Aix*Marseille université
Socialement engagée

NNT/NL : 2022AIXM0567/061ED352

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 14 décembre 2022 par

Léonard POLAT

Search for $\tau \to \ell\phi$ lepton flavour violating decays and fake tracks rate measurement with $e^+e^- \to \tau^+\tau^-$ events at the Belle II experiment

Discipline
Physique et Sciences de la Matière

Spécialité
Physique des Particules et Astroparticules

École doctorale
Physique et Sciences de la Matière (ED 352)

Centre de Physique des Particules de Marseille (CPPM) UMR 7346
Conseil européen de la recherche (ERC)

Composition du jury

Eli BEN-HAIM LPNHE (CNRS/IN2P3)	Rapporteur
Emi KOU IJCLab (CNRS/IN2P3)	Rapporteuse
Guillaume PIGNOL LPSC (CNRS/IN2P3)	Examinateur
Marlon BARBERO CPPM (CNRS/IN2P3)	Président du jury
Justine SERRANO CPPM (CNRS/IN2P3)	Directrice de thèse

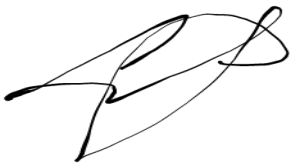
Belle II
cnrs
CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM
erc European Research Council Established by the European Commission

Affidavit

Je soussigné, Léonard Polat, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Justine Serrano, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 13 janvier 2023.



Remerciements

À Justine, pour m'avoir honoré de sa confiance ; pour sa patience surhumaine, sa disponibilité constante et salvatrice, la justesse de son jugement et le flegme de sa présence d'esprit ;

Aux membres du groupe Belle II du CPPM[1] (dans l'ordre d'apparition) : Laura, Robin, Klemens, Valerio, Arthur, Giampiero, Clotilde, pour m'avoir offert tout au long de ces trois années éprouvantes le baume de leur compagnonnage ; pour leur enthousiasme, leur soutien et leur bienveillance couronnée d'une indéfectible bonne humeur ;

Au CPPM et à tous ceux qui le rendent vivant, pour leur accueil chaleureux, même à temps partiel (circonstances sanitaires obligent) ;

Aux membres du jury de thèse, pour m'avoir accordé la denrée la plus précieuse : leur temps (et un diplôme, bien entendu) ; pour une séance de questions et réponses infiniment instructive et tout aussi agréable ;

À Jérôme, Isabelle et Giulio, pour m'avoir pris sous leur aile et introduit à la physique expérimentale il y a six ans ; pour m'avoir tant enseigné depuis, scientifiquement et humainement ; pour m'avoir, aussi, insufflé l'élan nécessaire à mon envol ;

À mes amis, qui se reconnaîtront, étoiles serties sur le canevas embrumé de l'existence ; comme les faces d'un dé, nous regardons ailleurs mais sommes faits de la même matière, et où que le sort puisse nous ballotter, nous allons ensemble, inséparables ;

À ma famille, mes parents, mes sœurs, pour avoir été là malgré la distance et cru en moi plus encore que moi-même ; pour m'avoir inculqué le noble sens du respect d'autrui ;

À tous mes enseignants, de la plus petite enfance à aujourd'hui, qui endossent la tâche la plus sacrée, si sacré il y a ;

Merci !

Marseille, 13 janvier 2023

[1] Aussi connu sous le nom de *Justine League*.

Résumé

La physique au-delà du modèle standard

Le modèle standard de la physique des particules (MS) est, à l'heure actuelle, la théorie décrivant les particules élémentaires et subatomiques avec la plus grande exactitude. C'est une théorie quantique relativiste des interactions électromagnétique, forte et faible.

Ce modèle a fait ses preuves depuis plusieurs décennies, de la découverte des quarks à celle plus récente du boson de Higgs, ce qui acheva de prouver l'existence de toutes les particules prédites par les théoriciens pour que le MS ait un socle solide et une forme cohérente. On peut toutefois suspecter qu'il ne s'agit que d'une "théorie effective", limitée a un certain domaine d'énergie, tandis qu'une théorie plus vaste encore l'engloberait. Il existe en effet quelques zones d'ombre en physique des particules que le MS ne parvient pas à éclairer, avec entre autres

- l'existence de la matière noire et de l'énergie sombre, prouvée par les observations de phénomènes astrophysiques et cosmologiques,
- les obstacles à une description de l'interaction gravitationnelle dans le cadre de la théorie quantique des champs,
- l'asymétrie d'abondance entre la matière et l'antimatière dans l'univers, que la source de violation CP présente dans l'interaction faible ne permet pas de rendre compte,
- le problème de la hiérarchie dans l'intensité des interactions forte et faible, qui semble se traduire par un ajustement fin de la masse du boson de Higgs,
- ou encore la faible masse des neutrinos, démontrée par le phénomène d'oscillation et qui ne trouve pas d'explication satisfaisante au sein du MS.

De fait, les physiciens développent de nombreuses théories allant au-delà du MS, regroupées sous l'expression générale de "nouvelle physique" (NP), afin de pallier ces lacunes.

Les modèles de nouvelle physique, qu'aucune découverte expérimentale n'a encore permis d'appuyer, consistent en l'introduction de champs quantiques supplémentaires, avec éventuellement de nouvelles symétries de jauge, dans le lagrangien du MS décrivant les interactions entre particules − ce à des échelles d'énergie

généralement au-delà des normes du MS et donc difficilement accessibles. Ces champs additionnels seront par exemple associés à des bosons vecteurs rattachés aux nouvelles symétries, à des particules supersymétriques telles que les bosons de Higgs chargés, ou encore à des particules à faibles couplages d'interaction, candidates à la matière noire.

Une conséquence potentielle des théories de nouvelle physique est l'apparition de processus favorisant des canaux de désintégration rares, voire inexistants dans le MS. C'est le cas des désintégrations violant la saveur leptonique (*lepton flavour violation*, LFV), a priori possibles via les oscillations de neutrinos mais à des taux proches de 10^{-50}. Observer un "événement" de cette nature dans un collisionneur de particules serait par conséquent la preuve indubitable qu'un phénomène précis échappe au MS, et un moyen de mettre à contribution la nouvelle physique pour en trouver l'origine.

Nous nous proposons ici d'étudier une désintégration LFV du lepton tau (τ), en l'occurrence $\tau \to \ell\phi$, où ℓ est un électron (e) ou un muon (μ)[2], et ϕ est un méson neutre formé d'une paire de quark-antiquark étranges ($s\bar{s}$), dont nous considérons seulement la désintégration en deux kaons chargés ($\phi \to K^+K^-$). Le canal de désintégration $\tau \to \ell\phi$, et plus particulièrement son mode muonique, a une importance théorique relative à l'hypothèse d'un leptoquark vecteur U_1 [1, 2, 3], capable de rendre compte d'une potentielle violation de l'universalité leptonique mise en lumière ces dernières années par LHCb [4, 5, 6], entre autres.

Nous utilisons pour ce faire le cadre expérimental de *Belle II* qui, sur le long terme, devrait être en mesure d'améliorer d'un à deux ordres de grandeur les limites actuelles sur les rapports d'embranchement de désintégrations LFV du tau (voir Figure 1). Cela inclut $\tau \to \ell\phi$, dont nous citons les limites attendues et observées par les expériences BaBar [7] et Belle [8] dans le Tableau 1, les meilleures étant celles qui furent déterminées par la seconde en 2011.

TABLEAU 1 : Limites supérieures attendues (att.) et observées (obs.) sur les rapports d'embranchements de $\tau \to e\phi$ et $\tau \to \mu\phi$ à un niveau de confiance de 90 %, par les expériences BaBar (451 fb^{-1}) [7] et Belle (854 fb^{-1}) [8].

Expérience	$\mathcal{B}^{90}_{\mathrm{LS}}(e\phi)$ ($\times 10^{-8}$)	$\mathcal{B}^{90}_{\mathrm{LS}}(\mu\phi)$ ($\times 10^{-8}$)
	att. / obs.	att. / obs.
BaBar	5.0 / 3.1	8.2 / 19
Belle	4.3 / 3.1	4.9 / 8.4

[2] Sauf mention contraire, nous utilisons l'appellation des particules pour désigner aussi bien les particules proprement dites que leurs antiparticules respectives.

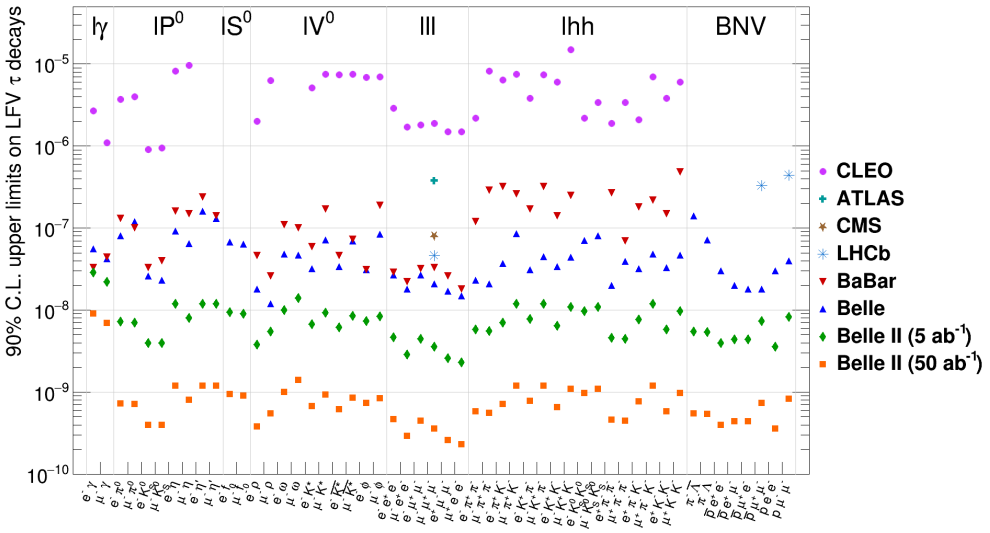


FIGURE 1 : Limites supérieures sur les rapports d'embranchements de diverses désintégrations LFV du tau à un niveau de confiance de 90 %, établies par les expériences CLEO, ATLAS, CMS, LHCb, BaBar et Belle, et estimées dans Belle II pour des luminosités intégrées de 5 (losanges verts) et 50 ab^{-1} (carrés oranges). Extrait de [9].

L'expérience Belle II

L'expérience Belle II, située au laboratoire KEK à Tsukuba (Japon), est dédiée à l'étude avancée du modèle standard comme à la recherche de nouvelle physique. Les installations se divisent en deux éléments principaux : le collisionneur SuperKEKB et le détecteur Belle II.

SuperKEKB est une amélioration du précédent collisionneur KEKB utilisé de 1998 à 2010 dans le cadre de l'expérience Belle, qui a précédé Belle II. Sa fonction est de faire entrer en collision des faisceaux d'électrons et de positrons ; il est ainsi composé de deux anneaux circulaires, *Low Energy Ring* (LER) et *High Energy Ring* (HER), le premier accélérant les positrons à 4 GeV et le second les électrons à 7 GeV. Cette différence en énergie permet de donner un "boost" aux particules produites durant une collision, c'est-à-dire de les propulser vers une direction donnée (l'axe du faisceau dans le sens des électrons) dans le référentiel du laboratoire, afin d'obtenir une meilleure séparation spatiale de leurs désintégrations.

L'avantage majeur de SuperKEKB est la multiplication du maximum de luminosité instantanée atteint dans KEKB par un facteur 30, jusqu'à $\mathcal{L} = 6{\times}10^{35}\ \text{cm}^{-2}\,\text{s}^{-1}$, grâce une avancée technologique récente réduisant fortement le volume du faisceau dans lequel les particules sont concentrées, tout en imposant une intensité électrique environ 1.5 fois plus grande. L'objectif est une luminosité intégrée de

50 ab^{-1} d'ici la fin de la prise de données, offrant aux chercheurs une quantité suffisamment importante d'événements à analyser pour accroître la précision des résultats expérimentaux obtenus auparavant.

Le collisionneur opère à une énergie dans le référentiel du centre de masse $\sqrt{s} = 10.58$ GeV, qui correspond à l'énergie de résonance du méson $\Upsilon(4S)$. Formé d'une paire de quark-antiquark beaux, ce méson se désintègre à 96 % [10] en paires de mésons $B\bar{B}$, un des intérêts principaux de Belle II étant l'étude de la violation de CP dans les B. Les collisions e^+e^- produisent également, entre autres, des paires de quarks plus légers ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$) ainsi que des paires de leptons : $e^+e^- \to e^+e^-$ (diffusion Bhabha), $\mu^+\mu^-$ ou $\tau^+\tau^-$. La section efficace de production de paires $\tau^+\tau^-$ est relativement proche de celle du méson $\Upsilon(4S)$ (0.919 nb contre 1.110 nb), ce qui fait de Belle II une expérience parfaitement adaptée à l'étude des taus.

Le détecteur, à l'instar de SuperKEKB, succède au précédent dispositif de l'expérience Belle en améliorant ses composants et sa technologie pour faire face à l'augmentation de la luminosité instantanée, donc des bruits de fond[3] et du taux d'occupation attendus. Il s'agit d'une structure cylindrique située au point de collision des faisceaux, faite d'un empilement de multiples couches de sous-détecteurs couvrant un large angle solide (17-150° pour les couches internes). Cette configuration, en plus de la connaissance précise de la masse invariante du système, identique entre les états initial et final des collisions e^+e^-, permet d'inclure tout processus impliquant des énergies manquantes (étude de la matière noire ou désintégrations avec neutrinos) dans le champ des analyses possibles à Belle II.

Les différents constituants du détecteur permettent de reconstruire la trajectoire des particules par combinaison des points d'impacts (dépôts d'énergie) en "traces", ce qu'on appelle *tracking* ; de restituer leur quadri-moment[4] et enfin d'identifier leur type. Des couches les plus internes aux plus externes, nous avons :

- le détecteur de vertex en silicium (*Vertex Detector*, VXD), qui se divise lui-même en deux couches de capteurs pixelisés (*Pixel Detector*, PXD) et quatre couches de capteurs silicium double-face à pistes (*Silicon Vertex Detector*, SVD), les premières étant les plus proches des faisceaux. La fonction du VXD est de reconstruire les vertex, c'est-à-dire les points de production des particules secondaires, et il participe d'autre part à la reconstruction des trajectoires (avec la chambre à dérive centrale), voire éventuellement à l'identification des particules (avec la mesure du dépôt d'énergie) ;

- la chambre à dérive centrale (*Central Drift Chamber*, CDC), contenant du gaz et qui se divise en petites "cellules" seulement délimitées par des fils disposés en réseau. Lorsqu'une particule chargée traverse une cellule, elle ionise le gaz et les électrons arrachés créent un signal dans les fils. La CDC permet ainsi

[3] Induits par les faisceaux.

[4] L'impulsion se mesure via le moment cinétique, ici dû au champ magnétique d'un solénoïde (1.5 T).

de reconstruire les traces et impulsions des particules, ou encore de donner des informations sur leur nature en mesurant la perte d'énergie dans le gaz ;

- le système d'identification des particules, qui consiste en un compteur de temps de propagation (*Time-Of-Propagation counter*, TOP) dans le tonneau, ainsi qu'un détecteur Tcherenkov (*Aerogel Ring-Imaging Cherenkov counter*, ARICH) dans le bouchon[5] avant du détecteur (dans la direction du boost). Les photons émis par effet Tcherenkov forment un cône, ou un anneau en deux dimensions, et traversent le TOP avec un temps de propagation qui donne indirectement la masse invariante de la particules émettrice. Le détecteur ARICH, fait d'aérogel, va quant à lui mesurer l'angle d'ouverture de l'anneau dont il reconstruit l'image et qui est unique pour chaque type de particule ;

- le calorimètre électromagnétique (*Electromagnetic Calorimeter*, ECL), dont l'utilité est de détecter les photons qui peuvent représenter une part importante de l'énergie dégagée dans les collisions, ou encore les électrons qu'il parviendra donc à distinguer des autres hadrons chargés, tout en assurant une mesure de leur énergie ;

- le détecteur de K-long (K_L^0) et de muons (KLM), constitué de scintillateurs permettant d'identifier ces particules qui, par nature, sont beaucoup moins susceptibles d'être stoppées par les couches plus internes du détecteur.

Les différentes étapes du traitement des données, de la reconstruction des particules et des événements s'effectuent par le biais de *basf2*, le logiciel d'analyse spécifique à Belle II. La collaboration fournit des échantillons de simulations Monte-Carlo (MC) pour les analyses, que nous exploitons ici à l'instar des données de collisions collectées entre 2019 et 2021, équivalentes à une luminosité intégrée de 189.88 fb^{-1}.

Mesure du taux de fausses traces dans les événements $e^+e^- \to \tau^+\tau^-$

Une étude secondaire réalisée dans le cadre de cette thèse a consisté en la mesure du taux de fausses traces dans les événements $e^+e^- \to \tau^+\tau^-$ produits par Belle II. Ce taux concerne :

- les fausses traces proprement dites, qui résultent d'une combinaison erronée de points d'impact dans le détecteur ;

[5] Le tonneau est la partie longeant le tuyau du collisionneur, alors que les bouchons avant et arrière sont placés dans le plan transverse.

- les clones, des traces produites par une seule et même particule mais que l'algorithme de reconstruction n'a pas su fusionner ;

- les traces que génère le bruit de fond induit par les faisceaux, bien qu'elles puissent être correctement reconstruites.

Nous mesurons le taux de fausses traces r_{fake} avec une méthode dite *tag-and-probe*, où l'on compare des désintégrations de paires $\tau^+\tau^-$ en quatre traces (*tag*), ici $\tau^\pm\,[\to \pi^+\pi^-\pi^\pm\nu_\tau] + \tau^\mp\,[\to e^\mp\bar{\nu}_e\nu_\tau]$, avec la présence éventuelle d'une cinquième fausse trace (*probe*). Nous obtenons alors

$$r_{\text{fake}} = \frac{N_5}{N_4 + N_5}, \tag{1}$$

où N_4 et N_5 sont respectivement les nombres d'événements qui contiennent quatre et cinq traces.

En plus des données, nous étudions des échantillons MC qui peuvent (*run-dependent MC*, MCrd) ou pas (*run-independent MC*, MCri) simuler les conditions réelles de la prise des données, entre chaque cycle (*run*) qui durera au plus 8 heures. Ces conditions ayant des répercussions sur les performances de la reconstruction des traces, on s'attend à ce que les échantillons de simulation MCrd rendent plus précisément compte de l'incidence des fausses traces dans les données.

Les événements de signal dans la simulation MC sont définis, aussi bien avec que sans fausse trace, comme deux taus se désintégrant respectivement en trois et une particules chargées, avec toutes les traces du *tag* qui proviennent effectivement d'une paire $\tau^+\tau^-$. Nous utilisons pour cela l'information de ce qui fut réellement généré dans la simulation (*truth-matching*).

Toutes les traces sont sélectionnées de telle sorte que leur origine se situe au plus à quelques centimètres du point de collision, tandis que les quatre traces du *tag* se voient imposer des conditions sur les impulsions transverses et les variables d'identification des particules. La suppression du bruit de fond s'opère via une série de sélections supplémentaires, notamment sur la topologie des événements, les masses invariantes des pions et les impulsions des taus.

La comparaison entre données et simulations montre un déficit d'événements MC, avec environ 12 % en moins pour MCrd et jusqu'à 25 % pour MCri si une cinquième fausse trace est présente ; pour quatre traces seulement, nous observons un déficit de 16 % et 6 % pour MCrd et MCri, respectivement. Il apparaît donc que le désaccord observé dans la simulation MCrd est davantage cohérent entre les deux types de reconstruction, comparé à la simulation MCri. Le taux de fausses traces mesuré dans MCrd est donc plus fidèle aux données.

Nous estimons le rendement du signal dans les données à partir de la pureté du signal dans la simulation, ce qui mène à deux résultats selon que ce facteur provienne des échantillons MCrd ou MCri. Nous considérons par défaut les rendements obtenus

dans le premier cas ; ainsi, la mesure du taux de fausses traces donne

$$r_{\text{fake}} = (1.30 \pm 0.05_{\text{stat}} \pm 0.14_{\text{sys}})\,\%, \tag{2}$$

avec les incertitudes statistique (stat) et systématique (sys). Cette dernière correspond à la différence maximale entre les taux mesurés dans les données, avant et après multiplication des rendements N_4 et N_5 par la pureté du signal dans la simulation.

Ce résultat permet d'évaluer les performances de la reconstruction des traces dans les événements $e^+e^- \to \tau^+\tau^-$; sur le long terme, il peut constituer un indicateur de la dégradation ou de l'amélioration des conditions expérimentales.

Recherche de désintégrations violant la saveur leptonique $\tau \to \ell\phi$

L'analyse de données que nous avons menée pour la recherche de désintégrations $\tau \to \ell\phi$ avec Belle II diffère des précédentes études produites par les collaborations de Belle et BaBar, en ce que nous reconstruisons des événements de collisions $e^+e^- \to \tau^+\tau^-$ où seul l'un des deux taus doit remplir les conditions du signal recherché, sans restriction quant à l'évolution du second tau.

Cette approche, dite "inclusive", a pour but d'accroître l'efficacité de sélection du signal par rapport la méthode "exclusive" de Belle et BaBar, qui requiert au plus quatre traces dans l'événement, trois pour le signal et une pour le tau de charge opposée, limitant de fait la proportion d'événements de signal reconstruits[6]. Cependant, cette approche inclusive introduit un surplus de bruit de fond qu'il est nécessaire de compenser par des critères de sélection performants, afin que le gain en efficacité du signal demeure significatif.

En plus des données expérimentales et des échantillons de simulation que nous avons décrits plus haut, nous disposons de simulations MC pour les deux canaux de désintégration $\tau \to e\phi$ et $\tau \to \mu\phi$, générées sans suivre aucun modèle de nouvelle physique[7].

L'analyse repose avant tout sur la définition d'une "région du signal" dans l'espace bidimensionnel (M_τ, ΔE_τ), où M_τ est la masse mesurée du tau reconstruit comme signal et ΔE_τ est la différence entre les énergies du tau et du faisceau ($\sqrt{s}$) dans le référentiel du centre de masse. En effet, les distributions du signal simulé forment un pic autour de $M_\tau \approx 1.777$ GeV/c^2 (masse du tau [10]) et de $\Delta E_\tau = 0$ GeV, une zone où l'on serait donc davantage susceptible de détecter une désintégration $\tau \to \ell\phi$ au sein des données. Les événements issus de ces données et appartenant à cette région sont dissimulés jusqu'à la fin de l'analyse, afin d'éviter tout biais

[6] Environ 15 % des désintégrations du tau mènent à un état final avec trois particules chargées [10], sans compter la présence possible de traces venant du bruit de fond.
[7] Désintégration dans l'espace des phases.

humain dans l'obtention des résultats.

Pour commencer, nous appliquons une sélection sur les traces similaire à celle de la section précédente quant à leur origine spatiale, à quoi nous ajoutons que toute trace supposée provenir

- d'un *électron*, doit être identifiable en tant que telle avec une probabilité supérieure à 0.9 (electronID$^{\text{noSVD, noTOP}}$[8], eID) ;
- d'un *muon*, doit être identifiable en tant que telle avec une probabilité supérieure à 0.99 (muonID$^{\text{noSVD}}$, μID) ;
- d'un *kaon*, doit être identifiable en tant que telle avec une probabilité supérieure à 0.5 (kaonID, KID) s'il s'agit du kaon issu de ϕ avec la plus grande impulsion transverse, et doit être contenue dans l'angle solide couvert par la CDC.

Ces conditions s'adaptent aux performances des différentes variables d'identification, celles du μID étant généralement plus élevées que les autres.

Les éventuelles particules non chargées dans l'événement, à savoir ici les photons et pions neutres π^0, sont également reconstruites pour évaluer leur nombre qui peut varier selon la nature des particules produites après la collision. Par ailleurs, l'électron du canal $\tau \to e\phi$ a une chance non négligeable d'émettre un photon (*rayonnement d'état final*) ; dans l'espace (M_τ, ΔE_τ), de tels événements se distribuent sous une forme d'"aile" partant du pic vers les valeurs basses des deux variables, donc hors de la région du signal. Nous pallions ce problème en combinant à l'électron tous les photons qui se propagent à l'intérieur d'un cône autour de son impulsion, dont nous avons optimisé les paramètres[9] afin de maximiser le gain d'efficacité dans cette région.

Deux kaons chargés sont combinés en un méson ϕ dont la masse mesurée M_ϕ doit se situer entre 1.014 et 1.024 GeV/c^2 ($m_\phi \simeq 1.019$ GeV/c^2 [10] pour la masse "officielle"). Les événements sont restreints à une fenêtre relativement large en M_τ et ΔE_τ et à un nombre de traces au plus égal à 6 ; topologiquement, le lepton et le méson ϕ doivent se trouver dans un seul et même hémisphère, dont l'axe est la direction de la "poussée" (*thrust*) totale de l'événement. Nous requérons par ailleurs qu'au moins une des conditions d'un ensemble de déclencheurs (*triggers*) soit satisfaite, parmi une liste de contraintes dédiées au rejet des événements à basse multiplicité (par exemple, les diffusions Bhabha).

Enfin, nous mesurons les propriétés du reste de l'événement (*Rest-Of-Event*, ROE), qui rassemble toutes les particules détectées ne participant pas à la reconstruction. De même que pour les particules neutres, les quantités liées au ROE peuvent éventuellement aider à distinguer le bruit de fond.

[8] Les détecteurs SVD et TOP ne contribuent pas à l'identification des électrons du fait de leurs faibles performances.

[9] Angle du cône et énergie minimale du photon.

Après cette étape, nous mettons en place un ajustement des distributions du signal pour les variables M_τ et ΔE_τ, afin de définir la région du signal. Cet ajustement se base sur la convolution de deux lois normales et d'une fonction *Crystal Ball*. La somme pondérée des écarts types nous donne la résolution σ [10], que nous utilisons pour établir trois régions dans l'espace (M_τ, ΔE_τ), en nombre de pas de résolution par rapport à la valeur moyenne du pic :

- la région "$\pm$ 20σ", qui redéfinit la fenêtre dans laquelle tous les événements doivent se situer ;
- la région "$\pm$ 5σ", qui à l'heure actuelle est dissimulée dans les données par principe de précaution ;
- la région "$\pm$ 3σ", qui est la région du signal à proprement parler, dans laquelle les limites sur les rapports d'embranchement seront obtenues.

Les résolutions et la définition exacte des régions sont détaillées dans le Tableau 2.

Tableau 2 : Régions du signal et résolutions obtenues par un ajustement sur les distributions des événements de signal.

Mode	Variable	Région 3σ	Région 5σ	Région 20σ	Résolution
$e\phi$	M_τ	$[1.743, 1.811]$	$[1.721, 1.834]$	$[1.551, 2.004]$	11.3 ± 0.2 MeV/c^2
	ΔE_τ	$[-0.125, 0.120]$	$[-0.207, 0.202]$	$[-0.8, 0.8]$	41.0 ± 0.8 MeV
$\mu\phi$	M_τ	$[1.760, 1.794]$	$[1.749, 1.805]$	$[1.665, 1.890]$	5.6 ± 0.2 MeV/c^2
	ΔE_τ	$[-0.083, 0.080]$	$[-0.137, 0.134]$	$[-0.5, 0.5]$	27.1 ± 0.6 MeV

À partir de variables notamment cinématiques et topologiques relatives aux différentes particules (lepton, kaons, ϕ, τ, photons et π^0), à l'énergie manquante, au ROE ou encore à l'événement en général, nous élaborons une stratégie de sélection du signal et suppression du bruit de fond comme suit :

1. une sélection préliminaire, ou présélection, est appliquée sur un nombre réduit de variables afin d'enlever les contributions au bruit de fond qui se distinguent le plus du signal dans leurs distributions ;
2. un arbre de décision optimisé (*Boosted Decision Tree*, BDT) est entraîné sur une partie des échantillons MC, avec en sortie une estimation de la probabilité pour chaque événement de correspondre au signal recherché ("probabilité du

[10] Cette résolution n'a pas les mêmes propriétés que l'écart type d'une loi normale, notamment en ce qui concerne la règle "68-95-99.7".

signal"), grâce à laquelle nous pourrons écarter ceux qui obtiennent un score plus faible qu'un certain critère à établir ;

3. dans l'espace bidimensionnel $E^{\mathrm{CM}}_{\mathrm{lepton}}$-$E^{\mathrm{CM}}_{\phi}$, respectivement les énergies du lepton et du ϕ dans le référentiel du centre de masse, nous conservons seulement la région qui répond à

$$E^{\mathrm{CM}}_{\phi} > -1 \times E^{\mathrm{CM}}_{\mathrm{lepton}} + 5.2, \tag{3}$$

dite "sélection 2D", avec un fort pouvoir de séparation entre signal et bruit de fond. Elle n'est appliquée que dans un dernier temps afin de ne pas réduire outre mesure les ressources d'entraînement du BDT.

Pour mettre en place ce BDT, nous divisons les échantillons de simulation en trois partie égales : un échantillon pour l'*entraînement* lui-même, un second pour la *validation* des performances du BDT[11] et enfin un troisième échantillon de *test*, qui retranscrit le rôle des données expérimentales dans le cadre de la simulation, c'est-à-dire qu'il n'est consulté que pour l'obtention des résultats dans le MC.

Le critère de sélection sur la probabilité du signal est fixé par minimisation de la limite supérieure sur le rapport d'embranchement du canal considéré, évaluée sur tout l'intervalle $[0, 1]$ en utilisant une moitié de l'échantillon de validation, réduite au préalable à la région 5σ. Le calcul de cette limite suit la méthode CLs dans sa formulation asymptotique [11] et est réalisé via le logiciel `pyhf` de Python [12].

Les résultats de l'entraînement et de l'optimisation du critère du BDT sont illustrés dans la Figure 2.

Dès lors que toutes les étapes de la sélection sont appliquées aussi bien sur les données que sur les échantillons de simulation, nous effectuons une première comparaison "données-MC", en prenant soin de dissimuler la région à 5σ dans les deux cas, ne laissant que les bandes latérales de 5 à 20σ. Nous constatons un écart dans les rendements totaux, visible dans la Figure 3 avec notamment des rapports "données sur simulation" égaux à $22 \pm 4_{\mathrm{stat}}$ (canal $\tau \to e\phi$) et $1.4 \pm 0.6_{\mathrm{stat}}$ (canal $\tau \to \mu\phi$).

Le rapport observé dans le canal électronique dévie largement de l'unité et ne peut être mis sur le compte des fluctuations statistiques. Cela semble en réalité dû à un excès de photoproduction du méson ϕ dans les données ($e^+e^- \to e^+e^-\phi$), que le MC ne reproduit pas fidèlement. Nous proposons comme solution l'entraînement d'un second BDT, avec pour mission de distinguer les événements de signal simulés des événements de données se situant dans la fenêtre $-0.1 < \Delta E_\tau < 0.4$ GeV, là où se concentre l'excès. Pour s'assurer qu'un éventuel signal enregistré ne soit accidentellement mis en évidence, l'entraînement s'opère sur un dixième de l'échantillon de données, dans les bandes latérales et sans appliquer de sélection après l'étape de reconstruction.

[11] Cette validation permet en particulier de réduire le risque de surapprentissage.

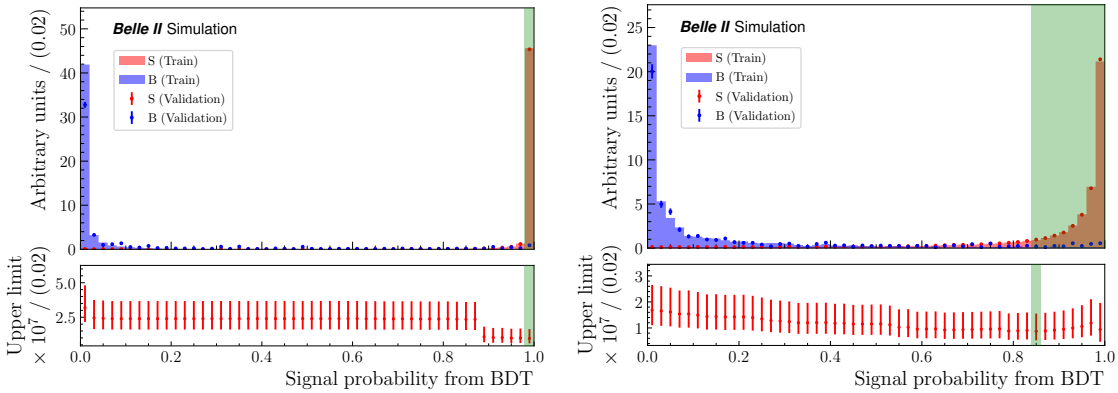


FIGURE 2 : Probabilité selon le BDT qu'un événement de l'échantillon d'entraînement (histogramme) ou de test (points) corresponde au signal recherché dans les canaux $\tau \to e\phi$ (gauche) et $\tau \to \mu\phi$ (droite). S et B signifient respectivement "signal" et "bruit de fond". L'encart inférieur indique la limite supérieure sur le rapport d'embranchement à un niveau de confiance de 90 %, évaluée sur l'échantillon de validation. La région verte correspond aux valeurs sélectionnées après minimisation de la limite.

Nous appliquons une coupure stricte sur le résultat du second BDT, en supprimant tout événement qui a une probabilité supérieure à 0.001 de se trouver dans ladite fenêtre. Ceci améliore sensiblement l'accord entre données et MC, cependant pas de façon suffisante pour atteindre un rapport compatible avec l'unité.

Les résultats majeurs de la sélection du signal sont présentés dans le Tableau 3, tandis que les distributions dans l'espace $(M_\tau, \Delta E_\tau)$ des événements de données et du bruit de fond MC sont illustrées dans la Figure 4. Le nombre d'événements de bruit de fond attendu dans les données au sein de la région du signal à 3σ, N_{att}, est estimé comme

$$N_{\text{att}} = N_{\text{obs}}^{\text{BLR}} \frac{N_{\text{MC}}^{\text{RS}}}{N_{\text{MC}}^{\text{BLR}}}, \tag{4}$$

où $N_{\text{MC}}^{\text{RS}}$ est le rendement du MC dans la région du signal, alors que $N_{\text{obs}}^{\text{BLR}}$ et $N_{\text{MC}}^{\text{BLR}}$ sont respectivement les rendements obtenus dans les données et la simulation à l'intérieur de la bande latérale réduite à $\pm 3\sigma$ en ΔE_τ (voir à nouveau la Figure 4).

Pour les deux canaux de désintégration, nous attendons environ 0.7 événement de bruit de fond dans les données à l'intérieur de la région du signal.

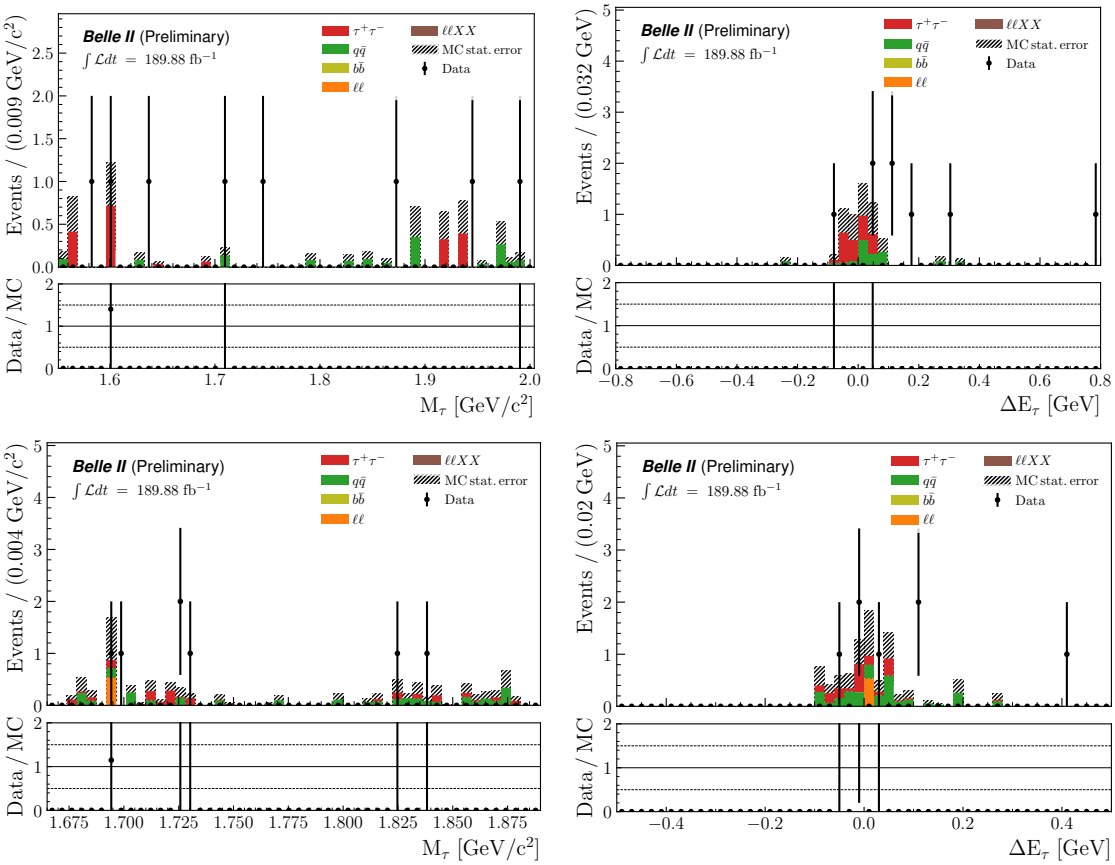


FIGURE 3 : Comparaisons données-MC des distributions de M_τ (gauche) et ΔE_τ (droite) dans les bandes latérales à $5 - 20\sigma$, dans les canaux $\tau \to e\phi$ (haut) et $\tau \to \mu\phi$ (bas), après application de la présélection, du critère du BDT et de la coupure 2D.

TABLEAU 3 : Résumé des résultats de la sélection du signal et suppression du bruit de fond, évalués avec la simulation (échantillon de test) et dans les données. Les résultats sont donnés dans des régions particulières de l'espace (M_τ, ΔE_τ) : la région du signal à 3σ (RS), les bandes latérales à $5 - 20\sigma$ (BL) ou les bandes réduites à la région 3σ en ΔE_τ (BLR). Les astérisques signalent les nombres issus de l'échantillon de validation. La proportion de bruit de fond (RS/BR) dans le MC est la moyenne des rapports dans les échantillons de test et de validation.

Résultat		Région	Mode	
			$e\phi$	$\mu\phi$
Efficacité du signal $\varepsilon_{\ell\phi}$		RS	6.0%	7.2%
Rendement du bruit de fond dans le MC		BL	3.34	5.06
		BLR	3.10 / 2.53*	4.10 / 4.06*
		RS	0.38 / 0.36*	0.83 / 0.76*
Proportion de bruit de fond dans le MC		RS / BLR	0.13	0.19
Rendement dans les données	N_{obs}	BL	8	7
		BLR	5	4
	N_{exp}	RS	0.66	0.78
Rapport Données / MC		BL	$2.4 \pm 1.0_{\text{stat}}$	$1.4 \pm 0.6_{\text{stat}}$
		BLR	$1.6 \pm 0.8_{\text{stat}}$	$1.0 \pm 0.5_{\text{stat}}$

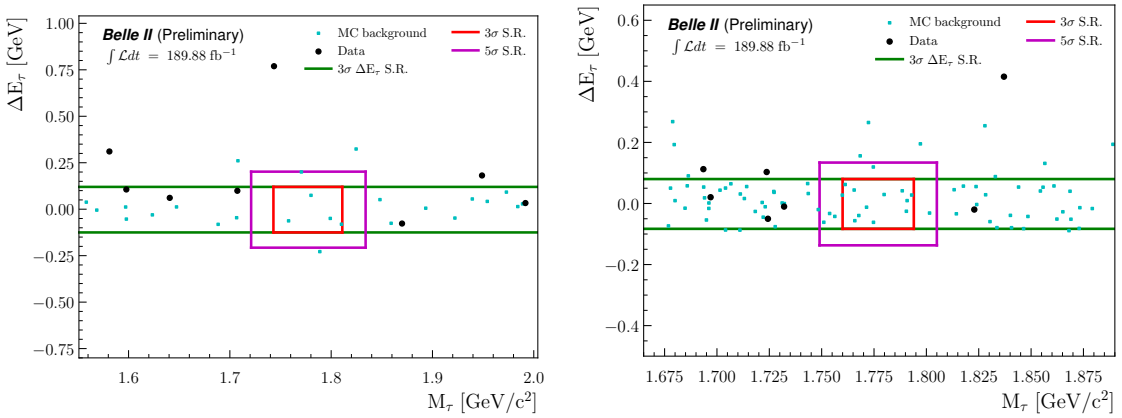


FIGURE 4 : Graphique en nuage de points des événements de données et de bruit de fond MC dans les canaux $\tau \to e\phi$ (gauche) et $\tau \to \mu\phi$ (droite), après application de toutes les sélections. Les rectangles rouge et magenta représentent les régions du signal à 3 et 5σ, dissimulées dans les données, tandis que les droites horizontales vertes délimitent la bande $\pm 3\sigma$ en ΔE_τ, qui sert à estimer le nombre d'événements attendus dans la région du signal.

La limite supérieure sur les rapports d'embranchement des canaux $\tau \to \ell\phi$ est calculée selon

$$\mathcal{B}_{\mathrm{LS}}(\tau \to \ell\phi) = \frac{s_{\mathrm{LS}}}{L \times 2\sigma_{\tau\tau} \times \varepsilon_{\ell\phi}}, \tag{5}$$

où L est la luminosité intégrée, $\sigma_{\tau\tau}$ la section efficace de production de paires $\tau^+\tau^-$, $\varepsilon_{\ell\phi}$ l'efficacité de sélection du signal dans la région à 3σ et s_{LS} la limite supérieure sur le rendement du signal dans cette même région. Avant de mener à bien ce calcul, il est nécessaire d'évaluer les incertitudes systématiques sur les différentes quantités concernées, résumées dans le Tableau 4.

En particulier, l'incertitude systématique sur le critère du BDT est obtenue grâce à une méthode de correction des distributions du signal, avant même l'entraînement du BDT, basée sur des "échantillons de contrôle" :

1. nous reconstruisons tout d'abord la désintégration $D_s \to \phi\pi$ dans les données et la simulation, avec un état final similaire à notre signal, avec la substitution du lepton par un pion chargé. Il s'agit ensuite d'extraire les distributions du rapport "données/MC" observé *dans cet échantillon de contrôle* pour quelques variables, afin de redimensionner lesdites distributions cette fois-ci *dans le signal* en fonction du résultat obtenu ;

2. nous répétons cette méthode avec un second échantillon de contrôle, $\tau \to 3\pi\nu_\tau$, afin de corriger les variables pour lesquelles l'usage de la désintégration $D_s \to \phi\pi$ n'était pas adapté (principalement les variables du ROE).

TABLEAU 4 : Origines et valeurs des incertitudes systématiques sur les quantités impliquées dans le calcul de la limite supérieure, en fonction du mode de désintégration. Les incertitudes citées sur l'identification des particules incluent également les contributions statistiques.

Quantité concernée	Source	Mode	
		$e\phi$	$\mu\phi$
$\varepsilon_{\ell\phi}$	Identification des particules	0.11%	0.16%
	Efficacité du tracking	0.52%	
	Efficacité des déclencheurs	0.4%	0.9%
	Critère du BDT	0.48%	0.95%
L	Luminosité	0.6%	
$\sigma_{\tau\tau}$	Section efficace $\tau^+\tau^-$	0.003 nb	
N_{att}	Facteur d'impulsion	$^{+0.0}_{-0.0}$	$^{+0.0}_{-0.2}$

L'idée ici est de réduire les potentielles différences entre ce qui serait du signal dans les données expérimentales (qu'on ne peut connaître sans une détection avérée) et la simulation des canaux $\tau \to \ell\phi$, à partir de ce que nous apprennent les deux canaux de contrôle. Cela nous permet ainsi d'évaluer les biais introduits dans l'élaboration du BDT, qui s'appuie nécessairement sur les distributions des variables en entrée. En l'occurrence, l'incertitude systématique due au BDT que nous affectons à $\varepsilon_{\ell\phi}$ est la différence entre les efficacités du signal obtenues après les deux sélections, corrigée et non corrigée.

Pour finir cette analyse, nous estimons les limites attendues sur les rapports d'embranchement en suivant à nouveau la méthode CLs à un niveau de confiance de 90 %, ce que représente la Figure 5. Les limites sont

$$\mathcal{B}^{90}_{\text{LS, att}}(\tau \to e\phi) = 1.18 \times 10^{-7}, \tag{6}$$

$$\mathcal{B}^{90}_{\text{LS, att}}(\tau \to \mu\phi) = 1.05 \times 10^{-7}. \tag{7}$$

En extrapolant aux luminosités des expériences précédentes (voir Tableau 1), la limite attendue sur le rapport d'embranchement de $\tau \to \mu\phi$ est égale à celle de Belle et légèrement meilleure que celle de BaBar (-17 %). En revanche, le résultat évalué pour le canal électronique pâtit du bruit de fond dans les données, qui n'a pu être suffisamment réduit sans trop compromettre l'efficacité du signal.

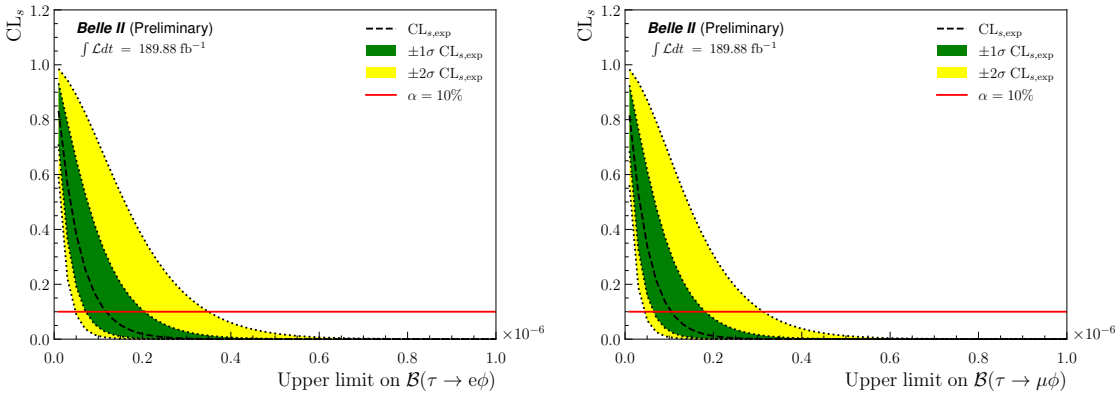


FIGURE 5 : CL$_s$ attendu en fonction de la limite supérieure sur les rapports d'embranchement de $\tau \to e\phi$ (gauche) et $\tau \to \mu\phi$ (droite). La droite horizontale rouge dénote un niveau de confiance à 90 %.

Abstract

We report the search of lepton flavour violating $\tau \to \ell\phi$ decays at the Belle II experiment, based on an integrated luminosity $\int \mathcal{L}dt = 189.88$ fb^{-1} of data collected from 2019 to 2021, at the $\Upsilon(4S)$ meson resonance energy $\sqrt{s} = 10.58$ GeV. We look for signal signatures that could result from $e^+e^- \to \tau^+\tau^-$ events, by reconstructing one tau going into a charged lepton and a meson $\phi \to K^+K^-$, while the oppositely charged tau is left unconstrained to maximise the signal selection efficiency. Using Monte Carlo simulation, we define a background suppression strategy that combines cut-based selections and boosted decision trees, taking as inputs kinematic, topological and other event related variables. These selections are tuned in order to minimise as much as possible the expected upper limits on the branching fractions.

Although the full data is not yet unblinded, we perform an estimation of the expected upper limits on $\tau \to e\phi$ and $\tau \to \mu\phi$ branching fractions of the order of 10^{-7}. At equivalent luminosities, our result on the muon decay mode is comparable to the expected limit of the previous Belle experiment's analysis and slightly better than the one from the BaBar experiment (-17%).

In addition, we conduct a measurement of the rate of fake tracks in $e^+e^- \to \tau^+\tau^-$ events. To this end, we devise an original method applied on τ-pairs decaying into four tracks, with the possible presence of an additional fake one. This rate, measured in data and simulation, is found to be around 1% in both cases.

Contents

Introduction

The Standard Model (SM) of particle physics is a theory that, to this day, gives the most precise description of fundamental and subatomic particles. It is a relativistic quantum field theory of the electromagnetic, weak and strong interactions, and of the particles subject to these interactions: leptons, quarks, gauge bosons, hadrons... Each particle has its own properties, whether it be the mass, the charge (electric or color), the spin, or miscellaneous quantum quantities that govern how they interact with each other. Among them is the lepton number, related to the lepton flavour, that is the very nature of these particles.

In the SM, lepton flavour appears to be in most cases accidentally conserved, except for neutrino oscillations. This phenomenon introduces charged lepton flavour violating (cLFV) decays through charged currents of the weak interaction, predicted to occur at rates close to 10^{-50}, far beyond the reach of current experimental sensitivities. Observing cLFV decays would therefore consist in an indisputable evidence of physics beyond the SM, or "new physics", which is already suspected since the observation of phenomena hardly explainable by the SM (dark matter, matter-antimatter asymmetry, hierarchy problem...).

Tau LFV decays like $\tau \to \ell V^0$, where ℓ is an electron or a muon and V^0 is a neutral vector meson, and $\tau \to \ell\phi$ in particular have gained some interest due to the B meson anomalies reported by LHCb and B-factories in $b \to s\ell^+\ell^-$ and $b \to c\ell\bar{\nu}$ transitions. Indeed, leptoquark models that could accommodate both anomalies predict large enhancements of the $\tau \to \mu\phi$ decay, with branching fractions close to the current upper limit. In that respect, these processes are possibly accessible to the sensitivity of the Belle II experiment, which exploits the SuperKEKB electron-positron collider and its detector at the KEK laboratory in Japan.

We discuss the SM and theories behind the possible new physics that motivate tau LFV searches in Chapter 1, while a presentation of the Belle II experiment and how it operates is taken care of in Chapter 2. This includes an explanation of what fake and clone tracks in Belle II are, and of which we measure the rate in Chapter 3 with $e^+e^- \to \tau^+\tau^-$ events[12]. Finally, we report in Chapter 4 the search for $\tau \to \ell\phi$ decays based on an untagged (or inclusive) selection.

[12]Qualification task required to obtain the rights of authorship in the Belle II collaboration.

1. Beyond the Standard Model with τ lepton flavour violation

One could argue that the very essence of existence is interaction. When it comes to the elementary components of what fills the universe, fundamental particles, the interactions that govern their course through time and space, if they don't proceed from any other, are called fundamental as well. The knowledge gathered so far on that matter forms a remarkably precise and consistent theory, the Standard Model of particle physics.

While it has been experimentally supported time and time again and rightly praised, the Standard Model is not without its flaws. Far from being a "theory of everything", it is yet unfit to account for some curious phenomena that have confidently been observed, if they don't contradict its predictions at all. For that reason, there is a clear need to refine and even look beyond the Standard Model.

Following the established theory, charged lepton flavour violation is a phenomenon that is not allowed. Although unobserved to this day, it can nonetheless occur as the rare result of neutrino oscillations, a behaviour that those particles display because of their unexpected non-zero masses. If some distinct and not yet discovered physical processes were to foster charged lepton flavour violation, specific particle decays could be detected in the near future. $\tau \to l\phi$ is one of them.

In this chapter, we will start off with an overview of the Standard Model and its description of elementary particles through their interactions, then we will discuss the problem of lepton flavour violation and why looking for it in charged particles could reveal hints of physics beyond the Standard Model, in particular with τ leptons and the $\tau \to l\phi$ decay.

1.1. Standard Model of particle physics

1.1.1. Elementary particles

The Standard Model of particle physics (SM) is a theory that describes the propagation and interactions of elementary particles[1].

[1] The first section of this chapter is based on many works that cover this topic: some literature [13, 14], numerous online resources that include [10, 15, 16, 17, 18] among others, and various lectures I attended in the past.

To this day, there are thirty-seven[2] elementary particles in the SM, summarised in Figure 1.1. They differ by at least one of the various physical properties that we measure, mainly the mass (in units of eV/c, where $1\,\text{eV} = 1.602\,176\,634\times10^{-19}\,\text{J}$ and $c = 299\,792\,458\,\text{m}\cdot\text{s}^{-1}$ is the speed of light), the electric charge (in units of the elementary charge $e = 1.602\,176\,634\times10^{-19}\,\text{C}$, omitted thereafter) and the spin (dimensionless, unless talking about the spin angular momentum which is proportional to the reduced Planck constant $\hbar \simeq 1.054\,571\,817\times10^{-34}\,\text{J}\cdot\text{s}$). There are also other quantum numbers which we shall discuss later on. Particles are firstly categorised either as fermions or bosons, depending on whether they have half-odd-integer or integer spin, respectively.

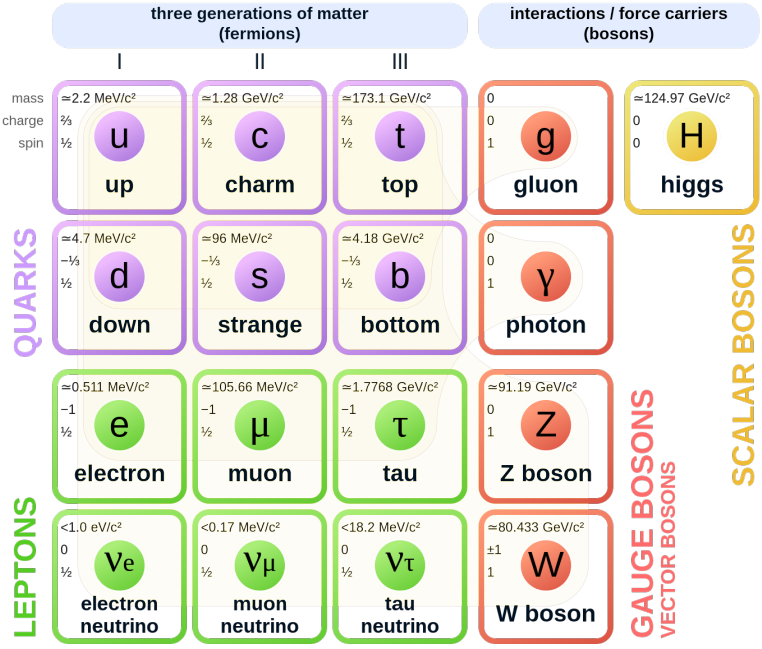


Figure 1.1: Particles of the Standard Model and their properties (mass, charge, spin). Taken from [19].

All elementary particles of the SM that are fermions have a spin equal to 1/2. Like all half-odd-integer spin particles, they obey the Fermi-Dirac statistics that results from the Pauli exclusion principle, which states that identical fermions in a quantum system cannot have the same quantum state at the same time. Elementary fermions are themselves either leptons or quarks.

As for the bosons, they are either scalar if the spin is equal to 0, or vector if it is equal to 1. The only scalar boson of the SM is the Higgs boson. Unlike fermions,

[2] If neutrinos are not their own antiparticles (see Majorana particles).

bosons are not subject to the Pauli exclusion principle and follow the Bose-Einstein statistics.

1.1.1.1. Leptons

Among the leptons are the electron (e^-), the muon (μ^-) and the tau (τ^-), with same electric charges q: $q_{e^-} = q_{\mu^-} = q_{\tau^-} = -1$, but increasing masses. The flavour of a lepton corresponds to its nature: either electron, muon or tau.

Neutrinos are also leptons, although with no electric charge (also said neutral) and were assumed massless until the discovery of the neutrino oscillation phenomenon. There are three types of neutrinos, one for each lepton flavour (ν_e, ν_μ, ν_τ).

Electron, muon, tau and neutrinos have antiparticles (e^+, μ^+, τ^+, $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$), named with the prefix "anti", although the antielectron is more commonly called positron. Antileptons have opposite sign electric charge (zero for antineutrinos) and opposite lepton number L, defined as the difference between the number of leptons and the number of antileptons. This quantum number is conserved in the SM, which means that L must be the same in both initial and final states of an interaction, which is also predicted to be true when considering each flavour separately.

Pairing leptons together according to their flavour gives three doublets called lepton generations or families, ranked in increasing order of charged lepton mass. As such, leptons tend to decay into generations of lower order. The MS, however, doesn't explain why there are exactly three generations of leptons.

1.1.1.2. Quarks

As in the case of leptons, quarks come in three generations, ranked by increasing masses, where each doublet has one quark with electric charge $q_u = +2/3$ ("up-type" quark), while it is $q_d = -1/3$ for the other ("down-type" quark). In this order, the first generation quarks are called up (u) and down (d) quarks, the second generation ones are the charm (c) and strange (s) quarks, finally the third generation corresponds to the top (t) and bottom (or sometimes beauty, b) quarks. However, unlike leptons, every quark has its own unique flavour. Each of them has also an antiparticle ($\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$), with opposite electric charge.

Quarks are always found in bound states with other quarks, except the top, the heaviest elementary particle which decays before it can form such a state. The resulting composite particles are called hadrons, or more specifically:

- baryons when they are made of three quarks or three antiquarks,
- mesons when they are made of one quark and one antiquark, or a quantum combination of quark-antiquark pairs (whatever the flavours).

The explanation to how quarks combine with each other is the colour charge, the

principle of which is quite similar to additive colour mixing: quarks are either red, blue or green, while antiquarks are antired, antiblue or antigreen, giving hadrons that are white or colourless.

The most common hadrons are the proton (uud) and the neutron (udd), both baryons. There are many other baryons and mesons which are unstable and decay into lighter hadrons or elementary particles, therefore only produced in specific environments involving high energies, like particle colliders or astrophysical processes.

Although quark flavour is not conserved in the SM, the baryon number B is. It is defined as the difference between the number of baryons and the number of antibaryons. Therefore, $B = 1$ for baryons, $B = -1$ for antibaryons and $B = 0$ for mesons.

1.1.1.3. Gauge bosons

Elementary vector bosons are referred to as gauge bosons, since they carry the three fundamental interactions included in the SM that arise from gauge symmetries. Those bosons are:

- the photon (γ), massless and neutral, carrying the electromagnetic interaction which involves only electrically charged particles,
- the W^+, W^- and Z^0 bosons, mediators of the weak interaction that concerns all the fermions,
- the gluons (g), eight different types of massless and neutral bosons that bind quarks together in hadrons, what is known as strong interaction.

1.1.1.4. Higgs boson

The Higgs boson is the last elementary particle discovered by physicists, in 2012 at the European Organisation for Nuclear Research (CERN) [20]. It is the boson associated to the scalar field of the same name, responsible for the masses of all elementary particles that interact with it through the Brout-Englert-Higgs (BEH) mechanism. This therefore concerns massive gauge bosons, quarks and leptons, with different coupling strengths: the stronger it is, the heavier the particle.

1.1.2. Free quantum fields

The SM is based on quantum field theory (QFT), which results from quantum mechanics, field theory and special relativity, key pillars of particle physics. The mathematical framework that models spacetime with the properties of special relativity is called Minkowski space, which has four dimensions: one for time and

three for space. To a point of the Minkowski space will therefore be associated a four-vector with its spacetime coordinates.

A transformation that changes a frame of the Minkowski space into another, while preserving the scalar product of two four-vectors, is called a Lorentz transformation. Adding to that spacetime translations gives us the Poincaré group of transformations that preserve distances. In QFT, particles are described not just as fields, but as field operators (canonical or second quantisation) that are invariant under the Poincaré group transformations, thus satisfying the requirements of special relativity. There are three types of fields that are of interest for us: scalar, spinor and vector fields. How the fields propagate or interact with each other is formulated thanks to Lagrangian densities $\mathcal{L}$.

In the following sections of this chapter only, the Einstein notation will be used for summation of covariant and contravariant indices, as well as the natural units: $c = \hbar = 1$. Unless specified otherwise, fields will be noted φ, or $\varphi^\dagger$ for the Hermitian conjugates, the dependence on the spacetime coordinates being implicit.

1.1.2.1. Scalar fields

Scalar fields correspond to particles with zero spin, thus to scalar bosons in the SM. A real or complex scalar field is equivalent to a neutral or charged scalar boson, respectively.

The Lagrangian density of a free complex scalar field is:

$$\mathcal{L}_{\text{KG-complex}} = \partial_\mu \varphi^\dagger \partial^\mu \varphi - m^2 \varphi^\dagger \varphi, \tag{1.1}$$

where ∂_μ is the four-gradient and m is the invariant mass of the particle. For a free real scalar field, the Lagrangian density is instead:

$$\mathcal{L}_{\text{KG-real}} = \frac{1}{2}\left(\partial_\mu \varphi \partial^\mu \varphi - m^2 \varphi^2\right). \tag{1.2}$$

There is no interaction terms in these Lagrangian densities, that is why we call them free. This is verified by the fact that the equation of motion (here, the Klein-Gordon equation), derived with the Euler-Lagrange equation, is linear in the field.

1.1.2.2. Spinor fields

Spinor fields correspond to particles with spin equal to 1/2, thus to fermions in the SM. To be more accurate, we can call them bispinors since they are four-dimensional objects that can be divided into two spinors of two components each. This is justified by the particular behaviour of spinors under Lorentz transformations.

The Lagrangian density of a free spinor field is the Dirac Lagrangian:

$$\mathcal{L}_{\text{Dirac}} = \bar{\varphi}\left(i\gamma^\mu \partial_\mu - m\right)\varphi, \tag{1.3}$$

where γ^μ are the four Dirac matrices and $\bar{\varphi} = \varphi^\dagger \gamma^0$. The Euler-Lagrange equation applied to $\mathcal{L}_{\text{Dirac}}$ gives the Dirac equation.

By writing φ in a specific basis, called Weyl basis, one can define two-dimensional spinors ψ_{L} and ψ_{R}, as well as two operators P_{L} and P_{R}:

$$\varphi = \begin{pmatrix} \psi_{\text{L}} \\ \psi_{\text{R}} \end{pmatrix}, \; P_{\text{L}} = \frac{1}{2}\left(\mathbb{1}_4 - \gamma^5\right), \; P_{\text{R}} = \frac{1}{2}\left(\mathbb{1}_4 + \gamma^5\right), \tag{1.4}$$

where $\mathbb{1}_4$ is the four-dimensional identity matrix and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, such that applying P_{L} and P_{R} on the bispinor gives:

$$P_{\text{L}}\varphi = \begin{pmatrix} \psi_{\text{L}} \\ 0 \end{pmatrix} = \varphi_{\text{L}}, \tag{1.5}$$

$$P_{\text{R}}\varphi = \begin{pmatrix} 0 \\ \psi_{\text{R}} \end{pmatrix} = \varphi_{\text{R}}. \tag{1.6}$$

P_{L} and P_{R} are the chirality projection operators, respectively for left-handed and right-handed chiralities. φ is therefore the sum of a left-handed bispinor φ_{L} and a right-handed bispinor φ_{R}. A particle is said left-handed or right-handed if, for some reason, only the corresponding component is involved in an interaction.

For massless particles, chirality is equivalent to helicity, which means that the directions of the spin and momentum vectors for a right-handed particle will be the same, while they will have opposite directions in the case of a left-handed particle.

1.1.2.3. Vector fields

Vector fields correspond to particles with spin equal to 1, thus to vector bosons in the SM. The Lagrangian density of a free massive vector field is:

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2\varphi_\mu\varphi^\mu, \tag{1.7}$$

where $F_{\mu\nu} = \partial_\mu\varphi_\nu - \partial_\nu\varphi_\mu$ is the field strength tensor[3], and is reduced to:

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.8}$$

if $m = 0$. However, because of particular issues that we won't detail here, the quantisation of the massless field requires the addition of a "gauge-fixing" term

[3] This is the expression for the simple case of an abelian theory, where gauge tranformations commute. Non-abelian theories require an additional term defined with the structure constants of the gauge group (the coefficients of the linear combination, expressed in a basis of Lie algebra, to which is equal the Lie bracket of two elements of the same basis).

$\mathcal{L}_{\mathrm{GF}} = -\left(1/2\xi\right)\left(\partial_\mu \varphi^\mu\right)^2$ to $\mathcal{L}_{\mathrm{Maxwell}}$, with $\xi \in \mathbb{R}^*$ (usually $\xi = 1$).

The corresponding equations of motion are the Proca equation (massive) or Maxwell's equations (massless).

1.1.3. Fundamental interactions

While there are four fundamental interactions that we know of, only three of them are included in the SM: electromagnetic, weak and strong. General relativity covers the gravitational interaction, of which there is still no satisfactory quantum description.

In the previous section, the Lagrangian densities correspond to free quantum fields, without any interaction. It happens that interaction terms arise by asking the Lagrangians to be invariant under gauge transformations.

Unlike the Poincaré group transformations which concern spacetime coordinates x, gauge transformations apply on the fields. They can, however, be a function of the position or not (local or global transformation).

If a Lagrangian density does not vary after applying a gauge transformation to the fields it contains, then it is said to have a gauge symmetry. The main interest in achieving a local gauge symmetry is the introduction of additional terms in the Lagrangian density, that couple the different fields to each other, thus providing a description of the interactions as we know them.

Mathematically, local gauge symmetries are formulated with Lie groups, which are continuous groups (in opposition to groups of discrete symmetries) of transformations with differentiable multiplication and inversion operations. A Lie group G has a certain number n of continuous parameters $(a_1, ..., a_n)$ which, together, can label each element g of the group. By performing an infinitesimal transformation (close to the identity) on one of the parameters a_j, one can identify an element I_j, called generator, such that the whole group is recoverable with the complete set of generators:

$$g(a_1, ..., a_n) = e^{\sum_{j=1}^{n} i a_j I_j}. \tag{1.9}$$

In fact, the number of generators is equal to the number of parameters of the Lie group.

The gauge group that describes a fundamental interaction being a Lie group, there are as many gauge bosons that mediate the interaction as the number of generators of the group, while the generators themselves correspond to the charges, associated to the interaction, that end up being conserved as a consequence of Noether theorem.

1.1.3.1. Quantum electrodynamics

Going back to the free Lagrangian density of fermions $\mathcal{L}_{\text{Dirac}}$ from Equation 1.3, a rotation of the bispinor fields in the form:

$$\varphi(x) \to e^{i\theta}\varphi(x), \tag{1.10}$$

where θ is a constant and the dependency on the spacetime coordinates x is made explicit, would leave $\mathcal{L}_{\text{Dirac}}$ unchanged: this is a global gauge symmetry. The corresponding Lie group is the unitary group $U(1)$, the group of complex numbers with absolute values equal to 1, which has only one generator. To achieve a local symmetry, we make θ a function of x and we write the transformation in the same fashion as in Equation 1.9:

$$\varphi(x) \to e^{ieQ\theta(x)}\varphi(x), \tag{1.11}$$

where Q is the generator of the group, in this case the electric charge operator[4], that is why we make the elementary charge e explicit in the exponent.

$\mathcal{L}_{\text{Dirac}}$ is not invariant under this local transformation, since the four-gradient also applies on $\theta(x)$ and generates an additional term in the transformed Lagrangian density. To restore the invariance, we need to expand $\mathcal{L}_{\text{Dirac}}$ with a gauge varying interaction term $\mathcal{L}_{\text{int}}$, so that the two parts that break the gauge symmetry in $\mathcal{L}_{\text{Dirac}}$ and $\mathcal{L}_{\text{int}}$ cancel each other:

$$\mathcal{L}_{\text{int}} = e\bar{\varphi}\gamma^{\mu}QA_{\mu}\varphi, \tag{1.12}$$

where A_{μ} is a vector field, or gauge field, that transforms as:

$$A_{\mu}(x) \to A_{\mu}(x) + \partial_{\mu}\theta(x). \tag{1.13}$$

The free Lagrangian density of the gauge field has to be taken into account as well: it can only be $\mathcal{L}_{\text{Maxwell}}$ since the mass term in $\mathcal{L}_{\text{Proca}}$ is not gauge invariant, hence the gauge field is massless.

The resulting Lagrangian density of quantum electrodynamics (QED), the QFT that describes the electromagnetic interaction, is:

$$\mathcal{L}_{\text{QED}} = \bar{\varphi}\left(i\gamma^{\mu}D_{\mu} - m\right)\varphi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.14}$$

with $D_{\mu} = \partial_{\mu} - ieQA_{\mu}$ being the covariant derivative. Once again, the quantisation of the gauge field would require the addition of the gauge-fixing term $\mathcal{L}_{\text{GF}}$ defined in Section 1.1.2.3.

The boson associated to the gauge field, massless, is the photon, while the quantity conserved in the electromagnetic interaction is the electric charge.

[4] If the field has an electric charge q in units of e, then $Q\varphi = q\varphi$.

1.1.3.2. Electroweak theory

The weak interaction consists in charged and neutral currents, carried by the $W^{\pm}$ and Z^0 bosons. The corresponding gauge group is the special unitary group $SU(2)$, the group of 2×2 complex unitary matrices with determinant equal to 1 [5]. The generators of the group are $T^a = \sigma^a/2$, $a = 1, 2, 3$ and σ^a are the three Pauli matrices.

Fermions and antifermions have both left-handed and right-handed components, except for neutrinos (antineutrinos) which are found to be completely left-handed (right-handed). Furthermore, while both chiralities are involved in neutral currents, only left-handed fermions and right-handed antifermions interact via charged currents. This means that the parity symmetry is violated by neutrinos and the weak interaction. Having that in mind, and given the nature of $SU(2)$ transformations, the Lagrangian density will include "mixed" fermionic fields: doublets of left-handed and singlets of right-handed fermions:

$$\ell_{\mathrm{L}} = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_{\mathrm{L}}, \quad q_{\mathrm{L}} = \begin{pmatrix} u \\ d \end{pmatrix}_{\mathrm{L}}, \quad \ell_{\mathrm{R}} = l_{\mathrm{R}}, \quad u_{\mathrm{R}}, \quad d_{\mathrm{R}}, \tag{1.15}$$

with $l = e, \mu, \tau$, while u means up-type quark and d means down-type quark. The singlets are not affected by $SU(2)$ transformations and don't interact with the gauge fields.

If we build a theory of the weak interaction in the same way as QED, we won't be able to include mass terms in the Lagrangian density, even though we know that the three weak bosons are massive. The electroweak theory, developed by Glashow, Salam and Weinberg, solves this problem by unifying the electromagnetic and weak interactions in a single QFT based on the $SU(2)\times U(1)$ gauge group.

First we start from a free and massless Dirac Lagrangian where the contributions of the chirality doublets and singlets are made explicit, then we restore the gauge symmetry with the introduction of four vector fields: W^a_μ ($SU(2)$ gauge fields) and B_μ ($U(1)$ gauge field), as well as the respective Maxwell Lagrangians. This gives us the Lagrangian density that describes the kinetics and the interactions of the fermions and gauge bosons:

$$\mathcal{L}_{\mathrm{EW}} = \bar{\ell}_{\mathrm{L}} i \not{D} \ell_{\mathrm{L}} + \bar{\ell}_{\mathrm{R}} i \not{D} \ell_{\mathrm{R}} + \bar{q}_{\mathrm{L}} i \not{D} q_{\mathrm{L}} + \bar{u}_{\mathrm{R}} i \not{D} u_{\mathrm{R}} + \bar{d}_{\mathrm{R}} i \not{D} d_{\mathrm{R}} - \frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.16}$$

where $\not{D} = \gamma^\mu D_\mu$ and:

$$D_\mu = \partial_\mu - i g T^a W^a_\mu - \frac{i}{2} g' Y B_\mu. \tag{1.17}$$

$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g \epsilon^{abc} W^b_\mu W^c_\nu$ is the field strength tensor of W^a_μ, where

[5] With this condition, we get three generators, one fewer than in the $U(2)$ group.

a, b, c run from 1 to 3, ϵ^{abc} are the structure constants of $SU(2)$ that define the commutation relations of the tensors, while $B_{\mu\nu}$ is the ($U(1)$-type) tensor of B_μ. T is the weak isospin operator, while Y is the weak hypercharge operator, the respective coupling constants (similar to e for QED) being written g and g'.

The hypercharge and the third component of isospin are related to the electric charge (Gell-Mann-Nishijima formula):

$$Q = T^3 + \frac{Y}{2}, \tag{1.18}$$

therefore weak isospin usually refers to T^3 instead of T. The absolute value of T^3 is equal to 1/2 for fermions, 1 for $W^\pm$ and 0 for Z^0, while the sign is the same as the electric charge of the particle (it is positive for neutrinos and negative for antineutrinos). Both T^3 and Y are conserved by the weak interaction.

At this point, the fields are still massless. The BEH mechanism is what solves this problem. We take a complex scalar field:

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.19}$$

the Higgs field that interacts with the electroweak gauge fields, and write its Lagrangian density, including potential terms as well:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi^\dagger)(D^\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 = |D_\mu \phi|^2 + \mu^2 |\phi|^2 - \lambda |\phi|^4, \tag{1.20}$$

with $\mu \in \mathbb{C}$ and $\lambda \in \mathbb{R}_+^*$.

In the case $\mu^2 \leq 0$, the minimum of the potential $V(|\phi|^2) = -\mu^2 |\phi|^2 + \lambda |\phi|^4$ is reached if the field is simply zero. However, if $\mu^2 > 0$, a zero field gives a local maximum, while the global minimum $V(|\phi_0|^2)$ corresponds to:

$$\phi_0 = \left(\frac{\mu^2}{2\lambda}\right)^{\frac{1}{2}} e^{i\theta} = \frac{v}{\sqrt{2}} e^{i\theta}, \tag{1.21}$$

with $0 \leq \theta < 2\pi$. If the field is non-zero, then the global $U(1)$ symmetry is spontaneously broken by choosing a value for θ.

We parametrise the Higgs field ("unitary gauge") so that it can be written:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.22}$$

$h(x)$ stands for the transverse[6] fluctuations of the field ($\langle h(x) \rangle = 0$), while v, which

[6] Orthogonal to the symmetry circle.

is strictly positive, is referred to as the vacuum expectation value (*vev*) of the Higgs field and can be computed thanks to the Fermi constant G_{F} [10]:

$$v = \left(\frac{(\hbar c)^3}{\sqrt{2} G_{\mathrm{F}}}\right)^{-1/2} \simeq 246.22 \text{ GeV}. \tag{1.23}$$

By rewriting ("mass eigenstates basis"):

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp W_\mu^2\right) \tag{1.24}$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \tag{1.25}$$

and developing $\mathcal{L}_{\text{Higgs}}$, mass terms that depend on v and the coupling constants arise for $W_\mu^\pm$ and Z_μ, which are the respective fields of the $W^\pm$ and Z^0 bosons, while the electromagnetic field A_μ remains massless. θ_w is the weak mixing angle, sometimes said Weinberg angle. The mass of the Higgs boson appears as well, since the potential term develops as a polynomial in powers of h.

Finally, an interaction between the Higgs field and fermions is necessary in order to explain fermion masses. Indeed, a Dirac mass term with left-handed doublets φ_{L} and right-handed singlet φ_{R} would be of the form: $m\left(\bar{\varphi}_{\mathrm{L}}\varphi_{\mathrm{R}} + \bar{\varphi}_{\mathrm{R}}\varphi_{\mathrm{L}}\right)$, which is not gauge invariant since the singlets don't transform under $SU(2)$. However, the addition of an $SU(2)$ scalar field preserves the gauge invariance. This is the idea behind the Yukawa interaction:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{i,j} \left[y_{ij}^\ell \bar{\ell}_{i\mathrm{L}} \phi \ell_{j\mathrm{R}} + y_{ij}^u \bar{q}_{i\mathrm{L}} \hat{\phi} u_{j\mathrm{R}} + y_{ij}^d \bar{q}_{i\mathrm{L}} \phi d_{j\mathrm{R}}\right] + \text{h.c.}, \tag{1.26}$$

where i, j cover the three lepton or quark generations, y_{ij} are the coupling constants of the interaction, $\hat{\phi}$ is the $SU(2)$ conjugate[7] of ϕ and h.c. stands for Hermitian conjugate. It is necessary to take the conjugate for up-type quarks since $\phi^+ = 0$, but we don't do so for leptons since neutrinos are not right-handed. Thus, they don't interact with the Higgs field and are assumed massless.

It is possible to formulate the part of the Yukawa Lagrangian that concerns quarks as the product of matrices, up-type and down-type quarks being combined in two triplets, while the *vev* and the coupling constants are stored in 3×3 matrices $\mathcal{M}^u$ and $\mathcal{M}^d$ defined as:

$$\mathcal{M}_{ij}^u = \frac{v}{\sqrt{2}} y_{ij}^u, \quad \mathcal{M}_{ij}^d = \frac{v}{\sqrt{2}} y_{ij}^d. \tag{1.27}$$

Using the four unitary matrices U_{L}, U_{R}, D_{L} and D_{R}, that transform the quark mass

[7] $\hat{\phi} = i\sigma_2\phi$, so that it transforms the same way as ϕ under $SU(2)$.

eigenstates into the weak interaction eigenstates, to diagonalise $\mathcal{M}^u$ and $\mathcal{M}^d$, the quark masses are obtained as the diagonal components of the resulting matrices:

$$U_{\mathrm{L}}^{-1}\mathcal{M}^u U_{\mathrm{R}} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}, \quad D_{\mathrm{L}}^{-1}\mathcal{M}^d D_{\mathrm{R}} = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}. \tag{1.28}$$

$V = U_{\mathrm{L}}^{\dagger} D_{\mathrm{L}}$, said Cabibbo–Kobayashi–Maskawa (CKM) matrix, has some importance due to its presence in the formulation of weak charged currents. Each value of the CKM matrix is related to the probability of transition from one quark flavour to another after interaction with a $W^{\pm}$ boson. Because of the existence of three quark generations instead of only two, it also introduces a complex phase that is responsible for the violation of charge and parity symmetries (CP violation) in the weak interaction[8].

1.1.3.3. Quantum chromodynamics

As it was already introduced in Section 1.1.1, the strong interaction, carried by gluons, binds quarks together in hadrons according to "rules" associated to the colour charge: this is called colour confinement, which doesn't allow quarks to be isolated. The colour charge of quarks comes in six possible states: red, blue or green ($|r\rangle$, $|b\rangle$, $|g\rangle$) for quarks, antired, antiblue or antigreen ($|\bar{r}\rangle$, $|\bar{b}\rangle$, $|\bar{g}\rangle$) for antiquarks, and is the quantity conserved by the strong interaction.

Gluons can produce quark-antiquark pairs and interact with each other too. For that reason, we can guess that these bosons also have colour charges. Since colour is conserved by the interaction, the charge of a gluon is a colour-anticolour pair, or a linear combination of such paired states depending on the chosen basis. While nine different combinations are possible, only eight of them are independent gluon colour states, therefore there are eight types of gluons. The ninth state, a colour singlet written $(|r\bar{r}\rangle + |g\bar{g}\rangle + |b\bar{b}\rangle)/\sqrt{3}$, is the colour state of hadrons and the only one that makes them free particles with respect to the strong interaction[9]. This is actually what is meant by saying that hadrons are white or colourless.

Quantum chromodynamics (QCD), the QFT of the strong interaction, is based on the $SU(3)$ gauge group which has eight generators $t^a = \lambda^a/2$, where λ^a $(a = 1, ..., 8)$ are the Gell-Mann matrices. The Lagrangian density of QCD, very similar to QED, is:

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\varphi}_{q\alpha} \left(i\gamma^\mu D_{\mu,\alpha\beta} - m_q \delta_{\alpha\beta} \right) \varphi_{q\beta} - \frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu}, \tag{1.29}$$

[8] Time reversal operator is anti-linear, which means that V becomes its conjugate under the action of this operator. Since $V \neq V^*$, T-symmetry is violated, therefore CP as well in accordance with the CPT theorem.

[9] As we know, hadrons like neutrons and protons can still be found in bound states through the nuclear force, which results from the strong interaction.

with $D_{\mu,\alpha\beta} = \partial_\mu \delta_{\alpha\beta} - i g_s t^a G^a_\mu$. Here we introduce the gluon fields G^a_μ and the respective tensors $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu$, where f^{abc} are the structure constants of $SU(3)$. a, b and c cover the eight gluon colour states, α and β the three quark colours and q the quark flavours. $\delta_{\alpha\beta}$ is the Kronecker delta and g_s (sometimes replaced by $\alpha_s = g_s^2/4\pi$) is the strong coupling constant.

1.1.4. The Standard Model as Theory and Representation

The SM is the combination of both electroweak theory and QCD: the product group $SU(3)_\mathrm{c} \times SU(2)_\mathrm{L} \times U(1)_\mathrm{Y}$ summarises the gauge symmetries of the model. The indices are often added to make explicit the colour, conserved charge of QCD, the P-violation in weak currents that involve only left-handed fermions, and finally the hypercharge.

The Lagrangian density of the SM is then:

$$\mathcal{L}_\mathrm{SM} = \mathcal{L}_\mathrm{EW} + \mathcal{L}_\mathrm{Higgs} + \mathcal{L}_\mathrm{Yukawa} + \mathcal{L}_\mathrm{QCD}, \tag{1.30}$$

if we remove from $\mathcal{L}_\mathrm{QCD}$ all the redundant terms (kinetics and mass of fermions) already present in $\mathcal{L}_\mathrm{EW}$ and $\mathcal{L}_\mathrm{Yukawa}$.

As one can guess, Lagrangians are of great use when it comes to modeling the evolution of quantum states that undergo interactions. Indeed, the interaction terms are exploited in the calculation of the scattering matrix (or S-matrix) that relates the states of free particles between times $t = -\infty$ and $t = +\infty$ (asymptotically free). The evolution of a state $|\psi(t)\rangle$ from an initial time t_i to a final time t_f is expressed thanks to the time-evolution operator U:

$$|\psi(t_f)\rangle = U(t_f, t_i)|\psi(t_i)\rangle, \tag{1.31}$$

and U is defined with the Hamiltonian of the interaction H_{int}:

$$U(t_f, t_i) = \mathcal{T}\left(e^{-i\int_{t_i}^{t_f} H_{int}\, dt}\right), \tag{1.32}$$

where $\mathcal{T}$ stands for the time-ordered product[10]. If there are no derivatives of the fields in the Hamiltonian, then $H_{int} = -\int \mathcal{L}_{int}\, dx^3$ and the S-matrix is simply:

$$S = \lim_{\substack{t_i \to -\infty \\ t_f \to +\infty}} U(t_f, t_i) = \mathcal{T}\left(e^{-i\int_{-\infty}^{+\infty} \mathcal{L}_{int}\, dx^4}\right). \tag{1.33}$$

Properties of physical processes like cross sections or decay widths are related to S and are computable starting from it. Furthermore, Feynman diagrams, a convenient

[10] Because $U(t_f, t_i) = 1 - \int_{t_i}^{t_f} H_{int}(t_k)\, U(t_k, t_i)\, dt_k$, $t_i < t_k < t_f$, the evolution operator ends up being developed as a Dyson series where the Hamiltonians have to be ordered from past (right) to future (left).

way of drawing interactions between particles, are essentially visual representations of the S-matrix if it is expanded as the sum of different terms, each formulating mathematically a specific interaction (in number and nature of particles involved).

In Feynman diagrams, the course of time is often depicted from left to right, although rotating a diagram gives equivalent physical processes. Fermions (antifermions) are drawn as straight lines with an arrow that points in the (opposite) direction of time, weak bosons and the photon as wavy lines, gluons as looped lines and the Higgs boson as a dashed line. A point where some lines converge is called a vertex and represents an interaction.

Each line can be translated as a part of the S-matrix term to which the diagram corresponds. A line that connects two vertices is called internal and a propagator, that makes the transition from initial to final states, is attached to it. The external lines, the initial state particles that interact and the resulting final state particles, are taken into account as fields, linked with couplings that are assigned to the vertices.

Depending on whether the line is external or internal, the respective particle is said real or virtual. Real particles have defined[11] masses and can be observed. On the other hand, virtual ones exist for a time so short that their energy cannot be precisely defined[12], as stated by Heisenberg's uncertainty principle. A virtual particle that connects directly the initial and final states of a process doesn't violate the conservation of energy and momentum of the system, but its mass, computed from the energy-momentum relation, doesn't have to be the same as its real counterpart. For example, the Feynman diagram in Figure 1.2 (left) illustrates the process $e^+e^- \to \tau^+\tau^-$, where a photon can play the role of a virtual mediator (internal line), of strictly positive mass.

In addition to that, a single real or virtual particle itself can emit a pair of virtual particles that are reabsorbed shortly after, as a quantum fluctuation. Such phenomena are described for example by loop or penguin diagrams (see for example Figure 1.2, diagram on the right) and don't imply the conservation of energy or momentum by each virtual particle. For this reason, it is necessary to integrate these quantities over all possible values when computing the associated probability amplitudes, resulting in unrealistic divergences.

However, the energies can only get larger if the time between emission and reabsorption becomes shorter, as the uncertainty principle justifies once again. This means that, passed a certain scale of energy, one can consider that a diagram with a loop is hardly distinguishable from the same diagram without the loop, from a physics point of view, and therefore cut off this contribution from the calculation, which is the idea behind the renormalisation of quantum field theories.

[11] Well predicted.

[12] The standard deviation of the energy increases, which means that its value cannot be predicted with as good precision as in the case of real particles.

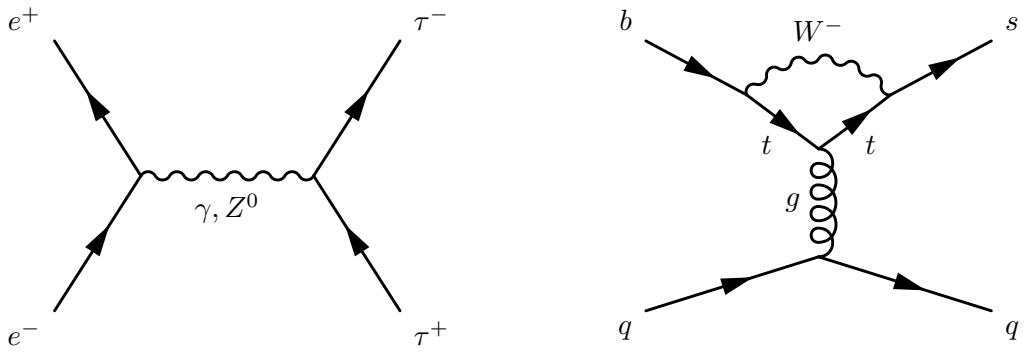


Figure 1.2: Feynman diagrams of the $e^+e^- \to \tau^+\tau^-$ process (left) and of a $b \to s$ transition via W boson loop, also said penguin diagram (right).

1.2. Lepton flavour violation, a path to new physics

1.2.1. Beyond the Standard Model

While the SM has been successfully tested for many decades, some theoretical problems arise that have not yet been elucidated and many experimental observations still need an explanation.

Although both were unpredicted and are even now undetected, the evidence of dark energy and dark matter accounting for most of the energy of the universe, from the observation of its accelerated expansion and the odd shape of galaxy rotation curves, might be the most famous example of a shortcoming of the SM. The fact that antimatter has vanished from the universe since its early moments, something that CP violation in the weak interaction cannot fully explain, or the failures to propose a satisfactory QFT of gravity, are further indications that the SM's potential to describe matter and interactions at cosmological scales is quite limited.

From the theoretical side, the values of some parameters of the SM are intriguing. Although the QCD Lagrangian density could involve a complex phase similar to the one of the CKM matrix, the fact that the strong interaction doesn't violate CP symmetry, as we know from experiments, suggests that the corresponding angle (theta vacuum angle) is very close to zero. There is no apparent reason for this angle to be so small, but it seems unlikely for such a fine-tuning to be present without a cause: this is called the strong CP problem. Another example of fine-tuning, referred to as the hierarchy problem, concerns the mass of the Higgs boson which, because of the large quantum corrections that arise from virtual particles, should be much larger than its current measured value unless those corrections are by chance very precisely (and therefore unlikely) canceled.

The flaws of our model push physicists to seek proofs of some "new physics"

(NP), beyond the SM (BSM), that would include new particles and interactions able to solve the aforementioned problems and many others. This led for example to the development of supersymmetry and string theories, physics beyond the SM being for a long time already one of the main area of research in particle physics.

1.2.2. Neutrino oscillations

In addition to the experimental observations cited before, the discovery of neutrino oscillations, the fact that one neutrino's flavour can change as it propagates, has also put the SM in a delicate position.

First hints of neutrino oscillations [21] emerged between the 1960s and 1990s, with the observation of solar[13] and atmospheric neutrino[14] deficits in experiments designed to detect these specific type of neutrinos, while the first evidence came in 1998 with the Super-Kamiokande experiment [22].

This phenomenon occurs if the neutrino flavour (ν_e, ν_μ, ν_τ) and mass (ν_1, ν_2, ν_3) eigenstates are mixed, in such a way that we can write:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \tag{1.34}$$

where U is the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix. Mathematically very similar to the CKM matrix, the PMNS matrix also introduces a CP violating phase in the mixing of neutrinos.

The implications of neutrino oscillations are that, not only lepton flavour violation (LFV) is induced, but at least two out of three neutrinos must have non-zero masses as well. Indeed, the probability for a neutrino of one flavour to be observed with a different flavour, after a certain amount of time, is non-zero only if both flavours have different masses.

As we have seen in the case of the Yukawa interaction through Equation 1.26, the SM doesn't include neutrino masses because right-handed neutrinos have never been observed. By postulating their existence, one can extend the SM in order to introduce an interaction with the Higgs field and neutrino mass terms as a consequence. These masses, however, seem to be so small that they still haven't been measured, which might suggest that evidence of a more sophisticated process[15] must be provided to explain the discrepancy between neutral and charged lepton masses.

[13] Electron neutrinos produced in proton-proton reactions.
[14] Electron or muon neutrinos produced in pion decays.
[15] For example, the "seesaw" mechanism; see Section 1.2.5.2.

1.2.3. Charged lepton flavour violation

Neutrino flavour violation caused by oscillations translates also as charged lepton flavour violation (cLFV) via charged weak currents. Without considering neutrino oscillations, weak interaction always compensates the decay or the creation of a charged lepton with a neutrino or antineutrino, thus preserving lepton flavour. The particular behaviour of neutrinos, however, makes possible transitions from one charged lepton flavour to another without even having any neutrinos in the initial and final states.

A simple example, illustrated in Figure 1.3, would be the transition $\mu \to e\gamma$, allowed if a muon decays into a muon neutrino via charged weak current, while the neutrino oscillates into an electron neutrino before reabsorbing the W boson, resulting in the production of an electron[16]. These processes may now be considered as part of the SM, extended so that it includes neutrino oscillations: in that case, it appears that the probability for this transition to occur, also said branching fraction or ratio and noted $\mathcal{B}$, is extremely small ($< 10^{-54}$ [23]), therefore one can safely assume that no current or future experiment would be able to observe it. This is in fact the case for many processes that involve the emission and reabsorption of a charged weak boson.

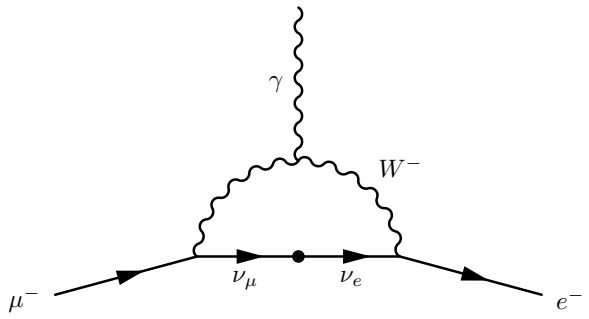


Figure 1.3: Feynman diagram of the decay $\mu^- \to e^-\gamma$ via neutrino oscillations.

This point leads us to the following epiphany: *observing cLFV decays in particle physics experiments would only be attributable to some unknown processes, hence constitute evidence of new physics.*

Theorists have been developing several NP models that, if they turn out to be accurate, would allow higher branching ratios for a certain number of LFV decays, at levels to which current experiments could be sensitive. A large amount of tau LFV decay channels are of interest in the context of BSM physics: among them is the $\tau \to \ell\phi$ decay, where l stands either for an electron or a muon, and $\phi = s\bar{s}$ is a

[16] A photon or any other possible particle must of course be emitted by the weak boson in order to conserve both energy and momentum.

neutral vector meson of mass $m_\phi = 1019.461 \pm 0.016$ MeV/c^2 [10], produced after emission of a Z^0 boson or a photon by one of the virtual particles. Vector mesons are spin 1 mesons, of which the fields transform as their opposites under the action of parity, therefore said to have odd parity (a spin 1 meson with even parity would be called pseudovector). A Feynman diagram of decay in the context of neutrino oscillations is shown in Figure 1.4.´

The purpose of this thesis is to study this decay channel as a potential probe of NP; to that end, let us first of all introduce the BSM theories that are relevant for tau LFV in general, and $\tau \to \ell\phi$ in particular.

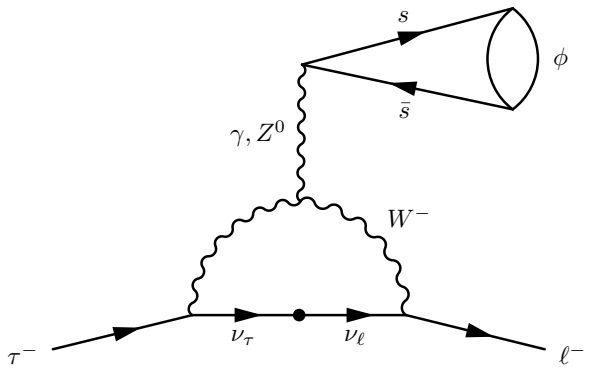


Figure 1.4: Feynman diagram of the decay $\tau^- \to \ell^-\phi$ ($\ell = e, \mu$) via neutrino oscillation.

1.2.4. Effective field theories

As it was discussed in regards to renormalisation, it is sometimes possible to work with a QFT by considering only a certain range of energy, where the dominant contributions to the studied physics are included and the remaining ones are negligible. This approach follows the idea that some very well known theories are in fact only good approximations of even more general theories, in the same way as Galilean relativity and Newton's gravitational laws were approximations of special relativity and general relativity, at "low" velocities and "small" masses. To define "low" and "small", one needs to set a scale below which the approximation becomes acceptable. When it comes to QFT, this corresponds to a maximum energy scale, usually referred to as Λ, and the resulting approximation is called an effective field theory (EFT).

This idea is applicable to the SM, in the sense that potential processes involving NP particles could very well be observable only at high energies, yet to be reached. Thus, the Lagrangian density of an EFT designed for the SM can be written as a

function of a chosen energy scale Λ, such that:

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda}\sum_i \mathcal{C}_i^{(5)}\mathcal{O}_i^{(5)} + \frac{1}{\Lambda^2}\sum_i \mathcal{C}_i^{(6)}\mathcal{O}_i^{(6)} + \frac{1}{\Lambda^3}\sum_i \mathcal{C}_i^{(7)}\mathcal{O}_i^{(7)} + ... \quad (1.35)$$

The effective Lagrangian is split into a summation of terms, each formulated depending on a dimension d as a linear combination of operators $\mathcal{O}_i^{(d)}$, with individual dimensionless factors $\mathcal{C}_i^{(d)}$ called Wilson coefficients and a global factor $1/\Lambda^{d-4}$. Expressing $\mathcal{O}_i^{(d)}$ in units of energy, to a power d, gives us the dimension. Partial derivative operators and bosonic fields are of dimension 1, while fermionic fields are of dimension 3/2. Then, it appears that the highest dimension present in $\mathcal{L}_{\text{SM}}$ is 4, that is why the BSM contributions start at dimension $d = 4 + 1$ and are suppressed by powers $d - 4$ of Λ [24].

For each dimension, the operators correspond to specific interactions. For example, in the case of τ LFV, a general low-scale[17] effective Lagrangian can be written as [25]:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{eff}}^{(D)} + \mathcal{L}_{\text{eff}}^{(lq)} + \mathcal{L}_{\text{eff}}^{(4l)} + \mathcal{L}_{\text{eff}}^{(G)} + ..., \quad (1.36)$$

where $\mathcal{L}_{\text{eff}}^{(D)}$ has dimension-5 effective dipole operators and stands for $\tau - \mu$ or $\tau - e$ transitions with a real photon emitted by the W boson, $\mathcal{L}_{\text{eff}}^{(lq)}$ and $\mathcal{L}_{\text{eff}}^{(4l)}$ are of dimension 6 and involve two additional quarks or leptons in the final state, while $\mathcal{L}_{\text{eff}}^{(G)}$ represents gluonic interactions with dimension-7 operators. In our case, only $\mathcal{L}_{\text{eff}}^{(lq)}$ is of interest, as it should contain two strange quark fields to model the $\tau \to \ell\phi$ LFV decay.

In practice, the benefit of effective field theory is that, by formulating the Lagrangian density of an interaction as an effective Lagrangian, one can identify Wilson coefficients that are relevant to the processes of interest. The coefficients, which are functions of the interaction's parameters, can be determined by measuring (or putting limits on) branching fractions, thus they allow to measure (or put constraints on) these parameters in order to test the accuracy of BSM theories.

1.2.5. The new physics of $\tau \to \ell\phi$

We will detail here two new physics models that could affect largely enough the $\tau \to \ell\phi$ decays that they would become accessible to current experimental sensitivities. We refer the interested reader to the documentation on other subjects, for example the littlest Higgs model with T-parity [26], $SO(10)$ GUT model and additional heavy Dirac neutrinos [27], unparticles [28] and Minimal Supersymmetric Standard Model [29].

[17] For masses around 2 GeV/c^2, far below the scale Λ taken at the electroweak symmetry breaking energy (*vev*) $v \simeq 246$ GeV or above.

1.2.5.1. Scalar and vector leptoquarks hypothesis

In the past years, experiments like BaBar, Belle and LHCb have performed tests of lepton flavour universality (LFU), which assumes that the couplings of electroweak gauge bosons to leptons doesn't depend on the flavour. To that end, they have considered semileptonic B meson decays, taking the ratio of their branching fractions and comparing it to SM predictions. Two observables $R_{K^{(*)}}$ and $R_{D^{(*)}}$ are defined as:

$$R_{K^{(*)}} = \frac{\mathcal{B}(B \to K^{(*)}\mu^+\mu^-)}{\mathcal{B}(B \to K^{(*)}e^+e^-)}, \quad R_{D^{(*)}} = \frac{\mathcal{B}(B \to D^{(*)}\tau\bar{\nu_\tau})}{\mathcal{B}(B \to D^{(*)}l\bar{\nu_l})}, \tag{1.37}$$

where $l = e, \mu$. Measurements of both observables show slight disagreements with the SM; in terms of significance, the deviations vary between 1.4 and 3.1σ, providing possible hints of LFU violations [4, 5, 6].

The leptoquark hypothesis is a fitting explanation for these anomalies. Leptoquarks (LQs), if they exist, would carry both baryon B and lepton L quantum numbers, joined together in a fermion number $F = 3B + L$, and introduce still undiscovered interactions between both types of fermions, converting quarks into leptons and vice versa. We show in Figure 1.5 possible Feynman diagrams for $b \to s\ell^+\ell^-$ transitions and $\tau \to \ell\phi$ decays via leptoquark interactions.

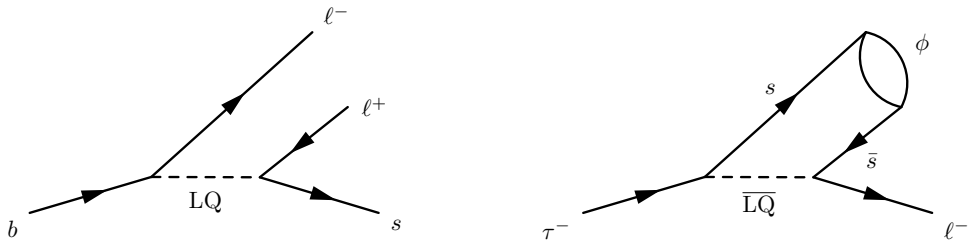


Figure 1.5: Feynman diagrams of a $b \to s\ell^+\ell^-$ ($\ell = e, \mu$) transition via a leptoquark LQ of electric charge $+2/3$ (left) and of the decay $\tau^- \to \ell^-\phi$ via a leptoquark of opposite electric charge $-2/3$ (right).

There are six scalar and six vector LQ variations. Depending on whether $|F| = 2$ or $F = 0$, scalar LQs are noted S or R and vector LQs are noted V or U. They are often written as multiplets that make explicit their properties under the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$: for example, the scalar LQ $R_2 = (\mathbf{3}, \mathbf{2}, 7/6)$ is a triplet under $SU(3)$, a doublet under $SU(2)$ and has a hypercharge $Y = 7/6$. The index differentiates LQs according to their $SU(2)$ representation, while they are given subscripts (tilde or bar) if they differ only by the hypercharge [30].

When it comes to B-anomalies, it appears that no single scalar LQ can accommodate both $R_{K^{(*)}}$ and $R_{D^{(*)}}$ deviations from the SM. It is however possible to do it by introducing interactions between SM fermions, R_2 and $S_3 = (\bar{\mathbf{3}}, \mathbf{3}, 1/3)$. The

interaction terms in the Lagrangian density can be written [30, 31]:

$$\mathcal{L} \supset Y_{\mathrm{R}}^{ij} \bar{q}_{\mathrm{L}i} \ell_{\mathrm{R}j} R_2 + Y_{\mathrm{L}}^{ij} \bar{u}_{\mathrm{R}i} \widetilde{R_2}^{\dagger} \ell_{\mathrm{L}j} + Y^{ij} \bar{q}_{\mathrm{L}i}^{C} i\sigma_2 (\sigma_k S_3^k) \ell_{\mathrm{L}j} + \mathrm{h.c.}, \tag{1.38}$$

where Y_{L}, Y_{R} and Y are Yukawa matrices, i and j specify the quark or lepton generations, $q_{\mathrm{L}}^{C} = i\gamma^2\gamma^0 \bar{q}_{\mathrm{L}}^{T}$, $\widetilde{R_2} = i\sigma_2 R_2^*$ and S_3^k $(k = 1, 2, 3)$ are expressed in $SU(2)$ representation. By writing the Lagrangian density in mass eigenstate basis, the unitary matrices U_{L}, U_{R}, D_{L} and D_{R}, introduced in Section 1.1.3.2 for the CKM matrix, come into play in such a way that the LQ $c - \mu$ and $c - \tau$ transition couplings are defined with parameters $y_{\mathrm{L}}^{c\mu}$ and $y_{\mathrm{L}}^{c\tau}$:

$$U_{\mathrm{R}} Y_{\mathrm{L}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & y_{\mathrm{L}}^{c\mu} & y_{\mathrm{L}}^{c\tau} \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.39}$$

Furthermore, writing $U_{\mathrm{R}}^{22} = \cos\theta$, constraints on the NP parameters can be put from experimental limits on $\tau \to \mu\phi$, thanks to the relation:

$$\mathcal{B}(\tau \to \mu\phi) \sim (y_{\mathrm{L}}^{c\mu} y_{\mathrm{L}}^{c\tau})^2 \left(\frac{\cos\theta}{m_{S_3}} \right)^4, \tag{1.40}$$

derived from an EFT approach of the LQ interactions.

Unlike scalar LQs, there is one vector LQ, $U_1 = (\mathbf{3}, \mathbf{1}, 2/3)$, that can explain both B-anomalies. The interaction Lagrangian density is written with currents J_μ^U [1]:

$$\mathcal{L} \supset \frac{g_U}{\sqrt{2}} (U_1^\mu J_\mu^U + \mathrm{h.c.}), \tag{1.41}$$

$$J_\mu^U = \beta_{\mathrm{L}}^{ij} (\bar{q}_{\mathrm{L}i} \gamma_\mu \ell_{\mathrm{L}j}) + \beta_{\mathrm{R}}^{ij} (\bar{d}_{\mathrm{R}i} \gamma_\mu \ell_{\mathrm{R}j}), \tag{1.42}$$

where g_U is the U_1 coupling, while β_{L} and β_{R} are matrices that contain couplings of specific transitions. In this case, it is assumed (as an approximation) that:

$$\beta_{\mathrm{L}} = \begin{pmatrix} 0 & 0 & \beta_{\mathrm{L}}^{d\tau} \\ 0 & \beta_{\mathrm{L}}^{s\mu} & \beta_{\mathrm{L}}^{s\tau} \\ 0 & \beta_{\mathrm{L}}^{b\mu} & 1 \end{pmatrix}, \quad \beta_{\mathrm{R}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{\mathrm{R}}^{b\tau} \end{pmatrix}. \tag{1.43}$$

From the EFT approach, the following relation is derived with the U_1 mass m_U:

$$\mathcal{B}(\tau \to \mu\phi) \propto \left| \left(\frac{g_U v}{2 m_U} \right)^2 \beta_{\mathrm{L}}^{s\tau} (\beta_{\mathrm{L}}^{s\mu})^* \right|^2. \tag{1.44}$$

The computation of this branching fraction depends on assumptions that are

made on $\beta_{\mathrm{R}}^{b\tau}$, with two possible scenarios: $\beta_{\mathrm{R}}^{b\tau} = 0$ or $\beta_{\mathrm{R}}^{b\tau} = -1$, both displayed in Figure 1.6 which shows the correlations between $\mathcal{B}(\tau \to \mu\phi)$ and $\mathcal{B}(B_s \to \tau^-\mu^+)$ in this model. The expected upper limit from the Belle II experiment on $\mathcal{B}(\tau \to \mu\phi)$, at the targeted integrated luminosity of 50 ab^{-1}, should be able to put further constraints on the U_1 model in the scenario $\beta_{\mathrm{R}}^{b\tau} = 0$, if it doesn't lead to a discovery. In two other recent studies [2, 3], theoreticians have made predictions on $\mathcal{B}(\tau \to \mu\phi)$ as a function of m_U, comparing them to the future sensitivity of Belle II. The results, shown in Figures 1.7 and 1.8, provide further evidence that $\tau \to \mu\phi$ is one of the most relevant τ LFV decay modes to analyse in order to test the leptoquark hypothesis as an explanation of the B-anomalies.

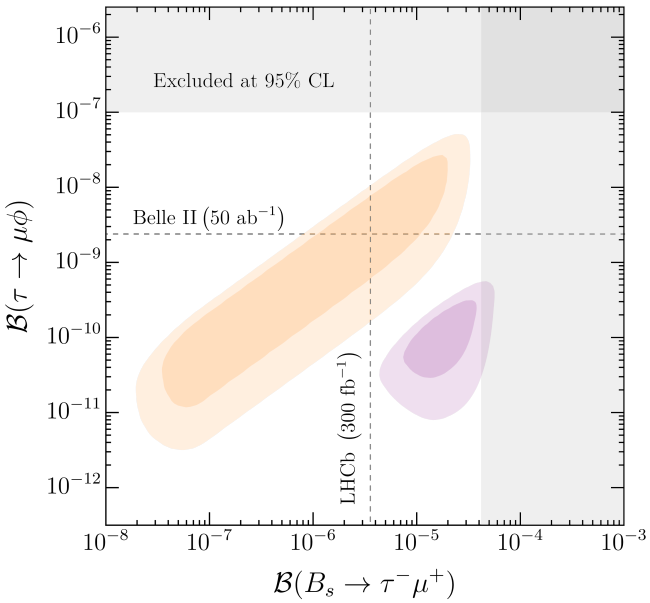


Figure 1.6: Two-dimensional branching fraction regions for $\tau \to \mu\phi$ and $B_s \to \tau^-\mu^+$ processes, with 1σ (dark colour) and 2σ (light colour) uncertainties, in the hypothesis of a U_1 vector leptoquark. Two scenarios are shown, in orange or purple, for $\beta_{\mathrm{R}}^{b\tau} = 0$ or $\beta_{\mathrm{R}}^{b\tau} = -1$ (see Equation 1.43) respectively. The grey bands correspond to 95% confidence level (C.L.) upper limits from past experimental results. Projected limits from Belle II and LHCb, depending on the integrated luminosity, are indicated with dashed lines [1].

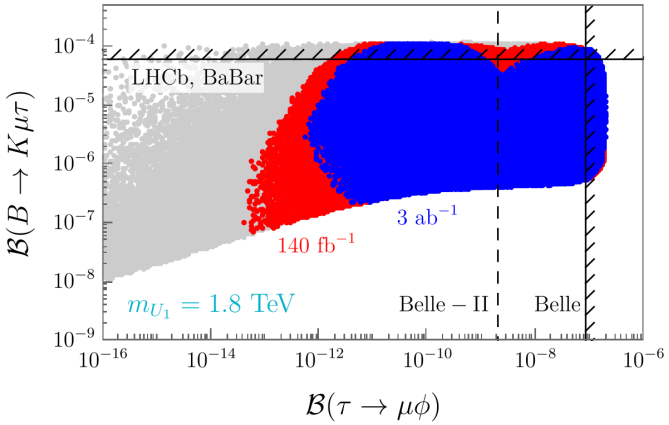


Figure 1.7: Two-dimensional branching fraction regions for $B \to K\mu\tau$ and $\tau \to \mu\phi$ processes, in the hypothesis of a U_1 vector leptoquark of mass $m_U = 1.8$ TeV/c^2. The grey, red and blue points correspond to different constraints, namely from low-energy observables, from LHC searches at 140 fb^{-1} of integrated luminosity, and from projections for when it will reach 3 ab^{-1}. Upper limits from the BaBar, Belle and LHCb experiments are indicated with hatched lines, and a dashed line for the future sensitivity expected at Belle II [2].

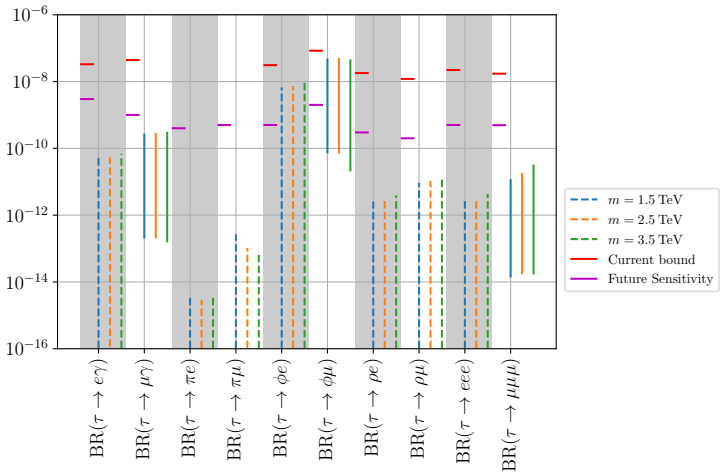


Figure 1.8: Current and expected upper limits (red and purple horizontal lines) on the branching fractions of lepton flavour violating τ decay modes studied at the Belle II experiment, and predicted ranges at 90% C.L. in the U_1 leptoquark hypothesis for input masses $m_U = 1.5$, 2.5, 3.5 TeV/c^2 (blue, yellow and green solid or dashed lines) [3].

1.2.5.2. Type-III seesaw mechanism

A possible solution to the problem of neutrino masses is the addition to the SM of heavy, self-conjugate $SU(2)_{\mathrm{L}}$ triplet fermions of zero hypercharge, that could have Yukawa interactions with the leptons we already know. Such a model is called type-III seesaw mechanism, with relations to $\tau \to \ell V^0$ decays that have been studied in 2010 [32].

A minimal extension of the SM is considered, where one triplet and one singlet fermions are integrated to the model (type I+III seesaw) [33]. From the Lagrangian that contains the new fermions, the elements of the neutrino mass matrix are derived after spontaneous symmetry breaking as

$$(m_\nu)^{ij} = -\frac{v}{2}\left(\frac{y_T^i y_T^j}{m_T} + \frac{y_S^i y_S^j}{m_S}\right), \tag{1.45}$$

where m_T and y_T (m_S and y_S) are respectively the mass and the Yukawa coupling of the fermion triplet (singlet), with $i, j = 1, 2, 3$. The couplings depend on the neutrino mass hierarchy: taking either $(a, b) = (2, 3)$ in the case of the normal hierarchy ($m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$) or $(a, b) = (1, 2)$ for the inverted order ($m_{\nu_3} < m_{\nu_1} < m_{\nu_2}$), we obtain

$$y_T^i = -i\frac{\sqrt{2m_T}}{v}\left(\sqrt{m_{\nu_a}}\cos z\, U_{ia}^* + \sqrt{m_{\nu_b}}\sin z\, U_{ib}^*\right), \tag{1.46}$$

$$y_S^i = -i\frac{\sqrt{2m_S}}{v}\left(-\sqrt{m_{\nu_a}}\sin z\, U_{ia}^* + \sqrt{m_{\nu_b}}\cos z\, U_{ib}^*\right). \tag{1.47}$$

U is the PMNS matrix and z a complex constant. The dependence of the couplings on z has been evaluated by taking into account mass constraints from neutrino experiments, with some examples shown in Figure 1.9 for the fermion triplet and looking at the imaginary part of z. The couplings remain of the order of 1 or below, thus no fine-tuning is required in order to explain the small neutrino masses.

Coming back to the question of LFV decays, the triplet fermion field is first chosen as

$$T = \begin{pmatrix} T^0/\sqrt{2} & T^- \\ T^+ & T^0/\sqrt{2} \end{pmatrix}. \tag{1.48}$$

The kinetic term of the Lagrangian is dominated at tree level by Z^0-mediated interactions, while the fermion singlet is not involved anymore since it doesn't couple to the gauge boson. The Lagrangian becomes

$$\mathcal{L}_Z = \frac{g}{2\cos\theta_w}\bar{\ell}\gamma^\mu\left(\cos 2\theta_w X_{\mathrm{L}} P_{\mathrm{L}} - 2\sin^2\theta_w X_{\mathrm{R}} P_{\mathrm{R}}\right)\ell Z_\mu, \tag{1.49}$$

where $\ell = (e, \mu, \tau, T^c)$, T^c being the charge conjugate of T^+, and $X_{\mathrm{L/R}}$ are two 4×4 matrices. By introducing $V_{\mathrm{L/R}}$, two unitary transformation matrices from the weak

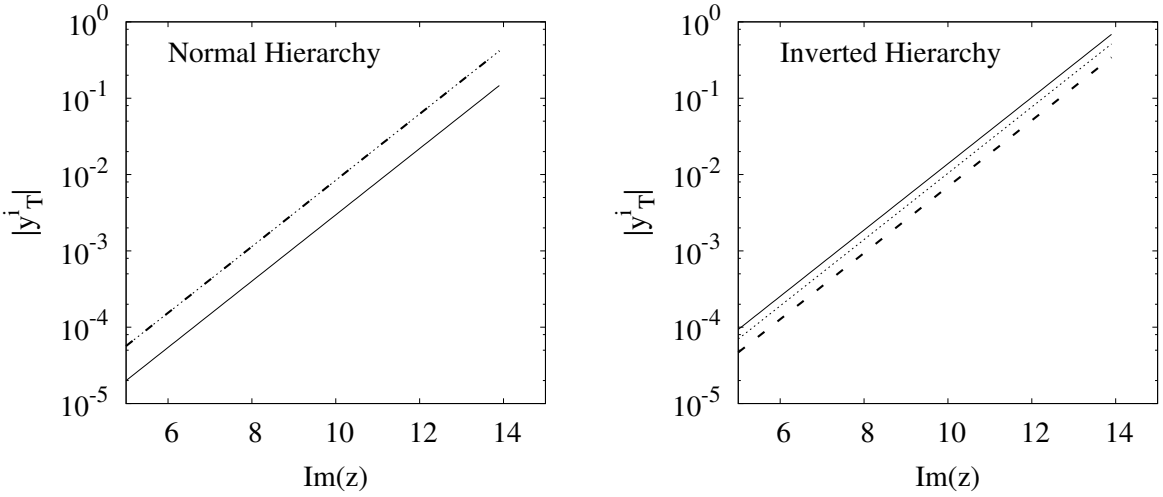


Figure 1.9: Absolute value of the Yukawa couplings y_T^i as a function of the imaginary part of z, in the case of the normal (left) and inverted (right) hierarchies, taking $m_T = 1$ TeV as well as a null Majorana phase. The solid, dashed and dotted lines stand for $i = 1, 2, 3$, respectively. Taken from [32].

to physical eigenstates of ℓ, one can write

$$Z_{(\mathrm{L/R})ij} = \chi_{\mathrm{L/R}} V_{(\mathrm{L/R})i4} V^*_{(\mathrm{L/R})j4}, \tag{1.50}$$

with $\chi_{\mathrm{L/R}} = X_{(\mathrm{L/R})44} - 1$. The unitary matrices are close to $\mathbb{1}_4$ and therefore expanded as $V_{\mathrm{L/R}} \approx \mathbb{1}_4 + \Delta_{\mathrm{L/R}}$.

The branching fraction of the $Z^0 \to \ell_i \bar{\ell}_i$ decay is now formulated as the sum of SM and NP contributions, with the help of the Fermi constant G_F and the Z^0 decay rate Γ_{Z^0}:

$$\mathcal{B}(Z^0 \to \ell_i \bar{\ell}_i) = \frac{G_\mathrm{F}^2 m_{Z^0}^3}{3\sqrt{2}\pi\Gamma_{Z^0}} \left[\underbrace{\left(\frac{\cos 2\theta_w}{2}\right)^2 + \sin^4\theta_w}_{\mathrm{SM}} + \underbrace{\frac{1}{2}\cos^2 2\theta_w \chi_\mathrm{L} |\Delta_{\mathrm{L}i4}|^2}_{\mathrm{NP}} \right]. \tag{1.51}$$

Furthermore, the branching fractions of the $\tau \to \ell_i \phi$ processes are computed according to

$$\mathcal{B}(\tau \to \ell_i \phi) = \frac{G_\mathrm{F}^2}{16\pi\Gamma_\tau} f_\phi^2 m_\tau^3 \cos^2 2\theta_w Y_\phi^2 |Z_{\mathrm{L}i3}|^2 \left(1 - \frac{m_\phi^2}{m_\tau^2}\right)^2 \left(1 - 2\frac{m_\phi^2}{m_\tau^2}\right), \tag{1.52}$$

where $f_\phi = 0.237$ is the ϕ meson's decay constant and

$$Y_\phi = -\frac{1}{2} + \frac{2}{3}\sin^2\theta_w. \tag{1.53}$$

Considering a single standard deviation error (1σ) on $\mathcal{B}(Z^0 \to \ell_i\bar{\ell}_i)$, the following new physics estimations are obtained:

$$\mathcal{B}_{\mathrm{NP}}(\tau \to e\phi) = 1.1{\times}10^{-8}, \quad \mathcal{B}_{\mathrm{NP}}(\tau \to \mu\phi) = 1.7{\times}10^{-8}. \tag{1.54}$$

1.2.6. Status of experimental searches for tau LFV decays

A few experiments have tried to detect signals of cLFV decays in the past, most of the contributions in the field of tau LFV coming from the CLEO, BaBar and Belle experiments, each improving the upper limits set in the previous searches. Up to now, however, none has been able to witness those long-awaited processes among the millions of events produced in their colliders; that is why the pursuit must go on.

Figure 1.10 shows the current and expected upper limits on various tau LFV decays. So far, the most constraining results have been established by Belle, with less than 1 ab^{-1} of integrated luminosity, and lie within the range of 10^{-8}-10^{-7}. Regarding the $\tau \to \ell\phi$ decay modes, they are set at 90% C.L. as [8]:

$$\mathcal{B}(\tau \to e\phi) < 3.1{\times}10^{-8}, \tag{1.55}$$
$$\mathcal{B}(\tau \to \mu\phi) < 8.4{\times}10^{-8}. \tag{1.56}$$

At 5 ab^{-1}, Belle II is expected to improve the limits so that they pass below 10^{-8} for most of the decay modes, and should be further lowered by one order of magnitude at the targeted luminosity of 50 ab^{-1}. In that case, the limits on $\tau \to e\phi$ and $\tau \to \mu\phi$ might reach the 10^{-9} threshold, giving hope that many of the aforementioned NP theories will be actively tested in the near future.

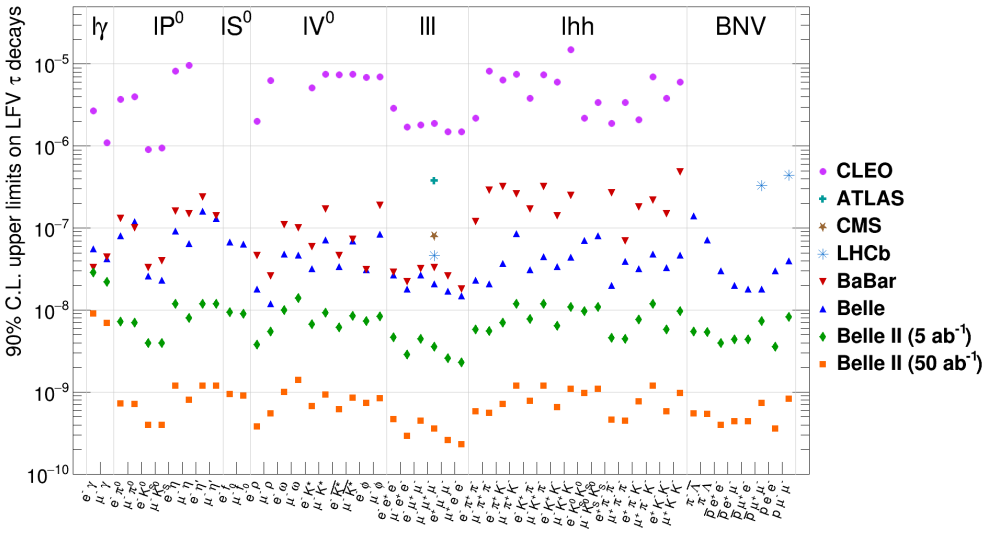


Figure 1.10: Upper limits on branching fractions at 90% C.L. for $\tau \to \ell\gamma$ ($\ell = e, \mu$), $\tau \to \ell P^0/S^0/V^0$ (neutral pseudoscalar, scalar or vector mesons), $\tau \to \ell\ell\ell$, $\tau \to \ell hh$ (two hadrons) and miscellaneous lepton flavour, lepton number and baryon number violating (BNV) decays. The green and orange marks correspond to expected upper limits at Belle II for integrated luminosities of 5 and 50 ab^{-1}, the other to current limits from CLEO, ATLAS, CMS, LHCb, BaBar and Belle. Taken from [9].

2. The Belle II experiment

Belle II [34, 35, 36] is a particle physics experiment operating at the SuperKEKB electron-positron collider, settled in the KEK laboratory (Tsukuba, Japan). SuperKEKB, as an upgrade of the previous KEKB collider, makes Belle II the successor of the Belle experiment, which ran from 1999 to 2010.

2.1. The SuperKEKB electron-positron collider

Before Belle II, the BaBar experiment (1999-2008), with the e^+e^- collider PEP-II (SLAC, USA), and Belle were primarily designed for the study of B mesons, composed of b quarks, and in particular CP violation induced in $B^0 - \bar{B}^0$ mixing systems by the complex phase of the CKM matrix. Capable of producing more than a million of B meson pairs per day, PEP-II and KEKB have been described as B-factories, a family that SuperKEKB has joined since it has started operating in 2018.

Their functional principle is the collision of asymmetric electron and positron beams at a center of mass energy $\sqrt{s} = 10.58$ GeV. This energy corresponds to the mass of the $\Upsilon(4S)$ resonance, a $b\bar{b}$ meson which decays into B^+B^- or coherent (entangled) $B^0\bar{B}^0$ pairs at a rate of approximately 96% [10]. The asymmetry, on the other hand, translates as a Lorentz boost in the laboratory frame that makes the produced particles travel some distance before decaying. This allows to measure the decay length that separates mixed B^0 mesons, giving in return the decay time required for the study of CP violation.

While the beam energies at KEKB were 8 GeV for electrons and 3.5 GeV for positrons, they are, respectively, 7 and 4 GeV at SuperKEKB. The boost $\beta\gamma$ is reduced from 0.425 to 0.284, decreasing at the same time the separation of B mesons, which must be compensated by better resolution performances in track and vertex reconstruction, but also beam-induced backgrounds that could otherwise weaken the performance of analyses.

SuperKEKB operates at the same facilities than KEKB, in a tunnel of 3 km of circumference. A first electron beam is generated in a pre-injector by sending photons from a short-pulsed laser onto a cathode target. The energy of the resulting beam is then increased up to 7 GeV in a linear accelerator (linac) before being injected inside the High Energy Ring (HER). Electrons from the same pre-injector are also used to irradiate a tungsten target halfway through the linac, producing a positron beam accelerated up to 4 GeV before injection inside the Low Energy

Ring (LER). Both beams end up colliding in a region called "interaction point" (IP), after a final focusing by magnetic field (QCS magnets), along an axis usually taken as "z". A schematic view of the collider is given in Figure 2.1.

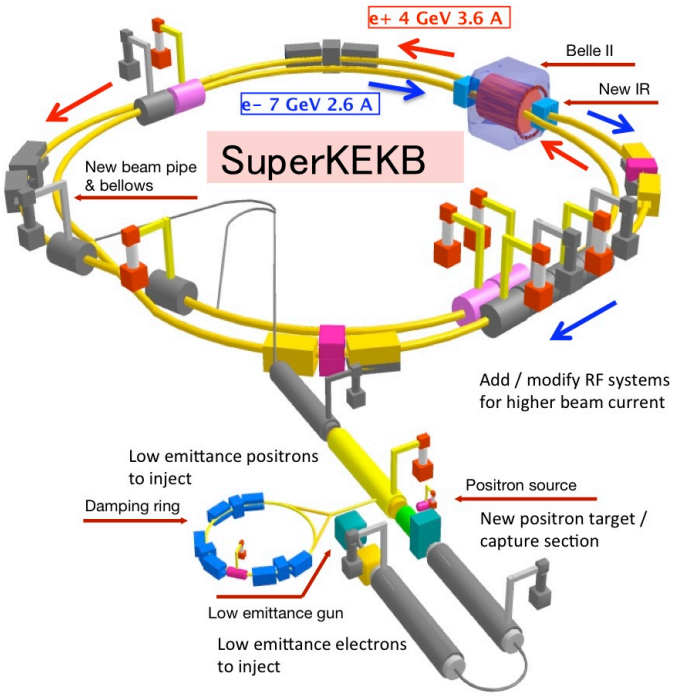


Figure 2.1: Schematic view of the SuperKEKB collider.

SuperKEKB seeks to achieve an instantaneous luminosity of $\mathcal{L} = 6{\times}10^{35}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, thirty times the record of KEKB. The time-dependent instantaneous luminosity is related to the number of collisions N as formulated in the expression

$$\frac{\mathrm{d}N}{\mathrm{dt}} = \mathcal{L}\sigma, \tag{2.1}$$

where σ is the cross section of any considered process. Therefore, the integrated luminosity is

$$L = \frac{N}{\sigma}. \tag{2.2}$$

Belle II targets an integrated luminosity[1] $L = 50\ \mathrm{ab}^{-1}$, fifty times more than what was collected at Belle.

To reach this unprecedented instantaneous luminosity, some parameters of the machine have to be modified compared to KEKB:

[1] Here expressed in barns: $1\ \mathrm{b} = 10^{-24}\ \mathrm{cm}^2$.

- the beam currents I will be increased by a factor ~ 2,
- the vertical beta function at the IP β_y will be reduced by a factor ~ 15, following the nano-beam scheme proposed for the abandoned project of a SuperB collider in Italy [37].

Indeed, one can also write the luminosity $\mathcal{L}$ as [38]

$$\mathcal{L} = \frac{\gamma^{\pm}\xi_y n^{\pm} f_c}{2 r_e \beta_y}\left(1 + \frac{\sigma_y}{\sigma_x}\right), \tag{2.3}$$

where the sign + (−) stands for the positron (electron), γ is the Lorentz factor, ξ_y the vertical tune shift scale[2], n the number of electrons or positrons in a bunch, f_c the frequency of collisions, r_e the electron radius and σ_x (σ_y) the horizontal (vertical) beam size. In particular:

$$n^{\pm} f_c \propto \frac{I^{\pm}}{e}. \tag{2.4}$$

Reducing β_y requires a smaller overlap region d of the beams when they collide, represented in Figure 2.2. Part of this is achieved with a non-zero crossing angle $2\phi = 83$ mrad in Belle II (four time larger than in Belle), thanks to the inverse proportionality between the two parameters:

$$d \simeq \frac{\sigma_y}{\phi}. \tag{2.5}$$

This choice, however, is mainly related to magnet design and reduction of detector background; the angle cannot be increased passed a certain point so that the design value of ξ_y is preserved, thus beam sizes are reduced instead.

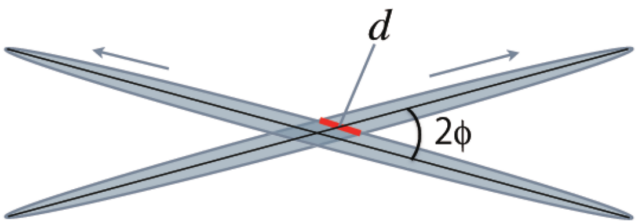


Figure 2.2: Schematic view of two beams colliding in SuperKEKB, following the nano-beam scheme.

Apart from the $\Upsilon(4S)$ resonance, many other processes are involved at Su-

[2] Beam-beam parameter, accounting for interactions between beams.

perKEKB, of which the production cross sections are described in Table 2.1. In fact, the Belle II experiment is also very suited for the study of $e^+e^- \to \tau^+\tau^-$ events, with approximately 46×10^9 τ-pairs produced at the target integrated luminosity of 50 ab^{-1}, to be compared to 55.5×10^9 B meson pairs. At the luminosity of 189.88 fb^{-1}, which corresponds to the amount of data analysed in this thesis, one can expect 174.5 millions of τ-pairs.

Table 2.1: Production cross section of physics processes resulting from e^+e^- collisions at $\sqrt{s} = 10.58$ GeV.

Physics process	Cross section (nb)
$\Upsilon(4S)$	1.110
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.40
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.30
$\tau^+\tau^-(\gamma)$	0.919
$e^+e^-(\gamma)$	300
$\gamma\gamma(\gamma)$	4.99
$\mu^+\mu^-(\gamma)$	1.148
$e^+e^-e^+e^-$	39.7
$e^+e^-\mu^+\mu^-$	18.9

2.2. The Belle II detector

The interaction point where electrons and positrons collide in SuperKEKB is surrounded by an eight-meter tall cylinder, the Belle II detector. It is for a large part an upgrade of the previous Belle experiment's device, designed as the piling of multiple layers, each component being a detector in itself with a function of its own. A simple 3D representation can be found in Figure 2.3 and a more detailed 2D description in Figure 2.4.

The purpose of the detector is the reconstruction of charged and neutral particles produced in the collisions, along with the missing energy (or momentum). To this

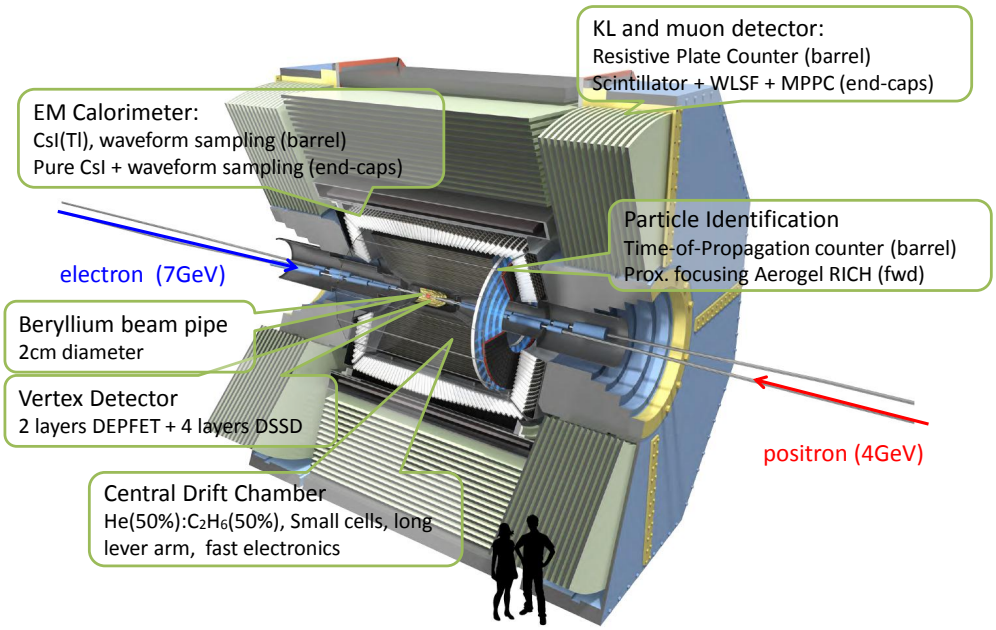


Figure 2.3: Schematic view of the Belle II detector.

end, it has to meet a few essential requirements: a precise measurement, with high resolutions, of a particle's four-momentum and space-time coordinates, including vertices of decaying particles; a large polar angle coverage; an efficient particle identification (PID) system; a high functioning speed to allow for good performance in trigger decision.

In order to measure the momenta, a superconducting solenoid placed inside the detector generates a uniform magnetic field of 1.5 T. Because of it, a charged particle will follow a helical trajectory of radius r, which can later be extracted from the track reconstruction. By equating the Lorentz and centripetal forces, one obtains the momentum p from

$$qvB = m\frac{v^2}{r} \Rightarrow p = Bqr, \tag{2.6}$$

where q is the particle's electric charge, m its mass, v its velocity and B the magnetic field.

In the following subsections, we describe the different parts of the Belle II detector from the innermost to outermost layers.

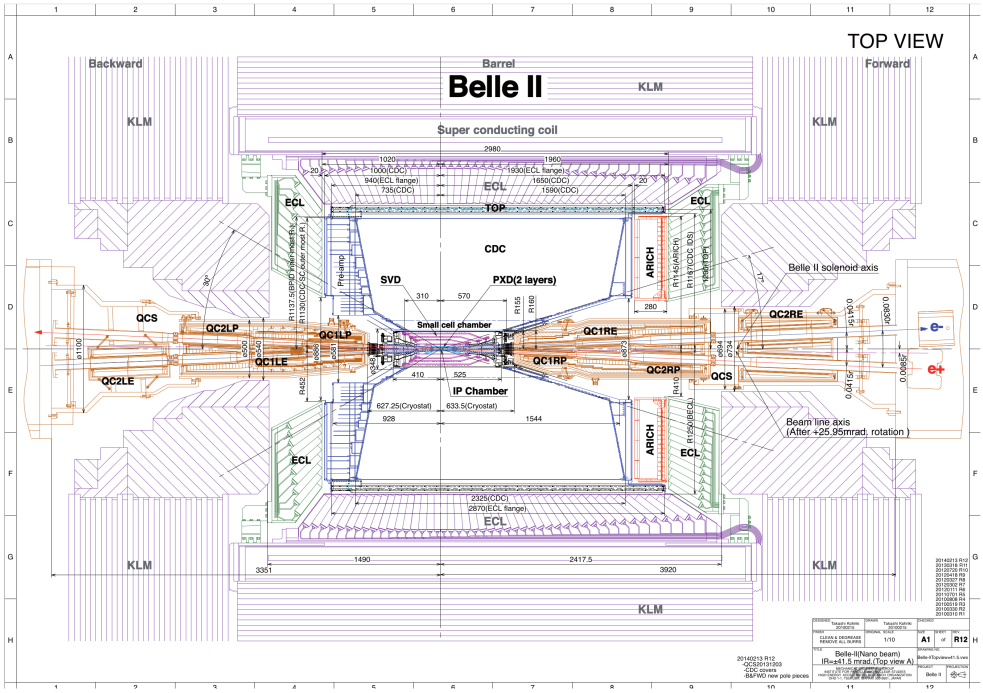


Figure 2.4: Detailed longitudinal view of the Belle II detector.

2.2.1. Vertex Detector

The Vertex Detector (VXD) serves to reconstruct vertices of decaying particles and provides information particle identification for low-momenta tracks, by measuring the energy loss from ionisation. It involves two sub-systems, the Pixel and Silicon Vertex Detectors, for a total amount of six layers of sensors. A global view of the VXD's geometry is presented in Figure 2.5.

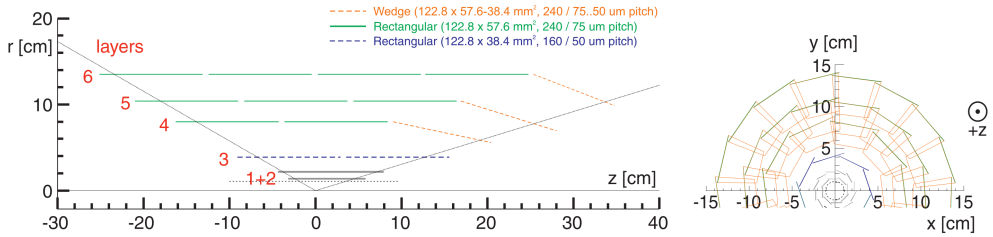


Figure 2.5: Schematic longitudinal (left) and transverse (right) view of half the VXD.

2.2.1.1. Pixel Detector

The Pixel Detector (PXD) is the part closest to the beam pipe. In order to improve the resolution on the vertex reconstruction, the PXD must be very close to the IP, which in return increases the occupancy of the detector, an effect mitigated by the use of pixel sensors instead of strips. As such, it combines two layers of pixelated silicon sensors based on the DEPFET (DEPleted Field Effect Transistor) technology, which allows to minimise the thickness of sensors down to 50 μm and thus the effects of multiple scattering.

The first and second layers are placed at radii of 14 and 22 mm with respect to the IP, respectively made of eight and 12 ladders, for a polar angle coverage from 17 to 150°. Here, a ladder is formed of two joined modules, one forward and one backward, each module corresponding to an arrangement of a sensor and readout devices. The sensitive regions of the sensors contain 1600 pixel rows, 800 on each side, the size of one pixel being 50×55-60 μm^2 (50×70-85 μm^2) in inner (outer) modules. The readout time per pixel row is 100 ns, while four rows are treated in parallel, which makes the PXD's readout time 20 μs per module.

At this time, the installation of the PXD has not been completed and only two of the second layer's ladders are in place. An illustration of the full PXD is shown in Figure 2.6.

2.2.1.2. Silicon Vertex Detector

The Silicon Vertex Detector (SVD) is a set of four layers of double-sided silicon strip detectors (DSSDs), of different sizes, at radii of 38, 80, 115 and 140 mm. The innermost to outermost layers are composed of 8, 10, 14 and 17 ladders, each ladder having 2, 3, 4 and 5 modules, respectively. As represented in Figure 2.5 and visible in a picture of the SVD shown in Figure 2.6, the three last layers present slanted modules in the forward region (boost direction) in order to maintain the same coverage as the previous ones, while minimising the material budget crossed by particles flying out from the IP.

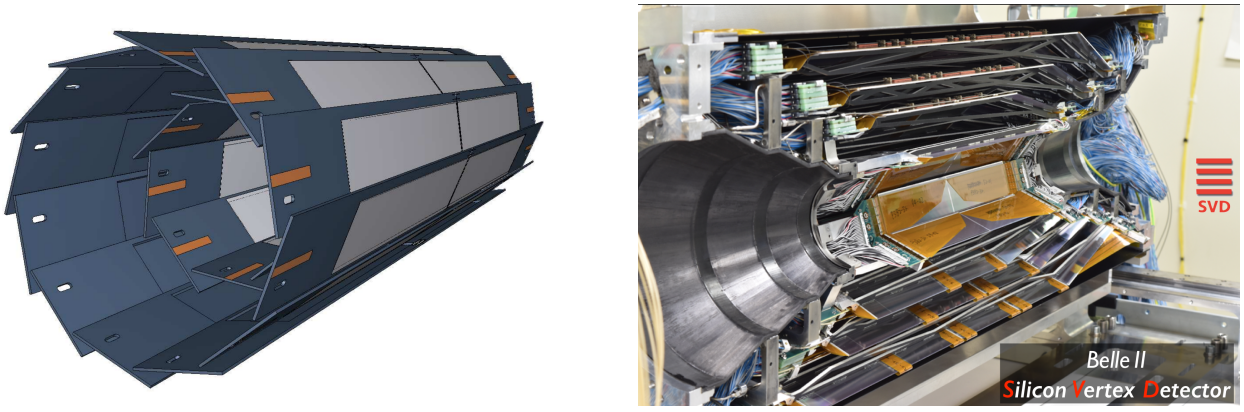


Figure 2.6: Schematic view of the PXD (left) and picture of the SVD [39] in longitudinal view before its installation (right).

A comparative study of the cluster position resolution in the SVD [40, 41], that exploits different measurement strategies, has shown a performance of 7 to 11 μm for the azimuthal resolution and 23 to 35 μm for the longitudinal one.

2.2.2. Central Drift Chamber

The Central Drift Chamber (CDC), which extends from a radius of 16 cm to 113 cm from the IP, is composed of 56 layers of 14 336 sense wires, all arranged in nine super-layers that alternate between axial (following the magnetic field of solenoids) and stereo (skewed by an angle of at most 80 mrad) wires. Such a layout enables the reconstruction of charged tracks in three dimensions. Additionally, the CDC performs precise momenta measurements and provides information on energy losses $\mathrm{d}E/\mathrm{d}x$ for particle identification purposes, along with trigger signals for charged particles.

The chamber is made of drift cells, from 6×10 mm^2 to 18×18 mm^2, filled with a mixture of helium and ethane at 50% proportion each, for a maximal drift time of

350 ns (corresponding to the largest cells). The spatial resolution is about 100 μm, while the energy loss resolution is 12% at 90° incident track angle.

2.2.3. Particle identification via Cherenkov radiations

The Time-Of-Propagation (TOP) counter, in the barrel, and the Aerogel Ring-Imaging Cherenkov (ARICH) counter, in the forward end-cap of the detector, contribute to the PID system. They both take advantage of the information extracted from Cherenkov radiations in order to identify charged particles crossing their material, in this case the Cherenkov angle retrieved from the photons' space-time coordinates. This gives access to the charged particle's velocity, therefore to the likelihoods of the different mass assumptions.

2.2.3.1. Time-Of-Propagation counter

The TOP counter is formed of 16 quartz radiators placed outside and around the CDC. Each radiator comprises two quartz bars joined together for a total 2.5 m length, 44 cm width and 2 cm thickness, a spherical mirror on the forward end of the bar and micro-channel plate photomultiplier tubes (MCP-PMTs) on the backward end. Photons reflected inside the quartz bars may either propagate to the MCP-PMTs, which provides the timing and two-dimensional spatial information (x, y), allowing to reconstruct the emitted Cherenkov ring, or travel in the opposite direction towards the forward end, before being sent back by the spherical mirror. The modules and their operating principle are illustrated in Figure 2.7.

The use of MCP-PMTs enhances significantly the time resolution performances, with transit time spreads[3] as low as ∼30 ps at a gain of $\sim 10^6$. Without the improvements granted by this method, the resolution would be close to 100 ps.

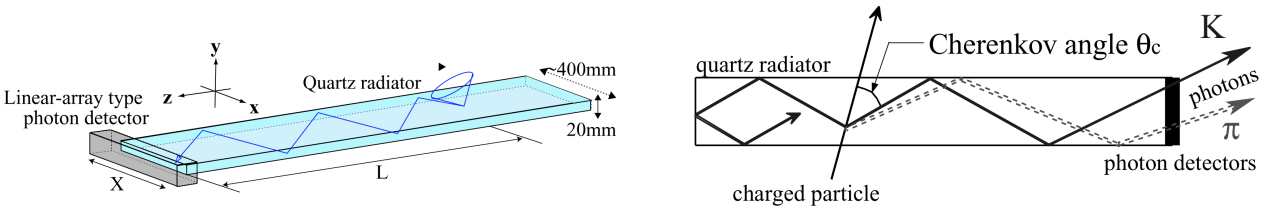


Figure 2.7: Schematic views of a single module of the TOP counter (left) and of its operating principle (right), consisting in identifying particles thanks to the characterisation of each reflected photon's trajectory and timing, emitted as Cherenkov radiations.

[3] Transit time of photo-electrons generated by Cherenkov photons entering an MCP-PMT.

2.2.3.2. Aerogel Ring-Imaging Cherenkov counter

The components of the ARICH counter, represented in Figure 2.8, consist in a Cherenkov photon radiator made of aerogel and an array of photon detectors measuring their (x, y) positions, with a 20 cm expansion volume in between to form well-defined Cherenkov rings, and finally a readout system for the detectors.

In order for the charged particle to emit a sufficient amount of photons, and thus to reconstruct the ring with high precision, the aerogel radiator is 4 cm long in the z direction and combines two layers of aerogel with refractive indices 1.045 and 1.055. Indeed, the higher the thickness, the larger the photon emission uncertainty, thus changing the refractive properties of the aerogel reduces the spread of photons on the detector's plane and makes the Cherenkov rings overlap.

The resolution on the angle is found to be around 3 mrad, measured as $\sigma_{\text{track}} = \sigma_\theta/\sqrt{N}$, where $\sigma_\theta \simeq 14$ mrad is the single photon resolution and N is the number of photons emitted by a single track, here assumed to be 20.

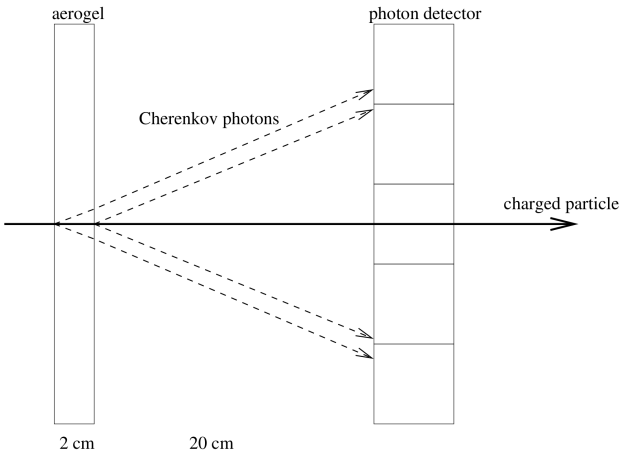


Figure 2.8: Schematic representation of the ARICH counter and its principle of operation.

2.2.4. Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) is designed for the detection of photons as well as the measurement of their energy and angular coordinates, while contributing to the identification of electrons and K_L^0, the trigger system and the luminosity determination.

Being merely an upgrade of the previous Belle device, essentially for the readout system, the ECL has an inner radius of 1.25 m and is a 3 m long barrel section

(including endcaps), from $z = -1.02$ m to $z = +1.96$ m. It covers polar angles from 12.4 to 155.1°, therefore a slightly larger region than the VXD and CDC, although it doesn't include two gaps of approximately 1° between the barrel and endcaps.

It contains 8736 CsI(Tl) crystals, with 98 different shapes corresponding to truncated pyramids of average volume $6\times6\times30$ cm^3. Two photodiodes of 10×20 mm^2 are attached to each crystal, providing two independent signals from scintillation light through their respective preamplifiers, the outputs being summed at a later stage.

In the Belle configuration, the intrinsic energy resolution σ_E/E ranged from 4% ($E = 100$ MeV) to 1.6% ($E = 8$ GeV), performances that are also expected in Belle II in the absence of background.

2.2.5. K-Long and Muon detector

The purpose of the K-Long and Muon detector (KLM) is mainly the detection and identification of muons and K_L^0 mesons. Depending on their momenta and polar angles, muons can cross the KLM in its entirety, while K_L^0 mesons might interact with the ECL and/or KLM, producing hadronic showers in one or both of them.

It is placed outside the superconducting solenoid as shown in Figure 2.9 and alternates 14 (15 in the barrel) layers of iron plates and 14 layers of active detector elements. The plates are incorporated not only as magnetic flux return for the solenoid, but also as a medium equivalent to 3.9 interaction lengths, in which K_L^0 mesons can shower hadronically.

The polar angle coverage is 45 to 125° in the barrel, 20 to 155° with the endcaps. The detection system in Belle was based on glass-electrode resistive plate chambers (RPCs) that are not suited for Belle II, given the amount the background expected in its new configuration. They have indeed long dead times after a discharge, that is why RPCs in the two innermost layers of the barrel and in all layers of the endcaps have already been replaced by scintillator strips, read out by silicon photomultipliers, with a time resolution around 0.7 ns.

The muon detection efficiency reaches 90% for momenta higher than 1 GeV/c, and 80% at 3 GeV/c in the case of K_L^0 mesons.

2.3. Track reconstruction and performance

The track reconstruction in Belle II [42], or "tracking", consists in the combination of single hits (spatial points) into a track (full trajectory) after a particle crosses one or multiple sub-detectors, before the properties of the said track are retrieved. The first step is called track finding; the second, track fitting.

The tracking relies on the information that the VXD and CDC provide. Two algorithms are considered:

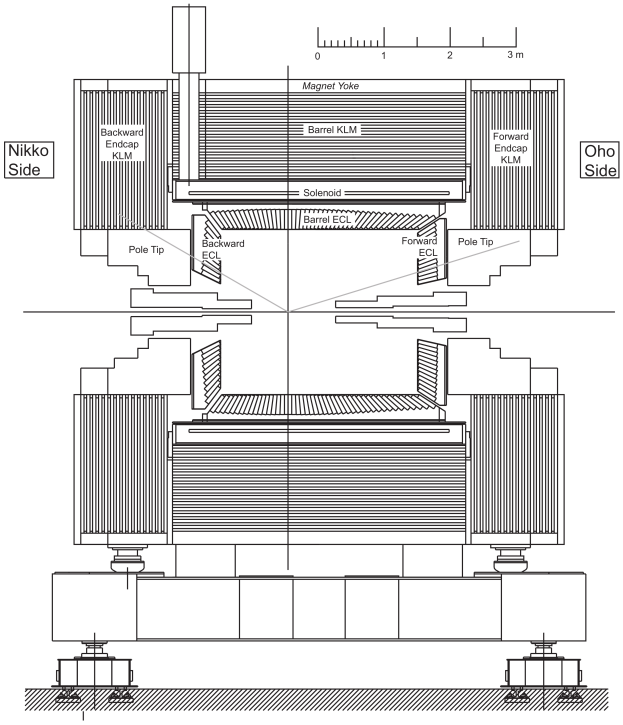


Figure 2.9: Schematic longitudinal view of the KLM.

- a *global Legendre algorithm*, used only for the CDC: a Legendre transformation is applied to the spatial coordinates of all hits at the same time, changing the position space into a $(\theta, \rho_{\pm})$ parameter space which represents all the circles traversing both the IP and the considered hit. θ is the angle of the tangent to the circle at the IP position, while $\rho_{\pm}$ is its signed curvature. Hits for which a common circle can be found in the $(\theta, \rho_{\pm})$ space may then be combined into a track.

- a *local Cellular Automaton*, used in both VXD and CDC: since the previous approach assumes that particles follow circular trajectories, this second algorithm aims at taking into account possible variations. Sets of three hits, or triplets, are put in relation to other triplets that share exactly two hits with them. These relations can be seen as paths from one triplet to another and are assigned weights according to a figure of merit. By combining the single paths into segments, or even a global path, with the highest total weights, one can either improve the track finding from the Legendre algorithm or reconstruct new tracks.

The track finding, alas, is not always perfect. A track might sometimes be

- a *fake track*, if it includes hits from beam-induced background, or combines hits from two different particles;
- a *clone track*, if other tracks are reconstructed from the same particle.

A measurement of the fake tracks rate (including clones) in $e^+e^- \to \tau^+\tau^-$ events is presented in Chapter 3.

The parameters of a track, here a helix, are determined thanks to a Kalman filter algorithm applied to the track fitting. This includes the position of the point of closest approach (POCA) to the IP, represented in Figure 2.10 by taking the IP as the origin of the reference frame, and in particular

- d_0, the signed distance of the POCA in the transverse plane (x, y), the sign being the same as the track's angular momentum at the POCA with respect to the magnetic field;
- z_0, the z-coordinate of the POCA.

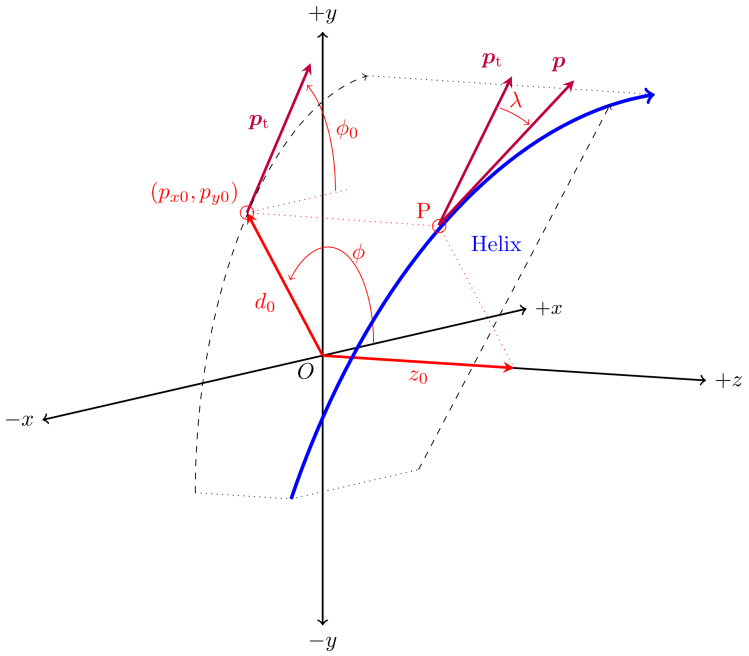


Figure 2.10: Three-dimensional representation of the helical trajectory of a track at the point of closest approach (P) to the IP (the origin O). p is the momentum of the track at the point P, p_t its transverse momentum and λ the angle between the two vectors. Taken from [43].

The tracking efficiency, that is the efficiency in reconstructing the track of a particle produced after a collision, has been studied in Belle II using $e^+e^- \to [\tau^\pm \to \pi^+\pi^-\pi^\pm\nu_\tau] + [\tau^\mp \to \ell^\mp\bar{\nu}_\ell\nu_\tau]$ $(\ell = e, \mu)$ events, by comparing the amounts of candidates where one of the charged pions is either missing or correctly reconstructed [44].

The tracking efficiency varies from 75% at low transverse momenta to 95% around 4 GeV/c. It degrades the closer the track is to the beam axis (small or large polar angles), while it remains mostly constant around 90% regardless of the azimuthal angle.

The calibrated discrepancy

$$\delta^* = 1 - \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} \tag{2.7}$$

between the efficiencies measured in data (ϵ_{data}) and Monte Carlo simulation (ϵ_{MC}) is shown in Figure 2.11, overall below 0.5%.

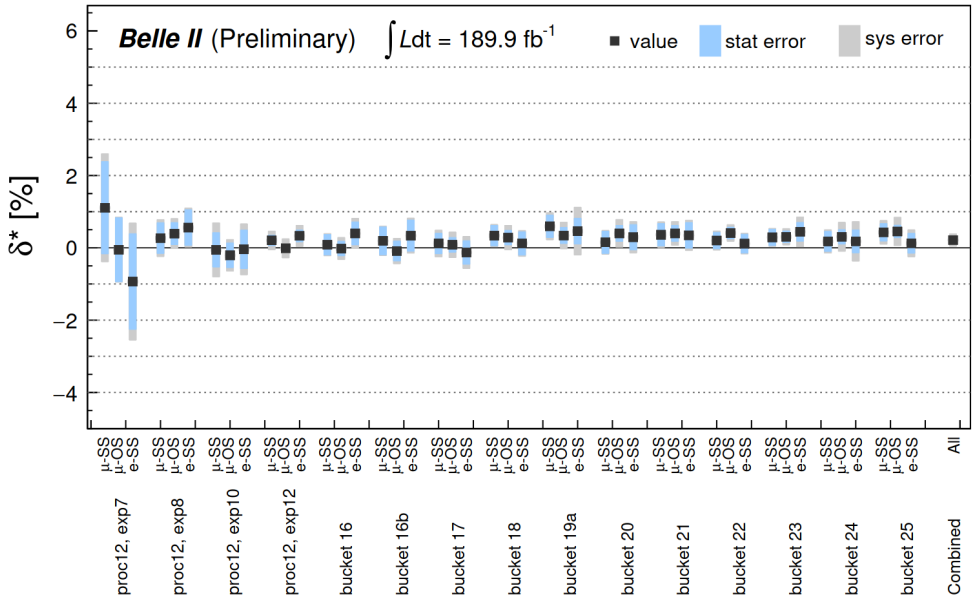


Figure 2.11: Discrepancy δ^* between data and simulation in the tracking efficiency, measured on 3×1-prong $[\tau^\pm \to \pi^+\pi^-\pi^\pm\nu_\tau] + [\tau^\mp \to \ell^\mp\bar{\nu}_\ell\nu_\tau]$ $(\ell = e, \mu)$ events, as a function of the data taking period (see Table 2.2) and of the studied channel (either electron (e) or muon (μ) 1-prong, and the two not missing pion tracks of opposite (OS) or same sign (SS) electric charge). The measurement is performed on 189.9 fb^{-1} of data collected from 2019 to 2021 and on simulation samples. The blue bands correspond to the statistical (stat) errors and the grey ones to the systematic (sys) errors. Taken from [45].

2.4. Trigger systems

Because of the high instantaneous luminosity that is targeted at SuperKEKB, the experiment must cope with large event rates, evaluated as 20 kHz for the main physics processes and at the previously planned luminosity $\mathcal{L} = 8{\times}10^{35}$ cm^{-2}s^{-1}. In addition, one must take into account the many beam-induced and luminosity-dependent background sources, that also affect Belle II more than Belle for the same reason, for example the Touschek effect or the radiative Bhabha ($e^+e^- \to e^+e^-\gamma$) and two-photon ($e^+e^- \to e^+e^-e^+e^-$) processes. Being able to discriminate these low-multiplicity background events is necessary for carrying out physics analyses on other events, like τ-pair studies and dark photon searches, where low amounts of tracks are also expected.

To this end, Belle II has a trigger system consisting in:

- a low-level or "Level 1" trigger (L1), implemented in hardware, with maximum output trigger rate and latency of 30 kHz and 5 μs, respectively: it intends to suppress beam-induced, radiative Bhabha and two-photon background events by combining information from the CDC and ECL;

- a High Level Trigger (HLT), software-based, which can further reduce the event rate of data stored offline to 10 kHz: it performs a full event reconstruction and defines a "region of interest" for events to be stored.

For most of the data taking period considered here, the HLT has been kept inactive except for the production of offline selections, also called "skims", that we don't consider here. However, it has been running in filter mode since March 2021 (*experiment 17*, see Section 2.7 below), without loss of efficiency for the analyses reported in the next chapters.

On the other hand, the L1 trigger, divided into multiple trigger lines or bits, already contributes to the filtering of the background through some lines that require the candidates to meet some selection criteria. On top of that, a given trigger bit t_{L1} is assigned a level of "prescale", an integer i, which defines the proportion $1/i$ of candidates effectively retained by t_{L1} among all the ones that passed its requirements. This helps to further reduce the event rate as the luminosity of SuperKEKB increases.

Details regarding the trigger lines that are considered in this document are given in Appendix A.

2.5. Particle identification and performance

The PID relies on information[4] provided by the SVD, CDC, TOP, ARICH, ECL and KLM, then combined together into likelihoods [46]. It is translated into a probability $P(h)$ for a given track to originate from a particle type h:

$$P(h) = \frac{\mathcal{L}_h}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}, \tag{2.8}$$

where $\mathcal{L}$ stands for the likelihood, taken as the product of the likelihoods from each considered sub-detector, while h is either of the electron (e), muon (μ), pion (π), kaon (K), proton (p) or deuteron (d) type. This PID probability is said "global" as opposed to "binary" calculations, which take only one particle type as reference, the denominator being simply the sum of the likelihoods for h and the reference particle type.

The electron and muon identification efficiencies have been found to be respectively 86% and 88.5% at a lower threshold of 0.9 on the PID probability, based on the study of $J/\psi \to \ell^+\ell^-$ ($\ell = e, \mu$) events [47]. Data and simulation are overall in good agreement, except at low particle momenta where discrepancies within 20% are observed. In addition, the pion misidentification rates have been measured according to the $K^0_S \to \pi^+\pi^-$ decay, as 0.4% for electron candidates and 7.3% for muon ones.

The performance of the kaon identification in Belle II [48] is summarised in Figure 2.12. The efficiency varies from 95% at the lowest thresholds to around 70% for the tightest selections. The pion misidentification rate follows a similar trend, from ∼25% to less than 5%. Data and simulation are the most consistent at the highest kaon identification efficiencies and lowest pion misidentification rates.

2.6. Belle II Analysis Software Framework

The Belle II Analysis Software Framework [49], or *basf2*, is one of the three parts of the core software of Belle II, with the *externals* that contain third-party code like ROOT, Geant4 or EvtGen, and all the *tools* that allow the installation, as well as configuration, of basf2 and externals [50].

The processing chain of data within basf2 is illustrated in Figure 2.13. Small processing blocks (`modules`) are linearly arranged in a specific order, which depends on the specific task that one intends to perform, inside a container (`path`). The `modules` of a `path` will therefore be executed one by one in that order.

The processed data is readable and writable from a common storage (`DataStore`) by all `modules`. Furthermore, a set of a conditions can be loaded from a central

[4] For example: energy measurements from the calorimeter; penetration distance of particles inside the detector; momenta and energy loss measurement from tracking devices; Cherenkov radiation angles...

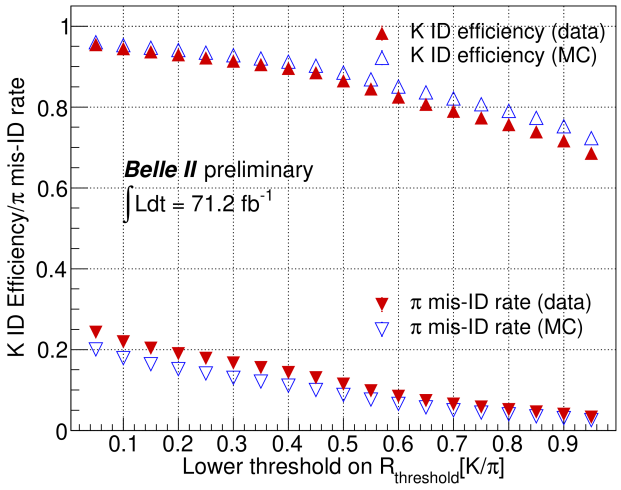


Figure 2.12: Kaon identification efficiencies and pion misidentification rates in data and Monte Carlo simulation (MC), as a function of the threshold on the binary kaon identification probability (with pion particle type as reference). The measurements are performed using events reconstructed as $D^{*+} \to [D^0 \to K^-\pi^+]\pi^+$, on 71.2 fb^{-1} of data collected in 2019 and early 2020 and on simulation samples. Taken from [48].

database and are accessible to the modules from the `DBStore`. Usually, this database consists of ROOT files (`payloads`), identifiable by a name and a revision number, to which are assigned "intervals of validity" (`iov`), indicating the data taking periods that are relevant for them.

A collection of `payloads` and `iov` is a `globaltag`, with a unique name and valid only for a given dataset. An example of the use of `globaltags` is the weighting of simulated events, in order to correct discrepancies with respect to the experimental data after a selection is applied on PID variables.

In physics analyses, the event reconstruction is performed thanks to basf2 functions within a Python script, often called "steering file". Lists of tracks which have already been reconstructed are loaded and combined together, from the final to the initial state, to form the studied decay.

Requirements can be set on both tracks and particles that result from their combinations, for example to select a desired particle type accurately or to suppress some background. All reconstructed events that remain at the end are stored in an output ROOT file, said "ntuple", that lists for each of them the values of the requested variables.

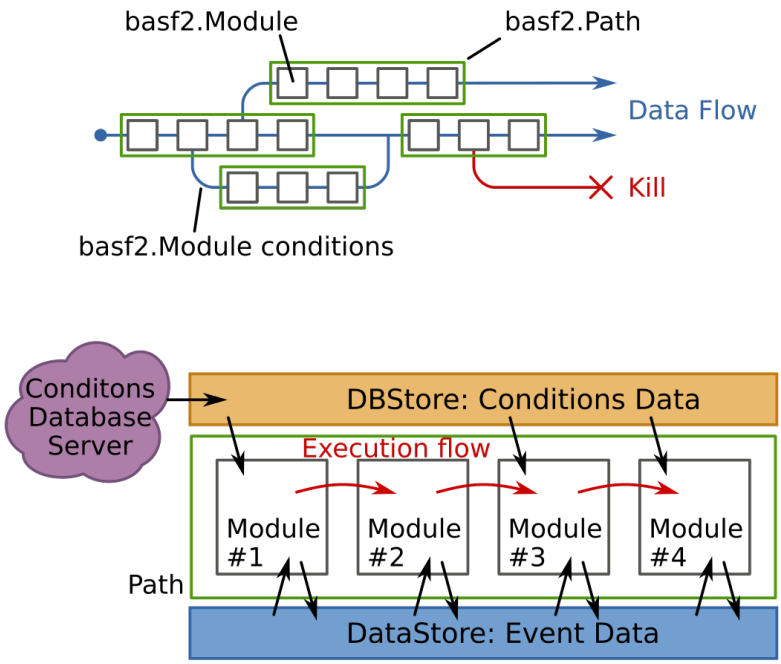


Figure 2.13: Schematic view of the data processing flow as it is executed in basf2.

2.7. Dataset production

2.7.1. Experimental data

The experimental datasets, or simply data, are first of all categorised according to the data taking period they belong to, with the experiment and run numbering. The former corresponds to long periods (usually a few months) with stable experimental conditions; the latter, to smaller time spans between the start of the data taking and the moment it has to be temporarily stopped, either because the 8-hour time limit has been reached, or an issue in the experimental setup has to be taken care of.

A more commonly used label refers to the processing of the data, which depends, in particular, on the software release of basf2 as well as on the calibration process. There are two main calibration steps [51]:

- the prompt calibration, carried out regularly after a period of the order of a few days, which leads to prompt processings of data called "buckets" (*bucketX* where *X* is a number), suitable for physics studies (conference notes or internal use);

- the offline calibration, with updated therefore better calibration constants compared to the previous step, leading to major data reprocessings called

"proc" (*procX* where X is a number) that take a few months before completion, suitable for journal publications.

The data analysed in Chapters 3 and 4 have been collected from March 2019 to July 2021, with more details given in Table 2.2. The total integrated luminosity is $189.88 \pm 0.01_{\text{stat}}$ fb^{-1}.

Table 2.2: Experimental datasets analysed in the next chapters. The uncertainties on the integrated luminosities are statistical.

Experiment number	Dataset	Integrated luminosity (pb)	Data taking period
7	proc12	509.8 ± 1.5	03/2019 – 05/2019
8		4463.6 ± 1.5	05/2019 – 06/2019
10		3635.8 ± 1.1	09/2019 – 12/2019
12		54642.2 ± 4.2	03/2020 – 07/2020
14	bucket16	10752.1 ± 3.7	10/2020 – 12/2020
	bucket16b	5666.3 ± 2.7	
16	bucket17	10321.0 ± 3.7	02/2021 – 03/2021
17	bucket18	10713.9 ± 3.7	03/2021 – 04/2021
18	bucket19a	8863.1 ± 3.2	04/2021 – 07/2021
	bucket20	9028.4 ± 3.2	
	bucket21	8737.7 ± 3.1	
	bucket22	17606.8 ± 4.5	
	bucket23	18011.2 ± 4.5	
	bucket24	11232.4 ± 3.6	
	bucket25	15696.1 ± 4.2	

2.7.2. Monte Carlo simulation

Monte Carlo simulations (MC) are identified as *MCX*, where X is the numbering of their production campaign. Indeed, the MC sample production depends more or less on the same conditions as in the data: it can either be run-dependent (MCrd)

or run-independent (MCri). MCrd samples mimic the data, with datasets dedicated to each data run, following the same experimental conditions (including calibration) and keeping the same proportions of integrated luminosities. On the other hand, MCri samples depend solely on the basf2 release they are produced with.

MC samples are generated for many physics processes, including the ones cited in Table 2.1 and some other low-multiplicity decays. A list of generators for the simulation of decays is given in Table 2.3. Beam-background events are added by background overlay and are either simulated (for MCri) or extracted from data (for MCrd).

The interactions with the detector are simulated thanks to `Geant4` [52], while the `Opera3D/TOSCA` [53] software models the magnetic field. Finally, trigger lines are included with the trigger simulation software `TSIM` [54].

MC samples for signal decays that are forbidden or highly suppressed in the SM may also be generated; in that case, the decay is modelled by imposing a Lorentz-invariant phase space, without any new physics effect.

Table 2.3: Simulated physics processes and their respective generators.

Physics process	Generator
B/D meson decays (exclusive final state)	`EvtGen` [55]
B/D meson decays (inclusive final state) Light $q\bar{q}$ continuum	`PYTHIA` [56]
$e^+e^- \to \tau^+\tau^-$	`KKMC` [57]
τ decays	`TAUOLA` [58]
$e^+e^- \to e^+e^-(\gamma)/\gamma\gamma(\gamma)$	`BABAYAGA.NLO` [59, 60]
$e^+e^- \to e^+e^-\ell^+\ell^-$ $(\ell = e, \mu)$	`AAFH` [61]

3. Fake tracks rate measurement with $e^+e^- \to \tau^+\tau^-$ events

As mentioned in Section 2.3, the track reconstruction is contaminated by fake and clone tracks, as well as tracks originating from beam-induced background. Here, we measure the rate of contamination under the name of *fake tracks rate*, although one must bear in mind that the word is used in a broad sense.

The measurement of the fake tracks rate is performed with a "tag-and-probe" method applied to $e^+e^- \to \tau^+\tau^-$ events, which consists in the reconstruction of τ-pair decays with exactly four tracks (*tag*), here $[\tau^\pm \to \pi^+\pi^-\pi^\pm\nu_\tau] + [\tau^\mp \to e^\mp\bar{\nu}_e\nu_\tau]$[1], with the possible presence of an additional fifth track (*probe*), therefore fake. The decays of the taus are also called "3×1-prong" decays, the three pion tag tracks being the 3-prong, while the electron is the 1-prong. This process is illustrated in Figure 3.1.

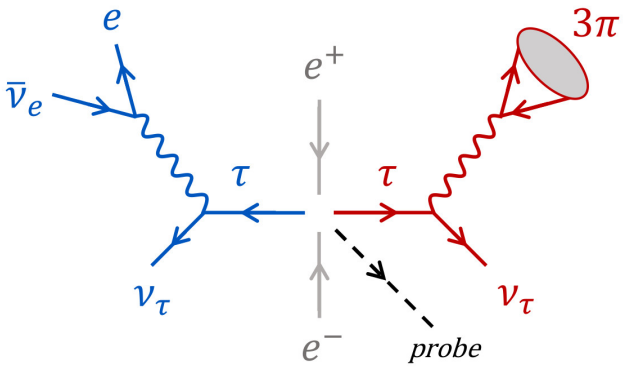


Figure 3.1: Diagram of the 3×1-prong τ-pair decays studied for the fake tracks rate measurement.

The rate r_{fake} is defined as

$$r_{\text{fake}} = \frac{N_5}{N_4 + N_5}, \tag{3.1}$$

[1] $\mathcal{B}(\tau^\pm \to \pi^+\pi^-\pi^\pm\nu_\tau) \simeq 9.3\%$, $\mathcal{B}(\tau^\pm \to e^\pm\bar{\nu}_e\nu_\tau) \simeq 18\%$ [10].

where N_4 and N_5 are the numbers of events where four and five tracks are found, respectively.

We don't consider the muon decay channel of tau leptons for the 1-prong track, since we observe they are more affected by background events coming from τ-pairs. A more in-depth study would therefore be required for a potential measurement with a muon 1-prong decay, in order to define an appropriate background suppression selection.

3.1. Data and MC samples

The data samples correspond to a total integrated luminosity of 189.88 fb^{-1}, collected from 2019 to 2021 and detailed in Table 2.2.

MC samples are taken from the 14$^{\text{th}}$ MC production campaign, with either run-independent (MCri) and run-dependent (MCrd) samples. They consist in 200 fb^{-1} (MCri) and 756.7 fb^{-1} (MCrd) of τ-pair and $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$ (continuum) samples. $b\bar{b}$ and low-multiplicity background samples (ee, $\mu\mu$, $eeee$, $ee\mu\mu$, $ee\pi\pi$, $eeKK$, $eepp$, $\mu\mu\mu\mu$) do not contribute to this study because of the harsh background suppression criteria mentioned below in Section 3.3.

3.2. Track and event selections

For this study, the track selection and multiplicity criteria are shown in Table 3.1 and 3.2, respectively. They involve the transverse momentum (p_T), the ECL cluster energy (E_{cluster}), the PID probability for a particle to be a muon (muonID), while z_0 and d_0 are already defined in Section 2.3.

All events are also required to fire at least one of the low-multiplicity ECL-based triggers (*lmlX*, with *X=0, 1, 2, 4, 6, 7 ,8, 9, 10, 12, 13*), defined in Appendix A, Table A.1. In the case of MCrd samples, however, trigger efficiencies have been derived from data and applied as rescale factors because of trigger simulation issues. The efficiencies are computed with respect to CDC reference triggers (*ffy*, *ffs* or *fyo*, defined in Appendix A, Table A.2), which rely on independent selections with respect to those that define the *lml* triggers. CDC and *lml* triggers can therefore be considered orthogonal.

3.3. Background suppression

Reaching a high signal purity (signal proportion) is essential for the tag-and-probe method reliability. The main background sources are radiative dilepton events other than τ-pairs, four-lepton final states and continuum hadronisation processes $e^+e^- \to q\bar{q}$ that can mimic the signal topology. To reject this contamination, we

Table 3.1: Track selection criteria.

Variable	Probe pion track	Tag pion track	Tag electron track
p_T [MeV]	–	> 200	> 200
$\lvert z_0 \rvert$ [cm]	< 3	< 3	< 3
$\lvert d_0 \rvert$ [cm]	< 1	< 1	< 1
$\frac{E_{\text{cluster}}}{p}$	–	< 0.6	$[0.8, 1.2]$
E_{cluster}	–	> 0	–
muonID	–	< 0.9	–

Table 3.2: Track multiplicity criteria.

Decay mode	$\text{N}_{\text{pion}}^{\text{probe}}$	$\text{N}_{\text{pion}}^{\text{tag}}$	$\text{N}_{\text{electron}}^{\text{tag}}$
electron channel, 5-track sample	5	≥ 3	≥ 1
electron channel, 4-track sample	4	3	1

exploit the kinematics and topological properties of τ-pairs events and devise a background suppression strategy, consisting in the following points:

- *Angular isolation*: The 1- and 3-prong decay products of the tau leptons tend to propagate back-to-back in the center of mass (CM) frame, with a large angular separation. To exploit this topology, we require the opening angle θ^{CM} between the 1-prong track and each of the 3-prong tracks in the CM frame to satisfy $\cos\left(\theta^{\text{CM}}\right) < -0.5$.

- *1-prong momentum*: The 1-prong track is required to have a momentum in the CM frame between 20 and 80% of the beam energy. The upper threshold significantly reduces the radiative dilepton background, while the lower threshold helps to suppress continuum and $e^+e^- \to e^+e^-\mu^+\mu^-$ backgrounds.

- *Hemispherical separation*: The thrust axis of the event is defined according to the vector $\mathbf{n}_T$ which maximises the magnitude

$$T = \max_{\mathbf{n}_T} \left(\frac{\sum_i |\mathbf{p}_i \cdot \mathbf{n}_T|}{\sum_i |\mathbf{p}_i|} \right), \tag{3.2}$$

  where $\mathbf{p}_i$ is the momentum vector of a final state particle i. The thrust axis

is reconstructed with the kinematics of the tag tracks and used to define two opposite hemispheres. We require that all the 3-prong and probe tracks belong to a same hemisphere, while the 1-prong track belongs to the opposite one.

- *Invariant mass*:
  - $|M_{\pi\pi}^{\mathrm{OS}} - m_\rho| < 100$ MeV/c^2, where $m_\rho = 775.26$ MeV/c^2 [10] and $M_{\pi\pi}^{\mathrm{OS}}$ is the invariant mass of the two tracks from the 3-prong tracks that have opposite-sign electric charges (OS). In our case, this configuration appears twice since we have three tracks. This target hadronic τ-decays with an intermediate ρ meson ($\rho \to \pi^+\pi^-$), and significantly reduces background from continuum.
  - 300 MeV/$c^2 < M_{\pi\pi}^{\mathrm{SS}} < m_\tau$, where $m_\tau = 1776.86$ MeV/c^2 [10] and $M_{\pi\pi}^{\mathrm{SS}}$ is the invariant mass of the two tracks from the 3-prong tracks that have same-sign electric charge (SS). This helps to reduce background from continuum.
  - The invariant mass of the 3-prong tracks combined $M_{\pi\pi\pi}$ is required to be smaller than 1.3 GeV.
- *3- and 1-prong transverse momenta*: The transverse momenta of the combined 3-prong tracks and of the 1-prong track must be higher than 3 GeV and 1 GeV respectively.
- *Particle identification*: We require the electron identification probability of the 1-prong track to be higher than 0.9. Furthermore, the kaon identification probability of the track which has an opposite-sign electric charge compared to the two other tracks of the 3-prong τ-decay must be lower than 0.6. The latter criterion has been added after we observed with Monte-Carlo samples that in the background, in many cases, this track corresponded to a kaon.
- *Neutral multiplicity*: In order to improve the signal purity of events with five tracks, we require that there are no π^0 in the event.

The requirements of the last three points as well as the cut on $M_{\pi\pi\pi}$ have been optimised according to the $S/\sqrt{B}$ figure of merit in 5-track Monte-Carlo samples (using the previous MC13 production campaign), where S (B) is the number of signal (background) events in the sample.

3.4. Definition of signal events in MC

Signal events in the simulated τ-pair samples have been defined by exploiting the available MC truth matching information.

Events must satisfy two requirements to be considered as signal:

1. One τ must decay into three charged final state particles while the other decays into only one.

2. All tag tracks must originate from a track truth-matched to a τ^+ or τ^-.

An event that doesn't meet these two requirements is tagged as background.

3.5. Data-MC comparison

The distributions of data and simulation, along with the bin-by-bin data-MC ratios for the 3-prong tracks invariant mass $M_{\pi\pi\pi}$, the azimuthal angle ϕ_{lab}, the polar angle θ_{lab} and the transverse momentum p_T of the 1-prong and probe tracks, are displayed in Figures 3.2, 3.3, 3.4 and 3.5, respectively. The signal (S) and background (B) yields in MC and total yields in data are shown in Table 3.3. The signal purities in simulation, defined as $S/(S+B)$, are presented in Table 3.4, as well as the estimated signal yields in data which are computed as the product of the overall data yields and the MC signal purities.

We observe a deficit of MC events compared to the data, around 16% (N_4) and 12% (N_5) in MCrd; in MCri, it is around 6% (N_4) and 25% (N_5). Therefore, we conclude that the fake rate measured in MCrd is more reliable than the one of MCri, since the data-MC disagreement in both reconstructions is more consistent.

Table 3.3: Data and MC yields in the 5- and 4-track samples, with statistical uncertainties. The MC yields are rescaled to the data luminosity of 189.9 fb^{-1}.

Sample	MC run dependence	MC signal yields	MC background yields	Data yields
N_5	MCri	495.6 ± 21.7	59.8 ± 7.5	737 ± 27
	MCrd	582.6 ± 12.4	68.3 ± 4.1	
N_4	MCri	47385.6 ± 212.1	536.4 ± 22.6	50805 ± 225
	MCrd	41979.4 ± 106.2	465.5 ± 10.7	

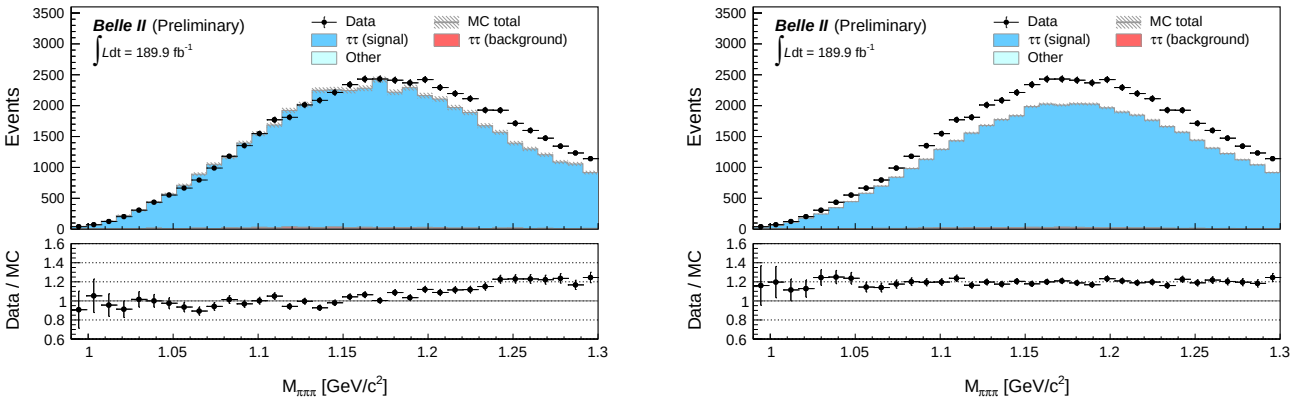


Figure 3.2: Distributions of $M_{\pi\pi\pi}$ for the combined 4- and 5-track samples, using MCri (a) or MCrd (b) simulation. The data (dots) are compared to the signal + background expectation scaled to the integrated data luminosity (stacked histograms). Statistical uncertainties for data and MC are shown, with the MC error band including also the trigger efficiency uncertainty. The lower panel shows the data/MC ratio and the error bars take into account contributions from both data and MC.

Table 3.4: Estimated signal yields in the 5- and 4-track data samples, obtained by the product of the overall data yields from Table 3.3 and the MC signal purities, calculated as $S/(S+B)$. The data yields correspond to an integrated luminosity of $189.9\,\mathrm{fb}^{-1}$.

Sample	MC run dependence	MC signal purity	Estimated data signal yields
N_5	MCri	0.892 ± 0.013	657.6 ± 26.0
	MCrd	0.895 ± 0.006	659.7 ± 24.7
N_4	MCri	0.989 ± 0.001	50236.3 ± 224.1
	MCrd	0.989 ± 0.001	50247.8 ± 223.3

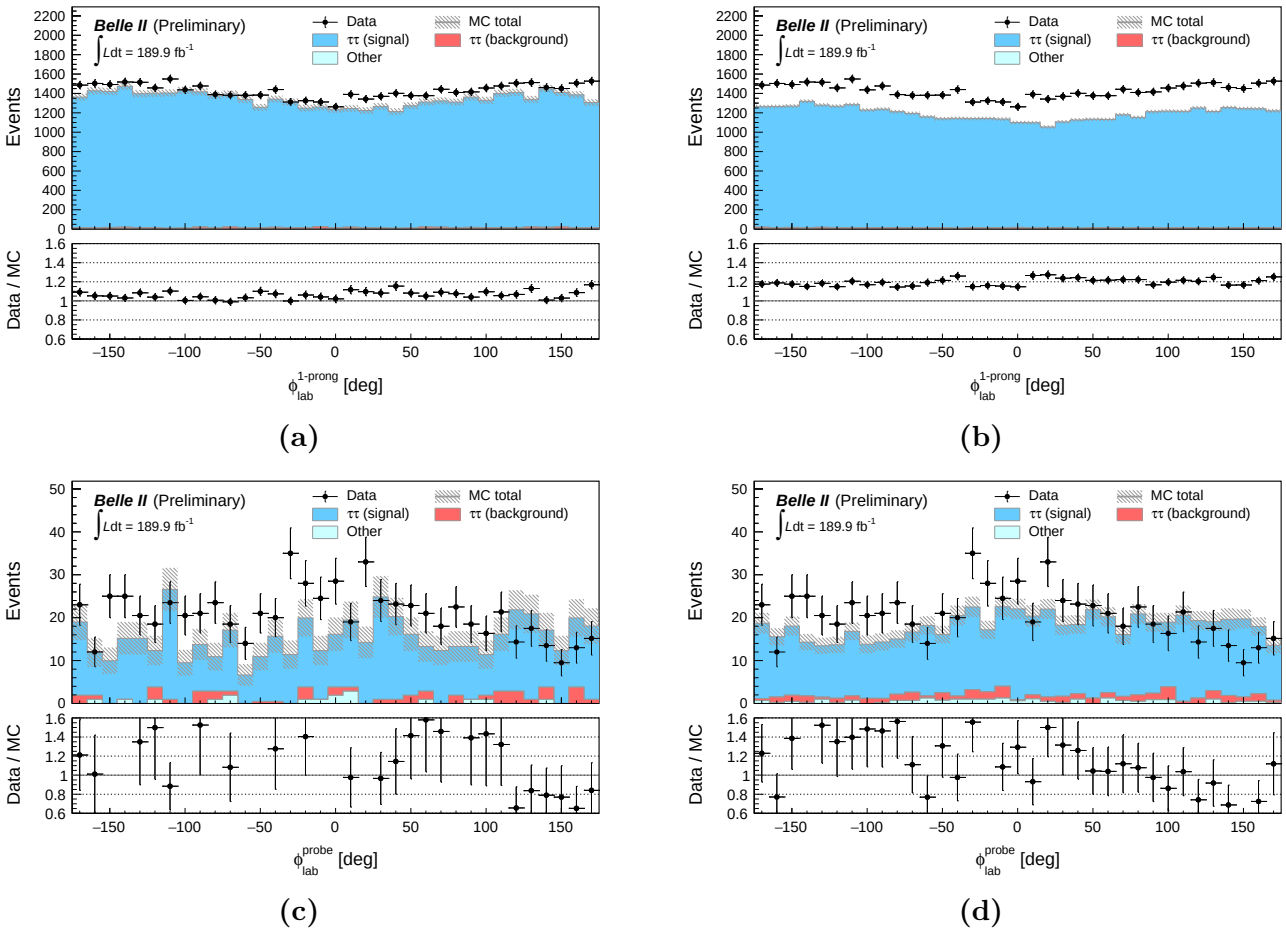


Figure 3.3: Distributions of the 1-prong track (a,b) and probe track (c,d) ϕ_{lab}, for the combined 4- and 5-track samples (a,b) or 5-track sample only (c,d), using MCri (a,c) or MCrd (b,d) simulation. The data (dots) are compared to the signal + background expectation scaled to the integrated data luminosity (stacked histograms). Statistical uncertainties for data and MC are shown, with the MC error band including also the trigger efficiency uncertainty. The lower panel shows the data/MC ratio and the error bars take into account contributions from both data and MC.

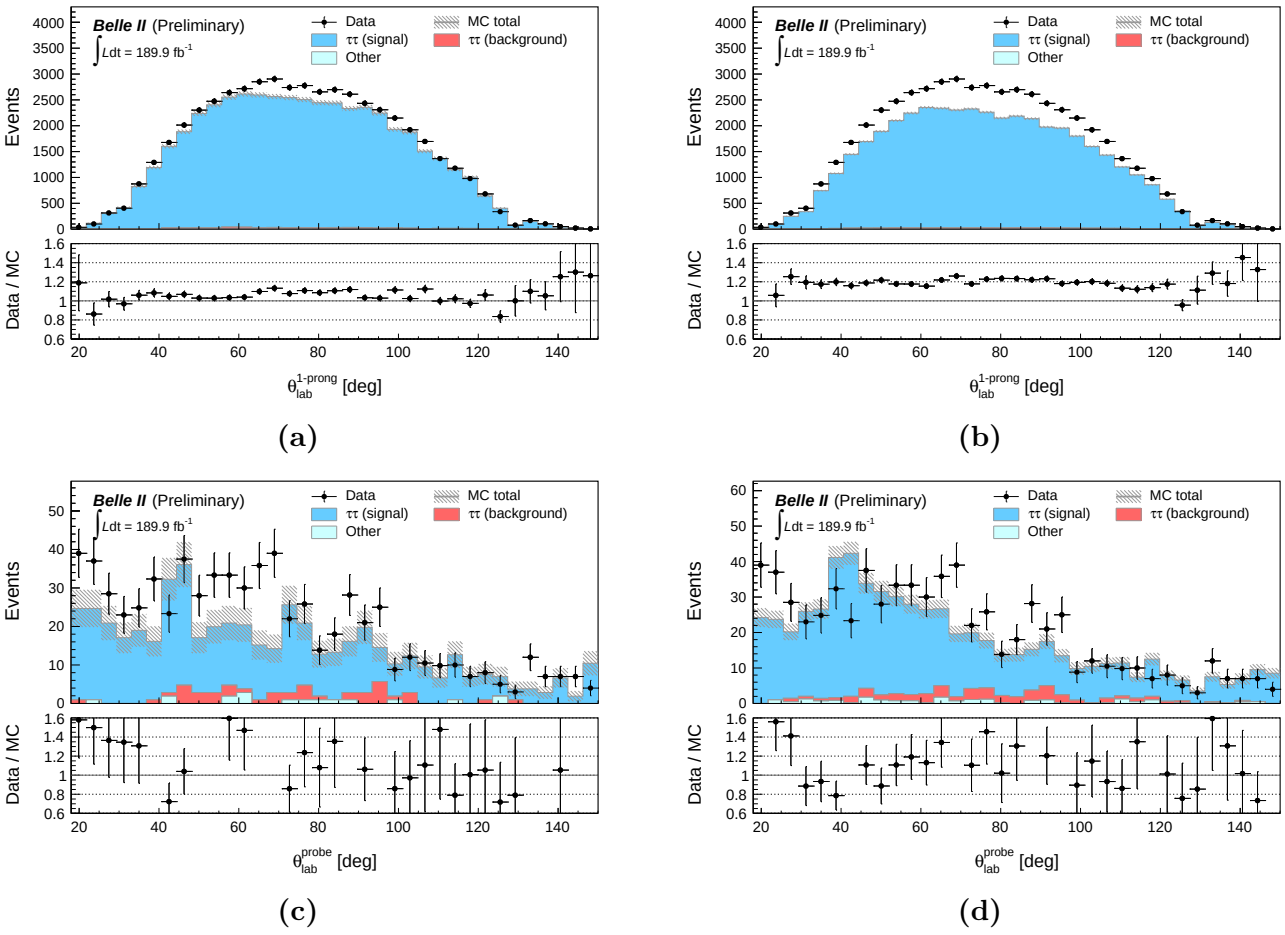


Figure 3.4: Distributions of the 1-prong track (a,b) and probe track (c,d) θ_{lab}, for the combined 4- and 5-track samples (a,b) or 5-track sample only (c,d), using MCri (a,c) or MCrd (b,d) simulation. The data (dots) are compared to the signal + background expectation scaled to the integrated data luminosity (stacked histograms). Statistical uncertainties for data and MC are shown, with the MC error band including also the trigger efficiency uncertainty. The lower panel shows the data/MC ratio and the error bars take into account contributions from both data and MC.

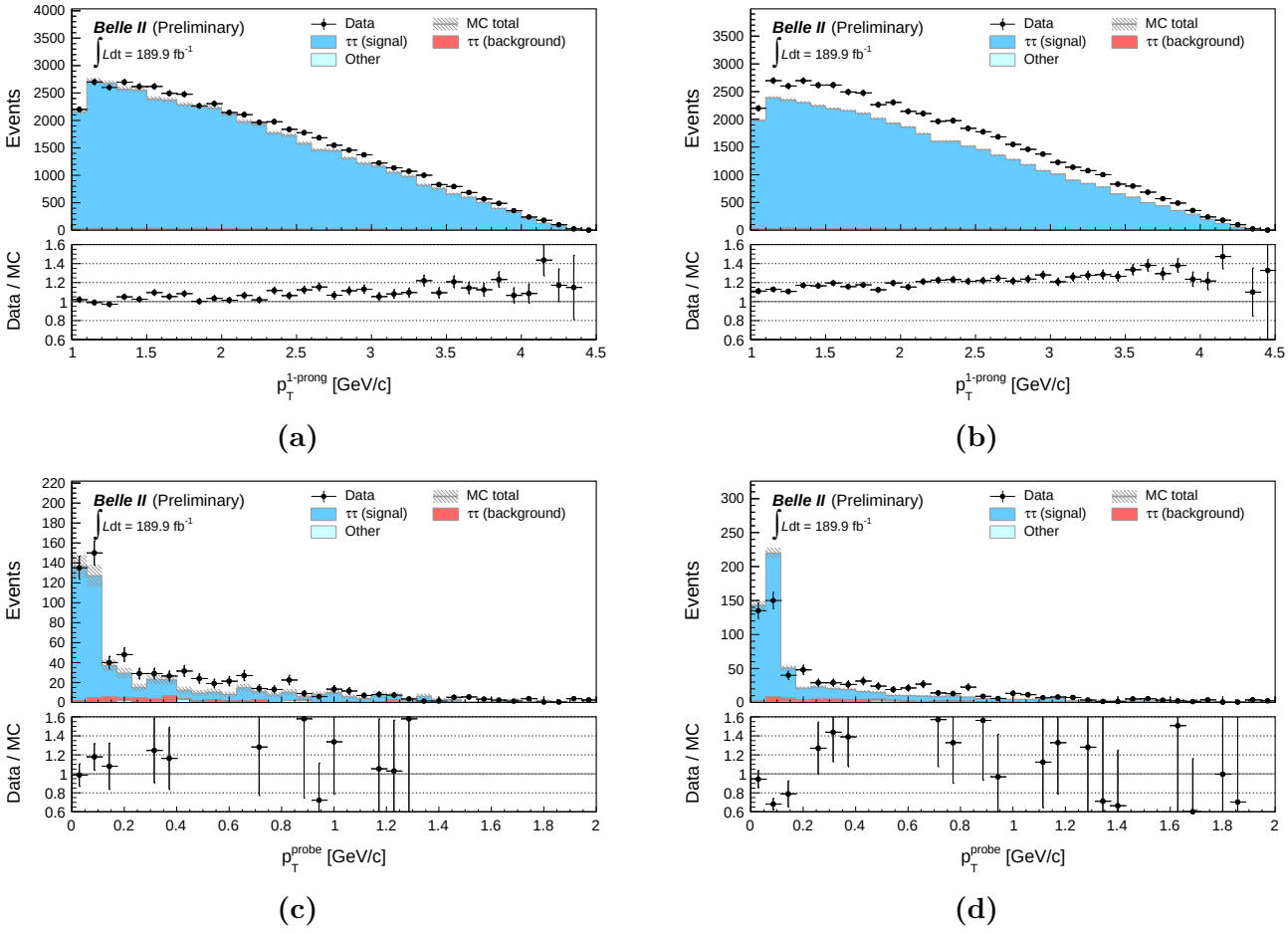


Figure 3.5: Distributions of the 1-prong track (a,b) and probe track (c,d) p_T, for the combined 4- and 5-track samples (a,b) or 5-track sample only (c,d), using MCri (a,c) or MCrd (b,d) simulation. The data (dots) are compared to the signal + background expectation scaled to the integrated data luminosity (stacked histograms). Statistical uncertainties for data and MC are shown, with the MC error band including also the trigger efficiency uncertainty. The lower panel shows the data/MC ratio and the error bars take into account contributions from both data and MC.

3.6. Results

The results of the fake tracks rate measurement in MC and data are presented in Table 3.5, including only statistical uncertainties.

Table 3.5: Fake tracks rates in MCri/MCrd ($r_{\text{fake}}^{\text{MC}}$) and data ($r_{\text{fake}}^{\text{data}}$), computed according to Equation 3.1, using the overall yields from Table 3.3 or the signal yields from Table 3.4. For the data, the signal estimation is done by considering either MCri or MCrd signal purities, therefore two results are given.

Type	MC run dependence	$r_{\text{fake}}^{\text{MC}}$	$r_{\text{fake}}^{\text{data}}$
Overall	MCri	$(1.15 \pm 0.05_{\text{stat}})\%$	$(1.43 \pm 0.05_{\text{stat}})\%$
	MCrd	$(1.51 \pm 0.03_{\text{stat}})\%$	
Signal	MCri	$(1.04 \pm 0.05_{\text{stat}})\%$	$(1.29 \pm 0.05_{\text{stat}})\%$
	MCrd	$(1.37 \pm 0.03_{\text{stat}})\%$	$(1.30 \pm 0.05_{\text{stat}})\%$

As expected, the rate measured in MCrd is closer to the data rate compared to MCri; in fact, they are compatible within one statistical standard deviation, that is why we consider the signal rate in data as the result obtained according to the MCrd signal purity. We estimate the systematic uncertainty on $r_{\text{fake}}^{\text{data}}$ as the largest difference between the overall rate and the signal rates (either from MCrd or MCri) in data: in this case, the final measurement of the fake tracks rate is

$$r_{\text{fake}}^{\text{data}} = (1.30 \pm 0.05_{\text{stat}} \pm 0.14_{\text{sys}})\,\%. \tag{3.3}$$

This result may serve as an indicator of the tracking performance, which depend on the background conditions that evolve with the instantaneous luminosity increase. We show in Figure 3.6 the evolution of the signal fake tracks rate in data as a function of the data taking period. The rate worsens (gets higher) within the more recent datasets, which is a consequence of the instantaneous luminosity increase thus higher beam-background levels. Dealing with the higher contamination by fake tracks in events might become a major challenge for Belle II in the future, as the current record luminosity of SuperKEKB, $\mathcal{L} \simeq 4.7{\times}10^{34}\ \text{cm}^{-2}\,\text{s}^{-1}$, is more than one order of magnitude smaller than the target peak luminosity.

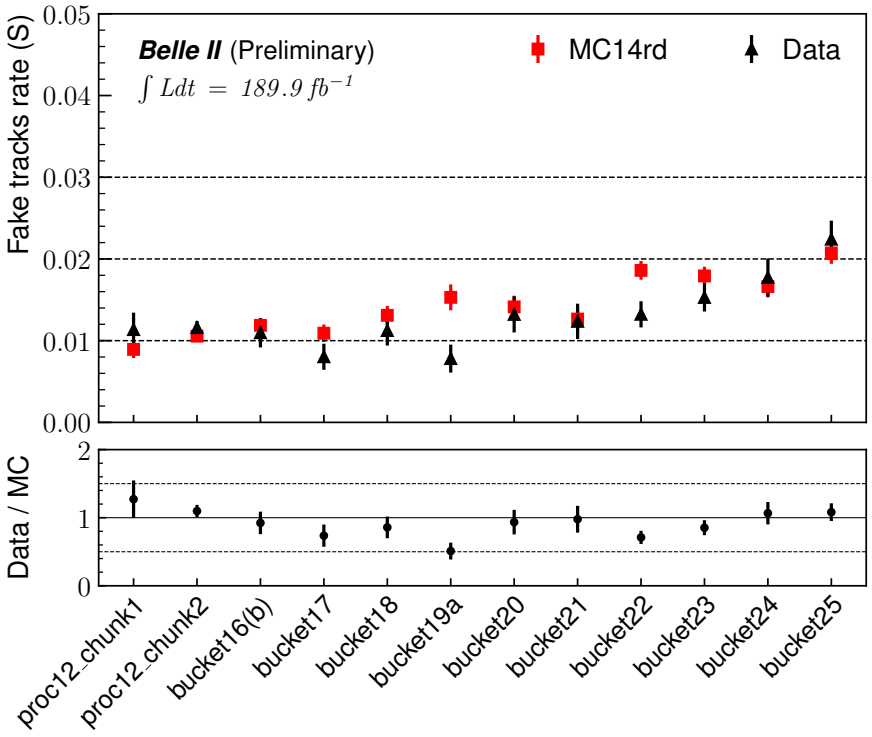


Figure 3.6: Evolution of the signal (S) fake tracks rate in data as a function of the data taking period. The meaning of the time labels is given in Table 2.2. In particular, *proc12_chunk1* refers to the experiments 7, 8 and 10, while *proc12_chunk2* refers to experiment 12.

4. Search for $\tau \to \ell\phi$ lepton flavour violating decays

The upper limits previously obtained by the BaBar and Belle experiments [7, 8] on the branching fractions of $\tau \to \ell\phi$ decays, using respectively 451 and 854 fb^{-1} of data, are presented in Table 4.1. Both analyses relied on a full reconstruction of $e^+e^- \to \tau^+\tau^-$ events, requiring one tau to decay into the searched LFV mode and the oppositely charged one into a single charged particle (1-prong tagging).

Table 4.1: Expected (exp.) and observed (obs.) upper limits on $\tau \to \ell\phi$ branching fractions at 90% confidence level, obtained by BaBar (451 fb^{-1}) and Belle (854 fb^{-1}) [7, 8].

Experiment	$\mathcal{B}^{90}_{\mathrm{UL}}(e\phi)$ ($\times 10^{-8}$) exp. / obs.	$\mathcal{B}^{90}_{\mathrm{UL}}(\mu\phi)$ ($\times 10^{-8}$) exp. / obs.
BaBar	5.0 / 3.1	8.2 / 19
Belle	4.3 / 3.1	4.9 / 8.4

Here, we report the search for the $\tau \to \ell\phi$ decay based on an untagged (or inclusive) selection.

4.1. Analysis Strategy

The untagged selection consists in reconstructing only the signal τ decay, applying as a first step a loose requirement on the number of tracks per event, whereas the previous Belle and BaBar analyses used 3×1-prong events where the opposite charge τ lepton decays into 1-prong.

In the inclusive reconstruction, removing the requirement of a tag track in a hemisphere opposite to the signal tracks and a number of good tracks (see Section 4.3.1) equal to 4 increases the signal efficiency by 32%, with respect to a 3×1-prong selection performed on Belle II simulation. This gain is coming from the addition of 3×3-prong events where the tag τ decays into 3-prong, which corresponds to $\sim$ 15%

of τ decays, and events where either the tag track is not (fully) reconstructed or an additional track (fake track or beam background) is wrongly included, resulting in a total number of good tracks different from 4. The distributions of the number of tracks and good tracks in signal simulation are shown in Figure 4.1.

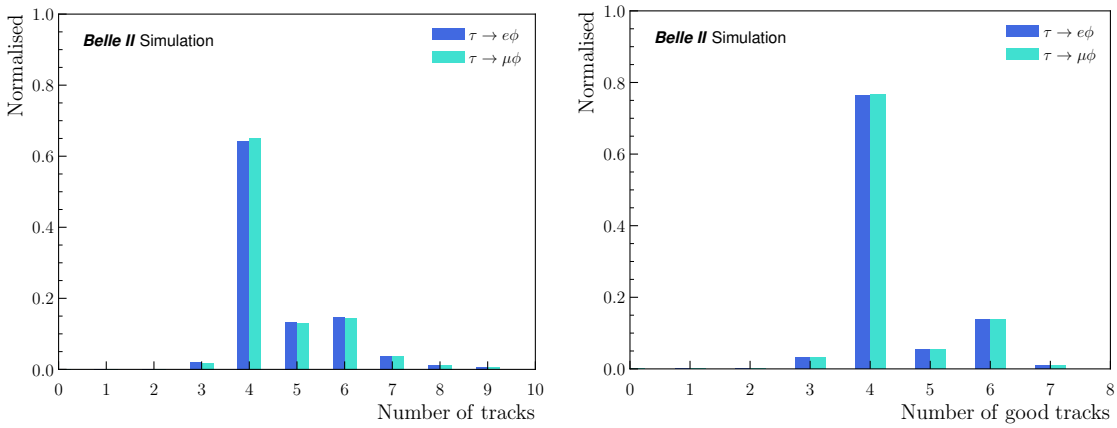


Figure 4.1: Number of tracks (left) and good tracks (right) in the electron and muon signal channels, obtained after the inclusive candidate reconstruction.

The downside of this approach is the larger amount of background events that are retained at the reconstruction level. However, the addition of variables characterising the rest of event, combined with a multivariate selection of signal events using boosted decision trees (BDTs), allows a background reduction to levels comparable to what can be obtained with a 3×1 reconstruction.

We work on data collected by the Belle II experiment, along with signal and background simulation samples produced by the collaboration. Details can be found in Section 4.2.

Just as in the previous analyses, we consider the two-dimensional (M_τ, ΔE_τ) space, where M_τ is the measured mass of the signal tau and ΔE_τ is the difference between the signal tau and beam energies in the center of mass frame. After the reconstruction of events, we use the resolutions on these variables to define several regions, in particular the signal region where the upper limit estimation will later be performed. These steps are described in Section 4.3.

The background suppression strategy is explained in Section 4.4, from which we derive the signal efficiencies and expected background yields. It consists of cut-based and BDT selections, optimised on simulation in order to obtain minimal upper limits. This part includes also the study of control channels, with the aim of defining systematic uncertainties on the selections.

The comparison of data and simulation, after the selections, is shown in Section 4.5. While the signal region in the (M_τ, ΔE_τ) space is blinded, the number of data

events that we expect to find in it is extracted from the sidebands.

Finally, in Section 4.6, we estimate systematic uncertainties on the different quantities that enter the computation of the upper limit, which at 90% confidence level (C.L.) is

$$\mathcal{B}^{90}_{\mathrm{UL}}(\tau \to \ell\phi) = \frac{s^{90}}{2L\sigma_{\tau\tau}\varepsilon_{\ell\phi}}, \tag{4.1}$$

where s^{90} is the upper limit on the signal yield, $L = \int \mathcal{L}dt$ the integrated luminosity of the data we use, $\sigma_{\tau\tau}$ the $\tau^+\tau^-$ production cross section calculated to be 0.919 ± 0.003 nb^{-1} [62] and $\varepsilon_{\ell\phi}$ the signal efficiency. The strategy to measure the limits and the final results are presented in Section 4.7.

4.2. Data and Monte Carlo Simulation Samples

4.2.1. Data

As detailed in Section 2.7 and Table 2.2, the analysis is based on data collected from 2019 to 2021, for a total integrated luminosity of $\int \mathcal{L}dt = 189.88$ fb^{-1}.

4.2.2. Signal simulation

We use the 14$^{\text{th}}$ official run-independent Monte Carlo (MC) production campaign for the simulation samples of the $\tau^\pm \to e^\pm\phi$ and $\tau^\pm \to \mu^\pm\phi$ decays, generated with `TauolaBelle` [58, 63]. Signal samples (produced for each positive and negative charges) correspond to one million of simulated signal events.

4.2.3. Background simulation

To study the event selection and background suppression, we use run-independent MC samples from the same production campaign and conditions as the signal simulation, which are described in Table 4.2. They include "generic" ($\tau^+\tau^-$, $q\bar{q}$, $B\bar{B}$) and "low-multiplicity" ($\ell^+\ell^-$ and $\ell^+\ell^-X^+X^-$, where $\ell = e, \mu$ and $X = e, \mu, \tau, \pi, K, p$) final states.

Since the background suppression strategy relies on the training of BDTs, mainly targeting the $q\bar{q}$ and the $\tau^+\tau^-$ background events, additional 5 ab^{-1} of skimmed simulation were produced for those background sources using the TauLFV skim (see definition in Section 4.3.6).

The `TauolaBelle` generator, used to produce the τ-pair simulation, does not include a τ decay potentially dangerous for this analysis, $\tau^\pm \to \phi K^\pm\nu_\tau$ which has a branching fraction of $(4.4 \pm 1.6) \times 10^{-5}$ [10] and could contribute to the background. In `TauolaBelle`, $\tau^\pm \to K^+K^-K^\pm\nu_\tau$ is only simulated according to a phase space distribution and does not account for the ϕ meson resonance. We verified the reconstruction efficiency of the resonant background in a τ-pair sample produced

with `TauolaBelle2`, a more recent generator which simulates this decay correctly. This sample, equivalent to 1 ab^{-1}, was skimmed to select only 3×1-prong decays. We find no remaining $\tau^\pm \to [\phi \to K^+K^-]\, K^\pm\nu_\tau$ events after our selections and conclude that this contribution can be ignored.

Table 4.2: List and integrated luminosities of the background simulation samples.

Background decay modes	Sample name	Luminosity per batch (unskimmed/skimmed)	Integrated luminosity
$\tau^+\tau^-$	`taupair`	100 fb^{-1}/1000 fb^{-1}	6000 fb^{-1}
$u\bar{u}$	`uubar`	100 fb^{-1}/1000 fb^{-1}	6000 fb^{-1}
$d\bar{d}$	`ddbar`	100 fb^{-1}/1000 fb^{-1}	6000 fb^{-1}
$s\bar{s}$	`ssbar`	100 fb^{-1}/1000 fb^{-1}	6000 fb^{-1}
$c\bar{c}$	`ccbar`	100 fb^{-1}/1000 fb^{-1}	6000 fb^{-1}
$B\bar{B}$	`mixed`, `charged`	100 fb^{-1}	1000 fb^{-1}
$e^+e^-\gamma$	`ee`	10 fb^{-1}	100 fb^{-1}
$\mu^+\mu^-$	`mumu`	100 fb^{-1}	1000 fb^{-1}
$e^+e^-e^+e^-$	`eeee`	100 fb^{-1}	200 fb^{-1}
$e^+e^-\mu^+\mu^-$	`eemumu`	100 fb^{-1}	200 fb^{-1}
$e^+e^-\pi^+\pi^-$	`eepipi`	100 fb^{-1}	1000 fb^{-1}
$e^+e^-K^+K^-$	`eeKK`	1000 fb^{-1}	1000 fb^{-1}
$e^+e^-p^+p^-$	`eepp`	1000 fb^{-1}	1000 fb^{-1}
$\mu^+\mu^-\mu^+\mu^-$	`mumumumu`	2000 fb^{-1}	2000 fb^{-1}
$e^+e^-\tau^+\tau^-$	`eetautau`	1000 fb^{-1}	2000 fb^{-1}
$\mu^+\mu^-\tau^+\tau^-$	`mumutautau`	2000 fb^{-1}	2000 fb^{-1}

4.3. Event reconstruction and ntuple production

The reconstruction of the decay has been performed with the Belle II Analysis Software `basf2`, release `light-2201-venus`.

4.3.1. Track and neutral particle selections

Photons and charged tracks are selected according to a few criteria listed in Tables 4.3 and 4.4, respectively. Photons are divided into three categories, depending on whether they are produced from π^0 decays, are emitted by electrons as Bremsstrahlung or are used to determine the shape and kinematics of the event.

These selections involve some variables that are defined as:

- `clusterHasNPhotons`: 1 if the cluster has the "N photons" hypothesis, 0 if not;
- `clusterNHits`: sum of weights of all crystals in an ECL cluster;
- `electronID_noSVD_noTOP`: electron identification probability, excluding information from SVD and TOP;
- `muonID_noSVD`: muon identification probability, excluding information from SVD.

Tracks are called "good" when they originate from around the interaction point: $|z_0| < 3$ cm and $|d_0| < 1$ cm.

Neutral pions in the event are also reconstructed for the purpose of removing background, selecting two photons from the first γ list in Table 4.3 and combining them with the following condition on the reconstructed π^0 mass: $0.115 < M_{\pi^0} < 0.152$ GeV/c^2.

Table 4.3: Photon selection criteria.

Variable	γ from π^0	γ from Bremsstrahlung	γ for event shape
clusterHasNPhotons	1	1	1
E [MeV]	> 100	> 20	> 200
$\cos\theta$	$[-0.8660, 0.9563]$	$[-0.8660, 0.9563]$	$[-0.8660, 0.9563]$
clusterNHits	> 1.5	> 1.5	> 1.5
comes from π^0	yes	no	no

4.3.2. Bremsstrahlung corrections

Because of final state radiations corresponding to Bremssthralhung, the kinematics of the lepton in the $\tau \to e\phi$ decay mode is corrected using a function developed by

Table 4.4: Charged track selection criteria.

Variable	Electron	Muon	Kaon	Pion
$\lvert z_0 \rvert$ [cm]	< 3	< 3	< 3	< 3
$\lvert d_0 \rvert$ [cm]	< 1	< 1	< 1	< 1
electronID$^{\text{noSVD, noTOP}}$	> 0.5	–	–	–
muonID$^{\text{noSVD}}$	< 0.9	> 0.9	–	< 0.9
in CDC acceptance	–	–	yes	–

the Belle experiment. The correction consists in adding to the four-momentum of the electron the ones of all photons within a cone around the track, selected from the second γ list in Table 4.3.

The cone is defined by an angle threshold, the maximum angle in radians between the electron and the photons, while another parameter of the function is the minimum energy that the photons must have in order to enter the correction.

Both parameters have been optimised by minimising the mean and the standard deviation of $(p_{\text{MC}} - p)/p_{\text{MC}}$, where p_{MC} and p are respectively the generated and reconstructed momenta of the electron, using signal simulation from the previous run-independent MC production campaign (`MC13`).

The parameters have been scanned by testing a few values: 0.05, 0.1, 0.15 and 0.2 rad for the angle; 20, 50, 100, 150 and 200 MeV for the energy. The means and standard deviations are shown in Figure 4.2. As a result, the maximum angle of the cone was set at 0.15 rad, while the energy of the photons must be at least 20 MeV. The distributions of signal events in the M_τ and ΔE_τ plane before and after the corrections are displayed in Figure 4.3. Most of the events in the diagonal tail are moved to the peak, increasing the signal efficiency by approximately 4.3% in the $1.6 < M_\tau < 1.9$ GeV/c^2 and $-0.6 < \Delta E_\tau < 0.4$ GeV region.

4.3.3. Shape and kinematics of the event

Good tracks and photons selected from the first and third lists defined in Table 4.3 are used to build the event's shape and kinematics.

By building the shape, we obtain variables related to the thrust axis of the event (see Equation 3.2) and the CLEO cones [64]. By building the kinematics, we obtain the total visible energy as well as information on the missing momentum of the event.

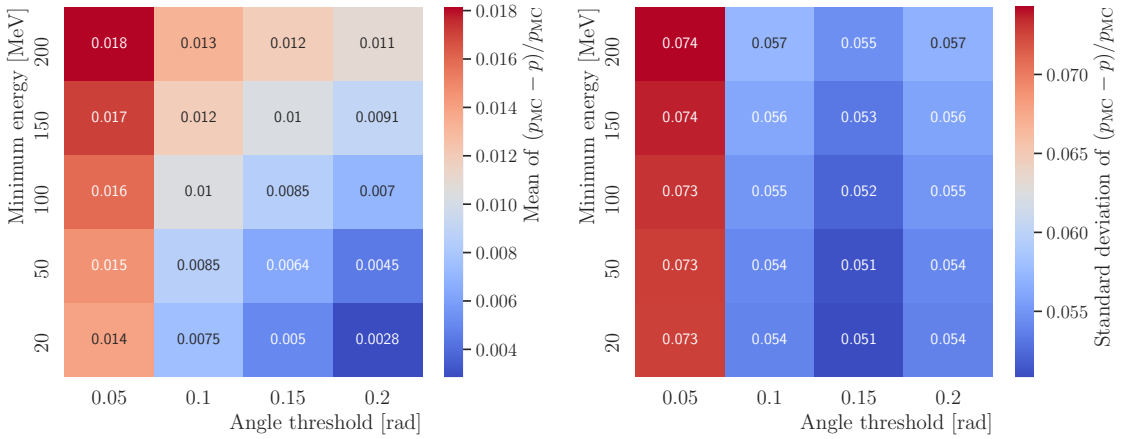


Figure 4.2: Mean (left) and standard deviation (right) of $(p_{\text{MC}} - p)/p_{\text{MC}}$, where p_{MC} and p are respectively the generated and reconstructed momenta of the electrons to be corrected against Bremsstrahlung, as a function of the maximum angle of the cone and the minimum energy of photons. Run-independent `MC13` signal simulation is used.

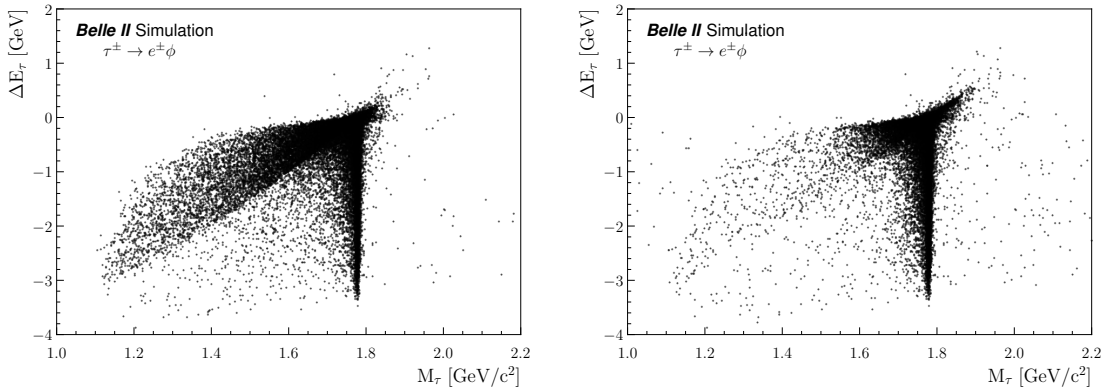


Figure 4.3: Scatter plots of $\tau \to e\phi$ events in the M_τ and ΔE_τ plane, before (left) and after (right) applying Bremsstrahlung corrections with optimised parameters. Run-independent `MC13` signal simulation is used.

4.3.4. Signal reconstruction

The signal events are reconstructed by combining two oppositely charged kaons and a lepton, requiring that they belong to the same event hemisphere defined by a plane perpendicular to the thrust axis. In particular, we ask that:

- the number of good tracks is at most 10;

- $0.95 < M_\phi < 1.30$ GeV/c^2;
- $1.4 < M_\tau < 2.1$ GeV/c^2 and $-0.8 < \Delta E_\tau < 0.8$ GeV for the electron decay mode;
- $1.5 < M_\tau < 2.0$ GeV/c^2 and $-0.5 < \Delta E_\tau < 0.5$ GeV for the muon decay mode.

τ and ϕ vertices are also fitted using the `TreeFitter` [65] tool, which updates the momenta of the mother particles as well as the vertex positions of all tracks concerned. The accuracy of the fit is evaluated thanks to a χ^2 test, which can be used as a variable to discriminate between signal and background events.

4.3.5. Rest of event

All the tracks and clusters that are not used in the signal reconstruction form the rest of event (ROE). The properties of the ROE are stored in variables that we harness for background suppression.

Here, a "mask" is applied to the ROE particles by requiring that:

- the tracks are in the CDC acceptance, with $p^{\mathrm{T}} > 0.05$ GeV/c, $|z_0| < 10$ cm and $|d_0| < 5$ cm;
- the ECL clusters are in the CDC acceptance, with $E > 0.05$ GeV/c.

4.3.6. Offline requirements

After the reconstruction of events, an additional selection on the ϕ meson's mass, a skimming and trigger requirements are applied.

4.3.6.1. ϕ mass

We ask for the ϕ meson's mass M_ϕ to be inside the range $[1.014, 1.024]$ GeV/c^2, a selection shown in Figure 4.4 and which was optimised against background according to the Punzi figure of merit [66]

$$\mathrm{FOM} = \frac{\varepsilon}{a/2 + \sqrt{B}}, \tag{4.2}$$

where ε is the signal efficiency, B the number of expected background events and $a = 3$ is the desired confidence level in units of a Gaussian distribution's standard deviation.

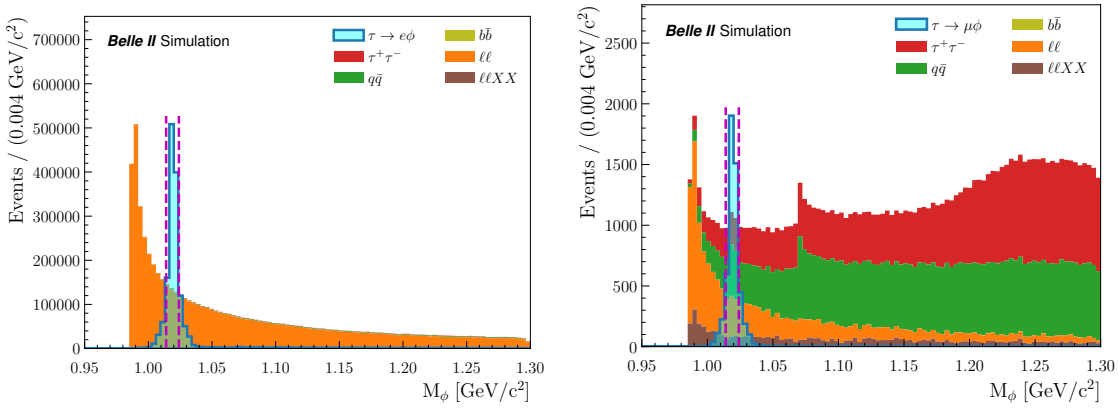


Figure 4.4: Distributions of the ϕ meson's mass M_ϕ, after the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) events reconstructions. The vertical dashed lines show the edges of the region selected as a result of the Punzi figure of merit optimisation.

4.3.6.2. Skimming

Events must pass the TauLFV skim: the number of good tracks is at most 6, $1.4 < M_\tau < 2.0$ GeV/c^2 and $-1.0 < \Delta E_\tau < 0.5$ GeV.

Although the selection performed by the TauLFV skim differs from the reconstruction explained in Section 4.3.4, only the requirement on the number of good tracks is truly effective, since the skim is based on a parallel reconstruction of events slightly different from ours. Therefore, an event passing this skim can still lie outside of the selected M_τ and ΔE_τ windows.

4.3.6.3. Trigger

Data and MC events are required to fire at least one of the dedicated low-multiplicity Level1 (L1) triggers *lmlX*, with *X=0, 1, 2, 4, 6, 7 ,8, 9, 10, 12, 13*, based on ECL information. Definitions are given in Appendix A, Table A.1.

The signal efficiencies and background yields after the reconstruction and after the offline requirements are given in Table 4.5.

4.3.7. Signal resolution

As in the previous Belle analysis, the signal is searched in the two-dimensional M_τ and ΔE_τ plane. The signal resolution of those two variables are used to define a signal region, here a box large as three times the resolutions on both sides of the M_τ and ΔE_τ mean values, and to which we refer as the $\pm 3\sigma$ region. This signal region is where the number of observed events is taken from in the computation

Table 4.5: Signal efficiencies and numbers of background candidates after the reconstruction (up to Section 4.3.5) and after the offline requirements. The yields correspond to the integrated luminosities cited in Table 4.2 and don't include any weighting or correction.

Signal and background decay modes	$\tau \to e\phi$ signal efficiency and background yields after reco. / after offline selections	$\tau \to \mu\phi$ signal efficiency and background yields after reco. / after offline selections
Signal	31.1% / 21.2%	25.7% / 18.5%
τ-pair	5268681 / 104766	2453626 / 34613
$q\bar{q}$	10524606 / 218416	2141517 / 45743
$B\bar{B}$	43551 / 119	2842 / 6
e^+e^-	7101298 / 286256	372 / 12
$\mu^+\mu^-$	50526 / 1138	105920 / 6416
$e^+e^-\ell^+\ell^-$	19957 / 178	6941 / 309
$e^+e^-h^+h^-$	246 / 0	22 / 0
$\mu^+\mu^-\mu^+\mu^-$	63 / 2	12928 / 635
$\ell^+\ell^-\tau^+\tau^-$	12184 / 269	1708 / 36

of upper limits, therefore blinded in data. Moreover, the $\pm 20\sigma$ region is used to optimise the background rejection and signal selection using simulation.

The resolution is obtained from an unbinned fit to the simulated signal using the sum of a Crystal Ball and two Gaussian probability density functions. The total resolution σ is then computed as the weighted sum of the three components' standard deviations. The fit results are shown in Figures 4.5, 4.6 and 4.7, and given in Table 4.6, which helps us define the different signal regions used throughout the analysis.

From there, we constrain ourselves to the 20σ region cited in Table 4.6. However, because of the selection applied on ΔE_τ at the event reconstruction level, this region is not exactly twenty times the resolution, albeit close enough.

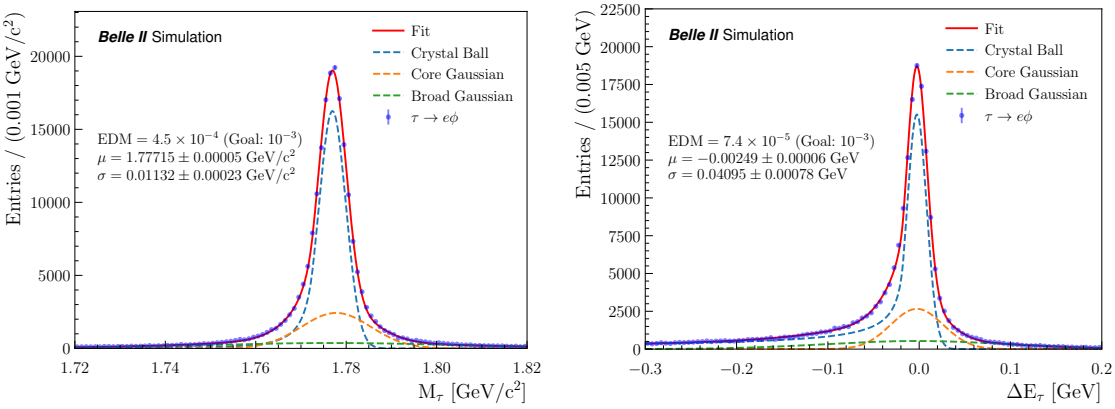


Figure 4.5: Fits of the M_τ (left) and ΔE_τ (right) distributions using $\tau \to e\phi$ signal simulation. The fit curve is represented with a solid line, while the different components of the fits are shown with dashed lines. The estimated vertical distance to the minimum (EDM), the mean (μ) and the standard deviation (σ) of the fit are also indicated.

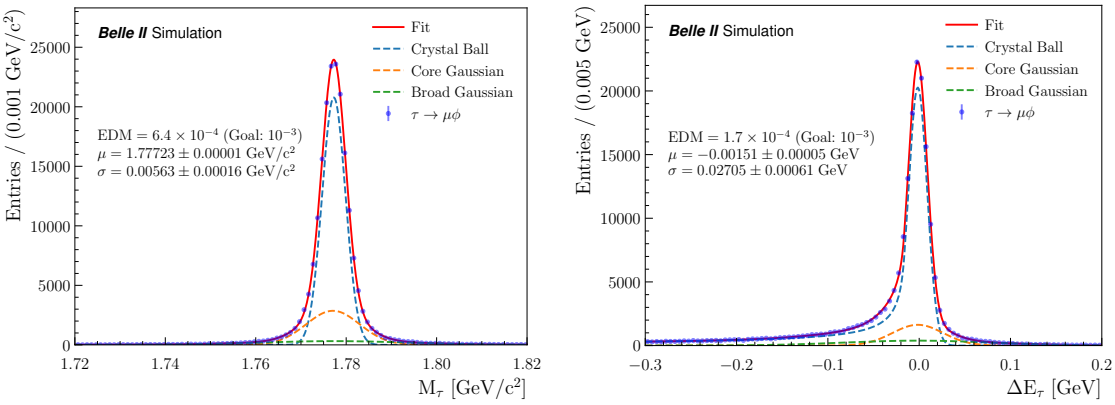


Figure 4.6: Fits of the M_τ (left) and ΔE_τ (right) distributions using $\tau \to \mu\phi$ signal simulation. The fit curve is represented with a solid line, while the different components of the fits are shown with dashed lines. The estimated vertical distance to the minimum (EDM), the mean (μ) and the standard deviation (σ) of the fit are also indicated.

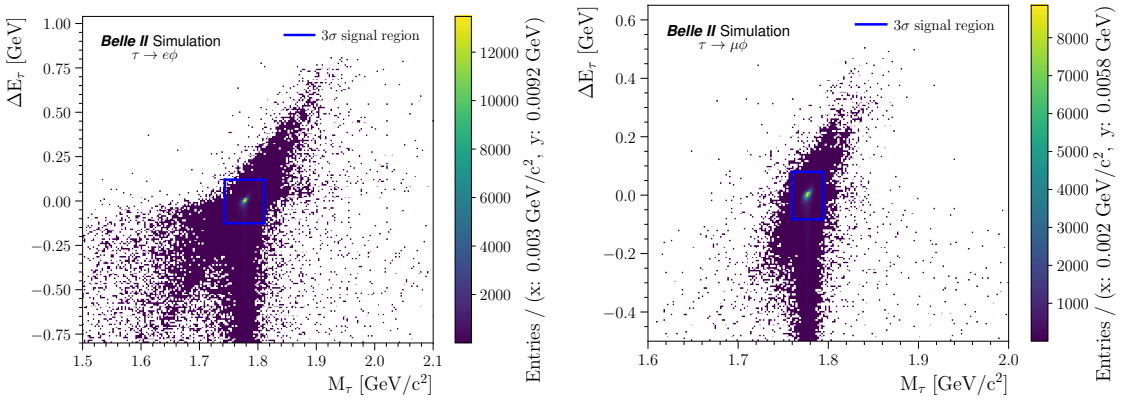


Figure 4.7: Distributions of $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) events in the M_τ and ΔE_τ plane. The edges of the $\pm 3\sigma$ signal regions defined from the previous fits are marked as blue rectangles.

Table 4.6: Signal regions and resolutions obtained from a fit to simulated signal.

Mode	Variable	3σ region	5σ region	20σ region	Resolution
$e\phi$	M_τ [MeV/c^2]	$[1.743, 1.811]$	$[1.721, 1.834]$	$[1.551, 2.004]$	11.3 ± 0.2
	ΔE_τ [MeV]	$[-0.125, 0.120]$	$[-0.207, 0.202]$	$[-0.8, 0.8]$	41.0 ± 0.8
$\mu\phi$	M_τ [MeV/c^2]	$[1.760, 1.794]$	$[1.749, 1.805]$	$[1.665, 1.890]$	5.6 ± 0.2
	ΔE_τ [MeV]	$[-0.083, 0.080]$	$[-0.137, 0.134]$	$[-0.5, 0.5]$	27.1 ± 0.6

4.4. Background suppression

Background events are first suppressed by a cut-based preselection aiming at removing, as much as possible, the most important contributions while maintaining a high signal efficiency. Secondly, a BDT that combines signal and background discriminating variables is trained on simulated events.

4.4.1. Discriminating variables

The discriminating variables considered for the background suppression are listed in Tables 4.7 and 4.8.

Among those variables, some are in the end not included in the BDT training:

- CLEO thrust 1, 2 and CLEO collision 0, 1, which are not discriminating enough to improve the BDT performances;
- the ROE thrust axis magnitude, which doesn't show a good data-MC agreement in the $\tau \to 3\pi\nu_\tau$ control channel and cannot be corrected without worsening the data-MC agreement of more important variables, as explained in Section 4.6.6.2;
- the τ thrust axis magnitude, which is found to be correlated to M_τ and is removed in order not to sculpt a peaking shape in the signal region;
- the kaonID of the leading kaon, which cannot be used without data-MC corrections provided by the collaboration for specific selection values, thus only included in the cut-based preselection;
- the number of neutral hadrons in masked ROE, because of large discrepancies observed between data and MC for background events in the sidebands;
- in the muon channel only: $M_{\mathrm{ROE(masked)}}$, $\Delta E_{\mathrm{ROE(masked)}}$ and the number of photons in masked ROE for the same reasons as in the previous bullet.

Table 4.7: List of signal and background discriminating variables related to the lepton, kaons, 3-prong, ϕ, τ, vertexing and neutral particles. Variables with no specific name are called by their definition.

Related to	Name	Definition
Lepton	$E^{\mathrm{CM}}_{\mathrm{lepton}}$	Energy in the center of mass frame (GeV).
	$p^{\mathrm{T}}_{\mathrm{lepton}}$	Transverse momentum (GeV/c).
	$\theta^{\mathrm{CM}}_{\mathrm{lepton}}$	Polar angle in the center of mass frame (rad).
Kaons	$\theta_{K_1-K_2}$	Angle between the two kaons (rad).
	KaonID of leading kaon	KaonID of the kaon with the highest transverse momentum.
3-prong	$p^{\mathrm{T,CM}}_{\mathrm{lead}}$, $p^{\mathrm{T,CM}}_{\mathrm{sub\text{-}lead}}$, $p^{\mathrm{T,CM}}_{\mathrm{third}}$	Transverse momentum in the center of mass frame of the leading, sub-leading and third 3-prong tracks (GeV/c).
ϕ meson	M_ϕ	Mass (GeV/c^2).
	E_ϕ	Energy (GeV).
	p^{T}_ϕ	Transverse momentum (GeV/c).
τ	–	τ thrust axis magnitude.
	$\cos\theta^{\mathrm{thrust}}_{\tau-z}$	Cosine of angle between τ thrust axis and z axis.
	$\theta_{\tau-\mathrm{closest}}$	Angle between τ and closest track (rad).
	$\cos\theta_{p-\mathrm{vertex}}$	Cosine of the angle between the τ momentum and the τ decay vertex vector with respect to the interaction point.
Vertexing	–	χ^2 probability of ϕ vertex fit result.
	–	χ^2 probability of τ vertex fit result.
Neutral particles	γ multiplicity	Number of photons in event.
	π^0 multiplicity	Number of neutral pions in event.
	γ from π^0 multiplicity	Number of photons produced by π^0 decays in event.

Table 4.8: List of signal and background discriminating variables related to the missing momentum, the event and the rest of event (ROE). Variables with no specific name are called by their definition.

Related to	Name	Definition
Missing momentum	M^2_{miss}	Missing mass squared (GeV^2/c^4).
	$E^{\text{CM}}_{\text{miss}}$	Energy in the center of mass frame (GeV).
	$p^{\text{T}}_{\text{miss}}$	Transverse momentum (GeV/c).
	θ_{miss}	Polar angle (rad).
	$\cos\theta_{\text{lepton}-\text{miss}}$	Cosine of angle between lepton and missing momentum.
Event	–	Number of tracks.
	–	Number of good tracks.
	–	Thrust axis magnitude.
	$E^{\text{CM}}_{\text{vis}}$	Visible energy in the center of mass frame (GeV).
	R_2	Reduced Fox-Wolfram moment $R_2 = H_2/H_0$ [a].
	CLEO thrust $0, 1, 2$	0^{th}, 1^{st}, 2^{nd} order CLEO cone with respect to thrust axis (GeV/c) [67].
	CLEO collision $0, 1, 2$	0^{th}, 1^{st}, 2^{nd} order CLEO cone with respect to collision axis (GeV/c) [67].
Rest Of Event	–	ROE thrust axis magnitude.
	$\cos\theta^{\text{thrust}}_{\tau-\text{ROE}}$	Cosine of angle between τ and ROE thrust axes.
	$M_{\text{ROE(masked)}}$	Mass of the masked ROE (GeV/c^2).
	$E_{\text{ROE(masked)}}$	Energy of the masked ROE (GeV).
	$\Delta E_{\text{ROE(masked)}}$	Difference between the energy of the masked ROE and half the center of mass energy (GeV).
	$p_{\text{ROE(masked)}}$	Momentum of the masked ROE (GeV/c).
	$p^{\text{T}}_{\text{ROE(masked)}}$	Transverse momentum of the masked ROE (GeV/c).
	–	Total charge of masked ROE.
	–	Number of charged particles in masked ROE.
	–	Number of photons in masked ROE.
	–	Number of neutral hadrons in masked ROE.

[a] $H_l = \sum_{i,j} \frac{|\boldsymbol{p}_i||\boldsymbol{p}_j|}{s} P_l(\cos\varphi_{ij})$, where i and j run over the final state particles, $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ are the momentum vectors of particles i and j, φ_{ij} is the angle between them and P_l is the Legendre polynomial of degree l [68].

4.4.2. Cut-based preselection

In order to lower down the low-multiplicity background contributions before the training of the BDT, a prior cut-based selection that we call "preselection" is applied, defined roughly by comparing the signal and background distributions in MC. The preselections on both electron and muon channels, with the corresponding signal retention and background rejection rates, are listed in Table 4.9 and illustrated in Figures 4.8 and 4.9.

Among all the requirements, only the one on the electronID for $\tau \to e\phi$ events is underperforming when applied alone, comparing the signal retention and background rejection rates relative to the initial yields. It is however helpful when combined to all other preselections.

The resulting signal efficiency is mainly lowered because of the kaonID selection. This variable's performance is to be compared to the one of the binary kaonID, which by its own could retain 80% of signal and reject 75% of background events. However, the binary kaonID introduces discrepancies between data and simulation which are not as well corrected as in the case of the kaonID.

Table 4.9: Cut-based preselections, with corresponding signal retention $\varepsilon_{\ell\phi}$ and background rejection r_{bkg} rates, depending on the decay mode. For each preselection, the rates are either relative to the initial yields ("init") or to the yields after all other preselections are applied ("presel"). The total ("tot") is computed after applying the full preselection.

Mode	Preselection	$\varepsilon_{\ell\phi}^{\text{init}}$ / $r_{\text{bkg}}^{\text{init}}$ (%)	$\varepsilon_{\ell\phi}^{\text{presel}}$ / $r_{\text{bkg}}^{\text{presel}}$ (%)
$e\phi$	$\theta_{\tau-\text{closest}} > 0.02$ rad	98.7 / 81.7	98.9 / 53.5
	CLEO thrust $0 < 8.5$ GeV/c	97.2 / 87.1	97.7 / 63.1
	$\theta_{\text{K}_1-\text{K}_2} > 0.01$ rad	99.8 / 6.4	99.8 / 5.5
	$\text{electronID}_{\text{lepton}}^{\text{noSVD, noTOP}} > 0.9$	90.0 / 7.5	91.0 / 59.7
	KaonID of leading kaon > 0.6	67.2 / 93.2	67.8 / 92.2
	Total (full preselection)	$\varepsilon_{\ell\phi}^{tot} = 58.7\%$ / $r_{bkg}^{tot} = 99.7\%$	
$\mu\phi$	Thrust axis magnitude < 0.99	98.1 / 31.6	97.8 / 9.6
	$\text{muonID}_{\text{lepton}}^{\text{noSVD}} > 0.99$	94.9 / 16.2	95.2 / 27.1
	KaonID of leading kaon > 0.6	69.2 / 84.9	69.3 / 80.8
	Total (full preselection)	$\varepsilon_{\ell\phi}^{tot} = 64.5\%$ / $r_{bkg}^{tot} = 89.8\%$	

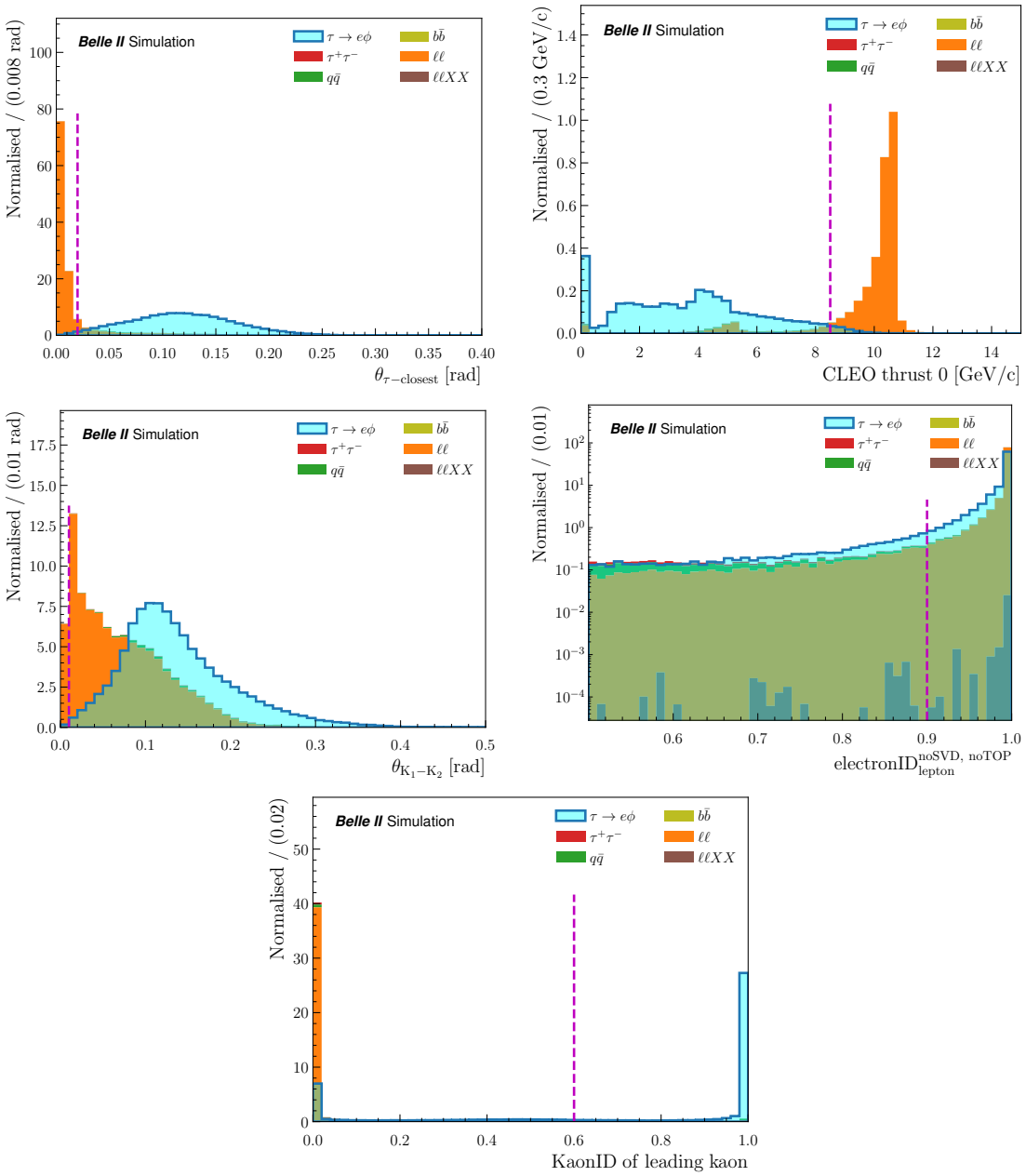


Figure 4.8: Signal and background distributions of $\tau \to e\phi$ events, for variables on which a preselection is performed and before any is applied. The vertical dashed lines indicate the background suppression criteria values.

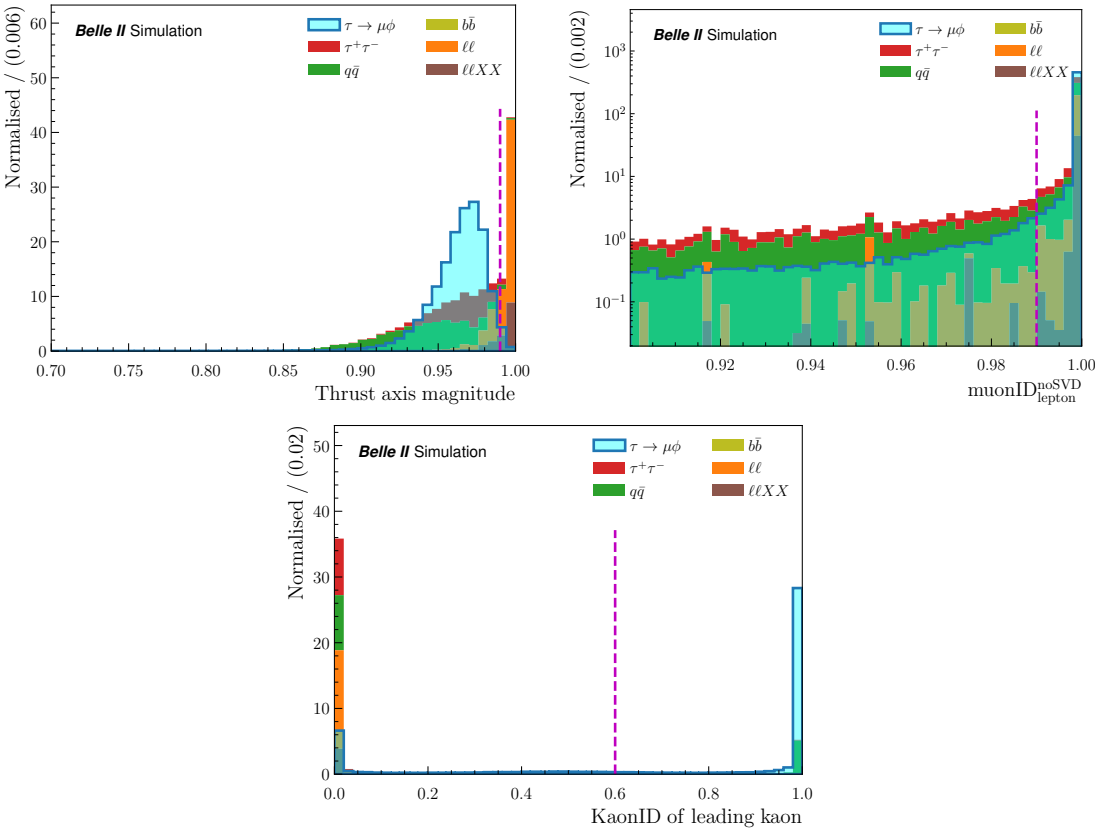


Figure 4.9: Signal and background distributions of $\tau \to \mu\phi$ events, for variables on which a preselection is performed and before any is applied. The vertical dashed lines indicate the background suppression criteria values.

4.4.3. Corrections of the particle identification

As shown in Table 4.9, preselections more stringent than the ones mentioned in Table 4.4 are applied on electron, muon and kaon identification variables, which are known not to be well described in the simulation. For this reason, we assign correction weights to all signal and background simulated events, provided by the collaboration. These particle identification (PID) weights w_{PID} are of the form

$$w_{\text{PID}} = w_{\text{LID}} \times w_{\text{HID}}, \tag{4.3}$$

where:

- w_{LID} are the *lepton identification* (LID) weights assigned to events selected with the electronID or muonID. They are either applied on the lepton selection efficiency if the track is truth-matched to the correct lepton, or on the pion (kaon) fake rate if the track is truth-matched to a pion (kaon);

- w_{HID} are the *hadron identification* (HID) weights assigned to events selected with the kaonID. They are either applied on the kaon selection efficiency if the track is truth-matched to a kaon, or on the pion fake rate if the track is truth-matched to a pion.

Since corrections on $\text{muonID}_{\text{lepton}}^{\text{noSVD}} > 0.99$ are not available at the time this analysis is performed, we use instead weights defined for a lower selection at 0.9.

4.4.4. BDT training and optimisation of the selection

Signal and background samples are split equally into three samples:

- a train sample, on which the BDT training is performed;
- a validation sample, used to validate the performances of the BDT and optimise the selection criterion on the BDT output;
- a test sample, from which the final signal efficiencies and background yields are derived: it isn't involved in any selection definition or optimisation.

Each one of these three samples has the same amount of event candidates as the others for all signal samples (see Section 4.2.2) and background decay modes listed in Table 4.2, therefore the equivalent of 2 ab^{-1} of τ-pair and $q\bar{q}$ background contributions after the preselection. Weights are assigned to events in all three samples, depending on the decay mode they come from, in order to rescale the MC luminosity to 189.88 fb^{-1}, as in the data.

The BDT is built using the XGBoost library [69]. Its output is the probability for an event to be signal, thus has a range between 0 and 1. The training is carried out on the full train sample, taking into account all events inside the 20σ region. The evaluation metric for the classification is the logarithmic loss function.

The validation sample, on the other hand, is split in half:

- one half to check the accuracy of the training, also in the 20σ region,
- the other to optimise, in the 5σ region, the selection criterion on the signal probability that gives the smallest expected upper limit on the branching fraction. The limit computation follows the CLs method implemented in the `pyhf` framework [12], using the asymptotic formula [11] and including only the statistical uncertainties on the background yields. Further details on this computation can be found below, in Section 4.7.

Some BDT parameters are optimised in order to counter overtraining, by keeping the set of values that give the highest area under the receiver operating characteristic (ROC) curve. The parameters, their definitions and the values tested are described in Table 4.10. The number of trees, although set as a constant, is tuned with an "early stopping" functionality that stops the training if the classification performances are not improved in the validation sample after a certain number of iterations.

The signal efficiencies and background yields after the preselection in the three samples are listed in Tables 4.11 and 4.12. The comparison of MC signal and background distributions for variables taken as inputs to the BDT is shown in Figures 4.10 and 4.11.

Table 4.10: List of parameters, their definitions and values tested to reduce the BDT overtraining with XGBoost. Definitions taken from [70].

Name	Definition	Values
`eta`	Step size shrinkage.	0.5
`n_estimators`	Number of gradient boosted trees.	1000
`max_depth`	Maximum tree depth for base learners.	[1,2,3,4,5,6]
`min_child_weight`	Minimum sum of instance weight needed in a child.	[0,2,4,6,8,10]
`gamma`	Minimum loss reduction required to make a further partition on a leaf node of the tree.	[0,1,5]

Table 4.11: $\tau \to e\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection and inside the 20σ and 3σ regions. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	$B\bar{B}$	Low multi.	Total backgr.	FOM
Train	20σ	10.9%	54.41	781.6	5.07	480.93	1322.01	0.003
	3σ	8.5%	2.2	1.95	0	32.31	36.46	0.011
Validation	20σ	10.8%	49.68	796.76	2.96	487.63	1337.03	0.003
	3σ	8.5%	1.04	2.71	0	21.57	25.32	0.013
Test	20σ	10.9%	52.89	796.26	4.14	481.44	1334.73	0.003
	3σ	8.5%	1.95	2.61	0	15.82	20.38	0.014

Table 4.12: $\tau \to \mu\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection and inside the 20σ and 3σ regions. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	$B\bar{B}$	Low multi.	Total backgr.	FOM
Train	20σ	10.3%	13.38	243.76	0.39	3.28	260.81	0.006
	3σ	8.4%	0.22	2.31	0	0	2.53	0.027
Validation	20σ	10.3%	13.51	249.48	0	2.73	265.72	0.006
	3σ	8.4%	0.76	2.18	0	0	2.94	0.026
Test	20σ	10.3%	13.65	247.7	0.42	3.28	265.05	0.006
	3σ	8.4%	0.38	3.56	0	0	3.94	0.024

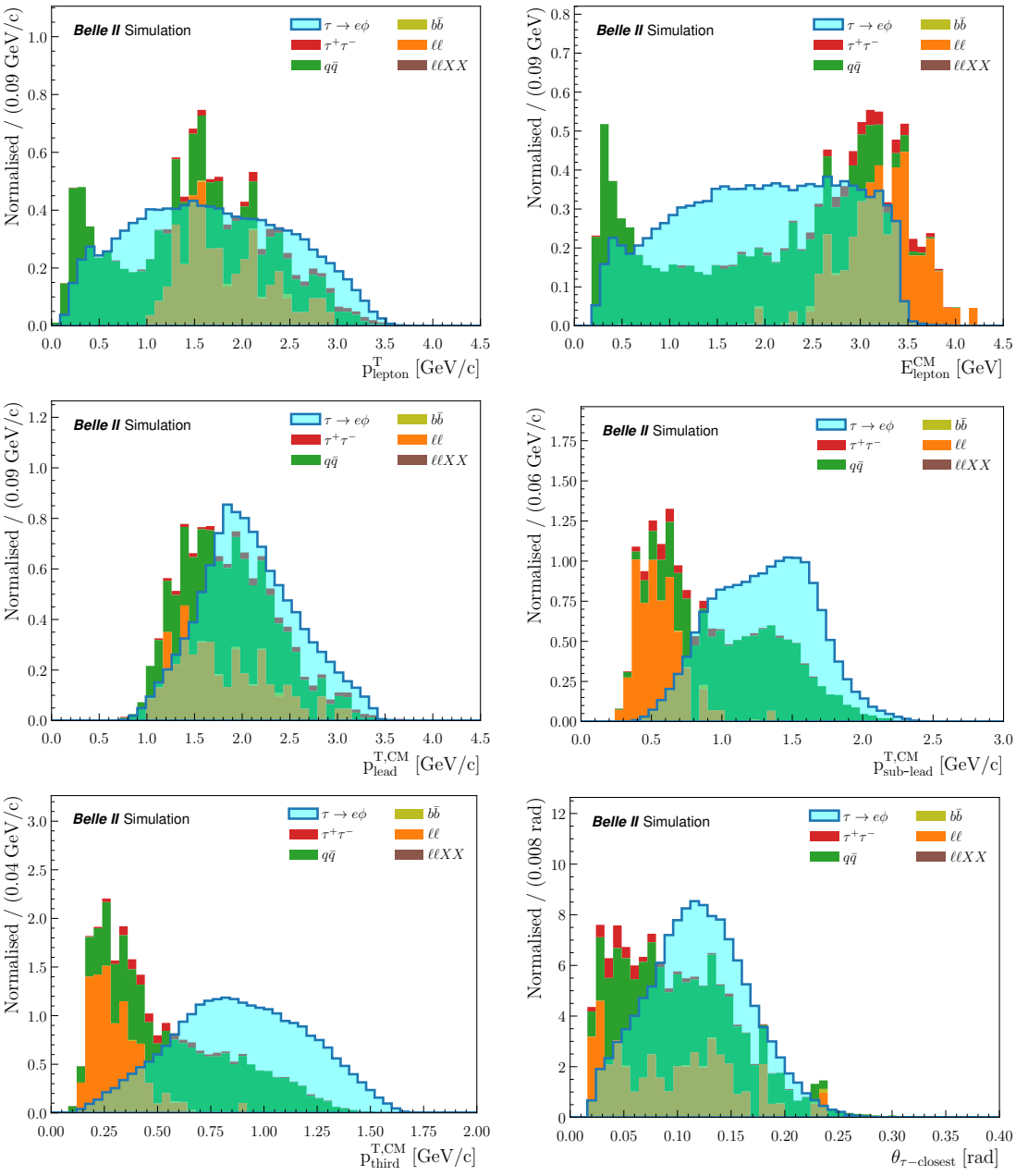


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

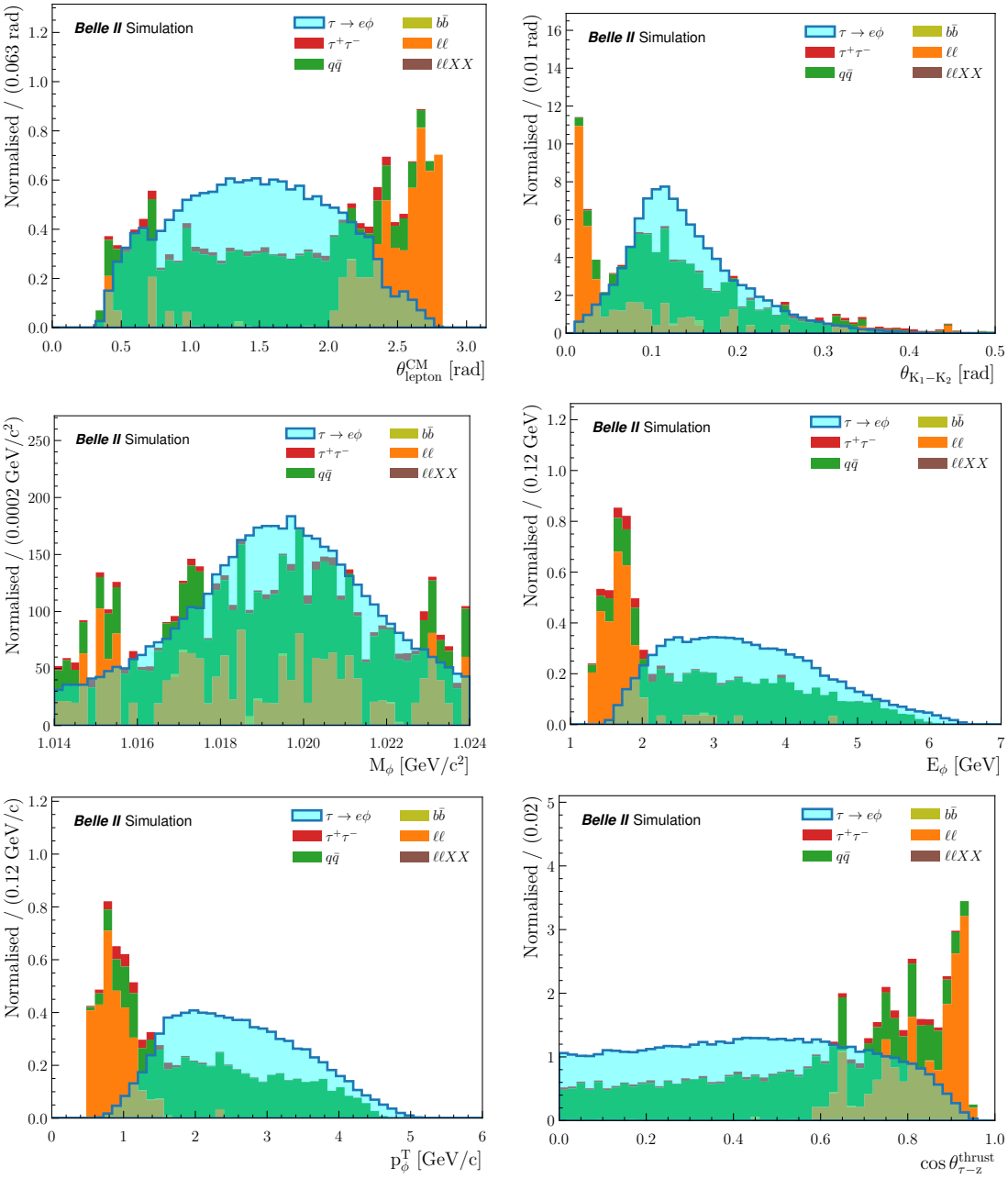


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

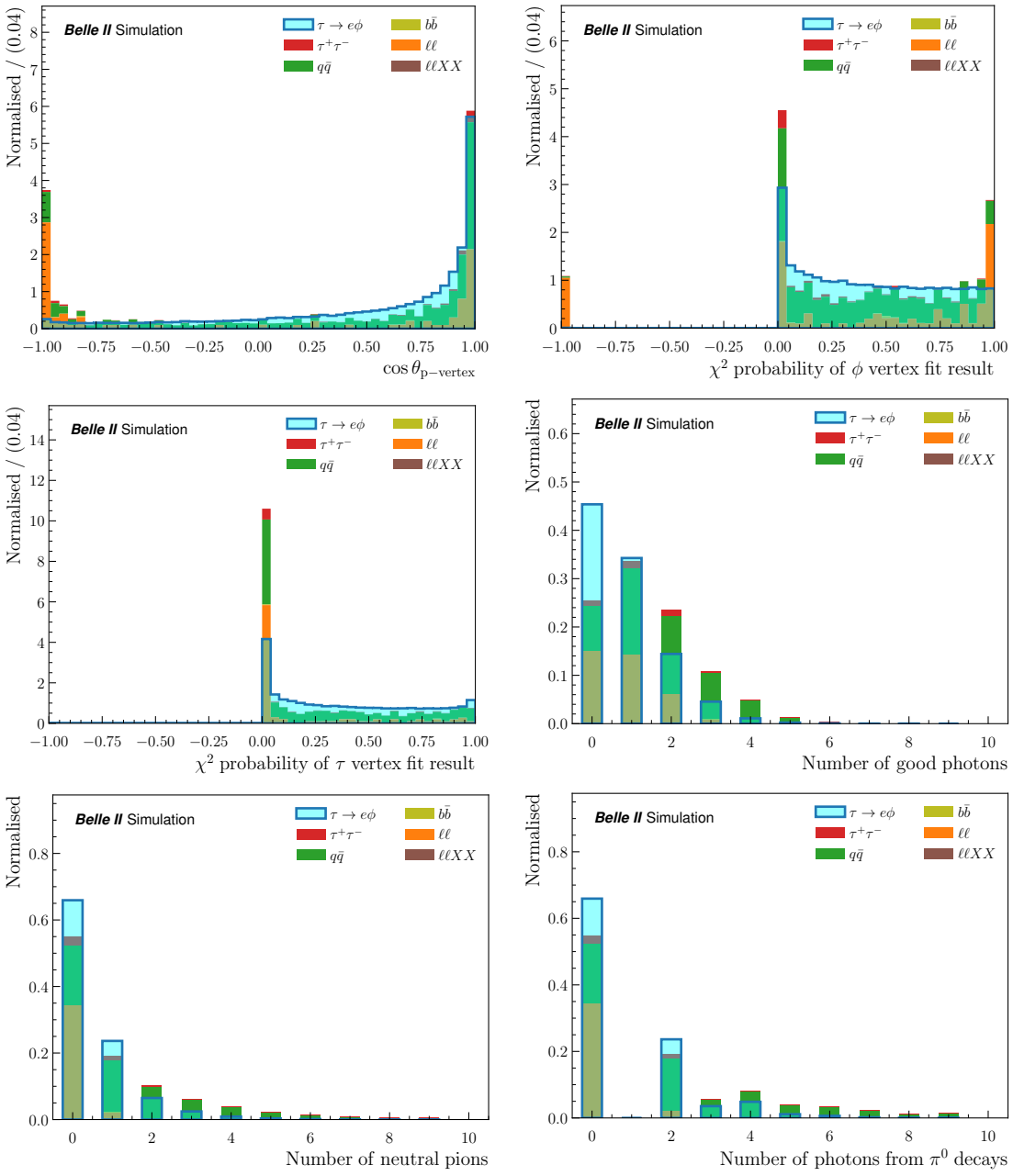


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

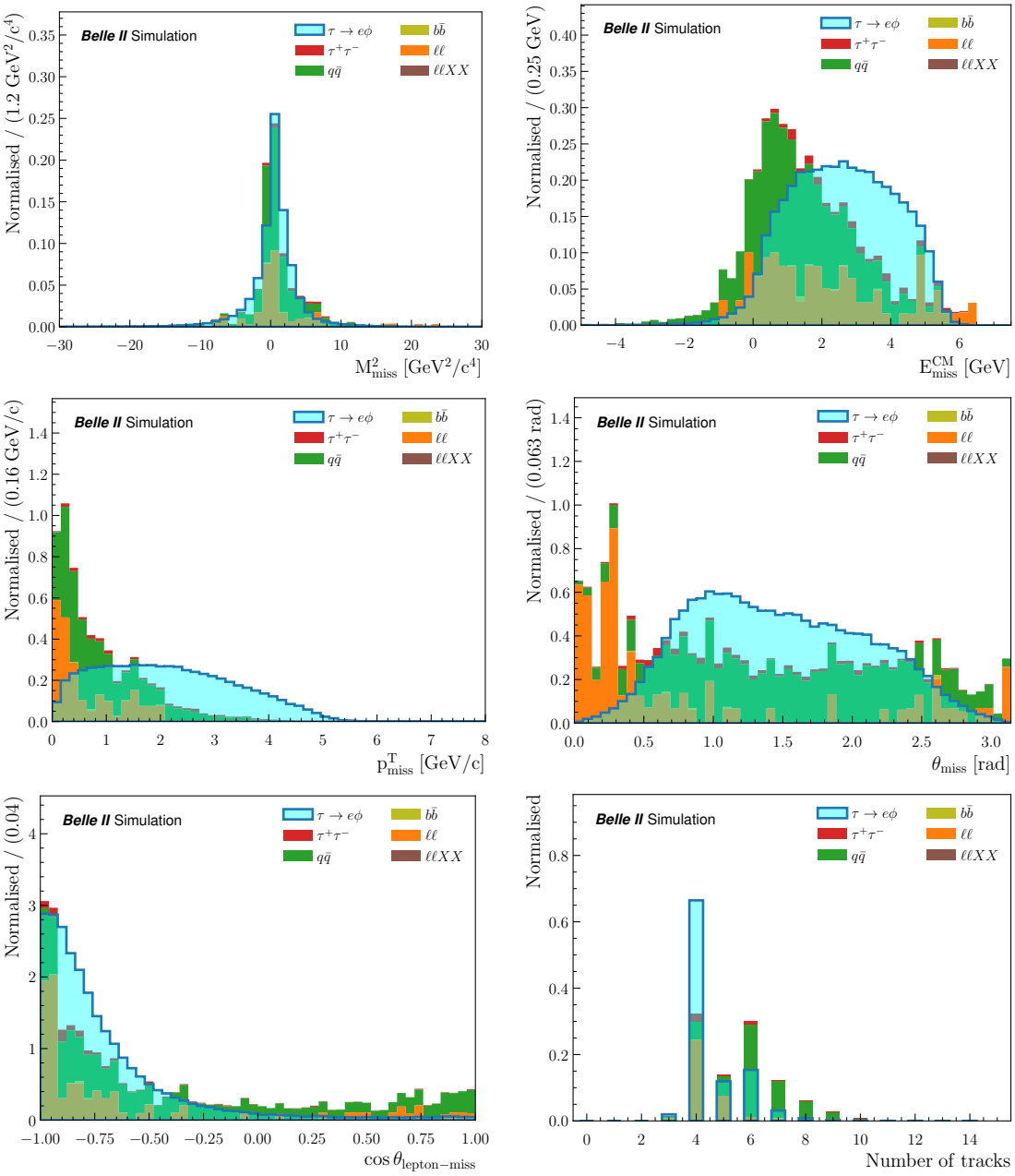


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

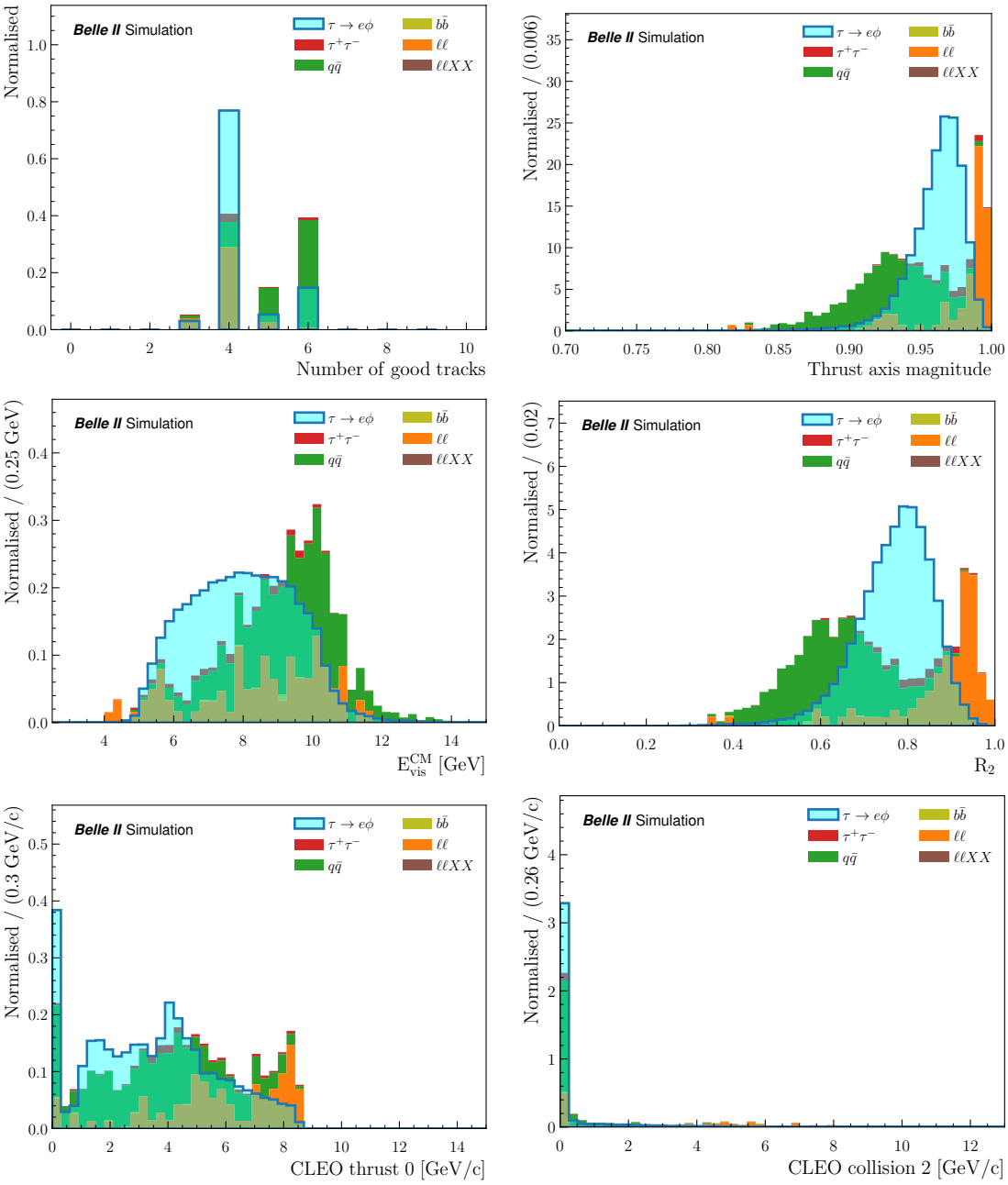


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

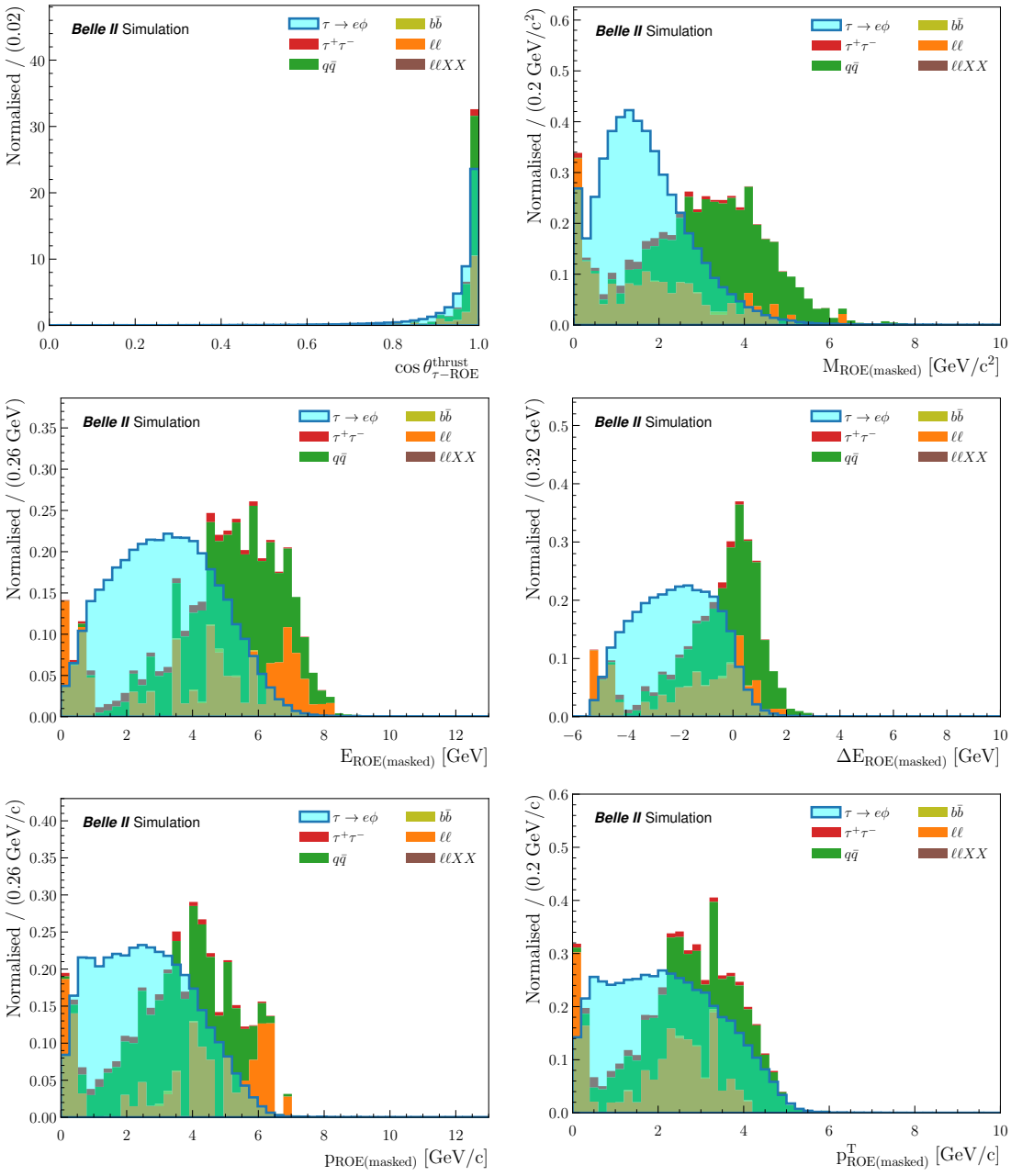


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

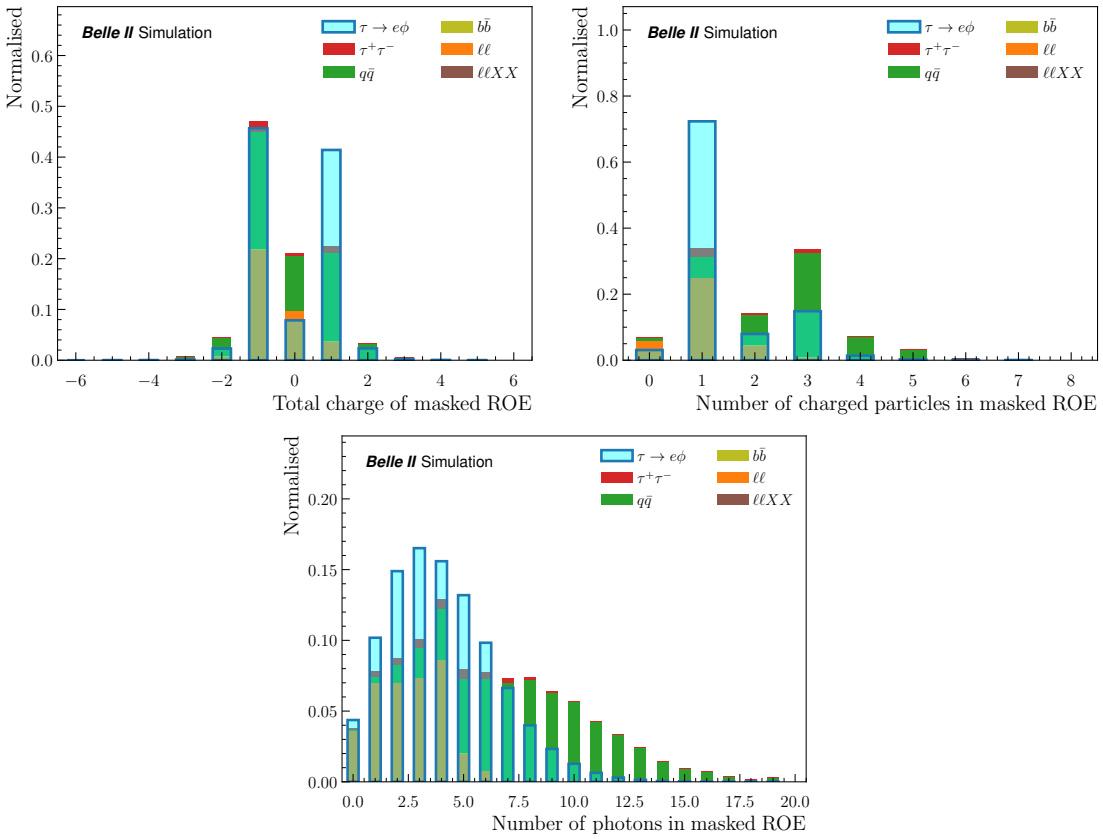


Figure 4.10: Normalised distributions of $\tau \to e\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

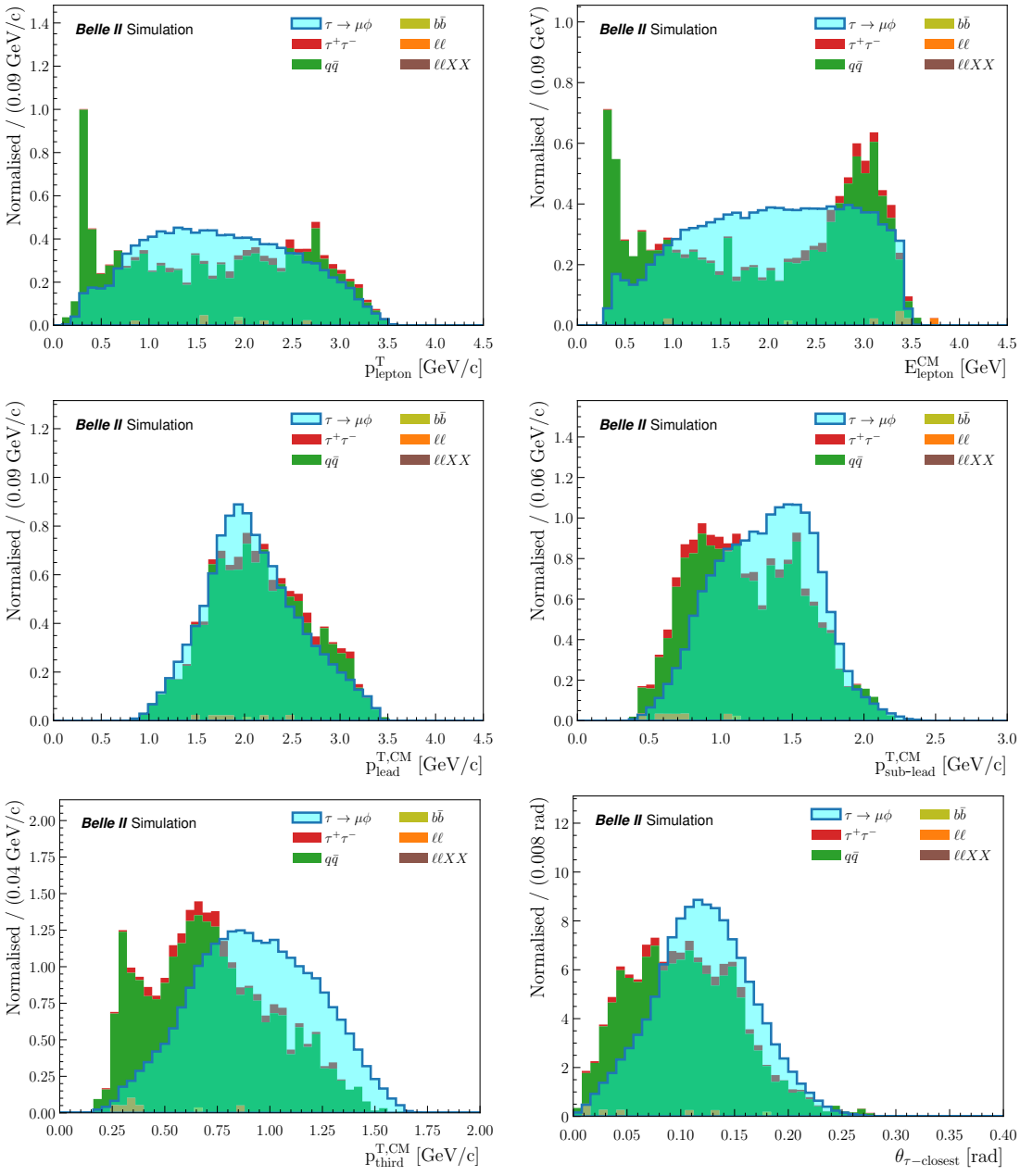


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

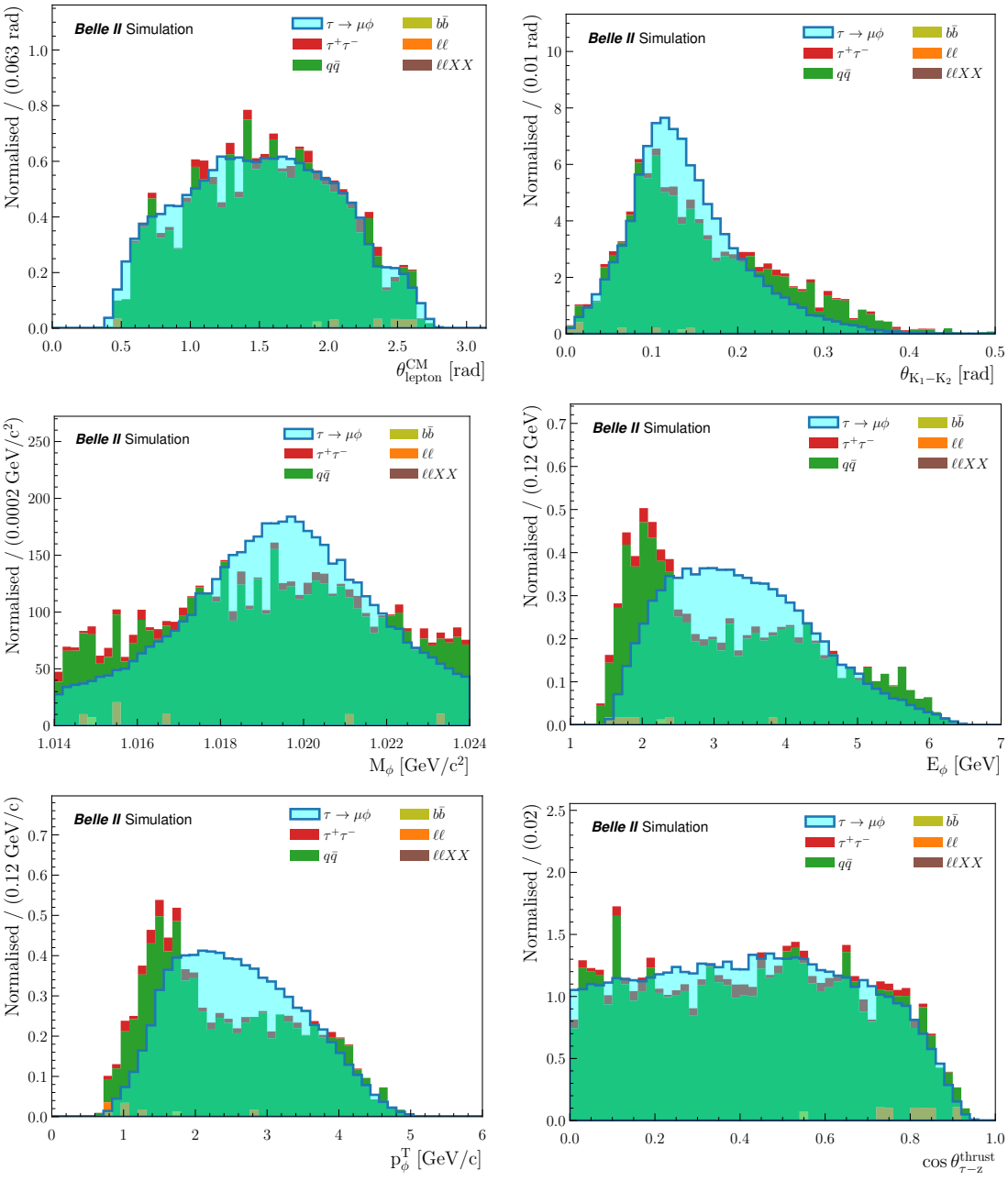


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

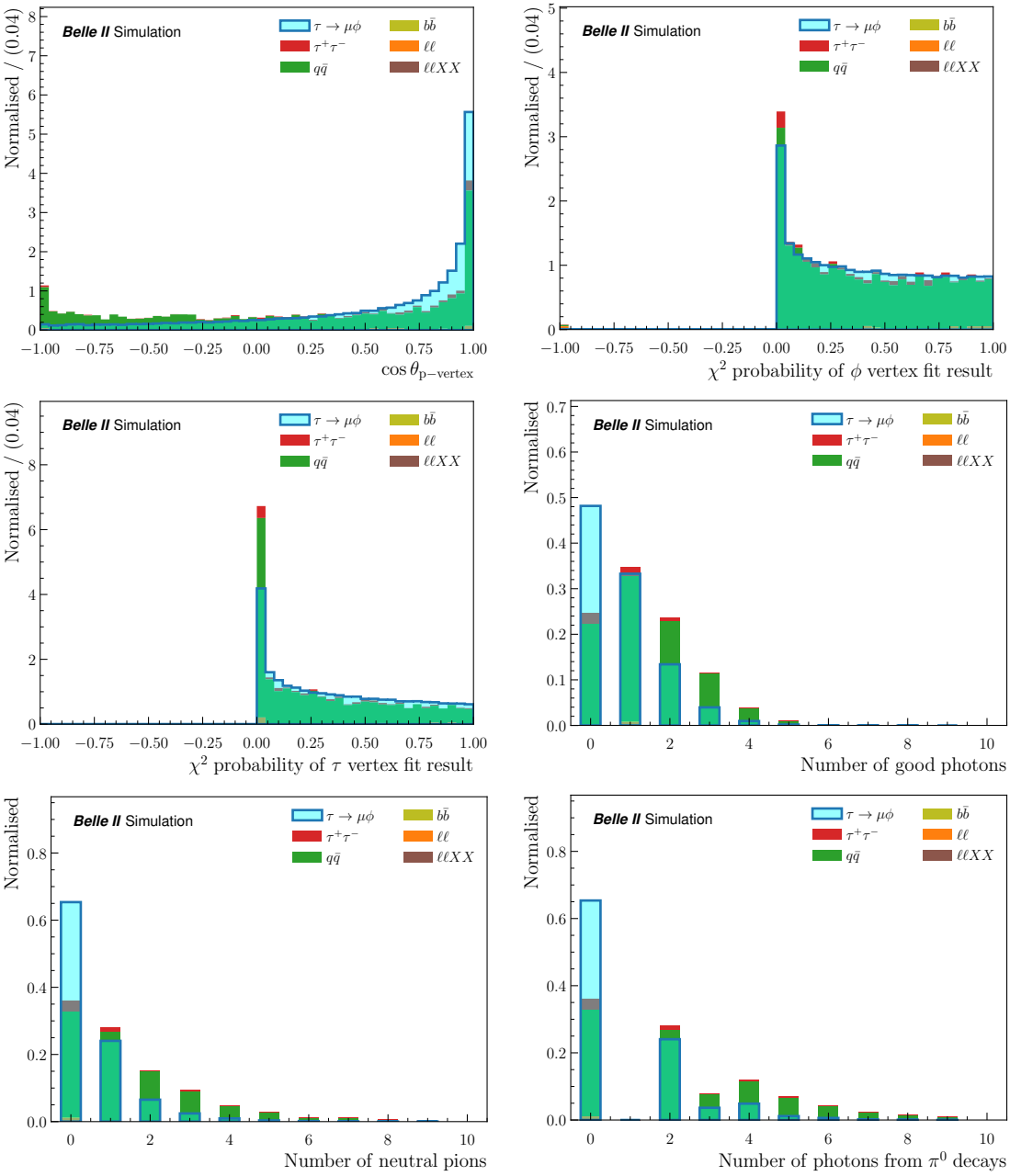


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

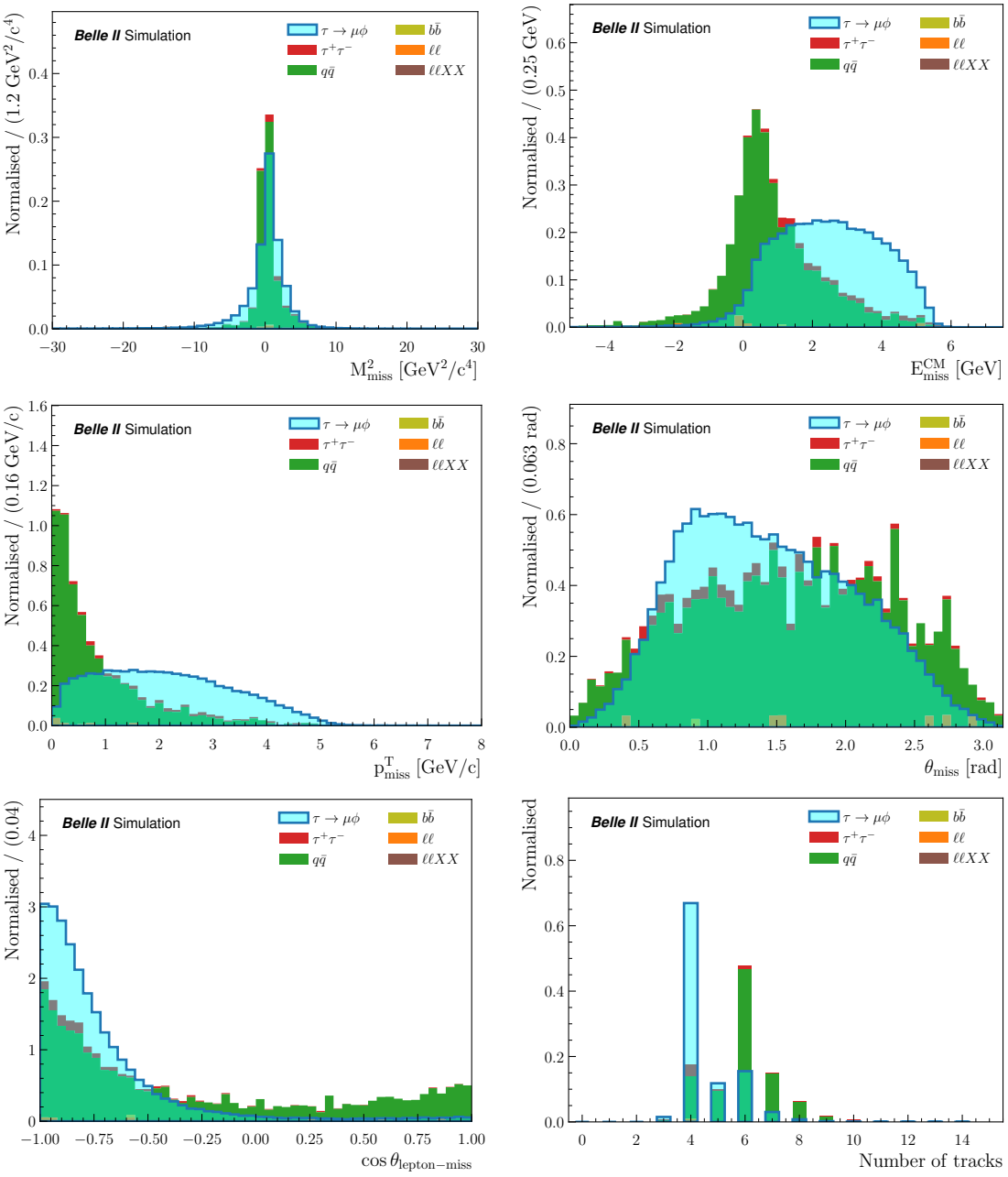


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

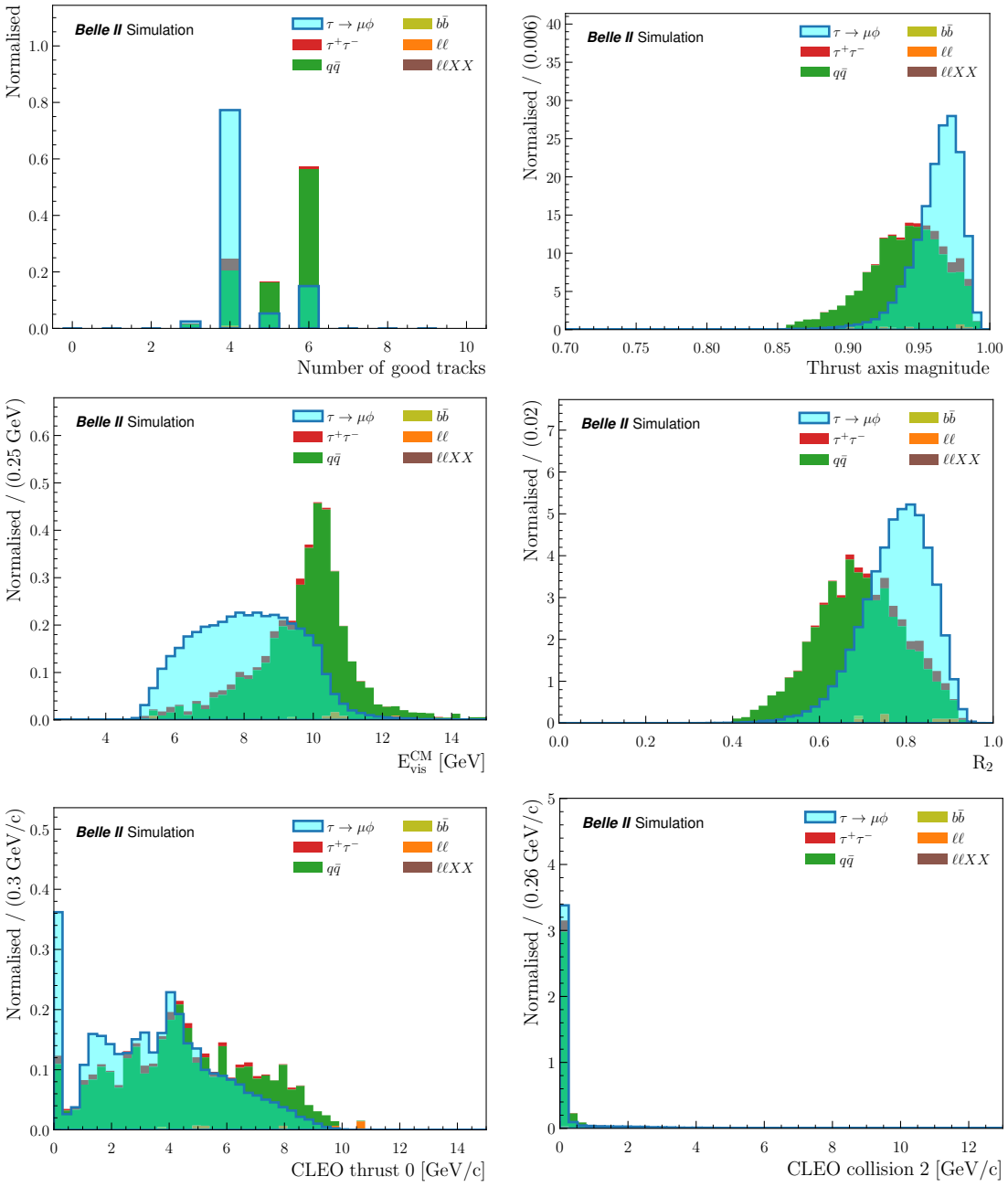


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

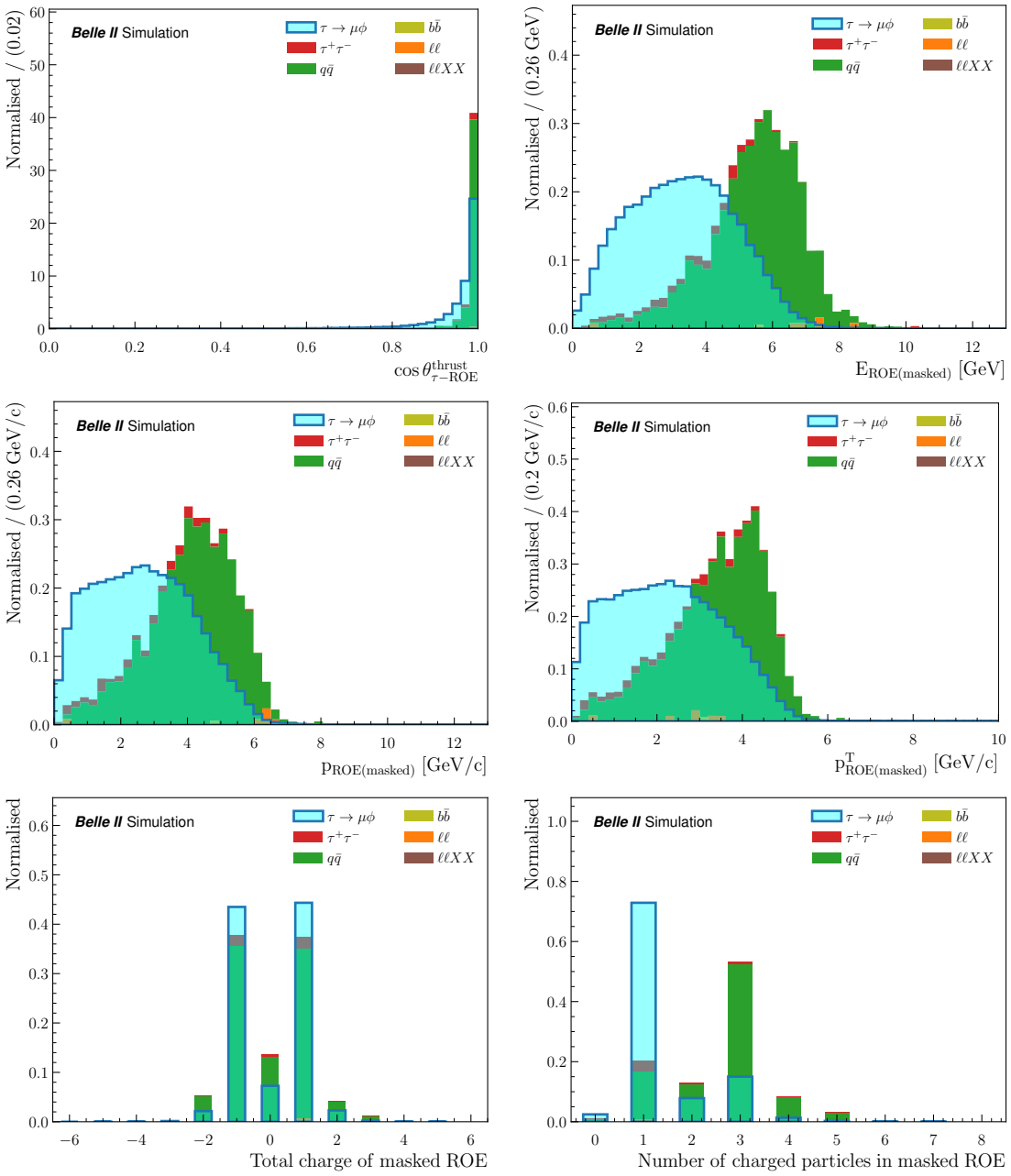


Figure 4.11: Normalised distributions of $\tau \to \mu\phi$ signal and background events in simulation, for all variables taken as inputs to the BDT.

4.4.5. Additional backgrounds

Background events from three body decays, mainly 3π (from τ decays) or $3K$ (from $q\bar{q}$ events) where one or several final state particles are misidentified, are further suppressed using a two-dimensional selection on the ϕ and lepton energies in the center of mass frame:

$$E_{\phi}^{\mathrm{CM}} > -1 \times E_{\mathrm{lepton}}^{\mathrm{CM}} + 5.2, \tag{4.4}$$

as illustrated in Figure 4.12. This cut, referred to as the "2D selection", is not applied at the stage of BDT training so that more background events are included, increasing the accuracy of the BDT. It is however applied on the validation sample before the optimal BDT selection is measured on it, and on all samples when looking at the final results.

If this selection were to be applied before the BDT selection, it would be 78% (84%) efficient on signal events and would remove 78% (88%) of all background events in the electron (muon) channel.

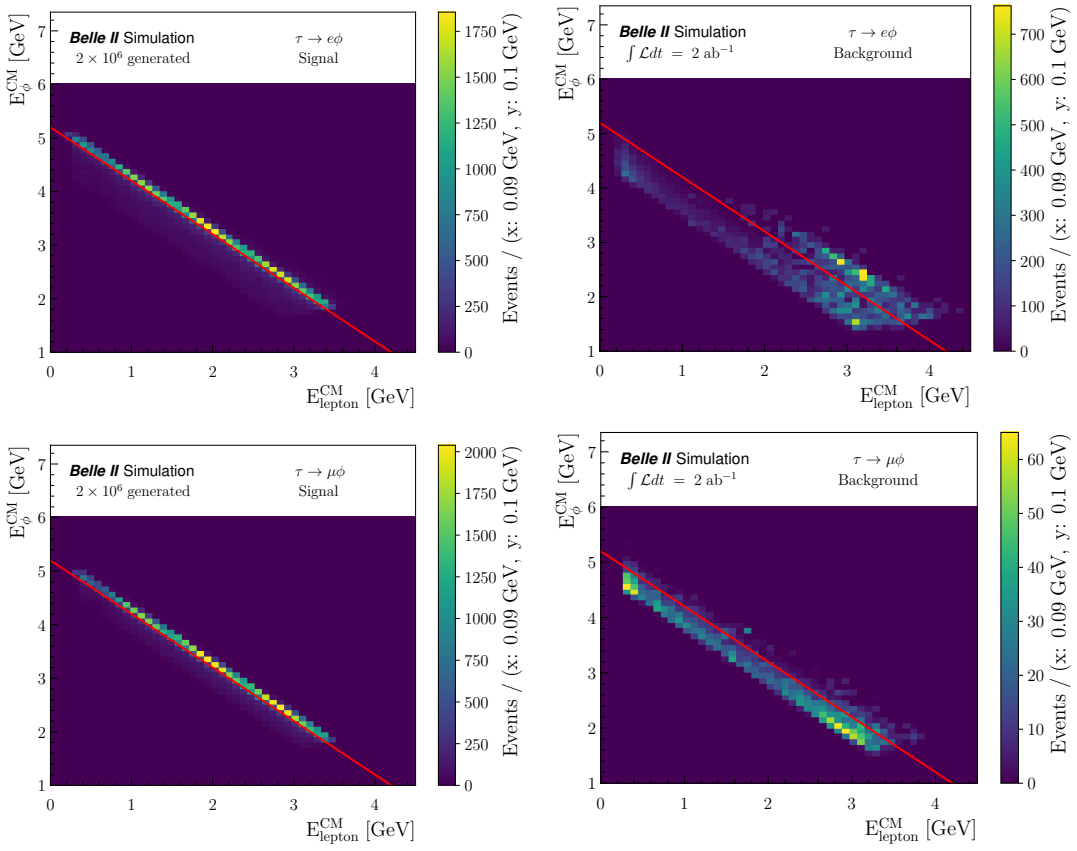


Figure 4.12: Signal (left) and background events (right) of the electron (top) and muon (bottom) channels in the $E_{\mathrm{lepton}}^{\mathrm{CM}}$ and E_{ϕ}^{CM} plane. The red diagonal line illustrates the 2D selection of Equation 4.4, keeping the upper right region.

4.4.6. Results of the background suppression strategy

The comparison of the signal probability distributions according to the BDT in the train and validation samples is shown in Figure 4.13. The optimal lower selection on the probability is 0.98 (0.84) in the $\tau \to e\phi$ ($\tau \to \mu\phi$) channel. The evolution of the logarithmic losses in Figure 4.14 demonstrates the absence of overtraining. The ranking by importance of variables used in the BDT is given in Appendix B.

The MC signal efficiencies, background yields and sensitivities (Punzi figure of merit) after the complete selection are presented in Tables 4.13 and 4.14. The details of the remaining background candidates truth-matched to specific final states are given in Table 4.15. As seen from this table, most of the remaining backgrounds in the 3σ region is coming from $\tau \to 3\pi$ for the $\mu\phi$ final state and $q\bar{q} \to KKK$ decays for the $e\phi$ final state.

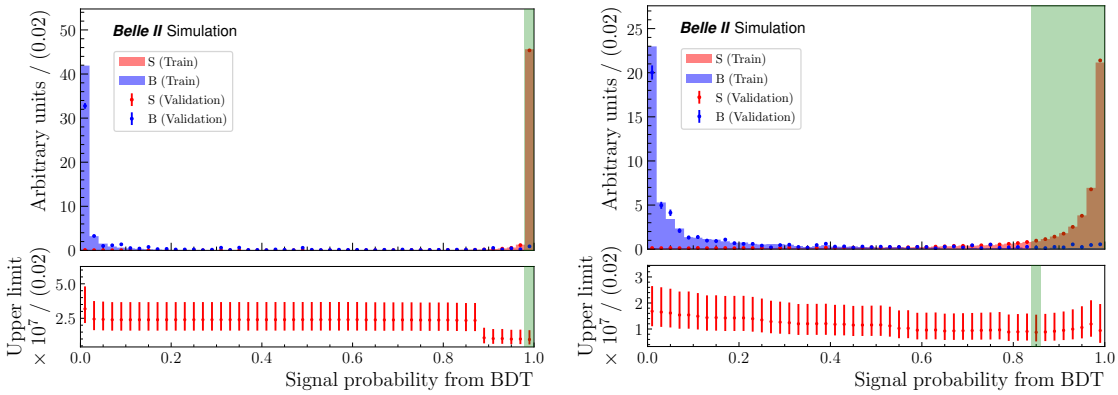


Figure 4.13: Probability of being a signal event according to the BDT, in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) channels. S means signal and B means background. The lower inset plot shows the 90% C.L. upper limit on the LFV channel's branching fraction, as evaluated on the validation sample. The green area corresponds to values selected as a result of the upper limit optimisation.

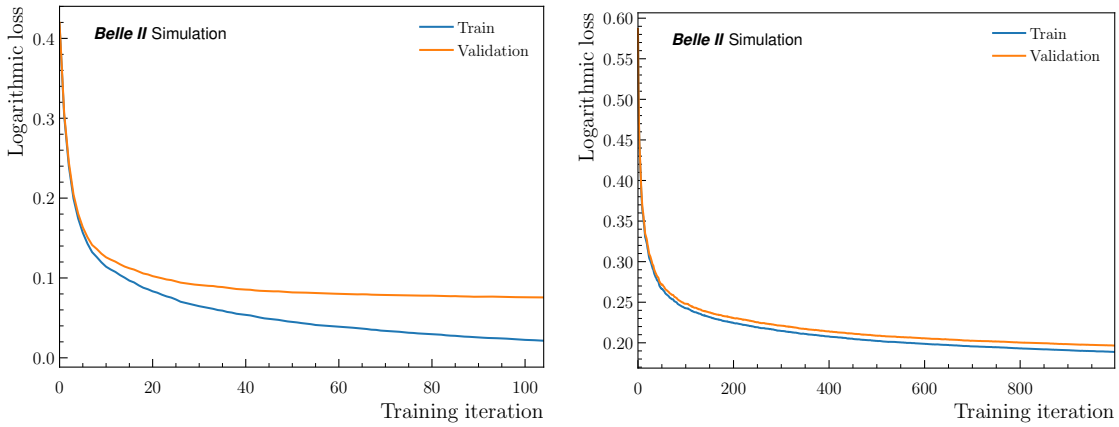


Figure 4.14: Logarithmic loss as a function of the training iteration, in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) channels. The ranges go up to the last iteration allowed by the early stopping.

Table 4.13: $\tau \to e\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	Low multi.	Total backgr.	FOM
Train	20σ	8.1%	0.0	0	0	0.0	0.054
	3σ	7.9%	0	0	0	0	0.052
Validation	20σ	8.1%	0.89	8.14	0	9.03	0.018
	3σ	7.8%	0.09	0.68	0	0.77	0.033
Test	20σ	8.1%	4.02	3.89	0	7.91	0.019
	3σ	7.9%	0.39	0.53	0	0.92	0.032

Table 4.14: $\tau \to \mu\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	Low multi.	Total backgr.	FOM
Train	20σ	7.4%	1.53	1.27	0	2.8	0.023
	3σ	7.2%	0.08	0.12	0	0.2	0.037
Validation	20σ	7.4%	2.12	5.13	0	7.25	0.018
	3σ	7.2%	0.55	0.21	0	0.76	0.031
Test	20σ	7.4%	2.41	3.42	0.53	6.36	0.018
	3σ	7.2%	0.38	0.45	0	0.83	0.030

Table 4.15: Numbers of remaining background candidates truth-matched to a final state in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions. The yields correspond to the integrated luminosities cited in Table 4.2 and don't include PID corrections.

Final state	Region	τ-pair	$q\bar{q}$	Low multi.
KKK	20σ	0	15	0
	3σ	0	3	0
KKπ	20σ	0	3	0
	3σ	0	0	0
K$\pi\pi$	20σ	1	3	0
	3σ	0	1	0
$\pi\pi\pi$	20σ	13	0	0
	3σ	0	0	0
eKK	20σ	0	5	0
	3σ	0	0	0
eeK	20σ	0	3	0
	3σ	0	0	0
e$\pi\pi$	20σ	0	3	0
	3σ	0	1	0
eeπ	20σ	5	1	0
	3σ	4	0	0

Final state	Region	τ-pair	$q\bar{q}$	Low multi.
KKK	20σ	0	20	0
	3σ	0	4	0
KKπ	20σ	3	9	0
	3σ	1	1	0
K$\pi\pi$	20σ	6	6	0
	3σ	2	0	0
$\pi\pi\pi$	20σ	26	1	0
	3σ	2	0	0
$\mu\pi\pi$	20σ	1	0	0
	3σ	1	0	0

4.5. Data-MC Comparison

4.5.1. Sidebands comparison

We check the data-MC agreement in the $5-20\sigma$ region, blinding both the signal (3σ) and $3-5\sigma$ regions. The M_τ and ΔE_τ distributions after the preselection are shown in Figure 4.15 (see Appendix E, Figures E.3 and E.4 for other variables), and in Figure 4.16 after the full selection (see Appendix E, Figures E.5 and E.7 for other variables). At this stage, the number of events according to the simulation is 6.33 (5.06) while we observe 138 (7) events in data for the $e\phi$ ($\mu\phi$) final state.

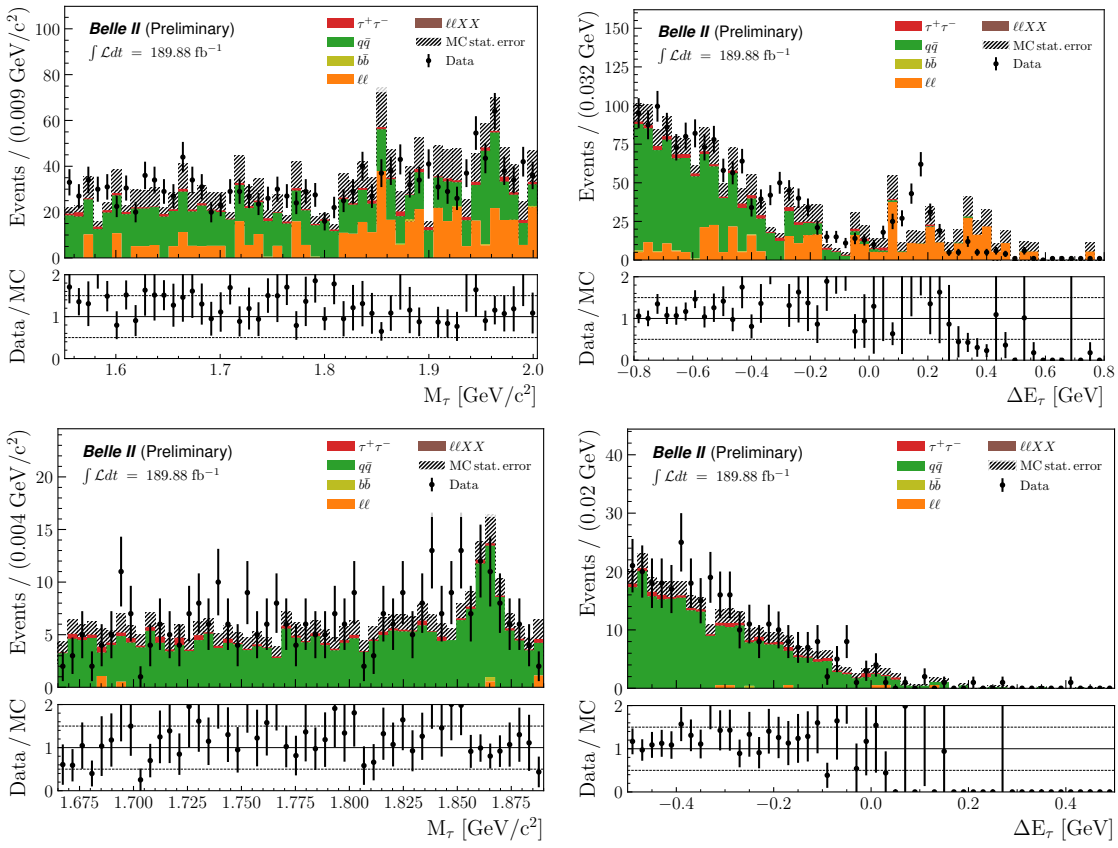


Figure 4.15: Data-MC comparison in the $5-20\sigma$ sidebands of M_τ (left) and ΔE_τ (right), for the $\tau \to e\phi$ (top) and $\tau \to \mu\phi$ (bottom) channels, after the preselection.

4.5.2. Reduction of data excess

The data-MC ratios on the sidebands yields are $22 \pm 4_{\text{stat}}$ ($e\phi$) and $1.4 \pm 0.6_{\text{stat}}$ ($\mu\phi$), thus mostly a large disagreement between data and MC in the $\tau \to e\phi$ channel,

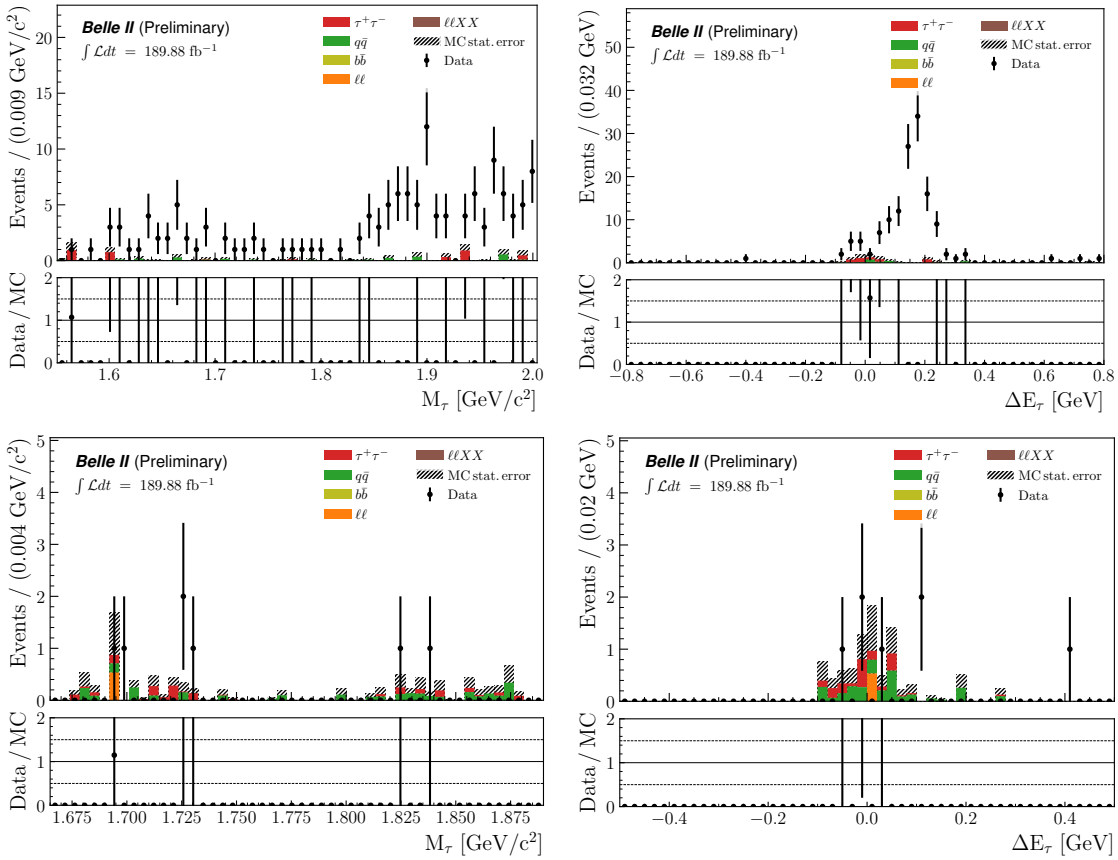


Figure 4.16: Data-MC comparison in the $5-20\sigma$ sidebands of M_τ (left) and ΔE_τ (right), for the $\tau \to e\phi$ (top) and $\tau \to \mu\phi$ (bottom) channels, after the preselection, the BDT and the 2D selections.

while the slight difference observed in $\tau \to \mu\phi$, although compatible with unity within the statistical uncertainty, seems to be a consequence of the kaonID selection, of which the effect is not completely mitigated by the hadronID corrections.

The large data excess in the electron channel, on the other hand, could be due to radiative Bhabha backgrounds ($e^+e^- \to e^+e^-\gamma$), because of the low statistics included in the MC sample on which the BDT is trained (one third of 100 fb^{-1}), and/or to mismodelling in this channel. As shown in Figure 4.17, the electronID of the kaons doesn't help much in removing this excess, but we apply nonetheless a veto by selecting events where the kaons have probabilities smaller than 0.95 of being electrons.

A more likely explanation would be a ϕ meson photoproduction ($e^+e^- \to e^+e^-\phi$), a process of which we don't have any available simulation. The preselection indeed reveals a clear peak in the data around the ϕ mass, if we enlarge the window to $0.95 < M_\phi < 1.3$ GeV/c^2 as displayed in Figure 4.18 (sidebands region).

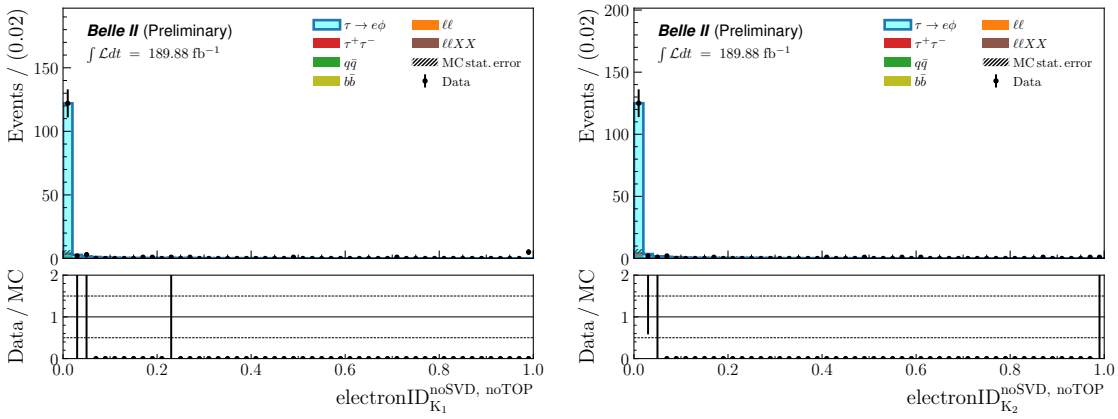


Figure 4.17: Data-MC comparison in the $5-20\sigma$ sidebands of the first kaon K_1 (left) and second kaon K_2 (right) electronID, defined without the SVD and TOP, after the preselection, the BDT and the 2D selections. The signal MC distribution is rescaled to the data bin with highest yield.

Furthermore, this excess shows some attributes specific to low-multiplicity events, as a number of tracks equal to 4 and a direction of propagation close to the beam axis (see Appendix E, Figure E.5).

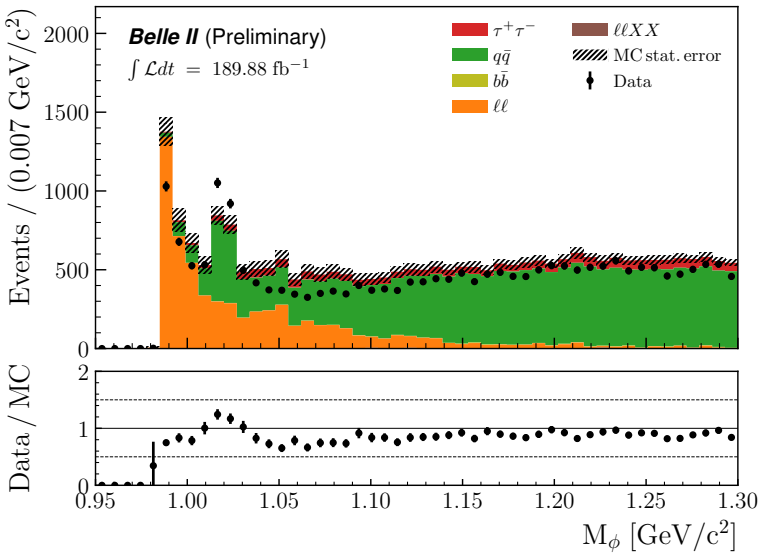


Figure 4.18: Data-MC comparison of the ϕ mass (M_ϕ) distributions in the $5-20\sigma$ sidebands, for the $\tau \to e\phi$ channel, after the preselection and with an enlarged mass window.

To alleviate this excess, a second BDT (that we call "data BDT" to distinguish it from the first BDT) is trained, with the purpose of targeting data events susceptible of not being well described by simulation, while retaining as much signal efficiency as possible. In our case, what is considered as "signal" for this BDT includes all data events in the $-0.1 < \Delta E_\tau < 0.4$ window, the rest (data events outside of the window and signal MC) being "background". The output is therefore the probability for an event to be in the data excess region *according to the data themselves*, thus is correlated to the probability for events to be mismodelled.

The datasets on which this data BDT is trained are one tenth of the full considered dataset (18.99 fb^{-1}) and half of the signal events comprised in the MC train sample. In both samples, the preselection, BDT, 2D and kaon's electronID selections are not applied, except for the requirement that the lepton's electronID remains larger than 0.9. In the data sample, the 5σ signal region is blinded.

The inputs of the data BDT are variables that appear to have the highest discriminating power between signal MC and data background after all selections, as one can judge from the comparative distributions in Figure 4.19. These are $p_{\text{lead}}^{\text{T,CM}}$, $\theta_{\text{lepton}}^{\text{CM}}$, $\cos\theta_{\tau-z}^{\text{thrust}}$, θ_{miss}, $E_{\text{ROE(masked)}}$, $p_{\text{ROE(masked)}}$, $E_{\text{vis}}^{\text{CM}}$, the thrust axis magnitude and R_2.

The parameters of the data BDT are optimised in the same manner as the first one, as explained in Section 4.4.4. An optimal selection is, however, not evaluated at this stage.

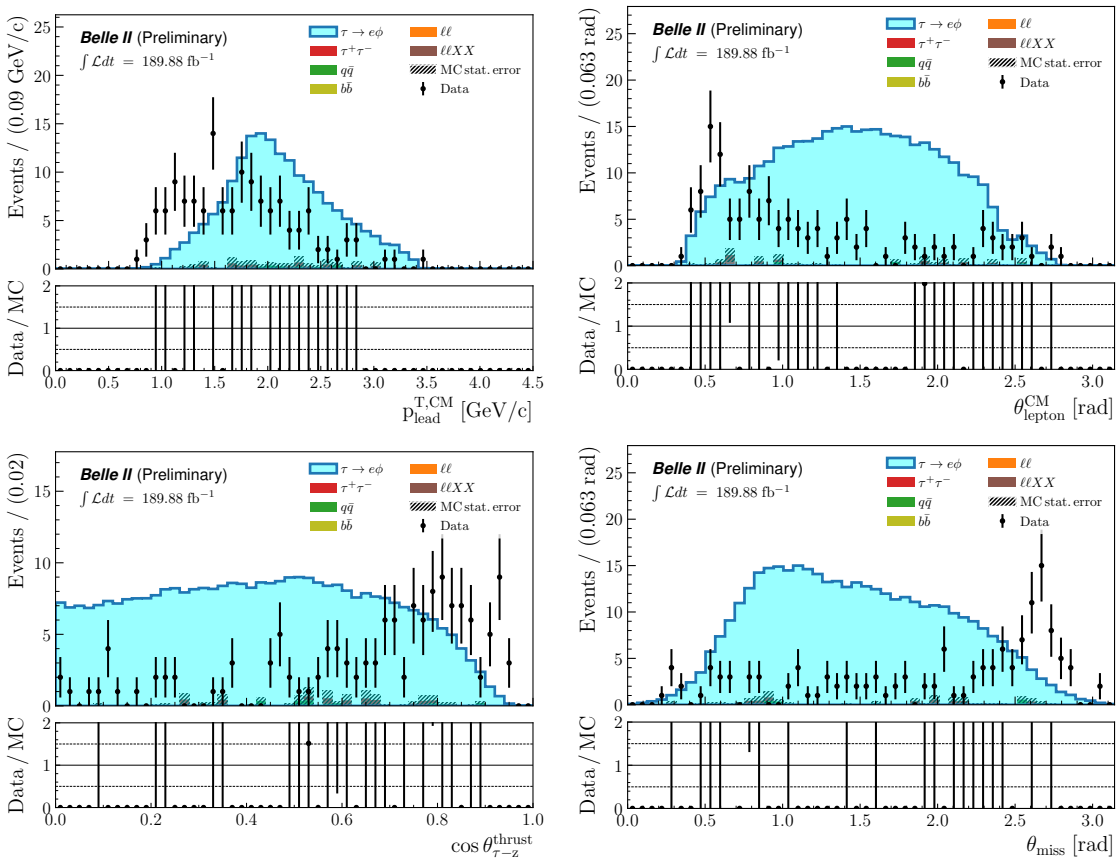


Figure 4.19: Data-MC comparison in the $5-20\sigma$ sidebands after the preselection, the BDT and the 2D selections of variables used in the training of the data BDT. The signal MC distribution is rescaled to the data bin with highest yield.

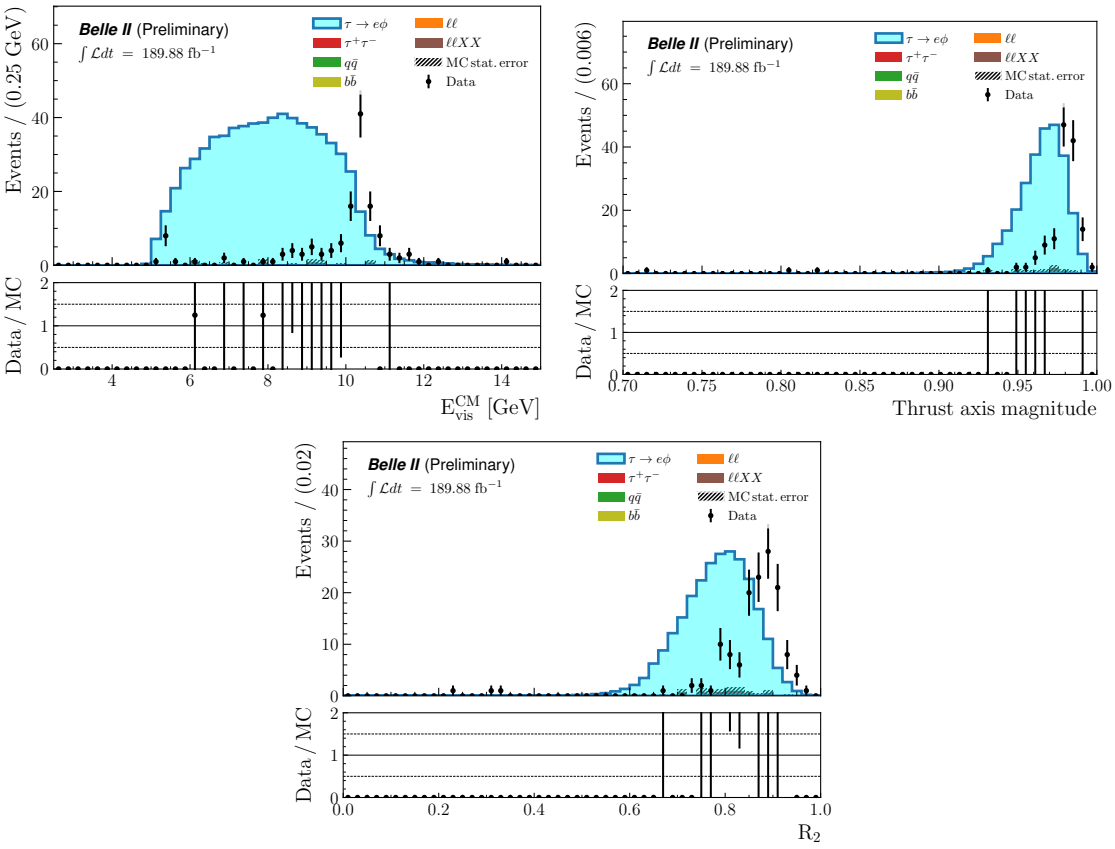


Figure 4.19: Data-MC comparison in the $5 - 20\sigma$ sidebands after the preselection, the BDT and the 2D selections of variables used in the training of the data BDT. The signal MC distribution is rescaled to the data bin with highest yield.

The distributions of the data BDT predictions on MC and data after the complete set of selections are shown in Figure 4.20. The corrected data BDT serves only in the computation of the signal efficiency systematic uncertainties. We decide to select events up to a probability of 0.001, in order to remove 94% of data events and keep 76% of signal. The data-MC ratio in the $5-20\sigma$ region drops to $2.4 \pm 1.0_{\text{stat}}$, greatly improved but still incompatible with unity.

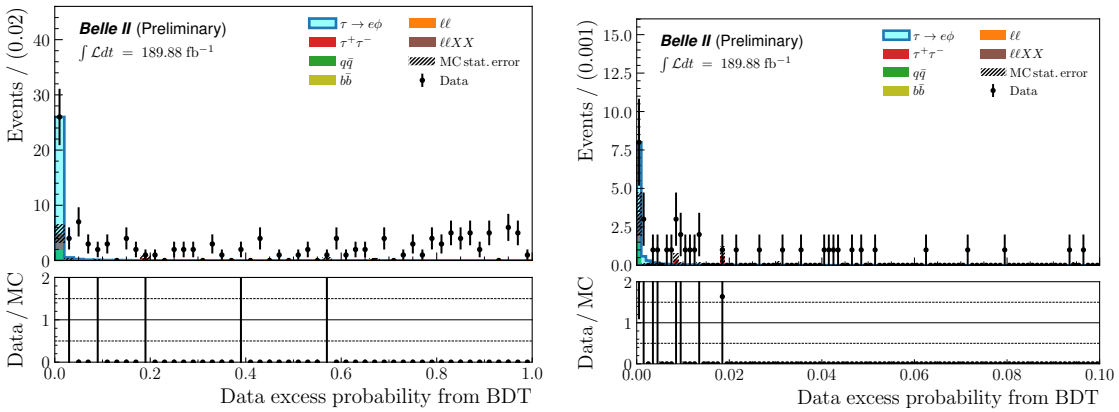


Figure 4.20: Probability, in the $[0, 1]$ (left) and $[0, 0.1]$ (right) ranges, for $\tau \to e\phi$ MC and data events to be in the data excess region, after the preselection, BDT, 2D and kaon's electronID selections. Signal distributions are rescaled to the data bin with highest yield.

The M_τ and ΔE_τ distributions after the criterium on the data BDT is applied are shown in Figure 4.21 (see Appendix E, Figure E.6 for other variables).

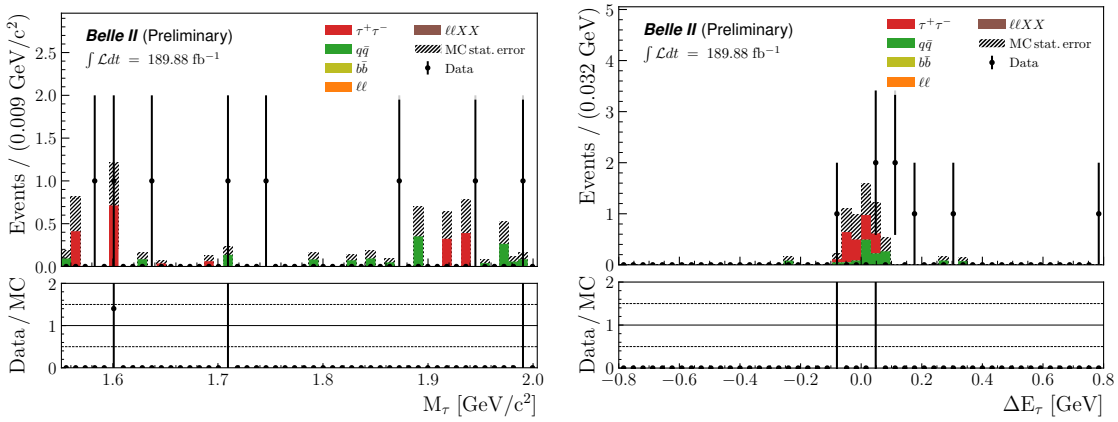


Figure 4.21: Data-MC comparison in the $5-20\sigma$ sidebands of M_τ (left) and ΔE_τ (right), for the $\tau \to e\phi$ channel, after the full selection.

4.5.3. Computation of the expected data yield and summary

The expected data yield $N_{\text{exp}}^{\text{SR}}$ in the signal region (SR) is estimated from the observed data yield $N_{\text{obs}}^{\text{RSB}}$ in the sidebands reduced to the $\pm 3\sigma$ ΔE_τ region (RSB) as

$$N_{\text{exp}}^{\text{SR}} = N_{\text{obs}}^{\text{RSB}} \frac{N_{\text{MC}}^{\text{SR}}}{N_{\text{MC}}^{\text{RSB}}}, \tag{4.5}$$

according to the ratio of MC background yields N_{MC} in both regions. In order to increase the accuracy of this ratio, because of the low amount of MC background events retained after the selections, we take the mean of the ratios in the validation and test samples: even if the validation sample contributes to the BDT optimisation, one can expect the possible bias that is introduced, compared to the test sample, to become negligible when computing a ratio.

A summary of the final efficiencies and yields obtained after all selections are applied is presented in Table 4.16. For comparison, signal efficiencies and expected background events from the previous BaBar and Belle experiments are reported in Table 4.17.

The distributions of data and MC events in the different regions are shown in Figure 4.22.

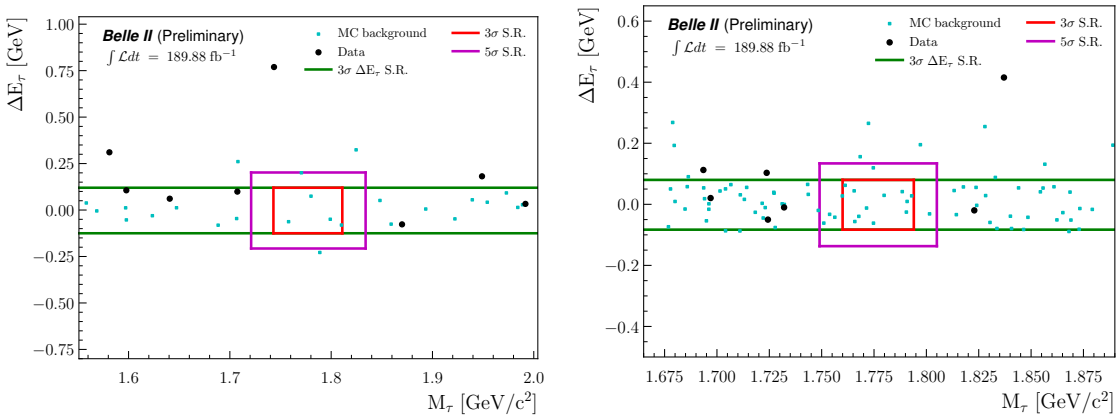


Figure 4.22: Scatter plots of MC background and data events in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) channels, after all selections. The red and magenta squares represent the 3 and 5σ signal regions blinded in data, while the green lines delineate the $\pm 3\sigma$ ΔE_τ band used for estimating the expected number of events in the signal region. The MC background events correspond to one third of the integrated luminosities cited in Table 4.2, without any weighting.

Table 4.16: Summary of the signal selection and background suppression results obtained on simulation (test sample) and data. The results are given in specific regions of the (M_τ, ΔE_τ) space: the 3σ signal region (SR), the $5-20\sigma$ sidebands (SB) or the sidebands reduced to the 3σ ΔE_τ region (RSB). Results marked with an asterisk are taken from the validation sample. The expected data yield in the SR N_{exp} is the product of the yield N_{obs} observed in the RSB and the MC background ratio. The latter is the mean of the test and validation ratios.

Result		Region	Mode	
			$e\phi$	$\mu\phi$
Signal efficiency $\varepsilon_{\ell\phi}$		SR	6.0%	7.2%
MC background yield		SB	3.34	5.06
		RSB	3.10 / 2.53*	4.10 / 4.06*
		SR	0.38 / 0.36*	0.83 / 0.76*
Background ratio in MC		SR / RSB	0.13	0.19
Data yield	N_{obs}	SB	8	7
		RSB	5	4
	N_{exp}	SR	0.66	0.78
Data / MC ratio		SB	$2.4 \pm 1.0_{\text{stat}}$	$1.4 \pm 0.6_{\text{stat}}$
		RSB	$1.6 \pm 0.8_{\text{stat}}$	$1.0 \pm 0.5_{\text{stat}}$

Table 4.17: Summary of signal efficiencies and expected background yields from BaBar (451 fb^{-1}) and Belle (854 fb^{-1}) [7, 8], in their respective signal regions.

Result	Experiment	Mode	
		$e\phi$	$\mu\phi$
Signal efficiency $\varepsilon_{\ell\phi}$	BaBar	6.43%	5.18%
	Belle	4.18%	3.21%
Data yield N_{exp}	BaBar	0.68	2.76
	Belle	0.47	0.06

4.6. Detector studies and systematic uncertainties

The systematic uncertainties arise from:

- detector effects, which affect quantities that have to be measured directly on data,
- differences between data and simulation, due to possible mismodelling in the generation of MC samples.

In both cases, the systematics need to be evaluated with dedicated performance and data validation studies: some are specific to this analysis and derived from data-MC comparisons performed in the previous chapters, while others are provided as uncertainty estimations common to all analyses done in Belle II.

We will discuss here the sources of systematic uncertainties that have been identified on the various inputs of the upper limit computation. A summary of these sources, along with their relation to the relevant quantities, is provided in Table 4.26.

4.6.1. Main systematic uncertainty sources

In the 3σ signal region, the upper limit on the signal branching fraction can be computed as

$$\mathcal{B}_{\mathrm{UL}}(\tau \to \ell\phi) = \frac{s}{L \times 2\sigma_{\tau\tau} \times \varepsilon_{\ell\phi}}, \tag{4.6}$$

where s is the upper limit on $N_{\mathrm{obs}} - N_{\mathrm{exp}}$, the difference between the observed and expected data yields. The other inputs are the integrated luminosity L corresponding to the analysed data; the τ-pair production cross section $\sigma_{\tau\tau}$ and the signal efficiency $\varepsilon_{\ell\phi}$. Their associated systematic uncertainties must be taken into account in order to perform a proper evaluation of the upper limit.

4.6.1.1. Systematic uncertainty in the selection efficiencies

The main source of systematic uncertainty comes from the estimate of the signal efficiency, with contributions from the particle identification, trigger and track reconstruction efficiencies, as well as from the many selections applied to reject background (see Sections 4.4 and 4.5):

1. *Particle identification efficiency and misidentification probability*: we take into account the fact that the lepton and hadron identification probabilities are not modelled perfectly in simulation, introducing differences with respect to data. Statistical and systematic uncertainties on PID efficiencies and fake rates are provided as part of the latest recommendations of the LeptonID and HadronID performance groups.

2. *Trigger efficiency*: as explained in Section 4.3.6, the low-multiplicity trigger selection is required in both data and MC. Therefore, the contribution of the trigger efficiency is directly taken into account in the estimate of the final selection efficiency on the simulated signal sample. Possible differences in the trigger efficiencies between data and simulation are accounted for in the data validation procedure, when comparing absolute yields in data and MC for the selected control samples (see studies in Section 4.6.6) and included in the shape correction applied to the input variables of the signal simulation entering the BDT selection. Additionally, the overall discrepancy in the trigger efficiency between data and simulation is also evaluated in the $\tau \to 3\pi\nu_\tau$ control sample. The measured effects and trigger related studies are reported in the section below, with further details in Appendix F. Conservatively, we assign on the signal efficiency, as systematic uncertainty due to the trigger effect, the total variation from unit observed in the ratio of the studied efficiencies.

3. *Tracking efficiency*: the difference in track reconstruction efficiencies between data and simulation is evaluated with a tag-and-probe performance study on τ-pair events. Such a deviation is used to assess the uncertainty assigned to each track explicitly required in the event reconstruction, with a total contribution equal to the sum in quadrature of each individual one.

4. *BDT selection efficiency*: the possible differences of BDT performances between data and simulation are evaluated thanks to a reweigthing technique, consisting in correcting the shapes of signal distributions in MC for variables involved in the BDT training. The discrepancy in signal efficiencies between corrected and uncorrected BDTs is then used as systematic uncertainty. The weigths determination is presented in the data validation procedure on control samples (see Sections 4.6.6.1 and 4.6.6.2).

4.6.1.2. Systematic uncertainty in the background yields

The systematics on the background yield, expected in the signal region before the unblinding, take into account:

1. *MC background ratio*: since this is the ratio of background yields in two different regions of the $M_\tau - \Delta E_\tau$ plane, each affected by uncertainties of the same nature, we consider that these contributions cancel out each other.

2. *Momentum scale*: we take into account the tracking momentum bias derived from the wrong magnetic field map (B-field map) used in the data reconstruction process, following prescriptions from the Belle II collaboration.

4.6.1.3. Other sources of systematic uncertainty

1. *Luminosity*: the systematic uncertainty on the integrated luminosity has been measured in [71], using Phase 2 data, on two samples corresponding to either Bhabha or digamma events. The combination of both measurements gives a relative uncertainty of 0.6%, that we chose to consider as relevant for the luminosity of the data collected afterwards.

2. *Tau-pair production cross section*: the uncertainty on this parameter was determined in [62] as 0.003 nb.

4.6.2. Trigger efficiency

We evaluate the systematic impact on the signal efficiency due to the trigger effect from two different sources:

1. the difference between data and simulation, estimated from the relative trigger efficiency measured in data and MC on the $\tau \to 3\pi\nu_\tau$ control sample, as shown in Appendix F, Figure F.1: 0.9%;

2. the bias in the trigger efficiency estimate as relative to the CDC reference lines, measured in signal MC samples of the $\tau \to \ell\phi$ LFV decay: 0.4% (0.9%) in the electron (muon) channel, as illustrated in Figure 4.23.

Conservatively, taking the sum in quadrature of the above measured effects would yield 1.0% and 1.3% systematic uncertainties on the $\tau \to e\phi$ and $\tau \to \mu\phi$ signal efficiencies, respectively. A less conservative option, but reasonable assuming that the difference in the trigger efficiency between data and simulation is already accounted for from the control sample studies and the correction of signal distributions, we assign only the measured bias as uncertainty in the signal efficiency due to trigger effect, thus 0.4% (0.9%) for $\tau \to e\phi$ ($\tau \to \mu\phi$).

4.6.3. Particle identification

The identification of signal charged particles involves LID and HID (see Sections 4.4.2 and 4.4.3) selection criteria. The uncertainties on the PID selections comprise four contributions: upward (↑) or downward (↓) statistical (stat) or systematic (sys) uncertainties, this for both LID and HID.

Keeping the notation introduced in Equation 4.3, we define four PID weights

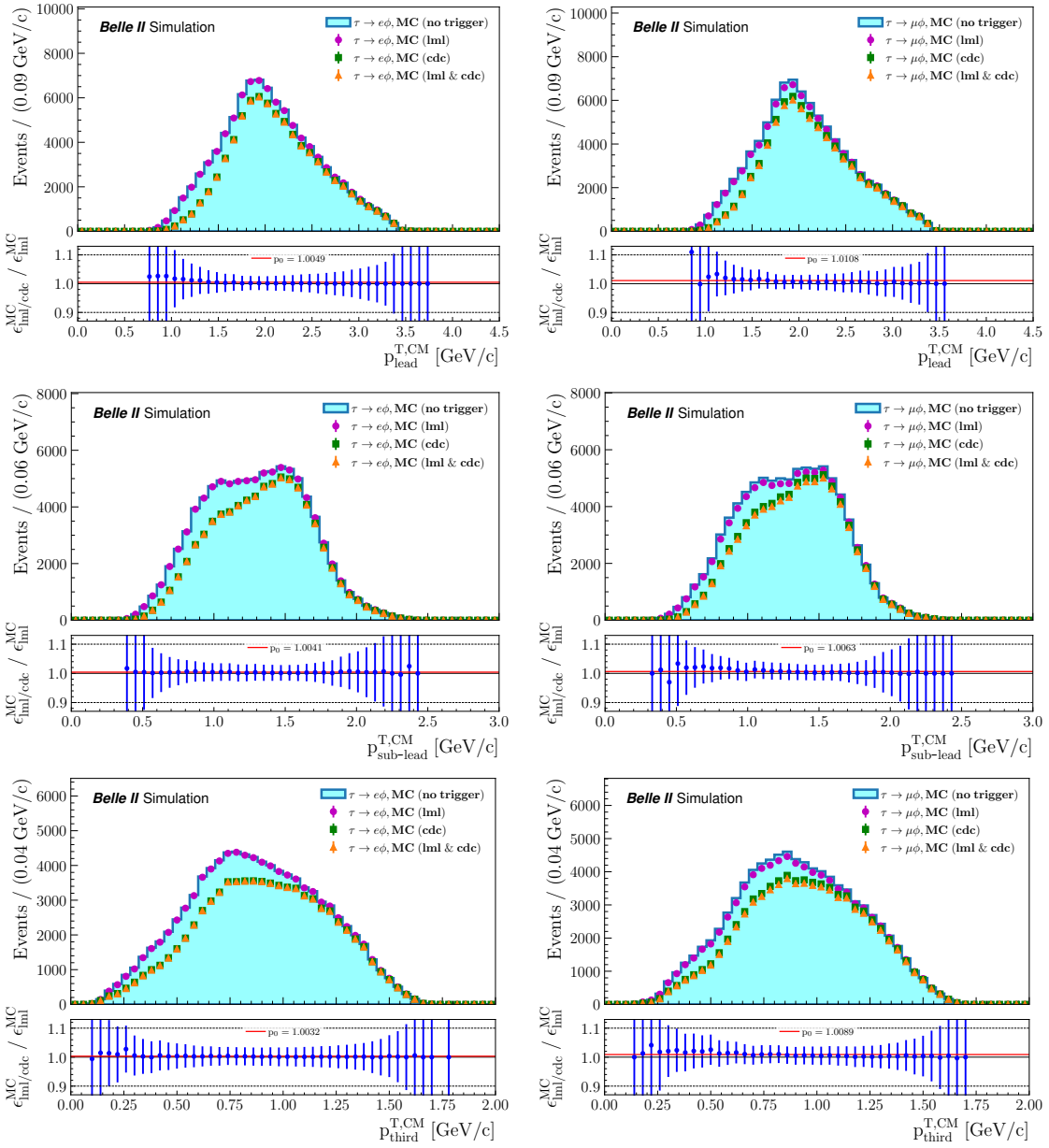


Figure 4.23: MC distributions of the leading (top), sub-leading (middle) and third (bottom) 3-prong tracks transverse momenta in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) LFV decay channels, including different trigger requirements.

modified by an uncertainty:

$$w^{\text{stat},\uparrow}_{\text{PID}} = \left(w_{\text{LID}} + w^{\text{stat},\uparrow}_{\text{LID}}\right) \times \left(w_{\text{HID}} + w^{\text{stat},\uparrow}_{\text{HID}}\right), \tag{4.7}$$

$$w^{\text{stat},\downarrow}_{\text{PID}} = \left(w_{\text{LID}} - w^{\text{stat},\downarrow}_{\text{LID}}\right) \times \left(w_{\text{HID}} - w^{\text{stat},\downarrow}_{\text{HID}}\right), \tag{4.8}$$

$$w^{\text{sys},\uparrow}_{\text{PID}} = \left(w_{\text{LID}} + w^{\text{sys},\uparrow}_{\text{LID}}\right) \times \left(w_{\text{HID}} + w^{\text{sys},\uparrow}_{\text{HID}}\right), \tag{4.9}$$

$$w^{\text{sys},\downarrow}_{\text{PID}} = \left(w_{\text{LID}} - w^{\text{sys},\downarrow}_{\text{LID}}\right) \times \left(w_{\text{HID}} - w^{\text{sys},\downarrow}_{\text{HID}}\right). \tag{4.10}$$

The statistical and systematic uncertainties on a yield n_{PID}, measured by applying the PID weights on all events, are

$$\sigma_{\text{PID}}^{\text{stat}} = \max(|n_{\text{PID}} - n_{\text{PID}}^{\text{stat},\uparrow}|, |n_{\text{PID}} - n_{\text{PID}}^{\text{stat},\downarrow}|), \tag{4.11}$$

$$\sigma_{\text{PID}}^{\text{sys}} = \max(|n_{\text{PID}} - n_{\text{PID}}^{\text{sys},\uparrow}|, |n_{\text{PID}} - n_{\text{PID}}^{\text{sys},\downarrow}|). \tag{4.12}$$

Finally, the total PID uncertainty on n_{PID} is

$$\sigma_{\text{PID}} = \sqrt{\sigma_{\text{PID}}^{\text{stat}\,2} + \sigma_{\text{PID}}^{\text{sys}\,2}}. \tag{4.13}$$

Here, the total PID uncertainty on the signal efficiency is 0.11% (0.16%) in the $\tau \to e\phi$ ($\tau \to \mu\phi$) decay channel.

4.6.4. Tracking recommendations for ICHEP

The data-MC discrepancy in the track reconstruction efficiency has been studied in a previous Belle II analysis on tau-pair events [72]. The most updated values on latest data are summarised in a document prepared by the Belle II Performance group for ICHEP 2022 [45], recommending to assign a per-track systematic uncertainty of 0.30% on the tracking efficiency.

4.6.5. Momentum scale

Imperfections in the magnetic field map used in data reconstruction can introduce a track momentum bias in data with respect to MC. To account for this, momentum scale correction factors are derived according to the observed D^0 mass shift in $D^{*+} \to [D^0 \to K^-\pi^+]\pi^+$ events [73].

The scale factor is applied directly on data, following the Performance group prescriptions [74] that set its value at 0.99971 $(+3.5/-5.5)\times10^{-4}$. Two additional data samples are produced similarly by applying the upward and downward variations of this factor. The differences with respect to the observed data yield N_{obs} in the $\pm3\sigma$ ΔE_τ sideband region are assigned as systematic uncertainties, found to be $+0.0/-0.0$ for $\tau \to e\phi$ and $+0.0/-0.2$ for $\tau \to \mu\phi$.

4.6.6. BDT selection efficiency

We consider two control channels, $D_s^\pm \to \phi\pi^\pm$ and $\tau^\pm \to \pi^+\pi^-\pi^\pm\nu_\tau$, to validate the data-MC agreement of some of the variables listed in Tables 4.7 and 4.8 that have ranges compatible with the signal LFV distributions. $D_s \to \phi\pi$ has the advantage of having a final state and kinematics close to our signal, with no missing energy. On the other hand, the distribution shapes of event and ROE related variables in $c\bar{c}$ can be different compared to $\tau^+\tau^-$, that is why we study the $\tau \to 3\pi\nu_\tau$ control channel as well.

Based on the data-MC agreement observed in the two control channels, we extract binned corrections, defined on the analysed variables, that can be assigned to $\tau \to \ell\phi$ events. This reweighting will be used to measure systematic uncertainties on the signal efficiencies.

4.6.6.1. Signal corrections from $D_s^\pm \to \phi\pi^\pm$ control channel

For this study, we use 189.88 fb^{-1} of data (described in Section 2.7, Table 2.2), as well as generic (200 fb^{-1}) and low-multiplicity (various luminosities) run-independent Monte Carlo simulations (`MC14`). The tracks and event reconstruction is mostly the same as the one described in Section 4.3. The selection specific to this channel is as follows:

- $1.92 < M_{D_s} < 2.02$ GeV/c,
- $-0.8 < \Delta E_{D_s} < 0.8$ GeV,
- $1.014 < M_\phi < 1.024$ GeV/c,
- no vertex fitting on the ϕ meson,
- no TauLFV skim required,
- at least one low-multiplicity trigger line must be fired.

The reconstructed D_s invariant mass is shown in Figure 4.24. There is a non-negligible amount of background contamination from $q\bar{q}$ and low-multiplicity events, that is why it is necessary to get rid of those in order to compare the signal distributions. This is done thanks to the *sPlot* technique [75], for which the D_s mass distribution is fitted to obtain *sWeights*, as shown in Figure 4.25. The fit is binned and uses a Cauchy function for the signal and a uniform distribution to model the background.

Signal $D_s \to \phi\pi$ events in MC are identified by truth-matching the 3-prong tracks to one pion and two kaons and the mother particle to a D_s meson. The variables used to define corrections are listed in Table 4.18, with the corresponding ranges. The comparison of sWeighted data and signal MC distributions for these

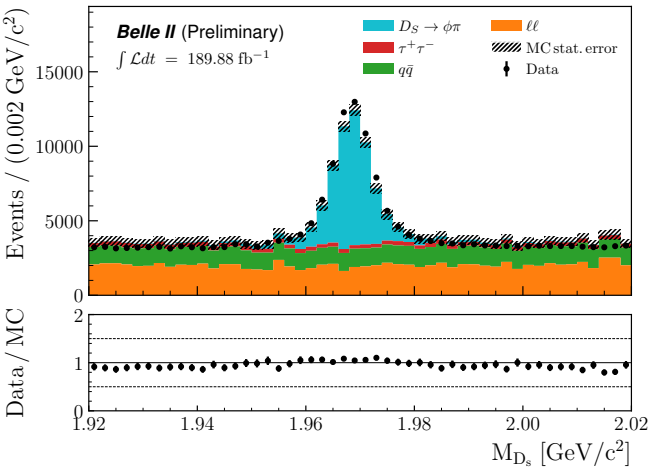


Figure 4.24: Distribution of the D_s mass M_{D_s} in simulation and data, after reconstruction of the $D_s \to \phi\pi$ decay channel. MC events displayed in cyan are truth-matched to this channel.

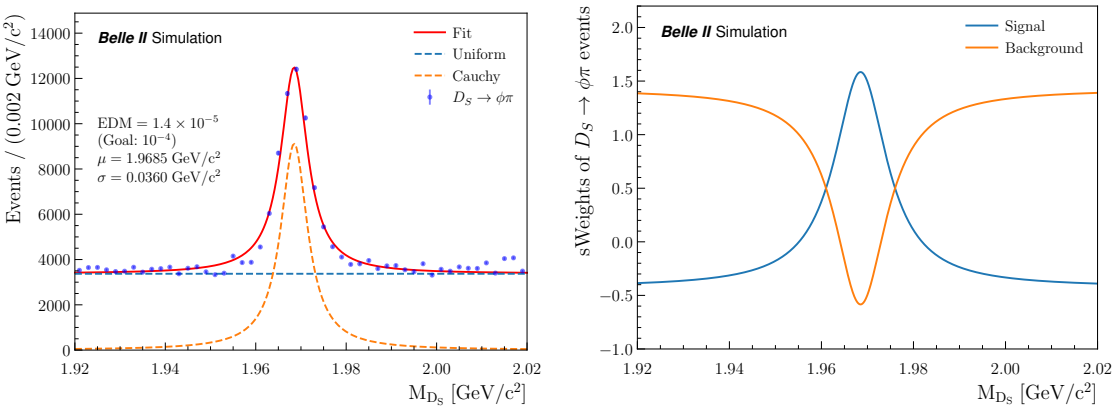


Figure 4.25: Left: binned fit of the D_s mass distribution. Right: signal and background sWeights computed from the fit to the D_s mass.

variables is shown in Figure 4.26, while others can be found in Appendix D, Figure D.1.

Table 4.18: Variables and corresponding ranges used to define corrections from the $D_s \to \phi\pi$ control channel.

Variable	Correction range
$p^{T,CM}_{third}$ [GeV/c]	[0.1, 1.3]
$p^{T,CM}_{sub\text{-}lead}$ [GeV/c]	[0.4, 1.8]
$p^{T,CM}_{lead}$ [GeV/c]	[0.8, 3.4]
E^{CM}_{lepton} [GeV]	[0.5, 3.3]
$\cos\theta^{thrust}_{\tau-z}$	[0, 1]
Number of neutral hadrons in masked ROE	[0, 6]
$\cos\theta_{lepton-miss}$	[-1, 1]
CLEO thrust 0 [GeV/c]	[0, 7]

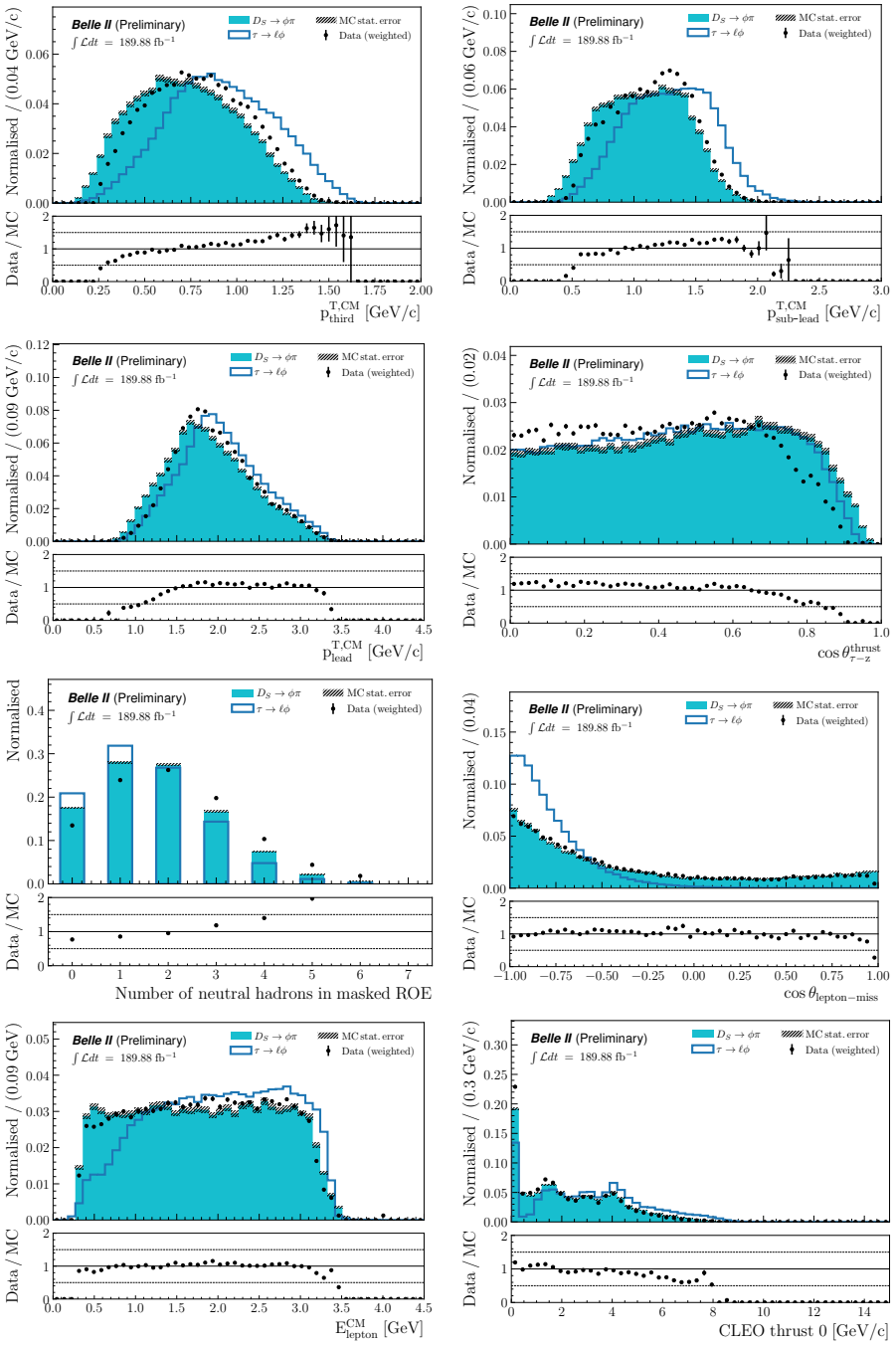


Figure 4.26: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables used to define corrections.

We reweight the events according to the following iterative procedure:

1. Taking one variable v from Table 4.18, the Data/MC ratio of the normalised distributions is computed for 50 bins inside the corresponding correction range. The binned ratio is a set of weights $w_{v,i}$ for each bin i.

2. To all events inside a bin i of the variable v, assign the weight $w_{v,i}$.

3. Repeat steps 1. and 2. with all variables listed in Table 4.18. For an event k, which belongs to a bin i_n of a variable v_n, $n = 1, ..., N$ where N is the total number of corrected variables, the final weight w_k is:

$$w_k = \prod_n w_{v_n,i_n}. \tag{4.14}$$

4. Recompute the data sWeights since there are correlations between the corrected variables and the D_s mass, and check whether the data-MC agreement is still satisfactory. If not, repeat the whole procedure.

In our case, the data-MC agreement is deemed acceptable for most of the variables even after recomputing the sWeights. The corrected distributions of variables listed in Table 4.18 are shown in Figure 4.27, the others in Appendix D, Figure D.2.

A quantitative evaluation of the agreement is possible with a Kolmogorov-Smirnov test. This test is performed thanks to the `TH1::KolmogorovTest` function from `ROOT`, which compares data and MC distributions before returning a value equivalent to a probability: close to 1, it means that the distributions are very much alike. However, the test is not valid for discrete variables and is approximate when dealing with binned distributions; that is why a large number of bins is taken, equal to 1000. The output of the test for each non-discrete variable, before and after the corrections, is given in Table 4.19.

It appears that the variables that present the worst data-MC agreements are the thrust axis magnitude, R_2 and variables related to the rest of event. Those cannot be corrected by using this control channel, as the ranges of the distributions are not similar enough to $\tau \to \ell\phi$. Instead, we consider the $\tau \to 3\pi\nu_\tau$ decay as a control channel for these variables.

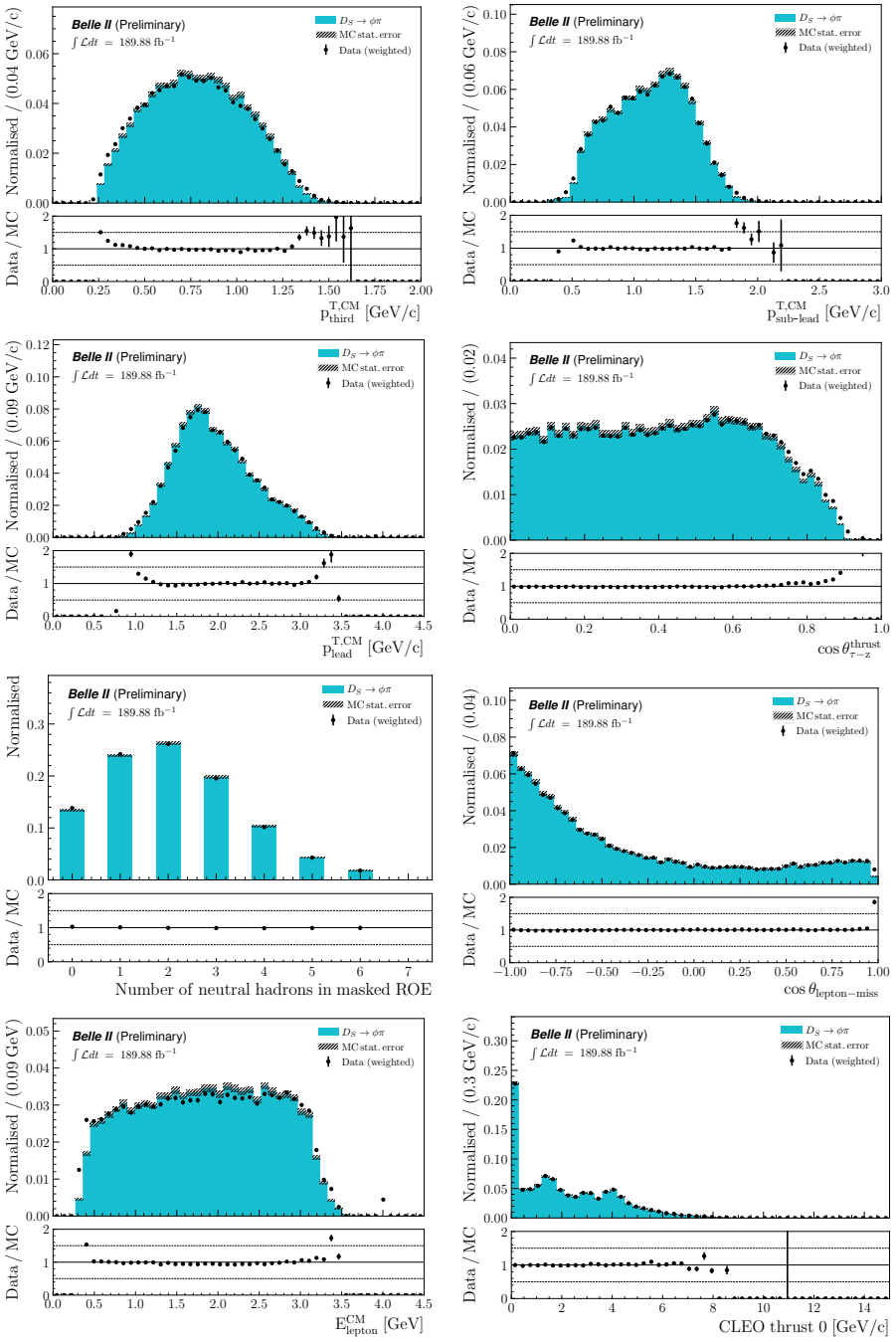


Figure 4.27: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables used to define the corrections.

Table 4.19: Results of the Kolmogorov-Smirnov tests in the $D_s \to \phi\pi$ channel for non-discrete variables, before and after applying the corrections.

Variable	K.-S. test output Before corrections	K.-S. test output After corrections
p^{T}_{lepton}	0.567	1.000
E^{CM}_{lepton}	1.000	0.997
θ^{CM}_{lepton}	0.204	1.000
$\theta_{K_1-K_2}$	1.000	1.000
$p^{T,CM}_{lead}$	0.180	1.000
$p^{T,CM}_{sub\text{-}lead}$	0.039	1.000
$p^{T,CM}_{third}$	0.008	0.999
M_{ϕ}	0.993	1.000
E_{ϕ}	0.347	0.948
p^{T}_{ϕ}	0.161	1.000
$\cos\theta^{thrust}_{\tau-z}$	0.000	1.000
$\theta_{\tau-closest}$	0.441	0.997
$\cos\theta_{p-vertex}$	1.000	1.000
χ^2 probability of ϕ vertex fit result	1.000	1.000
χ^2 probability of τ vertex fit result	1.000	1.000
M^2_{miss}	1.000	1.000
E^{CM}_{miss}	0.883	0.959
p^{T}_{miss}	0.891	0.997
θ_{miss}	0.731	1.000
$\cos\theta_{lepton-miss}$	0.991	1.000
Thrust axis magnitude	0.202	0.242
E^{CM}_{vis}	0.852	0.949
R_2	0.148	0.164
CLEO thrust 0	1.000	1.000
CLEO collision 2	1.000	1.000
ROE thrust axis magnitude	0.099	0.104
$\cos\theta^{thrust}_{\tau-ROE}$	1.000	1.000
$M_{ROE(masked)}$	0.197	0.182
$E_{ROE(masked)}$	0.312	0.252
$\Delta E_{ROE(masked)}$	0.393	0.239
$p_{ROE(masked)}$	0.996	1.000
$p^{T}_{ROE(masked)}$	0.273	1.000

4.6.6.2. Signal corrections from $\tau^{\pm} \to \pi^{+}\pi^{-}\pi^{\pm}\nu_{\tau}$ control channel

To study this control channel, we use 0.5% of the full dataset detailed in Table 2.2, thus equivalent to 0.9494 fb^{-1}, because of the large amount of reconstructed events. The simulation consists of generic (1% of 100 fb^{-1}, thus 1 fb^{-1}) and low-multiplicity (various luminosities) run-independent Monte Carlo samples (`MC14`). The tracks and event reconstruction is mostly the same as the one described in Section 4.3. The selection specific to this channel, roughly defined by comparing signal and background distributions, is listed in Table 4.20 and intends to remove a significant proportion of background contamination. In this case, signal $\tau \to 3\pi\nu_{\tau}$ events in MC are identified by truth-matching each 3-prong track to a pion, while requiring that they all originate from a same τ lepton decaying into exactly three charged particles.

Table 4.20: Cut-based background suppression selection in the $\tau \to 3\pi\nu_{\tau}$ channel.

$\text{M}_{\tau} \in [0.5, 1.7]$ GeV/c^2
$\Delta\text{E}_{\tau} \in [-5.0, 0]$ GeV
$\text{p}_{\text{lead}}^{\text{T,CM}} > 1$ GeV/c
$\text{p}_{\text{third}}^{\text{T,CM}} > 0.25$ GeV/c
τ thrust axis magnitude < 0.998
$\cos\theta_{\tau-\text{z}}^{\text{thrust}} < 0.95$
χ^2 probability of τ vertex fit result > 0.01
Number of photons from π^0 decays < 3
Number of good photons < 3
Number of good tracks < 7
Thrust axis magnitude $\in [0.85, 0.99]$
$\text{E}_{\text{vis}}^{\text{CM}} < 10$ GeV
$\text{M}_{\text{ROE(masked)}} < 5$ GeV/c^2
$\text{E}_{\text{ROE(masked)}} < 7$ GeV
Number of charged particles in masked ROE < 5
Number of photons in masked ROE < 8

Since the three pions are mainly coming from the a_1 resonance that has a broad mass distribution, it is difficult to use the sPlot method on this channel. We

perform instead a background subtraction on the data distributions using the background histograms from simulation. The variables used to define corrections are listed in Table 4.21, with the corresponding ranges.

The comparison between the signal $\tau \to 3\pi\nu_\tau$ and the background-subtracted data distributions is shown in Figure 4.28 for the variables to be corrected, while others can be found in Appendix D, Figure D.3.

The corrections are defined following the same steps as the first three ones described in Section 4.6.6.1. The distributions of the corrected variables as well as the ROE thrust axis magnitude, $\Delta E_{\text{ROE(masked)}}$ and $E_{\text{ROE(masked)}}$ are shown in Figure 4.29, the others in Appendix D, Figure D.4.

The agreement between data and simulation is also evaluated thanks to a Kolmogorov-Smirnov test, the results being presented in Table 4.22. Although the thrust axis magnitude has the highest score before $M_{\text{ROE(masked)}}$ and R_2 are corrected, its agreement worsens afterwards, thus we decide to include it in the correction process. Furthermore, we don't manage to reach a good agreement on the ROE thrust axis magnitude without compromising the one of $\text{M}_{\text{ROE(masked)}}$, which is found to be more important to the BDT predictive accuracy, that is why we decide to remove the former from the BDT inputs.

Table 4.21: Variables and corresponding ranges used to define corrections from the $\tau \to 3\pi\nu_\tau$ control channel.

Variable	Correction range
$M_{\text{ROE(masked)}}$ [GeV/c^2]	[0.0, 4.0]
R_2	[0.5, 1.0]
Thrust axis magnitude	[0.9, 1.0]

4.6.6.3. Correction of $\tau \to \ell\phi$ distributions

To estimate the systematic uncertainties on the signal efficiencies obtained after the background suppression selections, we apply the weights we obtained on signal events only, before the BDT training.

The comparisons of the $\tau \to \ell\phi$ signal distributions for M_τ, ΔE_τ and corrected variables, before and after the corrections, are shown in Figures 4.30 and 4.31. The other variables can be found in Appendix C, Figures C.3 and C.4.

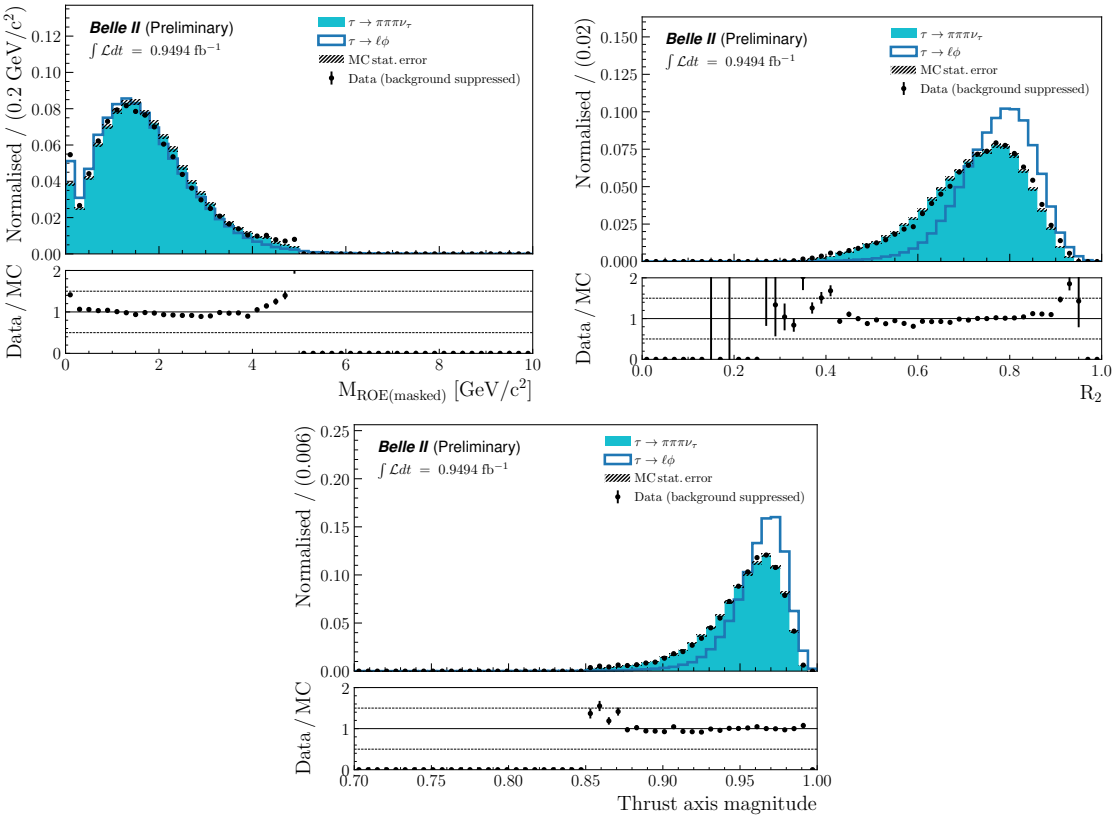


Figure 4.28: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables used to define corrections. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

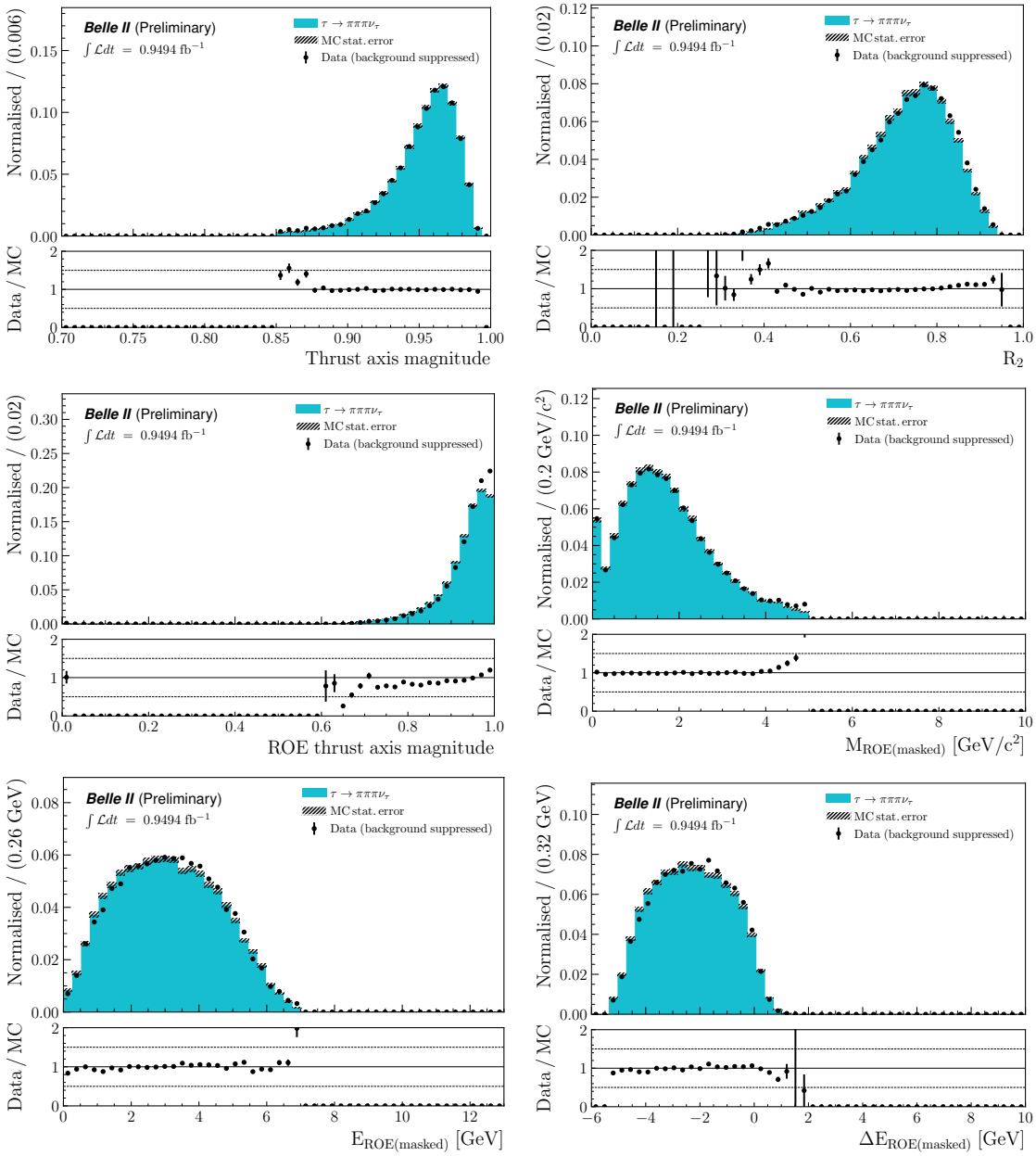


Figure 4.29: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables used to define corrections, the ROE thrust axis magnitude, $\Delta E_{\mathrm{ROE(masked)}}$ and $E_{\mathrm{ROE(masked)}}$.

Table 4.22: Results of the Kolmogorov-Smirnov tests in the $\tau \to 3\pi\nu_\tau$ channel for non-discrete variables, before and after applying the corrections.

Variable	K.-S. test output Before corrections	K.-S. test output After corrections
p^{T}_{lepton}	0.972	0.967
E^{CM}_{lepton}	0.949	0.926
θ^{CM}_{lepton}	1.000	1.000
$p^{T,CM}_{lead}$	0.997	0.990
$p^{T,CM}_{sub\text{-}lead}$	1.000	1.000
$p^{T,CM}_{third}$	1.000	1.000
$\cos\theta^{thrust}_{\tau-z}$	1.000	1.000
$\theta_{\tau-closest}$	1.000	1.000
χ^2 probability of τ vertex fit result	0.800	0.818
M^2_{miss}	1.000	1.000
E^{CM}_{miss}	1.000	1.000
p^{T}_{miss}	1.000	1.000
θ_{miss}	1.000	1.000
$\cos\theta_{lepton-miss}$	1.000	1.000
Thrust axis magnitude	1.000	1.000
E^{CM}_{vis}	1.000	1.000
R_2	0.893	0.996
CLEO thrust 0	1.000	1.000
CLEO collision 2	1.000	1.000
ROE thrust axis magnitude	0.588	0.920
$\cos\theta^{thrust}_{\tau-ROE}$	1.000	1.000
$M_{ROE(masked)}$	0.996	1.000
$E_{ROE(masked)}$	1.000	0.999
$\Delta E_{ROE(masked)}$	1.000	0.999
$p_{ROE(masked)}$	1.000	1.000
$p^{T}_{ROE(masked)}$	1.000	1.000

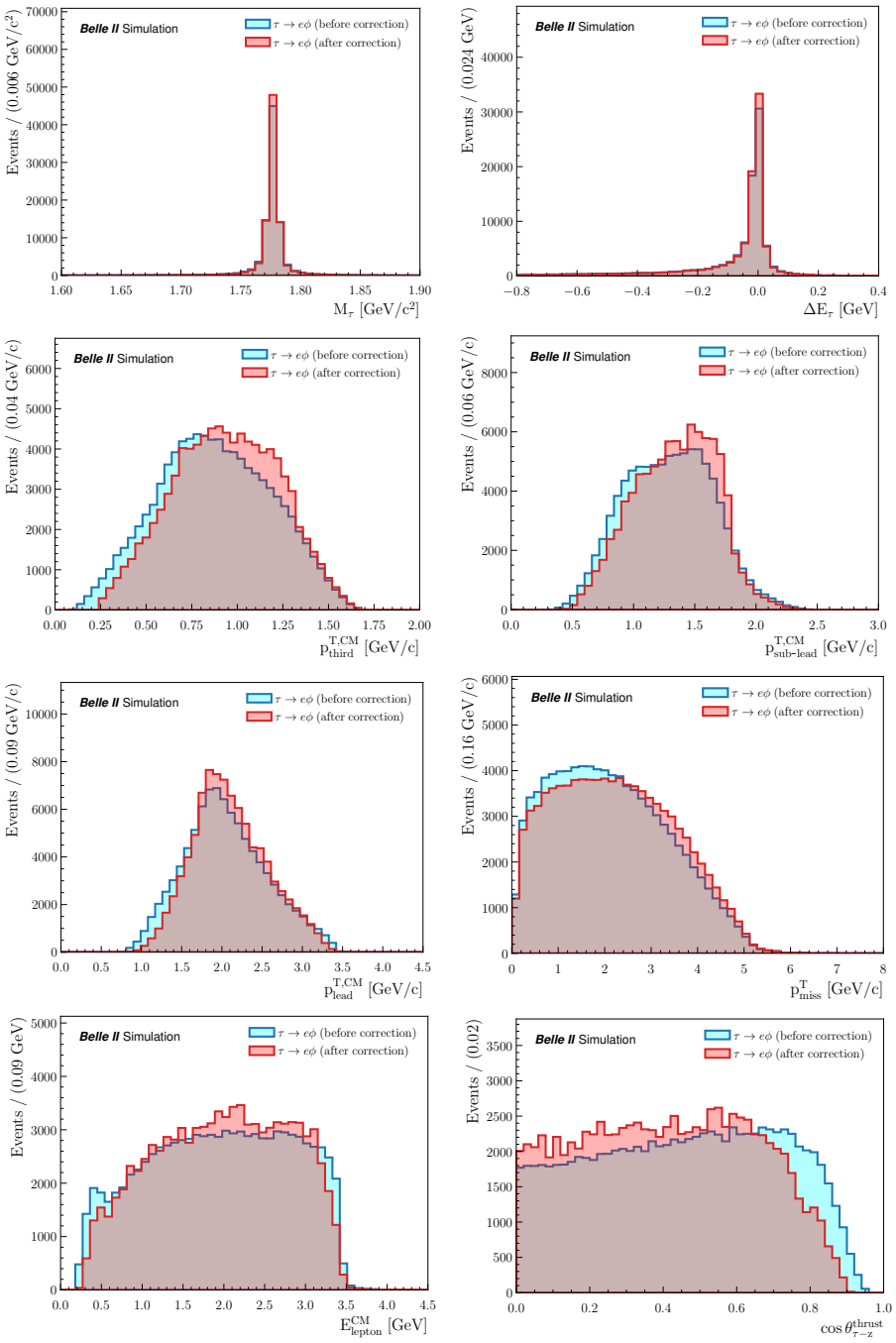


Figure 4.30: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and corrected variables, before and after the corrections are applied.

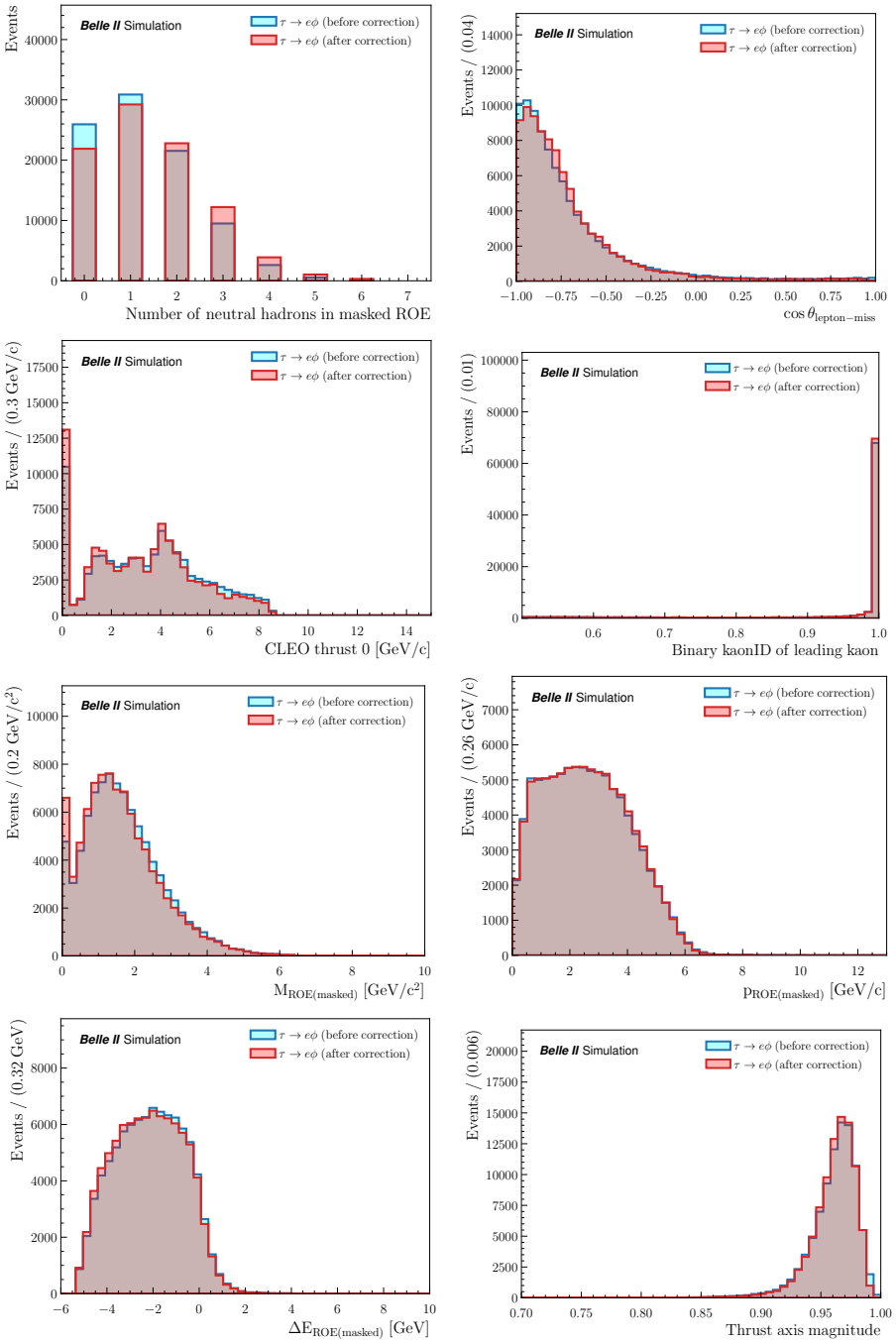


Figure 4.30: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and corrected variables, before and after the corrections are applied.

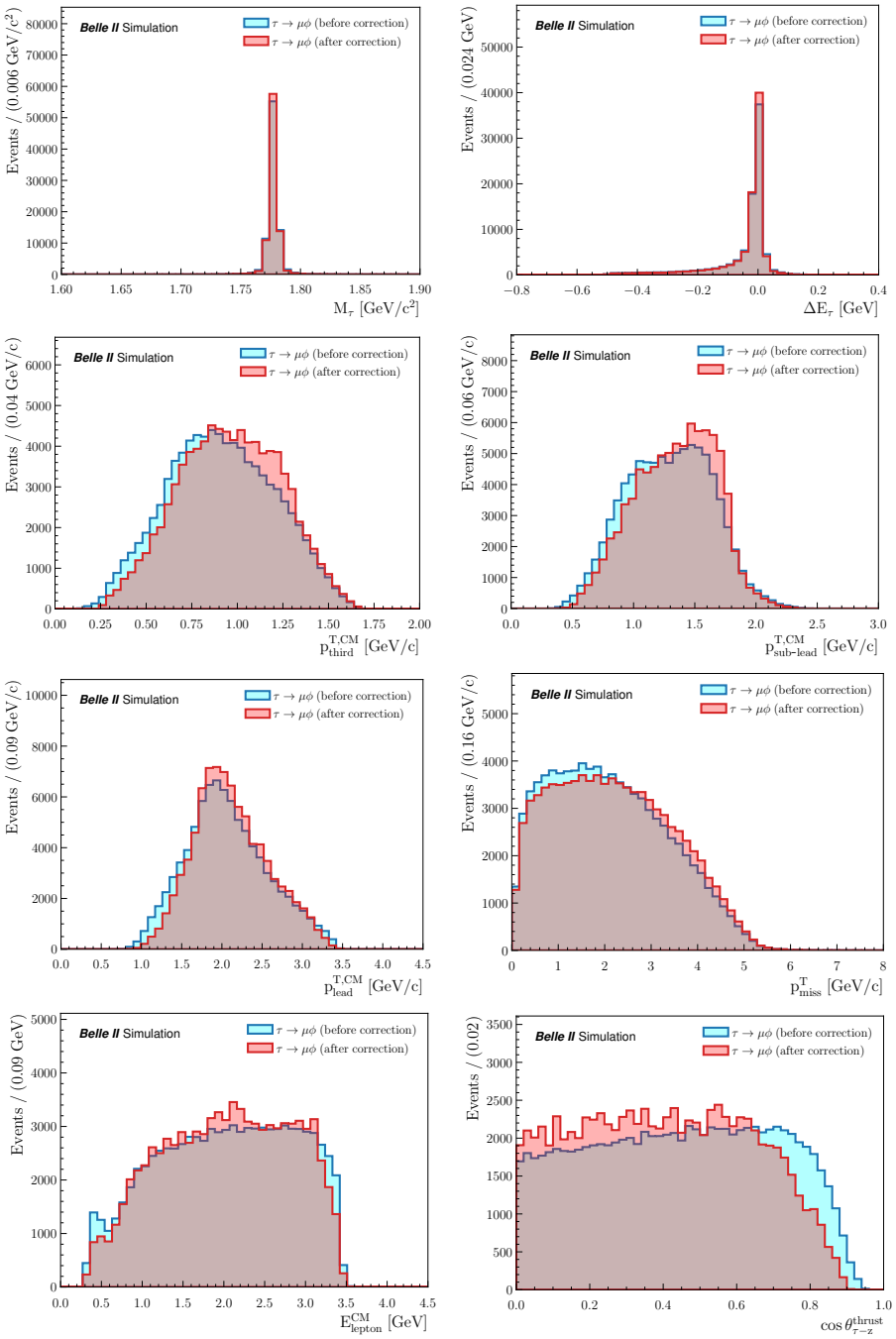


Figure 4.31: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and corrected variables, before and after the corrections are applied.

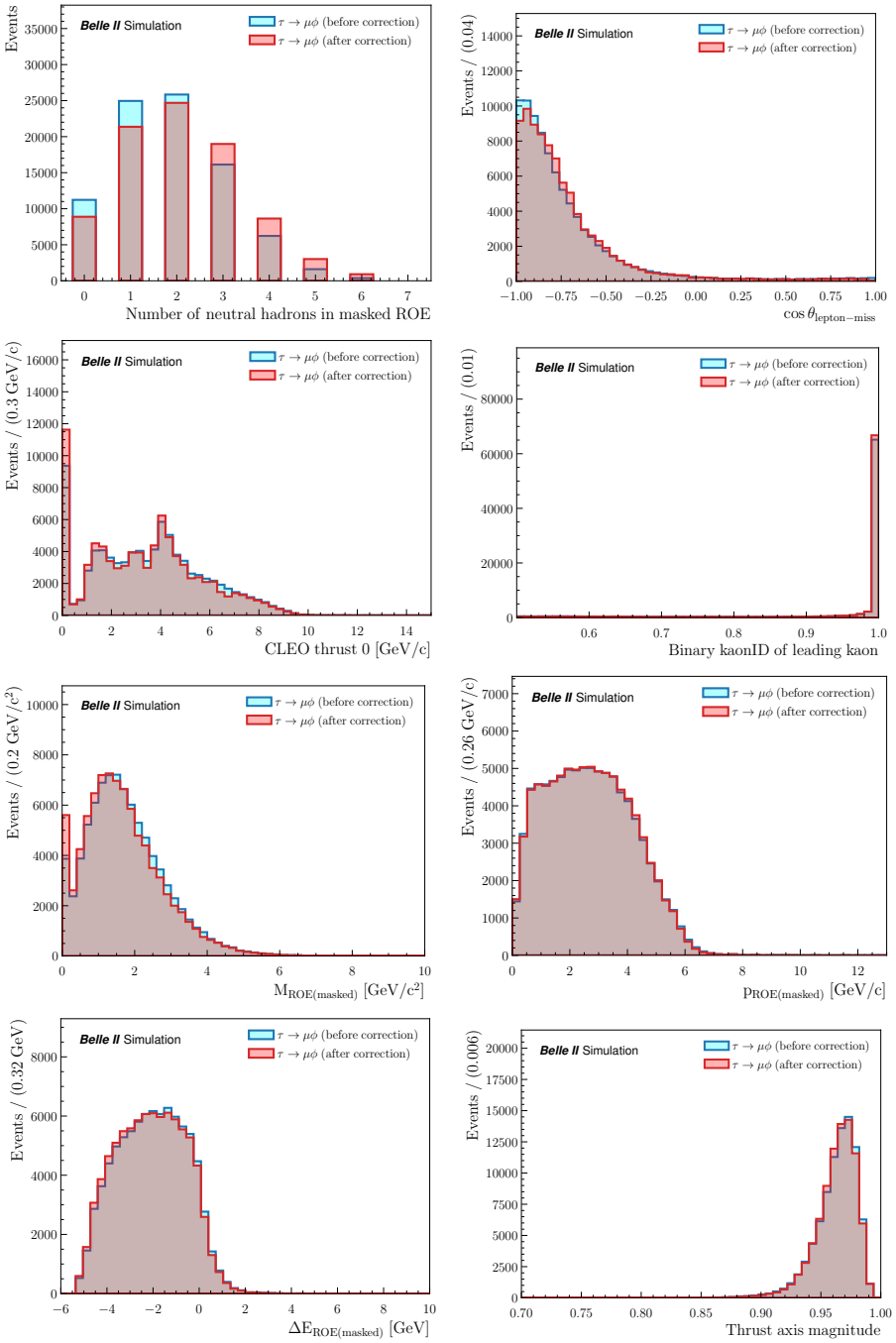


Figure 4.31: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and corrected variables, before and after the corrections are applied.

4.6.6.4. Results of the background suppression strategy after corrections

Following the correction of signal distributions, we train a BDT according to the strategy explained in Section 4.4.4, but this time taking into account the additional weights.

The comparison of the signal probability distributions according to the BDT in the train and validation samples is shown in Figure 4.32. The lower selections on the probabilities are kept the same as is in the previous section, 0.98 and 0.84 in the $\tau \to e\phi$ and $\tau \to \mu\phi$ channels, respectively. The evolution of the logarithmic losses in Figure 4.33 demonstrates the absence of overtraining. The ranking by importance of variables used in the BDT is given in Appendix C.

The MC signal efficiencies, background yields and sensitivities (Punzi figure of merit) after the complete selection are presented in Tables 4.23 and 4.24. The details of the remaining background candidates truth-matched to specific final states are given in Table 4.25.

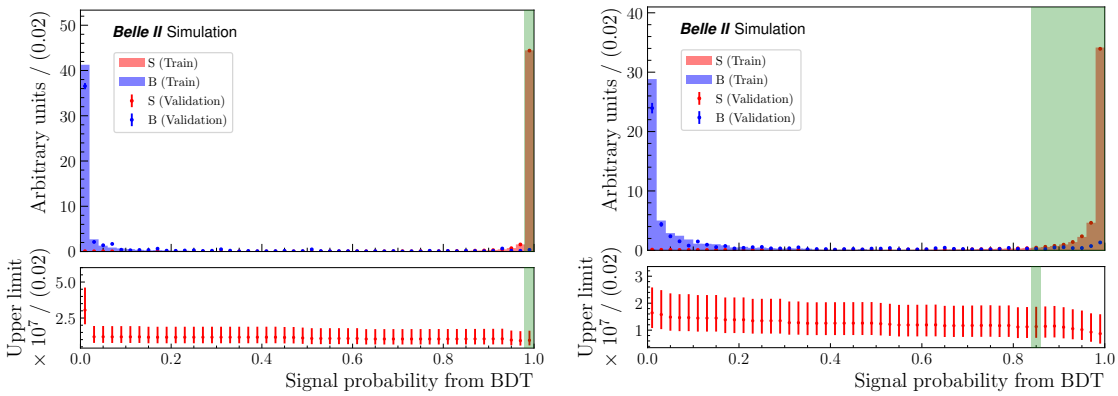


Figure 4.32: Probability of being a signal event according to the BDT, in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) channels, with weighted signal events. S means signal and B means background. The lower inset plot shows the 90% C.L. upper limit on the LFV channel's branching fraction, as evaluated on the validation sample. The green area corresponds to values selected as a result of the upper limit optimisation in Section 4.4.6.

We explained in Section 4.5 the need for an additional selection on $\tau \to e\phi$ events, due to an excess of data observed in the sidebands. This lead to a second BDT trained on data and signal simulation, of which we evaluate the contribution to the systematic uncertainties by retraining it once again with corrected signal distributions. The distributions of MC and data as a function of the corrected BDT output is shown in Figure 4.34; we apply a veto on the resulting probability at 0.001.

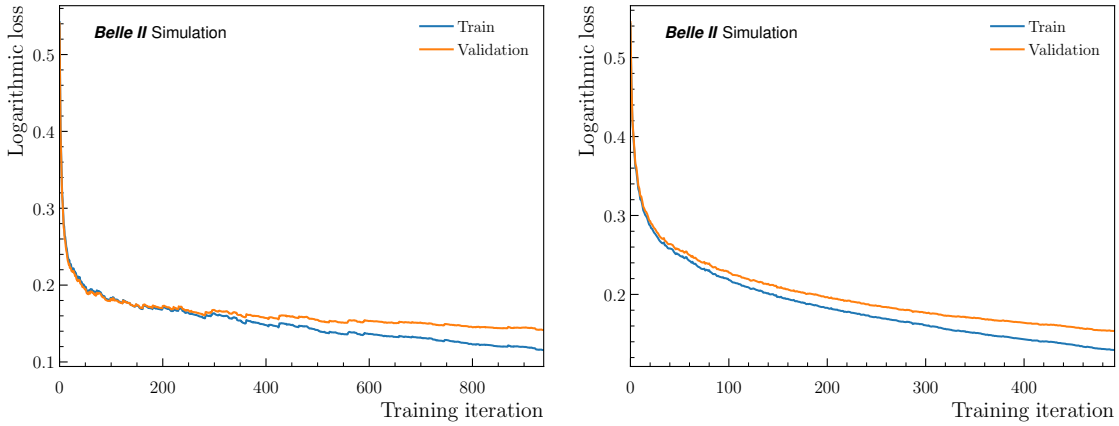


Figure 4.33: Logarithmic loss as a function of the training iteration, in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) channels, when signal events are weighted. The ranges go up to the last iteration allowed by the early stopping.

Table 4.23: $\tau \to e\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions and when signal events are weighted. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	Low multi.	Total backgr.	FOM
Train	20σ	8.3%	0	0.05	0	0.05	0.048
	3σ	8.1%	0	0	0	0	0.054
Validation	20σ	8.3%	0.77	3.95	5.14	9.86	0.018
	3σ	8.1%	0.24	0.52	0	0.76	0.034
Test	20σ	8.3%	3.62	3.21	0	6.83	0.020
	3σ	8.1%	0.47	0.22	0	0.69	0.035

Table 4.24: $\tau \to \mu\phi$ signal efficiencies, background yields and sensitivities for the train, validation and test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions and when signal events are weighted. The yields are rescaled to 189.88 fb^{-1}.

Sample	Region	Signal efficiency	τ-pair	$q\bar{q}$	Low multi.	Total backgr.	FOM
Train	20σ	8.4%	0.24	0	0	0.24	0.042
	3σ	8.2%	0	0	0	0	0.055
Validation	20σ	8.4%	2.37	8.23	0	10.6	0.018
	3σ	8.2%	0.61	0.87	0	1.48	0.030
Test	20σ	8.3%	2.57	6.22	0.53	9.32	0.018
	3σ	8.2%	0.38	0.69	0	1.07	0.032

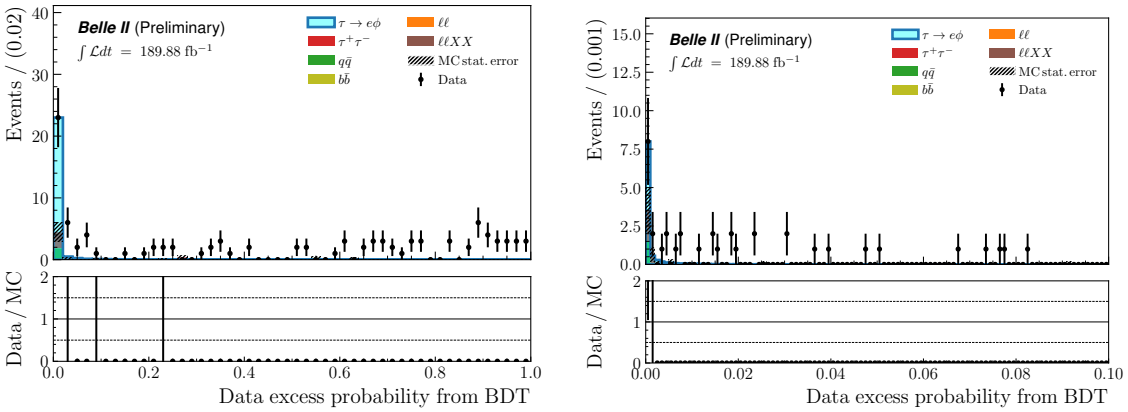


Figure 4.34: Probability, in the $[0, 1]$ (left) and $[0, 0.1]$ (right) ranges, for $\tau \to e\phi$ MC and data events to be in the data excess region, after the preselection, BDT, 2D and kaon's electronID selections, when signal events are weighted. Signal distributions are rescaled to the data bin with highest yield.

Table 4.25: Numbers of remaining background candidates truth-matched to a final state in the $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right) test samples after the preselection, the BDT and 2D selections, inside the 20σ and 3σ regions and when signal events are weighted. The yields correspond to the integrated luminosities cited in Table 4.2 and don't include PID corrections.

Final state	Region	τ-pair	$q\bar{q}$	Low multi.
KKK	20σ	0	9	0
	3σ	0	0	0
KKπ	20σ	0	2	0
	3σ	0	0	0
K$\pi\pi$	20σ	1	4	0
	3σ	0	1	0
$\pi\pi\pi$	20σ	12	0	0
	3σ	0	0	0
eKK	20σ	0	3	0
	3σ	0	0	0
eeK	20σ	0	1	0
	3σ	0	0	0
e$\pi\pi$	20σ	0	1	0
	3σ	0	0	0
eeπ	20σ	5	1	0
	3σ	4	0	0

Final state	Region	τ-pair	$q\bar{q}$	Low multi.
KKK	20σ	0	32	0
	3σ	0	5	0
KKπ	20σ	4	18	0
	3σ	1	2	0
K$\pi\pi$	20σ	7	8	0
	3σ	2	1	0
$\pi\pi\pi$	20σ	28	1	0
	3σ	2	0	0
μKK	20σ	0	1	0
	3σ	0	0	0
$\mu\pi\pi$	20σ	1	0	0
	3σ	1	0	0

The final signal efficiencies in the signal region, after signal events are weighted, are approximately 6.5% ($e\phi$) and 8.2% ($\mu\phi$). The systematic uncertainty coming from the BDT selection is taken as the absolute difference between these results and the efficiencies presented in Section 4.6.6.4, Tables 4.13 and 4.14. This corresponds to 0.48% and 0.95% of the $\tau \to e\phi$ and $\tau \to \mu\phi$ signal efficiencies, respectively.

4.6.7. Systematic uncertainty summary

A summary of the systematic uncertainty contributions accounted for in the 90% CL exclusion upper limit computation is shown in Table 4.26.

Table 4.26: Systematic uncertainties affecting the quantities used in the upper limit computation, depending on the decay mode. The measured values on the particle identification combine both statistical and systematic uncertainties.

Affected quantity	Source	Mode	
		$e\phi$	$\mu\phi$
$\varepsilon_{\ell\phi}$	Particle identification	0.11%	0.16%
	Tracking efficiency	0.52%	
	Trigger efficiency	0.4%	0.9%
	BDT selection	0.48%	0.95%
L	Luminosity	0.6%	
$\sigma_{\tau\tau}$	Tau-pair cross section	0.003 nb	
N_{exp}	Momentum scale	$^{+0.0}_{-0.0}$	$^{+0.0}_{-0.2}$

4.7. Limit estimate

An upper limit estimation is performed from the results obtained on simulation and data, using the CLs method with `pyhf`.

The number of expected events n follows a Poisson distribution, the expectation value of which is

$$E[n] = \mu s + b, \tag{4.15}$$

where μs (b) is the signal (background) yield. Since the signal yield is unknown but can be expressed as

$$\mu s = L \times 2\sigma_{\tau\tau} \times \varepsilon_{\ell\phi} \times \mathcal{B}(\tau \to \ell\phi), \tag{4.16}$$

we assign $\mu = \mathcal{B}(\tau \to \ell\phi)$ and $s = 2L\sigma_{\tau\tau}\varepsilon_{\ell\phi}$, μ being the parameter of interest that will be estimated.

This parameter is evaluated for 50 values evenly distributed between 10^{-8} and 10^{-6}. A 90% C.L. upper limit can be computed by setting $\mathrm{CL}_s = 10\%$, where

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{4.17}$$

CL_{s+b} and CL_b are respectively the "signal + background" and "background only" hypotheses p-values, obtained from the asymptotic formula [11], this for all values of μ. We consider that $\mu \geq 0$, thus we use the $\tilde{q}_\mu$ test statistic assuming the Wald approximation.

The signal efficiency and expected data yield in the 3σ region, which we use in this computation, are cited in Table 4.16. On top of the systematic uncertainties described in the previous section, we evaluate statistical uncertainties on:

- the signal efficiency $\varepsilon_{\ell\phi} = w_f n_f / n_i$, as the binomial error

$$\frac{w_f}{n_i}\sqrt{n_f\left(1 - \frac{n_f}{n_i}\right)}, \tag{4.18}$$

  where n_i is the initial (generated) signal yield, n_f the final (selected) *unweighted* signal yield and w_f is the total weight applied to n_f;

- the expected number of events in the signal region $N_{\mathrm{exp}} = \alpha N_{\mathrm{obs}}$ ($= b$ in Equation 4.15), where N_{obs} is the data yield observed in the $\pm 3\sigma$ ΔE_τ sideband and α is MC background ratio (see Table 4.16), as the combination via propagation of uncertainties of their respective Poisson distribution's error.

All uncertainties, statistical or systematic, related to a same quantity are then combined in quadrature. The total uncertainties are listed in Table 4.27: in the case of the signal input $L \times \sigma_{\tau\tau} \times \varepsilon_{\ell\phi}$, the final value is obtained via propagation

of errors.

Table 4.27: Total uncertainties (statistical and systematic) affecting the signal and background inputs of the upper limit computation, depending on the decay mode.

Affected quantity	Mode	
	$e\phi$	$\mu\phi$
$L \times \sigma_{\tau\tau} \times \varepsilon_{\ell\phi}$	8.25%	13.39%
N_{exp}	$^{+0.4}_{-0.4}$	$^{+0.4}_{-0.5}$

The expected CL_s as a function of the upper limit is displayed in Figure 4.35: at 90% C.L., the expected upper limits are

$$\mathcal{B}^{90}_{\text{UL,exp}}(\tau \to e\phi) = 1.18 \times 10^{-7}, \tag{4.19}$$
$$\mathcal{B}^{90}_{\text{UL,exp}}(\tau \to \mu\phi) = 1.05 \times 10^{-7}. \tag{4.20}$$

Comparing these results to the previous expected limits by Belle and BaBar, cited in Table 4.1, our expected limit for $\tau \to \mu\phi$ is 30% higher than the one from BaBar and close to twice as much as the Belle one. However, at equivalent luminosities, we expect the same limit as Belle (854 fb^{-1}) and one slightly better than BaBar (-17%) at 451 fb^{-1}. After unblinding, an upper limit on $\tau \to \mu\phi$ from Belle II could lie between the observed limits from BaBar and Belle, if no unexpected data excess is found inside the signal region.

For the $\tau \to e\phi$ channel, the situation is less optimistic since the limit we would expect at the same luminosities is 53% (29%) larger than the BaBar (Belle) one. This can be attributed to the poor description of the background by the simulation, as revealed by the sideband comparison. The selection used for those searches consists only of a loose cut on the leading kaon's kaonID, since a more stringent requirement would lower the signal efficiency too much. A better kaonID would certainly help improving the analysis performance.

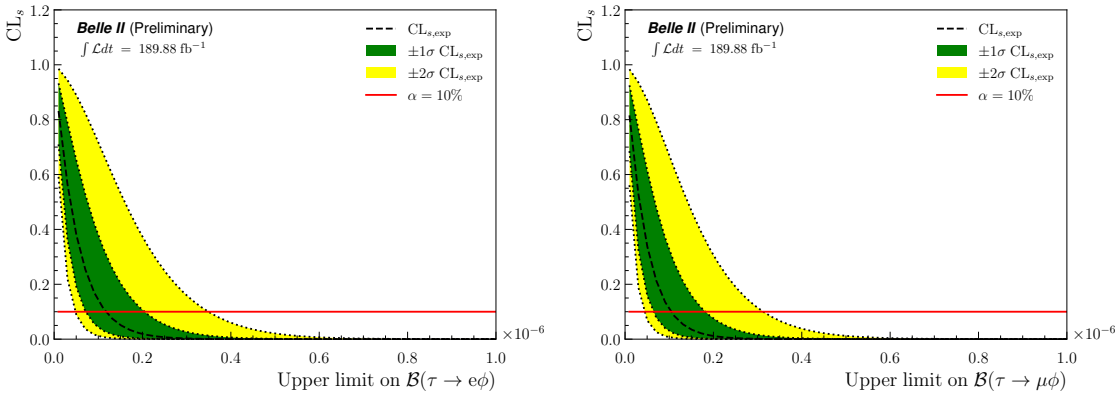


Figure 4.35: Expected CL_s as a function of the upper limit on the branching fraction of $\tau \to e\phi$ (left) and $\tau \to \mu\phi$ (right). The red lines correspond to the 90% confidence level.

Conclusion

We carried out the search for lepton flavour violating $\tau \to \ell\phi$ decays ($\ell = e, \mu$) with the Belle II experiment, using 189.88 fb^{-1} of data, collected from 2019 to 2021 at the SuperKEKB electron-positron collider. This analysis relies on an "inclusive" approach, where only one tau produced in $e^+e^- \to \tau^+\tau^-$ events is reconstructed so that it corresponds to the signal decay, the other tau being left unconstrained. The purpose of this strategy is to increase the signal efficiency compared to a tagged reconstruction. Furthermore, we devise cut- and boosted-decision-trees-based selections, defined on Monte Carlo simulation samples in such a way that it maximises the background rejection and signal efficiency inside the signal region. This region corresponds to three times the resolution (3σ) around the expected signal peak in the $(M_\tau, \Delta E_\tau)$ space.

We estimated upper limits on the branching fractions of both lepton decay modes according to the CLs method, evaluated at 90% confidence level, and without unblinding the 5σ region. The expected upper limits before full unblinding are 1.18×10^{-7} for $\tau \to e\phi$ and 1.05×10^{-7} for $\tau \to \mu\phi$. Compared to the previous results obtained by the BaBar and Belle experiments, our expected upper limit on the $\tau \to \mu\phi$ branching fraction is better than the one of the former (-17%) and equal to the one of the latter, at equivalent luminosities. On the other hand, the limit obtained for $\tau \to e\phi$ is larger than what was achieved by both BaBar and Belle.

The presence of mismodelled data events hindered the good performance one could have expected in the electron decay mode analysis, knowing that the expected signal efficiencies and background yields in simulation are similar between both decay modes. In addition, larger sets of simulated background events may have helped in enhancing the signal and background separation performed by the BDT. Those are two of the points that should be solved in order to improve this study in the next years; in particular, Belle II expects to record 50 ab^{-1} of data by $\sim$2035 (see Figure 4.36), therefore it will also be necessary to cope with higher amounts of beam-background events by reoptimising the selections.

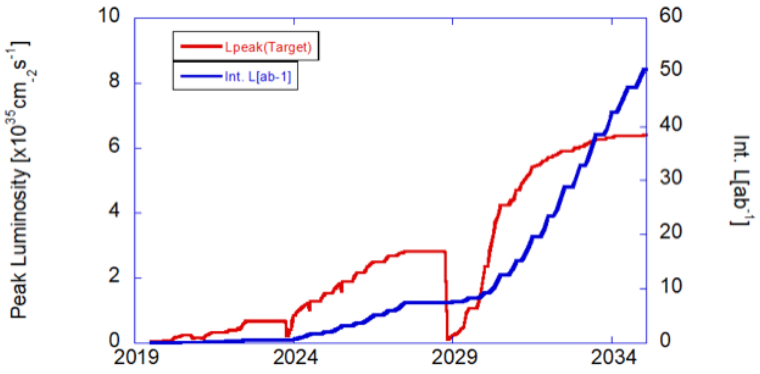


Figure 4.36: Instantaneous (red, left y-axis) and integrated (blue, right y-axis) luminosity projections up to 2035.

Finally, we measured the rate of fake tracks with $e^+e^- \to \tau^+\tau^-$ events at Belle II, using 189.88 fb^{-1} of data as well. We reconstructed the τ-pairs as $\tau^\pm [\to \pi^+\pi^-\pi^\pm\nu_\tau] + \tau^\mp [\to e^\mp\bar{\nu}_e\nu_\tau]$, with the possibility of an additional fake track. After we established a background rejection and signal retention selection on Monte Carlo simulation, we computed the fake rate as $N_5/(N_4 + N_5)$, where N_4 (N_5) is the number of remaining events with 4 (5) tracks. We found the rate to be around 1% in both data and simulation. This measurement can be used as an indicator of the tracking performance: here, given that the rate is low, the amount of fake and clone tracks shouldn't be worrisome, but could be so in the future with the expected increase of background events.

Appendices

A. L1 trigger lines

In Chapters 3 and 4, we consider the ECL low-multiplicity trigger lines *lmlX* and some CDC lines used as reference for the trigger efficiency studies, defined in Tables A.1 and A.2, respectively.

The *lmlX* trigger lines have remained active throughout the considered data taking period (see Section 2.7, Table 2.2). Some lines have been prescaled in the meantime, the details being summarised in Table A.3.

The implemented trigger logic for the low-multiplicity lines is further described in [76] and the performance of the used triggers has been extensively studied within the Tau Working Group of Belle II in [77].

Table A.1: Definitions of L1 low-multiplicity trigger lines (*lmlX*) [76, 77]. "CM" means center of mass frame.

Trigger line	ECL cluster(s)	Energy of cluster(s) E	Polar angle of cluster(s) θ	Opening angle in CM $\Delta\phi^{\mathrm{CM}}$	Other requirements
lml0	≥ 3	≥ 1 with $E > 300$ MeV	$[18.5° \,;\, 139.3°]$	–	not an ECL Bhabha
lml1	≥ 1	≥ 1 with $E^{\mathrm{CM}} > 2$ GeV	$[32.2° \,;\, 124.6°]$	–	–
lml2	≥ 1	≥ 1 with $E^{\mathrm{CM}} > 2$ GeV	$[18.5° \,;\, 32.2°]$ or $[124.6° \,;\, 139.3°]$	–	not an ECL Bhabha
lml4	≥ 1	≥ 1 with $E^{\mathrm{CM}} > 2$ GeV	$< 18.5°$ or $> 139.3°$	–	not an ECL Bhabha
lml6	1	$E^{\mathrm{CM}} > 1$ GeV	$[32.2° \,;\, 128.7°]$	–	no other cluster with $E > 300$ MeV
lml7	1	$E^{\mathrm{CM}} > 1$ GeV	$[18.5° \,;\, 31.9°]$ or $[128.7° \,;\, 139.3°]$	–	no other cluster with $E > 300$ MeV
lml8	2	$E > 250$ MeV	–	$[170° \,;\, 190°]$	no 2 GeV (CM) cluster
lml9	2	$E > 250$ MeV $E < 250$ MeV	–	$[170° \,;\, 190°]$	no 2 GeV (CM) cluster
lml10	2	–	$\Sigma\, \theta^{\mathrm{CM}} \in [160° \,;\, 200°]$	$[160° \,;\, 200°]$	no 2 GeV (CM) cluster
lml12	≥ 3	≥ 1 with $E > 500$ MeV	$[18.5° \,;\, 139.3°]$	–	not an ECL Bhabha
lml13	1	$E^{\mathrm{CM}} > 500$ MeV	$[44.2° \,;\, 94.8°]$	–	no other cluster with $E > 300$ MeV

Table A.2: Definitions of the relevant CDC trigger lines [76].

Trigger line	Full track(s)	Short track(s)	Neuro 3D track(s) with $\lvert z\rvert < 20$ cm	Opening angle $\Delta\phi$	Other requirements
fff	≥ 3	–	–	–	–
ffs	≥ 2	≥ 1	–	–	–
ffy	≥ 3	–	≥ 1	–	–
ffo	≥ 2	–	–	$> 90°$	not an ECL Bhabha
fyo	≥ 2	–	≥ 1	$> 90°$	not an ECL Bhabha
ff30	≥ 2	–	–	$> 30°$	not an ECL Bhabha

Table A.3: Prescale changes to low-multiplicity trigger lines after the start of Belle II data taking [78, 79].

Trigger line	Prescale change	Experiment	Run
lml1	$1 \rightarrow 2$	18	66
lml4	$1 \rightarrow 10$	18	66
lml0	$1 \rightarrow 10$	20	915
lml13	$1 \rightarrow 10$	20	915

B. Feature importance in BDT

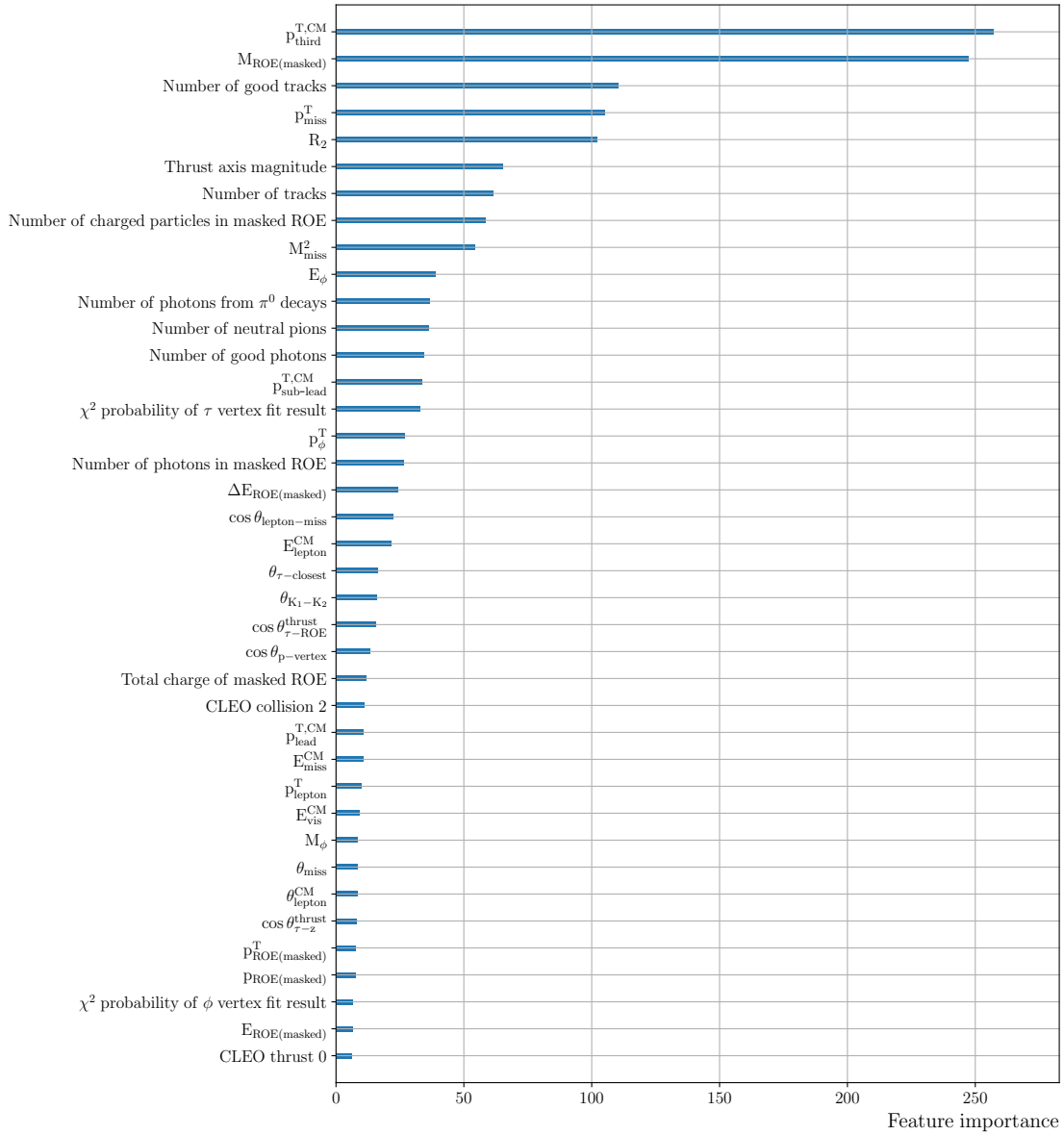


Figure B.1: Feature importance for the BDT predictive accuracy, in the $\tau \to e\phi$ channel. The importance of a feature is computed as "the average gain of splits which use the feature" [70].

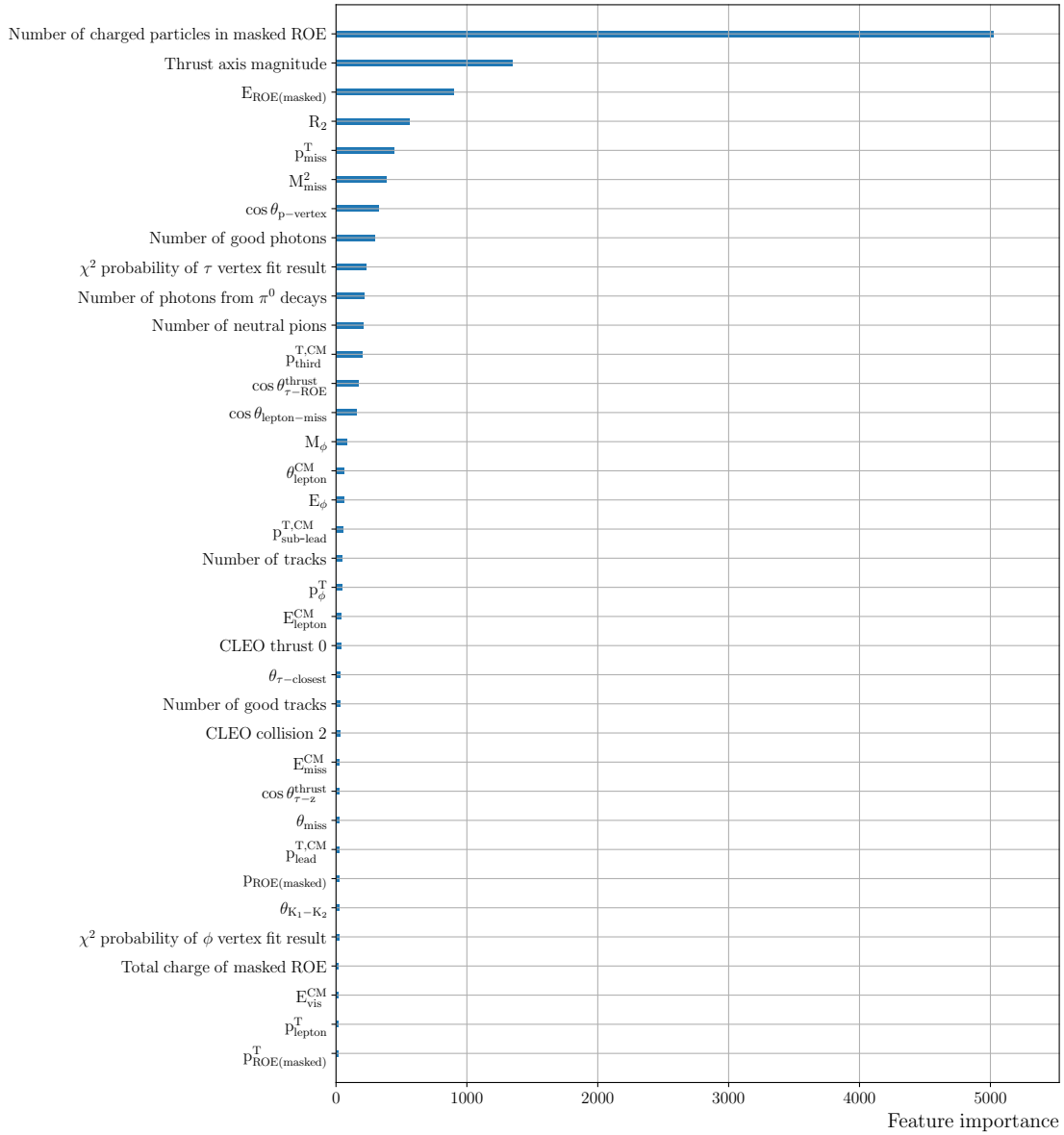


Figure B.2: Feature importance for the BDT predictive accuracy, in the $\tau \to \mu\phi$ channel. The importance of a feature is computed as "the average gain of splits which use the feature" [70].

C. Corrected signal distributions

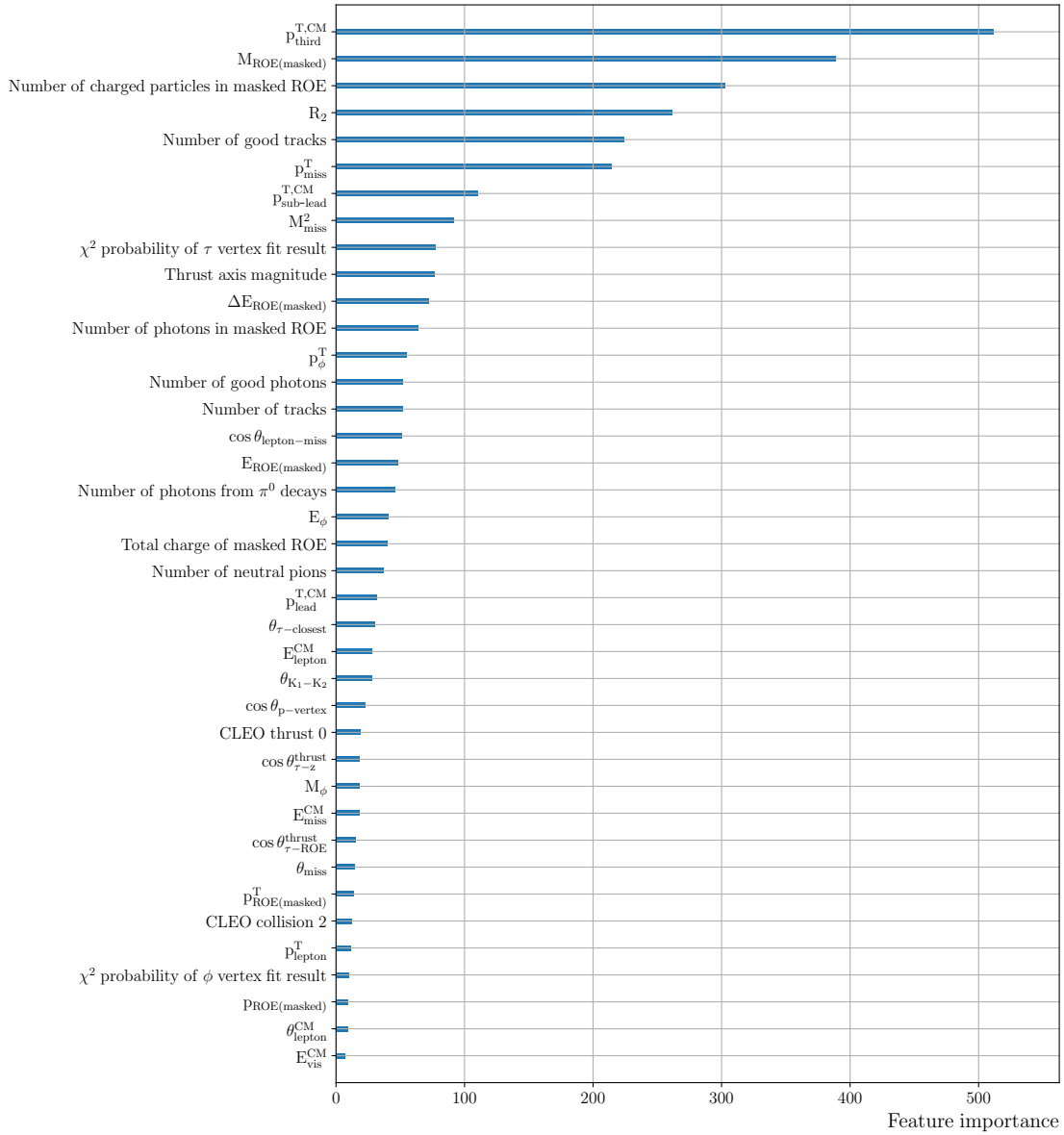


Figure C.1: Feature importance for the BDT predictive accuracy, in the $\tau \to e\phi$ channel, with weighted signal distributions. The importance of a feature is computed as "the average gain of splits which use the feature" [70].

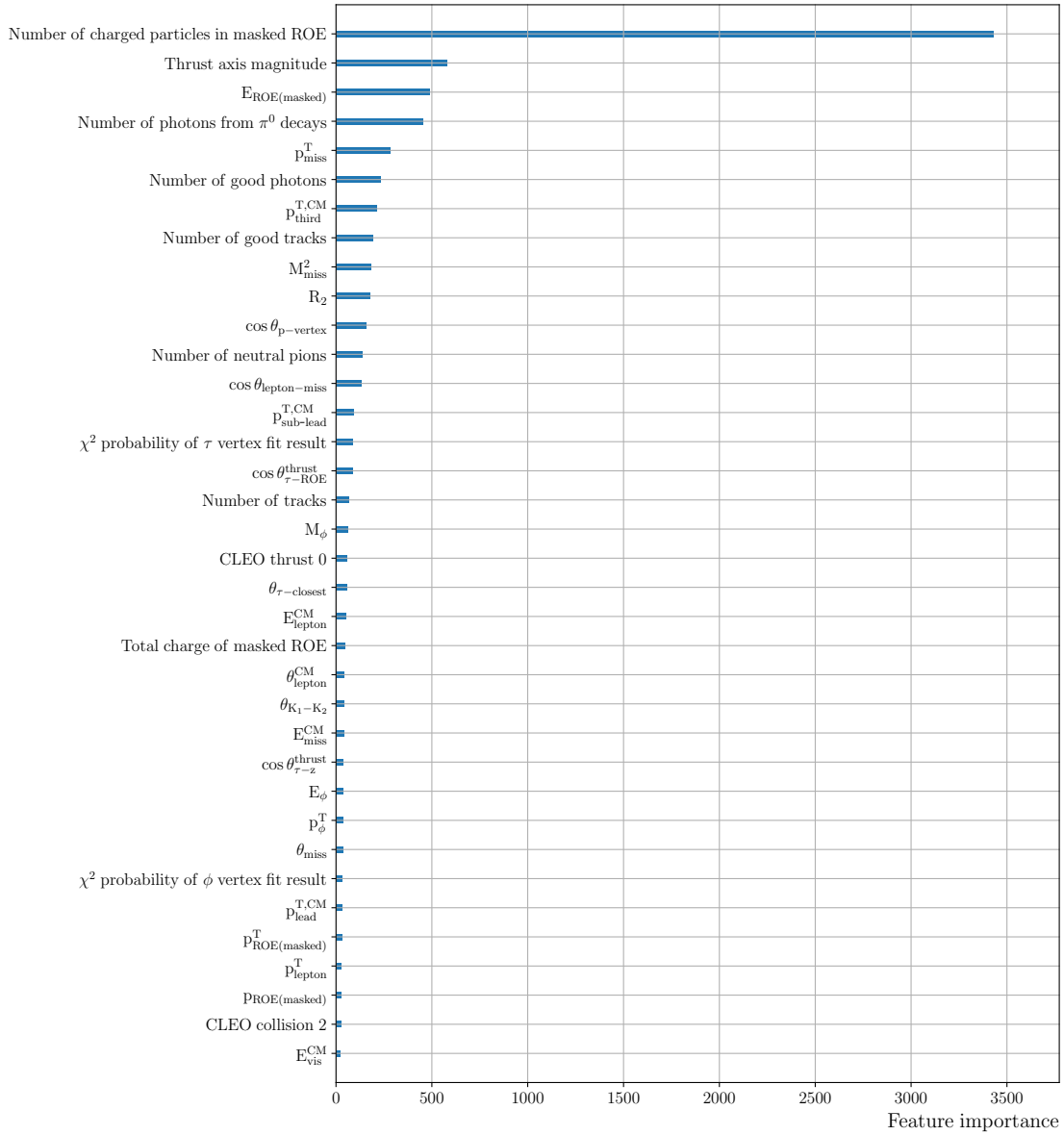


Figure C.2: Feature importance for the BDT predictive accuracy, in the $\tau \to \mu\phi$ channel, with weighted signal distributions. The importance of a feature is computed as "the average gain of splits which use the feature" [70].

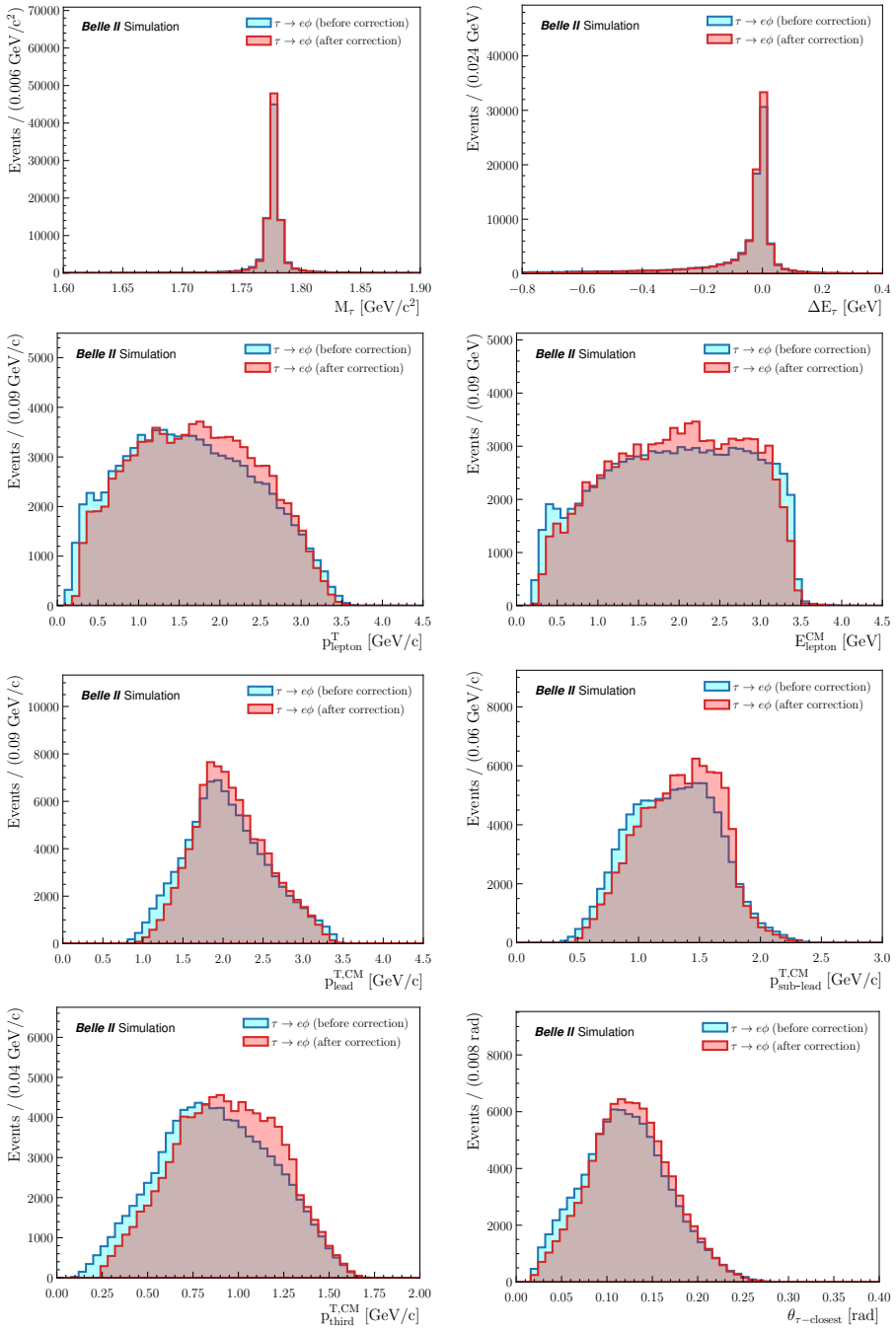


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

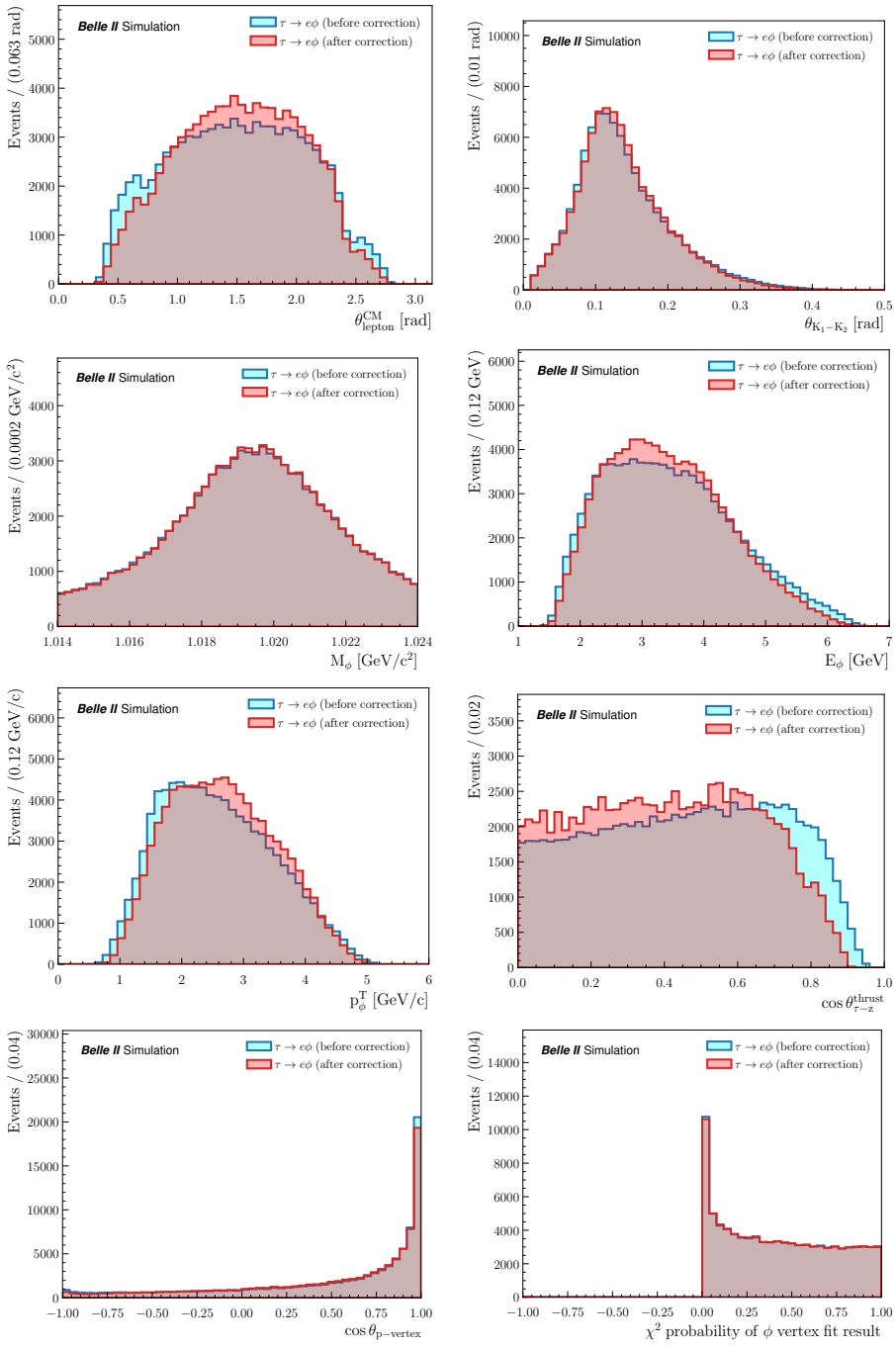


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

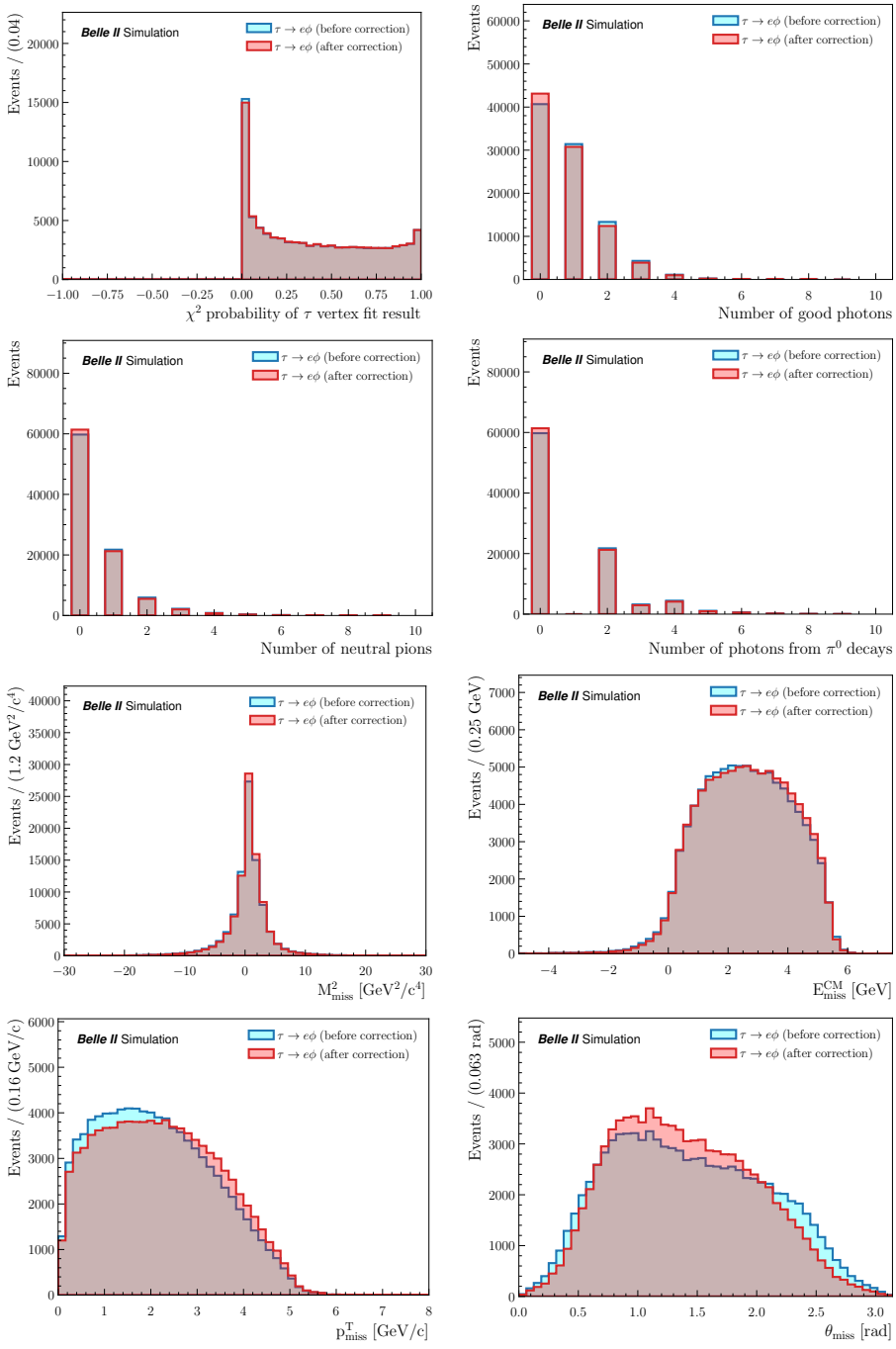


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

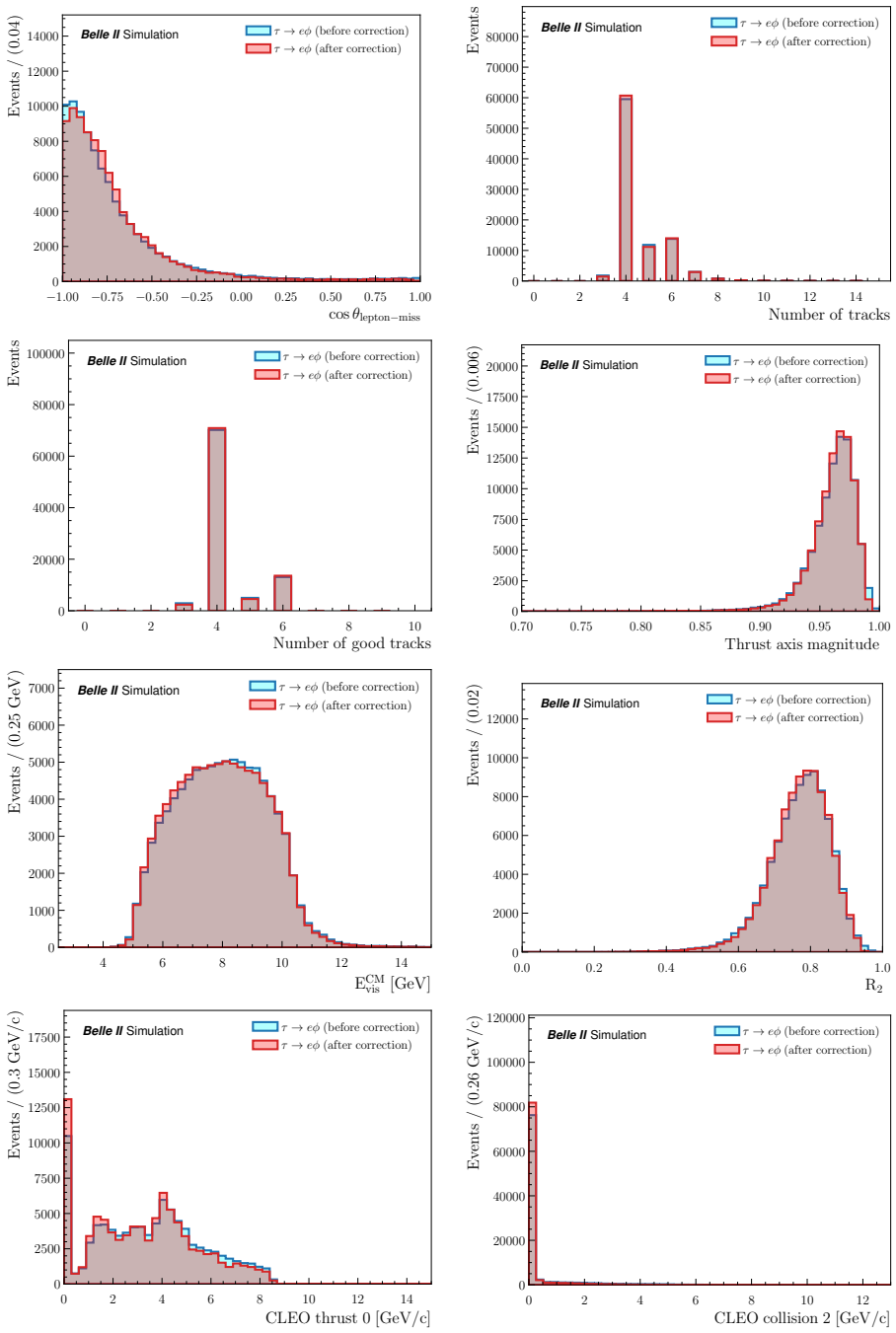


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

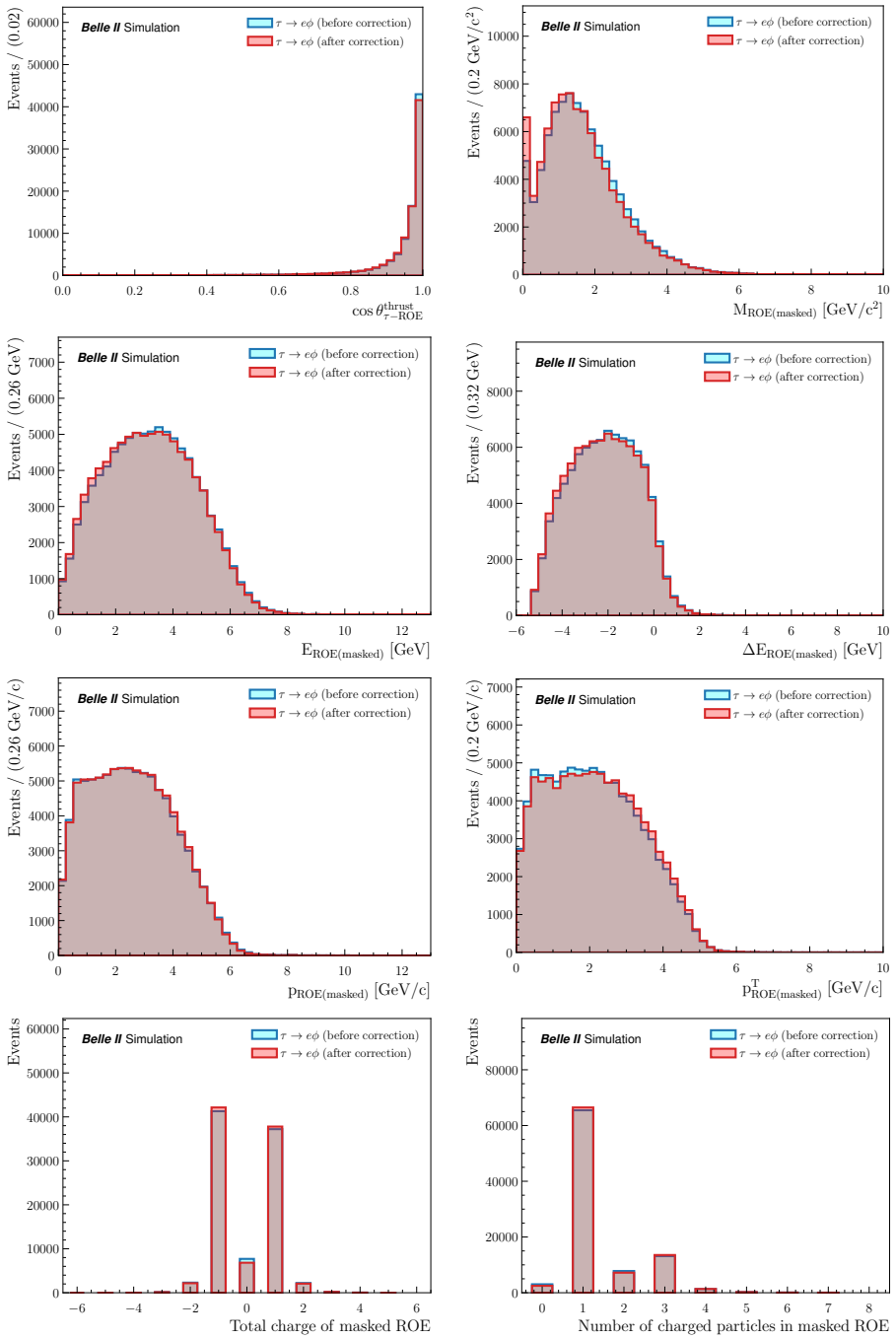


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

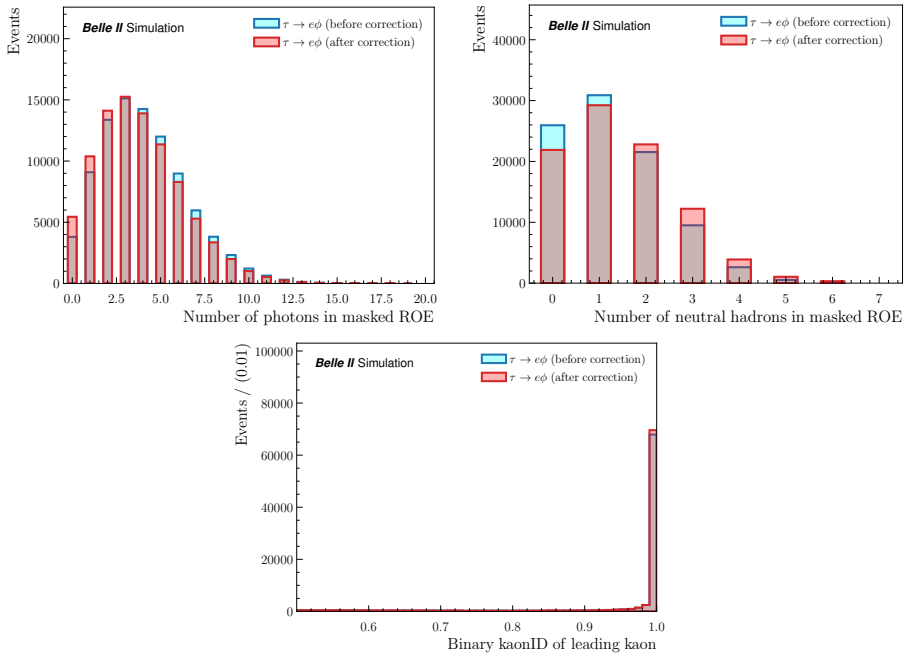


Figure C.3: Distributions of $\tau \to e\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

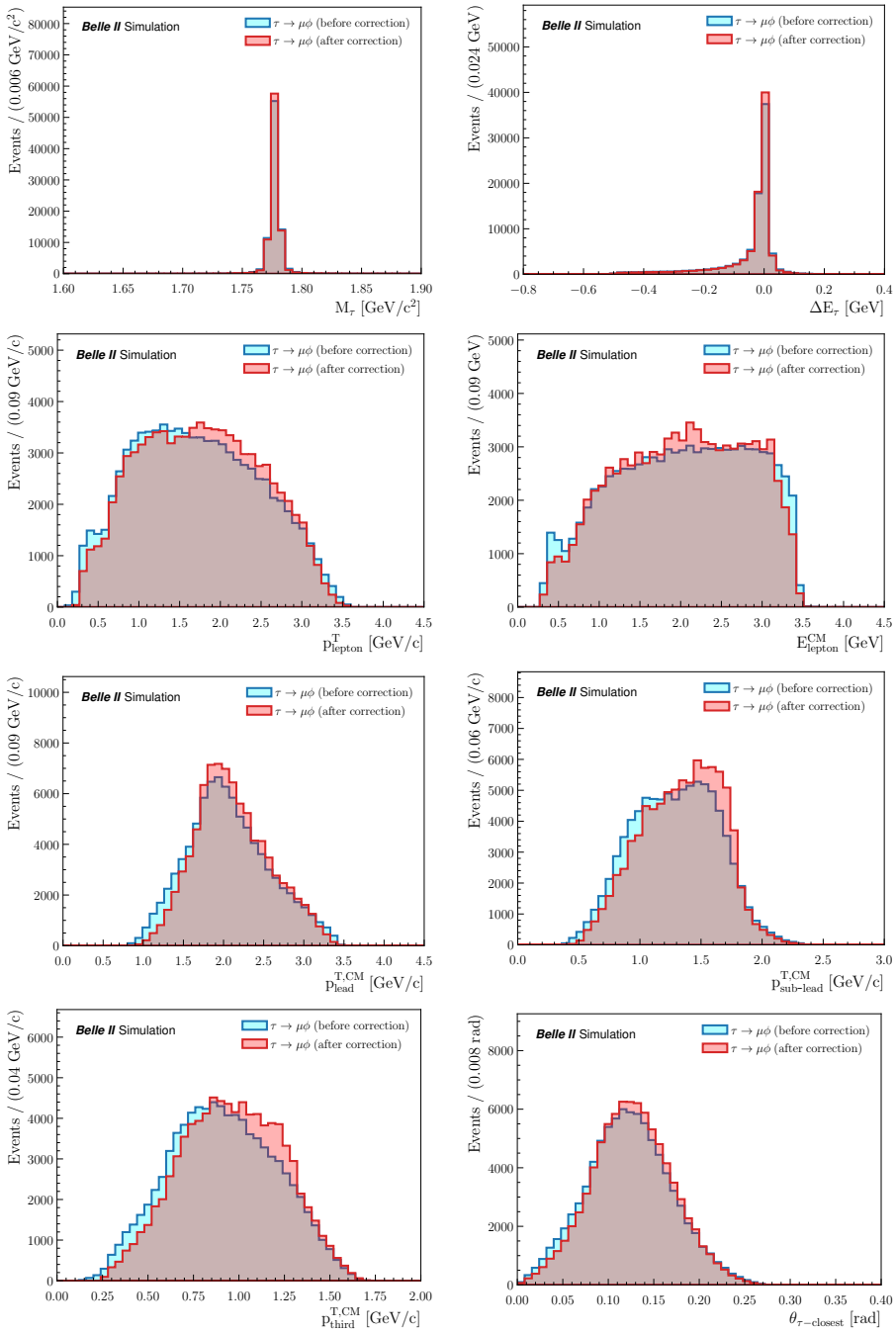


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

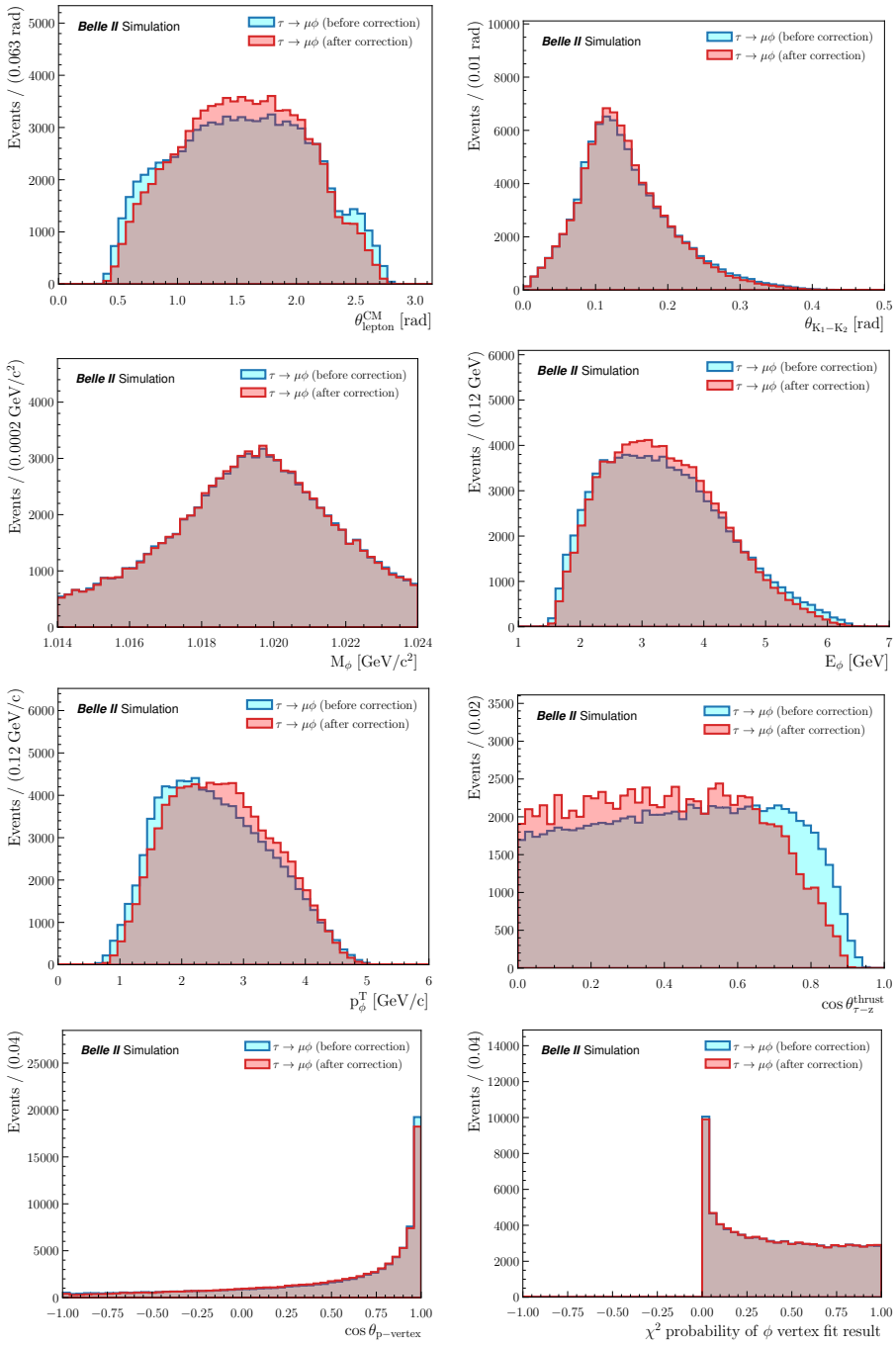


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

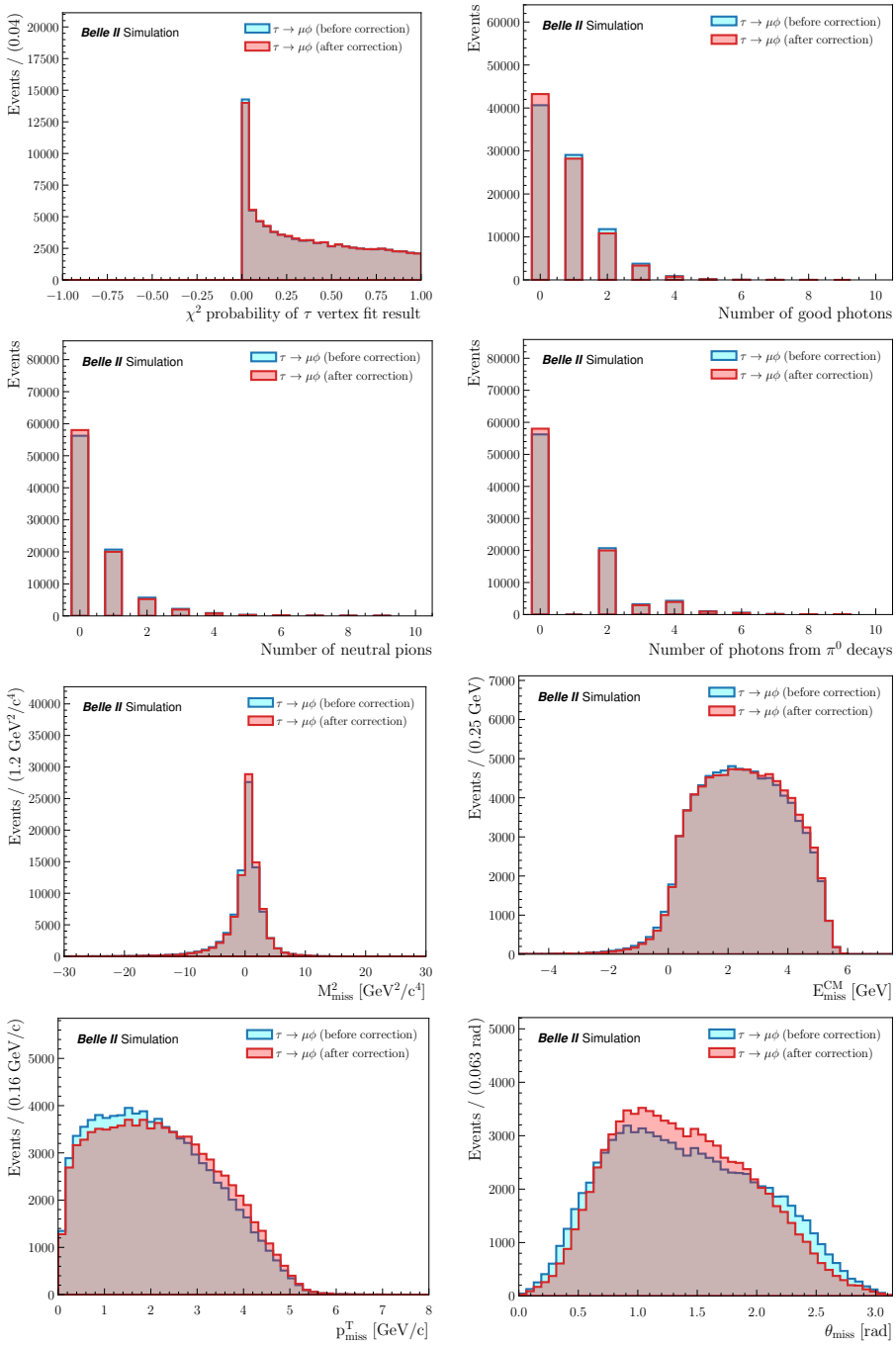


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

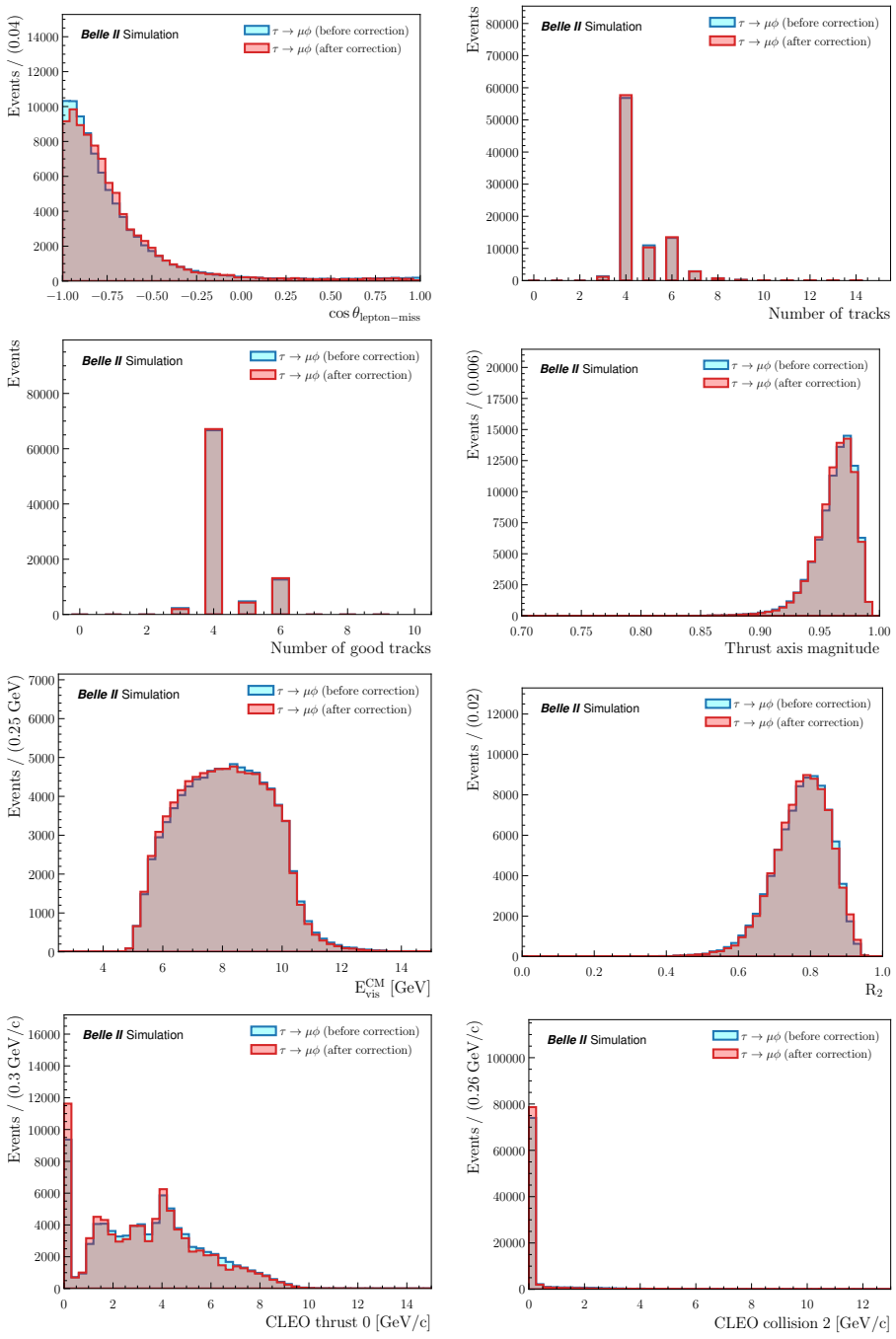


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

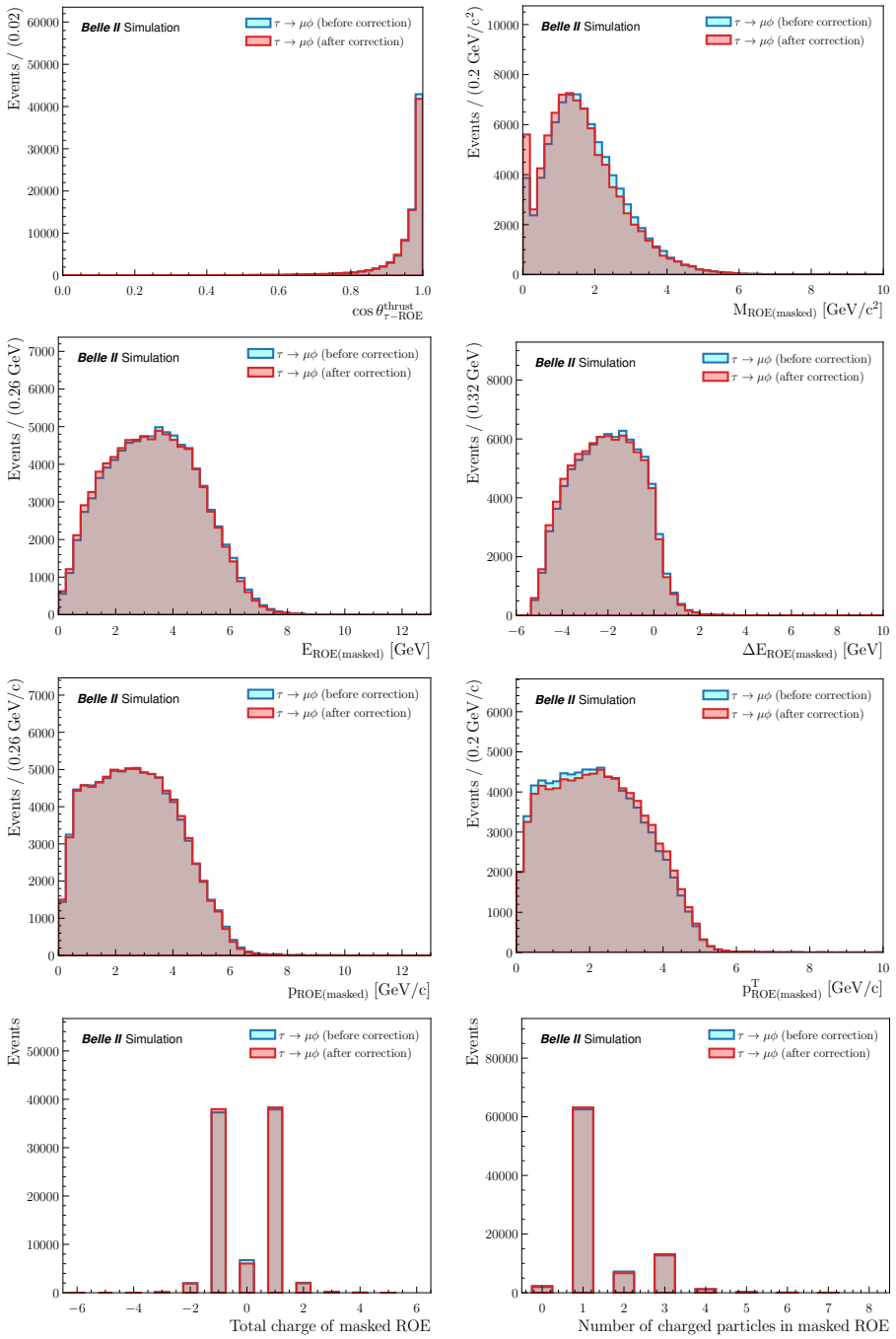


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

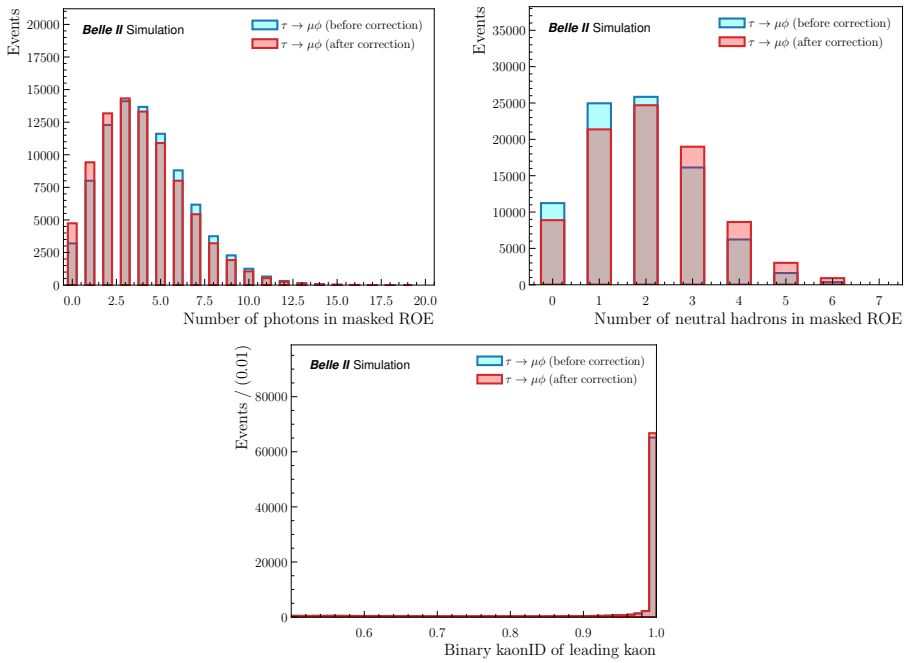


Figure C.4: Distributions of $\tau \to \mu\phi$ signal events for M_τ, ΔE_τ and variables of interest, before and after the corrections are applied.

D. Control samples study

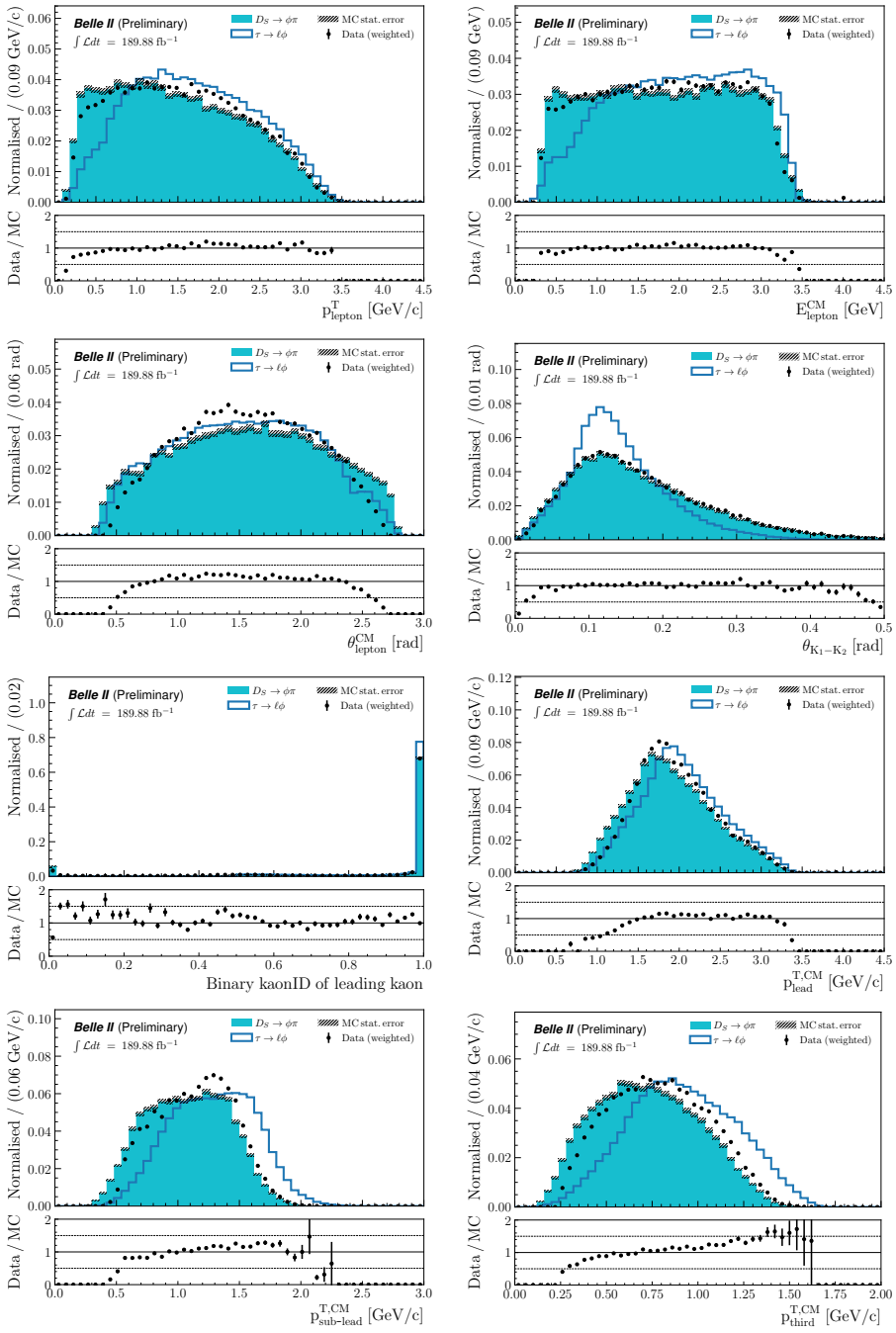


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

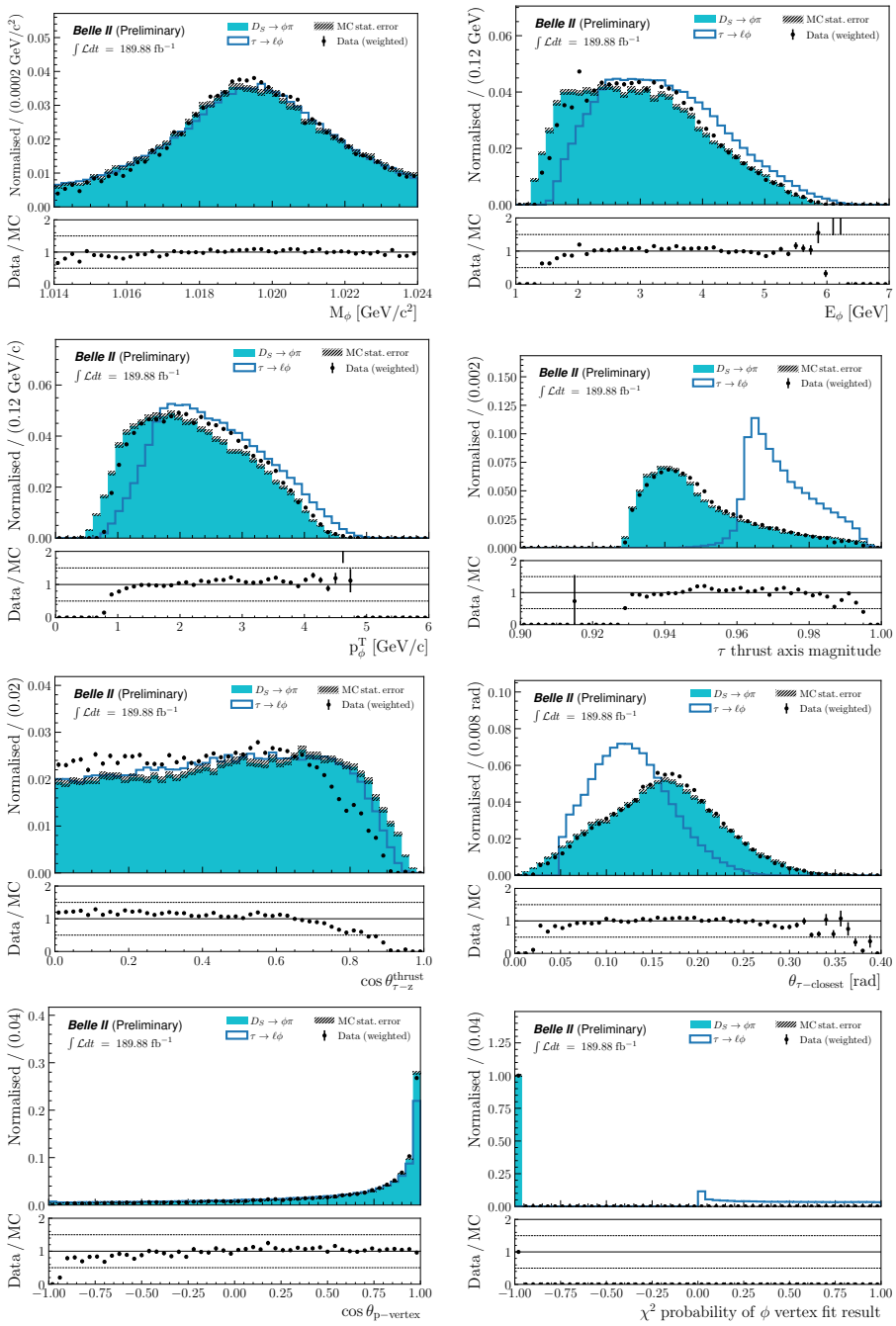


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

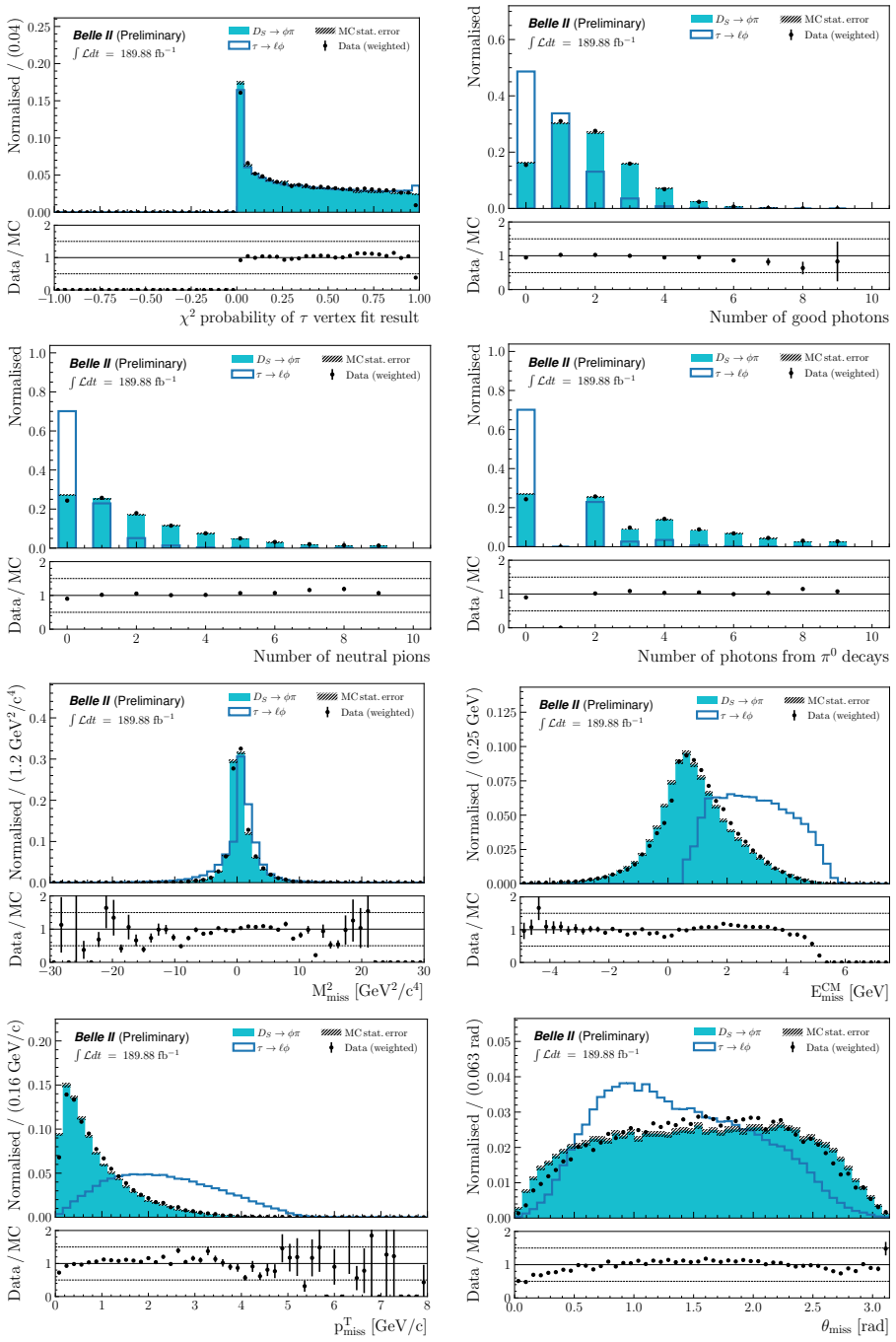


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

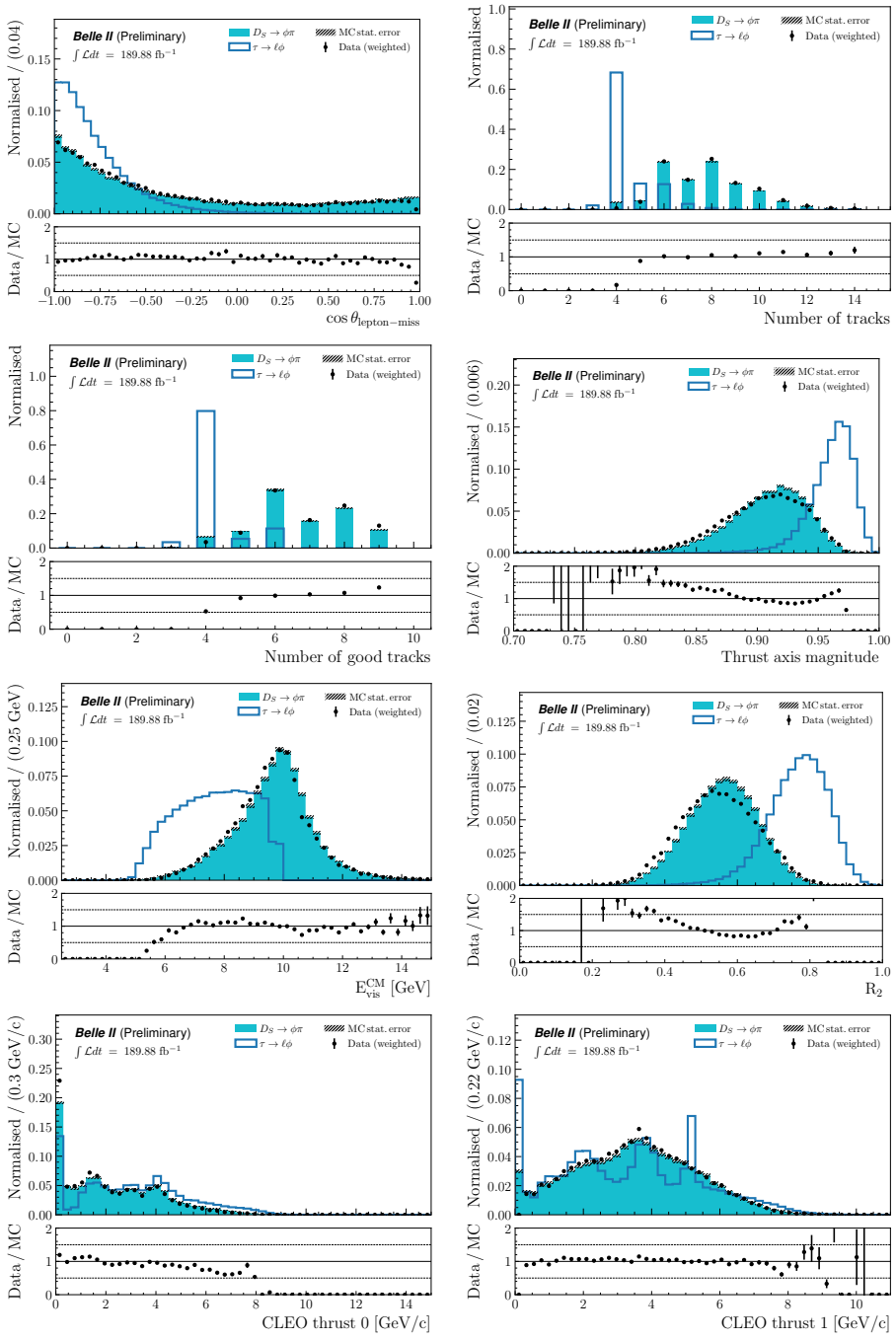


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

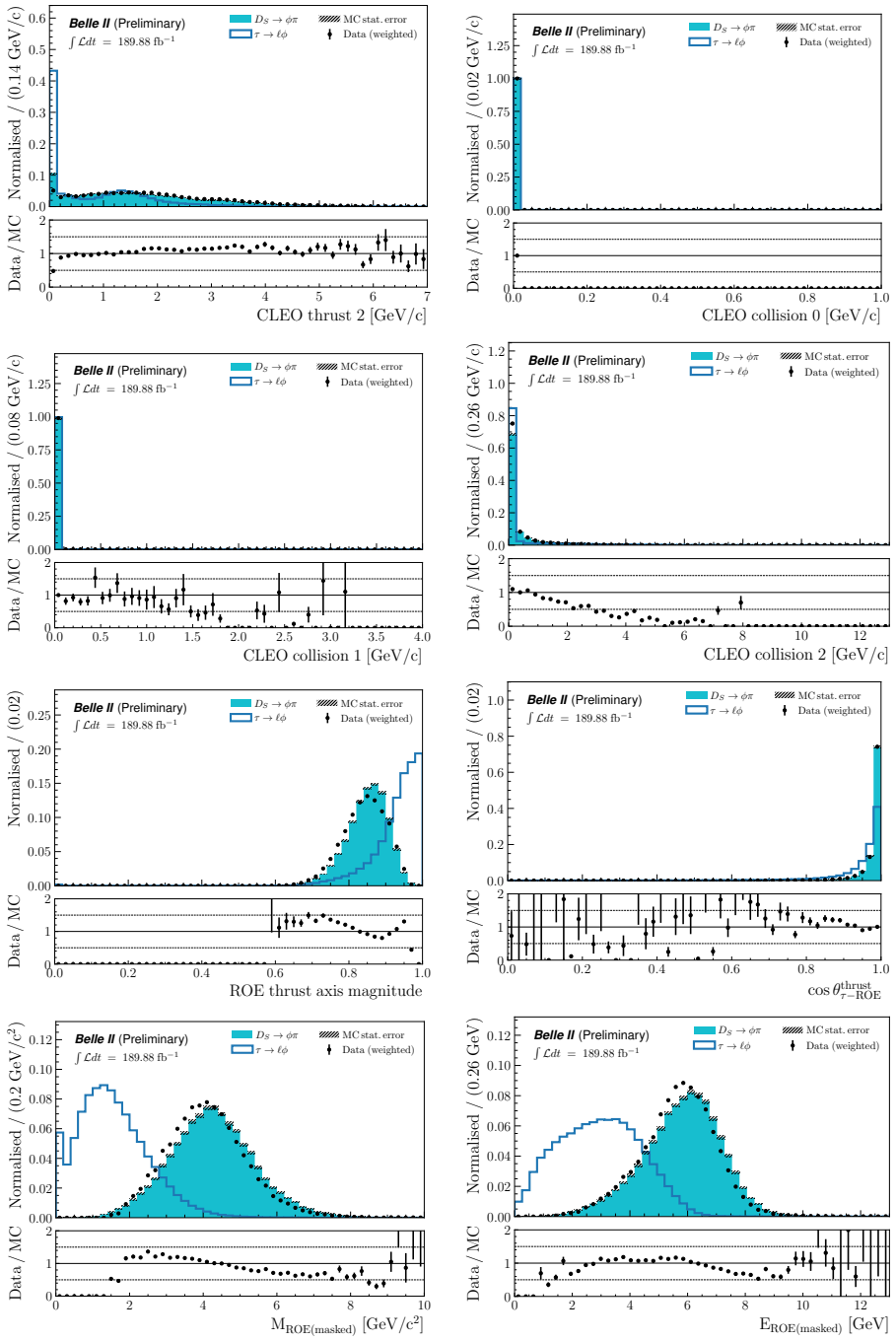


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

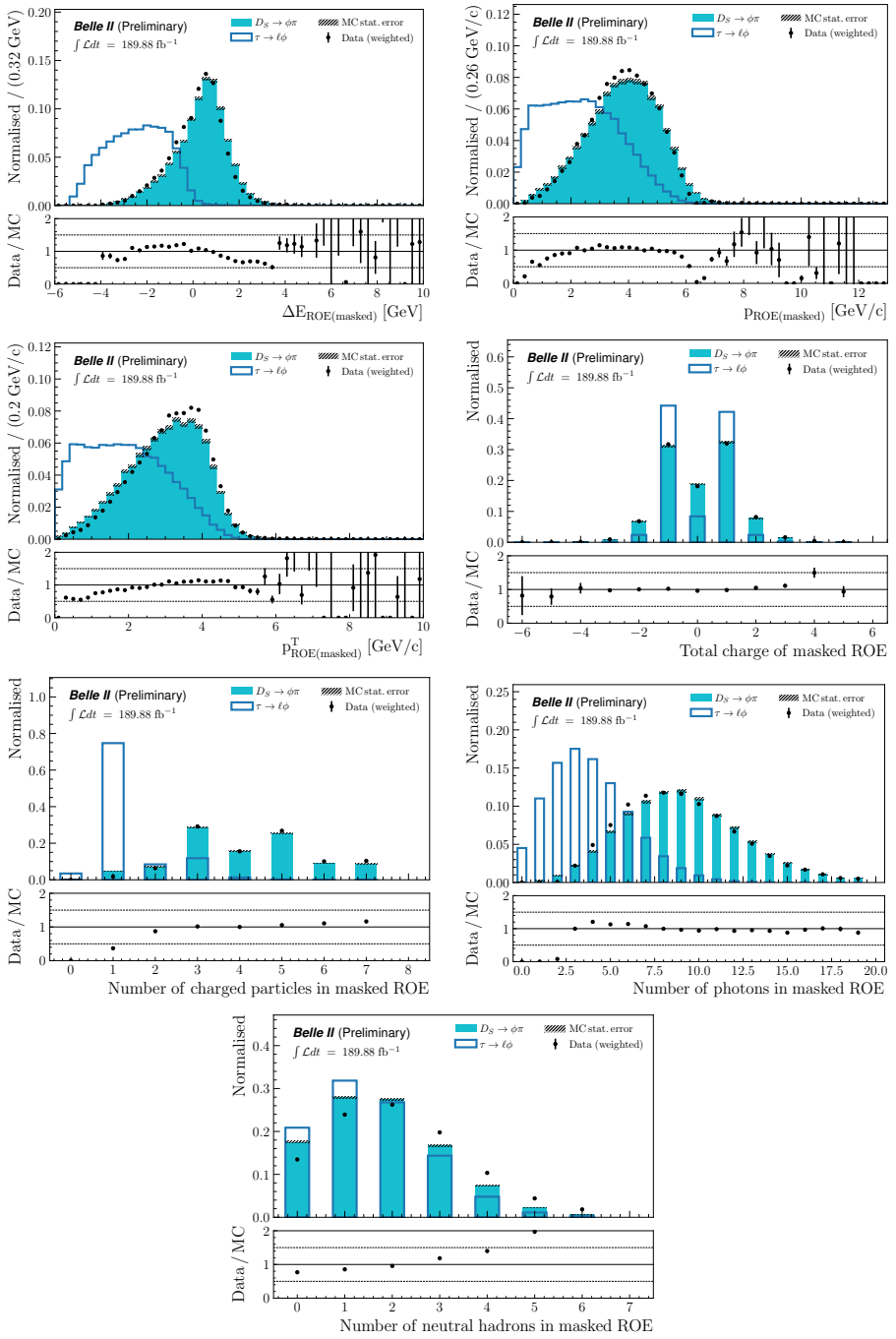


Figure D.1: Normalised $D_s \to \phi\pi$ sWeighted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest.

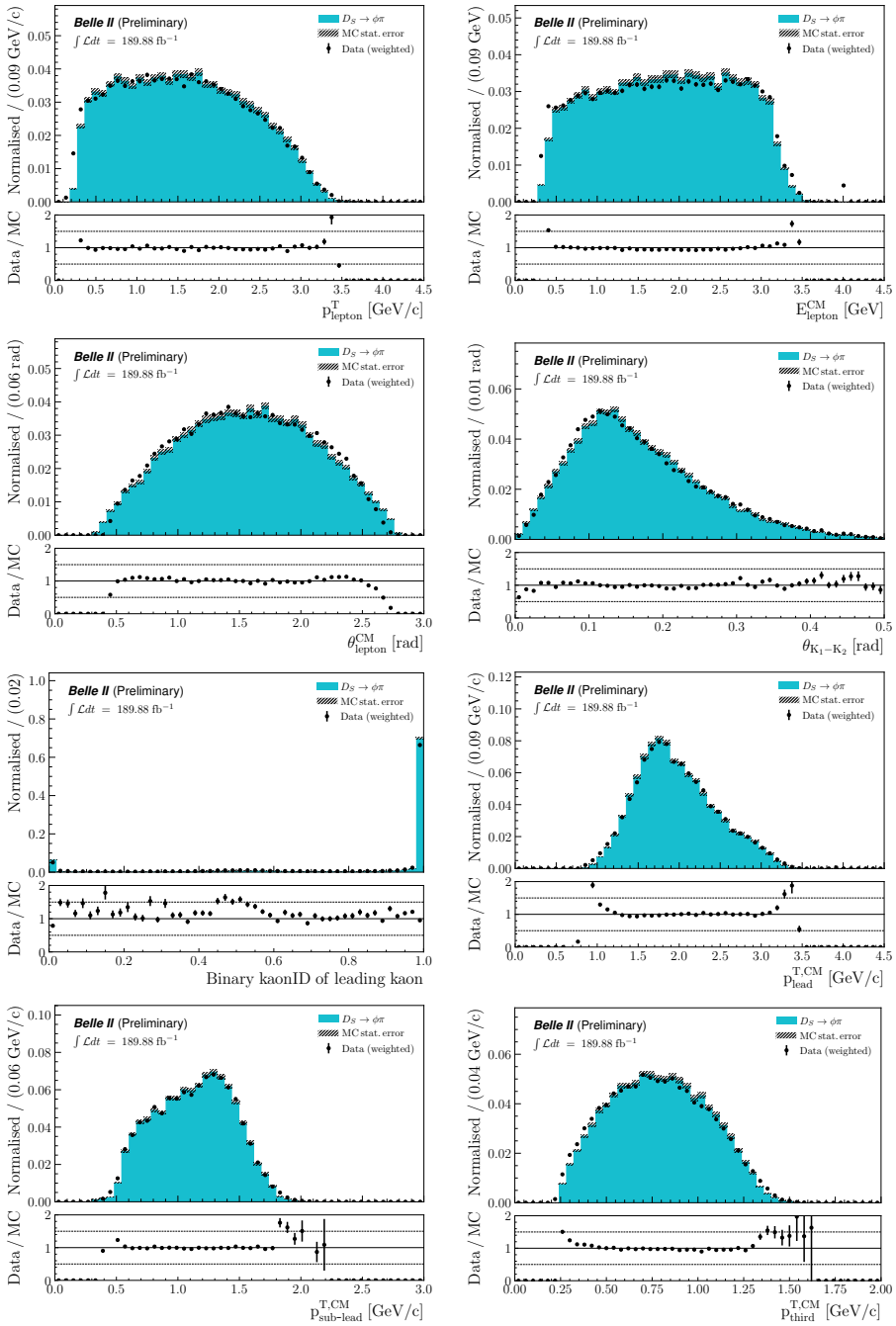


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

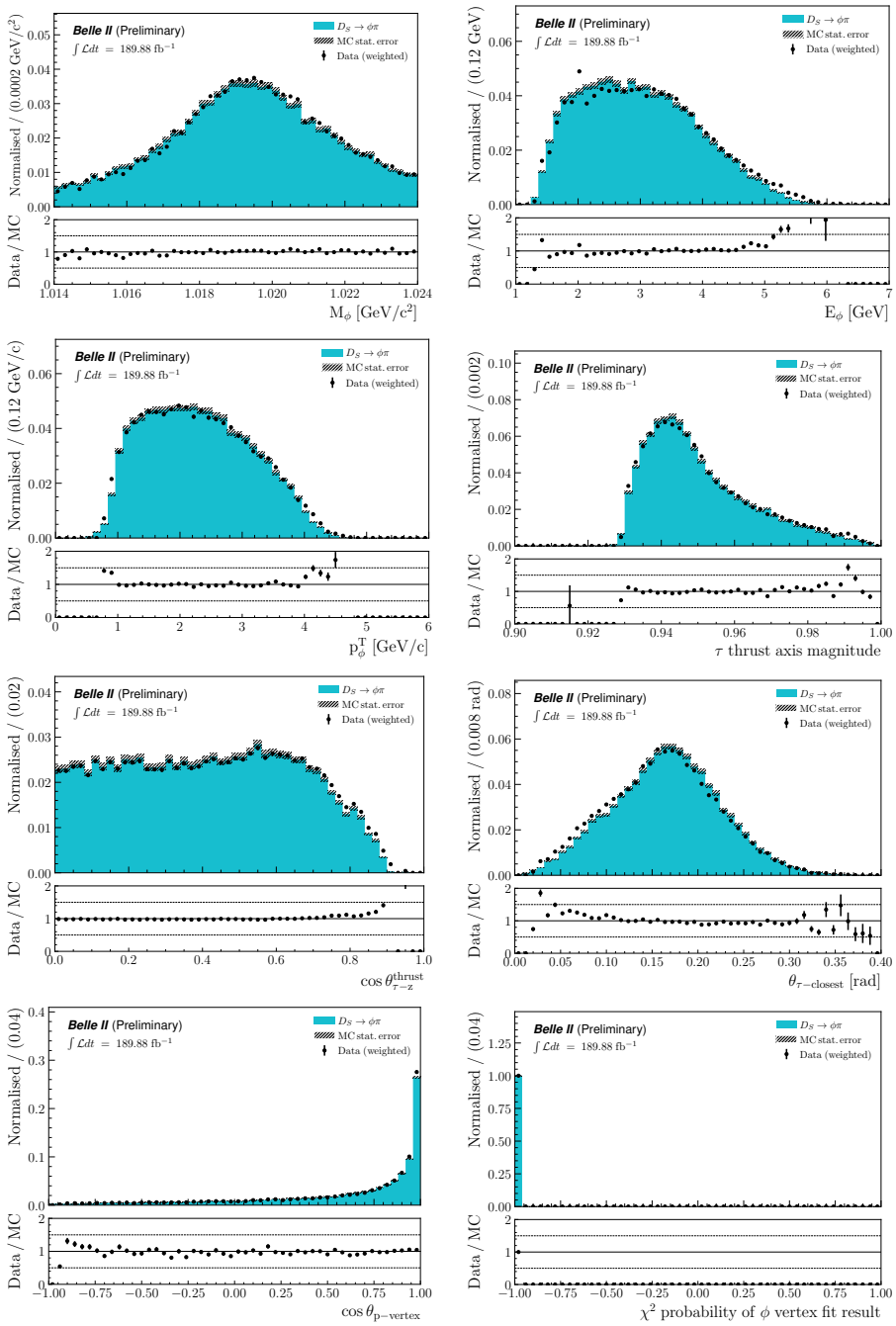


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

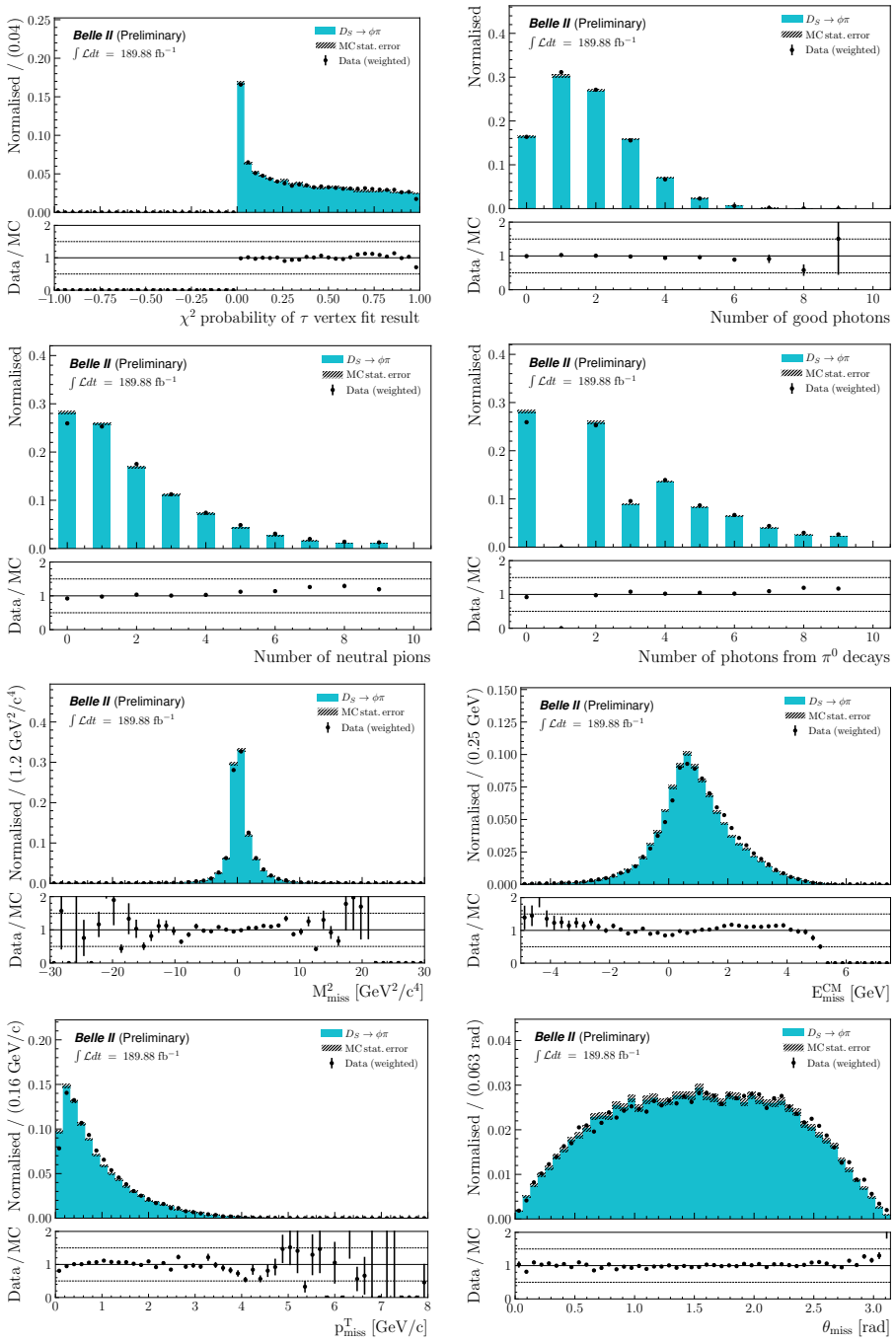


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

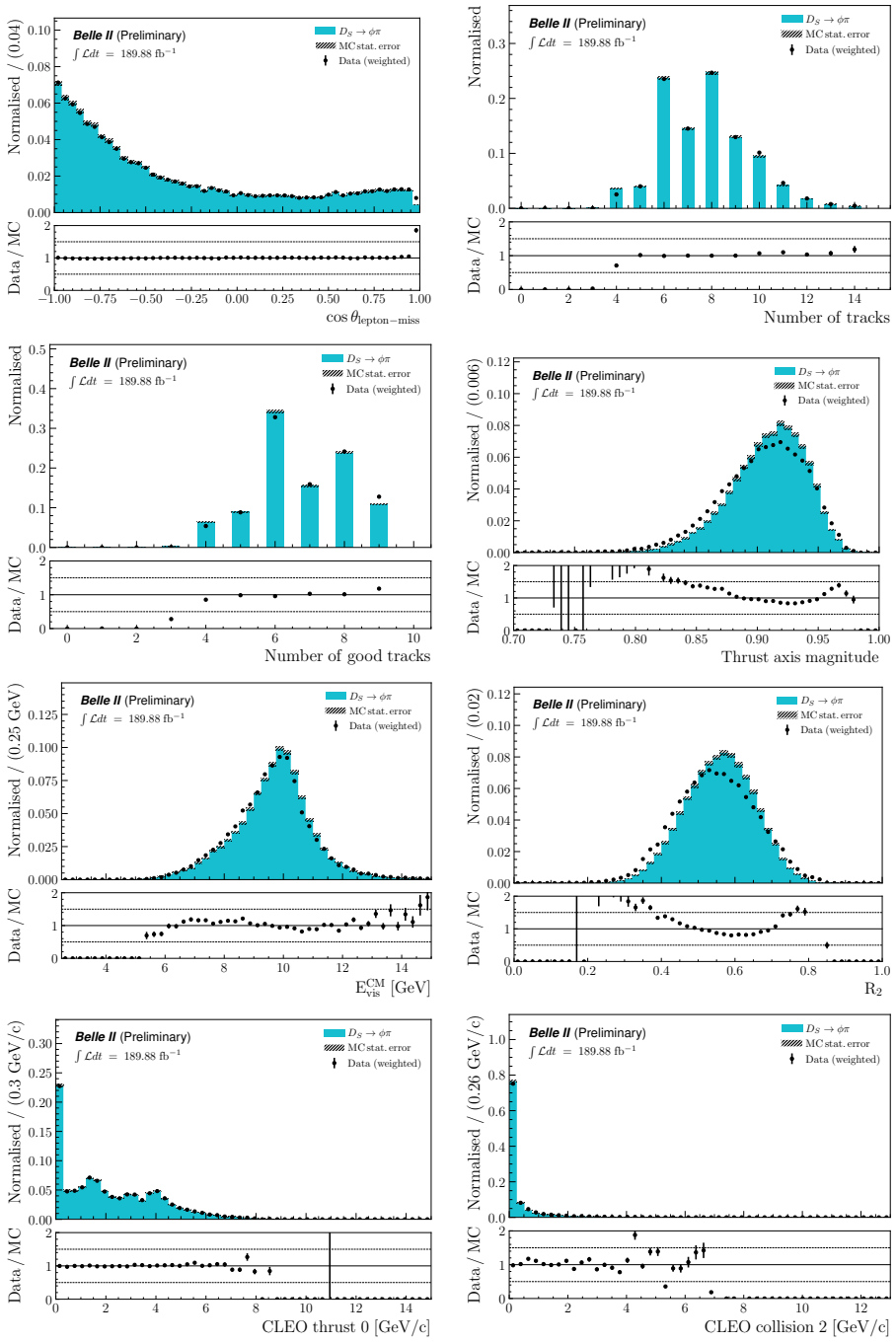


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

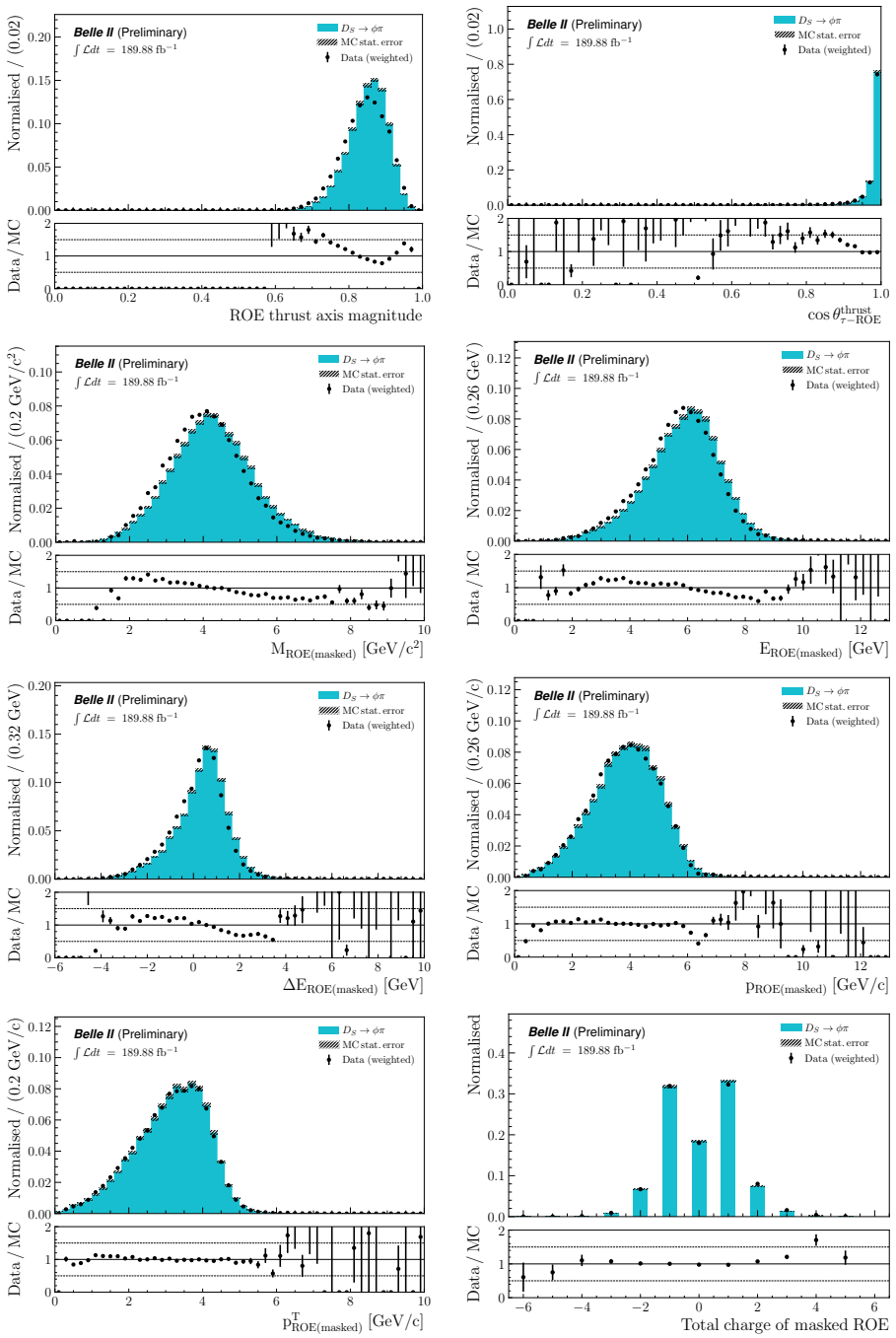


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

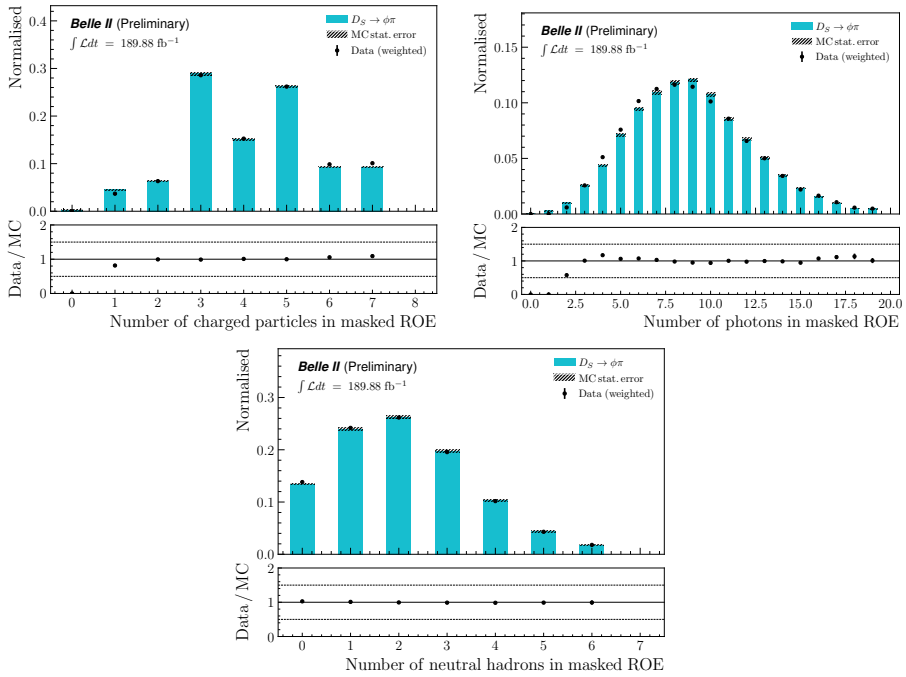


Figure D.2: Normalised $D_s \to \phi\pi$ sWeighted data and corrected signal MC distributions of variables of interest.

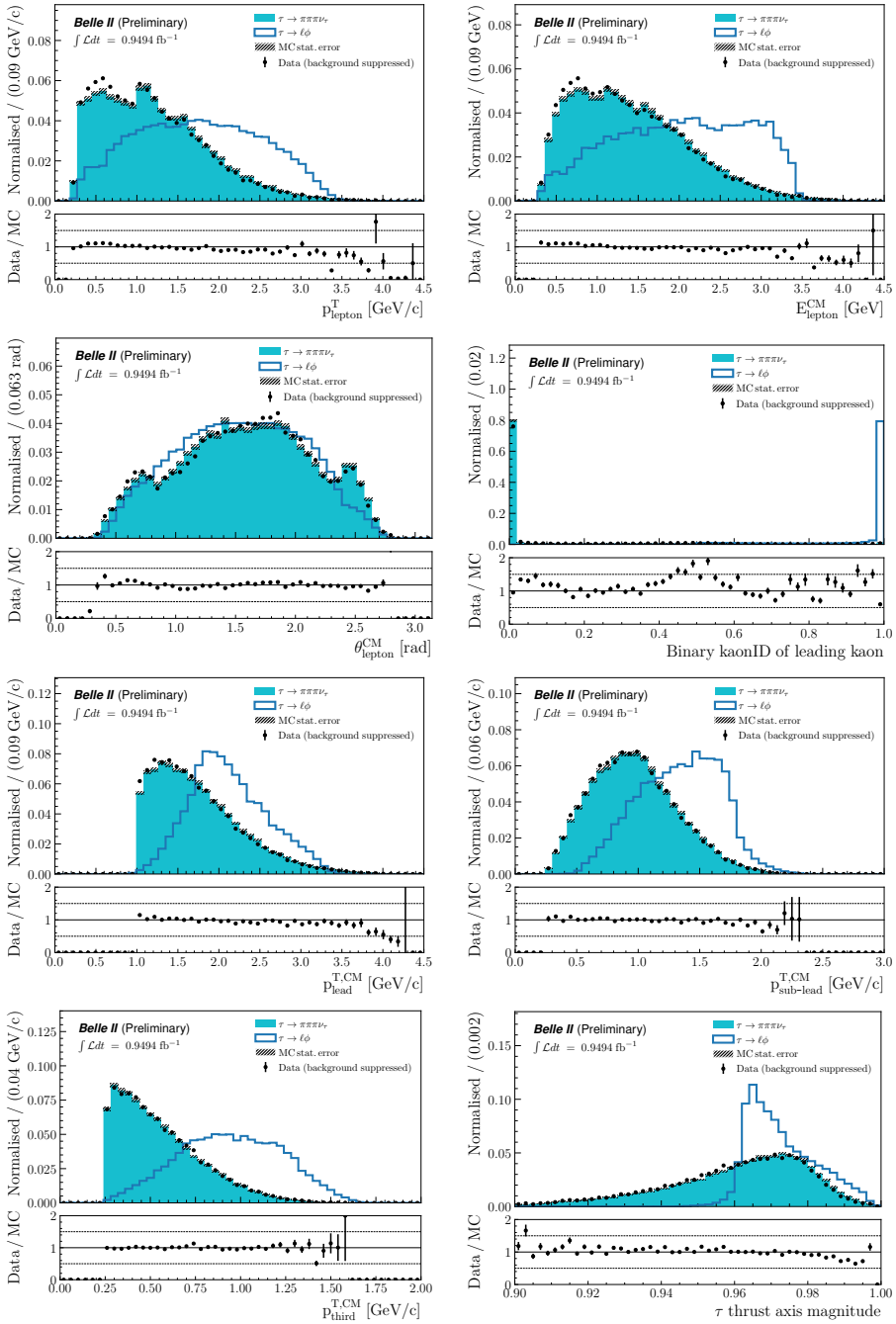


Figure D.3: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

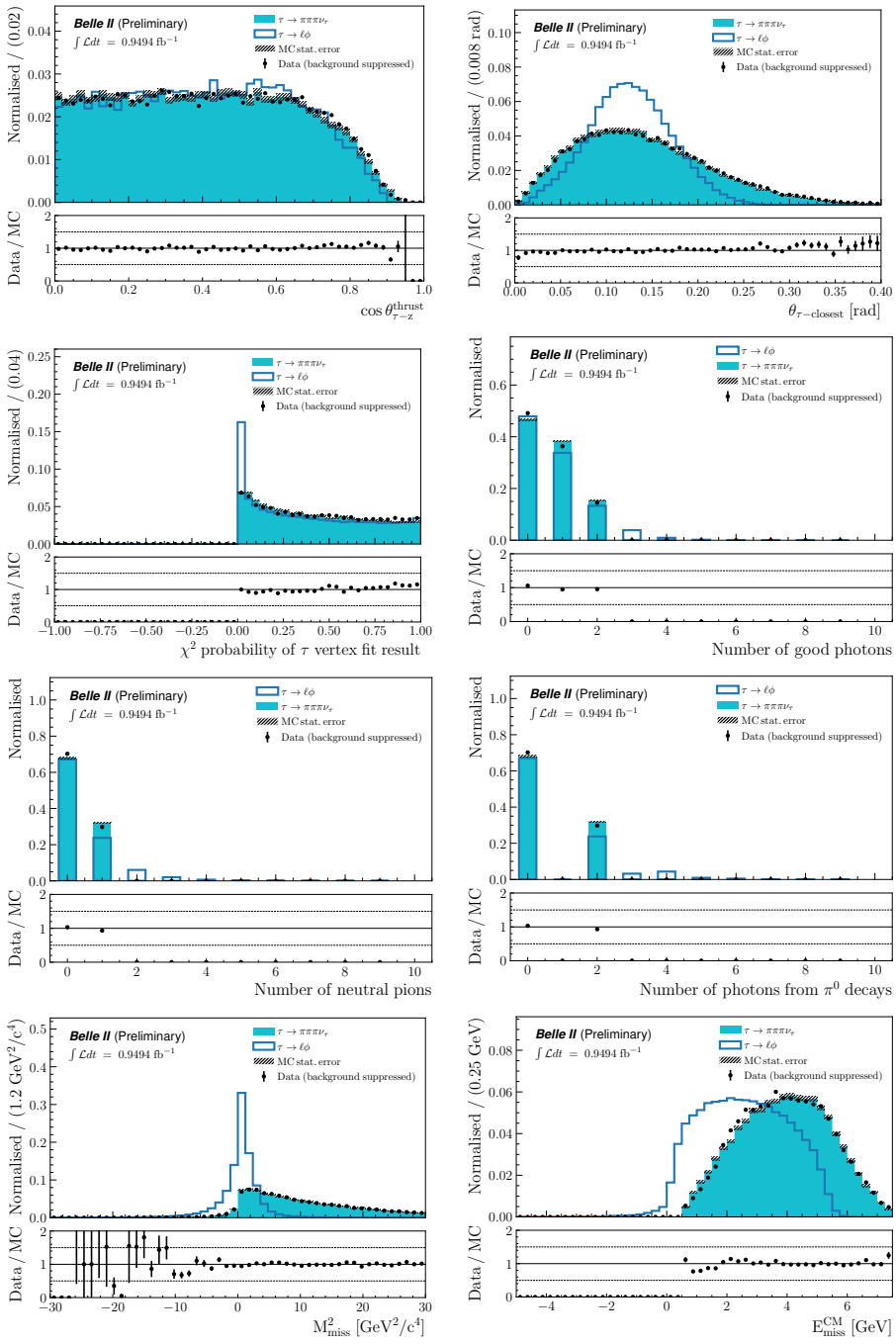


Figure D.3: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

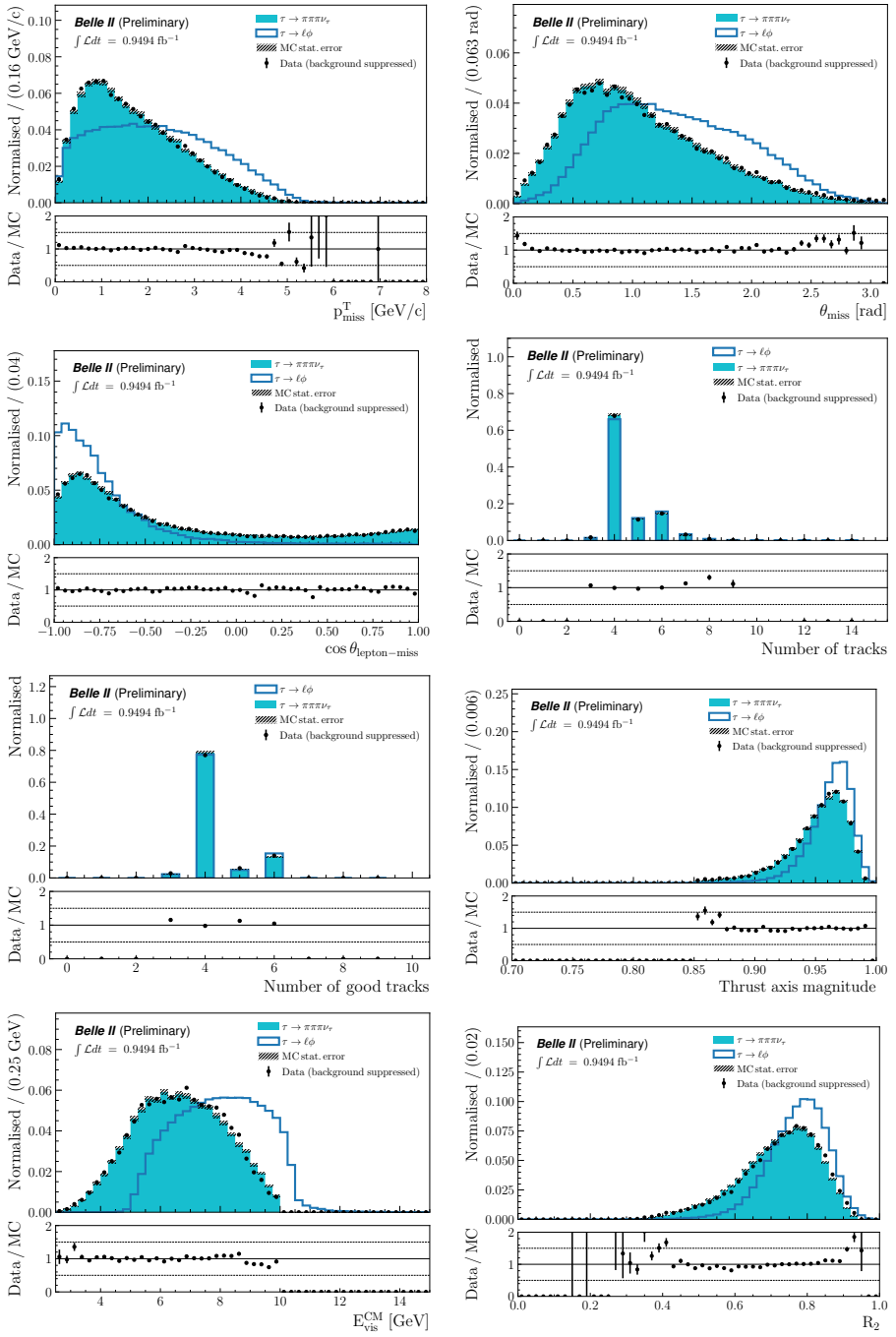


Figure D.3: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

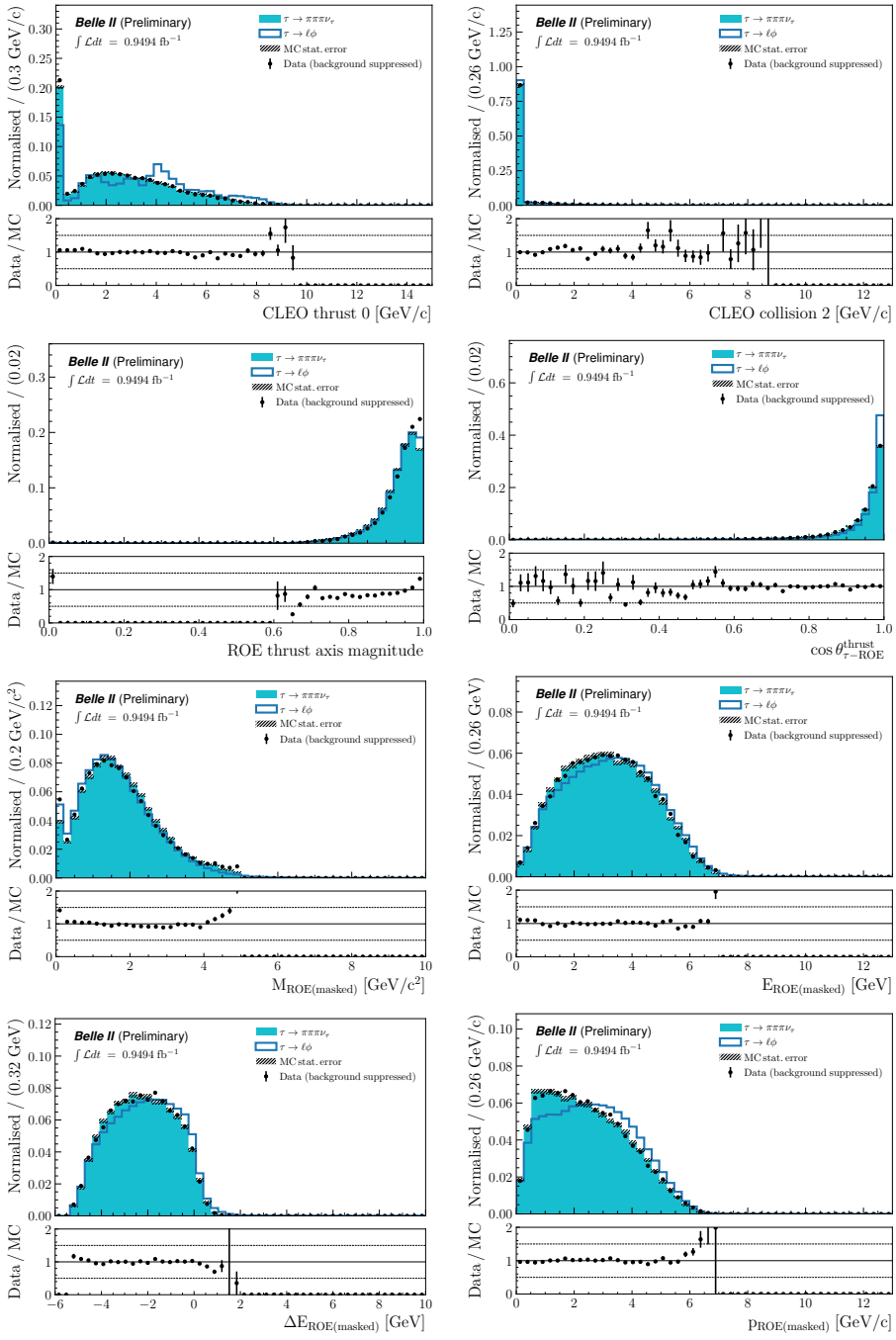


Figure D.3: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

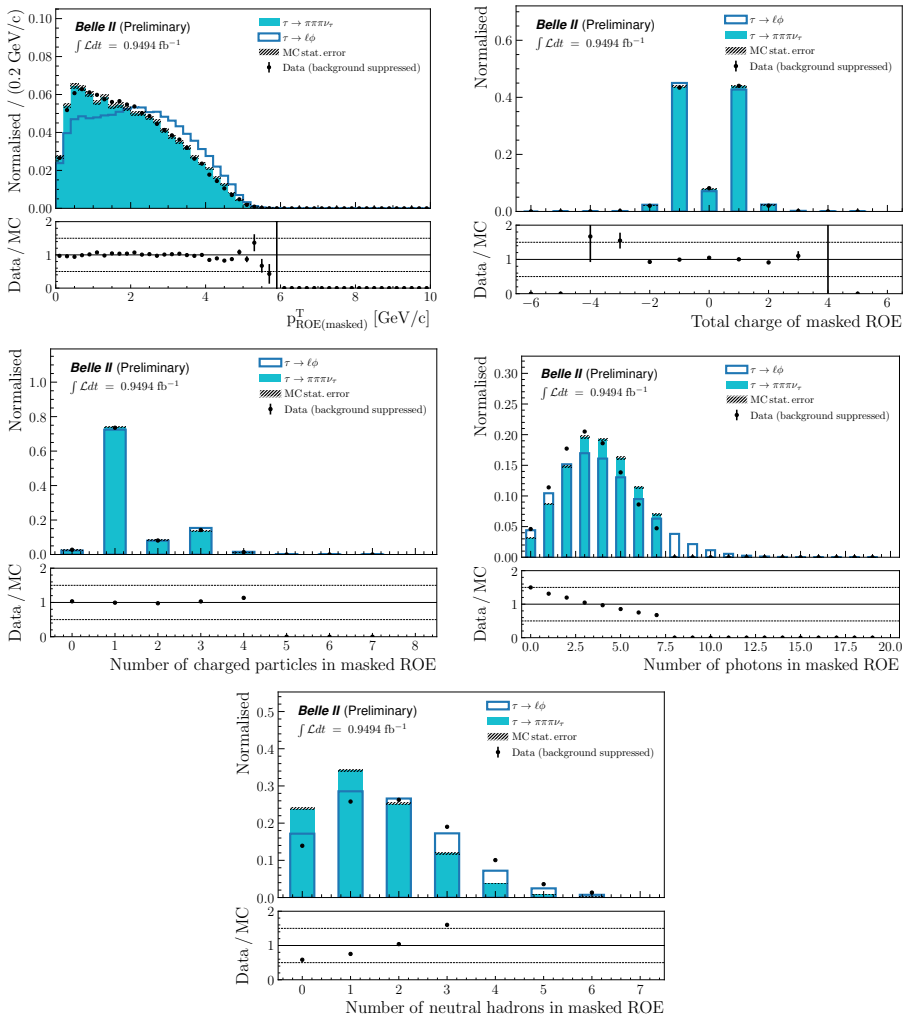


Figure D.3: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and signal MC distributions, along with $\tau \to \ell\phi$ distributions of variables of interest. $\tau \to \ell\phi$ signal MC distributions are weighted according to the $D_s \to \phi\pi$ corrections.

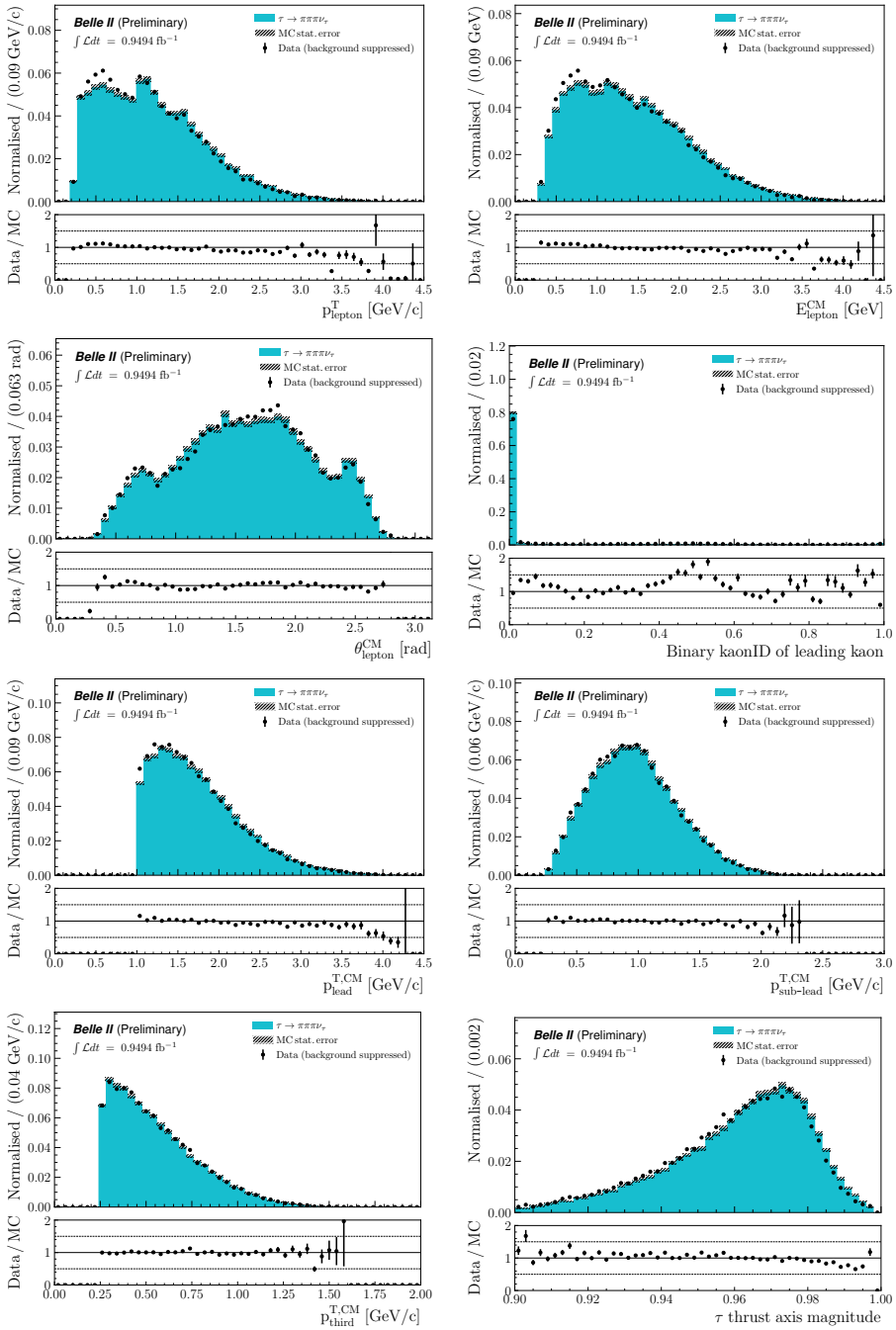


Figure D.4: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables of interest.

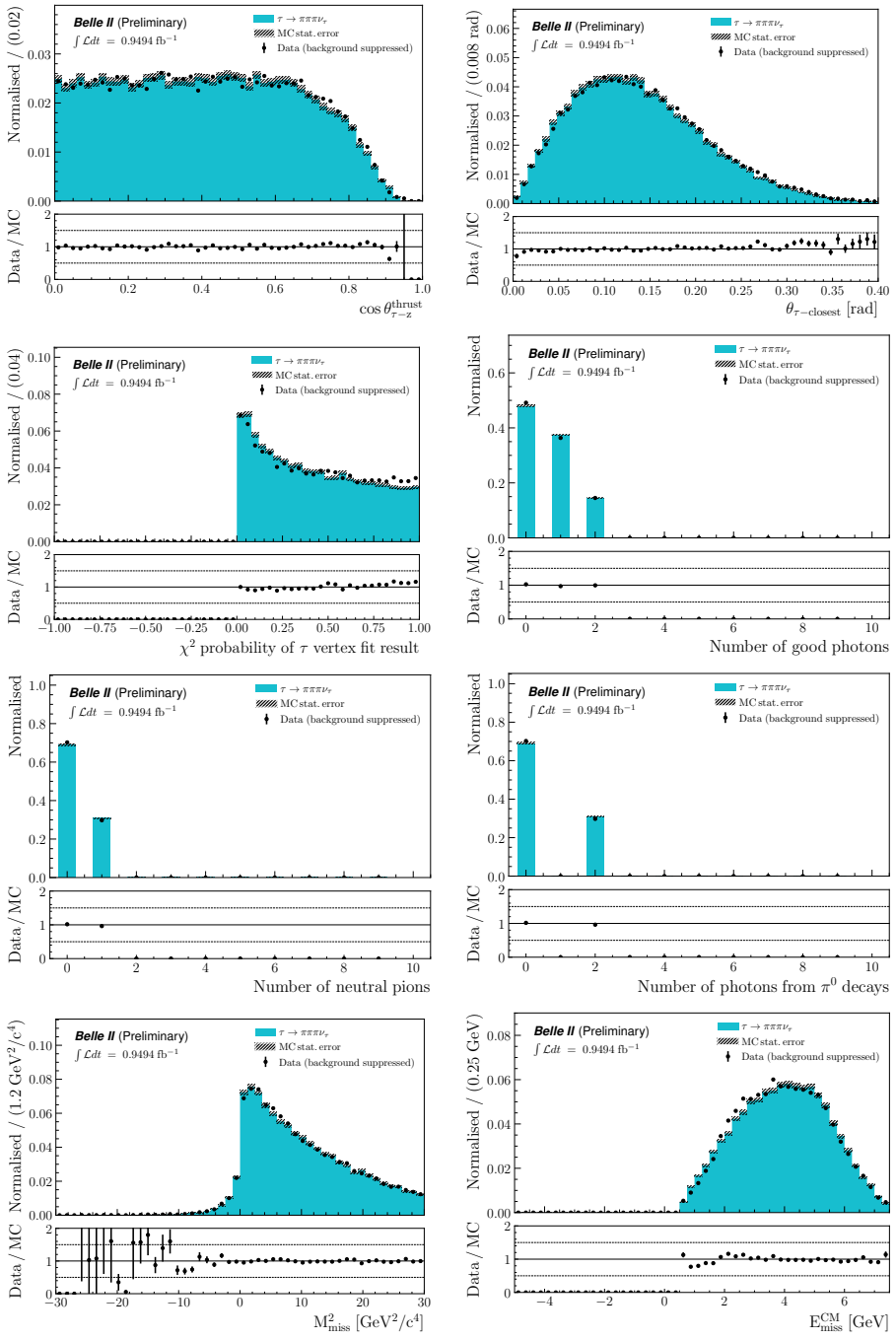


Figure D.4: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables of interest.

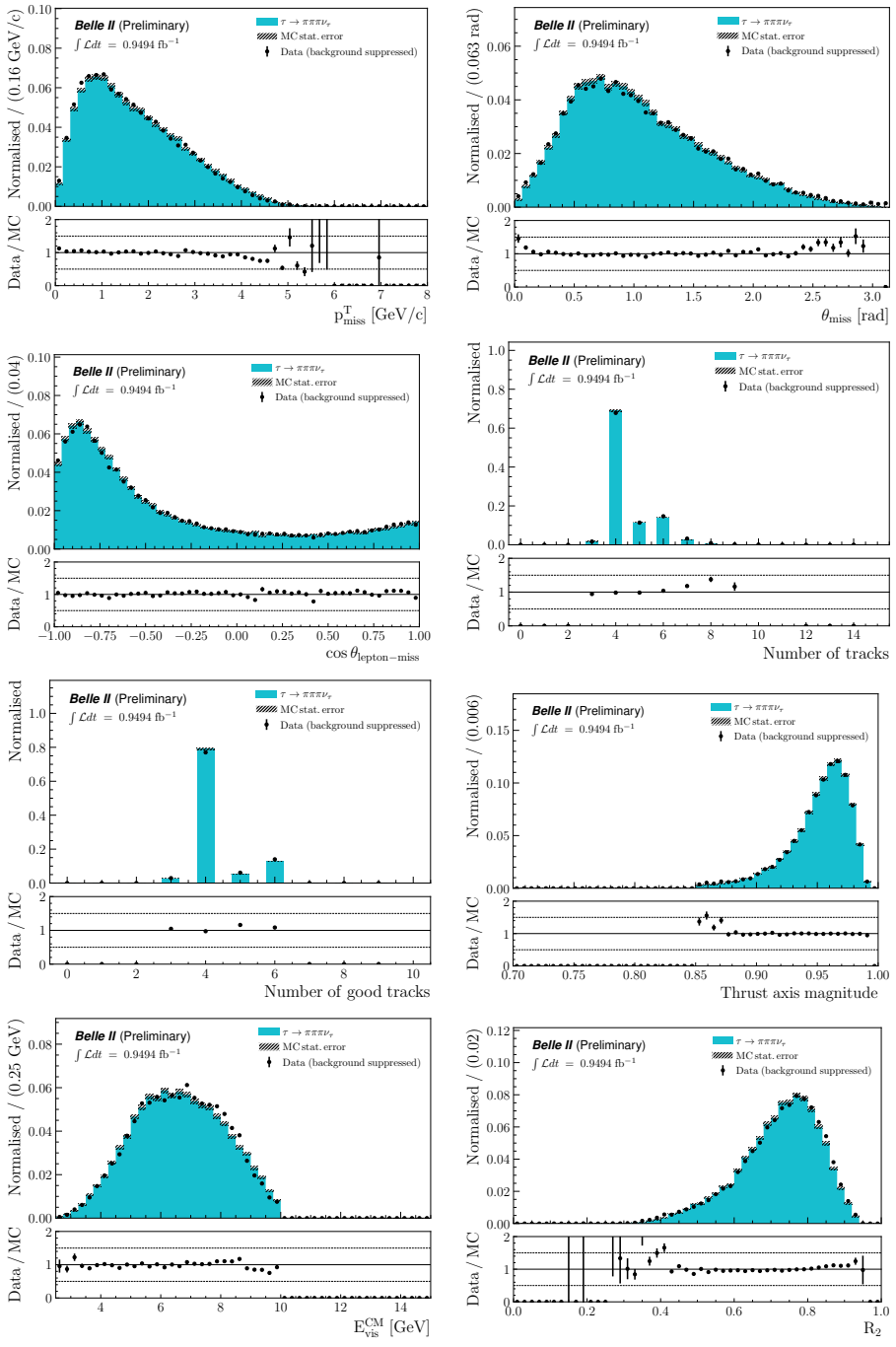


Figure D.4: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables of interest.

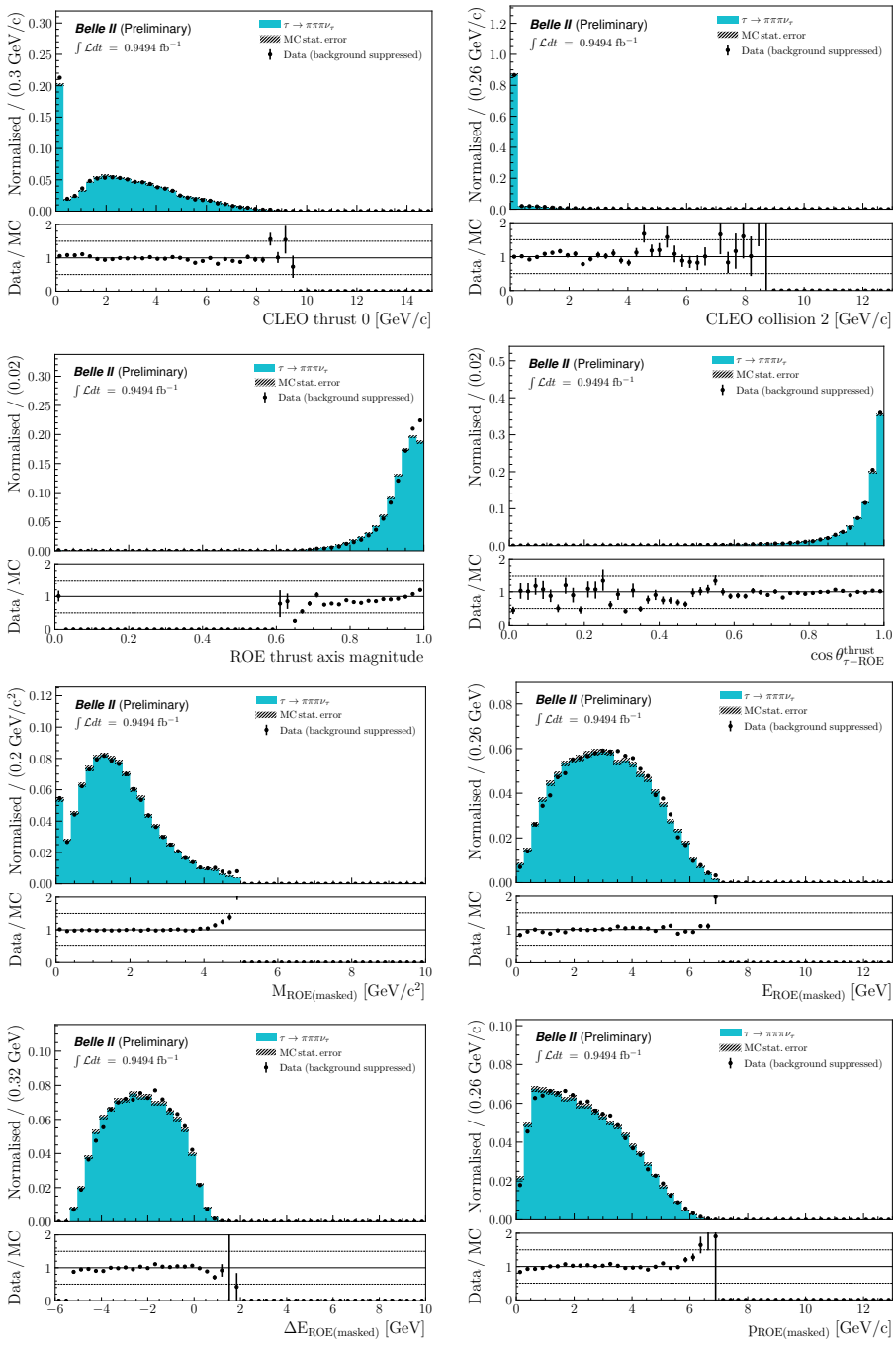


Figure D.4: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables of interest.

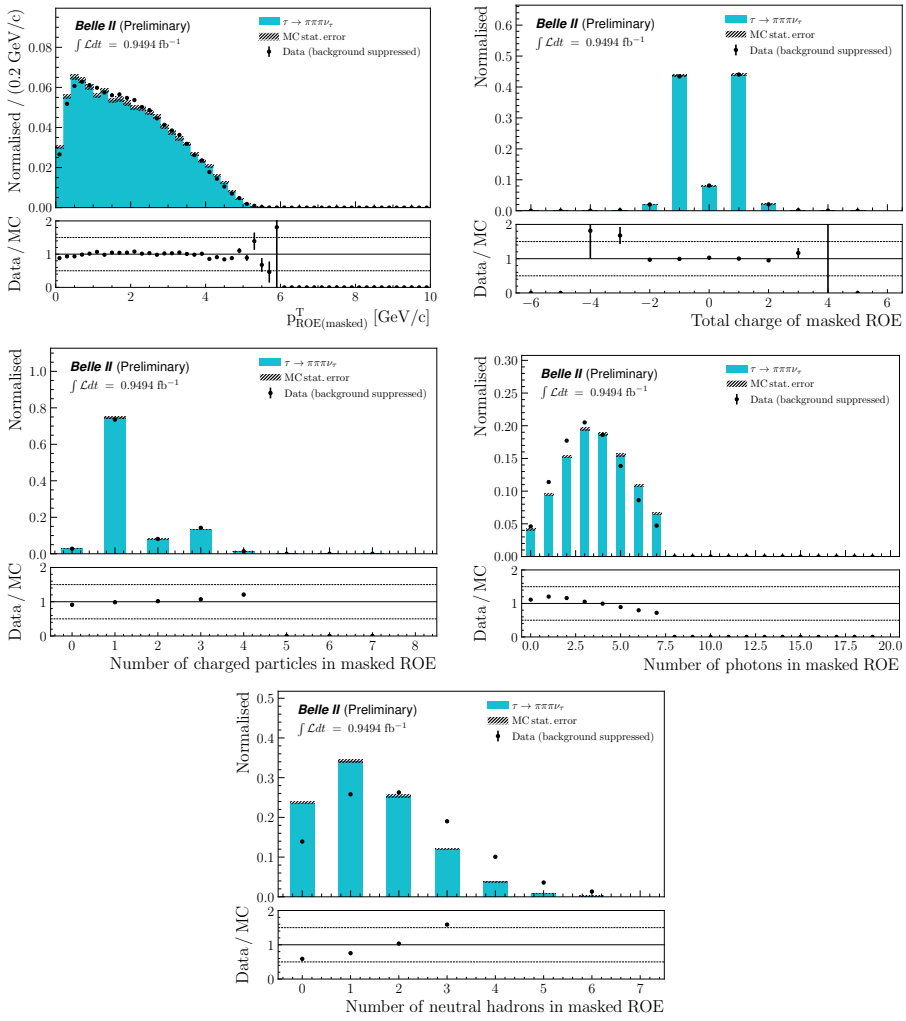


Figure D.4: Normalised $\tau \to 3\pi\nu_\tau$ background-subtracted data and corrected signal MC distributions of variables of interest.

E. Data-MC comparison

Before the preselection

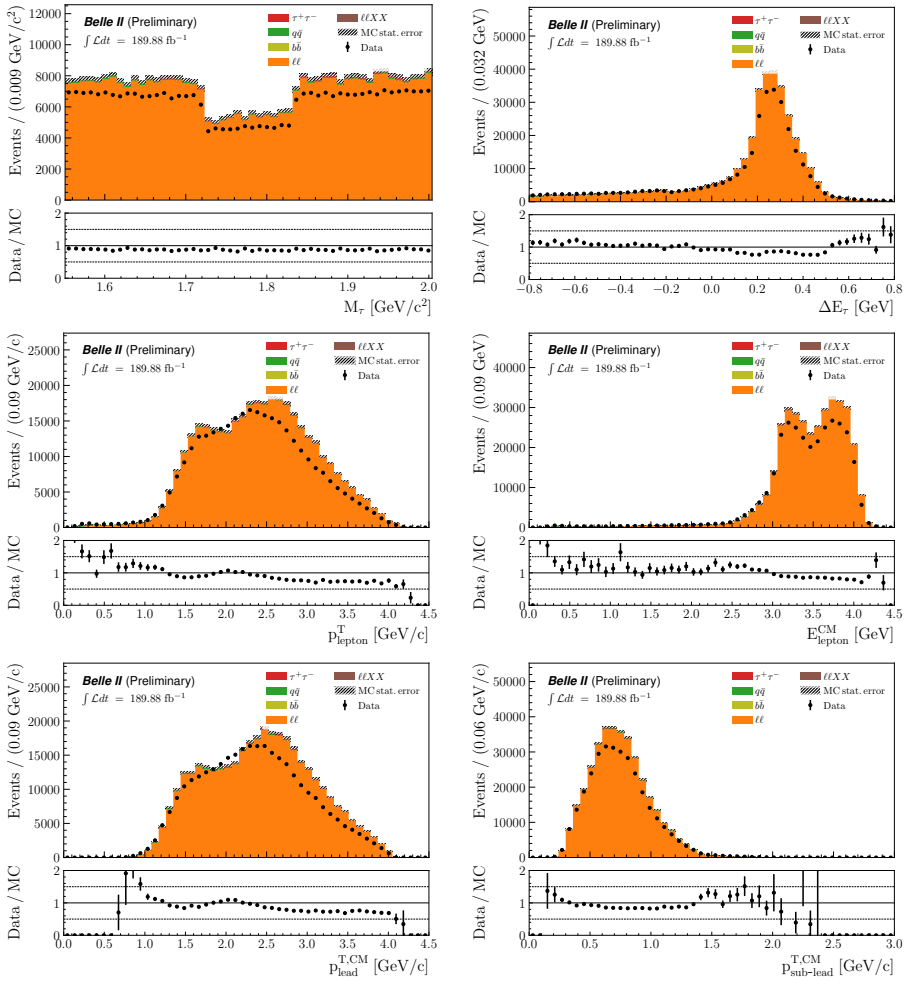


Figure E.1: Data-MC comparison in the $5 - 20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

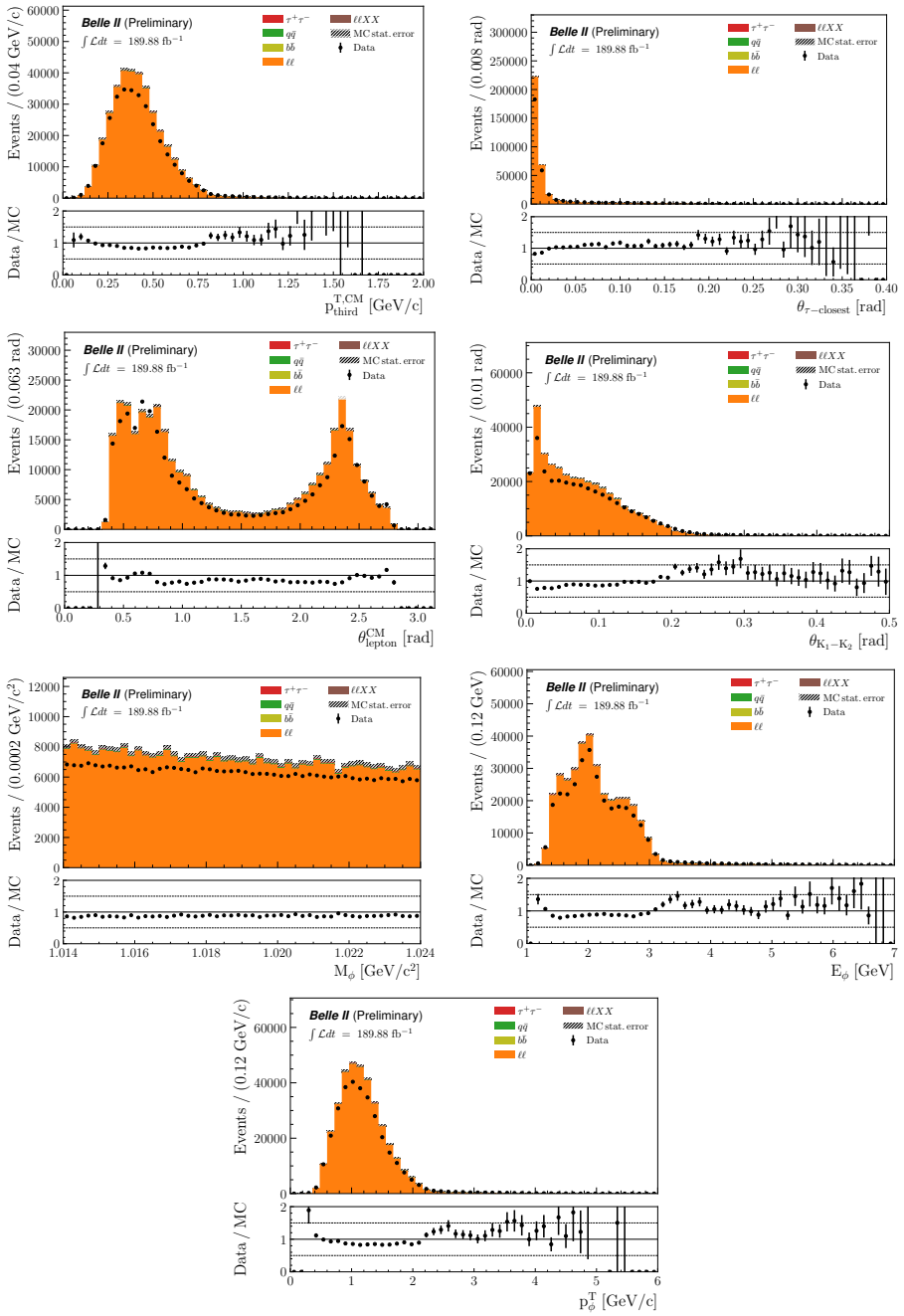


Figure E.1: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

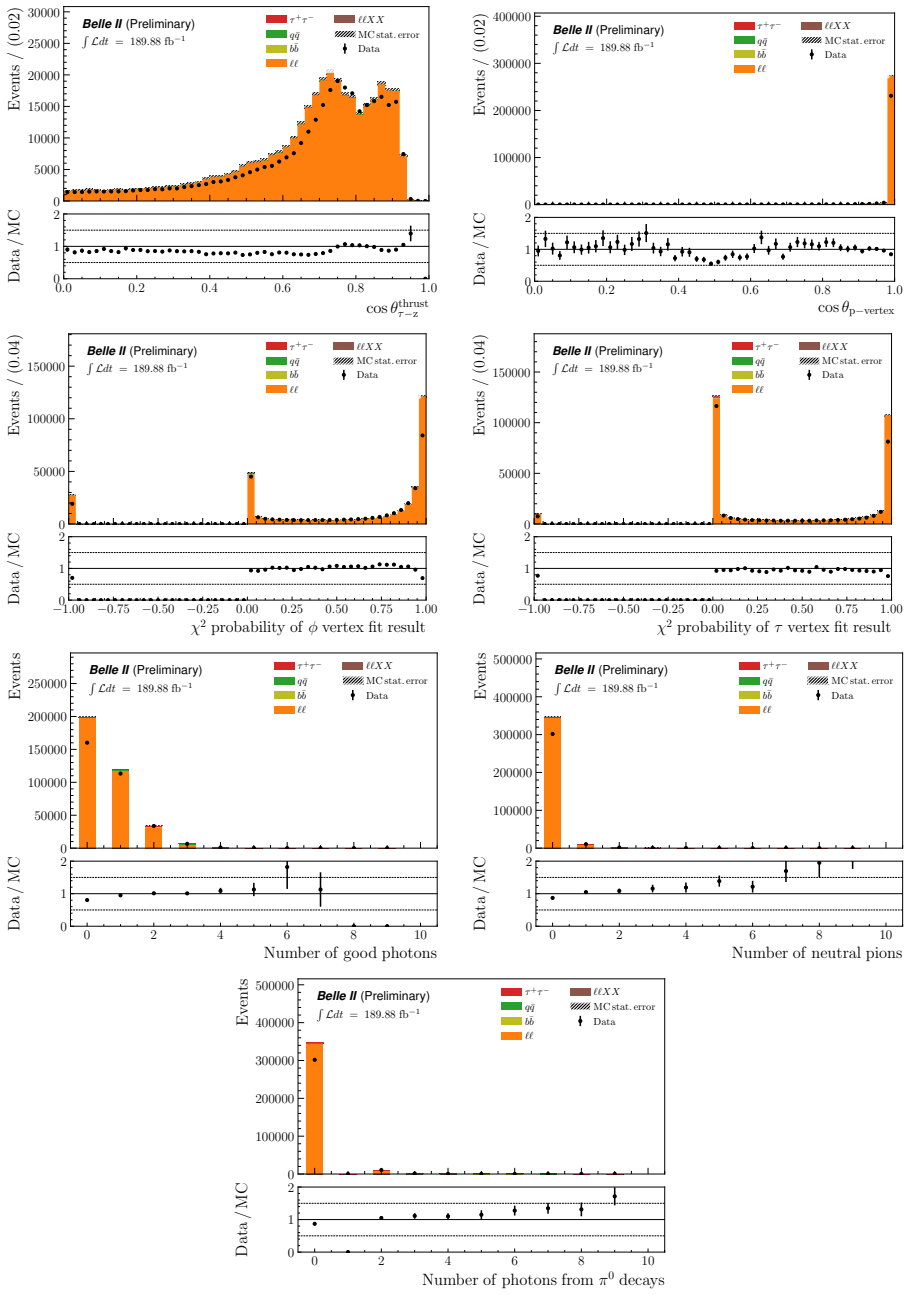


Figure E.1: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

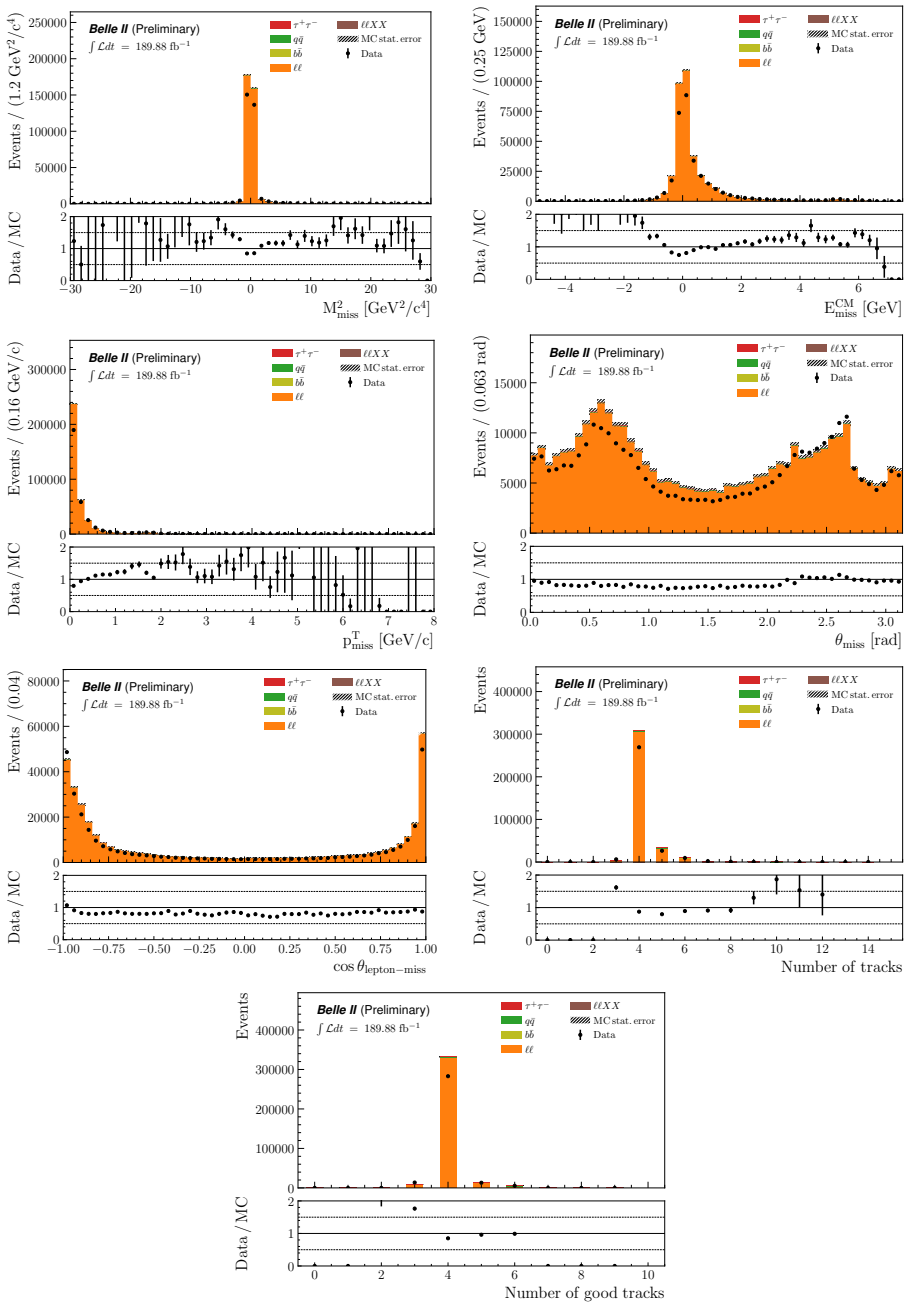


Figure E.1: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

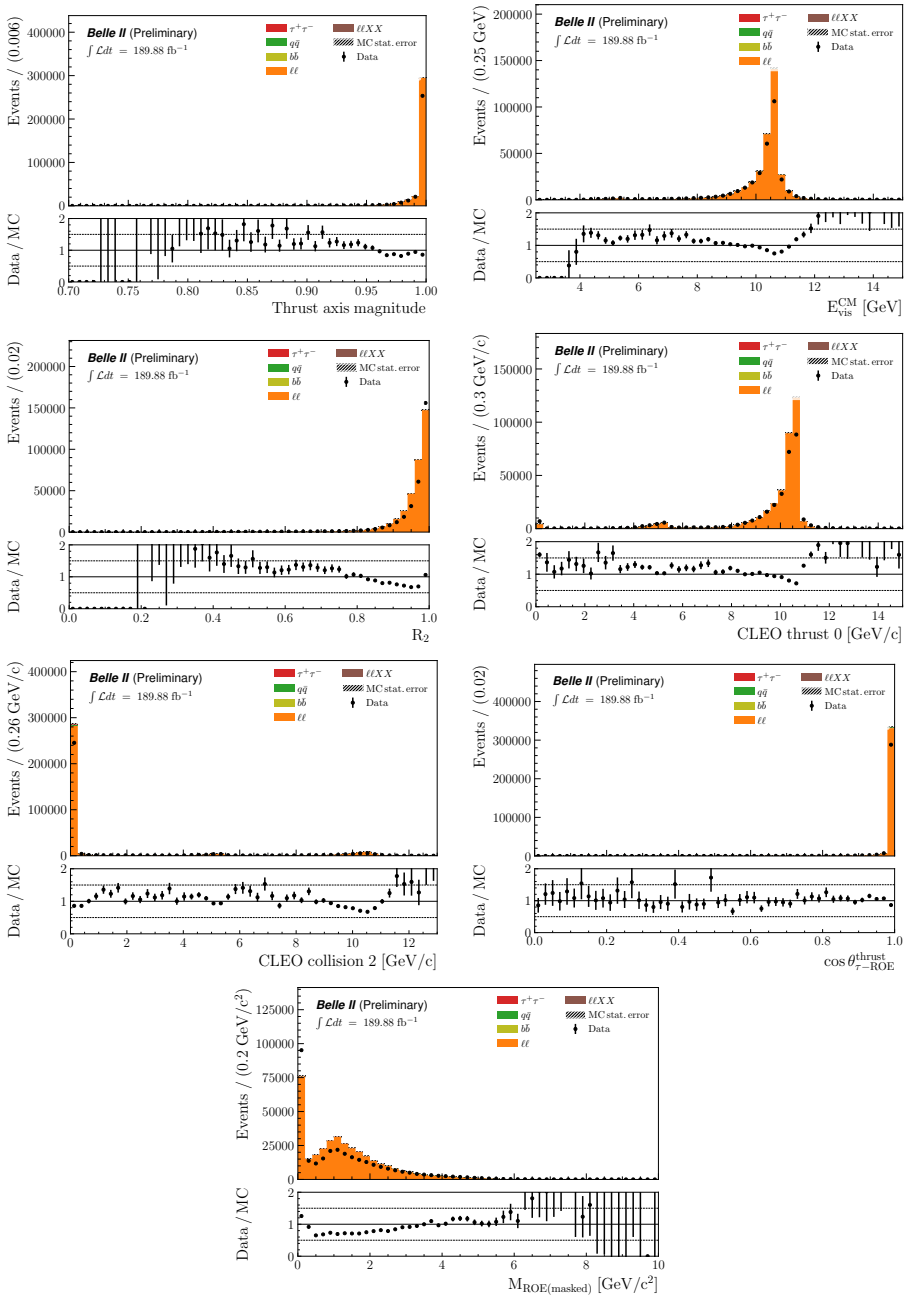


Figure E.1: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

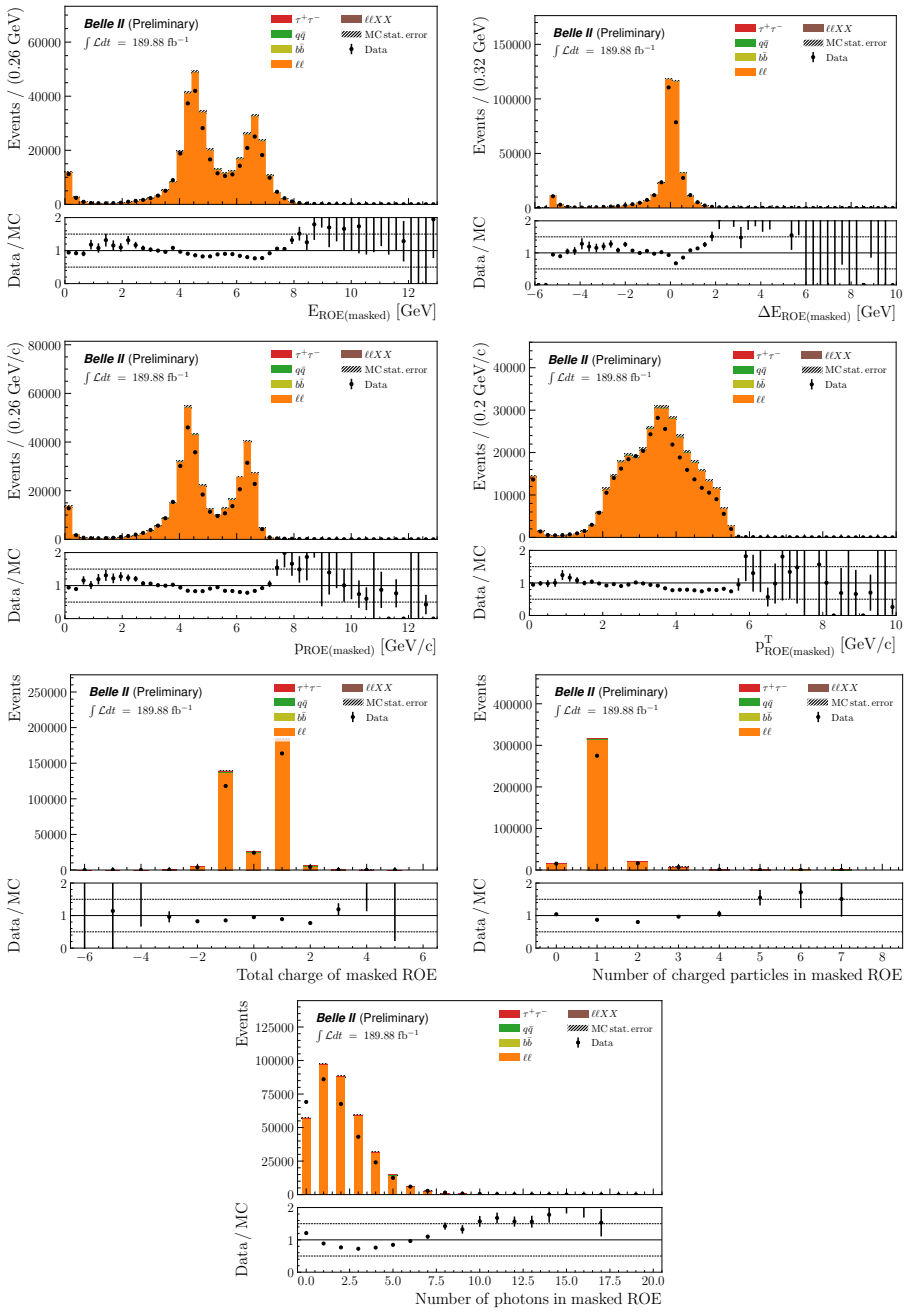


Figure E.1: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

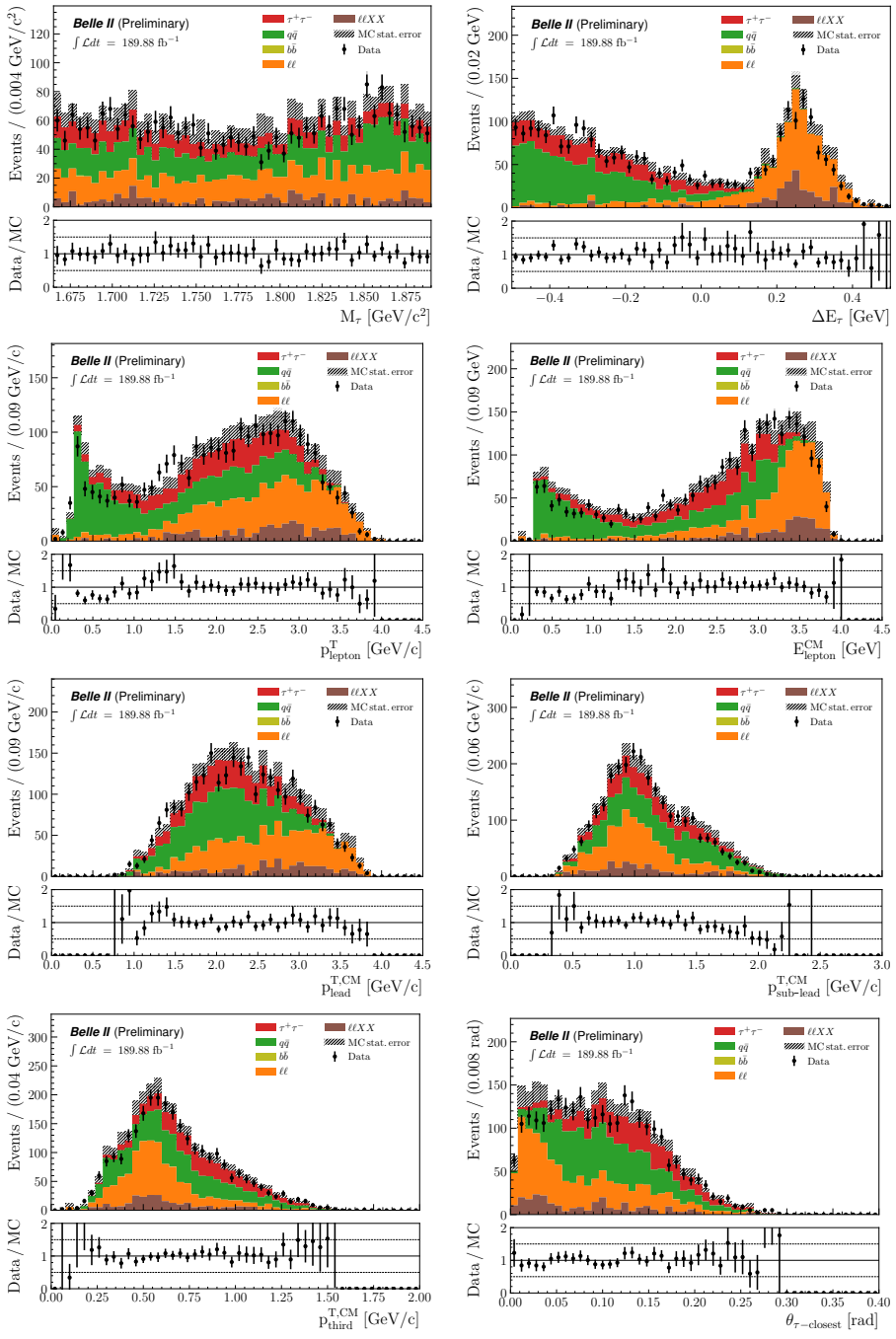


Figure E.2: Data-MC comparison in the $5 - 20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

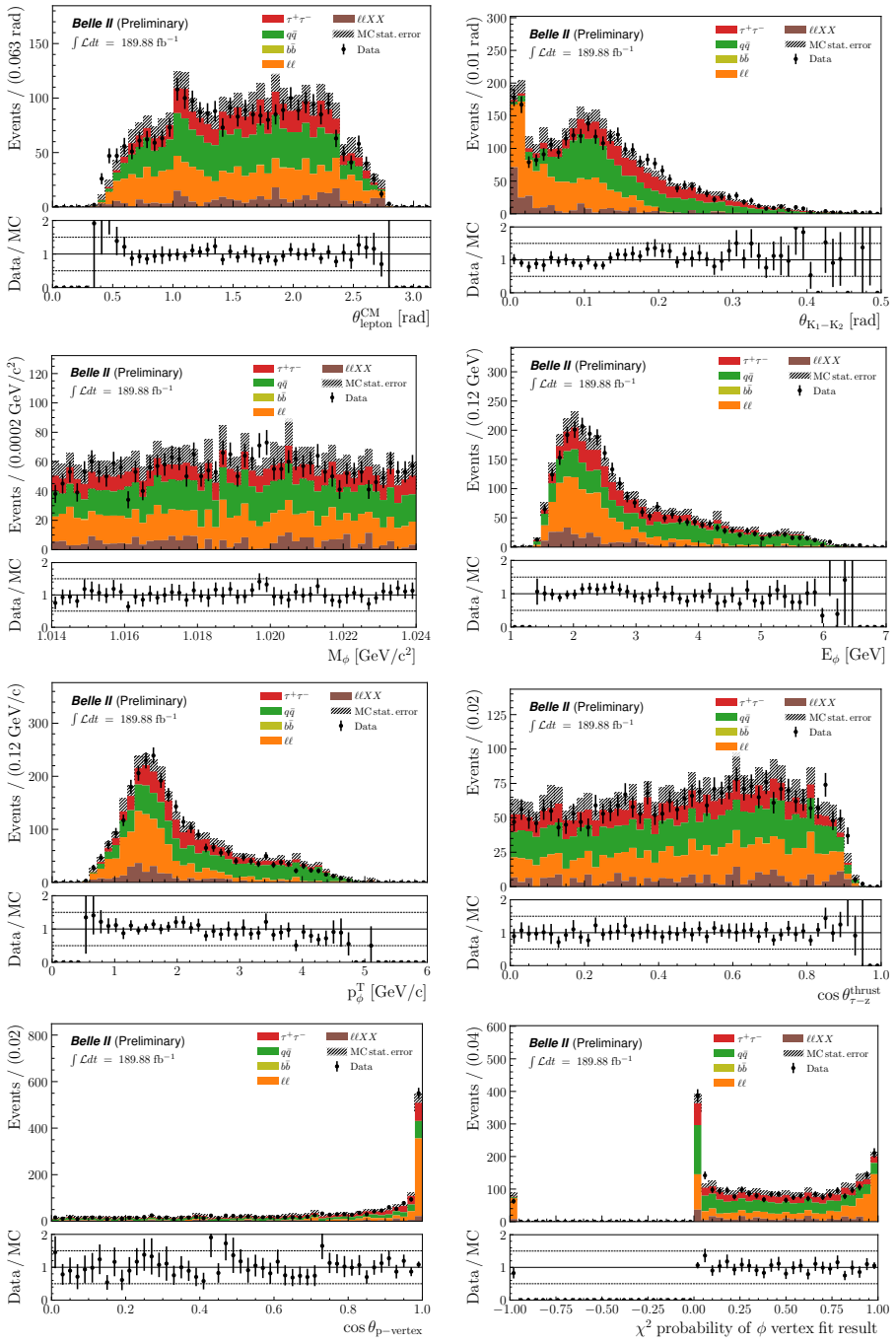


Figure E.2: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

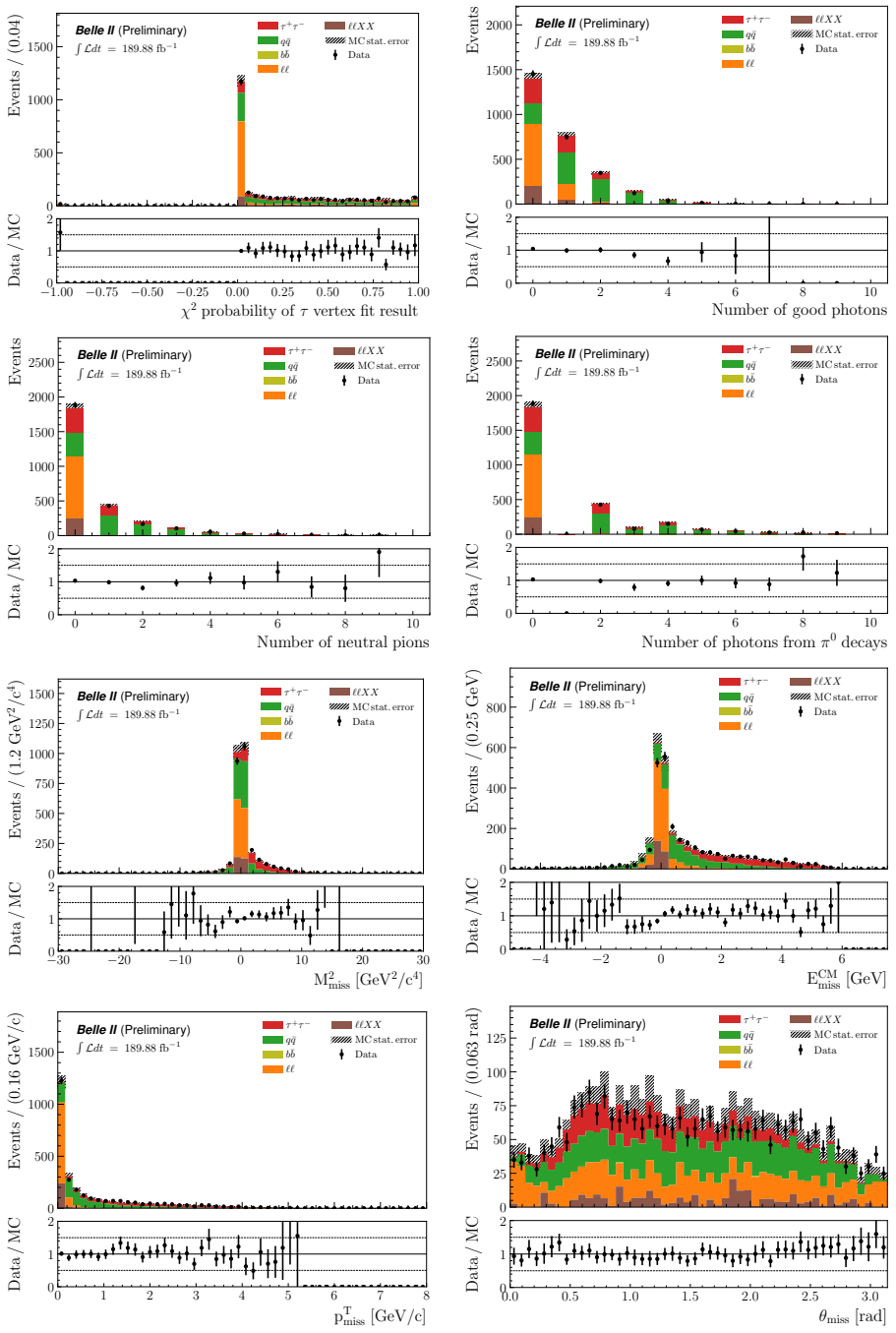


Figure E.2: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

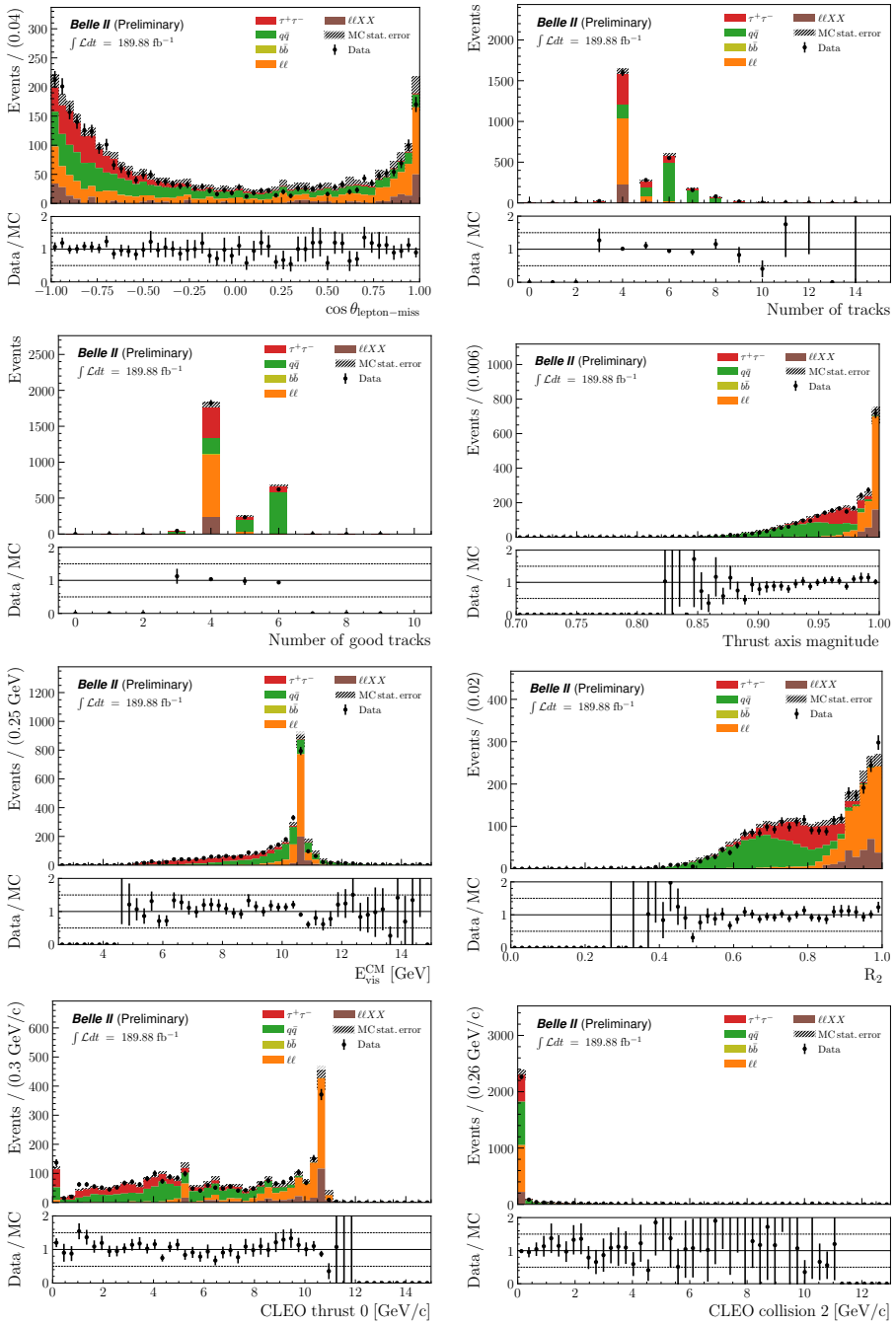


Figure E.2: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

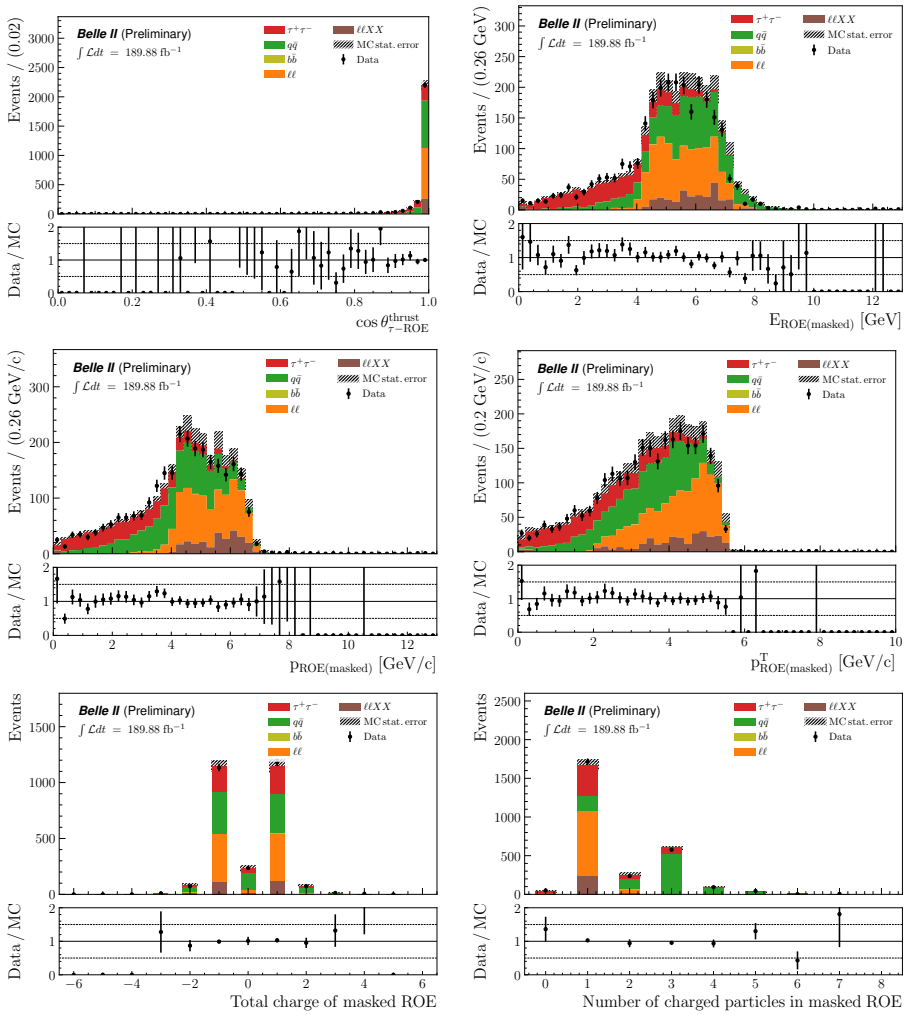


Figure E.2: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, before the preselection.

After the preselection

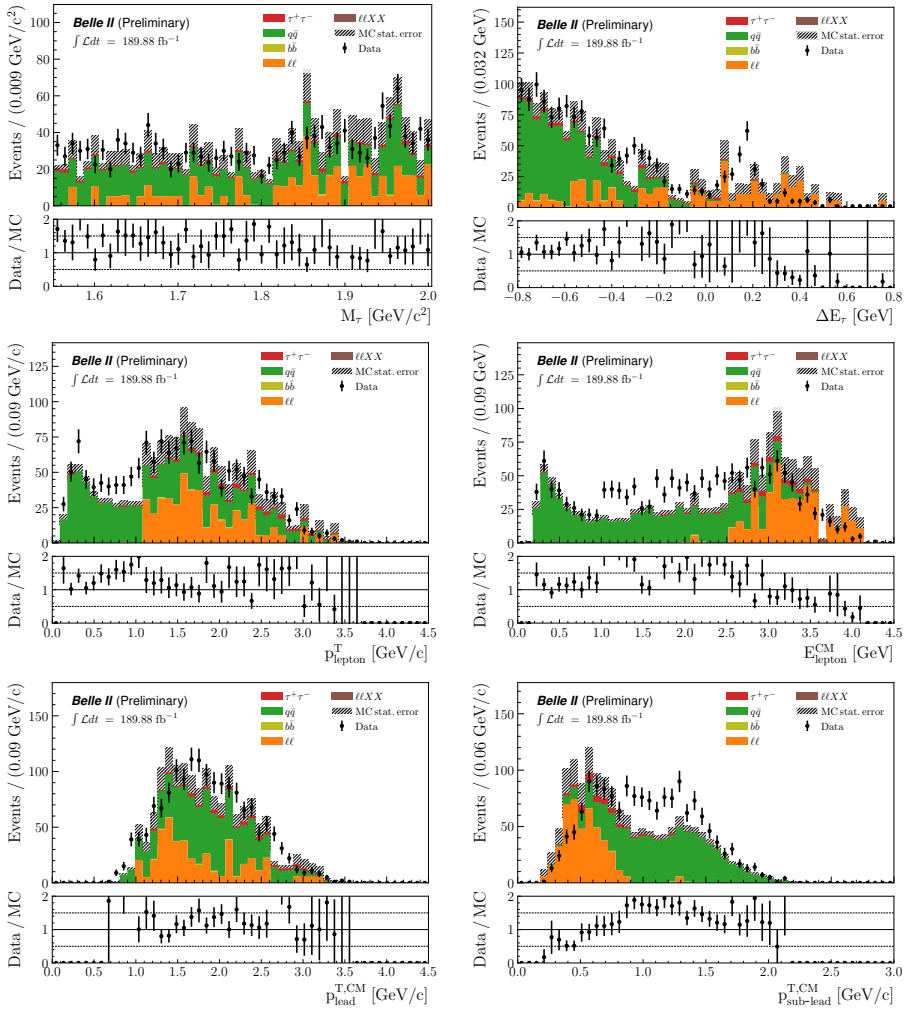


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

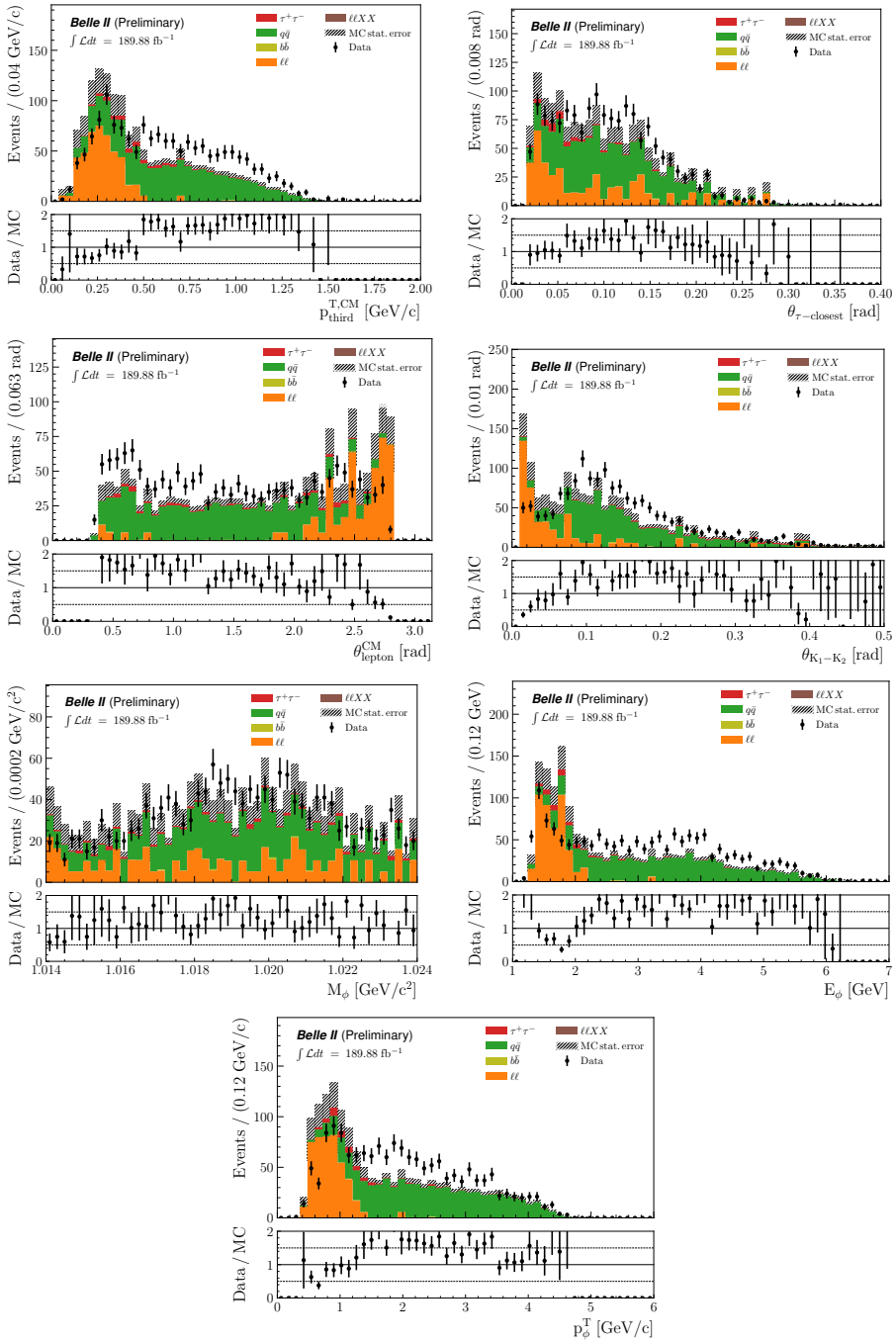


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

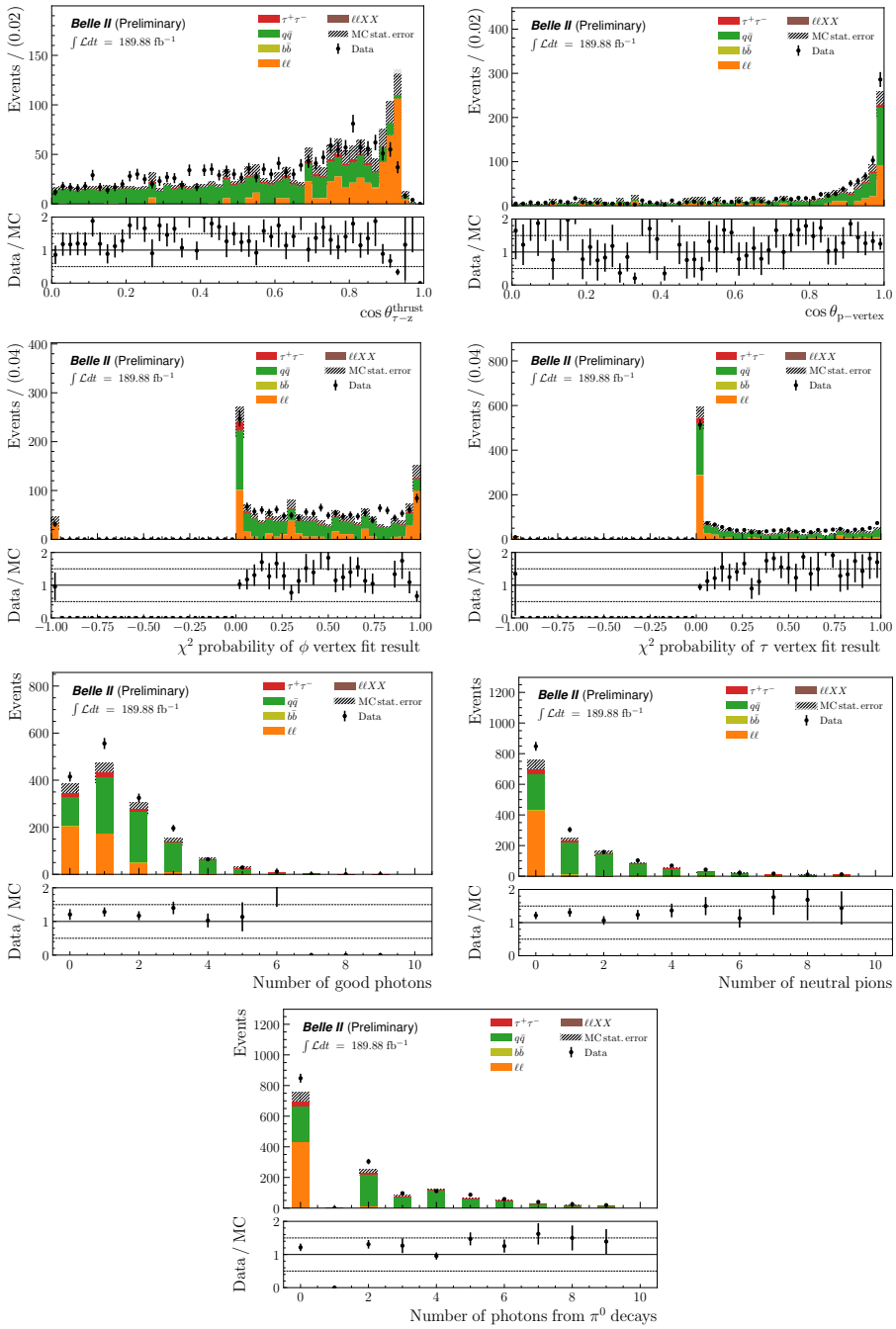


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

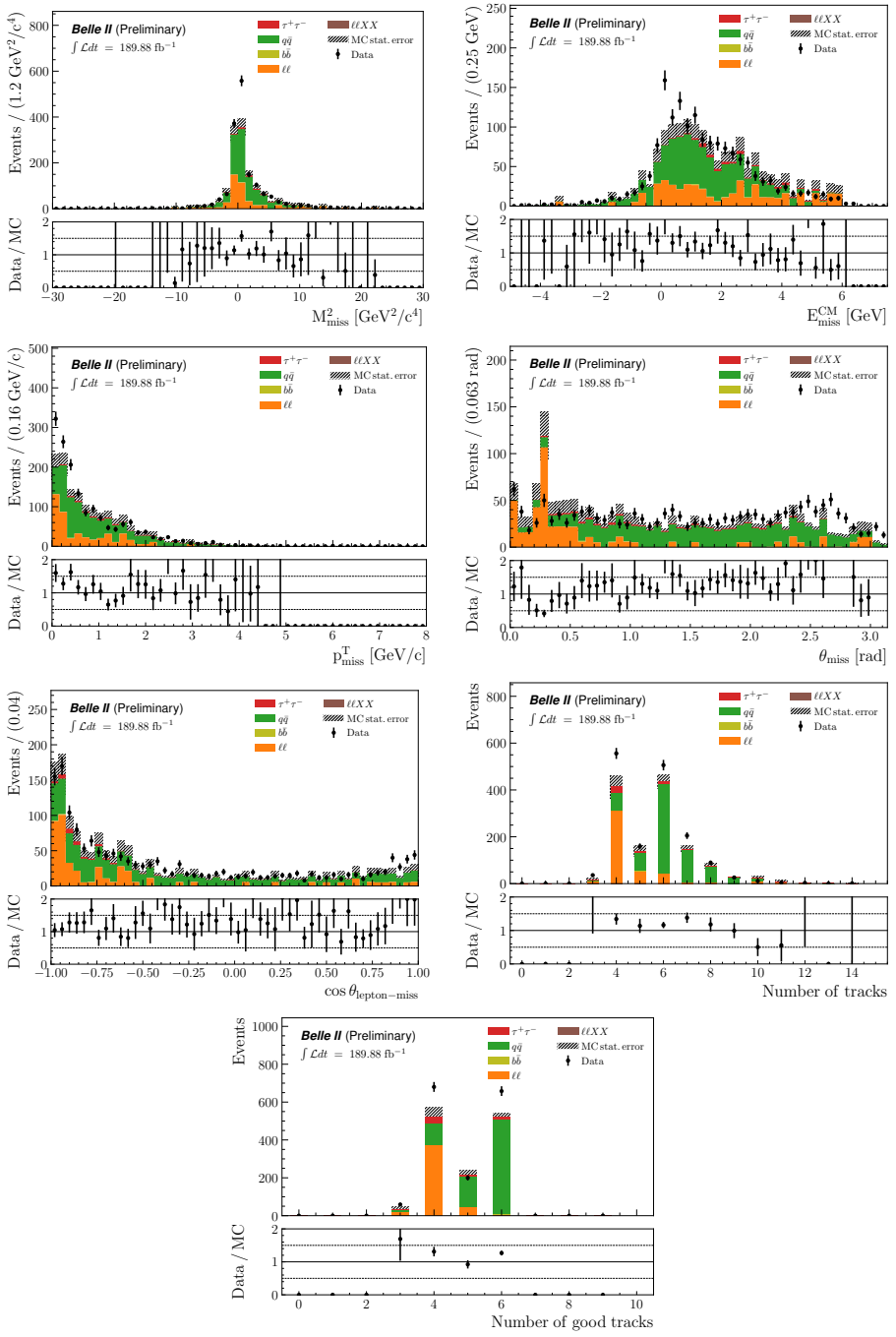


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

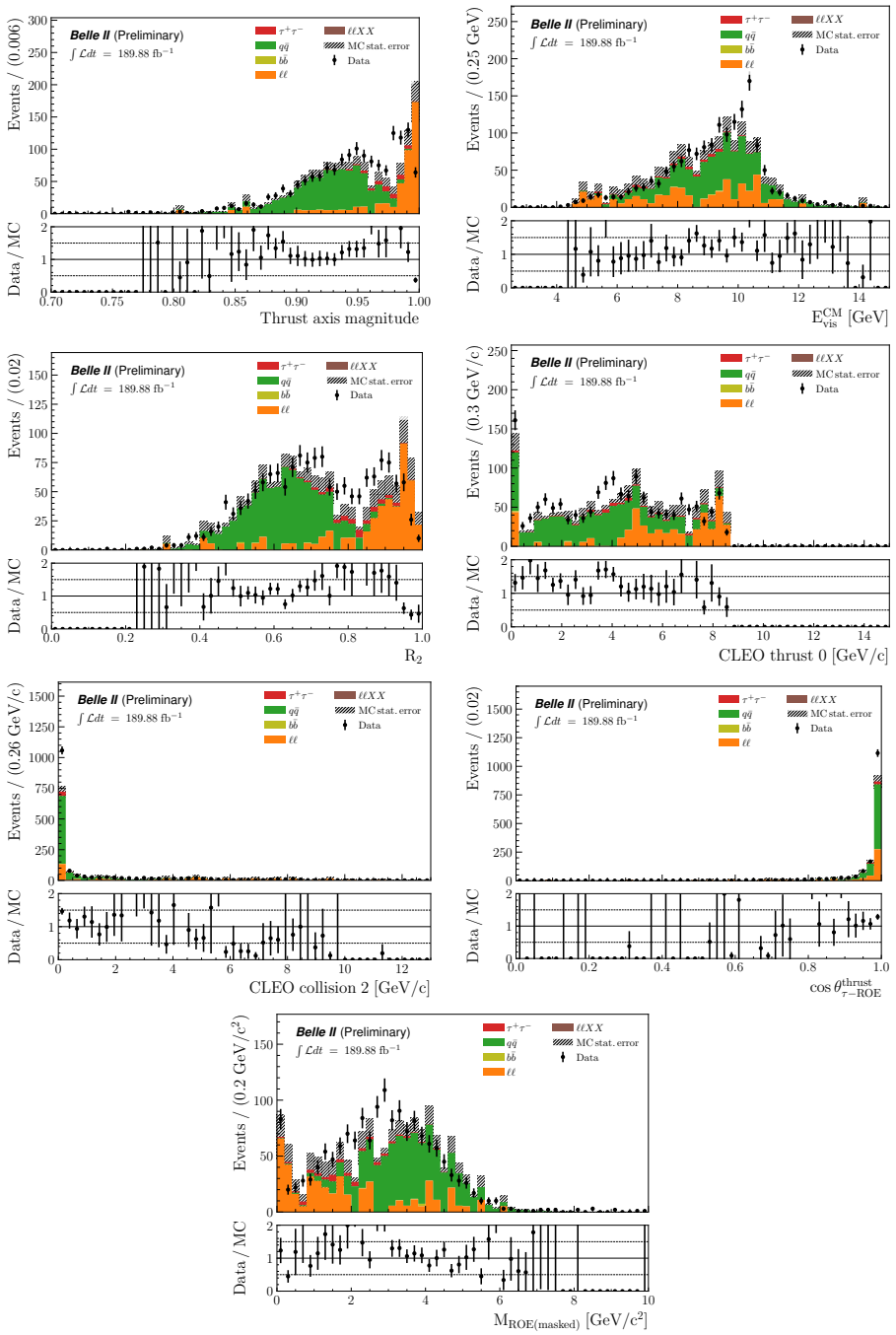


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

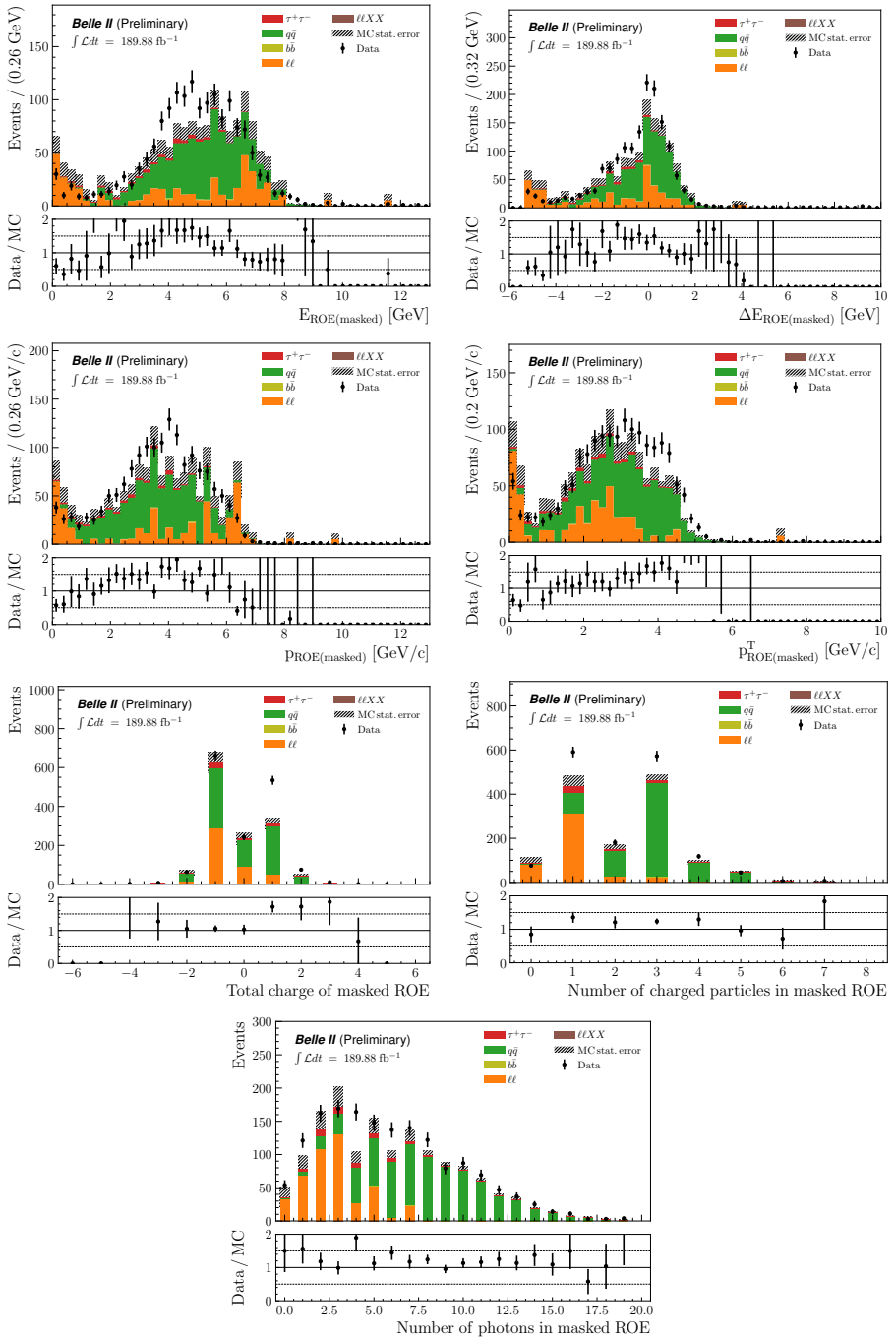


Figure E.3: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

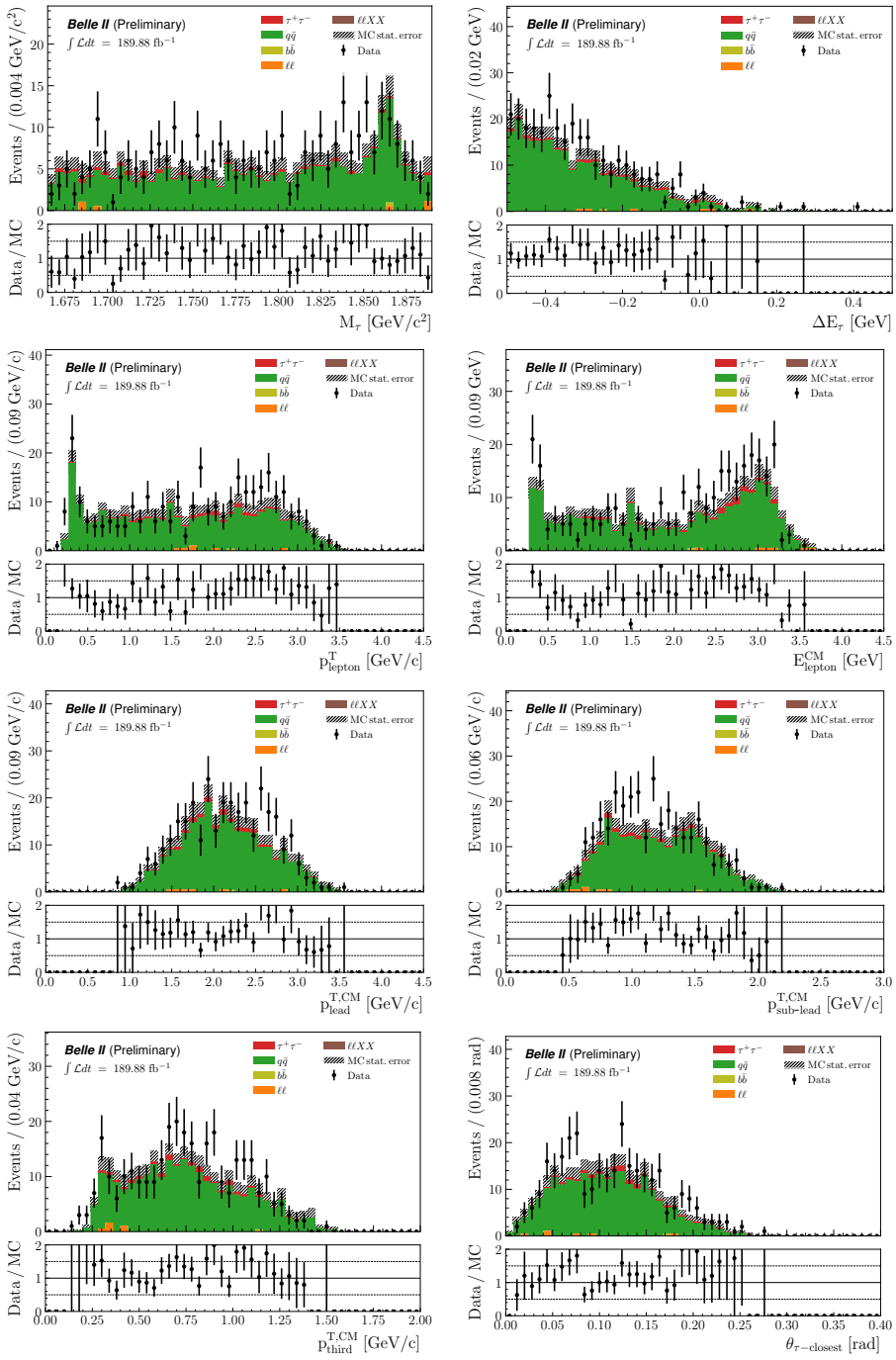


Figure E.4: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

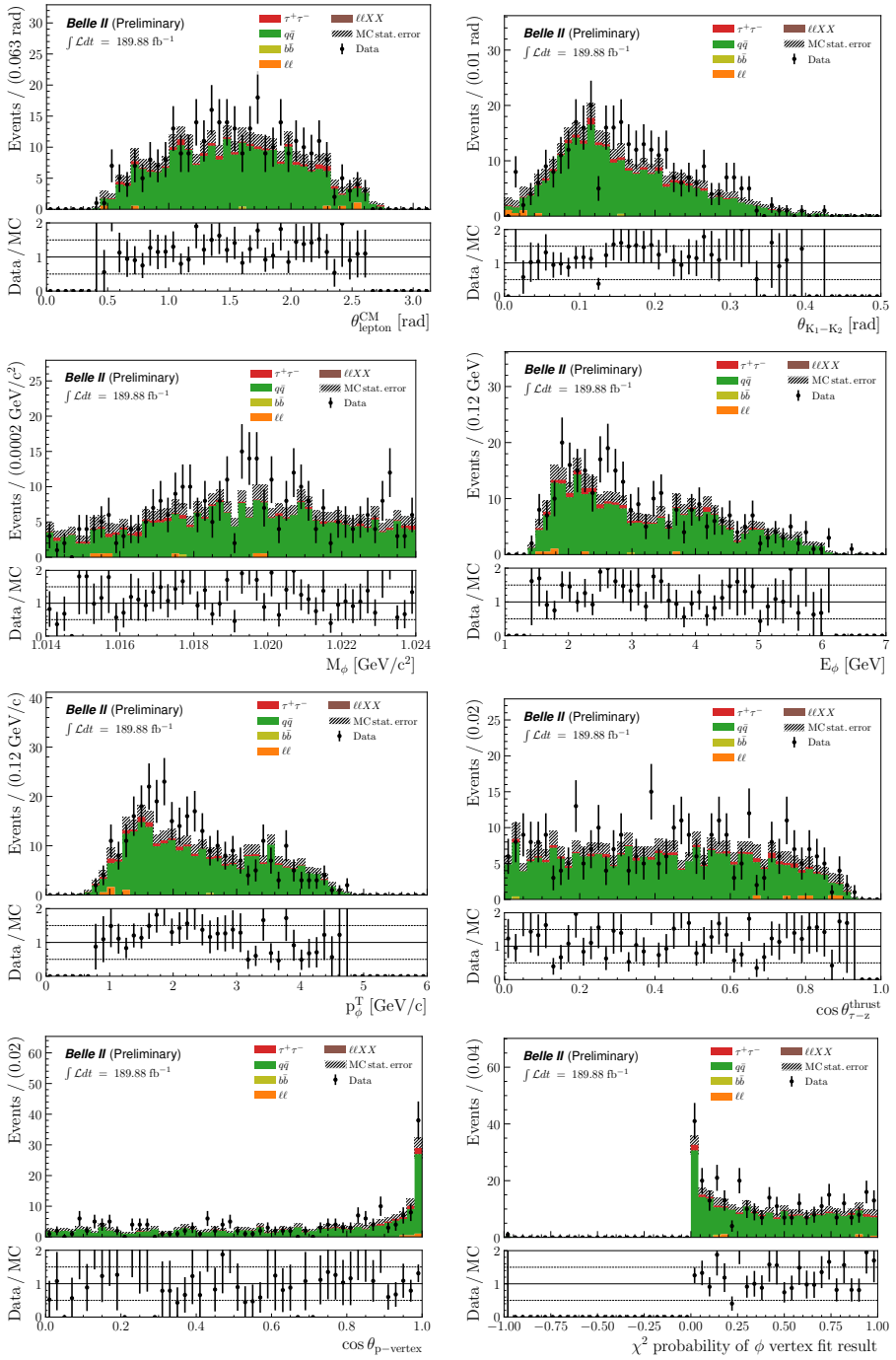


Figure E.4: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

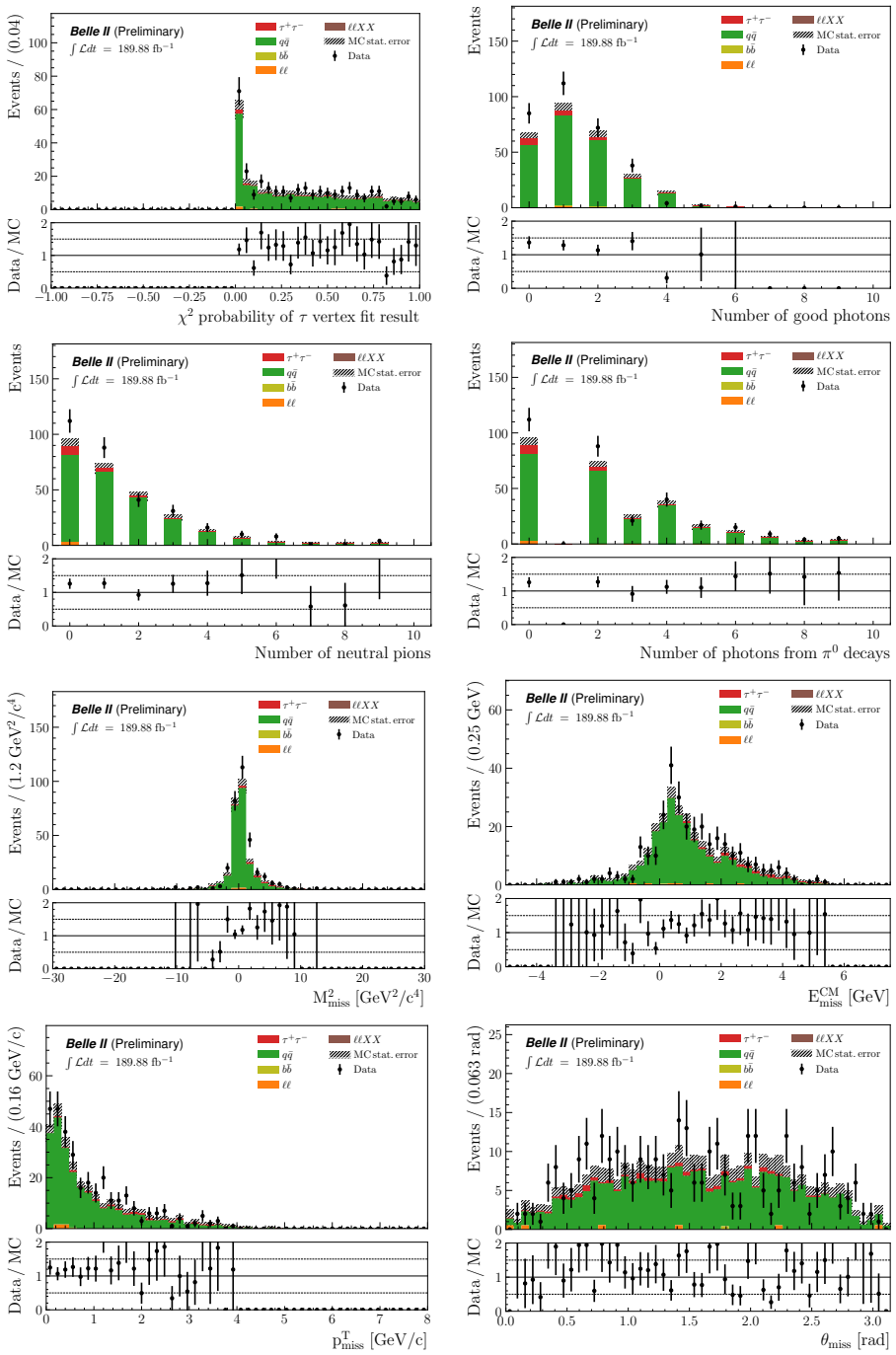


Figure E.4: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

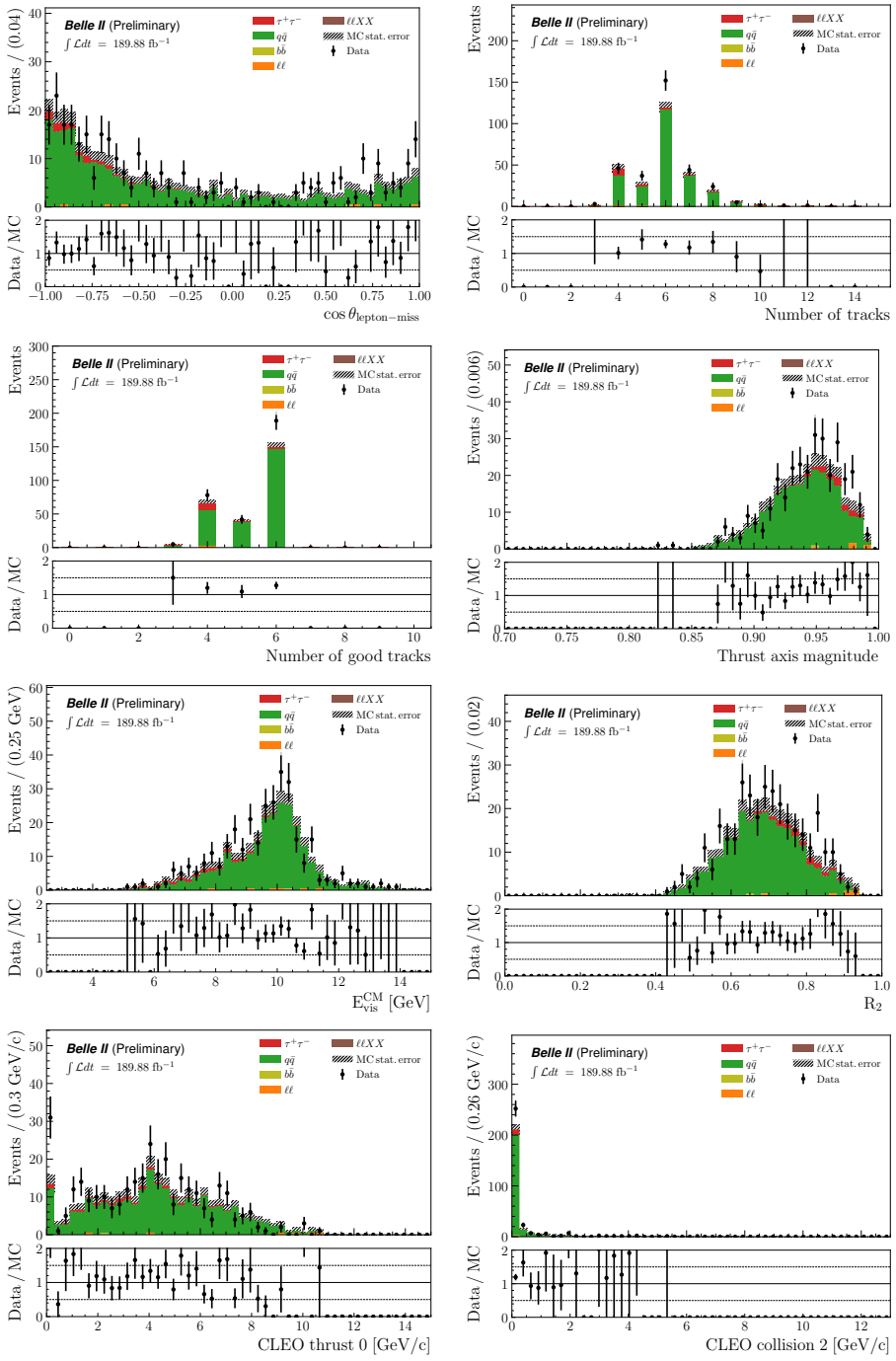


Figure E.4: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

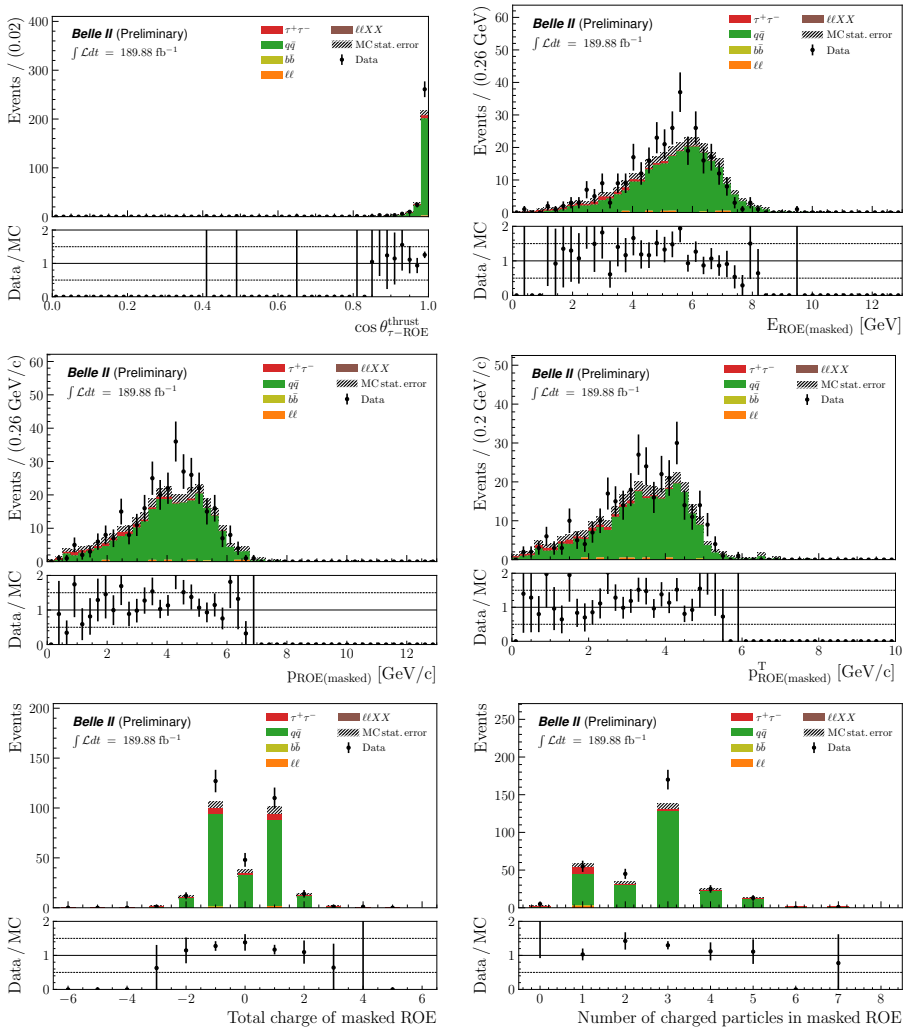


Figure E.4: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the preselection.

Before the "data BDT" ($\tau \to e\phi$)

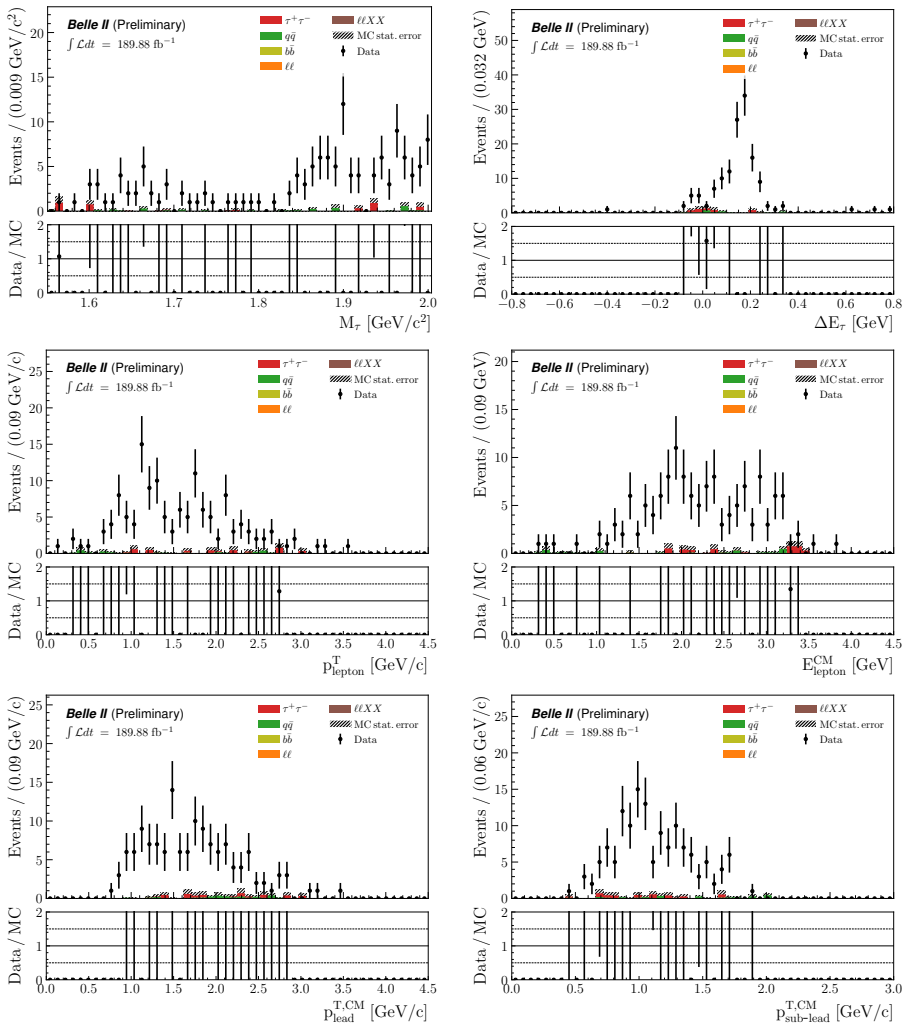


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

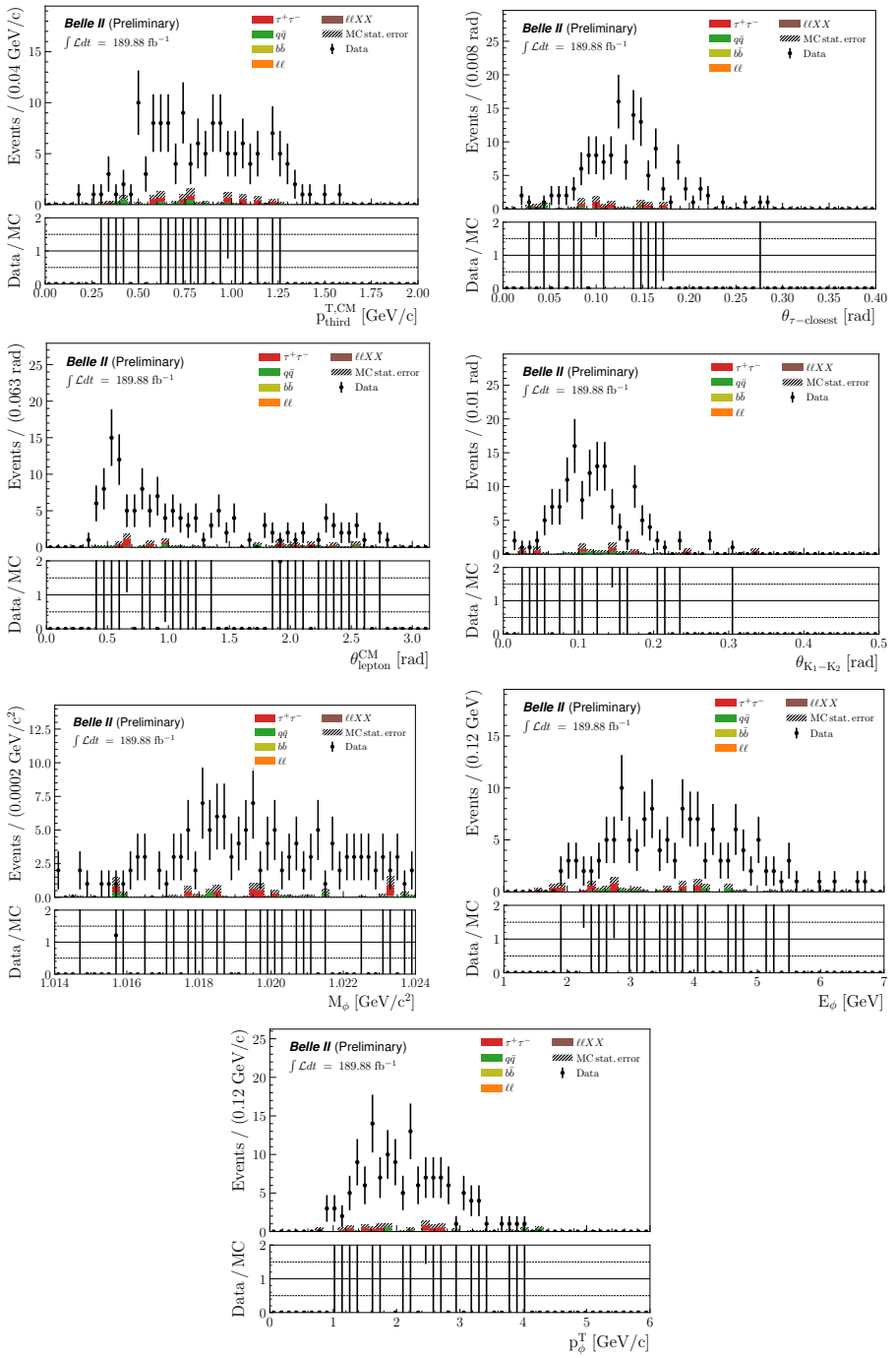


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

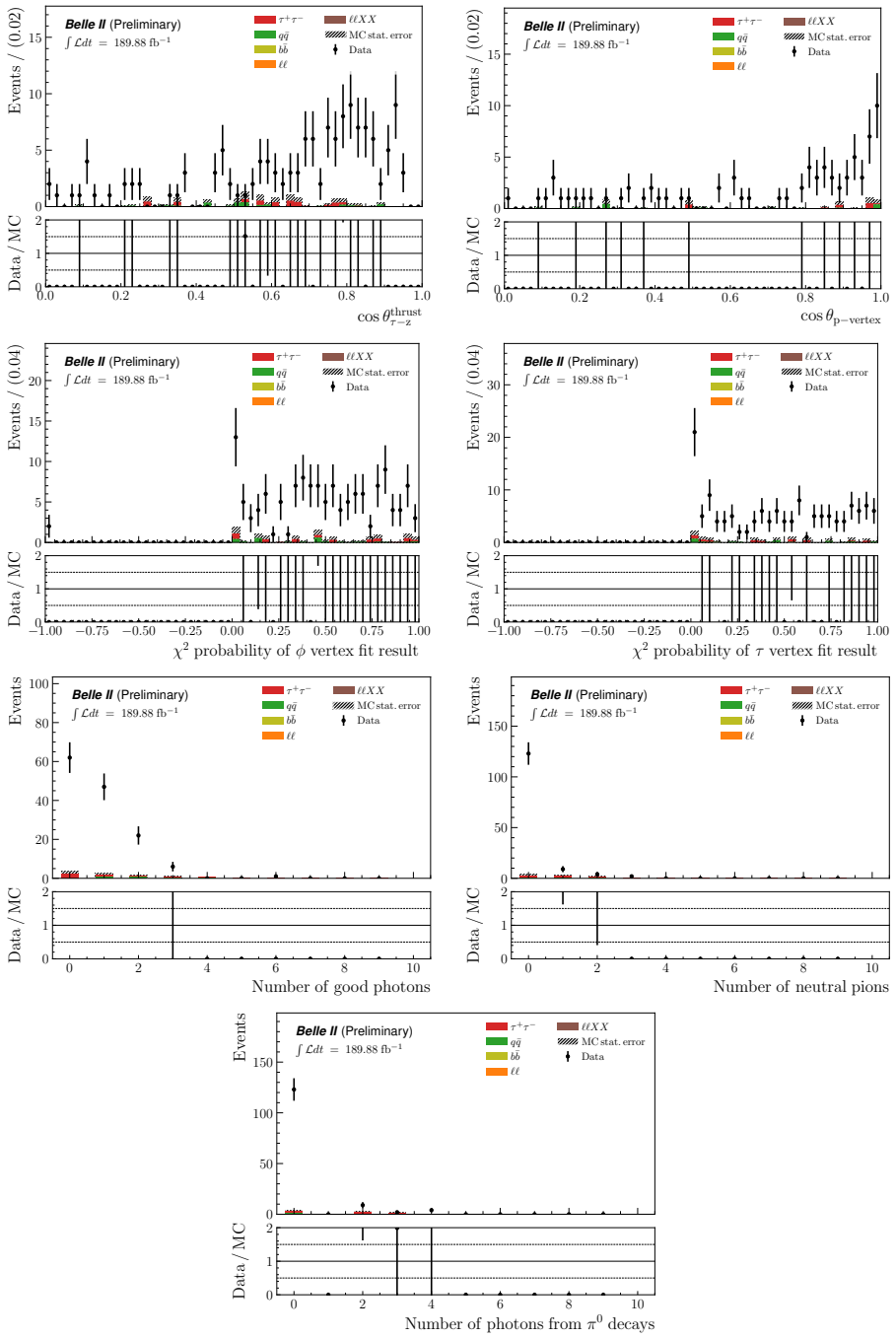


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

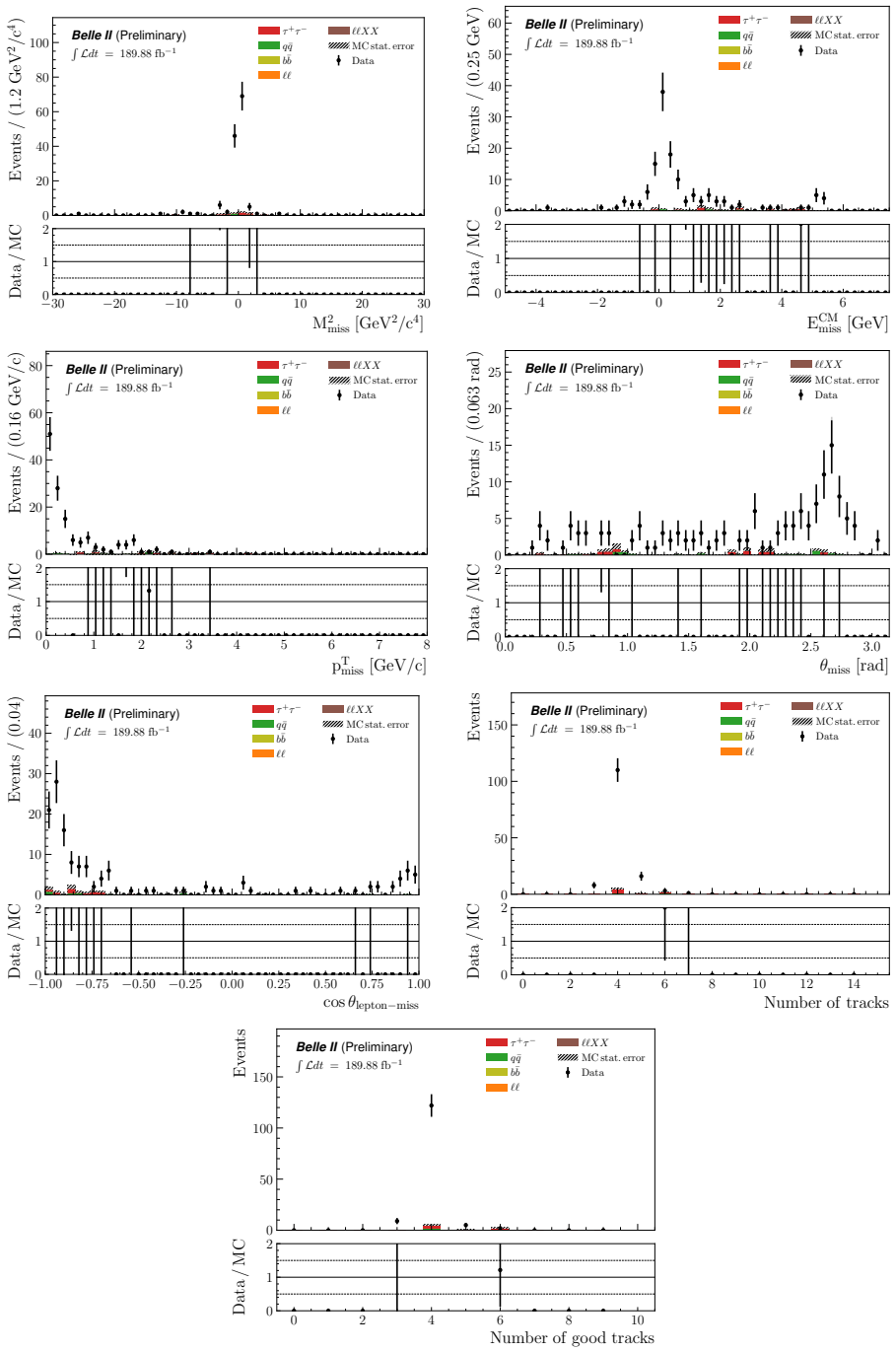


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

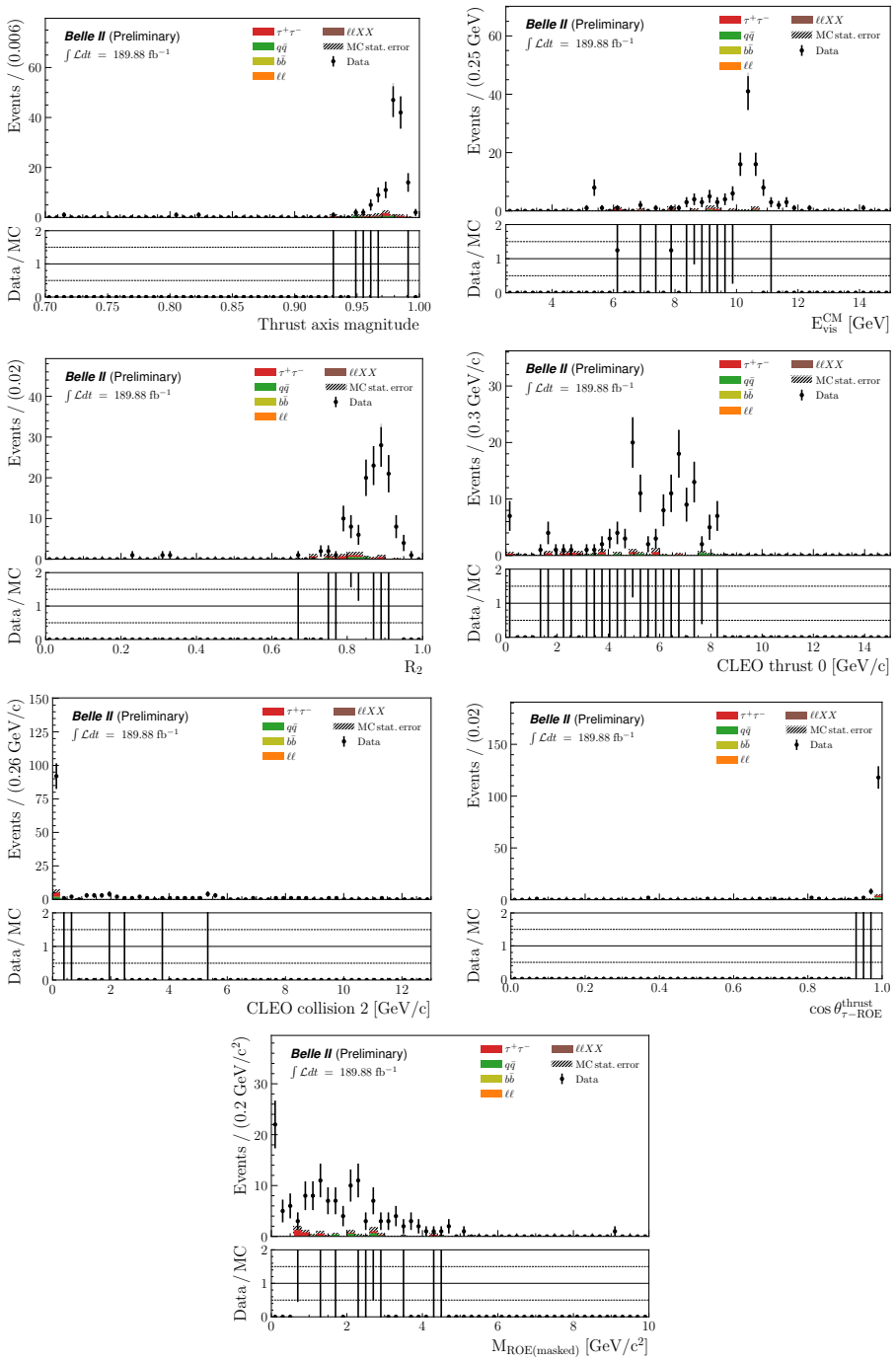


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

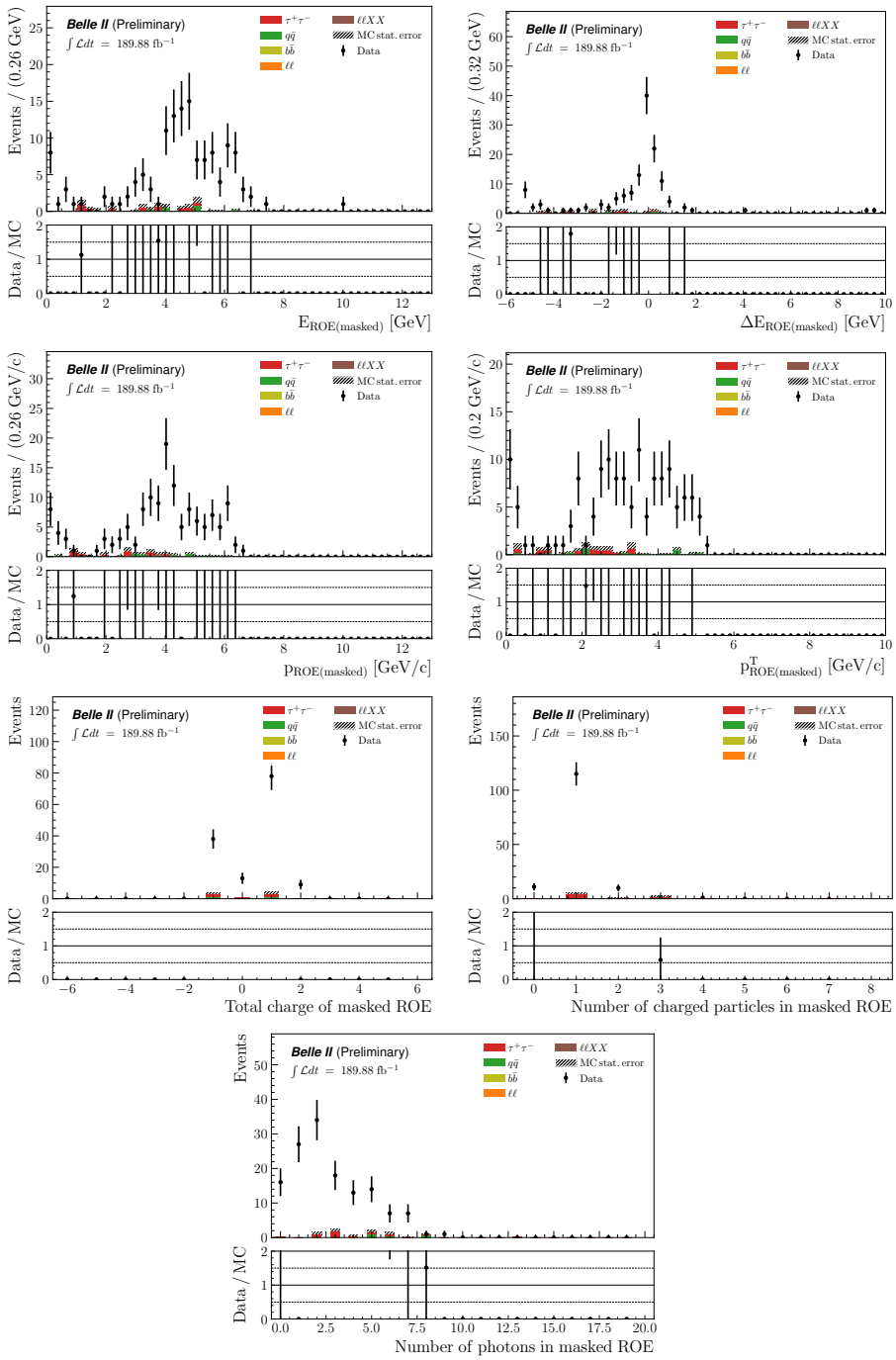


Figure E.5: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the first BDT, before applying the "data BDT" criterium.

After the full selection

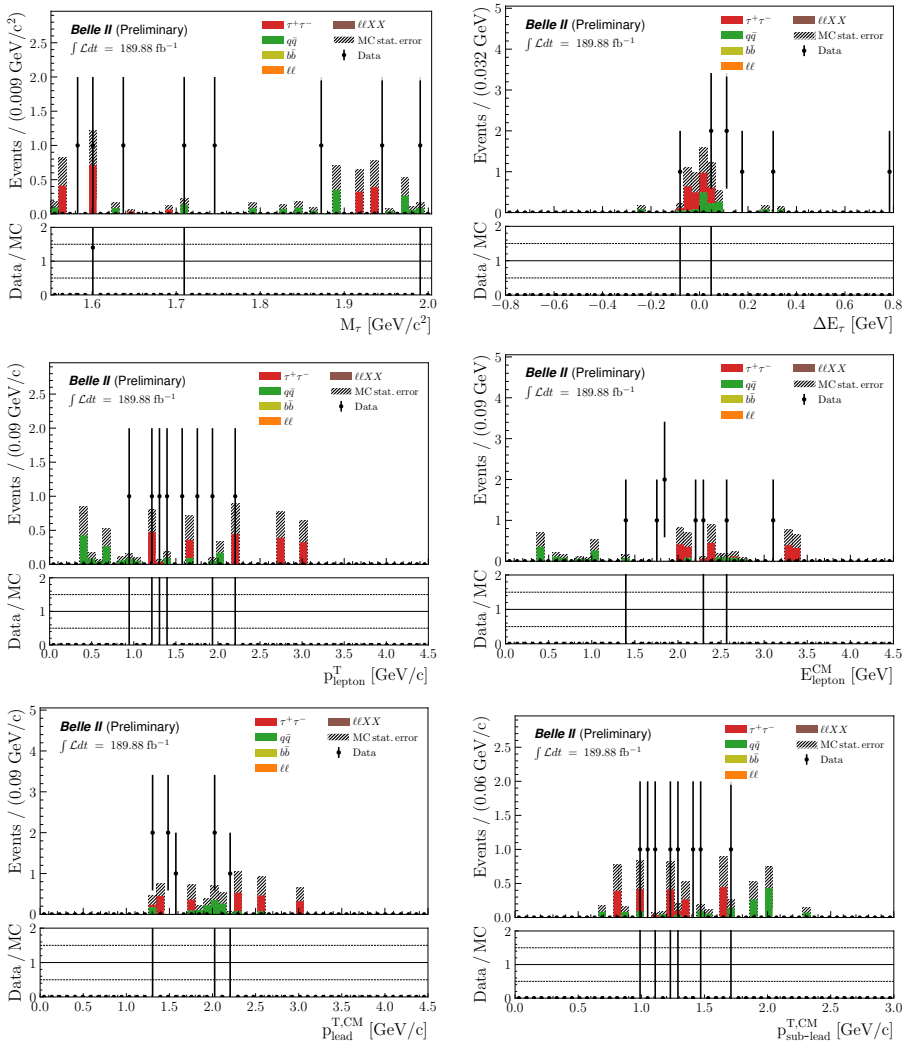


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

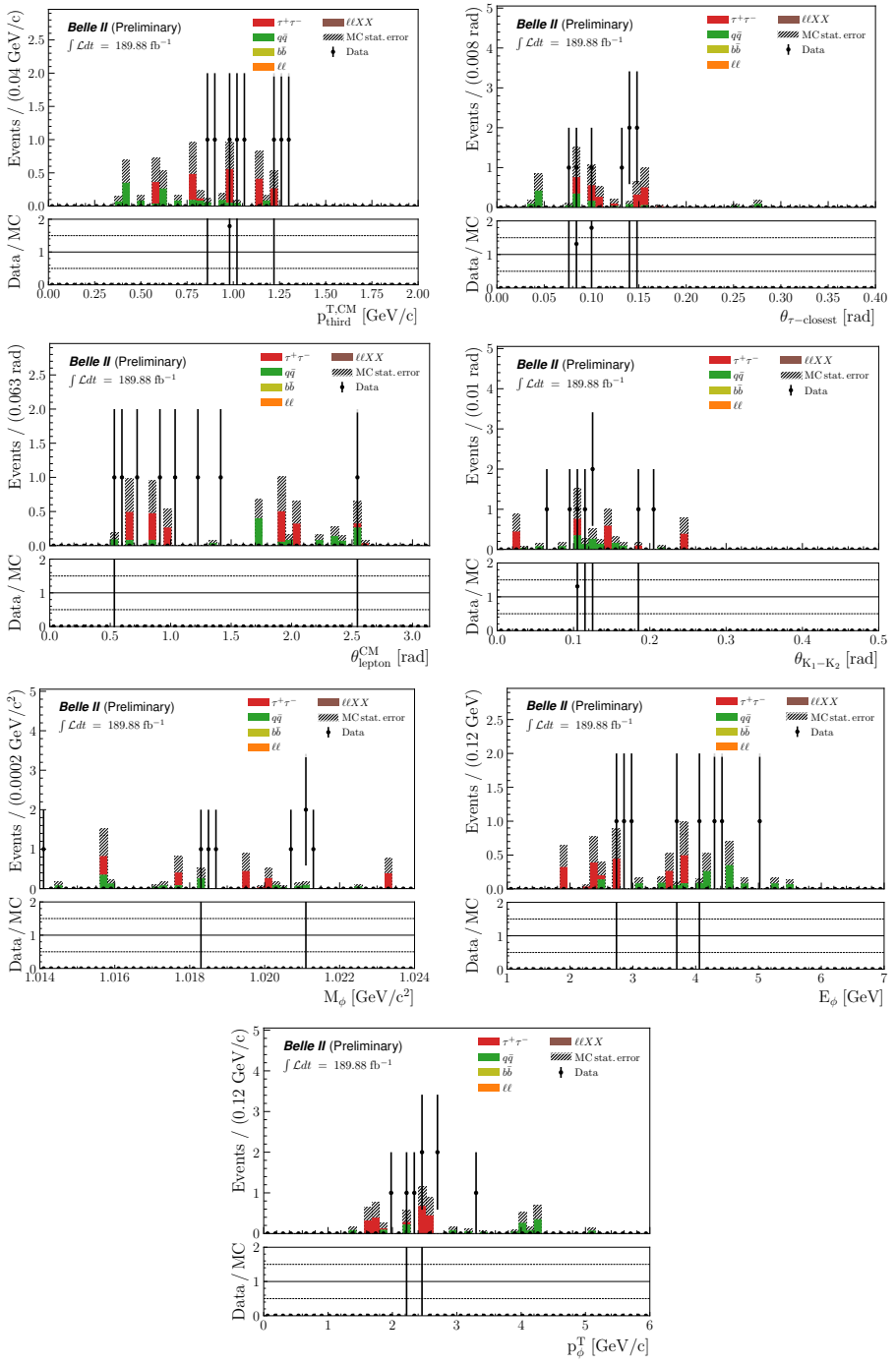


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

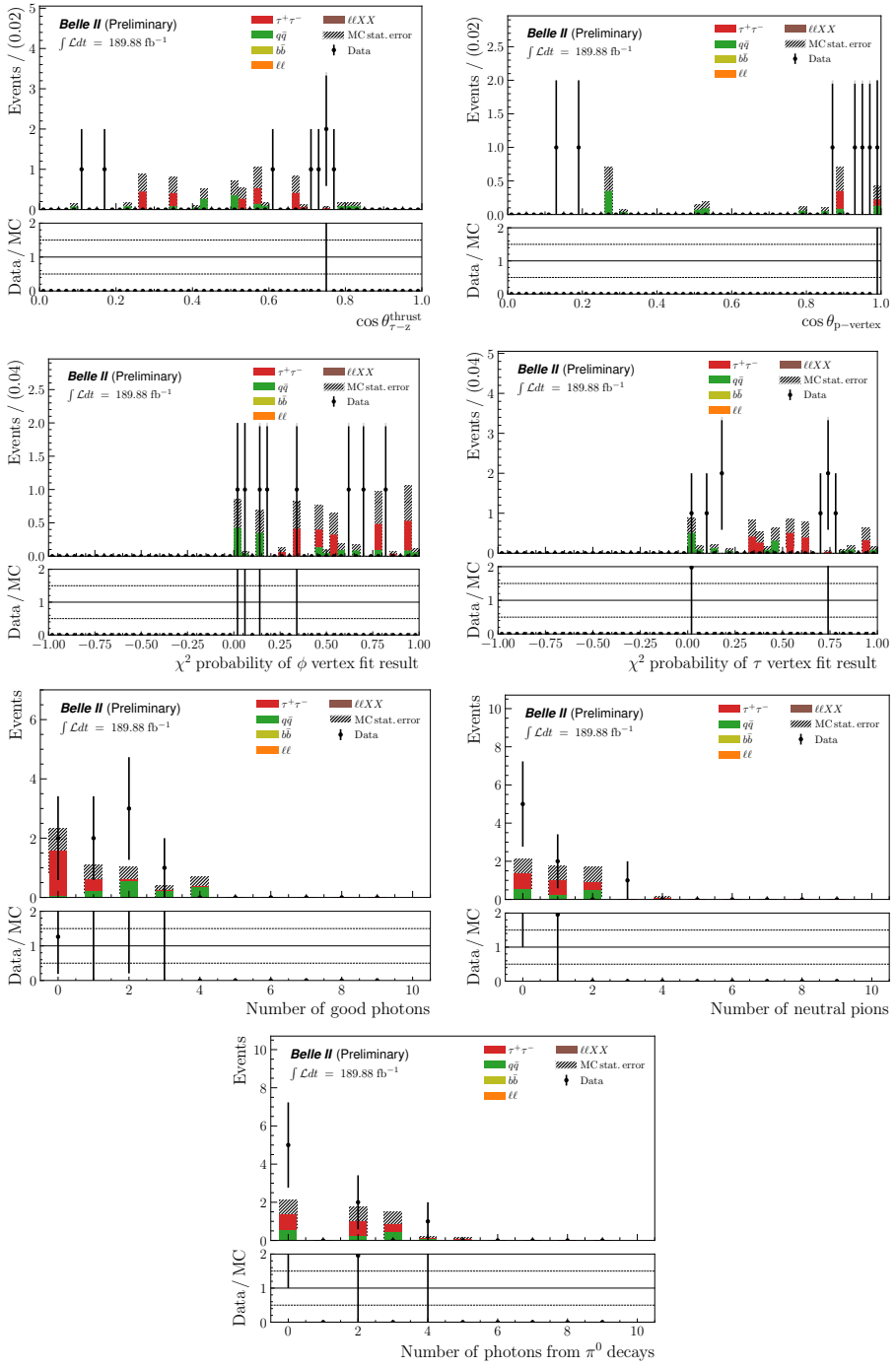


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

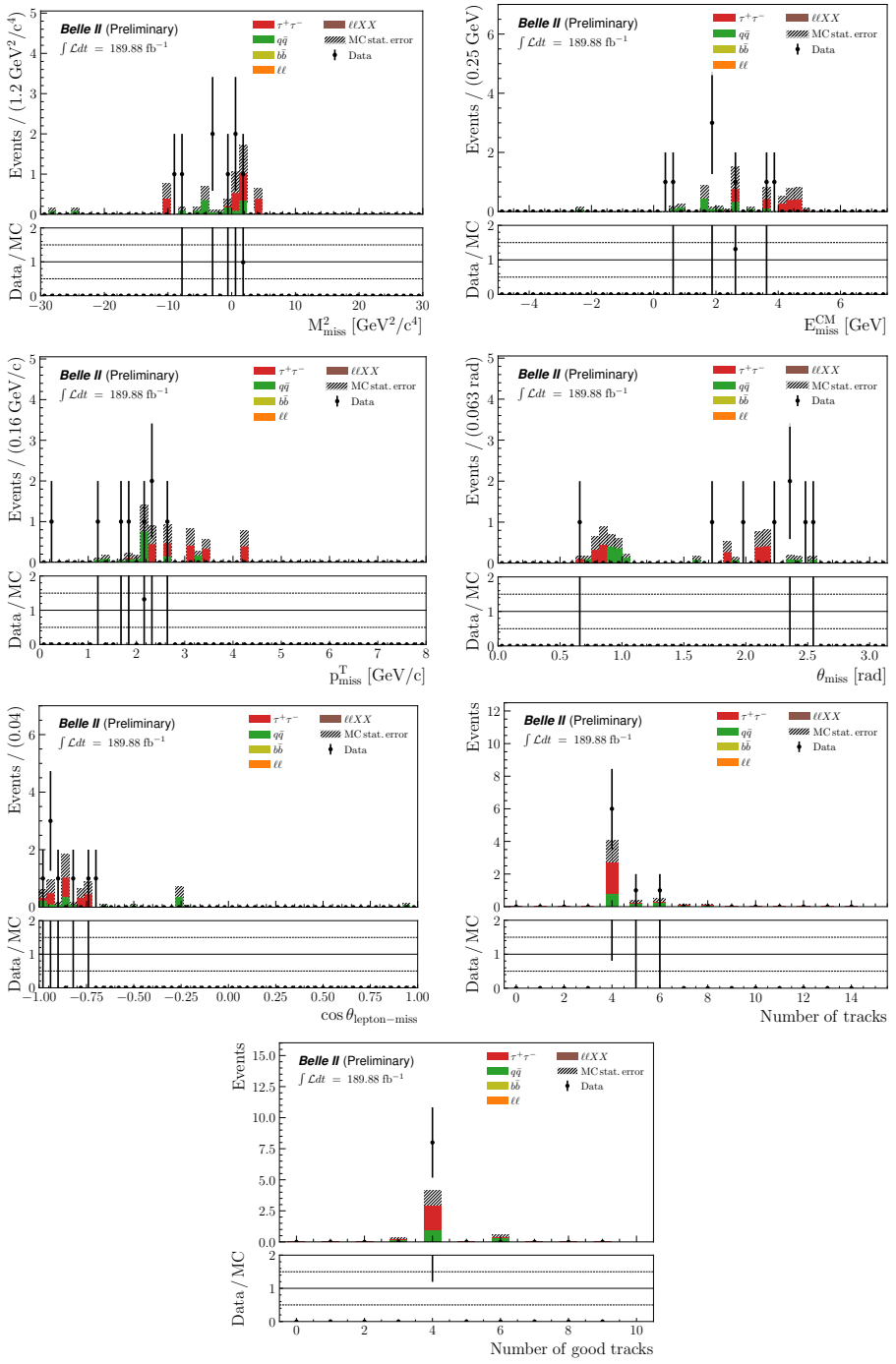


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

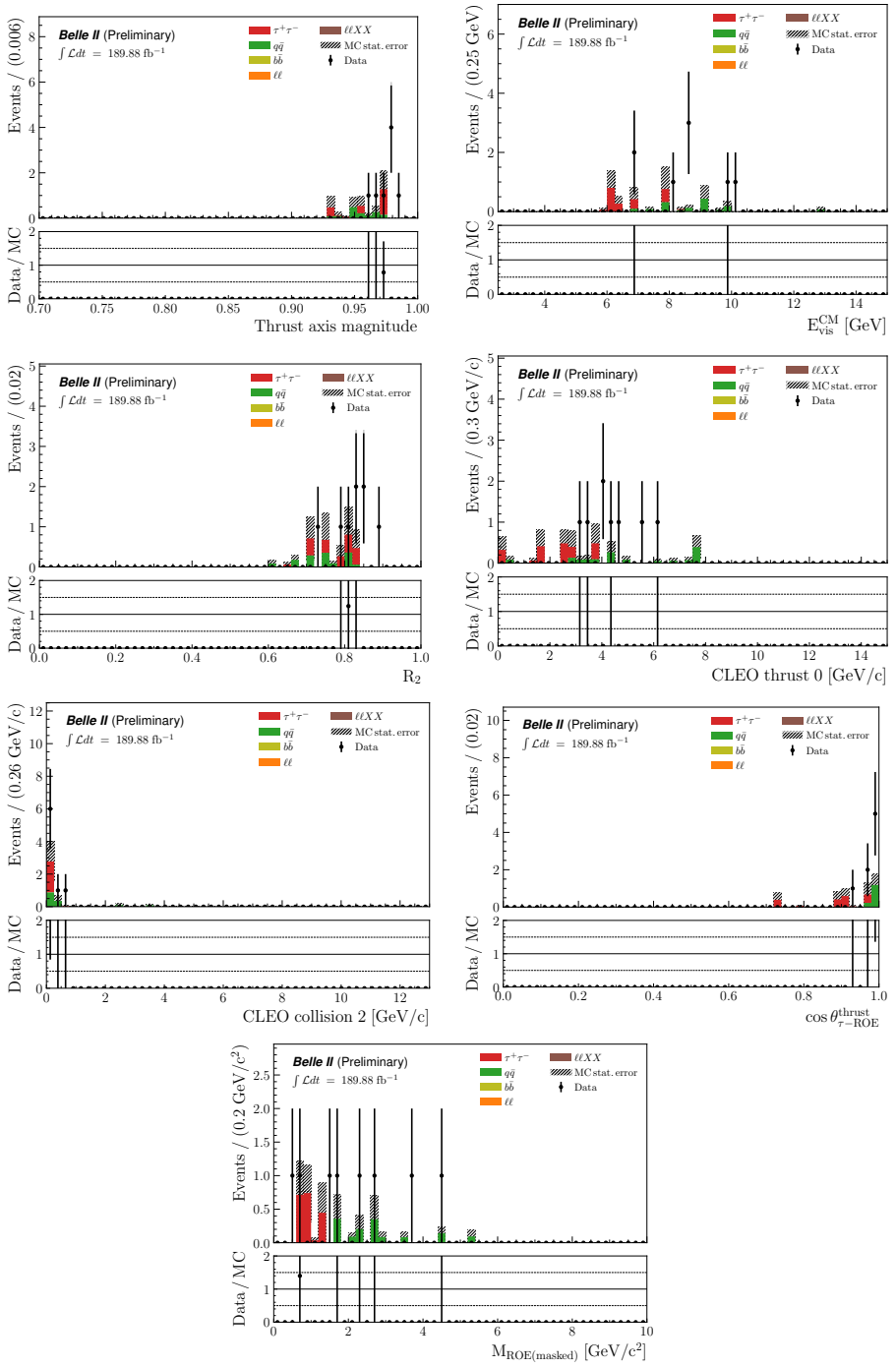


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

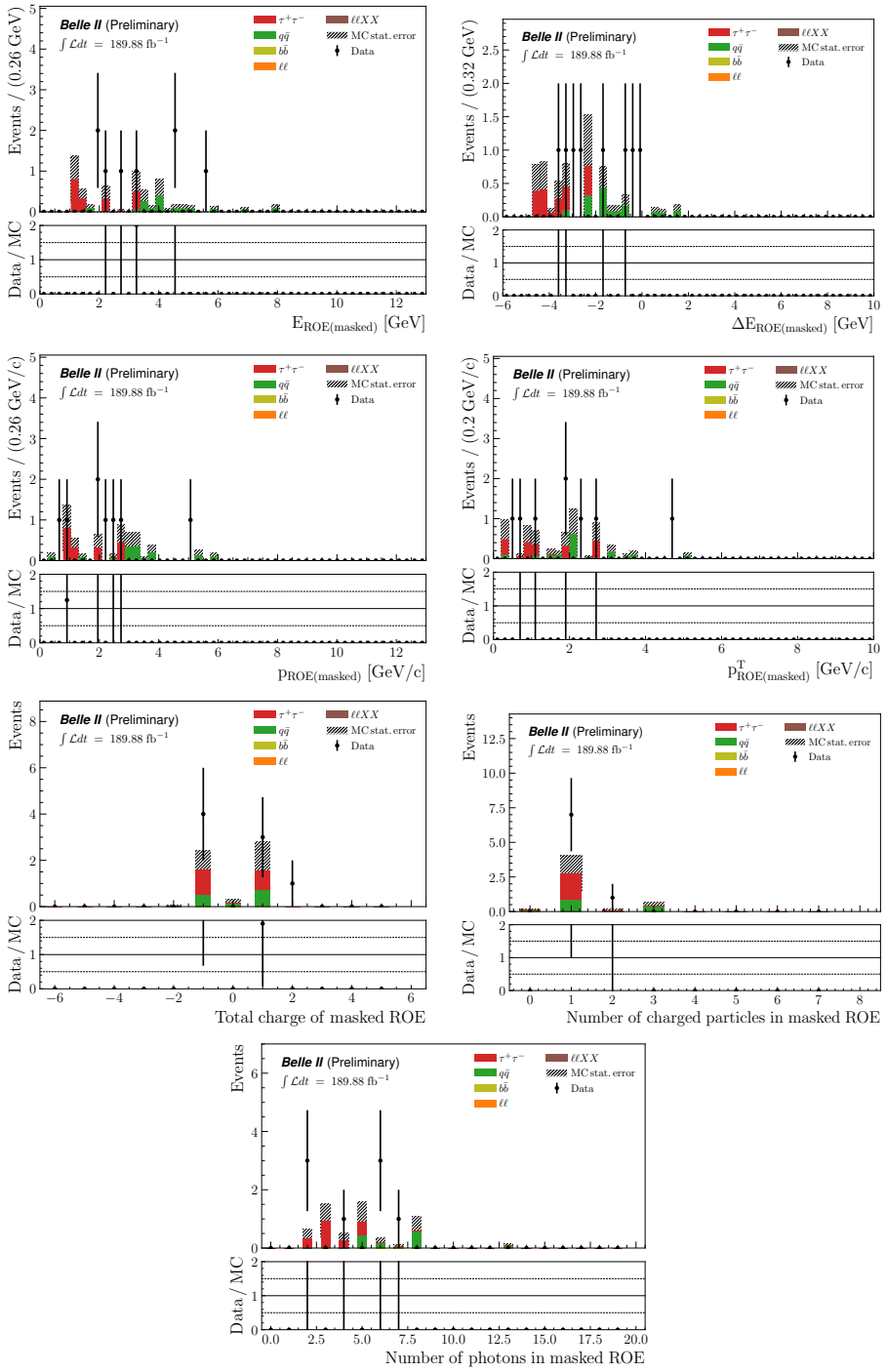


Figure E.6: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to e\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

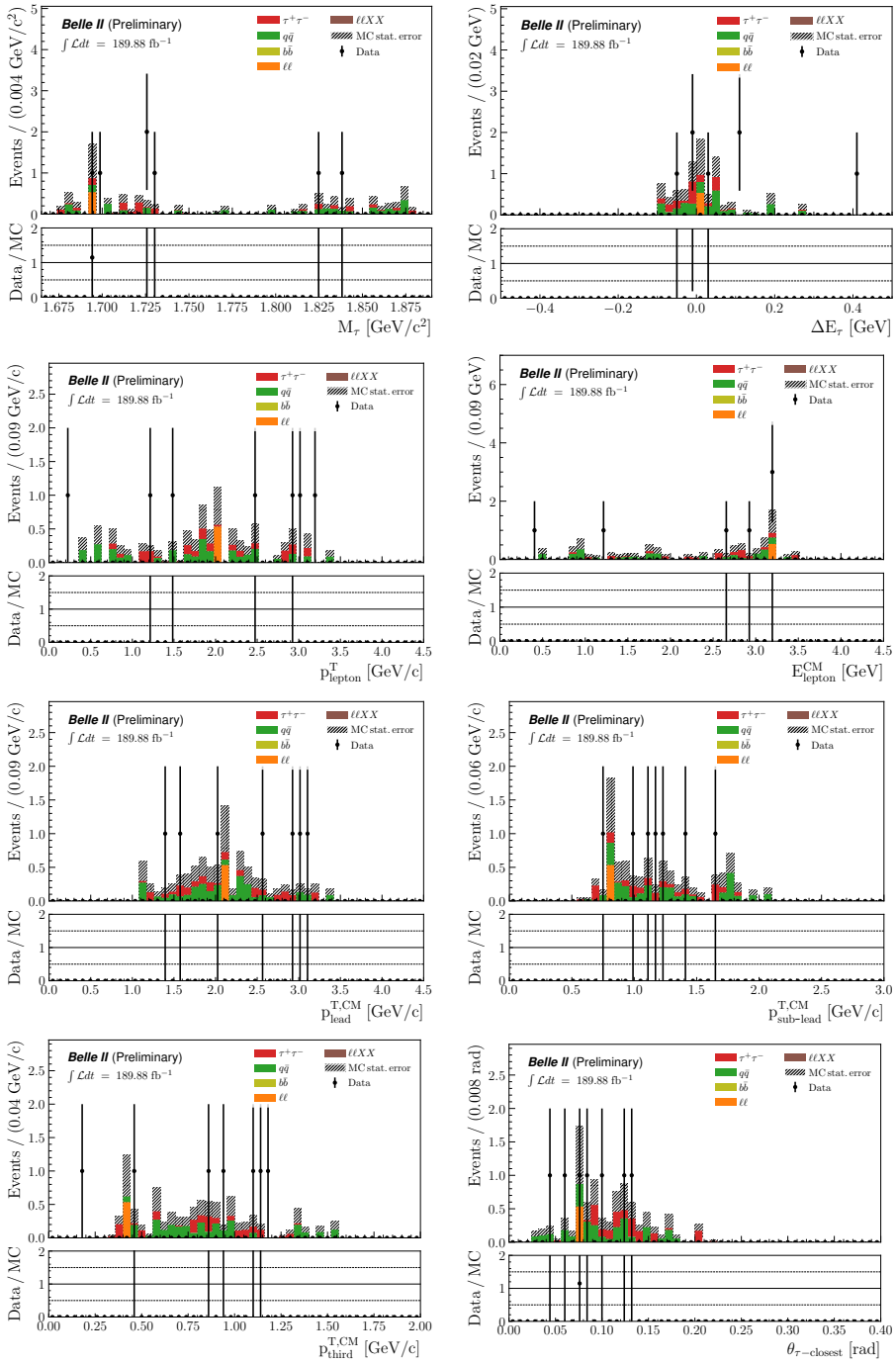


Figure E.7: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

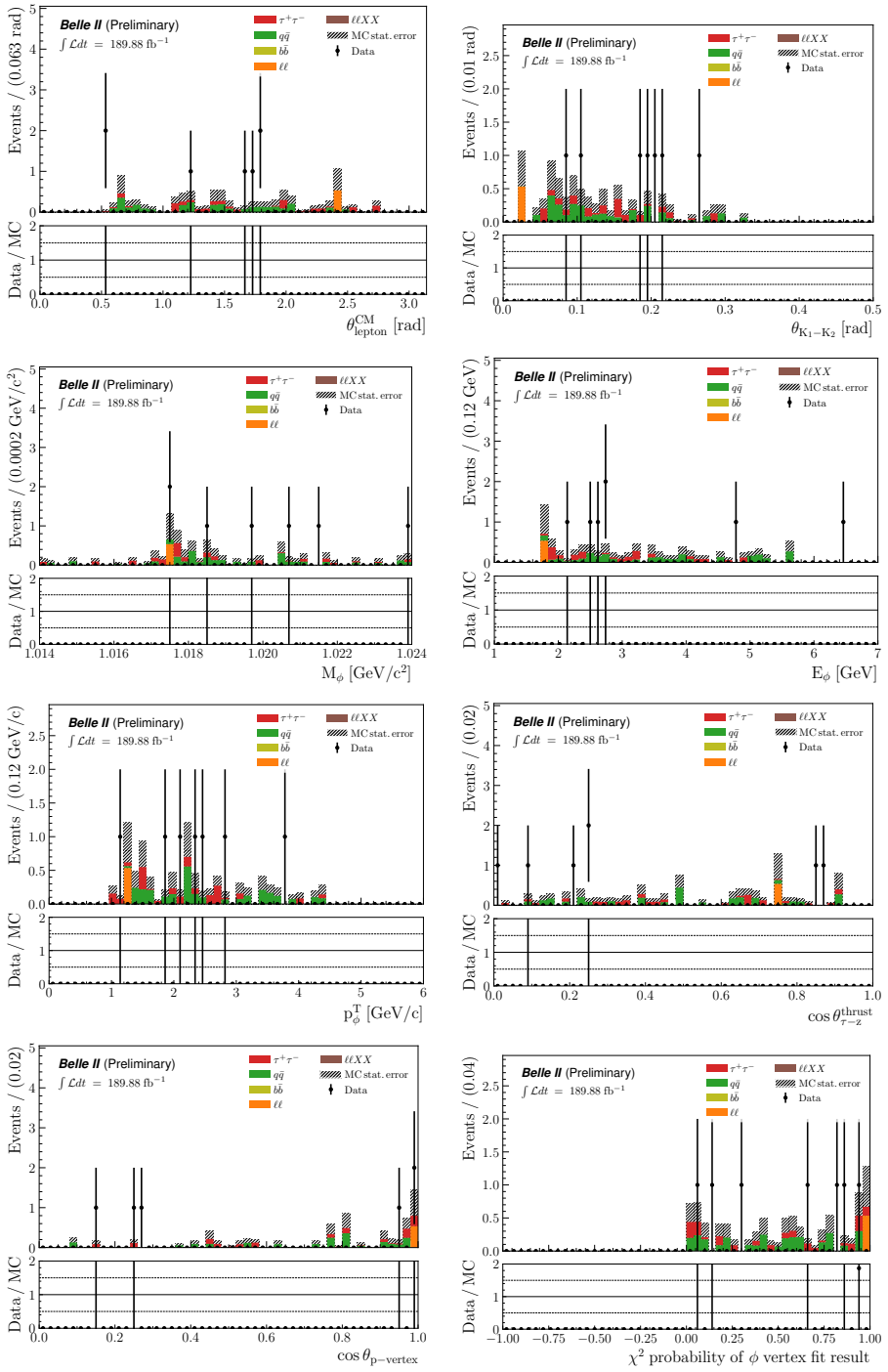


Figure E.7: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

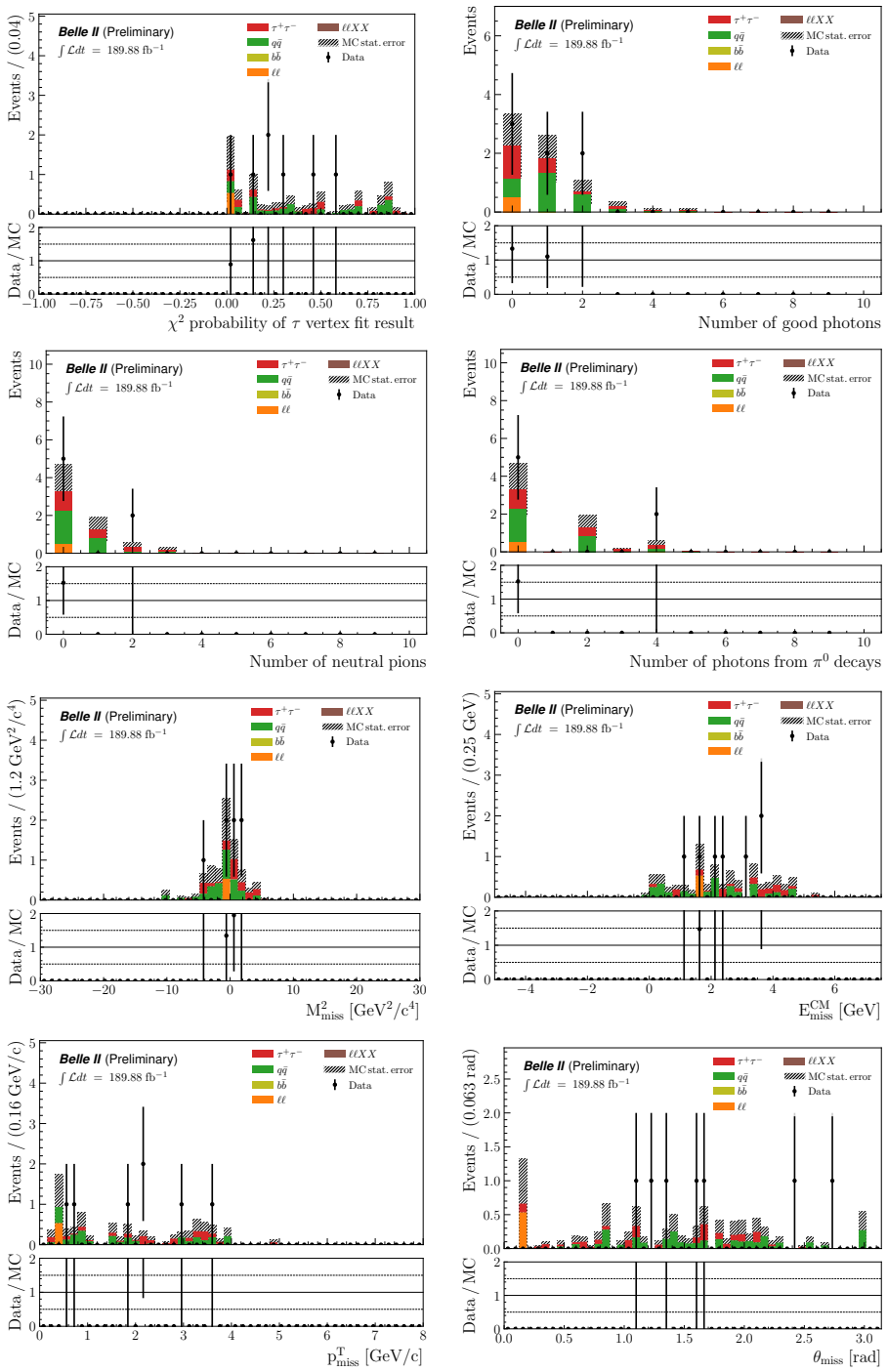


Figure E.7: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

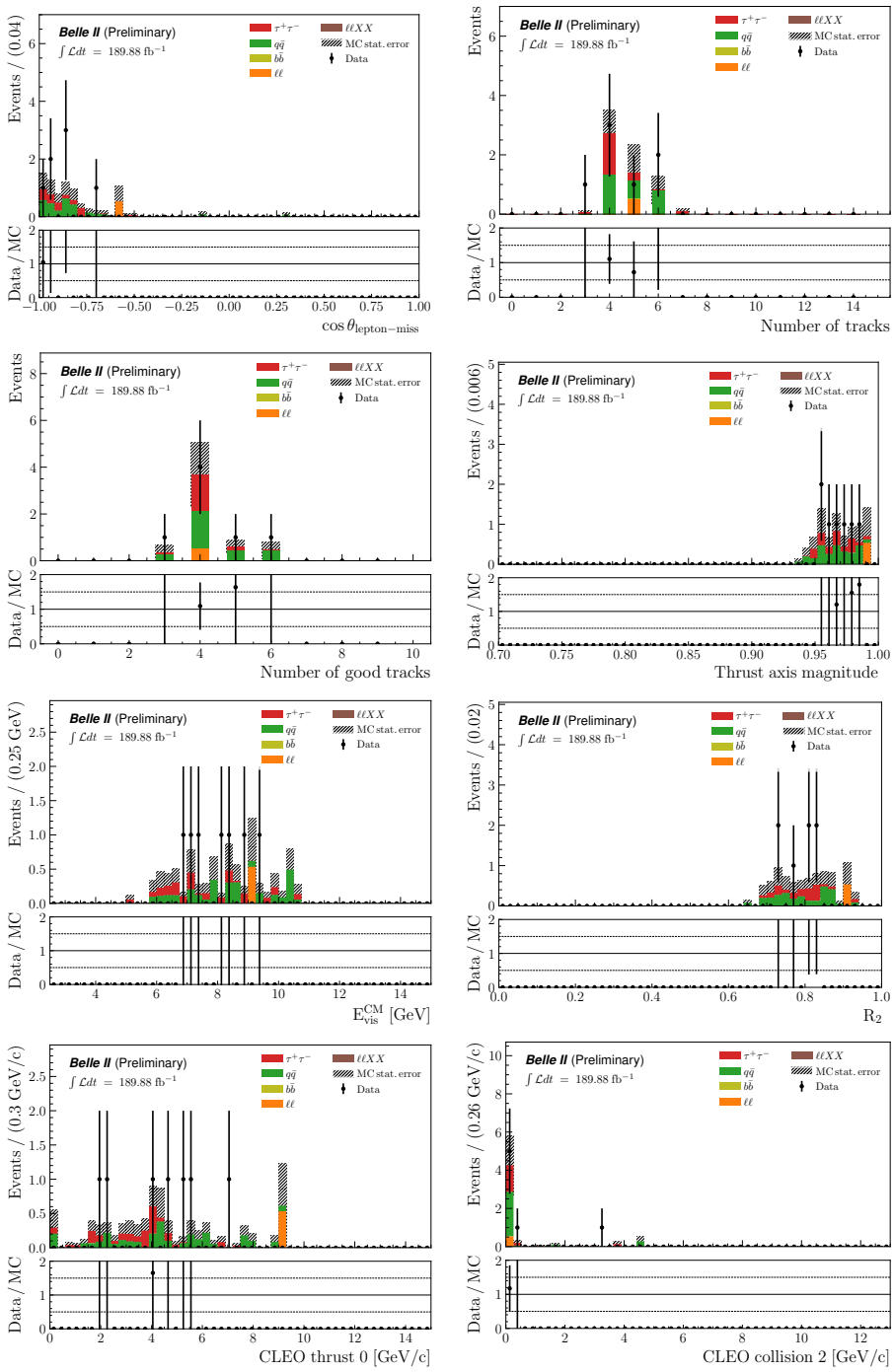


Figure E.7: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

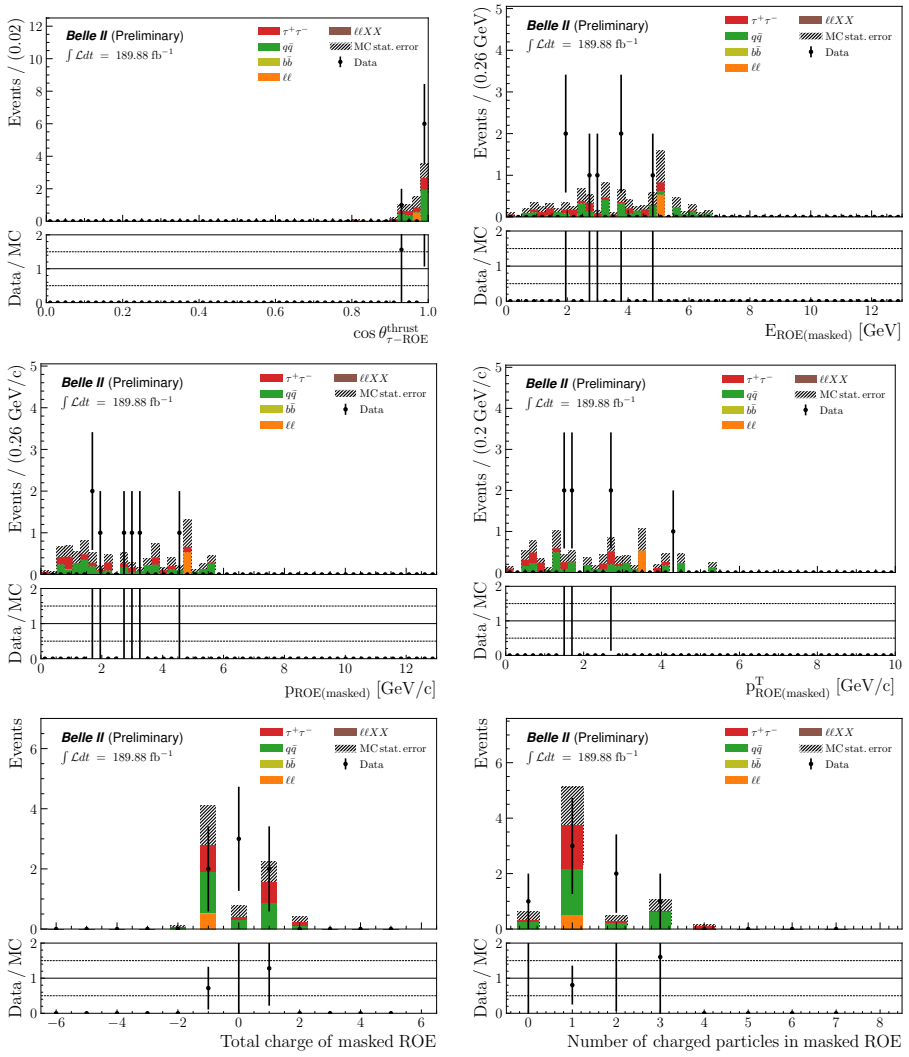


Figure E.7: Data-MC comparison in the $5-20\sigma$ sidebands of reconstructed $\tau \to \mu\phi$ events for M_τ, ΔE_τ and variables taken as inputs to the BDT, after the full selection.

F. Trigger studies

By "relative trigger efficiency", we mean the efficiency of the trigger selection of interest, for example in this case the low-multiplicity trigger bits *lmlX* combined as explained in Section 4.3.6, measured with respect to a reference trigger selection that involves trigger bits independent from (or orthogonal to) the *lmlX* lines. Namely, since the low-multiplicity bits are all ECL-based, we use as reference trigger bits based on CDC-only information (*fff, ffs, ffo, ff30*, defined in Table A.2). The trigger efficiency $\epsilon_{\text{lml/cdc}}$ is then measured by taking the ratio of events passing both the trigger selection of interest and the reference trigger selection, to the total number of events that passed the trigger reference selection:

$$\epsilon_{\text{lml/cdc}} = \frac{(fff \mid ffs \mid ffo \mid ff30) \ \& \ lmlX}{(fff \mid ffs \mid ffo \mid ff30)}. \tag{F.1}$$

This method is the only one applicable to the data, where we cannot rely on the truth information at generator level like in the simulation. Similarly, for simulated samples lacking the simulation of the trigger bits, the relative efficiency strategy is the only option to perform such a study.

The difference between the efficiencies measured in data and MC is assigned as systematic uncertainty on the signal efficiency due to the trigger selection.

Datasets

To be able to compare to the data without any risk of unblinding, the study is performed on the $\tau \to 3\pi\nu_\tau$ control sample. A set of 0.9279 fb^{-1} of data[1] is exploited, along with 1 fb^{-1} of simulation samples that include generic and low-multiplicity processes (portion of samples detailed in Section 4.2).

Selections and results

The same event reconstruction and offline selections are applied both on the data and simulation. They consist of the inclusive event reconstruction described in Section 4.3, with LID vetoes so that all three signal tracks are selected as pion candidates, suppressing the risk of accidentally unblinding the analysis. The background suppression criteria consist in the same cut-based selection as the one presented in Section 4.6.6.2, Table 4.20, except for the requirements on the ranked transverse momenta of the 3-prong tracks, replaced with $R_2 > 0.6$ and $0.3 < \theta_{\text{miss}} < 2.5$ rad. In the same way as what we did in the control channel study, we only take into account MC events truth-matched as signal, while the analysed data distributions are background-subtracted.

[1] 10% of experiment 12, *proc12*, good runs 5888-6427.

We show in Figure F.1 the distributions of the ranked transverse momenta in data and MC for the different types of trigger selections. The bin-by-bin ratio of the two measured efficiencies $\epsilon^{\text{Data}}_{\text{lml/cdc}}$ and $\epsilon^{\text{MC}}_{\text{lml/cdc}}$, as a function of the transverse momentum, is displayed in the inset bottom canvas of each sub-figure, while the red line corresponds to the result of the fit with a constant p_0 to the values between 0.9 and 1.1.

We observe a good agreement between the efficiencies in data and MC, with an average deviation of 0.9% to unity. This value is assigned as systematic uncertainty in the signal efficiency due to possible differences between the simulation and the data in the trigger behaviour.

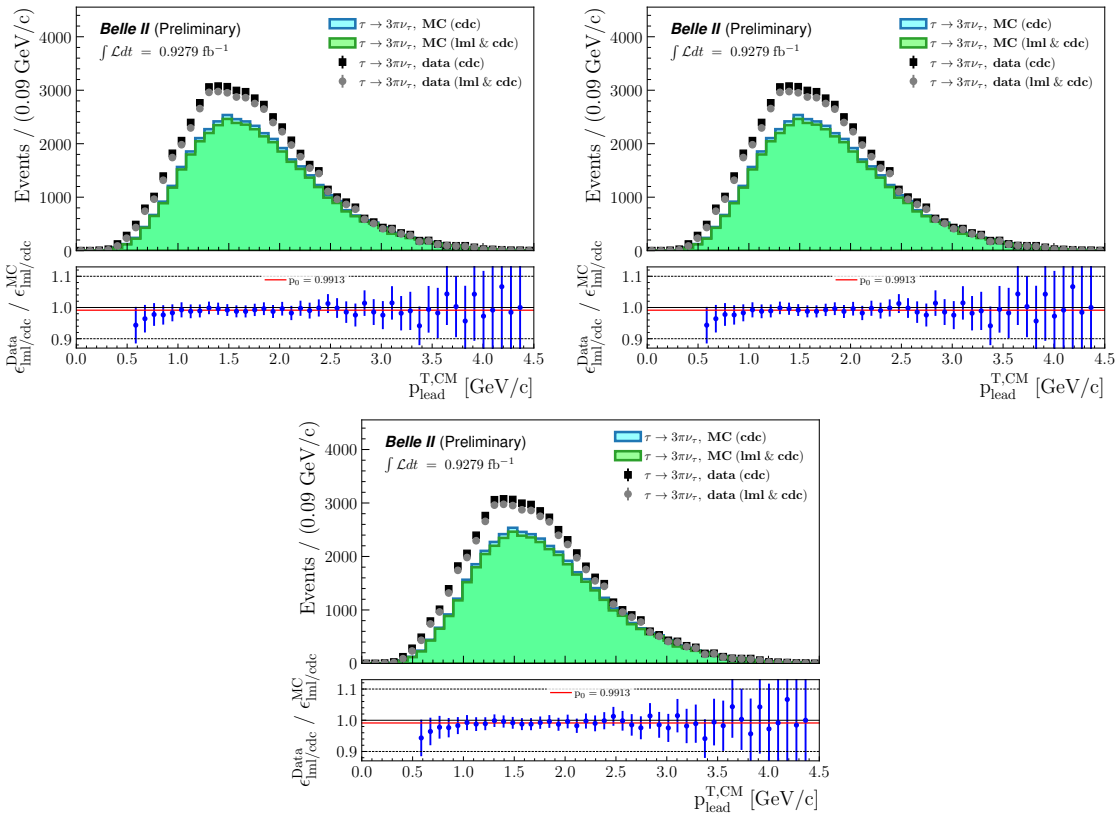


Figure F.1: Comparison of data and MC distributions for the leading (top left), sub-leading (top right) and third (bottom) 3-prong tracks transverse momenta in the $\tau \to 3\pi\nu_\tau$ decay channel, including different trigger requirements.

List of Figures

List of Tables

Bibliography

[1] C. Cornella, D. A. Faroughy, J. Fuentes-Martín, G. Isidori and M. Neubert, *Reading the footprints of the B-meson flavor anomalies*, Journal of High Energy Physics **2021** (2021) no. 8, 50, arXiv:2103.16558 [hep-ph].

[2] A. Angelescu, D. Bečirević, D. A. Faroughy, F. Jaffredo and O. Sumensari, *Single leptoquark solutions to the B-physics anomalies*, Physical Review D **104** (2021) no. 5, 055017, arXiv:2103.12504 [hep-ph].

[3] C. Hati, J. Kriewald, J. Orloff and A. M. Teixeira, *The fate of V_1 vector leptoquarks: the impact of future flavour data*, The European Physical Journal C **81** (2021) no. 12, 1066, arXiv:2012.05883 [hep-ph].

[4] LHCb Collaboration, R. Aaij et al., *Test of lepton universality with $B^0 \to K^{*0} l^+ l^-$ decays*, Journal of High Energy Physics **2017** (2017) no. 8, 55, arXiv:1705.05802 [hep-ex].

[5] LHCb Collaboration, R. Aaij et al., *Test of lepton universality in beauty-quark decays*, 2021. arXiv:2103.11769 [hep-ex].

[6] Heavy Flavor Averaging Group (HFLAV), Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, 2022. arXiv:2206.07501 [hep-ex].

[7] BaBar Collaboration, B. Aubert et al., *Improved Limits on Lepton Flavor Violating Tau Decays to $\ell\phi$, $\ell\rho$, ℓK^*, and $\ell\bar{K}^*$*, Phys. Rev. Lett. **103** (2009) 021801, arXiv:0904.0339 [hep-ex].

[8] Belle Collaboration, Y. Miyazaki et al., *Search for Lepton-Flavor-Violating tau Decays into a Lepton and a Vector Meson*, Phys. Lett. B **699** (2011) 251–257, arXiv:1101.0755 [hep-ex].

[9] S. Banerjee et al., *Snowmass 2021 White Paper: Charged lepton flavor violation in the tau sector*, 2022. arXiv:2203.14919 [hep-ph].

[10] Particle Data Group, P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020 and 2021 update) no. 8, 083C01.

[11] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) no. 2, 1554, arXiv:1007.1727v3 [physics.data-an].

[12] *Documentation of pyhf*, https://pyhf.readthedocs.io/en/v0.6.3/index.html.

[13] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[14] L. H. Ryder, *Quantum field theory*. Cambridge University Press, 1986.

[15] J. Shifflett, *Standard Model Lagrangian (including neutrino mass terms)*, https://www.einstein-schrodinger.com/Standard_Model.pdf. From *An Introduction to the Standard Model of Particle Physics*, 2nd Edition, W. N. Cottingham and D. A. Greenwood, Cambridge University Press, Cambridge, 2007.

[16] H. E. Logan, *TASI 2013 lectures on Higgs physics within and beyond the Standard Model*, 2017. arXiv:1406.1786 [hep-ph].

[17] J. Iliopoulos, *Introduction to the Standard Model of the Electro-Weak Interactions*, 2013. arXiv:1305.6779 [hep-ph].

[18] P. Aurenche, J. P. Guillet and E. Pilon, *Lagrangien QED, Règles de Feynman*, https://lectures.lapth.cnrs.fr/qed_qcd/cours/chapitre2.pdf.

[19] *Standard Model — Wikipedia, The Free Encyclopedia*, 2022. https://en.wikipedia.org/wiki/Standard_Model.

[20] ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) no. 1, 1–29, arXiv:1207.7214 [hep-ex].

[21] A. Bettini, *Introduction to Elementary Particle Physics*. Cambridge University Press, 2008.

[22] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Physical Review Letters **81** (1998) no. 8, 1562–1567, arXiv:hep-ex/9807003 [hep-ex].

[23] M. Blanke, A. J. Buras, B. Duling, A. Poschenrieder and C. Tarantino, *Charged lepton flavour violation and* $(g-2)_\mu$ *in the Littlest Higgs model with T-Parity: a clear distinction from Supersymmetry*, Journal of High Energy Physics **2007** (2007) no. 05, 013–013, arXiv:hep-ph/0702136v3 [hep-ph].

[24] A. Falkowski, *Lectures on Effective Field Theories*, https://indico.in2p3.fr/event/22195/contributions/86017/attachments/59873/81148/eftlectures.pdf.

[25] A. Celis, V. Cirigliano and E. Passemar, *Model-discriminating power of lepton flavor violating τ decays*, Physical Review D **89** (2014) no. 9, 095014, arXiv:1403.5781 [hep-ph].

[26] I. Pacheco and P. Roig, *Lepton flavour violation in hadron decays of the tau lepton within the littlest Higgs model with T-parity*, Journal of High Energy Physics **2022** (2022) no. 9, 144, arXiv:2207.04085 [hep-ph].

[27] T. Fukuyama, A. Ilakovac, T. Kikuchi and S. Meljanac, *Lepton Flavour Violation in the Minimal $SO(10)$ GUT Model and in the Standard Model with additional Heavy Dirac Neutrinos*, Nuclear Physics B - Proceedings Supplements **144** (2005) 143–148, arXiv:hep-ph/0411282 [hep-ph].

[28] Z.-H. Li, Y. Li and H.-X. Xu, *Unparticle-Induced Lepton Flavor Violating Decays $\tau \to \ell(V^0, P^0)$*, Physics Letters B **677** (2009) no. 3-4, 150–156, arXiv:0901.3266 [hep-ph].

[29] T. Fukuyama, A. Ilakovac and T. Kikuchi, *Lepton flavor violating leptonic and semileptonic decays of charged leptons in the minimal supersymmetric standard model*, The European Physical Journal C **56** (2008) no. 1, 125–146, arXiv:hep-ph/0506295 [hep-ph].

[30] I. Doršner, S. Fajfer, A. Greljo, J. Kamenik and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, Physics Reports **641** (2016) 1–68, arXiv:1603.04993 [hep-ph].

[31] D. Bečirević, I. Doršner, S. Fajfer, D. A. Faroughy, N. Košnik and O. Sumensari, *Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies*, Physical Review D **98** (2018) no. 5, 055003, arXiv:1806.05689 [hep-ph].

[32] A. Arhrib, R. Benbrik and C.-H. Chen, *Lepton flavor violating tau decays in type-III seesaw mechanism*, Physical Review D **81** (2010) no. 11, 113003, arXiv:0903.1553 [hep-ph].

[33] J. F. Kamenik and M. Nemevšek, *Lepton flavor violation in type I + III seesaw*, Journal of High Energy Physics **2009** (2009) no. 11, 023–023, arXiv:0908.3451 [hep-ph].

[34] E. Kou et al., *The Belle II Physics Book*, Progress of Theoretical and Experimental Physics **2019** (2019) no. 12, 123C01, arXiv:1808.10567 [hep-ex].

[35] A. J. Bevan, B. Golob, T. Mannel, S. Prell and B. D. Yabsley, *The Physics of the B Factories*, The European Physical Journal C **74** (2014) no. 11, 3026, arXiv:1406.6311 [hep-ex].

[36] T. Abe et al., *Belle II Technical Design Report*, arXiv, 2010. arXiv:1011.0352 [physics.ins-det].

[37] P. Raimondi, *Status on SuperB effort*, 2nd LNF Workshop on SuperB, Frascati, Italy, 2006. http://www.lnf.infn.it/conference/superb06/talks/raimondi1.ppt.

[38] SuperB Collaboration, *SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report*, arXiv, 2007. arXiv:0709.0451 [hep-ex].

[39] SVD WebHome. https://confluence.desy.de/display/BI/SVD+Basics.

[40] G. Casarosa, G. Dujany, C. Finck, R. Leboucher, L. Martel, L. Polat and L. Zani, *Measurement of the SVD Cluster Position Resolution*, 2022. BELLE2-NOTE-TE-2022-005.

[41] Belle II SVD Collaboration, R. Leboucher et al., *Measurement of the cluster position resolution of the Belle II Silicon Vertex Detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1033** (2022) 166746, arXiv:2209.03006 [physics.ins-det].

[42] N. Braun, *Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment*. PhD thesis, Karlsruhe, 2018.

[43] J.-F. Krohn et al., *Global decay chain vertex fitting at Belle II*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **976** (2020) 164269, arXiv:1901.11198 [hep-ex].

[44] The Belle II Tracking Group, *Tracking Performance in Early Belle II Data*, 2022. BELLE2-NOTE-TE-2022-006.

[45] T. B. I. P. P. Group, *Performance highlights from Belle II at ICHEP 2022*, 2022. BELLE2-NOTE-PH-2022-035.

[46] C. Hainje, A. Albert, C. Hadjivasiliou and J. Strube, *A comprehensive study of Belle II particle identification performance, efficiencies, and detector effects*, 2022. BELLE2-NOTE-TE-2021-022.

[47] Charged PID Group, The Belle II Collaboration, *Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022*, 2022. BELLE2-CONF-PH-2022-003.

[48] S. Sandilya and A. Schwartz, *Study of Kaon and Pion Identification Performances in Phase III data using the decay $D^{*+} \to D^0[K^-\pi^+]\pi^+$*, 2021. BELLE2-NOTE-PH-2019-048.

[49] *Documentation of basf2*, https://software.belle2.org.

[50] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, *The Belle II Core Software*, Computing and Software for Big Science **3** (2018) no. 1, 1, arXiv:1809.04299 [physics.comp-ph].

[51] J. Bennett, *Calibration data flow summary*, BELLE2-NOTE-PH-2016-003, 2016.

[52] GEANT4, S. Agostinelli et al., *GEANT4–a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250–303.

[53] Cobham Technical Services, Vector Fields Software (Oxford, England), *Opera-3D User Guide Ver. 15R3*, 2012.

[54] *Trigger Simulation (TSIM)*, https://confluence.desy.de/pages/viewpage.action?pageId=47822559.

[55] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152–155.

[56] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[57] S. Jadach, B. Ward and Z. Was, *The precision Monte Carlo event generator $\mathcal{KK}$ for two-fermion final states in e^+e^- collisions*, Computer Physics Communications **130** (2000) no. 3, 260–325, arXiv:hep-ph/9912214 [hep-ph].

[58] N. Davidson, G. Nanava, T. Przedziński, E. Richter-Was and Z. Was, *Universal interface of TAUOLA: Technical and physics documentation*, Computer Physics Communications **183** (2012) no. 3, 821–843, arXiv:1002.0543 [hep-ph].

[59] C. C. Calame, C. Lunardini, G. Montagna, O. Nicrosini and F. Piccinini, *Large-angle Bhabha scattering and luminosity at flavour factories*, Nuclear Physics B **584** (2000) no. 1-2, 459–479, arXiv:hep-ph/0003268 [hep-ph].

[60] G. Balossini, C. Bignamini, C. C. Calame, G. Montagna, O. Nicrosini and F. Piccinini, *Photon pair production at flavour factories with per mille accuracy*, Physics Letters B **663** (2008) no. 3, 209–213, arXiv:0801.3360 [hep-ph].

[61] R. K. F.A. Berends, P.H. Daverveldt, *Complete lowest-order calculations for four-lepton final states in electron-positron collisions*, Nuclear Physics B **253** (1985) 441–463.

[62] S. Banerjee, B. Pietrzyk, J. M. Roney and Z. Was, *Tau and muon pair production cross sections in electron-positron annihilations at* $\sqrt{s} = 10.58$ *GeV*, Physical Review D **77** (2008) no. 5, 054012, arXiv:0706.3235 [hep-ph].

[63] S. Banerjee, D. Biswas, T. Przedzinski and Z. Was, *Monte Carlo Event Generator updates, for tau pair events at Belle II energies*, 2021. arXiv:2111.05914 [hep-ph].

[64] D. M. Asner et al., *Search for exclusive charmless hadronic B decays*, Physical Review D **53** (1996) no. 3, 1039–1050, arXiv:hep-ex/9508004v1 [hep-ex].

[65] *Tree Fitter*, https://software.belle2.org/development/sphinx/analysis/doc/TreeFitter.html.

[66] G. Punzi, *Sensitivity of searches for new signals and its optimization*, arXiv:physics/0308063 [physics.data-an].

[67] D. M. Asner et al., *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (1996) 1039–1050, arXiv:hep-ex/9508004v1 [hep-ex].

[68] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in* e^+e^- *Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581–1585.

[69] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. ACM, New York, NY, USA, 2016. arXiv:1603.02754 [cs.LG].

[70] *Documentation of XGBoost*, https://xgboost.readthedocs.io/en/stable/index.html.

[71] F. Abudinén et al., *Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment*, Chinese Physics C **44** (2020) no. 2, 021001, arXiv:1910.05365 [hep-ex].

[72] Glazov, A. and others, *Measurement of the tracking efficiency in Phase 3 data using tau-pair events.*, 2020. BELLE2-NOTE-PH-2020-006.

[73] Qi-Dong Zhou, *Correction for tracking momentum bias based on invariant mass peak studies*, 2020. BELLE2-NOTE-PH-2020-030.

[74] T. B. I. P. P. Group, *Conference Readiness - ICHEP 2022*, https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Conference+readiness.

[75] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A **555** (2005) 356–369, arXiv:physics/0402083.

[76] U. Yuji, *Data Format from ETM to GDL*, https://confluence.desy.de/display/BI/Data+Format+from+ETM+to+GDL.

[77] P. Rados, A. Rostomyan and A. Martini, *Measurement of the trigger efficiency in Phase 3 data using tau-pair events*, 2020. BELLE2-NOTE-PH-2020-028.

[78] T. Koga, H. Nakazawa and N. Ping, *Trigger Summary in experiment 16,17,18*, 2021. BELLE2-NOTE-TE-2021-018.

[79] T. Koga, H. Nakazawa and N. Ping, *Trigger Summary in experiment 20,21,22*, 2022. BELLE2-NOTE-TE-2022-001.