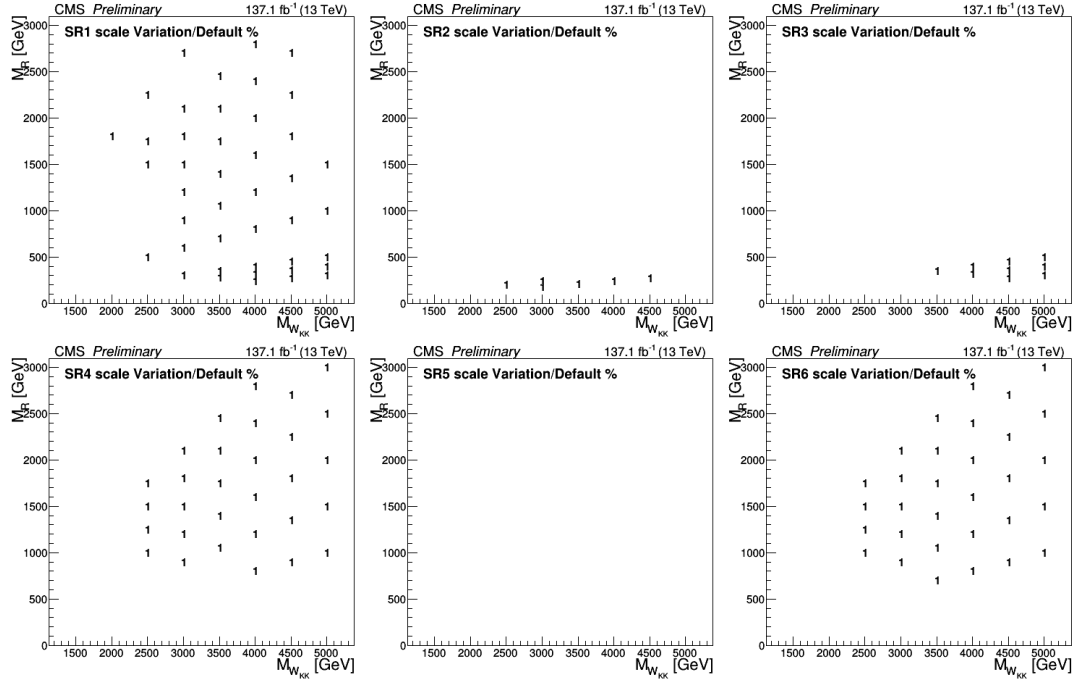


图 6.5 不同信号在不同信号区 QCD 缩放误差。

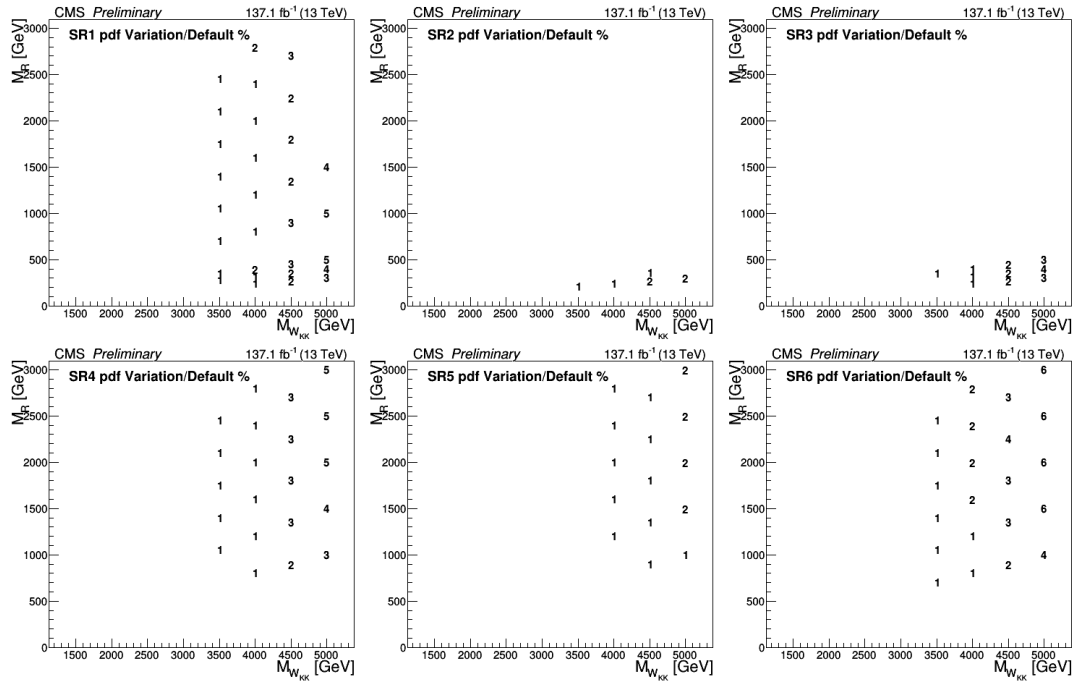


图 6.6 不同信号在不同信号区 pdf 误差。

喷注能量修正和分辨误差

喷注能量修正 (JEC) 和分辨 (JER) 会影响重建喷注的横动量及能量, 继而影响丢失横动量和最终的观测量。基于喷注横动量并按照官方推荐来计算 JEC 和 JER 的修

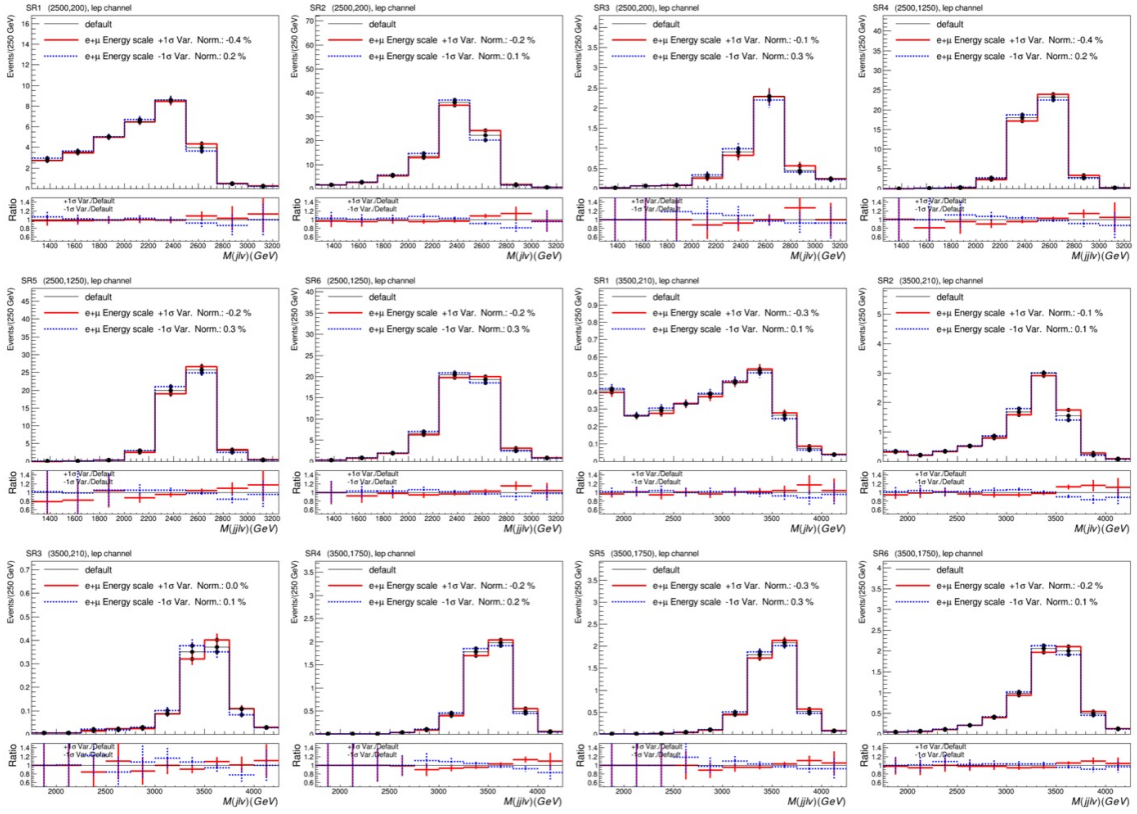


图 6.7 全轻子道轻子能量缩放刻度误差误差对信号观测分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

正因子误差，并传递到观测测量，如图6.8和6.9所示。

喷注质量刻度误差

分析使用 soft drop 质量，CMS 只有对两夸克的 W 喷注质量的修正。而对于更复杂的，比如本文中 t^2 、 $t^{3,4}$ 、 R^3 、 R^4 和 R^{lqq} 并没有官方的修正及相应的误差。基于其他分析过程的经验，喷注质量刻度 (JMS) 的误差大约为 1%，因此，本文通过将喷注质量变动 $\pm 1\%$ 来研究刻度误差的影响，如附录B.18所示，分析中使用 $\pm 1\sigma$ 对应的形状作为该项的误差。

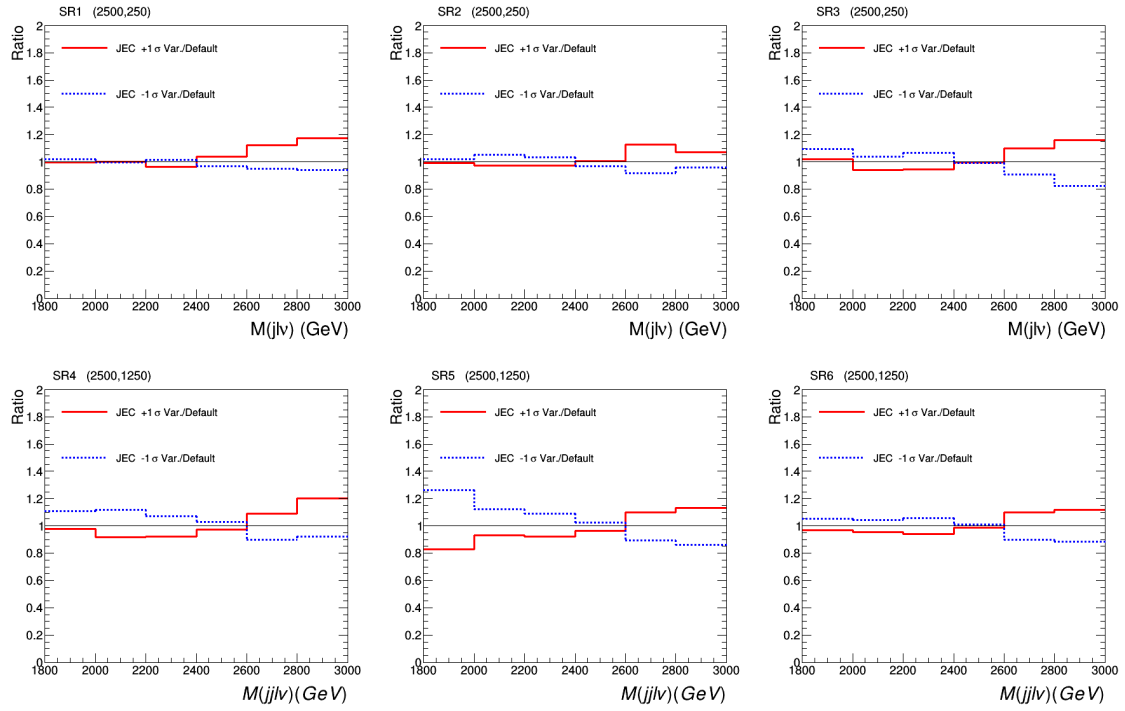


图 6.8 JEC $\pm 1\sigma$ 误差与初始形状的对比，第一行是合并信号 (2500, 250) 在信号区 1-3 的对比，第二行是分立信号 (2500, 1250) 在信号区 4-6 的对比。

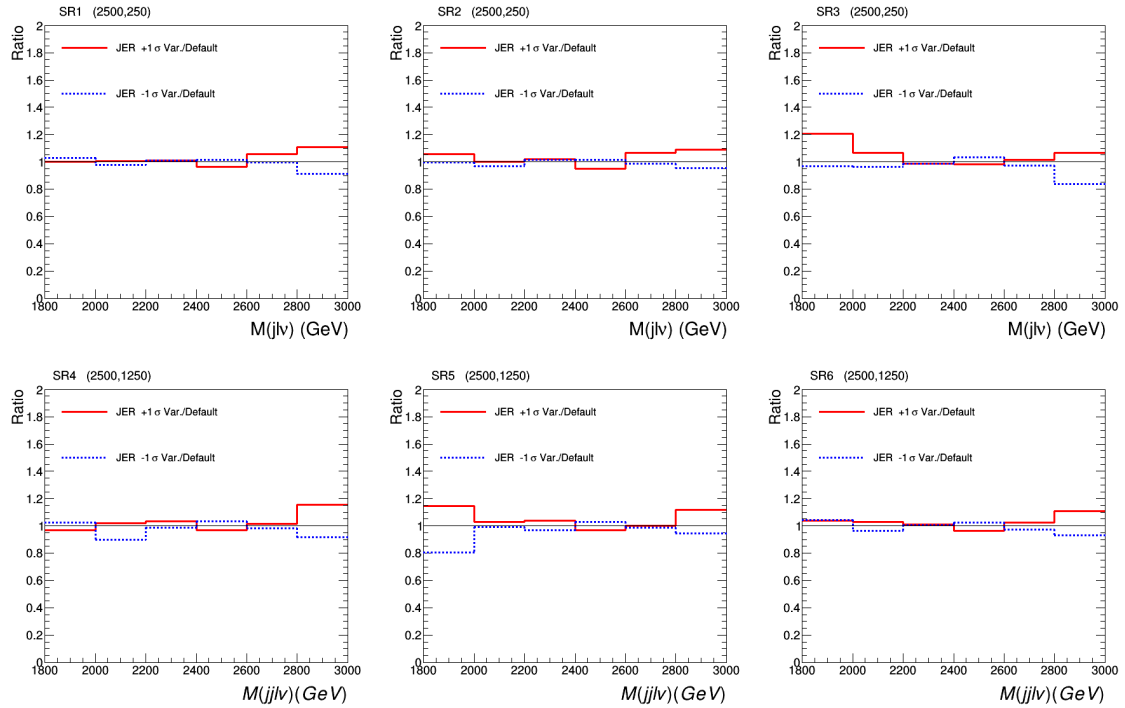


图 6.9 JER $\pm 1\sigma$ 误差与初始形状的对比，第一行是合并信号 (2500, 250) 在信号区 1-3 的对比，第二行是分立信号 (2500, 1250) 在信号区 4-6 的对比。

6.4 拟合与排除限设置

6.4.1 预拟合与后拟合分布

确定本底估计方法和系统误差后，将用到的形状作为模板输入 CMS 官方的统计工具 Higgs Combine Tool[28] 可以得到 6 个信号区预拟合与后拟合观测分布图，如6.10和6.11所示，在拟合后，系统误差得到了很好地限制，而且数据与蒙卡基本吻合，没有新粒子的迹象。

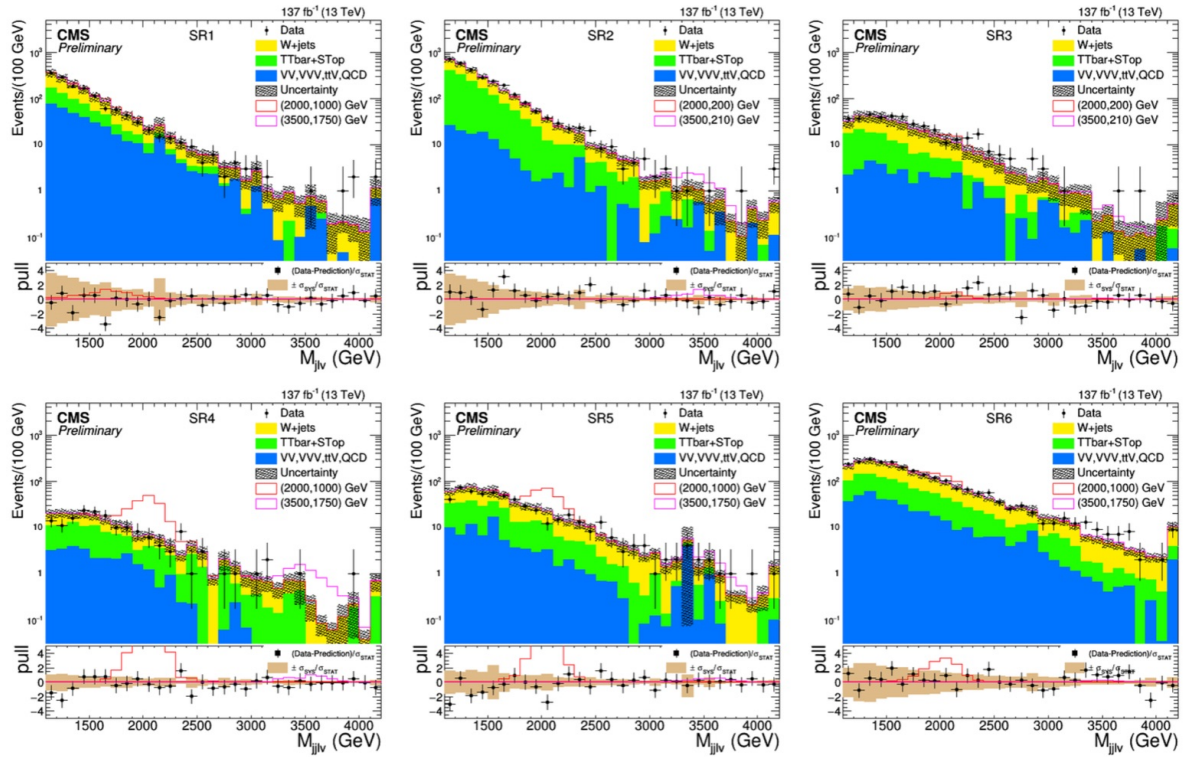


图 6.10 观测量预拟合分布图，第一行为信号区 1-3，第二行为信号区 4-6。

6.4.2 拟合结果评价

为了评价拟合结果，采用 Saturated Model[47] 产生 toys 来进行统计建议，类似于 χ^2 分布，是用于检验数据是否与预期的模型吻合。图6.12展示了统计检验结果的分布，并用高斯拟合。定义 p 值如下：

$$p = \int_{t=t_0}^{+\infty} f(t)dt, \quad (6.1)$$

其中 t 是横轴统计检验值， t_0 是数据得到的检验结果。最后 p 值为 0.27，说明拟合时合理的。

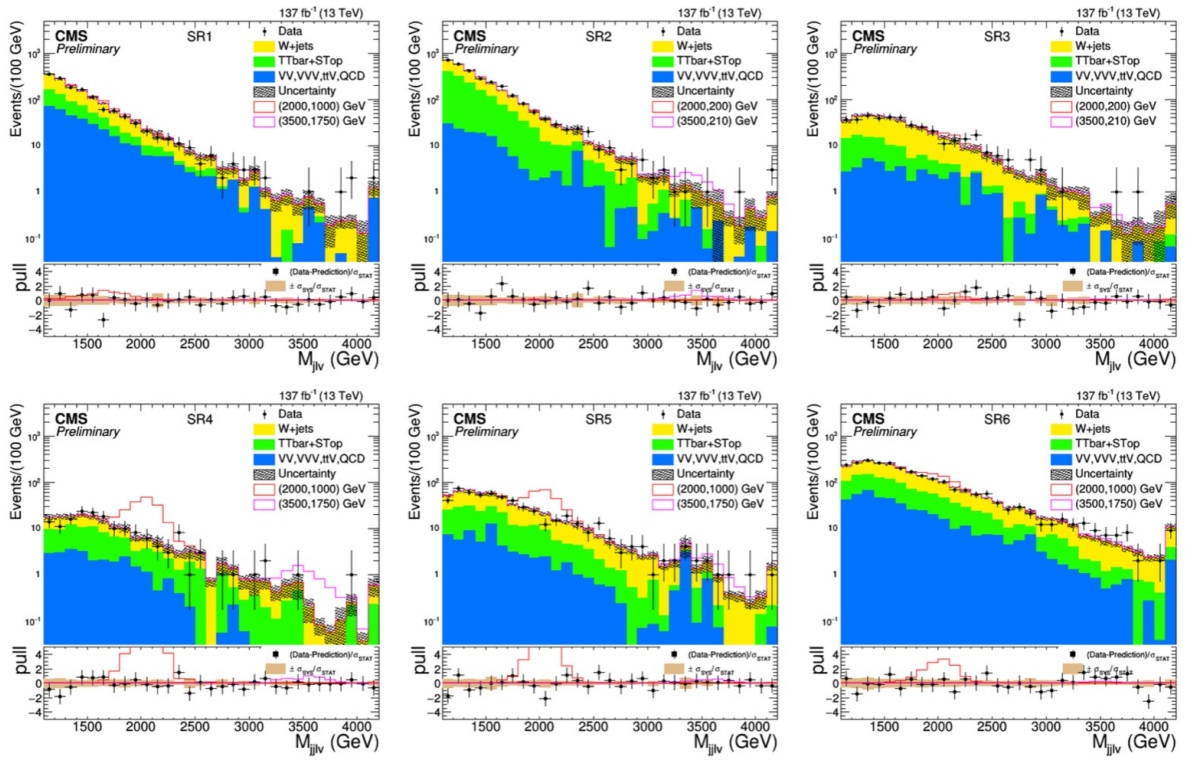


图 6.11 观测量后拟合分布图，第一行为信号区 1-3，第二行为信号区 4-6。

GOF test (SR1,2,3,4,5,6)

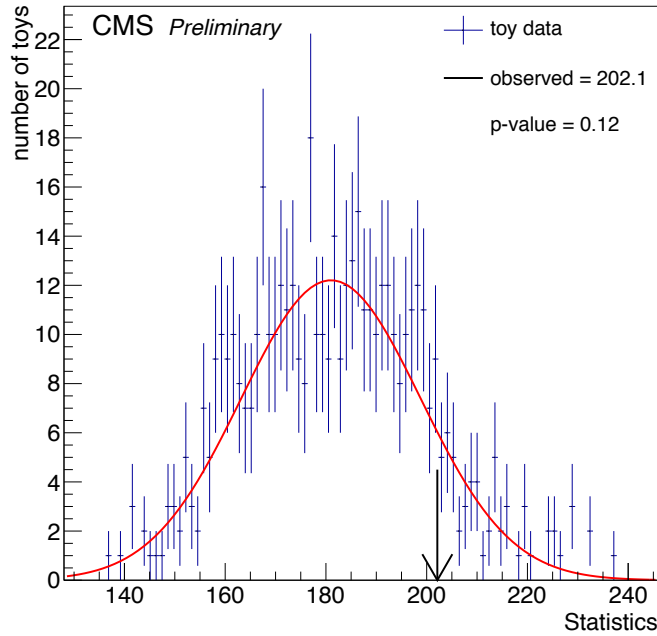


图 6.12 采用 Saturated Model 得到的统计检验结果分布。

6.4.3 信号注入偏离检验

在 CMS 统计检验中，通常要观测注入一定信号后拟合得到的信号强度是否相对注入值相差较大来判定是否存在偏离。在注入信号点，通过本底加信号拟合得到新的信号强度的分布，定义偏移量 $r_{bias} = \frac{r_{fit} - r_{injected}}{r_{Err}}$ 。固定 (3000,300)GeV 信号点，注入不同的信号强度，分别为 0.1、0.2、0.3、0.6、3 和 5，对应的偏移量 r_{bias} 分布如图6.13。在所有的图中中心值都小于 10%，成功通过检验。

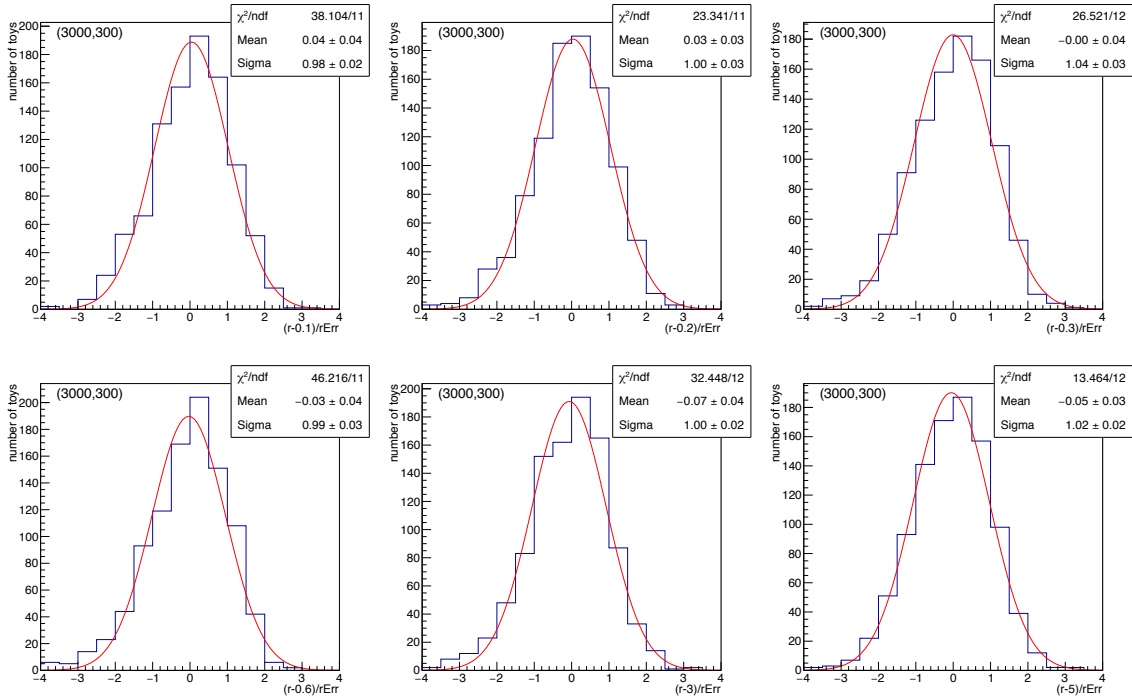


图 6.13 对 (3000,300)GeV 信号点，注入不同的信号强度，分别为 0.1、0.2、0.3、0.6、3 和 5 时偏移量分布

6.4.4 排除限

在拟合时，使用数据可以得到信号强度的上限。图6.14为 95% 置信度下信号强度 (signal strength) 的上限、预测的显著度及信号过程截面的上限 [86]。整个 2 维质量分布被分为 3 个区域，沿 $M_R \sim 0.2W_{KK}$ 线排除限略微下降，这是因为分析使用喷注尺寸参数 $R=0.8$ ，而这条线对应分立和合并 radion 情形转变的区域，造成敏感度的下降。在上方沿大 radion 质量有明显的分界线，这是由于信号截面在这条线上下有大约 30 倍的差异，另外，这种情形下，radion 衰变得到的 W 横动量较小，因此会被重建为单个的夸克喷注而非 W 喷注，因此不在信号区的接收范围内。在这些图中使用了线性插值的方法，具体如附录B.6。图6.15展示了部分冗余参数后拟合的分布。

与预测的排除限相比,观测的结果要弱 400GeV。这是由于在信号区 6 的后拟合图中 $M_{jj\ell\nu} = 3500 \pm 300 \text{ GeV}$ 处有连续的数据超出,因此合并后导致观测结果变差。图6.16展示了只合并信号区 1-5 的结果,不考虑 SR6 时观测得到的上限与预期相近。

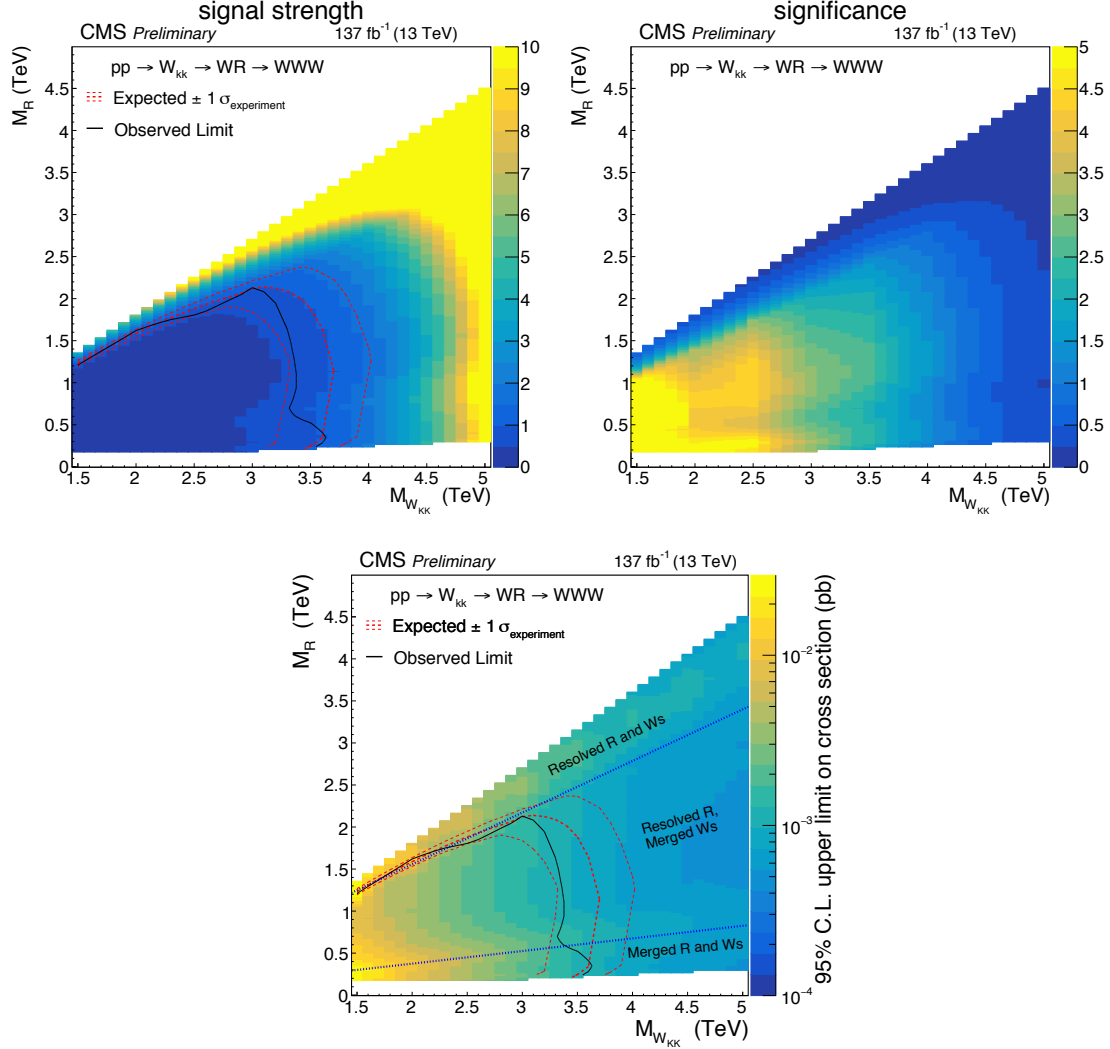


图 6.14 95% 置信度下二维平面 (W_{KK}, M_R) 上预期信号强度、显著度与信号截面的预期值和观测值。

除 W_{KK} 信号之外,还研究了 $Z_{KK} \rightarrow ZR \rightarrow ZWW$ 的信号过程,插值后的信号接收度、反应截面和 95% 置信水平的截面排除限如图6.17。由于 Z_{KK} 信号反应截面只有 W_{KK} 信号的 1/10,并且在合并信号区 2 和 3 接受度很低,因此只在 $Z_{KK}=1.5\text{TeV}$ 时预测能排除少数分立信号点,而数据不能排除任何信号点。

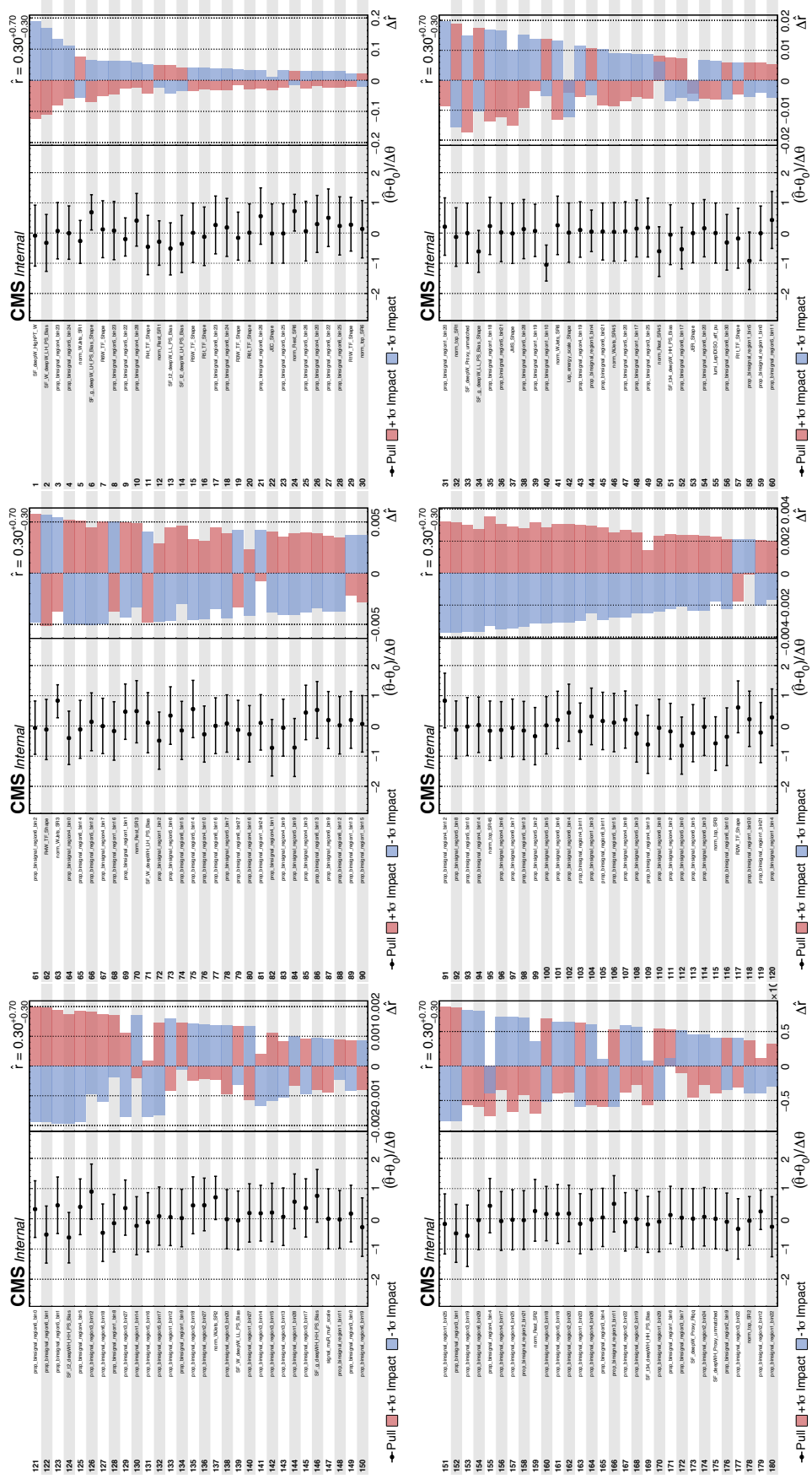


图 6.15 信号 (3000,300) 中冗余参数最终值相对预拟合值的误差及对信号强度的影响。

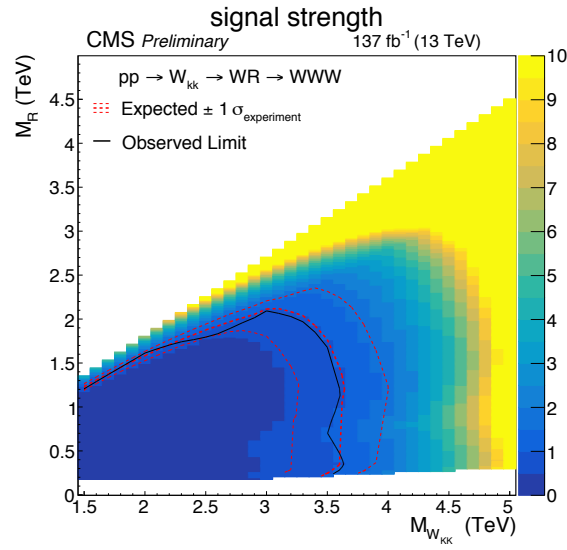
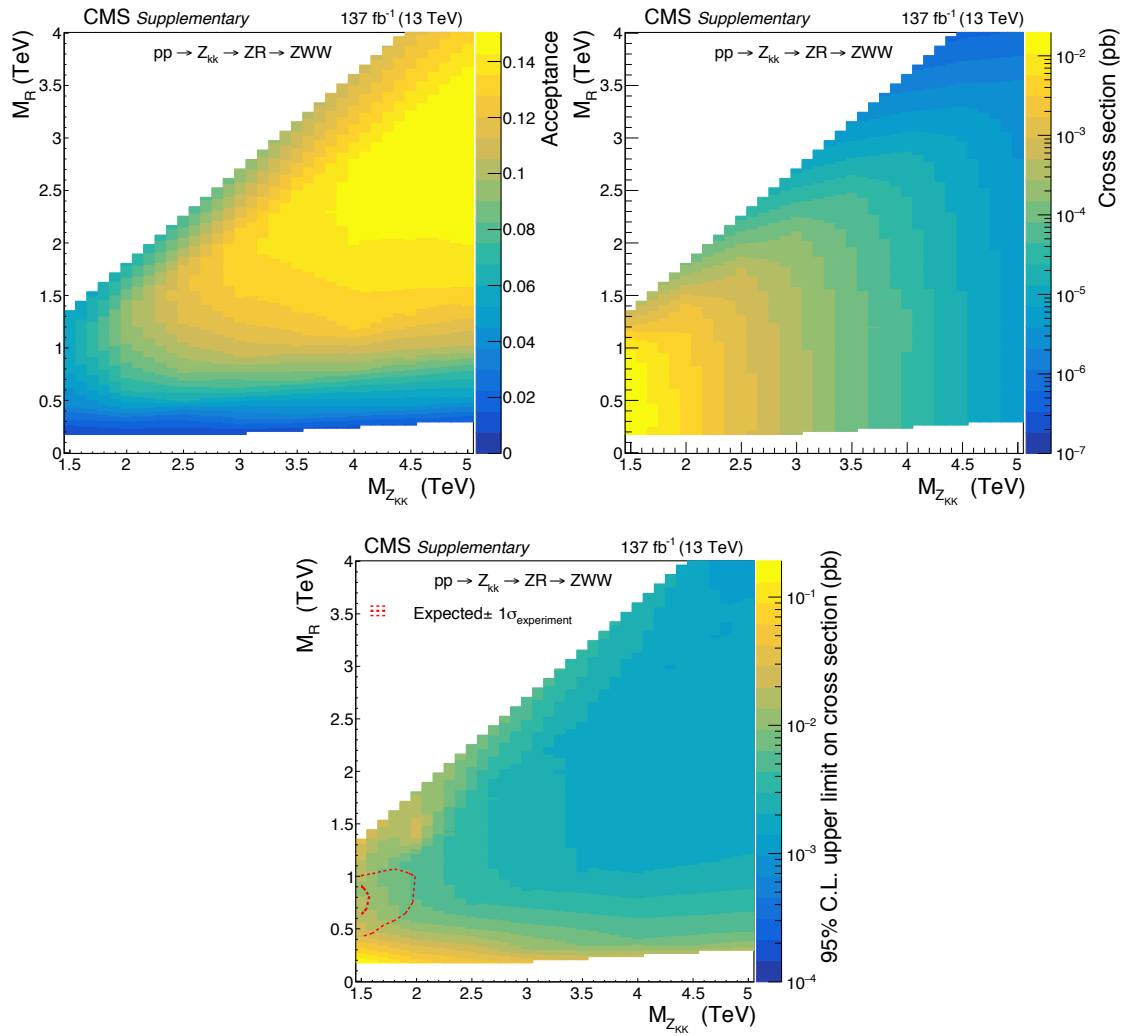


图 6.16 只合并信号区 1-5 时 95% 置信度下预期信号强度上限和排除限的分布。


 图 6.17 $Z_{KK} \rightarrow ZR \rightarrow ZWW$ 的信号接收度、反应截面和 95% 置信水平的截面排除限。

6.4.5 一维排除限

为了更加直观的展示排除限结果，固定 radion 的质量，只改变 W_{KK} 或 Z_{KK} 的质量得到的排除限结果如图6.18、6.19所示，预期排除限中心值都在 2 倍 σ 误差带内，对 W_{KK} 信号，数据分别对合并和分立情形排除上限为 3.5 和 3.3TeV；而对于 Z_{KK} 信号不能排除任何质量点。

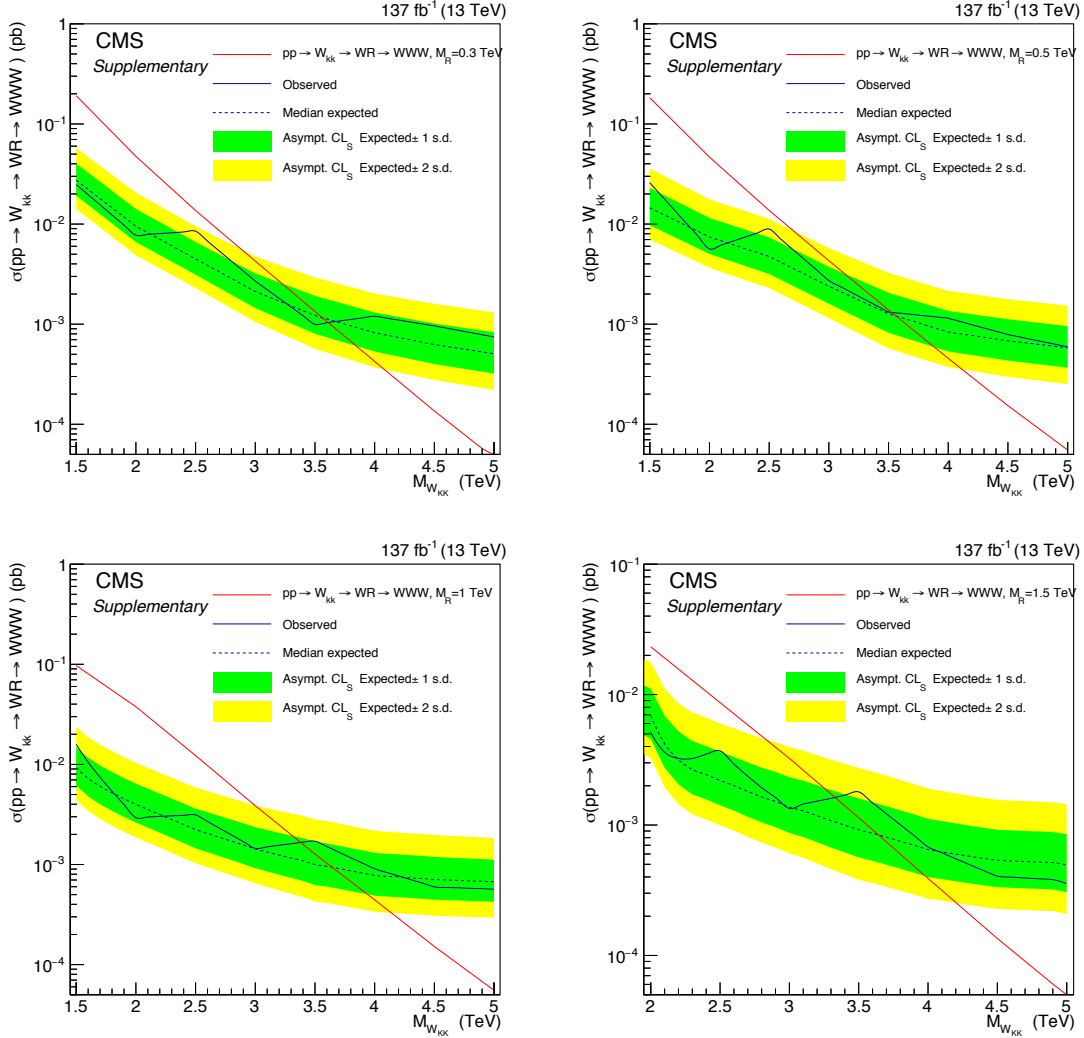


图 6.18 $W_{KK} \rightarrow WR \rightarrow WWW$ 的在 radion 质量为 0.3TeV、0.5TeV、1TeV 和 1.5TeV 时 95% 置信水平的截面排除限。

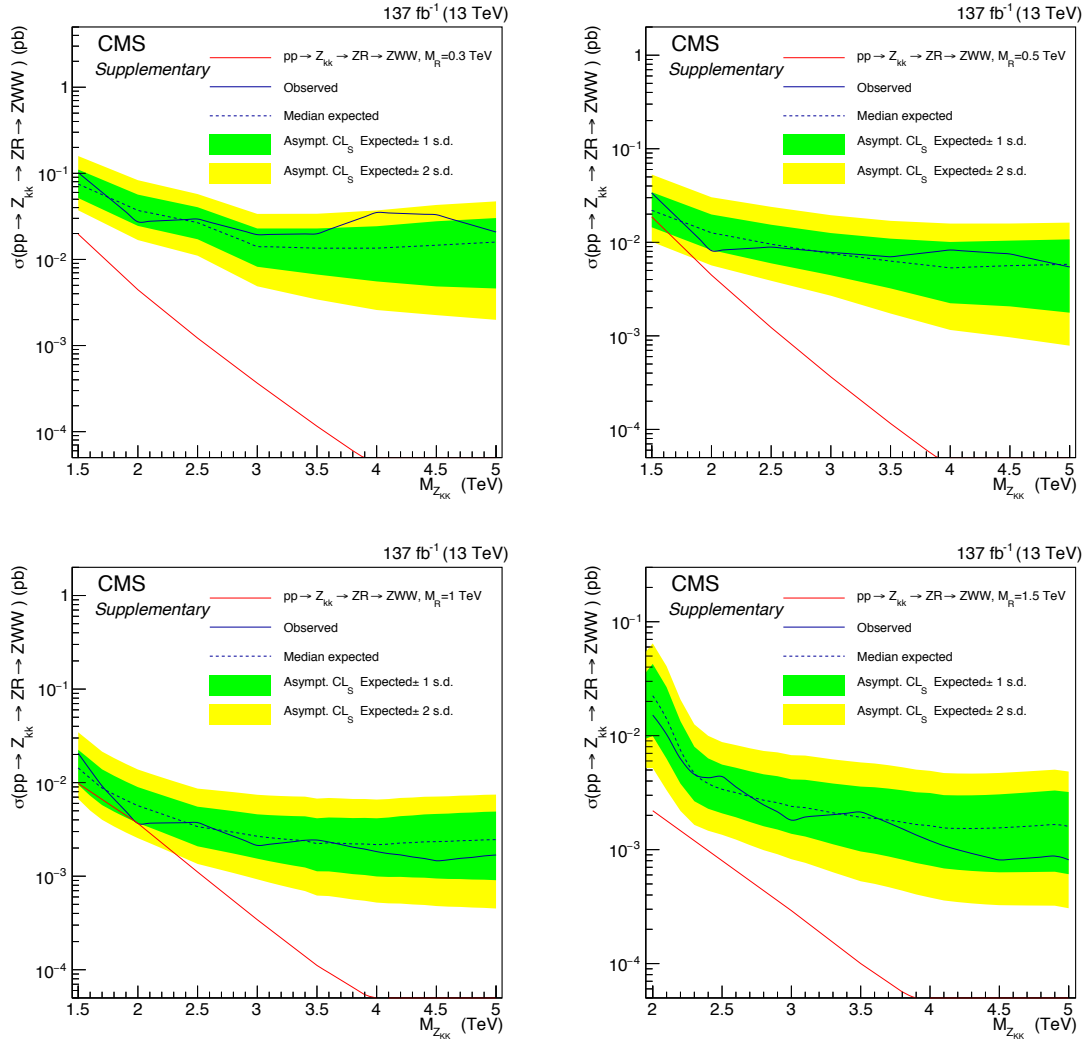


图 6.19 $Z_{KK} \rightarrow ZR \rightarrow ZWW$ 的在 radion 质量为 0.3TeV、0.5TeV、1TeV 和 1.5TeV 时 95% 置信水平的截面排除限。

第七章 蒙特卡洛 contact 工作

7.1 蒙卡样本的产生过程

CMS 为蒙特卡洛样本的模拟提供了一套中心产生系统，主要流程如图7.1[38] 所示，即样本产生链，主要包括以下几个过程：

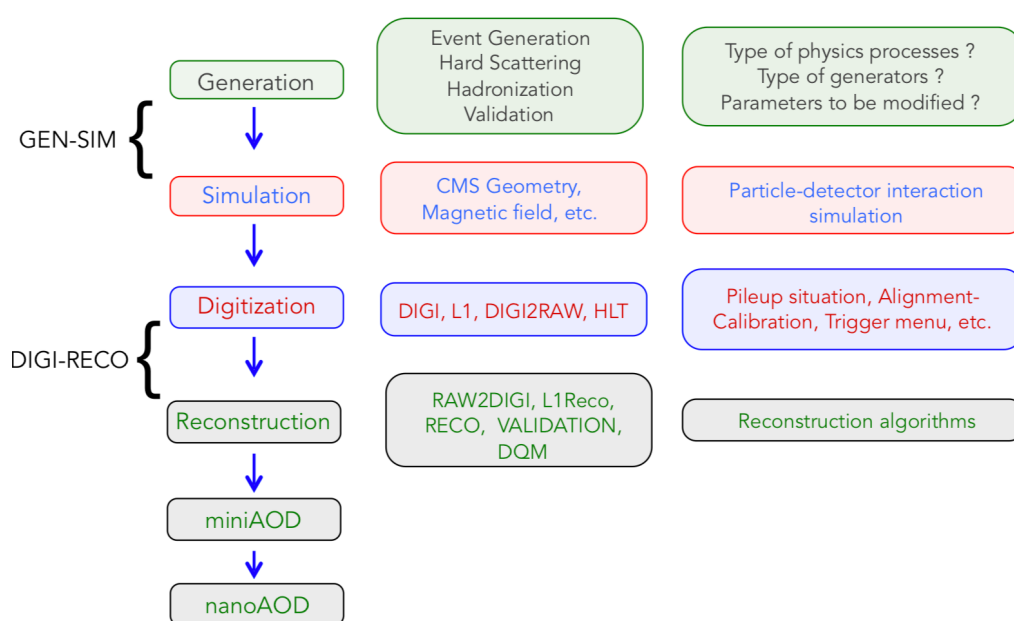


图 7.1 CMS 产生蒙卡模拟样本的流程。

GEN-SIM: 从蒙特卡洛产生子开始产生事例，并模拟粒子与探测器的相互作用。这一步的关键参数包括束流点、指定产生的过程（fragment）以及探测器的几何尺寸。其中常用产生子包括 Madgraph、Powheg 等产生硬散射过程，之后经过 Pythia 进行部分子簇射和强子化过程，所有粒子在探测器中的响应都是由 GEANT4[11] 进行全模拟。

DIGI-RECO: 将模拟的探测器信号数字化，并在此基础上采用重建算法。在此过程中，需要添加 pileup 方案，常用的方案包括经典混合（classic mix）、预混合（premix）[26]、水平分布 PU(FlatPU) 等。pileup 事例主要是软 QCD 过程，可以用 Pythia 产生子产生最小偏差事例来模拟，之后与要求的硬散射过程混合。经典混合中每个核心事例与大量 pileup 事例同时混合，需消耗大量的时间、CPU 和内存。预混合方案中，pileup 样本提前进行单独数字化，因此可以节约大约 5 倍的时间，而且极大降低了内存消耗。出于

特殊的需求,也可以产生 0 到 75 水平分布的 PU 样本。另外,触发信息包括 L1 和 HLT 也会作为这一步的输出。

MINIAOD[78]: CMS 实验事例中物理对象信息存储的一种格式。在 RunI 阶段,数据存储格式为 AOD (Analysis Object Data),每个事例占用 400kB 空间。在 RunII 阶段随着亮度的提升,CMS 开发了 MiniAOD 模块对经数字化重建后事例进行压缩以降低存储占用,与 AOD 记录事例尽可能多的方式不同,MINIAOD 是按物理对象分类如轻子、光子、喷注和丢失恒能量进行存储,并且通过加如横动量阈值等基础要求来减少事例数,每个事例占用空间减小为 35-60kB。该模块保存的物理对象的信息可直接用于物理分析。

NANOAOD: MINIAOD 事例被进一步删减得到纯粹的 root 文件,采用 TTree 的形式存储,每一个 entry 对应一个事例,可以在任何 CMSSW 环境之外中进行分析。

7.2 MC contact 的工作任务

蒙特卡样本产生的第一阶段 GEN-SIM 是由 MC contact 负责，样本的产生是基于 MCM 网站（Monte Carlo Management）网站，通常由所在的分析组提出产生样本的申请，之后 contact 需在 MCM 上提交 request 并进行设置，主要流程如图7.2[38] 所示。主要包括两类：

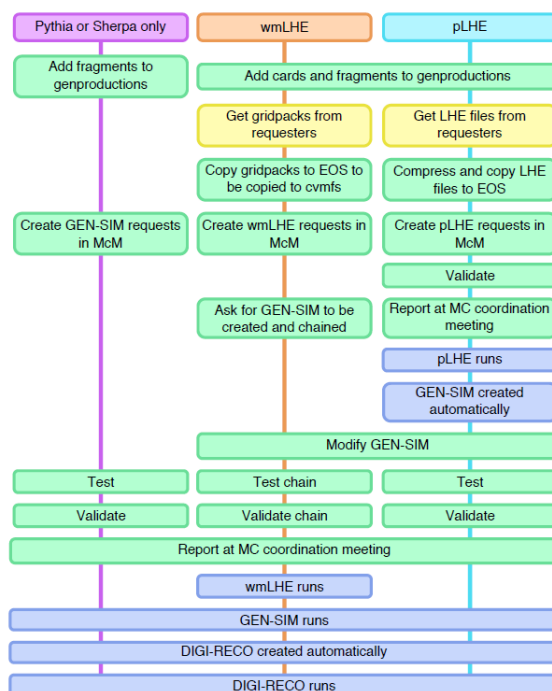


图 7.2 CMS contact 产生 request 的流程。

由 Pythia 或 Herwig 直接产生事例

对于指定的物理过程，需编辑 request 里 fragment 设置。图7.3为使用 Herwig7 产生子直接产生喷注横动量在 15 到 7000GeV 之间水平分布的 QCD 样本对应的 fragment 设置，其中 Herwig7CH3SettingsBlock、Herwig7PSWeightsSettingsBlock 和 Herwig7StableParticlesForDetectorBlock 分布为 Herwig7 产生子对应的 CH3 tune、部分子簇射权重和稳定粒子表模块，这三部分是默认设置。hw_user_settings 是对特定样本的设置，包括对质心系能量、费曼图水平及喷注横动量等参数的设置。

由 Madgraph 或 Powheg 等产生硬散射，然后接 Pythia 或 Herwig 进行部分子簇射

Fragment ②	<pre> generator = cms.EDFilter("Herwig7GeneratorFilter", herwig7CH3SettingsBlock, herwig7PSWeightsSettingsBlock, herwig7StableParticlesForDetectorBlock, configFiles = cms.vstring(), crossSection = cms.untracked.double(-1), dataLocation = cms.string('\${HERWIGPATH:-6}'), eventHandlers = cms.string('/Herwig/EventHandlers'), filterEfficiency = cms.untracked.double(1.0), generatorModule = cms.string('/Herwig/Generators/EventGenerator'), hw_user_settings = cms.vstring('cd /Herwig/EventHandlers', 'set EventHandler:LuminosityFunction:Energy 13000*GeV', 'cd /', 'read snippets/PPCollider.in', 'mkdir /Herwig/Weights', 'cd /Herwig/Weights', 'create ThePEG::ReweightMinPT reweightMinPT ReweightMinPT.so', 'cd /Herwig/MatrixElements/', 'insert SubProcess:MatrixElements[0] MEQCD2to2', 'insert SubProcess:Pweights[0] /Herwig/Weights/reweightMinPT', 'cd /') </pre>
---------------	---

图 7.3 Herwig7 产生子直接产生喷注横动量在 15 到 7000GeV 之间水平分布的 QCD 样本对应的 fragment。

这种事例的产生过程对应图7.2[38]的第二和第三列，首先需要产生感兴趣过程的卡片，包括过程卡片、运行卡片、参数卡片和模型卡片等。运行卡片中需设置运行事例数、样本精度（比如 LO 或 NLO）和对物理对象的基本选择条件，参数卡片用于修改不同于所用模型的参数。在确定卡片之后，使用官方 `genproduction`[35] 包产生 Gridpack。图7.4展示了 `madgrapMLM+Herwig7` 产生 QCD 样本的 fragment，需添加 Gridpack 路径和额外喷注的匹配设置。

设置 `request` 之后，需要本底运行来检查设置是否正确，同时记录过滤效率、匹配效率、平均每个事例时间和空间占用等信息。在提交 `request` 之后，要监测是否正常运行，如果运行异常需及时反馈和处理。

Fragment ②	<pre>import FWCore.ParameterSet.Config as cms externalLHEProducer = cms.EDProducer("ExternalLHEProducer", args = cms.vstring('/cvmfs/cms.cern.ch/phys_generator/gridpacks/2017/13TeV/madgraph/V5_2.6.5/QCD_HT_I nEvents = cms.untracked.uint32(5000), numberOfParameters = cms.uint32(1), outputFile = cms.string('cmsgrid_final.lhe'), scriptName = cms.FileInPath('GeneratorInterface/LHEInterface/data/run_generic_tarball_cvmfs.sh')) #Link to datacards: #https://github.com/cms-sw/genproductions/tree/master/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/ import FWCore.ParameterSet.Config as cms from Configuration.Generator.Herwig7Settings.Herwig7StableParticlesForDetector_cfi import * from Configuration.Generator.Herwig7Settings.Herwig7CH3TuneSettings_cfi import * from Configuration.Generator.Herwig7Settings.Herwig7PSWeightsSettings_cfi import * from Configuration.Generator.Herwig7Settings.Herwig7LHECommonSettings_cfi import *</pre>
Fragment ②	<pre>generator = cms.EDFilter("Herwig7GeneratorFilter", herwig7StableParticlesForDetectorBlock, herwig7CH3SettingsBlock, herwig7PSWeightsSettingsBlock, herwig7LHECommonSettingsBlock, configFiles = cms.vstring(), crossSection = cms.untracked.double(-1), dataLocation = cms.string('\${HERWIGPATH:-6}'), eventHandlers = cms.string('/Herwig/EventHandlers'), filterEfficiency = cms.untracked.double(1.0), generatorModule = cms.string('/Herwig/Generators/EventGenerator'), hw_user_settings = cms.vstring('cd /Herwig/EventHandlers', 'set EventHandler:LuminosityFunction:Energy 13000*GeV', 'cd /', 'read snippets/PPCollider.in', 'cd /Herwig/Shower', 'set FxFxHandler:njetsmax 4', 'set FxFxHandler:MergeMode TreeMG5', 'cd /', 'set /Herwig/Particles/h0.NominalMass 125.0'</pre>

图 7.4 使用 gridpack 产生样本对应 request 中 fragment 设置。

7.3 contact 工作中的创新

在 JME 和 B2G 组中，会有大量相似的样本，比如同一样本按不同的横动量划分为多个样本或新物理寻找中使用多个不同的质量点，如图7.5。此时可能需要产生成百上千相似的 request，如果在 MCM 网页手动修改是很消耗时间的。为了减少重复性工作，我采用两种不同的技术。

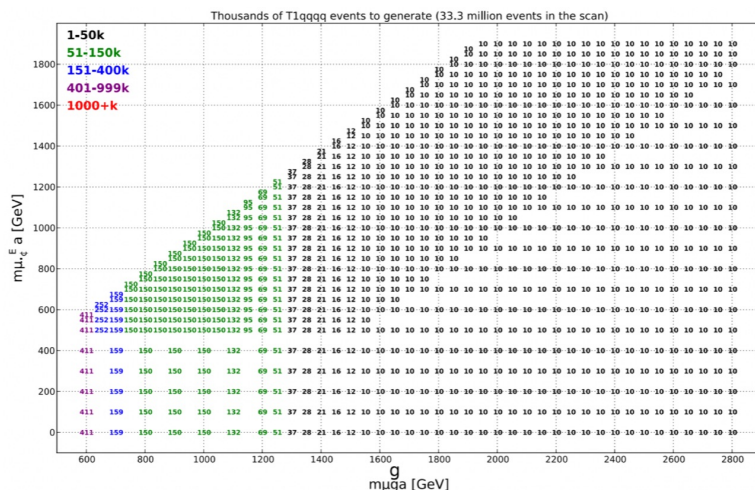


图 7.5 两个共振态质量在二维平面上信号点选取示例。

7.3.1 multi-gridpack

对于图7.5中的样本，这些样本除了质量不同外其他参数都相同，因此没必要使用传统的 1 个 Gridpack 对于 1 个 request 的形式，可以整合多个 Gridpack 产生一个混合的大的样本，fragment 设置如图7.6[70]。使用循环在一个 request 中包含多个 Gridpack，并用不同的字符串 ConfigDescription 来区分不同的过程，processParameters 可以按需求设置粒子的衰变模式，图中要求 Higgs 衰变到轻夸克对、 $b\bar{b}$ 和 $c\bar{c}$ 各占 1/3。虽然此技术有很大的优势，但是无法使用图7.4中 LHE 模块，LHE 输出中包含 QCD scale 和不同 pdf 的权重，如果采用这项技术产生样本则无法计算这两项系统误差，因此该技术适用于不需要单独计算 pdf 误差的情形。

7.3.2 自动化产生 request 脚本

另一个简化工作量的方法是使用脚本来批量修改网页内容，图7.7是其中一个例子，可以对不同年份、质量点循环，另外可以设置所有的参数，比如线程数、内存、每个事

```

for point in mpoints:
    generator.RandomizedParameters.append(
        cms.PSet(
            ConfigWeight = cms.double(1),
            GridpackPath =
cms.string('/cvmfs/cms.cern.ch/phys_generator/gridpacks/2017/13TeV/madgraph/V5_2.6.5/BulkG
all.tar.xz' % (model,point[0], point[1])),
            ConfigDescription = cms.string('%s_MX%s_MH%s' % (model, point[0], point[1])),
            PythiaParameters = cms.PSet(
                pythia8CommonSettingsBlock,
                pythia8CP5SettingsBlock,
                pythia8PSWeightsSettingsBlock,
                processParameters = cms.vstring('25:onMode = off',
                                                '25:oneChannel = 1 0.33333 100 5 -5',
                                                '25:addChannel = 1 0.33333 100 4 -4',
                                                '25:addChannel = 1 0.11111 100 3 -3',
                                                '25:addChannel = 1 0.11111 100 2 -2',
                                                '25:addChannel = 1 0.11111 100 1 -1',
                                                'ResonanceDecayFilter:filter = on'),
            ),
            parameterSets = cms.vstring('pythia8CommonSettings',
                                        'pythia8CP5Settings',
                                        'pythia8PSWeightsSettings',
                                        'processParameters',

```

图 7.6 采用 multi-gridpack 技术设置 fragment。

例需要存储空间和时间等，最后使用 update 函数在网页端更新所有的修改。除此之外，request 的批量删除、提交等操作都可以通过脚本形式实现，具有很强的通用性。如今，B2G 分析组采用了脚本的方法，避免重复性的操作，节约了大量的时间。

```

for year in ["16","17","18"]:
    for im in range(len(MXWX)):
        request=mcm.get('requests',range_of_requests[0]);
        if year in ["16"]:request['member_of_campaign']="RunIISummer%swmLHEGS"%(15)
        if year in ["17","18"]:request['member_of_campaign']="RunIIFall%swmLHEGS"%(year)
        clone_answerid = mcm.clone_request(request)
        clone_answer = mcm.get('requests', clone_answerid['prepid'])
        clone_answer['dataset_name']="Wprimetotb_%s_%s_TuneCP5_13TeV-madgraph-pythia8"%(MXWX[im],pro[im])
        clone_answer['fragment']=clone_answer['fragment'].replace('/afs/cern.ch/work/a/agapitos/public/MC4/Au
fs/cern.ch/user/x/xulyu/work/public/mcrequest/Wprimetotbnew.py').replace("exo_diboson/Spin-0/Radion_hh_htatahbb_i
0_slc7_amd64_gcc700_CMSSW_10_6_19_tarball.tar.xz","Wprimetotb/Wprimetotb_%s_%s.tar.xz"%(MXWX[im],pro[im])).replac
/tree/master/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/exo_diboson/Spin-0/Radion_hh_htatahbb_incl_narrow
ee/master/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/Wprimetotb')
        if year=="16":clone_answer['fragment']=clone_answer['fragment'].replace("from Configuration.Generator
import *","")
        if year=="16":
            clone_answer['time_event'][0]= 70
            clone_answer['size_event'][0]= 975
        if year=="17":
            clone_answer['time_event'][0]= 50
            clone_answer['size_event'][0]= 990
        if year=="18":
            clone_answer['time_event'][0]= 40
            clone_answer['size_event'][0]= 950
        clone_answer['memory']=2300
        clone_answer['sequences'][0]['nThreads']=1

```

图 7.7 自动化产生 request 的部分脚本。

第八章 总结与展望

本文设计了寻找 $W_{KK} \rightarrow RW \rightarrow WWW$ 信号的末态为三玻色子过程，并且可以同时探测合并和分立 radion 喷注情形的分析。玻色子标记采用了近三年开发的基于深度学习的 deepAK8 技术，并且开发了全新的技术用于标记值校准。针对合并情形的复杂情况，通过喷注与产生子水平粒子匹配确定喷注的结构，基于喷注的质量和横动量对不同标记值进行修正。在修正因子 SFs 的计算中使用到的新的矩阵算法的思路可以应用于其他分析过程，其中一个要点是将样本分为多个不同类型喷注占主导的子样本，将不同类喷注按比例缩放后与数据匹配，相关系数可列为矩阵，最终转化为矩阵的求解；另一个要点是提供了含四夸克的胖喷注的代理对象，标准模型中很难找到统计量较高的同类喷注用于计算 SFs，本文在全合并高能 top 的即三夸克的喷注基础上，额外要求一个不是来自于 top 或 W 的夸克或胶子在喷注锥角内从而组成四夸克结构，这样新的代理可以用于研究新物理中的新粒子或标准模型的 Higgs，比如高能 Higgs 衰变到 2 个 W 继而到 4 个夸克，与本文中的 radion 即为相似，可以说这种 SFs 的计算方法有很大的应用前景。另外，也提供了这种方法的计算系统误差的方法，具有很强的推广性和借鉴意义。最终信号区域被分为 6 个正交的样本，具有互补的敏感度，对于分立和合并情形分别用 AK8 喷注数量为 1 或 2 对应观测量 $M_{j\ell\nu}$ 和 $M_{jj\ell\nu}$ 来揭示三玻色子的共振态。对于本底的估计使用了数据驱动的方法，为主要的本底 W+jets 和 top 分别设计了和信号区运动学相似的控制区来修正本底的事例数并估算误差。得到本底和信号和数据观测量形状并估计系统误差后，使用 CLs 的方法限制冗余参数得到后拟合分布从而预测本底过程。遗憾的是数据和预测的本底吻合较好，并没有发现新物理的迹象。尽管在信号区 6 有一定的数据超出，观测的显著度接近 3σ ，但是需要更高统计量样本的检验。最终 W_{KK} 排除限在分立信号处约为 3.3TeV，在合并信号处达到了 3.5TeV，这是 LHC 上第一次对三个重玻色子的共振态的寻找，并且设置了排除限，在未来 LHC 升级后更高亮度下的结果值得期待。

除了 W_{KK} 信号之外，其它过程也可能会有有一定的敏感度，比如 $Z_{KK} \rightarrow RZ \rightarrow ZWW$ 过程。在信号区 1 中可以高效探测 W 或 Z 玻色子，文中也给出了相应的研究结果，由于 Z_{KK} 过程的产生截面只有 W_{KK} 的 1/10，因此 Z_{KK} 信号给出的预期排除限很低，使用数据不能排除任何质量点。另外，其他的信号过程，比如 $X \rightarrow tW \rightarrow bq\bar{q}, W$ 和 $X \rightarrow WH \rightarrow W+4q$ 会在信号区 2 有一定的分布，预期会有较高的敏感度，是后续分析的一个方向。目前使用的 deepAK8 标记的深度学习是卷积神经网络，最新有使用 particleNET[81] 技术得到有更强分辨能力的鉴别器，可以进一步提高敏感度。本文所研

究的过程还有计算 SFs 的矩阵算法可以延伸到更多的分析，比如 boosted $H \rightarrow WW$ 、 $X \rightarrow HW \rightarrow WWW$ 、 $HH \rightarrow WWbb$ 等，该多夸克探针技术会有广泛的应用前景。

参考文献

- [1] M. Aaboud, G. Aad, B. Abbott *et al.* “Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 35.9fb^{-1} of proton-proton collision data at $\sqrt{s}=13\text{TeV}$ with the ATLAS detector”. *Physical Review D*, 2018-09. <http://dx.doi.org/10.1103/PhysRevD.98.052008>.
- [2] P. e. A. Adzic. *Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter* [techreport]. Geneva, 2006-11. <http://cds.cern.ch/record/1009081>.
- [3] K. Agashe, J. H. Collins, P. Du *et al.* “Dedicated strategies for triboson signals from cascade decays of vector resonances”. *Physical Review D*, 2019-04. <http://dx.doi.org/10.1103/PhysRevD.99.075016>.
- [4] K. Agashe, J. H. Collins, P. Du *et al.* “Dedicated strategies for triboson signals from cascade decays of vector resonances”. *Physical Review D*, 2019-04. <http://dx.doi.org/10.1103/PhysRevD.99.075016>.
- [5] K. Agashe, J. H. Collins, P. Du *et al.* “Detecting a boosted diboson resonance”. *Journal of High Energy Physics*, 2018-11. [http://dx.doi.org/10.1007/JHEP11\(2018\)027](http://dx.doi.org/10.1007/JHEP11(2018)027).
- [6] K. Agashe, H. Davoudiasl, G. Perez *et al.* “Warped gravitons at the CERN LHC and beyond”. *Phys. Rev. D*, 2007-08: 036006. <https://link.aps.org/doi/10.1103/PhysRevD.76.036006>.
- [7] K. Agashe, H. Davoudiasl, G. Perez *et al.* “Warped gravitons at the CERN LHC and beyond”. *Physical Review D*, 2007-08. <http://dx.doi.org/10.1103/PhysRevD.76.036006>.
- [8] K. Agashe, P. Du, S. Hong *et al.* “Flavor universal resonances and warped gravity”. *Journal of High Energy Physics*, 2017-01. [http://dx.doi.org/10.1007/JHEP01\(2017\)016](http://dx.doi.org/10.1007/JHEP01(2017)016).
- [9] K. S. Agashe, J. H. Collins, P. Du *et al.* “LHC signals from cascade decays of warped vector resonances”. *Journal of High Energy Physics*, 2017-05. [http://dx.doi.org/10.1007/JHEP05\(2017\)078](http://dx.doi.org/10.1007/JHEP05(2017)078).
- [10] K. S. Agashe, J. H. Collins, P. Du *et al.* “LHC signals from cascade decays of warped vector resonances”. *Journal of High Energy Physics*, 2017-05. [http://dx.doi.org/10.1007/JHEP05\(2017\)078](http://dx.doi.org/10.1007/JHEP05(2017)078).
- [11] S. Agostinelli *et al.* “GEANT4—a simulation toolkikt. GEANT4. A Simulation toolkit”. *Nucl. Instrum. Methods Phys. Res., A*, 2002-07: 250–303. 54 p. <https://cds.cern.ch/record/602040>.
- [12] J. A. Aguilar-Saavedra. “Profile of multiboson signals”. *Journal of High Energy Physics*, 2017-05. [http://dx.doi.org/10.1007/JHEP05\(2017\)066](http://dx.doi.org/10.1007/JHEP05(2017)066).
- [13] A. S. et. Al. “Particle-flow reconstruction and global event description with the CMS detector”. *Journal of Instrumentation*, 2017-10: P10003–P10003. <https://doi.org/10.1088/1748-0221/12/10/p10003>.

- [14] B. A. et al. “*Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*”. *Phys. Rev. Lett.* 2021-04: 141801. <https://link.aps.org/doi/10.1103/PhysRevLett.126.141801>.
- [15] G. W. B. et al. “*Final report of the E821 muon anomalous magnetic moment measurement at BNL*”. *Phys. Rev. D*, 2006-04: 072003. <https://link.aps.org/doi/10.1103/PhysRevD.73.072003>.
- [16] S. Alioli, P. Nason, C. Oleari et al. “*A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*”. *Journal of High Energy Physics*, 2010-06. [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043).
- [17] J. Alwall, R. Frederix, S. Frixione et al. “*The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*”. *Journal of High Energy Physics*, 2014-07. [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- [18] T. Aoyama, M. Hayakawa, T. Kinoshita et al. “*Complete Tenth-Order QED Contribution to the Muon $g-2$* ”. *Phys. Rev. Lett.* 2012-09: 111808. <https://link.aps.org/doi/10.1103/PhysRevLett.109.111808>.
- [19] R. D. Ball, V. Bertone, S. Carrazza et al. “*Parton distributions from high-precision collider data*”. *The European Physical Journal C*, 2017-10. <http://dx.doi.org/10.1140/epjc/s10052-017-5199-5>.
- [20] V. Barger, W. Y. Keung and E. Ma. “*Gauge model with light W and Z bosons*”. *Phys. Rev. D*, 1980-08: 727–737. <https://link.aps.org/doi/10.1103/PhysRevD.22.727>.
- [21] D. Bertolini, P. Harris, M. Low et al. “*Pileup per particle identification*”. *Journal of High Energy Physics*, **2014**, 2014(10). [http://dx.doi.org/10.1007/JHEP10\(2014\)059](http://dx.doi.org/10.1007/JHEP10(2014)059).
- [22] P. Billoir. “*Progressive track recognition with a Kalman-like fitting procedure*”. *Computer Physics Communications*, **1989**, 57(1): 390–394. <https://www.sciencedirect.com/science/article/pii/001046558990249X>.
- [23] J. Butterworth, S. Carrazza, A. Cooper-Sarkar et al. “*PDF4LHC recommendations for LHC Run II*”. *Journal of Physics G: Nuclear and Particle Physics*, 2016-01: 023001. <http://dx.doi.org/10.1088/0954-3899/43/2/023001>.
- [24] J. M. Butterworth, A. R. Davison, M. Rubin et al. “*Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider*”. *Physical Review Letters*, 2008-06. <http://dx.doi.org/10.1103/PhysRevLett.100.242001>.
- [25] M. Cacciari, G. P. Salam and G. Soyez. “*The anti- k_T jet clustering algorithm*”. *Journal of High Energy Physics*, 2008-04: 063–063. <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [26] CMS. *A New Pileup Mixing Framework for CMS*, **2021** (accessed February 20, 2021). <https://indico.cern.ch/event/304944/contributions/1672382/attachments/578591/796743/CHEP2015-premixing.pdf>.
- [27] CMS. *CMS Luminosity public results*, **2021** (accessed February 20, 2021). https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots.

-
- [28] CMS. *Common Statistical Methods*, **2021 (accessed February 20, 2021)**. <https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part3/commonstatmethods/>.
- [29] CMS. *Current recommendations for luminosity estimations*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM#CurRec>.
- [30] CMS. *Electron efficiencies and scale factors*, **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Efficiencies%5C_and%5C_scale%5C_factors.
- [31] CMS. *Event reweighting using scale factors and MC b-tagging efficiencies*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods>.
- [32] CMS. *HEEP v7.0 electrons*, **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/viewauth/CMS/HEEPElectronIdentificationRun2#Selection%5C_Cuts%5C_HEEP%5C_V7%5C_0%5C_recomme.
- [33] CMS. *HEM15/16 issue in jets with $-3.2 < \eta < -1.3$ and $-1.57 < \phi < -0.87$* , **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMET#Quick_links_to_current_recommend.
- [34] CMS. *High-pt muons*, **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#HighPt%5C_Muon.
- [35] CMS. *How to produce gridpacks*, **2021 (accessed February 20, 2021)**. <https://github.com/cms-sw/genproductions/tree/master/bin>.
- [36] CMS. *Jet identification*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/view/CMS/JetID>.
- [37] CMS. *L1 Prefiring*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe>.
- [38] CMS. *Monte Carlo Management (McM): introduction*, **2021 (accessed February 20, 2021)**. <https://cms-pdmv.gitbook.io/project/monte-carlo-management-mcm-introduction>.
- [39] CMS. *Noise filter*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe>.
- [40] CMS. *Recommendation for Using b-tag Objects in Physics Analyses*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation>.
- [41] CMS. *Recommended cross section*, **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData#Recommended_cross_section.
- [42] CMS. *Recommended Jet Energy Corrections*, **2021 (accessed February 20, 2021)**. <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- [43] CMS. *Smearing procedures*, **2021 (accessed February 20, 2021)**. https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#Smearing_procedures.
- [44] T. C. Collaboration. “The CMS experiment at the CERN LHC”. *Journal of Instrumentation*, 2008-08: S08004–S08004. <https://doi.org/10.1088/1748-0221/3/08/s08004>.

- [45] T. C. collaboration. “*Determination of jet energy calibration and transverse momentum resolution in CMS*”. *Journal of Instrumentation*, 2011-11: P11002–P11002. <http://dx.doi.org/10.1088/1748-0221/6/11/P11002>.
- [46] T. C. collaboration. “*Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*”. *Journal of Instrumentation*, 2012-10: P10002–P10002. <https://doi.org/10.1088/1748-0221/7/10/p10002>.
- [47] R. Cousins. *Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models with Application to Histograms*, 2013. http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.
- [48] C. Csáki, A. Falkowski and A. Weiler. “*The flavor of the composite pseudo-goldstone Higgs*”. *Journal of High Energy Physics*, 2008-09: 008–008. <http://dx.doi.org/10.1088/1126-6708/2008/09/008>.
- [49] M. Dasgupta, A. Fregoso, S. Marzani *et al.* “*Towards an understanding of jet substructure*”. *Journal of High Energy Physics*, 2013-09. [http://dx.doi.org/10.1007/JHEP09\(2013\)029](http://dx.doi.org/10.1007/JHEP09(2013)029).
- [50] Y. Dokshitzer, G. Leder, S. Moretti *et al.* “*Better jet clustering algorithms*”. *Journal of High Energy Physics*, 1997-08: 001–001. <http://dx.doi.org/10.1088/1126-6708/1997/08/001>.
- [51] S. D. Ellis and D. E. Soper. “*Successive combination jet algorithm for hadron collisions*”. *Physical Review D*, 1993-10: 3160–3166. <http://dx.doi.org/10.1103/PhysRevD.48.3160>.
- [52] M. Gell-Mann. “*Symmetries of Baryons and Mesons*”. *Phys. Rev.* 1962-02: 1067–1084. <https://link.aps.org/doi/10.1103/PhysRev.125.1067>.
- [53] C. Gnendiger, D. Stöckinger and H. Stöckinger-Kim. “*The electroweak contributions to $(g-2)_\mu$ after the Higgs-boson mass measurement*”. *Phys. Rev. D*, 2013-09: 053005. <https://link.aps.org/doi/10.1103/PhysRevD.88.053005>.
- [54] J. Goldstone. “*Field theories with "superconductor" solutions*”. *Nuovo Cimento*, 1960-08: 154–164. <http://cds.cern.ch/record/343400>.
- [55] D. J. Griffiths. “*Introduction to elementary particles; 2nd rev. version*”. 2008. <https://cds.cern.ch/record/111880>.
- [56] C. Grojean, E. Salvioni and R. Torre. “*A weakly constrained W at the early LHC*”. *Journal of High Energy Physics*, 2011-07. [http://dx.doi.org/10.1007/JHEP07\(2011\)002](http://dx.doi.org/10.1007/JHEP07(2011)002).
- [57] B. Keren-Zur, P. Lodone, M. Nardecchia *et al.* “*On Partial Compositeness and the CP asymmetry in charm decays*”. *Nuclear Physics B*, 2013-02: 394–428. <http://dx.doi.org/10.1016/j.nuclphysb.2012.10.012>.
- [58] A. Keshavarzi, D. Nomura and T. Teubner. “ *$g - 2$ of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium*”. *Phys. Rev. D*, 2020-01: 014029. <https://link.aps.org/doi/10.1103/PhysRevD.101.014029>.
- [59] V. Khachatryan, A. M. Sirunyan, A. Tumasyan *et al.* “*Event generator tunes obtained from underlying event and multiparton scattering measurements*”. *The European Physical Journal C*, 2016-03. <http://dx.doi.org/10.1140/epjc/s10052-016-3988-x>.

- [60] V. Khachatryan, A. M. Sirunyan, A. Tumasyan *et al.* “Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7\text{TeV}$ ”. *Journal of High Energy Physics*, 2011-01. [http://dx.doi.org/10.1007/JHEP01\(2011\)080](http://dx.doi.org/10.1007/JHEP01(2011)080).
- [61] M. S. Kim. “CMS reconstruction improvement for the muon tracking by the RPC chambers”. *PoS*, 2012-09: 045. 9 p. <https://cds.cern.ch/record/1477844>.
- [62] R. Kogler, B. Nachman, A. Schmidt *et al.* “Jet substructure at the Large Hadron Collider”. *Reviews of Modern Physics*, 2019-12. <http://dx.doi.org/10.1103/RevModPhys.91.045003>.
- [63] D. Krohn, J. Thaler and L.-T. Wang. “Jet trimming”. *Journal of High Energy Physics*, 2010-02. [http://dx.doi.org/10.1007/JHEP02\(2010\)084](http://dx.doi.org/10.1007/JHEP02(2010)084).
- [64] Y.-P. Kuang, H.-Y. Ren and L.-H. Xia. “Model-independent probe of anomalous heavy neutral Higgs bosons at the LHC”. *Physical Review D*, 2014-12. <http://dx.doi.org/10.1103/PhysRevD.90.115002>.
- [65] A. Kurz, T. Liu, P. Marquard *et al.* “Hadronic contribution to the muon anomalous magnetic moment to next-to-next-to-leading order”. *Physics Letters B*, **2014**, 734: 144–147. <https://www.sciencedirect.com/science/article/pii/S0370269314003529>.
- [66] P. Kusch and H. M. Foley. “The Magnetic Moment of the Electron”. *Phys. Rev.* 1948-08: 250–263. <https://link.aps.org/doi/10.1103/PhysRev.74.250>.
- [67] K. Lane and L. Pritchett. “The light composite Higgs boson in strong extended technicolor”. *Journal of High Energy Physics*, 2017-06. [http://dx.doi.org/10.1007/JHEP06\(2017\)140](http://dx.doi.org/10.1007/JHEP06(2017)140).
- [68] A. J. Larkoski, S. Marzani, G. Soyez *et al.* “Soft drop”. *Journal of High Energy Physics*, 2014-05. [http://dx.doi.org/10.1007/JHEP05\(2014\)146](http://dx.doi.org/10.1007/JHEP05(2014)146).
- [69] H. Liu, L. Yu and M. Huang. “Survival of charged condensation at high temperature and density”. *Chinese Physics C*, 2016-02: 023102. <http://dx.doi.org/10.1088/1674-1137/40/2/023102>.
- [70] X. Lyu. *New multi-gridpack workflow*, **2019**. https://indico.cern.ch/event/783811/contributions/3458071/attachments/1864431/3065252/GEN_6.18.pdf.
- [71] A. D. Martin, W. J. Stirling, R. S. Thorne *et al.* “Parton distributions for the LHC”. *The European Physical Journal C*, 2009-07: 189–285. <http://dx.doi.org/10.1140/epjc/s10052-009-1072-5>.
- [72] S. P. MARTIN. “A SUPERSYMMETRY PRIMER”. *Advanced Series on Directions in High Energy Physics*, 1998-07: 1–98. http://dx.doi.org/10.1142/9789812839657_0001.
- [73] D. Marzocca, M. Serone and J. Shu. “General composite Higgs models”. *Journal of High Energy Physics*, 2012-08. [http://dx.doi.org/10.1007/JHEP08\(2012\)013](http://dx.doi.org/10.1007/JHEP08(2012)013).
- [74] E. Mobs. “The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018”. 2018-08. <https://cds.cern.ch/record/2636343>.
- [75] L. G. Pantelis Kontaxakis Huilin Qu and *et. Al.* *DeepAK8 Top/W WPs | SFs for V2 Training*, **2020**. https://indico.cern.ch/event/877167/contributions/3744193/attachments/1989744/3379280/DeepAK8_Top_W_SF_V2.pdf.

- [76] M. Perelstein. “*Little Higgs models and their phenomenology*”. *Progress in Particle and Nuclear Physics*, 2007-01: 247–291. <http://dx.doi.org/10.1016/j.pnpnp.2006.04.001>.
- [77] M. E. Peskin and D. V. Schroeder. “*An introduction to quantum field theory*”. **1995**. <https://cds.cern.ch/record/257493>.
- [78] G. Petrucciani, A. Rizzi and C. Vuosalo. “*Mini-AOD: A New Analysis Data Format for CMS*”. *Journal of Physics: Conference Series*, 2015-12: 072052. <http://dx.doi.org/10.1088/1742-6596/664/7/072052>.
- [79] “*Pileup Removal Algorithms*”. **2014**, (CMS-PAS-JME-14-001). <http://cds.cern.ch/record/1751454>.
- [80] M. B. Popovic. “*The Standard Model hierarchy, fine-tuning, and negativity of the Higgs mass squared*”. 2002-04. <https://cds.cern.ch/record/549513>.
- [81] H. Qu and L. Gouskos. “*Jet tagging via particle clouds*”. *Physical Review D*, 2020-03. <http://dx.doi.org/10.1103/PhysRevD.101.056019>.
- [82] L. Randall and R. Sundrum. “*Large Mass Hierarchy from a Small Extra Dimension*”. *Phys. Rev. Lett.* 1999-10: 3370–3373. <https://link.aps.org/doi/10.1103/PhysRevLett.83.3370>.
- [83] M. Robinson, K. Bland, G. Cleaver *et al.* *A Simple Introduction to Particle Physics*, **2008**.
- [84] E. Salvioni, G. Villadoro and F. Zwirner. “*MinimalZ models: present bounds and early LHC reach*”. *Journal of High Energy Physics*, 2009-11: 068–068. <http://dx.doi.org/10.1088/1126-6708/2009/11/068>.
- [85] J. Schwinger. “*On Quantum-Electrodynamics and the Magnetic Moment of the Electron*”. *Phys. Rev.* 1948-02: 416–417. <https://link.aps.org/doi/10.1103/PhysRev.73.416>.
- [86] “*Search for resonances decaying to triple W-boson final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*”. **2021**, (CMS-PAS-B2G-20-001). <https://cds.cern.ch/record/2759857>.
- [87] S. Sharma. “*Understanding the performance of CMS calorimeter*”. *Pramana*, 2007-12: 1069. <https://ui.adsabs.harvard.edu/abs/2007Prma..69.1069S>.
- [88] A. Sirunyan, A. Tumasyan, W. Adam *et al.* “*Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons*”. *Physics Letters B*, 2019-11: 134952. <http://dx.doi.org/10.1016/j.physletb.2019.134952>.
- [89] A. Sirunyan, A. Tumasyan, W. Adam *et al.* “*Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*”. *Journal of Instrumentation*, 2020-06: P06005–P06005. <http://dx.doi.org/10.1088/1748-0221/15/06/P06005>.
- [90] A. Sirunyan, A. Tumasyan, W. Adam *et al.* “*Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton–proton collisions at $\sqrt{s}=13$ TeV*”. *Physics Letters B*, 2018-06: 244–269. <http://dx.doi.org/10.1016/j.physletb.2018.03.084>.
- [91] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “*Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements*”. *The European Physical Journal C*, 2020-01. <http://dx.doi.org/10.1140/epjc/s10052-019-7499-4>.

-
- [92] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for a heavy resonance decaying into a Z boson and a vector boson in the $\nu\bar{\nu}q\bar{q}$ final state”. *Journal of High Energy Physics*, 2018-07. [http://dx.doi.org/10.1007/JHEP07\(2018\)075](http://dx.doi.org/10.1007/JHEP07(2018)075).
- [93] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for a heavy resonance decaying into a Z boson and a Z or W boson in 22q final states at $\sqrt{s} = 13$ TeV”. *Journal of High Energy Physics*, 2018-09. [http://dx.doi.org/10.1007/JHEP09\(2018\)101](http://dx.doi.org/10.1007/JHEP09(2018)101).
- [94] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for a heavy resonance decaying to a pair of vector bosons in the lepton plus merged jet final state at $\sqrt{s} = 13$ TeV”. *Journal of High Energy Physics*, 2018-05. [http://dx.doi.org/10.1007/JHEP05\(2018\)088](http://dx.doi.org/10.1007/JHEP05(2018)088).
- [95] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos and b quarks at $\sqrt{s} = 13$ TeV”. *Journal of High Energy Physics*, 2018-11. [http://dx.doi.org/10.1007/JHEP11\(2018\)172](http://dx.doi.org/10.1007/JHEP11(2018)172).
- [96] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for heavy resonances decaying into two Higgs bosons or into a Higgs boson and a W or Z boson in proton-proton collisions at 13 TeV”. *Journal of High Energy Physics*, 2019-01. [http://dx.doi.org/10.1007/JHEP01\(2019\)051](http://dx.doi.org/10.1007/JHEP01(2019)051).
- [97] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV”. *The European Physical Journal C*, 2017-09. <http://dx.doi.org/10.1140/epjc/s10052-017-5192-z>.
- [98] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at $\sqrt{s}=13$ TeV”. *Physical Review D*, 2018-04. <http://dx.doi.org/10.1103/PhysRevD.97.072006>.
- [99] A. M. Sirunyan, A. Tumasyan, W. Adam *et al.* “Search for ZZ resonances in the 22 final state in proton-proton collisions at 13 TeV”. *Journal of High Energy Physics*, 2018-03. [http://dx.doi.org/10.1007/JHEP03\(2018\)003](http://dx.doi.org/10.1007/JHEP03(2018)003).
- [100] J. Thaler and K. Van Tilburg. “Identifying boosted objects with N-subjettiness”. *Journal of High Energy Physics*, 2011-03. [http://dx.doi.org/10.1007/JHEP03\(2011\)015](http://dx.doi.org/10.1007/JHEP03(2011)015).
- [101] S. Weinberg. “The making of the Standard Model”. *The European Physical Journal C*, 2004-05: 5–13. <http://dx.doi.org/10.1140/epjc/s2004-01761-1>.

附录 A 博士期间的研究成果

A.1 CMS 公开文档

Search for resonances decaying to triple W-boson final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, CMS-B2G-20-001-001, <https://cds.cern.ch/record/2757280>

The target journal is PRL.

A.2 国际会议报告

Pheno2021: Phenomenology 2021 Symposium, "Searches for heavy resonances decaying into Z, W, and Higgs bosons at CMS", 24-26 May 2021, University of Pittsburgh, Video only (Virtual World)

附录 B 附录

B.1 不同轻子道其他物理量数据与模拟对比

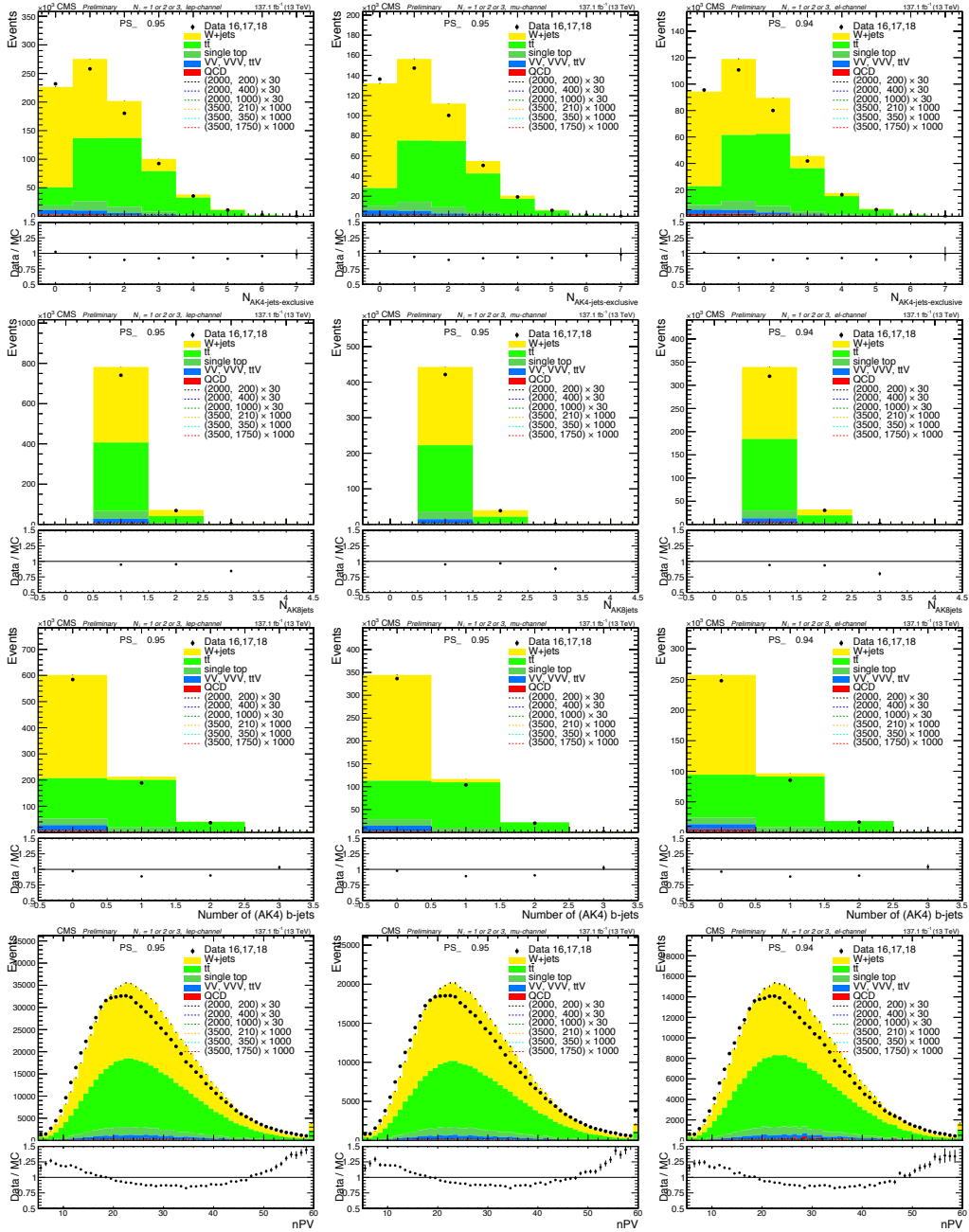


图 B.1 RunII 3 年内的数据得到的部分物理量的分布图，按 μ 子 + 电子道、 μ 子和电子道从左到右的顺序。

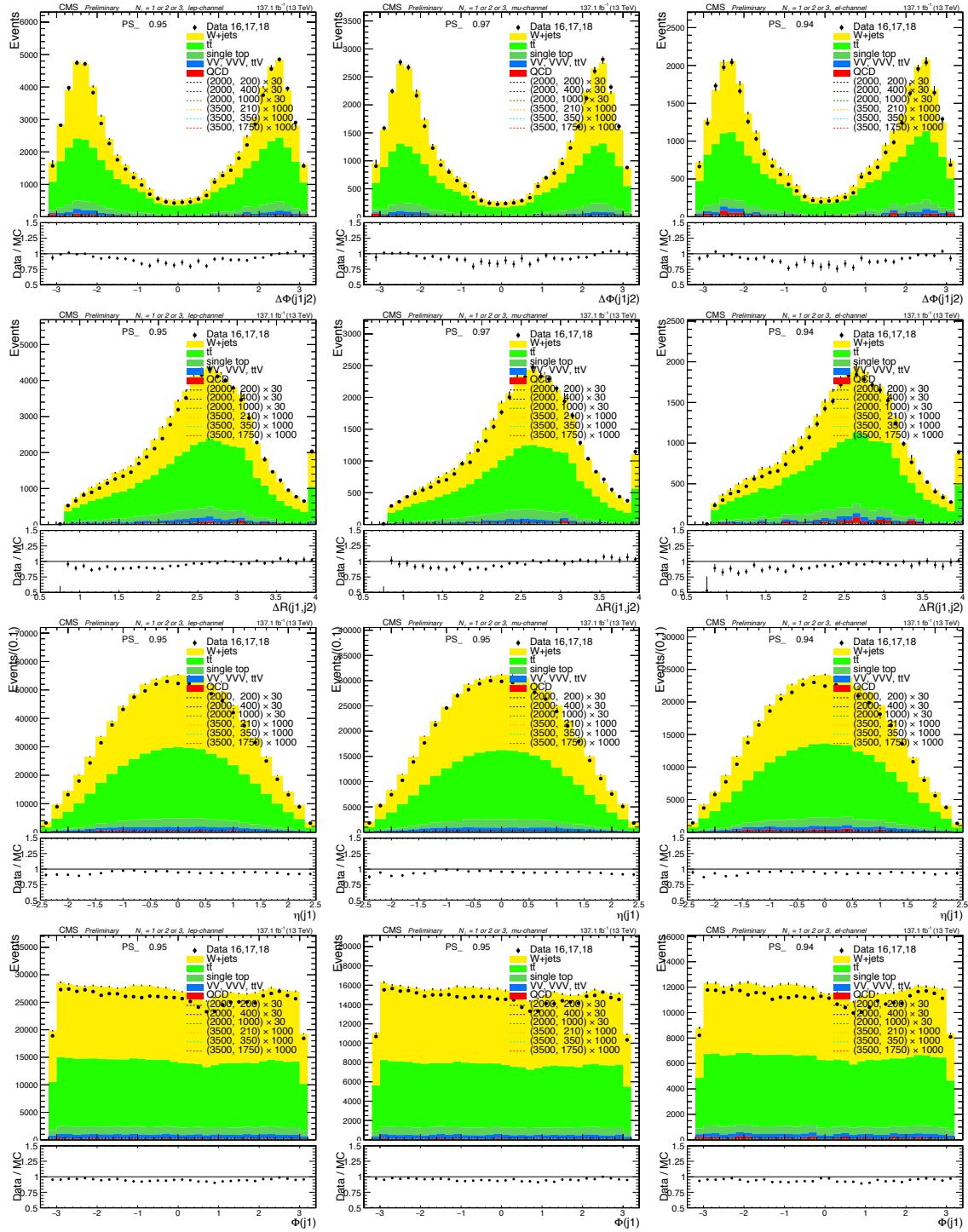
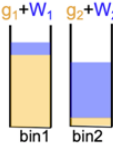


图 B.2 RunII 3 年内的数据得到的部分物理量的分布图，按 μ 子 + 电子道、 μ 子和电子道从左到右的顺序。

B.2 矩阵方法中由于子样本挑选条件带来的偏差

出于简化的考虑，仅使用 2 种类型的喷注：W 喷注 (W) 和胶子 (g)；然后矩阵方法可简化为 2×2 的矩阵方差程。假设对应两种喷注设计的 2 个子样本中的 SFs 分别为 a_1 、 b_1 和 a_2 、 b_2 。对应于这两种喷注的真实 SFs， a' 和 b' ，可表示为 a_1 、 a_2 和 b_1 、 b_2 的加权平均。图 B.3 计算了 SFs 测量值相对真实值的偏离。真正的分析使用的 3×3 矩阵如图 B.4 所示，结论为如果对于 W、g、t 的真实 SFs 在不同子样本中接近或者每个子样本中其成分纯度非常高，测量值与真实值的偏离会较小。在矩阵方法使用时会尽量满足后者的要求从而降低偏离。

- We assume only 2 types of jets: W, g, are present in a single tagger-score bin, split in 2 buckets $\rightarrow$ 2 equations, 2×2 problem.
- We assume purities as in the cartoon:
 

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \begin{bmatrix} g_1 & w_1 \\ g_2 & w_2 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$
- The **Measured SFs** a, b with the matrix method are: $a = \frac{D_1 w_2 - D_2 w_1}{g_1 w_2 - g_2 w_1}$, $b = \frac{D_2 g_1 - D_1 g_2}{g_1 w_2 - g_2 w_1}$
- The **True SFs** a', b' are given by considering different SFs in the 2 bins i.e.: a_1, a_2, b_1, b_2 , and by taking the weighted average: $a' = \frac{a_1 g_1 + a_2 g_2}{g_1 + g_2}$, $b' = \frac{b_1 w_1 + b_2 w_2}{w_1 + w_2}$
- The Matrix problem then looks like: $\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \begin{bmatrix} a_1 g_1 & b_1 w_1 \\ b_2 g_2 & b_2 w_2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and can't be solved as this has 4 unknowns.
 $\rightarrow$ But we can extract useful expressions for the differences $(a' - a)$, $(b' - b)$: biases

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \begin{bmatrix} g_1 & w_1 \\ g_2 & w_2 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{bmatrix} a_1 g_1 & b_1 w_1 \\ b_2 g_2 & b_2 w_2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{cases} g_1 a_1 + w_1 b_1 = a g_1 + b w_1 & (1) \\ g_2 a_2 + w_2 b_2 = a g_2 + b w_2 & (2) \end{cases}$$
 From true SFs definition: $\begin{cases} a'(g_1 + g_2) = g_1 a_1 + g_2 a_2 \rightarrow \frac{g_2}{g_1} = \frac{a' - a_1}{a_2 - a'} & (3) \\ b'(w_1 + w_2) = w_1 b_1 + w_2 b_2 \rightarrow \frac{w_1}{w_2} = \frac{b' - b_2}{b_1 - b'} & (4) \end{cases}$
- We define: $det = g_1 w_2 - g_2 w_1$; and rearrange the problem **solving for bias** as:

$$(1)_{(4)}^{(3)} \rightarrow (a' - a) = (w_1 + w_2) \left[\frac{w_1}{det} (b' - b_1) - \frac{g_2}{det} (a' - a_2) \right] \quad (5)$$

$$(2)_{(4)}^{(3)} \rightarrow (b' - b) = -(g_1 + g_2) \left[\frac{w_1}{det} (b' - b_1) - \frac{g_2}{det} (a' - a_2) \right] \quad (6)$$
- Bias depends to det^{-1} and to true SFs deviation in low-purity subsets $(a' - a_2)$, $(b' - b_1)$**
- The [...] is the same in (5), (6) and so $\rightarrow (a' - a) = -|K|(b' - b)$ i.e. anticorrelations
- The 2 terms are independent and we need both to be ~ 0 to suppress bias
- Assuming high purities: $g_1 w_2 \gg g_2 w_1 \rightarrow det \approx g_1 w_2$ and bias can be written:

$$(1)_{(4)}^{(3)} \rightarrow (a' - a) \approx \frac{w_1 + w_2}{w_2} \left[\frac{w_1}{g_1} (b' - b_1) - \frac{g_2}{g_1} (a' - a_2) \right] \quad (5')$$

$$(2)_{(4)}^{(3)} \rightarrow (b' - b) \approx -\frac{g_1 + g_2}{g_1} \left[\frac{w_1}{w_2} (b' - b_1) - \frac{g_2}{w_2} (a' - a_2) \right] \quad (6')$$
- Bias depends directly to purities and to SF deviations $(a' - a_2)$, $(b' - b_1)$**
- To suppress bias we rely on "small impurities" i.e.: $\frac{w_1}{g_1}, \frac{g_2}{g_1}, \frac{w_1}{w_2}, \frac{w_1}{w_2} \ll 1$
- If this is not true, bias is subjected to how the true SFs differs i.e. depends on cuts

图 B.3 简化 2×2 矩阵方法中测量值相对真实值偏离的计算。

bin1 eq.1 bin2 eq.2 bin3 eq.3

$$\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \begin{bmatrix} g_1 & w_1 & t_1 \\ g_2 & w_2 & t_2 \\ g_3 & w_3 & t_3 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{bmatrix} g_1 a_1 & w_1 b_1 & t_1 c_1 \\ g_2 a_2 & w_2 b_2 & t_2 c_2 \\ g_3 a_3 & w_3 b_3 & t_3 c_3 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$D_1 = g_1 a_1 + w_1 b_1 + t_1 c_1 = a g_1 + b w_1 + c t_1 \quad (1)$$

$$D_2 = g_2 a_2 + w_2 b_2 + t_2 c_2 = a g_2 + b w_2 + c t_2 \quad (2)$$

$$D_3 = g_3 a_3 + w_3 b_3 + t_3 c_3 = a g_3 + b w_3 + c t_3 \quad (3)$$

(1) $\rightarrow g_1(a' - a) + w_1(b' - b) + t_1(c' - c) = -g_1(a' - a_1) - w_1(b' - b_1) - t_1(c' - c_1) \quad (1')$

(2) $\rightarrow g_2(a' - a) + w_2(b' - b) + t_2(c' - c) = -g_2(a' - a_2) - w_2(b' - b_2) - t_2(c' - c_2) \quad (2')$

(3) $\rightarrow g_3(a' - a) + w_3(b' - b) + t_3(c' - c) = -g_3(a' - a_3) - w_3(b' - b_3) - t_3(c' - c_3) \quad (3')$

$$\begin{bmatrix} g_1 & w_1 & t_1 \\ g_2 & w_2 & t_2 \\ g_3 & w_3 & t_3 \end{bmatrix} \begin{bmatrix} a' - a \\ b' - b \\ c' - c \end{bmatrix} = \begin{bmatrix} -g_1(a' - a_1) - w_1(b' - b_1) - t_1(c' - c_1) \\ -g_2(a' - a_2) - w_2(b' - b_2) - t_2(c' - c_2) \\ -g_3(a' - a_3) - w_3(b' - b_3) - t_3(c' - c_3) \end{bmatrix} \equiv \begin{bmatrix} -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} g_1 & w_1 & t_1 \\ g_2 & w_2 & t_2 \\ g_3 & w_3 & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -g_1 x_1 - w_1 y_1 - t_1 z_1 \\ -g_2 x_2 - w_2 y_2 - t_2 z_2 \\ -g_3 x_3 - w_3 y_3 - t_3 z_3 \end{bmatrix} \equiv \begin{bmatrix} -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \end{bmatrix} \quad \text{simplification changing variable to } x, y, z, \dots$$

Solving for bias we get:

$$\text{biases} \begin{cases} x = \det^{-1}[-\lambda_1(w_2 t_3 - w_3 t_2) + \lambda_2(w_1 t_3 - w_3 t_1) - \lambda_3(w_1 t_2 - w_2 t_1)] \\ y = \det^{-1}[-\lambda_1(g_2 t_3 - g_3 t_2) + \lambda_2(g_1 t_3 - g_3 t_1) - \lambda_3(g_1 t_2 - g_2 t_1)] \\ z = \det^{-1}[-\lambda_1(g_2 w_3 - w_2 g_3) + \lambda_2(g_1 w_3 - w_1 g_3) - \lambda_3(g_1 w_2 - w_1 g_2)] \end{cases}$$

While det is: $\det = \begin{vmatrix} g_1 & w_1 & t_1 \\ g_2 & w_2 & t_2 \\ g_3 & w_3 & t_3 \end{vmatrix} = g_1 \begin{vmatrix} w_2 & t_2 \\ w_3 & t_3 \end{vmatrix} - w_2 \begin{vmatrix} g_1 & t_1 \\ g_3 & t_3 \end{vmatrix} + t_3 \begin{vmatrix} g_1 & w_1 \\ g_2 & w_2 \end{vmatrix}$

(Which in the high purities regime: $\det \approx g_1 w_2 t_3$) and λ_i are given by the:

$$\begin{bmatrix} -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \end{bmatrix} \equiv \begin{bmatrix} -g_1(a' - a_1) - w_1(b' - b_1) - t_1(c' - c_1) \\ -g_2(a' - a_2) - w_2(b' - b_2) - t_2(c' - c_2) \\ -g_3(a' - a_3) - w_3(b' - b_3) - t_3(c' - c_3) \end{bmatrix}$$

The exact solution for the biases then are:

$$a' - a = x = -\lambda_1 \frac{w_2 t_3 - w_3 t_2}{\det} + \lambda_2 \frac{w_1 t_3 - w_3 t_1}{\det} - \lambda_3 \frac{w_1 t_2 - w_2 t_1}{\det}$$

$$b' - b = y = -\lambda_1 \frac{g_2 t_3 - g_3 t_2}{\det} + \lambda_2 \frac{g_1 t_3 - g_3 t_1}{\det} - \lambda_3 \frac{g_1 t_2 - g_2 t_1}{\det}$$

$$c' - c = z = -\lambda_1 \frac{g_2 w_3 - w_2 g_3}{\det} + \lambda_2 \frac{g_1 w_3 - w_1 g_3}{\det} - \lambda_3 \frac{g_1 w_2 - w_1 g_2}{\det}$$

$$x = (x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1) \frac{\begin{vmatrix} w_2 & t_2 \\ w_3 & t_3 \end{vmatrix}}{\det} + (x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3) \frac{\begin{vmatrix} w_1 & t_1 \\ w_3 & t_3 \end{vmatrix}}{\det} + (z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3) \frac{\begin{vmatrix} w_1 & t_1 \\ w_2 & t_2 \end{vmatrix}}{\det}$$

$$y = (x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1) \frac{\begin{vmatrix} g_2 & t_2 \\ g_3 & t_3 \end{vmatrix}}{\det} + (x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3) \frac{\begin{vmatrix} g_1 & t_1 \\ g_3 & t_3 \end{vmatrix}}{\det} + (z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3) \frac{\begin{vmatrix} g_1 & t_1 \\ g_2 & t_2 \end{vmatrix}}{\det}$$

$$z = (x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1) \frac{\begin{vmatrix} g_2 & w_2 \\ g_3 & w_3 \end{vmatrix}}{\det} + (x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3) \frac{\begin{vmatrix} g_1 & w_1 \\ g_3 & w_3 \end{vmatrix}}{\det} + (z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3) \frac{\begin{vmatrix} g_1 & w_1 \\ g_2 & w_2 \end{vmatrix}}{\det}$$

The same exact solution in a more explicit form (substituting λ_i) are here:

In the high purities approximation (where $\det \approx g_1 w_2 t_3$) we have the bias:

$$x \approx -\frac{\lambda_1}{g_1} + \frac{\lambda_2 w_1}{g_1 w_2} + \frac{\lambda_3 t_1}{g_1 t_3}$$

$$y \approx -\frac{\lambda_1}{g_1} + \frac{\lambda_2}{g_1} + \frac{\lambda_3 t_2}{g_1 t_3}$$

$$z \approx \frac{\lambda_1 w_3}{g_1 t_3} + \frac{\lambda_2 w_3}{w_2 t_3} - \frac{\lambda_3}{t_3}$$

While substituting λ_i we get:

$$x \approx -\frac{x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1}{g_1} + \frac{(x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3) w_1}{g_1 w_2} + \frac{(z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3) t_1}{g_1 t_3}$$

$$y \approx -\frac{(x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1) g_2}{g_1} + \frac{x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3}{g_1} - \frac{(z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3) t_2}{g_1 t_3}$$

$$z \approx \frac{(x_2 g_2 + x_3 g_3 - t_1 x_1 - w_1 y_1) g_3}{g_1 t_3} + \frac{(x_2 g_2 - w_1 y_1 + t_2 z_2 - w_3 y_3) w_3}{w_2 t_3} - \frac{z_2 t_2 + t_1 z_1 - w_3 y_3 - x_3 g_3}{t_3}$$

These equations degenerates in the following 2 if we constrain to 2x2 problem.

Result consistent with the 2x2 calculations

$$x \approx -x_2 \frac{g_2}{g_1} - x_3 \frac{g_3}{g_1} + x_1 \frac{t_1}{g_1} + y_1 \frac{w_1}{g_1}$$

$$y \approx +x_2 \frac{g_2}{w_2} - y_1 \frac{w_1}{w_2} + z_2 \frac{t_2}{w_2} - y_3 \frac{w_3}{w_2}$$

$$z \approx -z_2 \frac{t_2}{t_3} + z_1 \frac{t_1}{t_3} - y_3 \frac{w_3}{t_3} - x_3 \frac{g_3}{t_3}$$

图 B.4 3×3 矩阵方法中测量值相对真实值偏离的计算。

B.3 t^3 和 t^4 关于 deep-W (WH) 的差异

图B.5比较了两种 top: t^3 和 t^4 在不同横动量区间的 deep-WH 和 deep-W 分布谱的对比. 可以发现 t^3 和 t^4 在所有的横动量区间都是相似的, 为了描述两者的微小差异, 计算两个直方图在 deep-W (WH) 在值大于 0.7 (为信号区 deepAK8 标记的选择条件截断值) 每个分区的事例数的绝对差值之和与 t^3 的事例数之和的比值, 即 $\Sigma(t_i^4 - t_i^3)/\Sigma(t_i^3)$, 标记子右上角图例的中括号中. 使用最大值作为 t^3 和 t^4 的差异, 对于 deep-W 取 10%, 对于 deep-WH 取 5%.

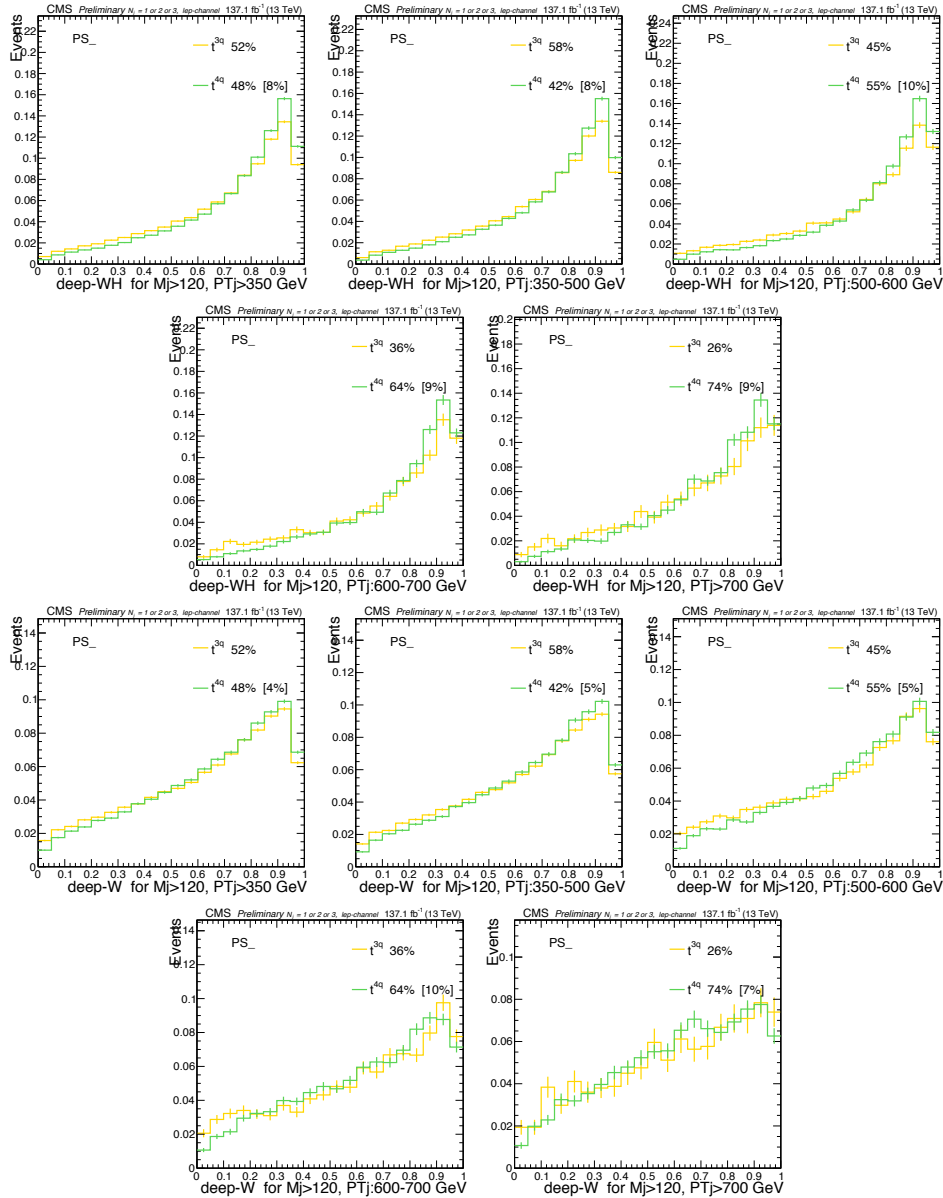


图 B.5 t^3 和 t^4 在不同横动量区间的 deep-WH (上两行) deep-W (下两行) 分布对比, 所有直方图都归一到 1, t^3 和 t^4 后第一个数为相对占比, 它们在取值大于 0.7 时的差异被标记在右上角图例的方括号中。

B.4 信号 R^{3q} 、 R^{4q} 和 $t^{3,4}$ 在添加代理误差后的对比

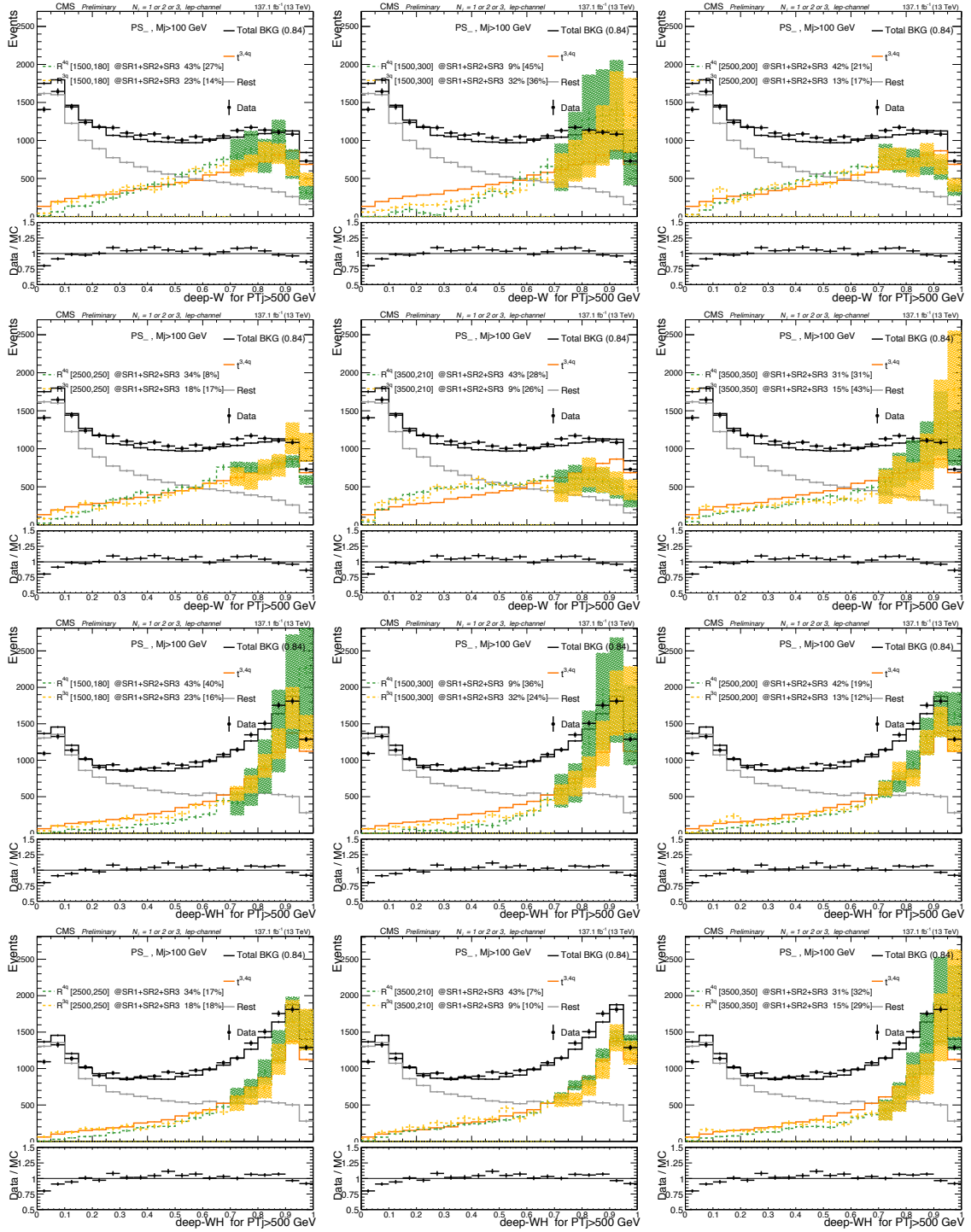


图 B.6 6 个有代表性的合并 radion 信号点 R^{3q} 、 R^{4q} 和 $t^{3,4}$ 在 deep-W (WH) 大于 0.7 加代替误差后的对比，可以发现误差带可以很好地覆盖其形状的差异，前两行为 deep-W，后两行为 deep-WH。

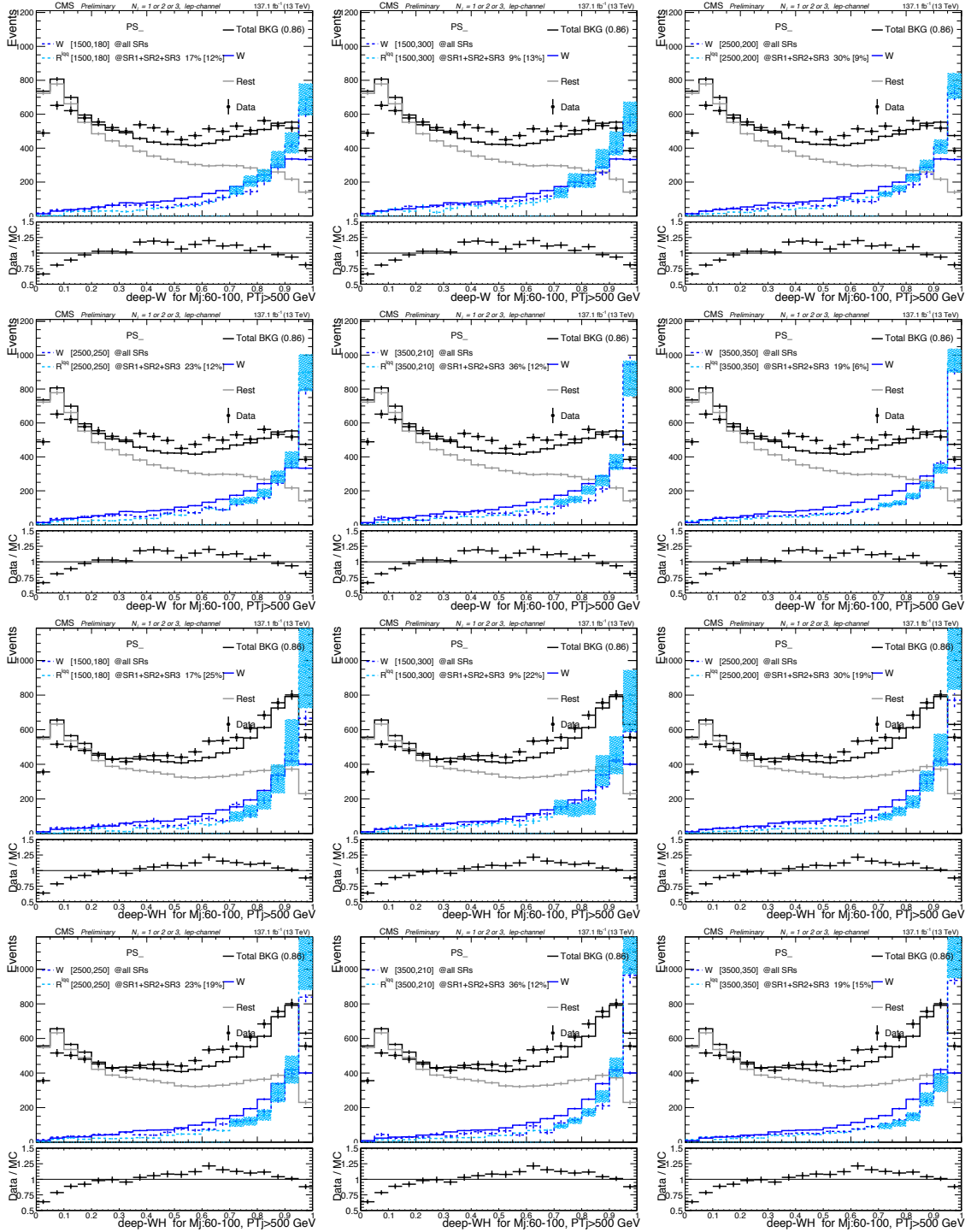


图 B.7 6 个有代表性的合并 radion 信号点 R^{lqq} 、W 和本底 W 喷注在 deep-W (WH) 大于 0.7 加代替误差后的对比，可以发现误差带可以很好地覆盖其形状的差异，前两行为 deep-W，后两行为 deep-WH。

B.5 top 与 W+jets 控制区传递系数线性拟合

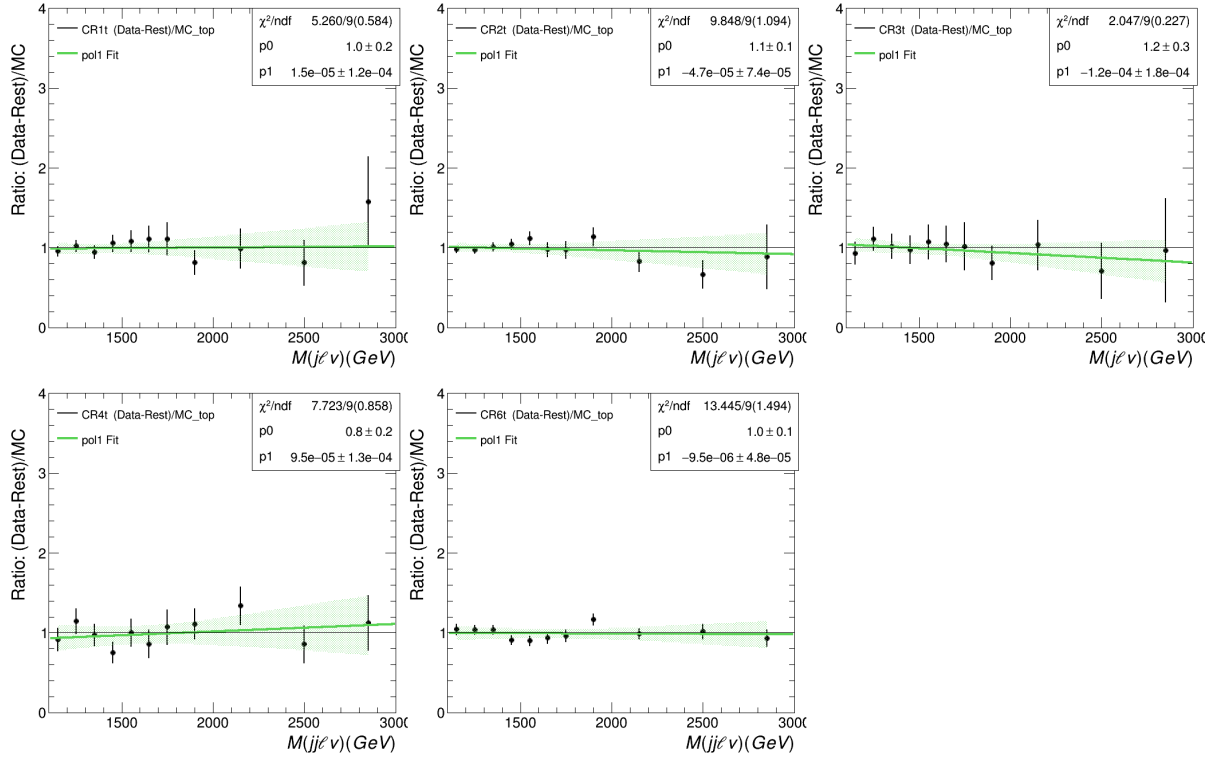


图 B.8 top 本底控制区 1、2、3、45、6 的形状传递系数线性拟合，为了保证足够的统计量，观测量是按照 [1100、1200、1300、1400、1500、1600、1700、1800、2000、2300、2700、……]GeV 分区。

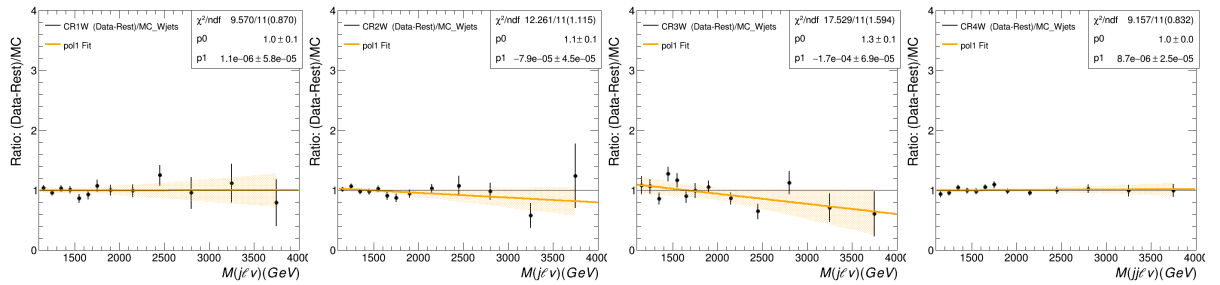


图 B.9 W+Jets 控制区 1、2、3、456 传递系数线性拟合，为了保证足够的统计量，观测量是按照 [1100、1200、1300、1400、1500、1600、1700、1800、2000、2300、2600、3000、3500、……]GeV 分区。

B.6 二维质量平面的插值方法

二维共振态质量平面上在 75 个信号点间插值方法如图B.10，使用信号强度的 $\log$ 值间线性插值，具体中间步骤如图B.11。

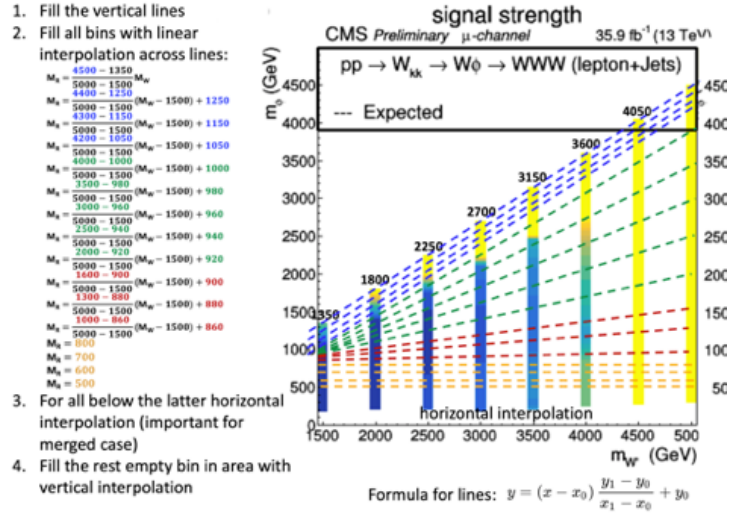


图 B.10 2D 平面上线性插值流程。

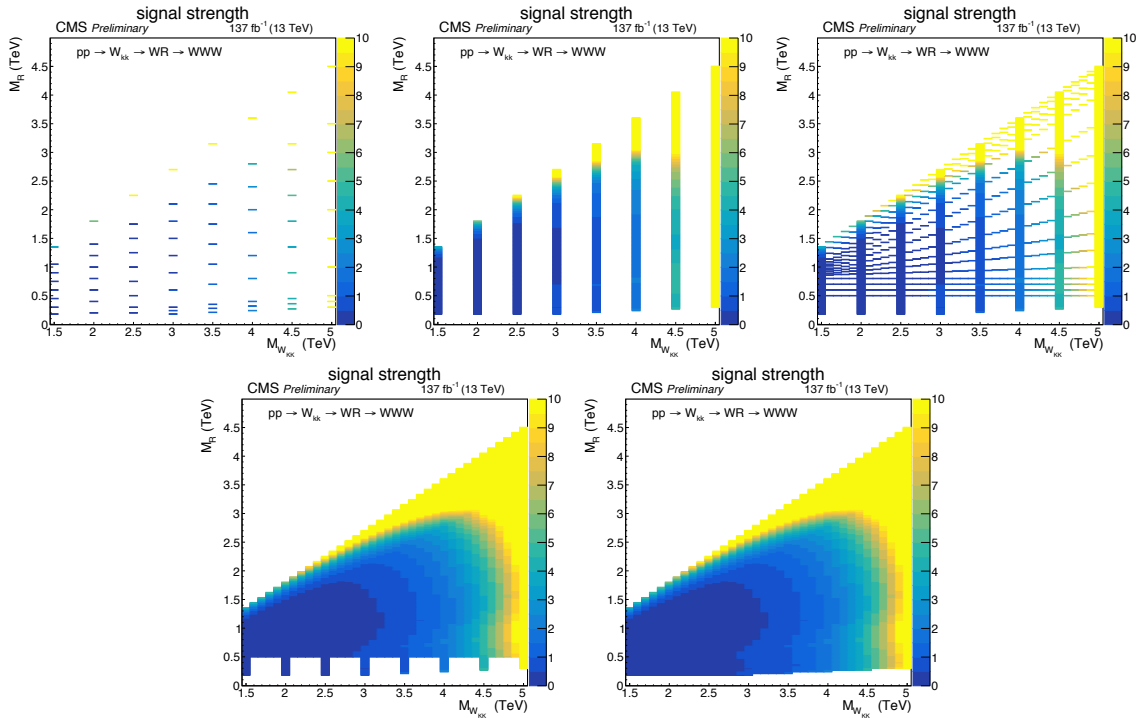


图 B.11 对信号强度的线性插值方法的具体步骤。

对于信号截面，使用 Madgraph 计算上图第一行中间图中的每一列对应的截面然后在水平方向截面的 $\log$ 值间线性插值。

B.7 信号系统误差的研究

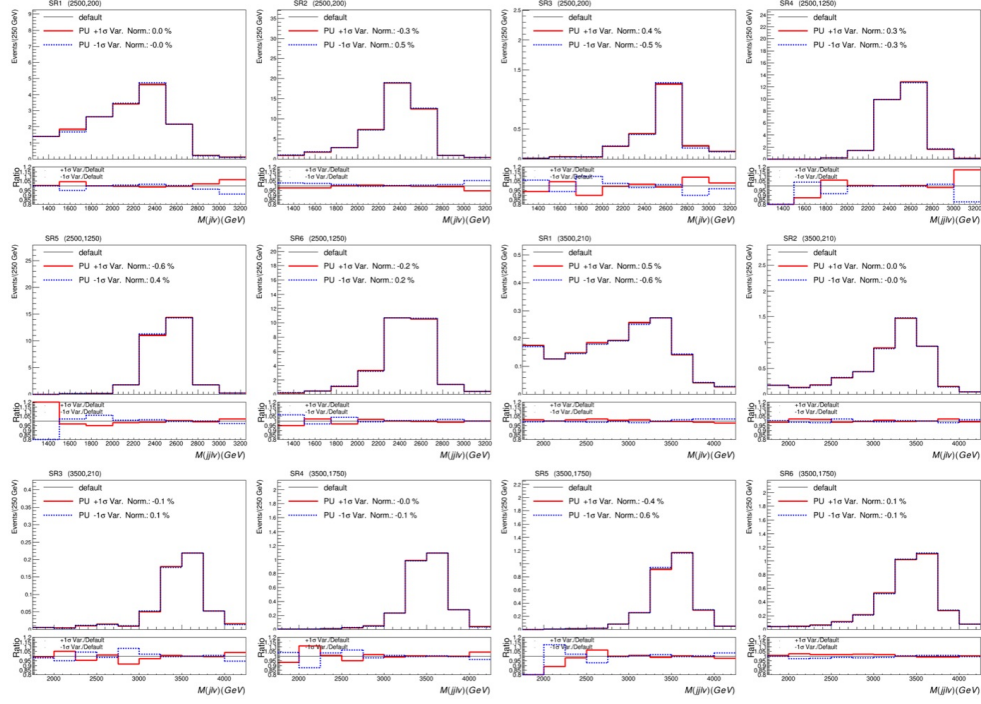


图 B.12 μ 子道 pileup 误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

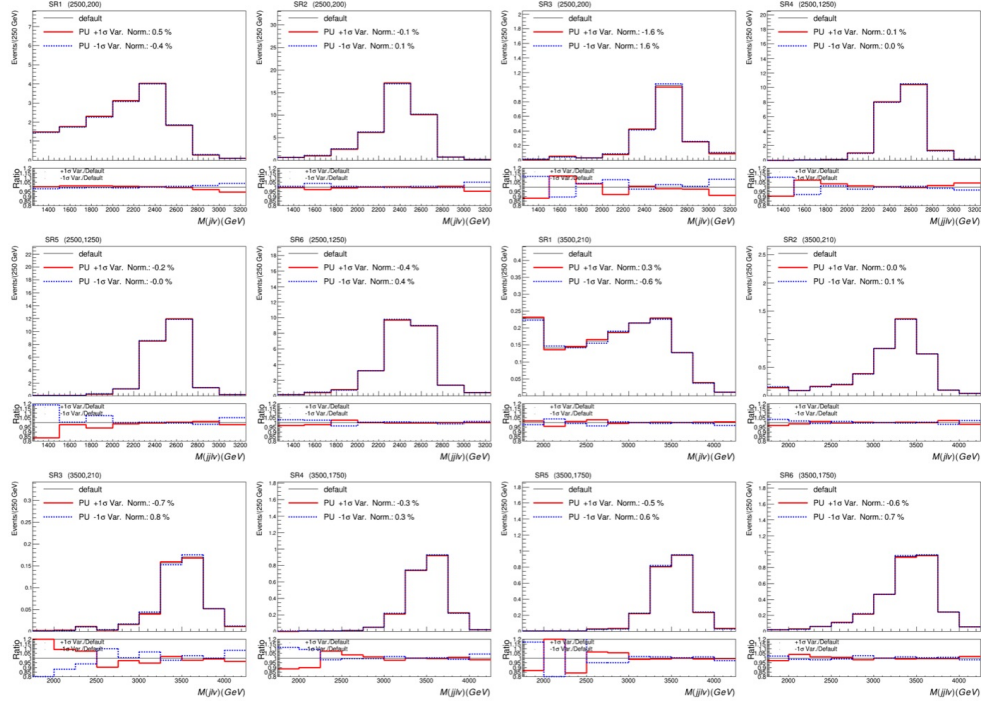


图 B.13 电子道 pileup 误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

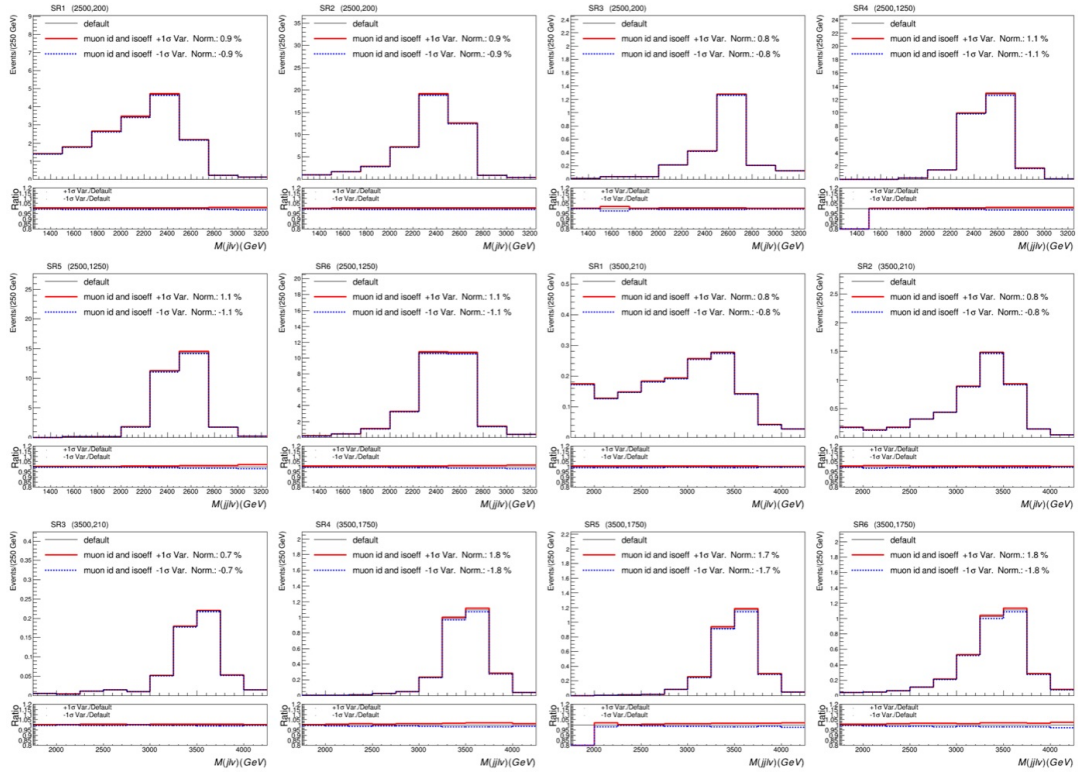


图 B.14 μ 子道轻子鉴别与分辨误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

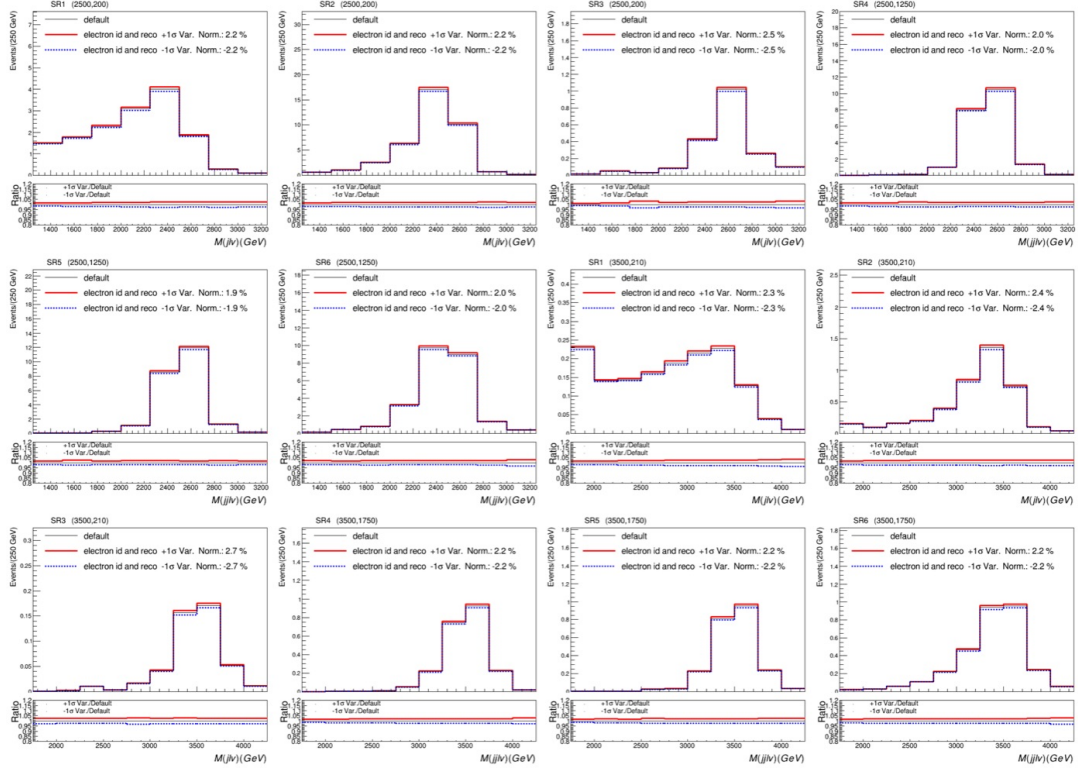


图 B.15 电子道轻子鉴别与分辨误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

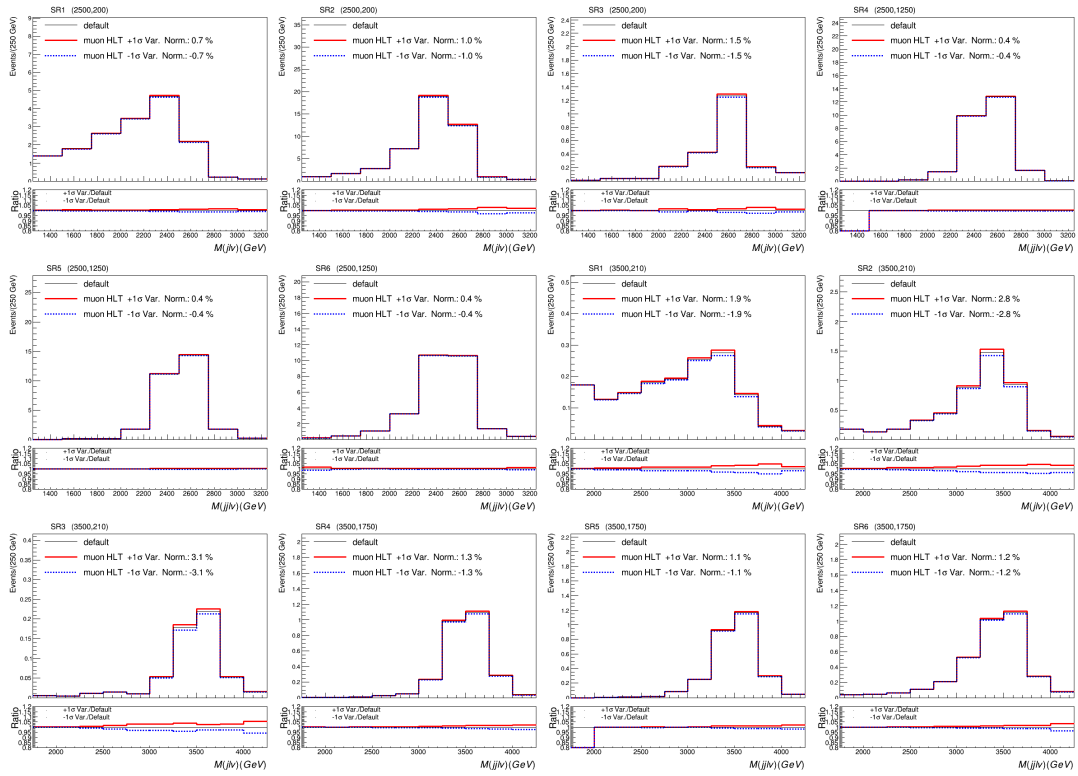


图 B.16 μ 子道触发效率误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

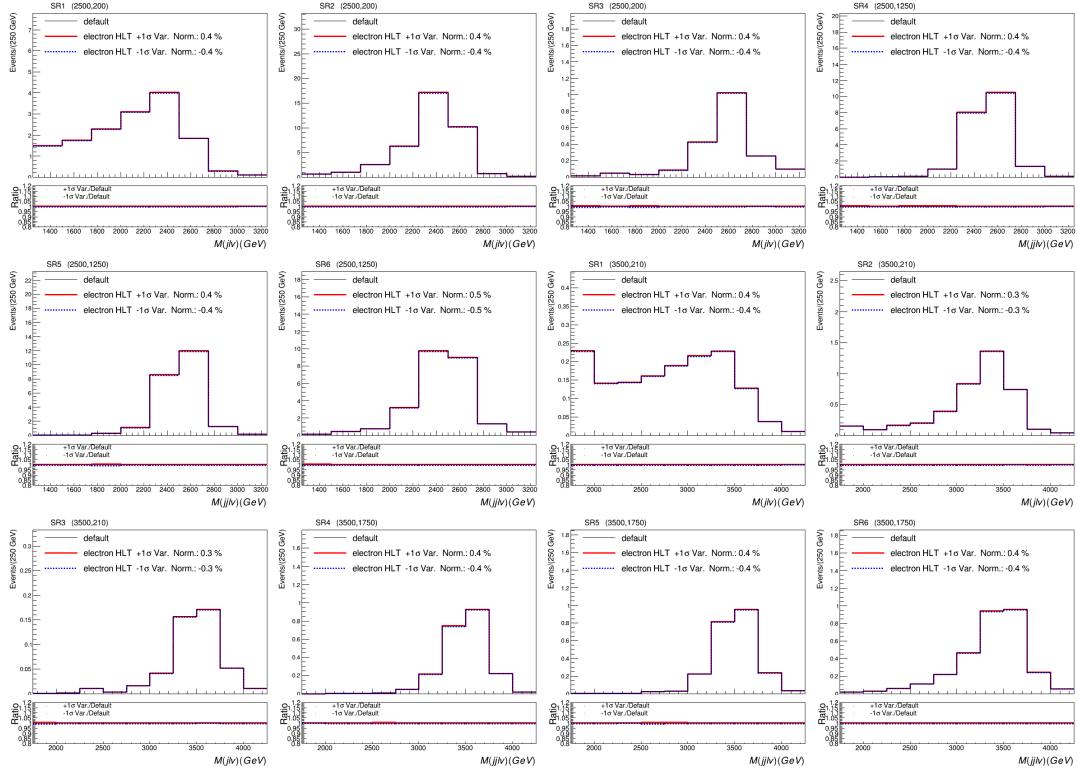


图 B.17 电子道触发效率误差对信号观测量分布的影响，分别为信号 (2500,200) 在信号区 1-3, (2500,1250) 在信号区 4-6, (3500,210) 在信号区 1-3, (3500,1750) 在信号区 4-6 的分布。

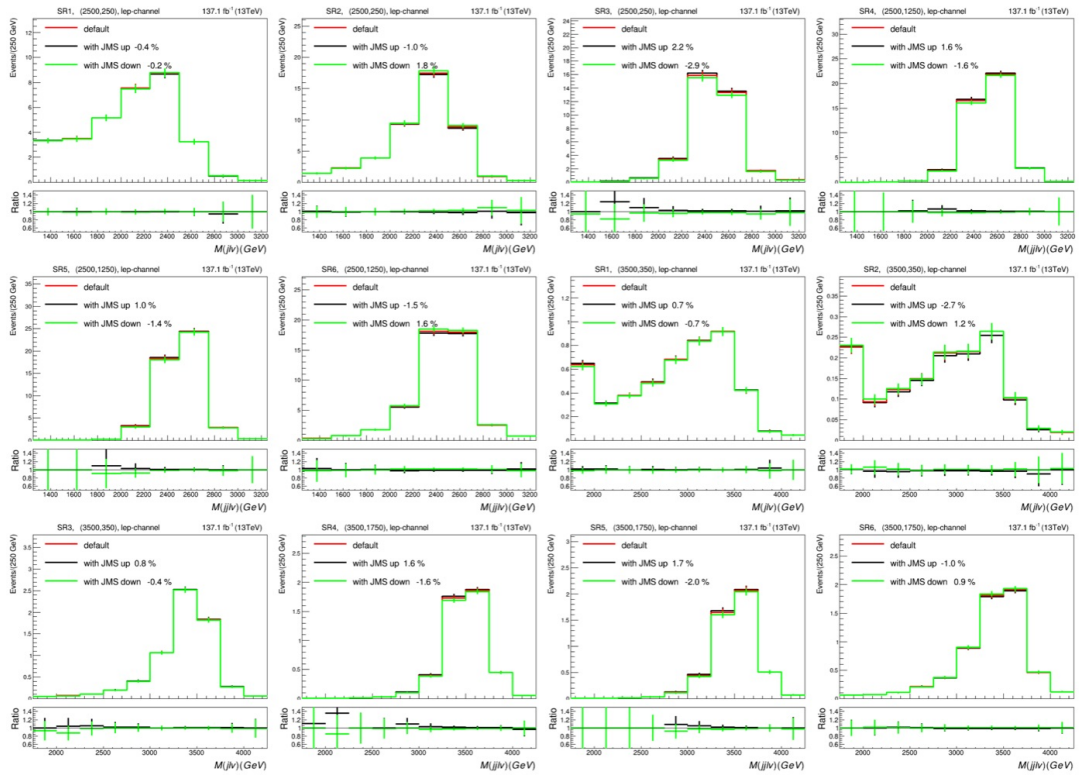


图 B.18 JMS $\pm 1\sigma$ 误差与初始形状的对比, 分别为合并信号 (2500, 250) 和分立信号 (2500, 1250) 在信号区 1-6 的对比。

致谢

不知不觉五年的博士生涯即将结束，毕业论文写作也接近尾声。在五年的博士生活和研究中得到了很多人的帮助，谨以这一章表达我的感激之情。

感谢我的导师冒亚军教授和李强长聘副教授的悉心指导，在五年前是由冒老师引入高能物理实验方向，后来受李老师指导完成了 CMS 上三波色子共振态的寻找分析。在与老师的讨论中增进了我对高能物理的了解，掌握了高能物理实验的研究方法。老师们都十分和蔼，在遇到问题时总会耐心点拨、循循善诱，感谢在毕业论文选题、分析工作中的引领及写作过程中提供的帮助。他们的博闻强识、一丝不苟的工作态度以及亲和的人格魅力给我留下了深刻的印象，在此向我的指导老师表示诚挚的谢意！

感谢国家留学基金委和北大物院在我于 CERN 两年学习生活中的资金支持。在瑞士留学期间李老师在让我认识了很多专家，感谢外导 Maurizio Pierini 以及 B2G 会议召集人 Clemens Lange 对我分析工作的帮助。在分析工作外，我参与了 CMS 上触发速率的测量和蒙特卡洛模拟样本的产生等工作，感谢触发负责人 Nadir Daci、JME 会议召集人 Andreas Hinzmann 对我的指导和帮助。

在瑞士生活的最后一年经历了逐渐爆发的疫情，鉴于国外很少戴口罩的情形，公共场所不太安全，因此大部分时间都在家中，生活相对沉闷单调。房东范姐经常陪聊天，她是一个非常热心的人，会主动帮忙解决遇到的问题，比如帮忙买东西、接送出行等，在此表示衷心的感谢。另外，也感谢驻瑞士大使馆提供的口罩和在瑞士期间小伙伴们的陪伴。

感谢合作组来自德州理工大学的 Jordan Damgov 和 Sungwon Lee 对于分析程序和方法建议方面的帮助，感谢博士后 Antonios Agapitos 和同组的陈城同学的合作，Antonios 在工作时精益求精的精神值得学习。感谢北大组所有老师和同学对生活中的照顾和学习中的答疑解惑。特别感谢黄璜和卢梦师兄在刚接触 CMS 分析时对代码和分析框架的指引，感谢郭启隆师弟对部分工作的检验和论文的检查。

最后感谢父母和所有朋友的无私关怀和支持，正是有你们在学习和生活上的鼓励和陪伴，我才能够走的更远，在此表示衷心的感谢！

北京大学学位论文原创性声明和使用授权说明

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。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：吕旭东 日期：2021年4月18日

学位论文使用授权说明

(必须装订在提交学校图书馆的印刷本)

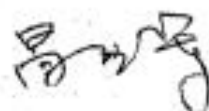
本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
- 因某种特殊原因需要延迟发布学位论文电子版，授权学校 ☐ 一年 / ☐ 两年 / ☐ 三年以后，在校园网上全文发布。

(保密论文在解密后遵守此规定)

论文作者签名：吕旭东

导师签名：



日期：2021年4月19日

