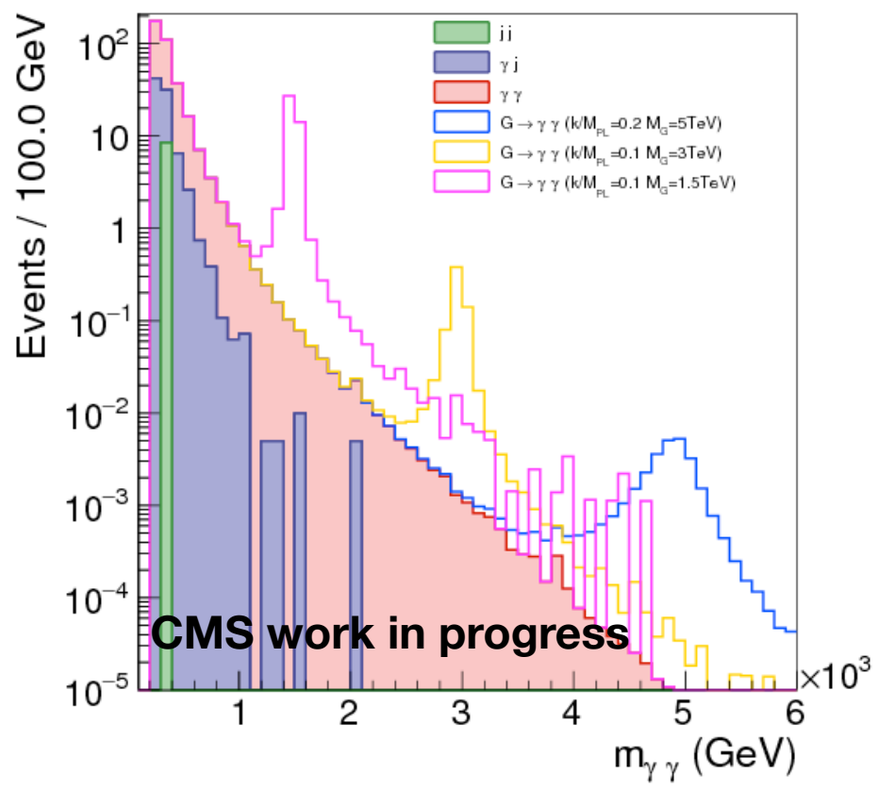
ETH Institute for Particle Physics

Search for extra dimensions in the high mass diphoton spectrum at 13 TeV at CMS



M. Quittnat

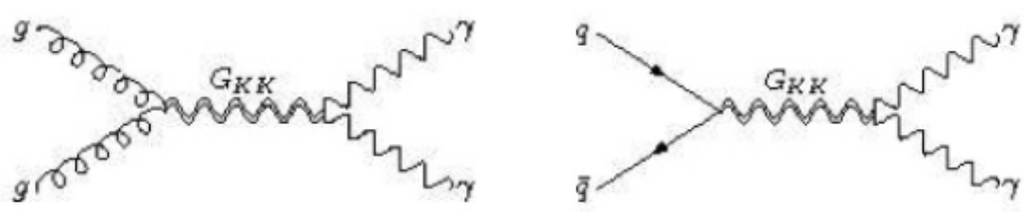
August 27th, 2015

motivation

extradimensional models motivated by hierarchy problem

http://phdcomics.com/comics/archive.php?comicid=1522

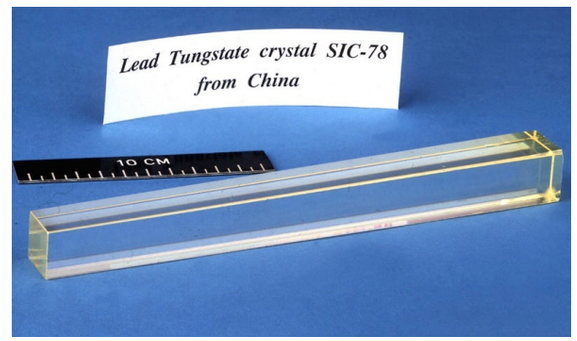
- gravity diluted by extra spatial dimensions
- two important models for searches at TeV scale:
  - **RS model (Randall-Sundrum):**
    - gravitons well separated Kaluza-Klein (KK) resonances
    - coupling κ/M_{Pl}
  - **ADD model (Arkani-Hamed, Dimopoulos, Dvali):**
    - KK resonances- discrete energy levels 1-100meV ~ 1/size ED
    - continuum spectrum in diphoton mass



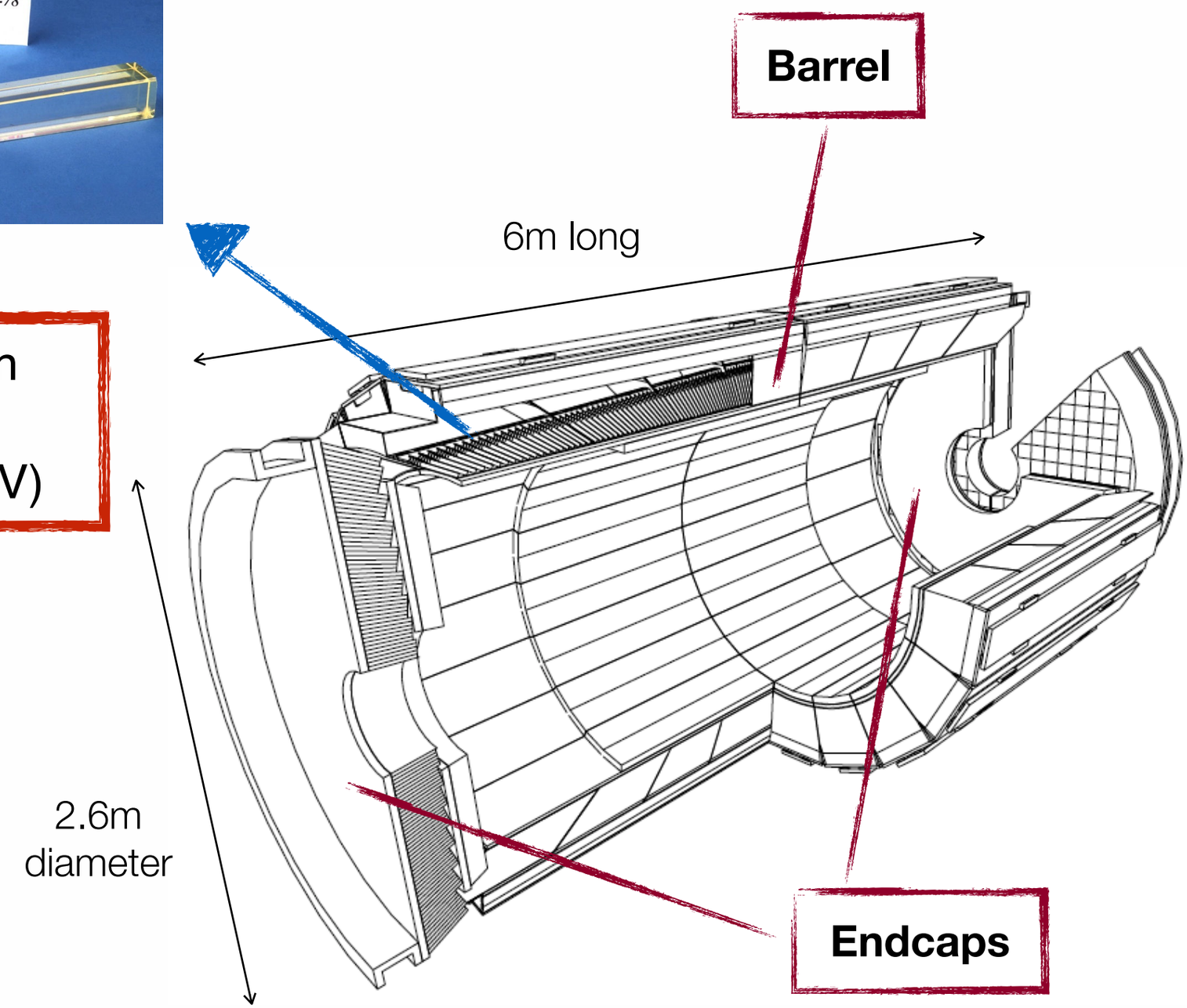
- excited graviton preferentially decays into 2 gauge bosons (Spin 2 particle)
- preferred over dilepton final state (branching ratio ≈ 2x higher)

electromagnetic calorimeter CMS

photon candidates from energy deposits "superclusters" in ECAL



- ECAL provides excellent diphoton energy resolution (~ 1% for masses above 100 GeV)



high energy photon ID

- **reconstruction from detector:**
  - triggering: double photon $p_{t\gamma} > 85$ GeV, single photon $p_{t\gamma} > 200$ GeV
  - energy corrections
  - identification of photon candidates

  - **event selection optimised for analysis sensitivity:**
    - photon pair with same primary vertex
    - $p_{t\gamma} > 100$ GeV
    - efficiency of 80-85 % with:

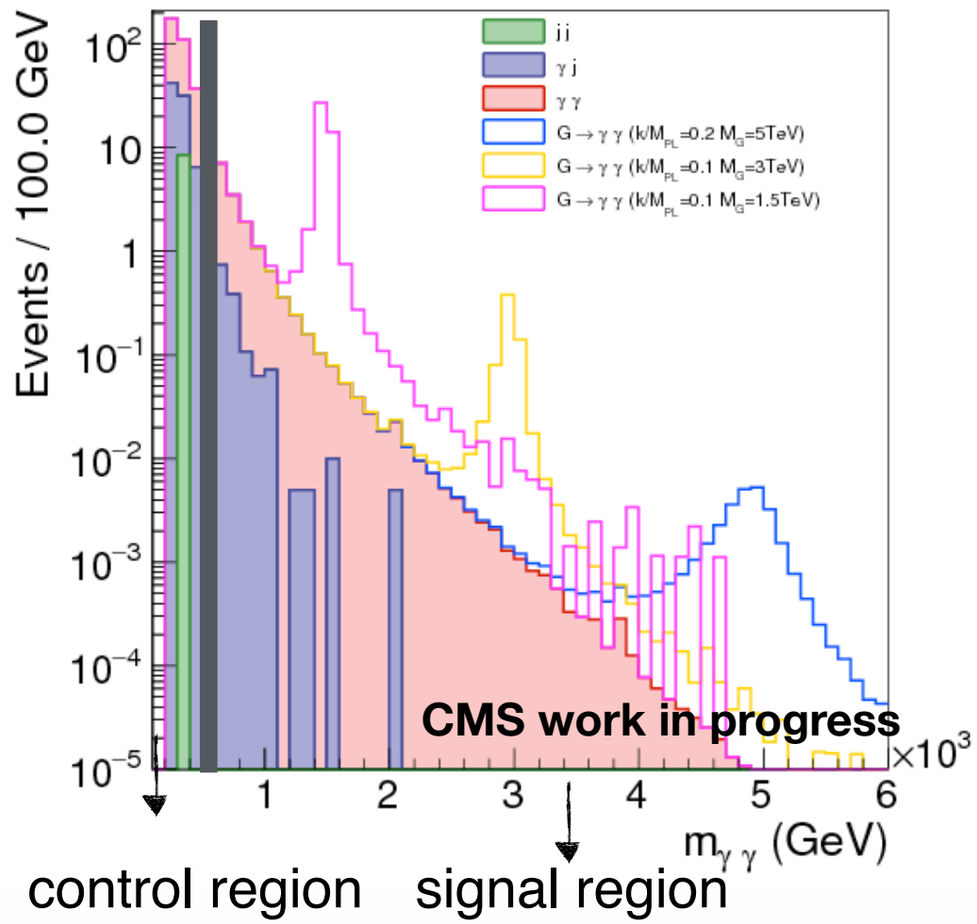
photon category	$\mathrm{Iso_{Ch}}$ cut (GeV)	$\bar{\mathrm{Iso}}_\gamma$ cut (GeV)	H/E cut	$\sigma_{i\eta i\eta}$ cut
$\eta_{SC} < 1.4442$	5	2.5	5×10^{-2}	0.0105
$\eta_{SC} > 1.566$	5	2.0	5×10^{-2}	0.028

analysis strategy

- **2 categories:**
  - both photons in barrel (EB)
  - at least one photon in endcap (EE)

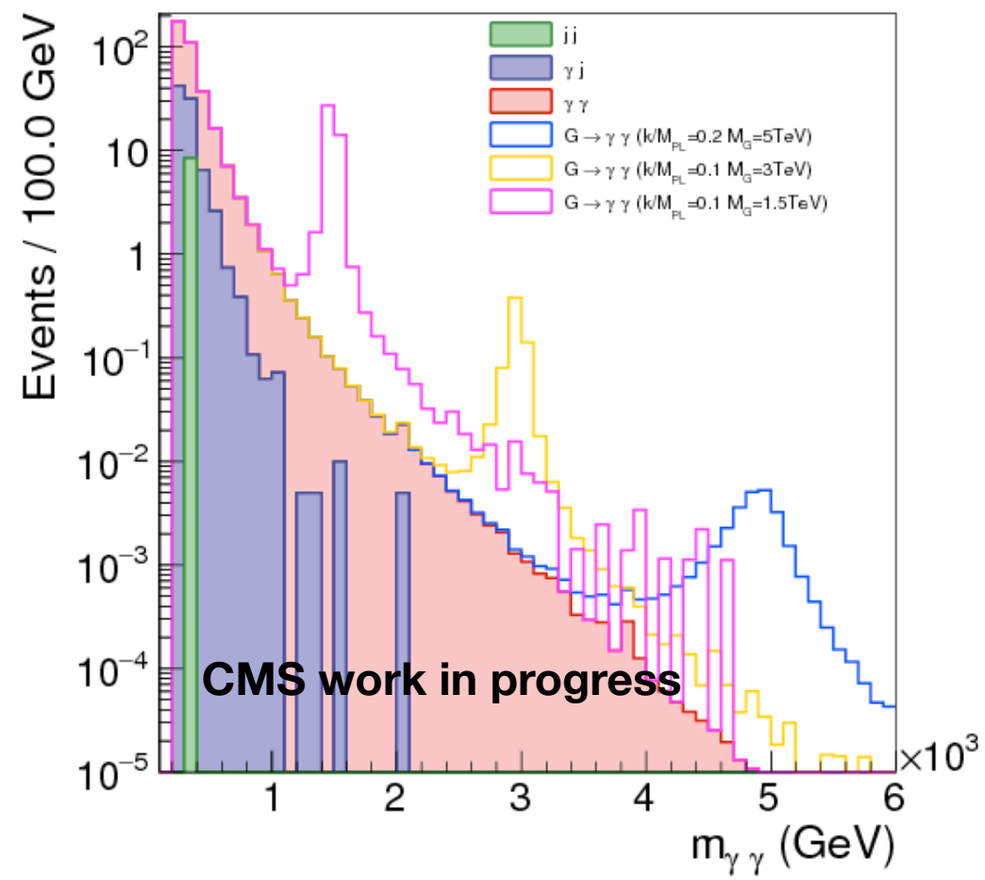
- control region: $m_{\gamma\gamma} > 300$ GeV
- search region: $m_{\gamma\gamma} > 500$ GeV

- background modelling
- signal modelling
- statistical interpretation



background description

- background:
  - irreducible: prompt photons (p) from SM
  - reducible: fakes (jets) passing photon ID
    → pf, ff
    - ca 30 %



use all available information to model the background

semi-parametric model

novum for exo searches!

- **why semi-parametric approach?**
  - parametric in $m_{\gamma\gamma}$ and non-parametric (histogram) in isolation of the two photons
  - additional data driven constraints to control and improve (reducible) bkg description

    → development of 3-dim fit to break down 3 bkg - components

$$SR(m_{\gamma\gamma}, ChIso_1, ChIso_2) = SR^{bkg}(m_{\gamma\gamma}, ChIso_1, ChIso_2) + SR^{sig}(m_{\gamma\gamma}, ChIso_1, ChIso_2)$$

↓

$$\begin{aligned} SR^{bkg}(m_{\gamma\gamma}, ChIso_1, ChIso_2) = & N^{pp} \cdot g^{pp}(m_{\gamma\gamma}) \cdot pdf^{pp}(ChIso_1, ChIso_2 | m_{\gamma\gamma}) \\ & + N^{pf} \cdot g^{pf}(m_{\gamma\gamma}) \cdot pdf^{pf}(ChIso_1, ChIso_2 | m_{\gamma\gamma}) \\ & + N^{ff} \cdot g^{ff}(m_{\gamma\gamma}) \cdot pdf^{ff}(ChIso_1, ChIso_2 | m_{\gamma\gamma}) \end{aligned}$$

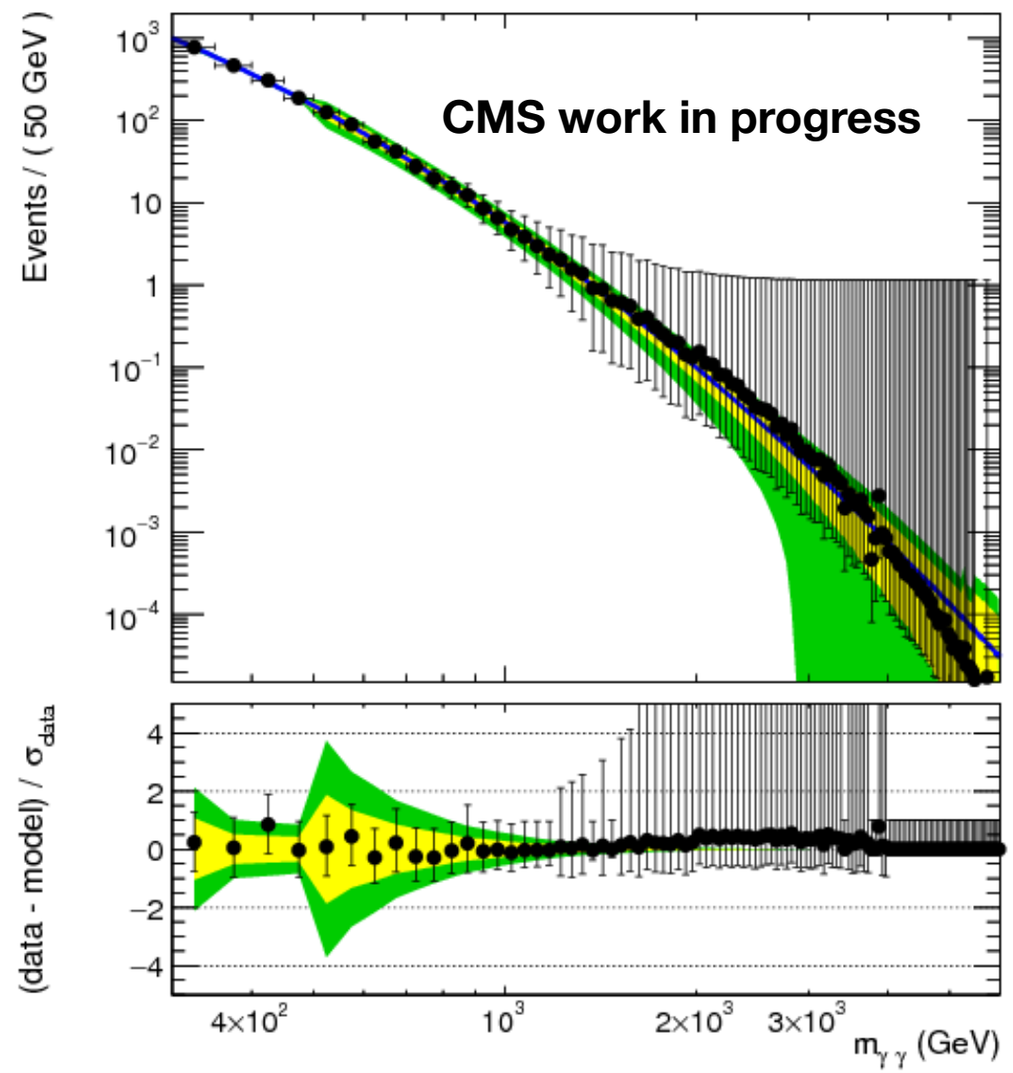
parametric functions → shape pdfs with ChPfIso → purity estimate

background model : parametric

$$SR^{bkg}(m_{\gamma\gamma}, ChIso_1, ChIso_2) = N^{pp} \cdot g^{pp}(m_{\gamma\gamma}) \cdot pdf^{pp}(ChIso_1, ChIso_2|m_{\gamma\gamma})$$
$$+ N^{pf} \cdot g^{pf}(m_{\gamma\gamma}) \cdot pdf^{pf}(ChIso_1, ChIso_2|m_{\gamma\gamma})$$
$$+ N^{ff} \cdot g^{ff}(m_{\gamma\gamma}) \cdot pdf^{ff}(ChIso_1, ChIso_2|m_{\gamma\gamma})$$

- for semi-parametric fit:
  - independent shapes for background components in $m_{\gamma\gamma}$
- baseline parametrisation:

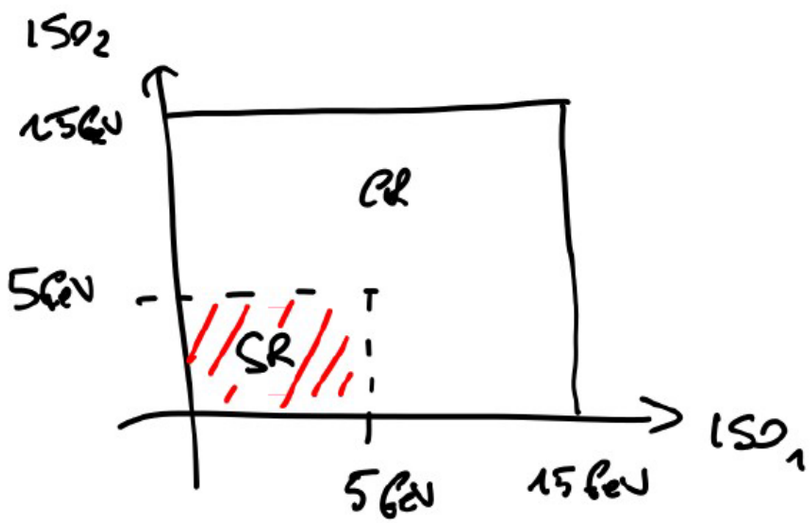
$$model_{bkg}(m_{\gamma\gamma}) = m_{\gamma\gamma}^{a+b\log(m_{\gamma\gamma})}$$



how to build the histograms?

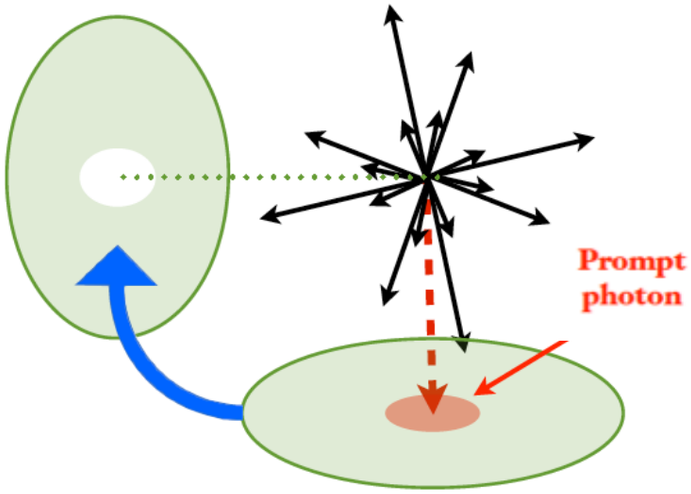
$$
\begin{aligned}
SR^{bkg}(m_{\gamma\gamma}, ChIso_1, ChIso_2) = {} & N^{pp} \cdot g^{pp}(m_{\gamma\gamma}) \cdot pdf^{pp}(ChIso_1, ChIso_2 | m_{\gamma\gamma}) \\
& + N^{pf} \cdot g^{pf}(m_{\gamma\gamma}) \cdot pdf^{pf}(ChIso_1, ChIso_2 | m_{\gamma\gamma}) \\
& + N^{ff} \cdot g^{ff}(m_{\gamma\gamma}) \cdot pdf^{ff}(ChIso_1, ChIso_2 | m_{\gamma\gamma})
\end{aligned}
$$

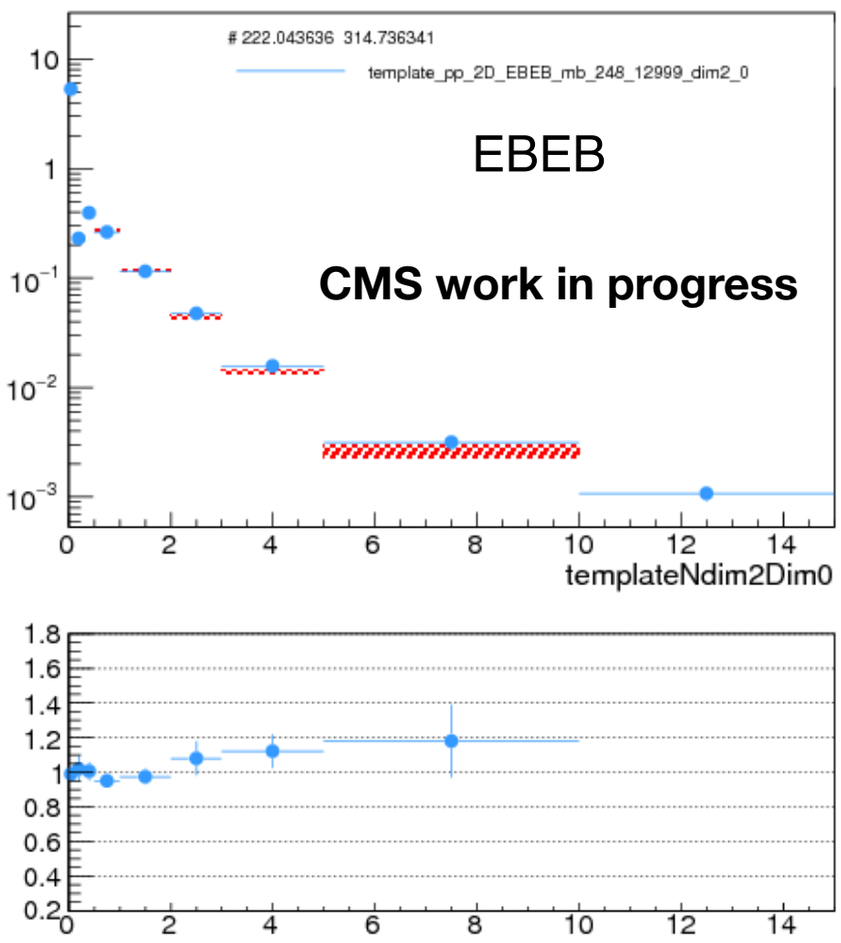
- **template variable**: charged particle flow isolation
  - same photon ID, but relax charged particle flow isolation (ChIso) cut from 5 to 15 GeV
- **data-driven!**



prompt photon: random cone isolation

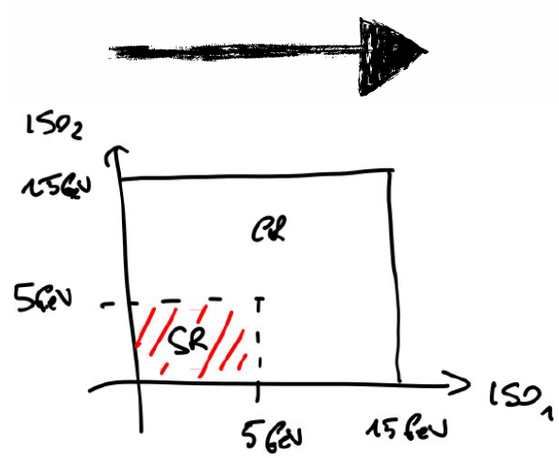
- assumption: prompt photon is isolated
- throw random cone from each photon candidate
  - keeping η, throw cone randomly in ϕ-direction
  - two random cones for each diphoton event

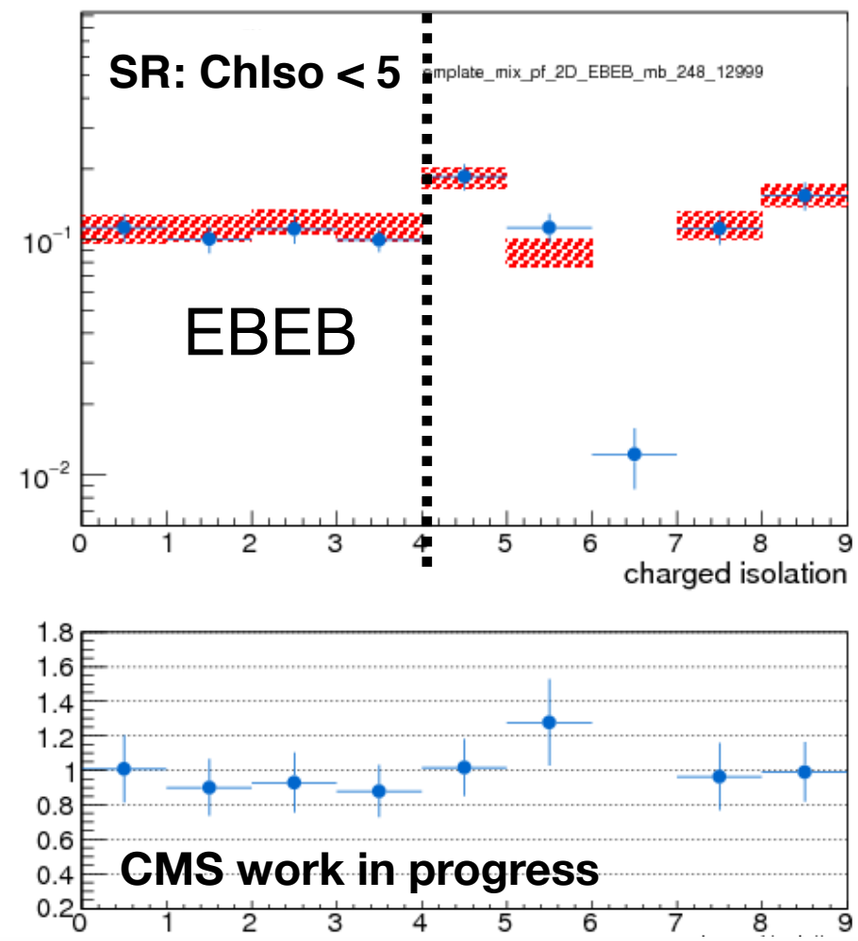




projection of leg1 ChIso

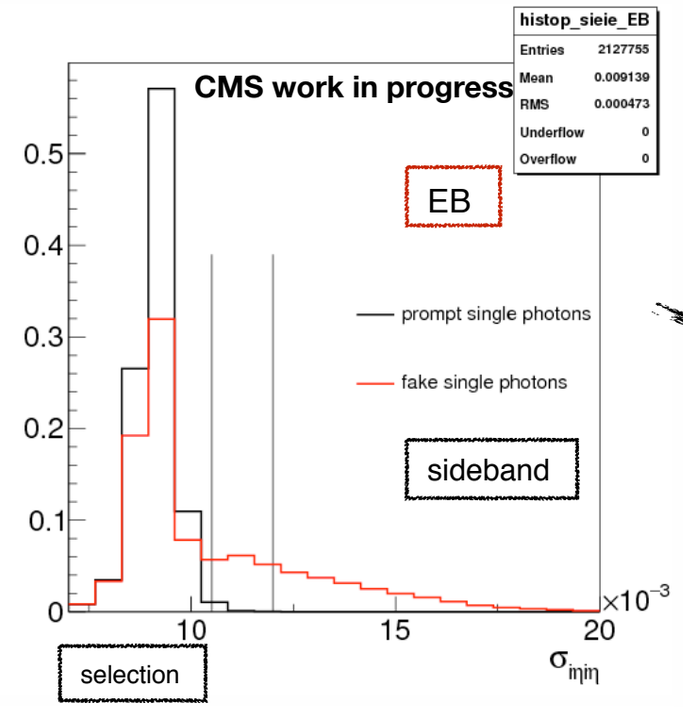
"unroll"

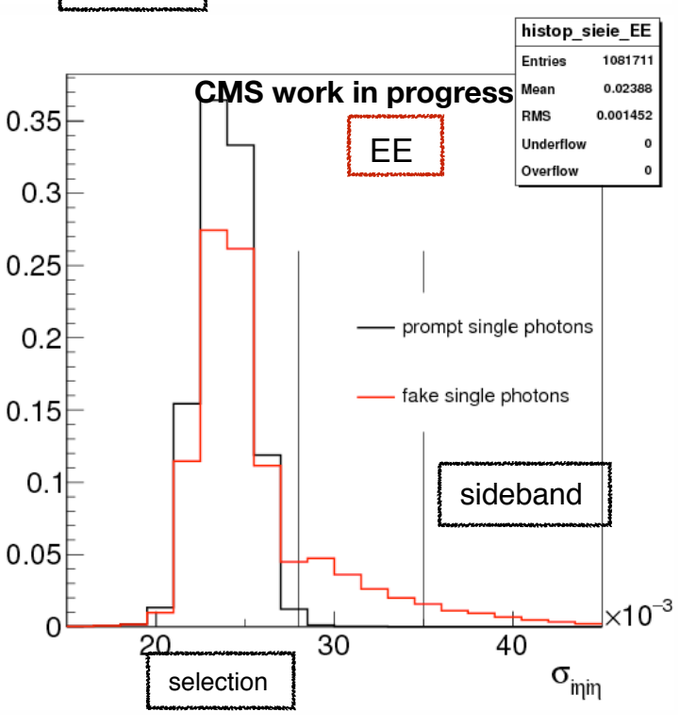


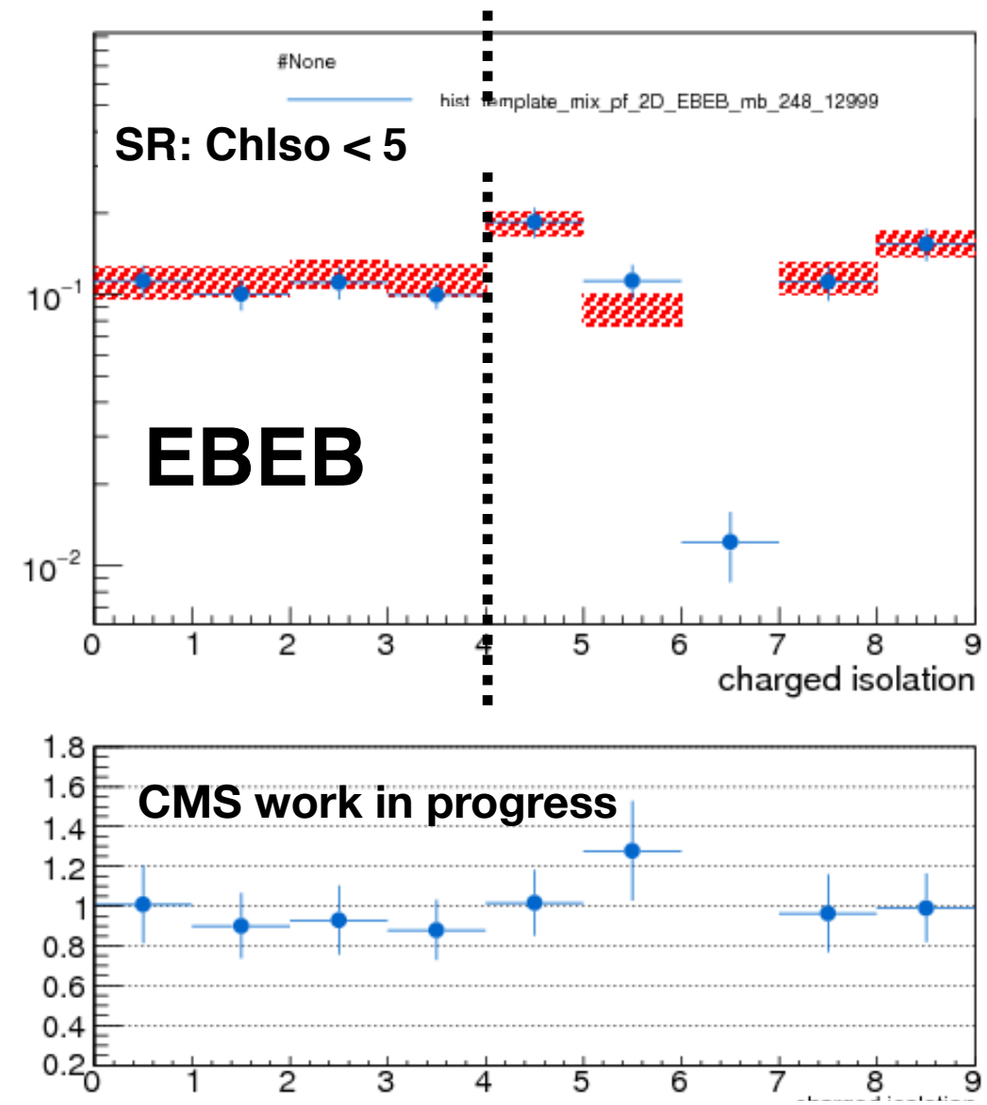


$\sigma_{i\eta i\eta}$-sideband definition

- $\sigma_{i\eta i\eta}$ and ChIso are uncorrelated
- increase statistics for reducible bkg templates with event mixing
  - ADVANTAGE: can build templates with enough statistics in high p_t region



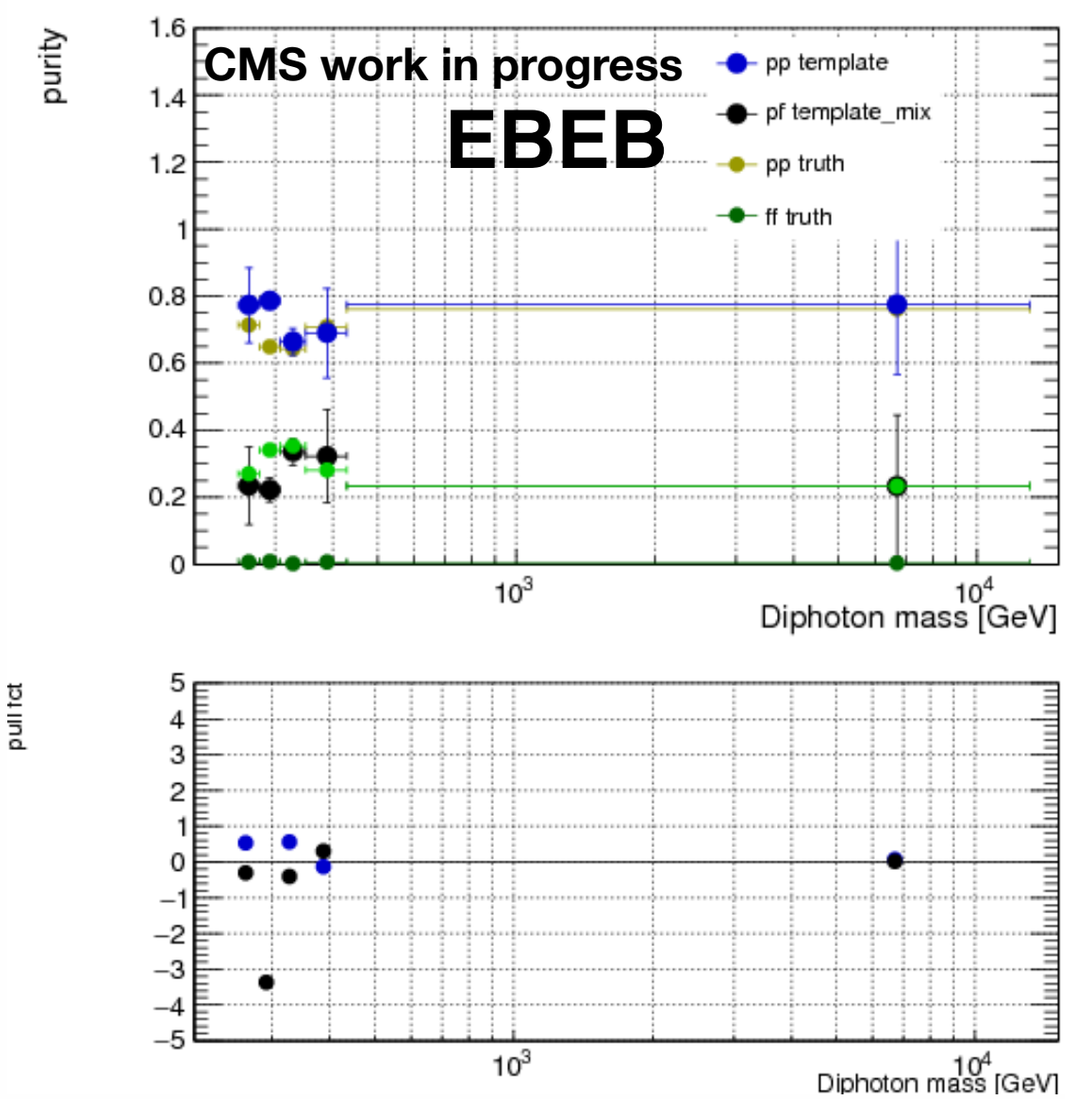


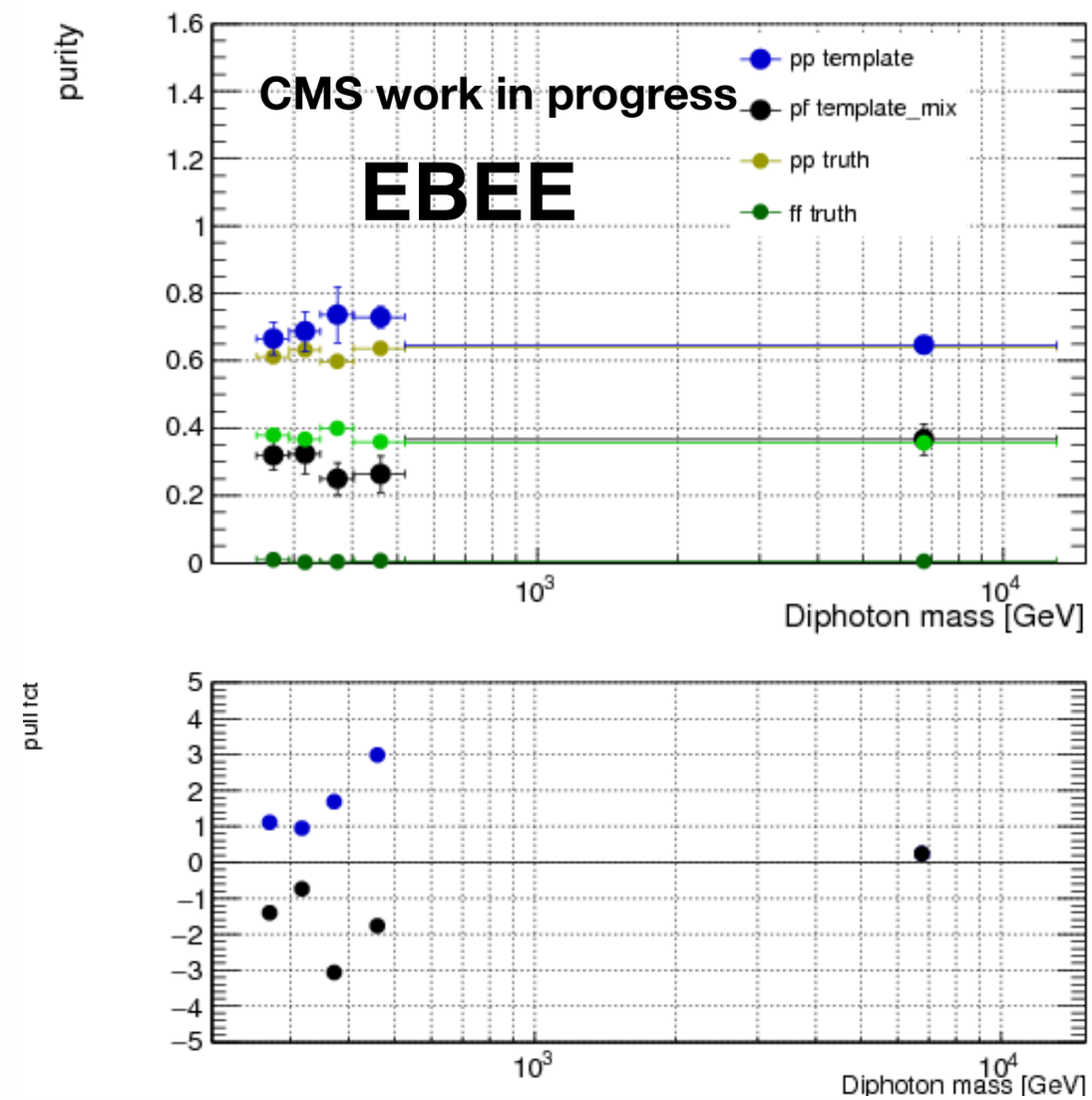


can we rely on these ChPfIso templates?

determine purity

- 10 % statistical error for 1fb^{-1}
- closure within 5-10 %

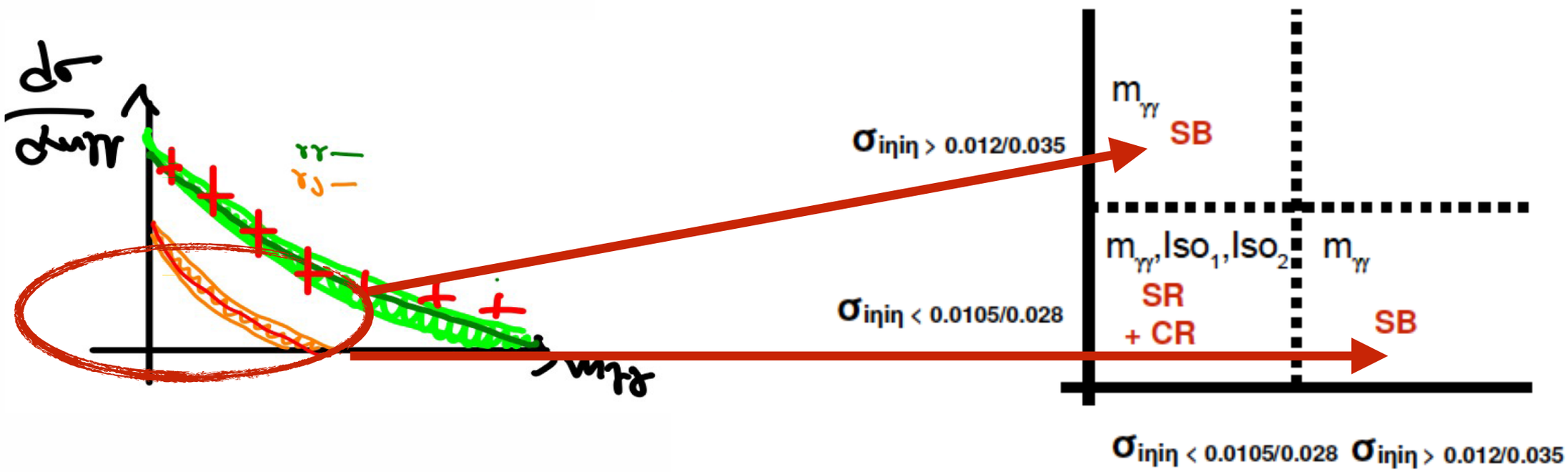




control region in mass spectrum

- back to full semi-parametric background description:
  - to further constrain $m_{\gamma\gamma}$ shape for fake background use $\sigma_{i\eta i\eta}$- sideband

$$SB^{bkg}(m_{\gamma\gamma}) = N^{pf} \cdot g^{pf}(m_{\gamma\gamma})$$

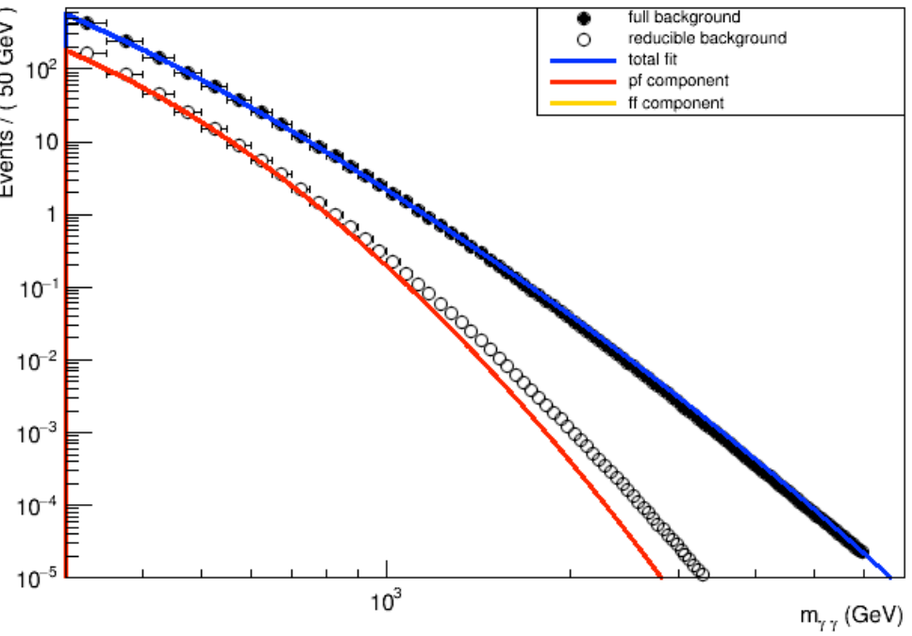


» Final fit: simultaneous fit in relaxed charged isolation region ("SR+CR") and sideband

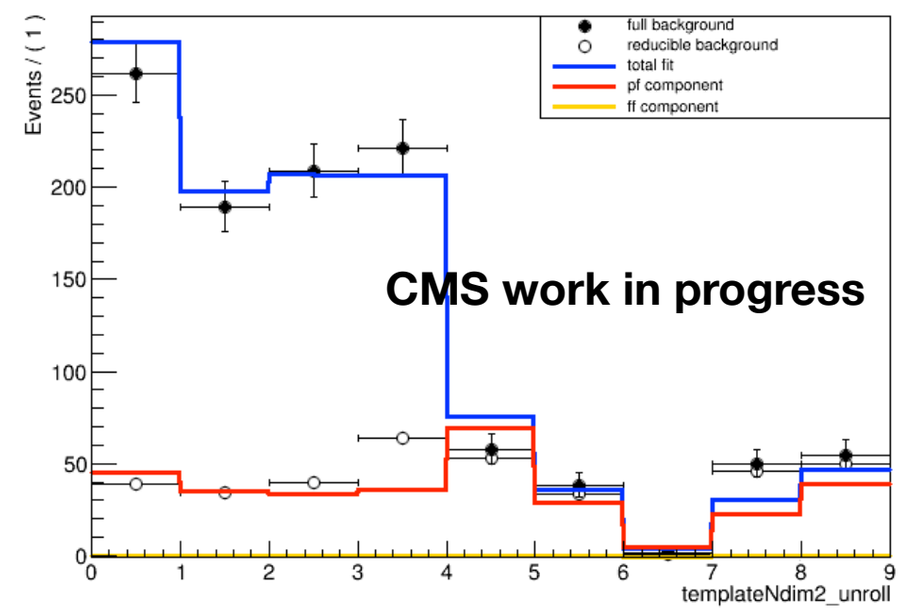
projection for 5 /fb

$$SR^{bkg}(m_{\gamma\gamma}, ChIso_1, ChIso_2) = N^{pp} \cdot g^{pp}(m_{\gamma\gamma}) \cdot pdf^{pp}(ChIso_1, ChIso_2|m_{\gamma\gamma}) + N^{pf} \cdot g^{pf}(m_{\gamma\gamma}) \cdot pdf^{pf}(ChIso_1, ChIso_2|m_{\gamma\gamma}) + N^{ff} \cdot g^{ff}(m_{\gamma\gamma}) \cdot pdf^{ff}(ChIso_1, ChIso_2|m_{\gamma\gamma})$$

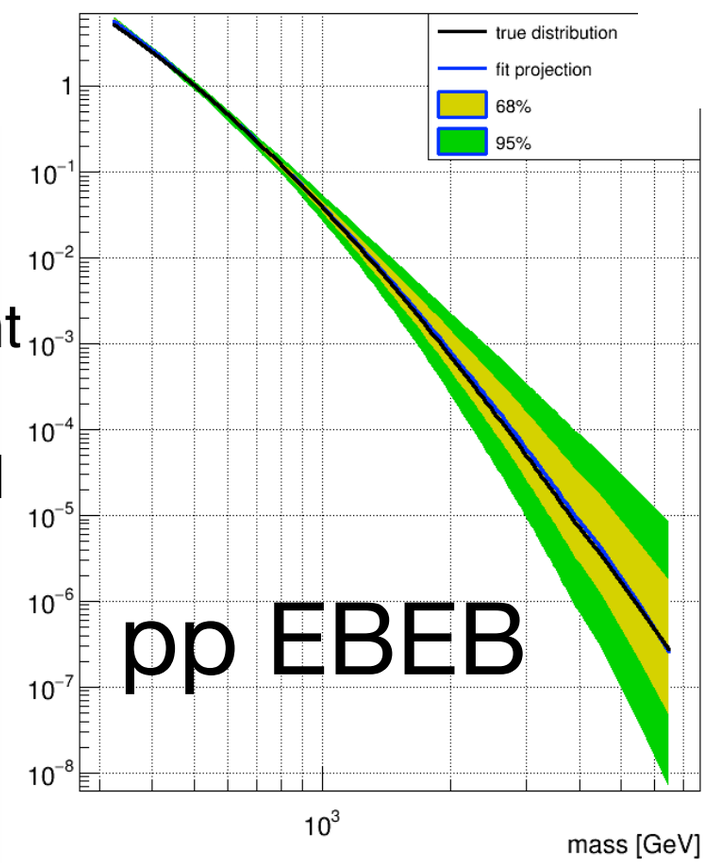
EBEB



templateEBEB



- fit converges
- bias on the non prompt component O(5%)
- pp-component important for signal sensitivity
  - good accuracy



summary

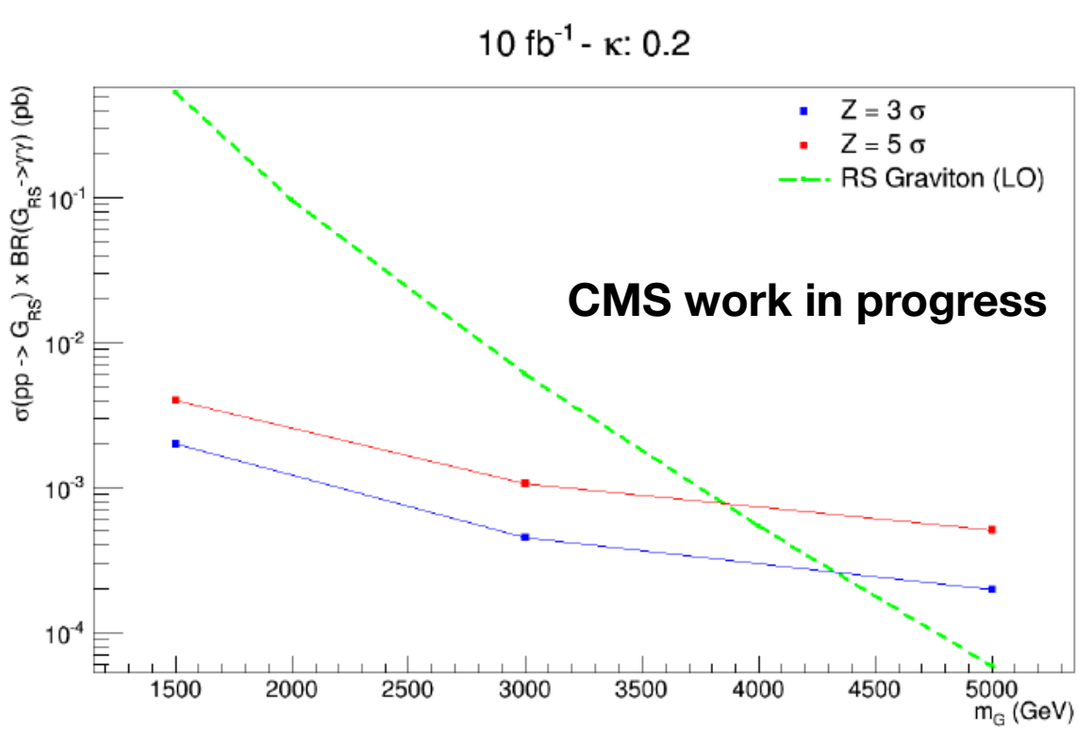
- Focus on background description of search for extra dimensions in high mass diphoton spectrum
- developed semi-parametric background model
  - fully data driven
  - break-down of background components with templates
    - for 2d-isolation templates - purity within 5-10%

- started to use semi-parametric model for hypothesis test in search
- fit converges and gives purity estimates with a bias of 5%

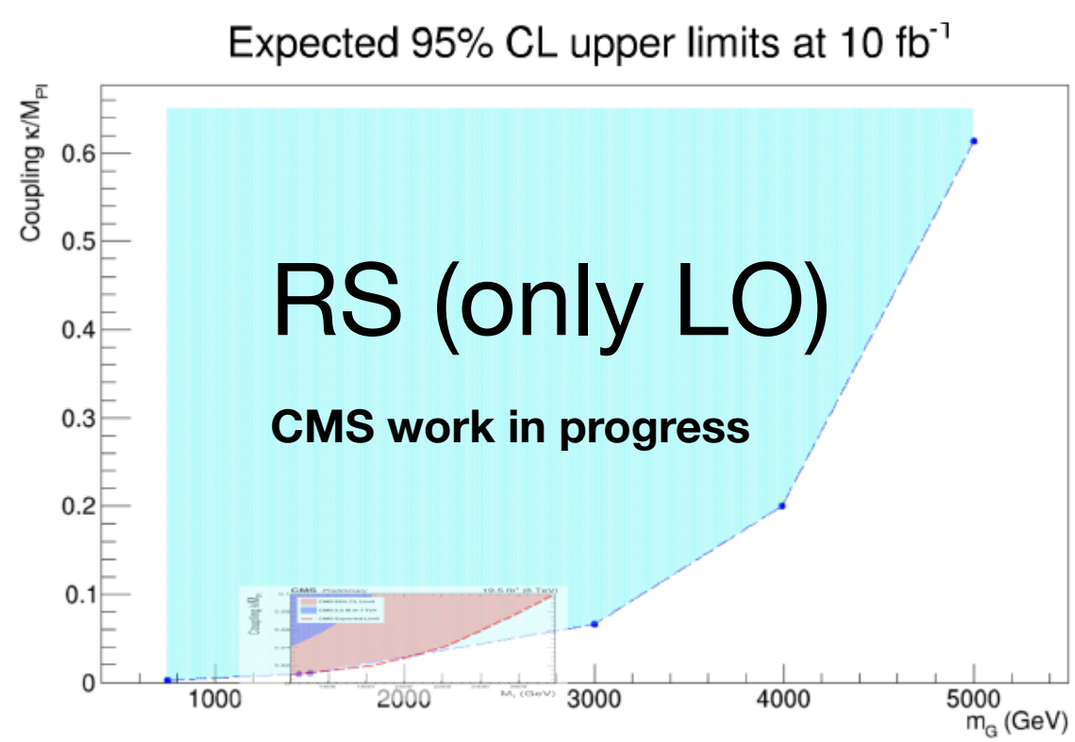
outlook

- these days data at 25 ns is arriving
  - first validations with 50 ns done

☞ take data and look for the graviton!

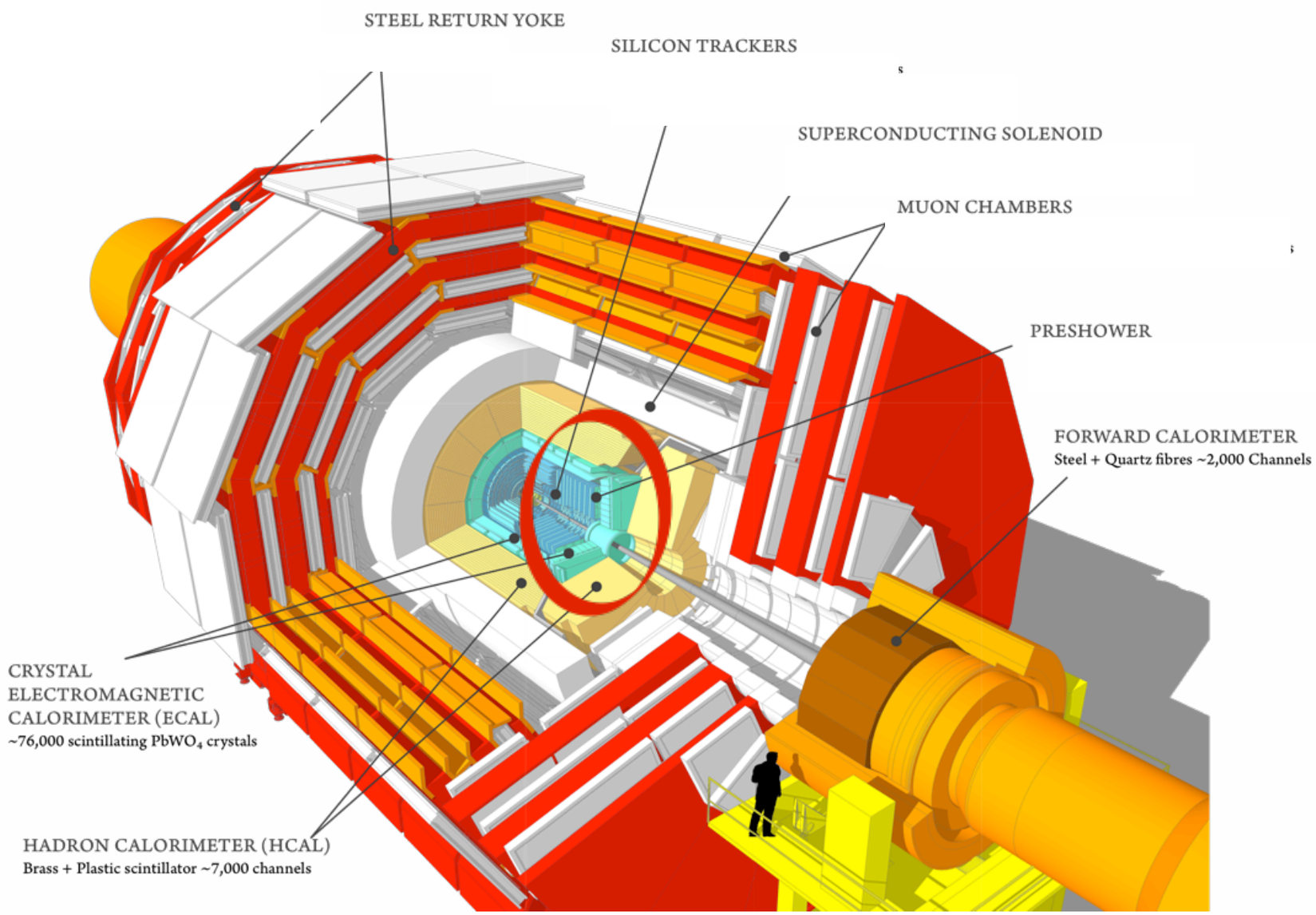


3 and 5σ discovery potential for G_{RS}



backup

CMS at the LHC



High energy photon ID

- Final working point selection set optimized for analysis sensitivity.
  - Taking broad set of signal hypoteses to avoid over-optimization.

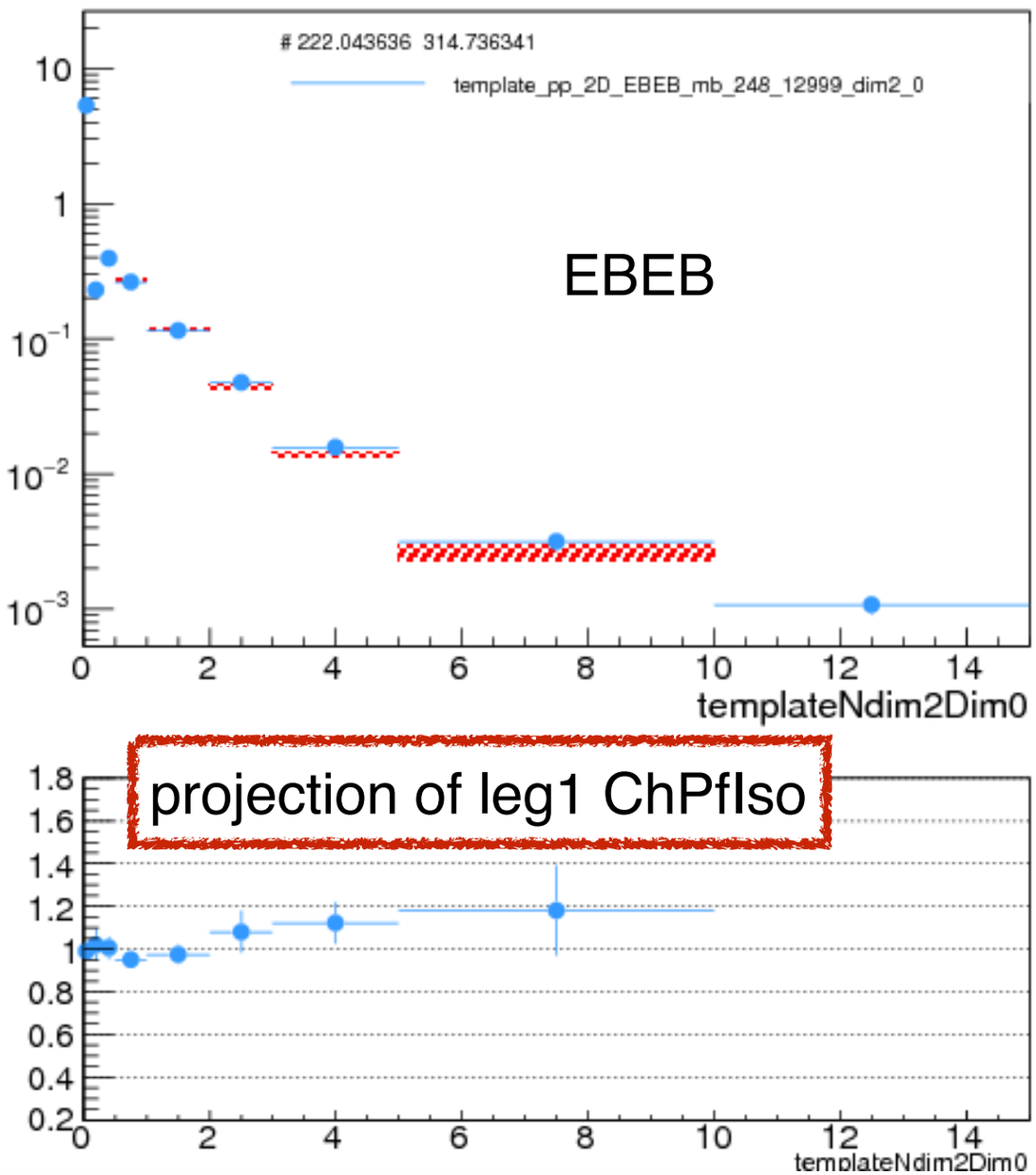
	$\overline{\text{Iso}}_\gamma$ (GeV)	ISO_{Ch} (GeV)	$\sigma_{i\eta i\eta}$	H/E	Electron veto
EB HighR9	2.5	5	0.0105	0.05	Conversion-safe
EB Low R9	2.5	5			
EE High R9	2	5	0.028		
EE Low R9	2	5			

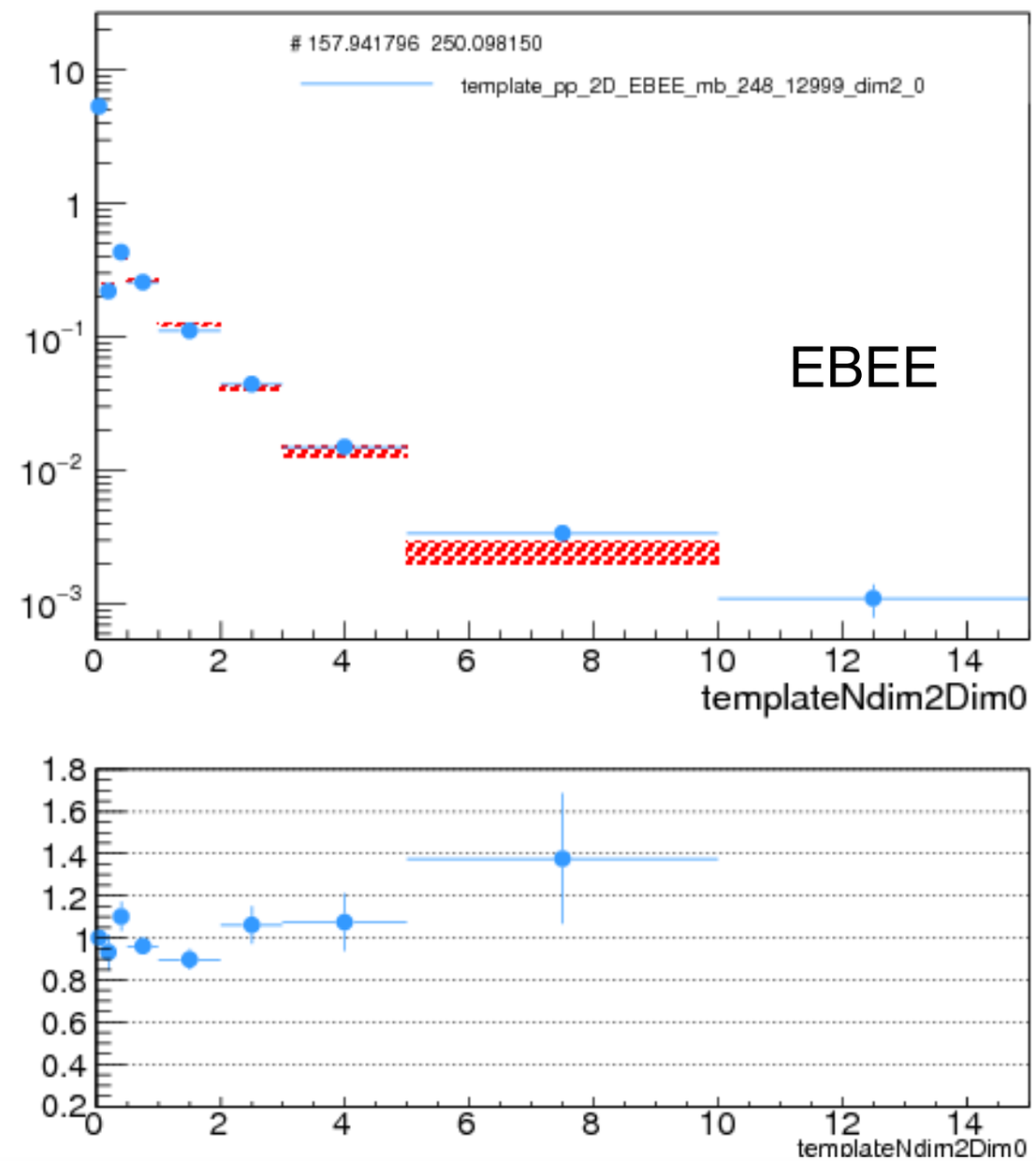
$$\overline{\text{Iso}}_\gamma = \alpha + \text{Iso}_\gamma - \rho \cdot A - \kappa \cdot p_T$$

Eta bin	0-0.9	0.9-1.5	1.5-2	2.-2.2	2.2-2.5
A	0.21	0.2	0.14	0.22	0.31
α	1.5	1.5	2	2	2
κ	2×10^{-3}	2×10^{-3}	2×10^{-3}	2×10^{-3}	2×10^{-3}

Prompt-prompt template

- EBEB: ChPFIso(random cone) of each leg randomly projected on the two axes of the 2d histogram
- EBEE: ChPFIso(random cone) of EB photon candidate on 1st axis

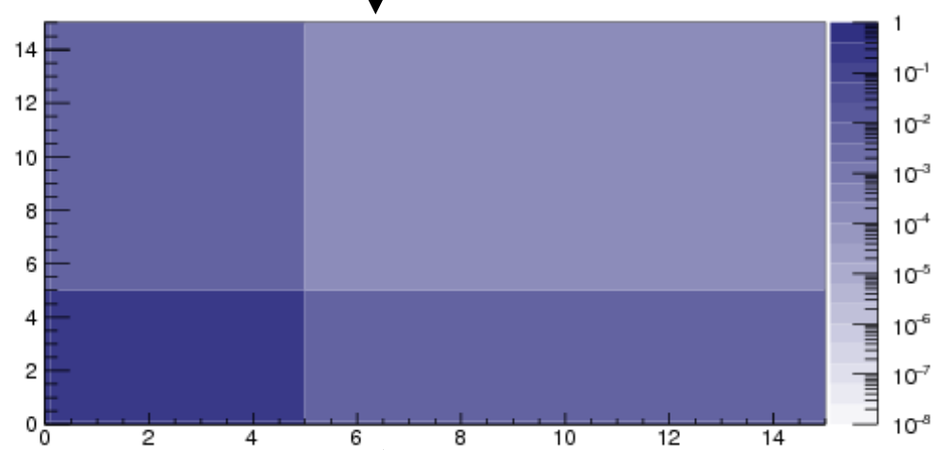




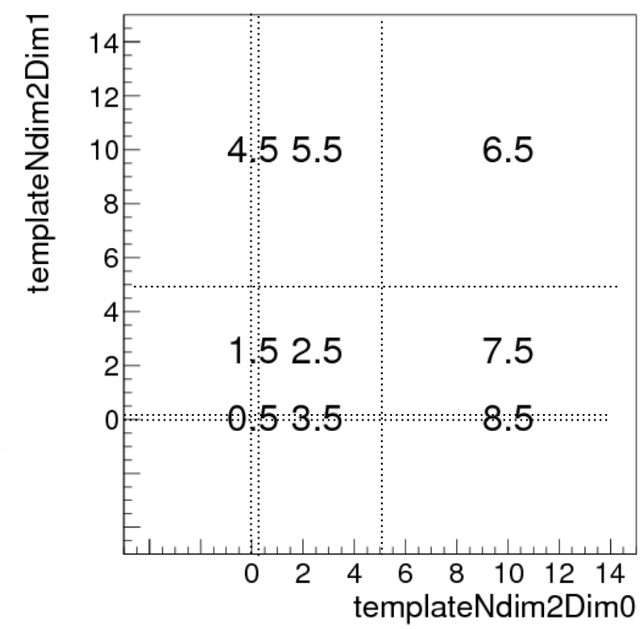
Technicality: “unroll” templates

goal: use these histograms in semi-parametric fit with combine tool

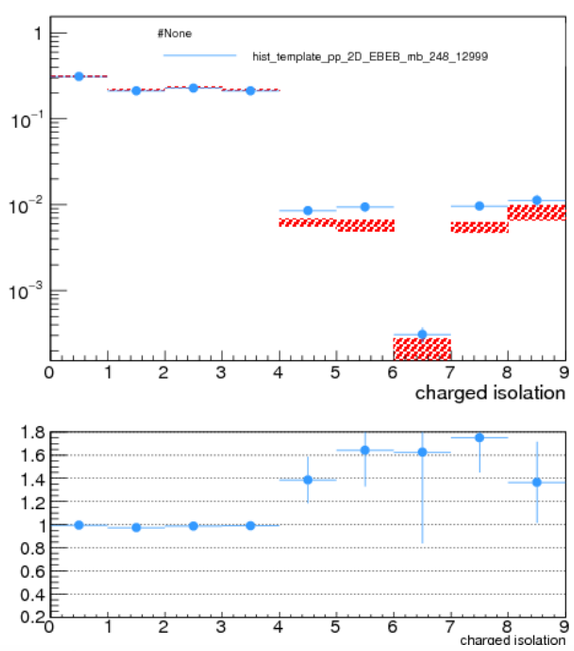
start from 2d ChPfIso-random cone Isolation with 9 bins

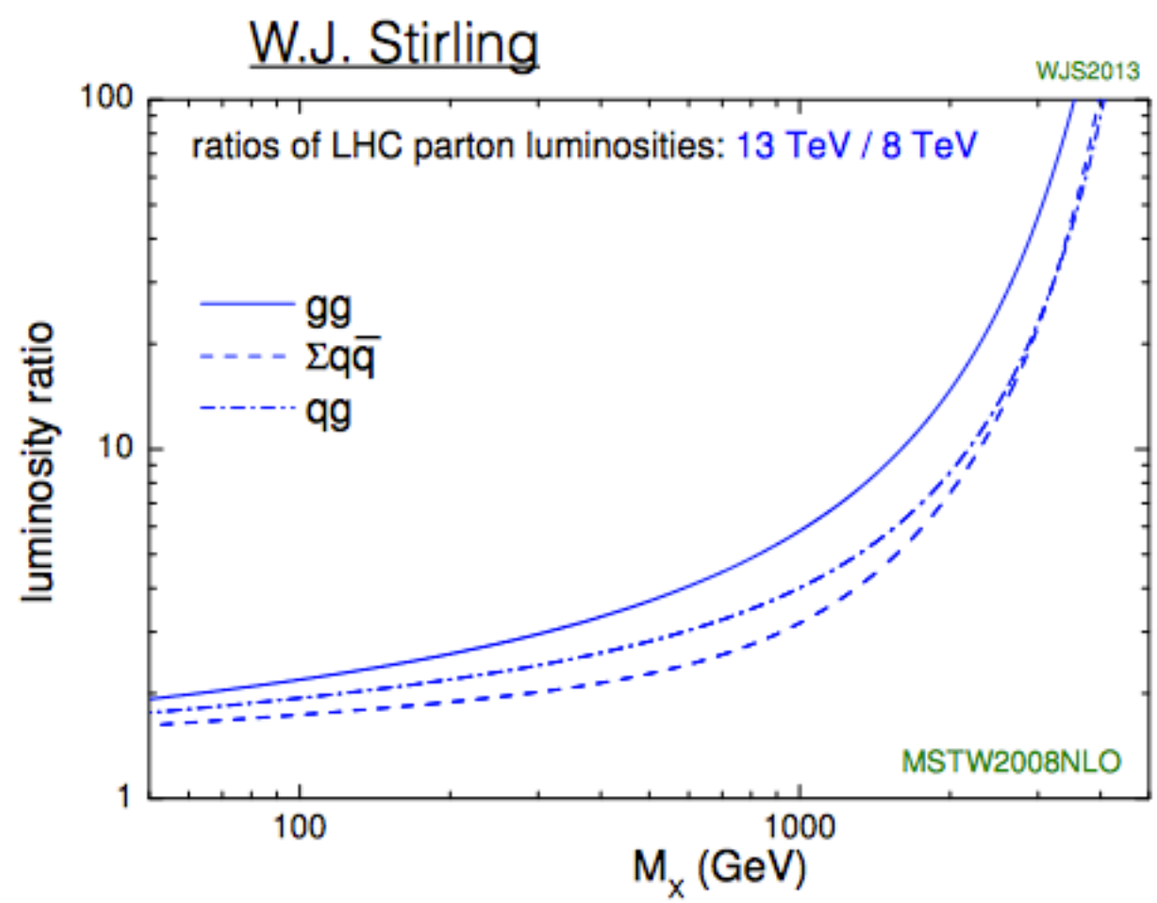


unroll this histogram in its bins



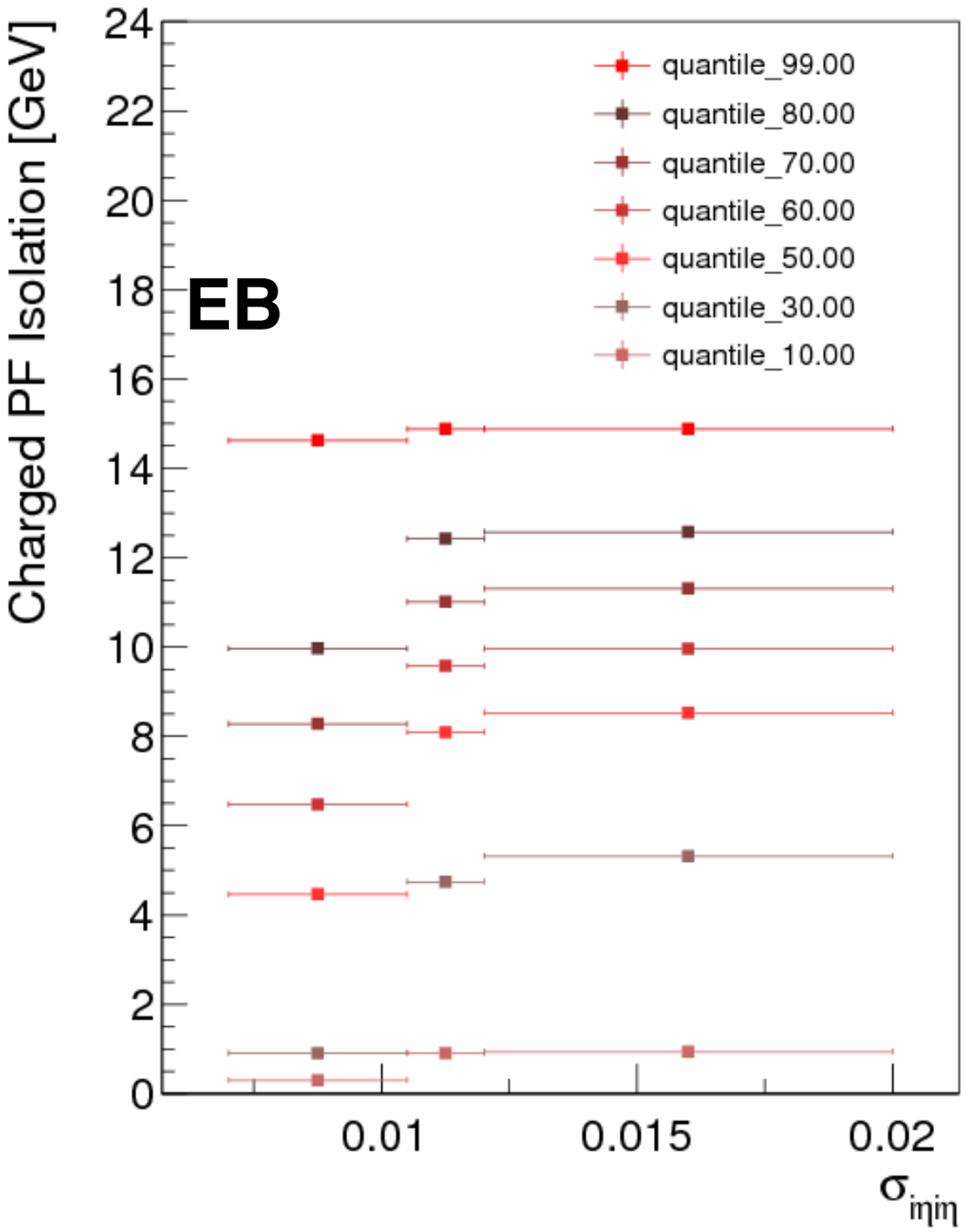
and fill the bin content in a 1d histogram

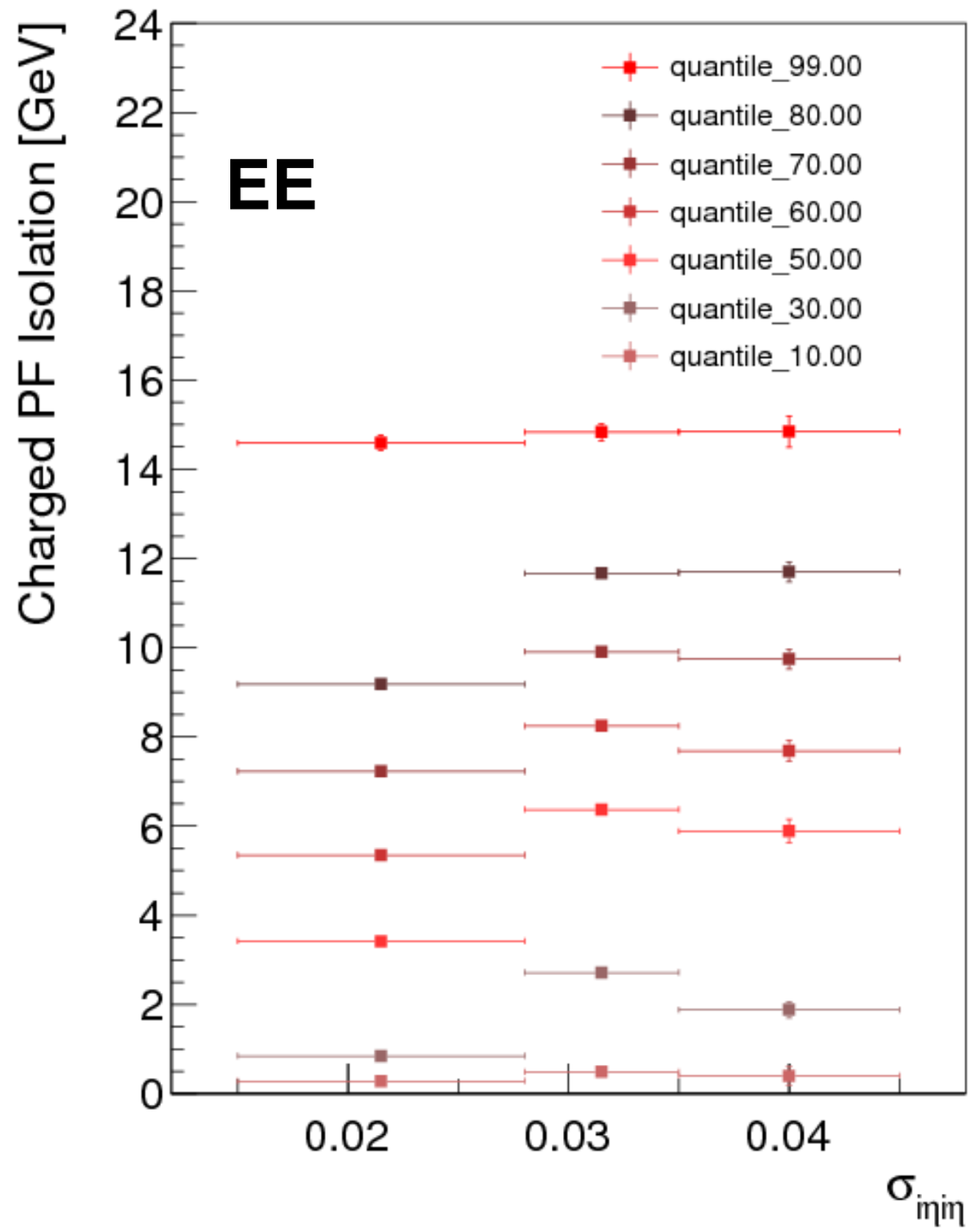




August 27th, 2015

Works only if ChPfIso & $\sigma_{i\eta i\eta}$ uncorrelate





Not enough events for f-templates?

- problem for pf and ff templates:
  - if only “pool” of direct diphoton events from data
  - not enough events to populate the (pf and) ff-template with reasonable statistics
  - especially for high p_t photon candidates in sideband

Expected number of diphoton events for the 2-dimensional templates per fb.

category	expected events in 2d templates	
	prompt-fake	fake-fake
EBEB	82	86
EBEE	418	61

event mixing

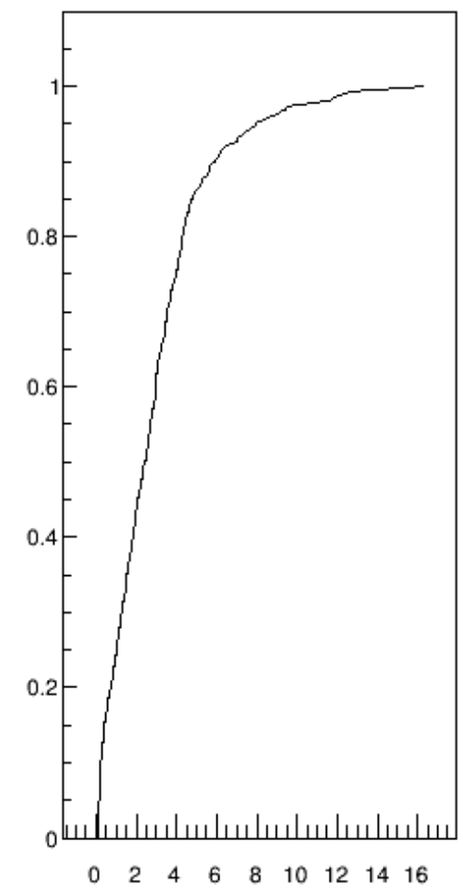
the art of event mixing

take single photon selection as “pool” for suitable photon candidates for templates:

- $p_t > 170$ GeV because of single photon trigger
- p (with ChPFIso random cone) & sideband single photons with similar phase space values

match photons from diphoton event to single photons:

- with kNN: k-dimensional nearest neighbour algorithm
  - ca 20 neighbours
  - k=3 variables for matching
    - EBEB: $\log(p_t)$, η, nvtx
    - EBEE: $\log(p_t)$, η, phoIso
  - build cdf for each variable -> same metric
  - match photon to single photon with smallest cdf distance



  - include correlations between two legs
    - pf EBEB: p 80% leading photon
    - pf EBEE: p 60% in barrel

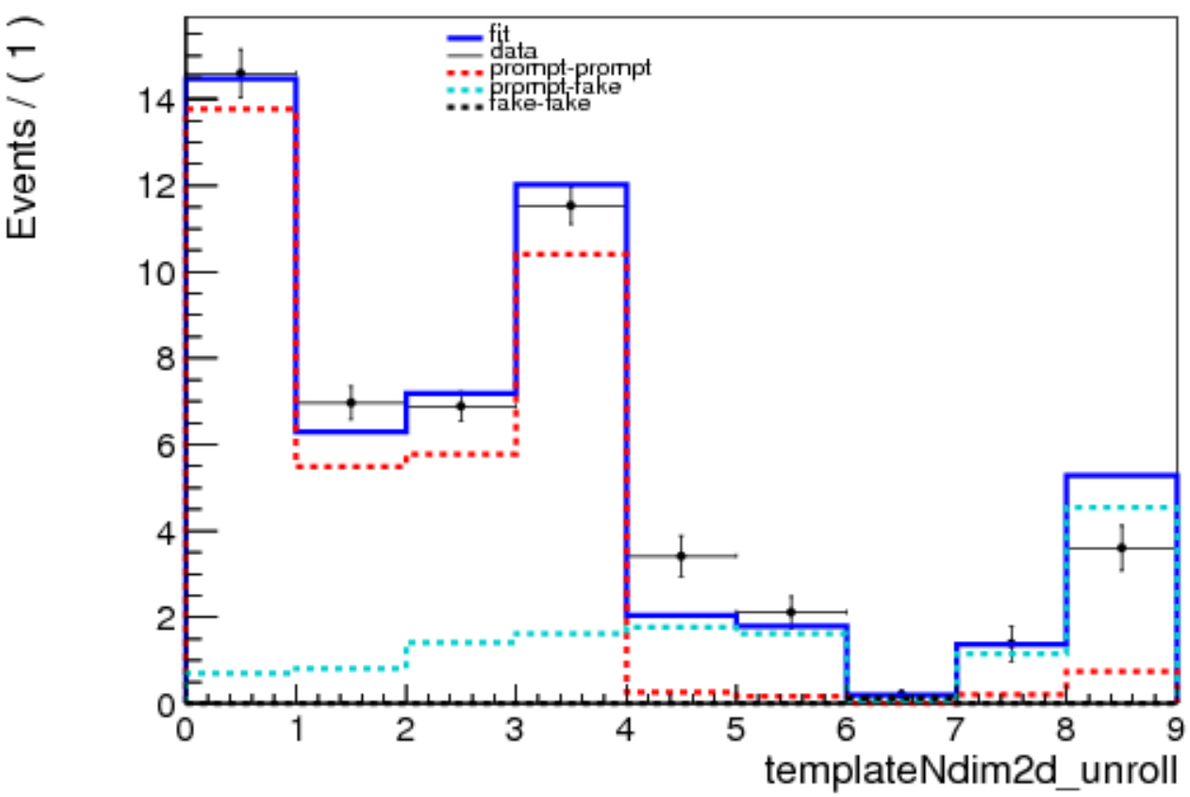
build template and project ChPfIso (random cone for p) for the categories as for pp template

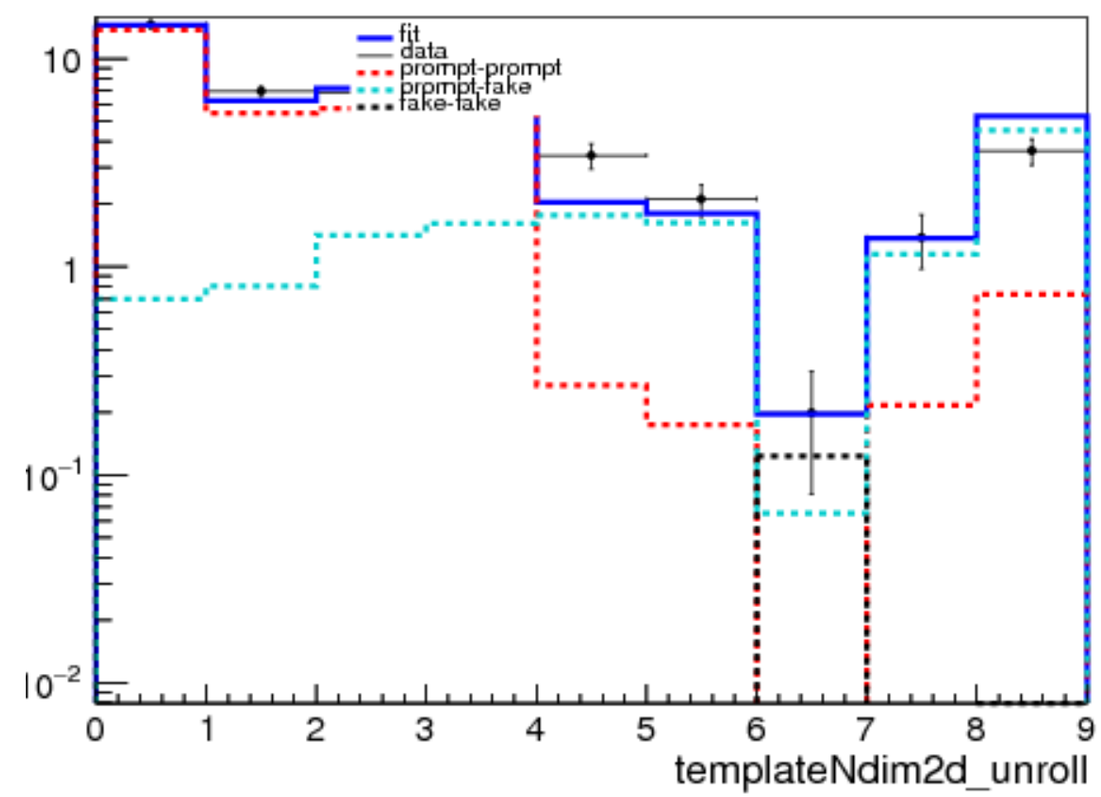
Example of “2d template “ fit

- create templates in bins of m$\gamma\gamma$
- with limited statistics for pf and ff: 5 bins
- 2d fit with templates

$$c^{pp} \cdot pdf^{pp}(ChIso_1, ChIso_2) \cdot + c^{pf} \cdot pdf^{pf}(ChIso_1, ChIso_2) + (1 - c^{pp} - c^{pf}) \cdot pdf^{ff}(ChIso_1, ChIso_2)$$

example: fit for highest mass bin EBEE (518 GeV - 7 TeV)



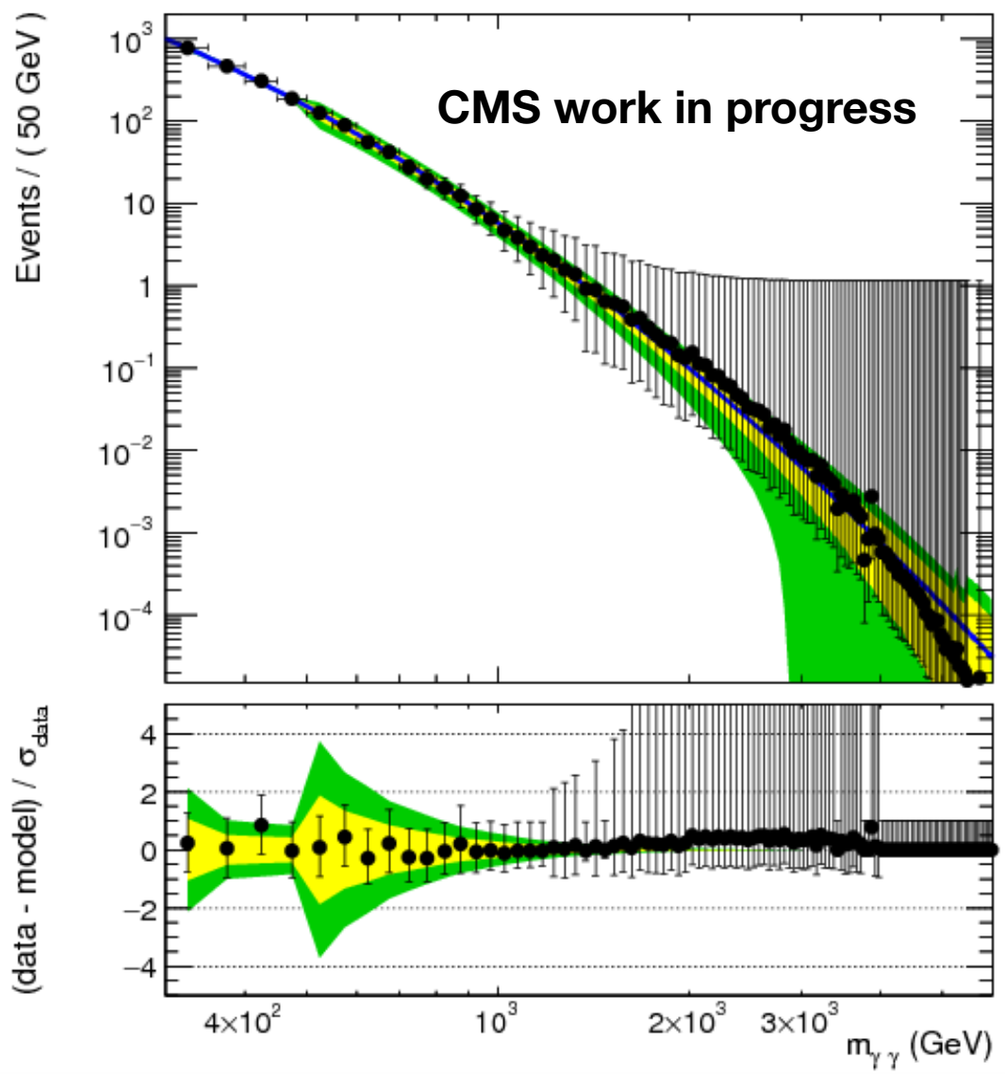


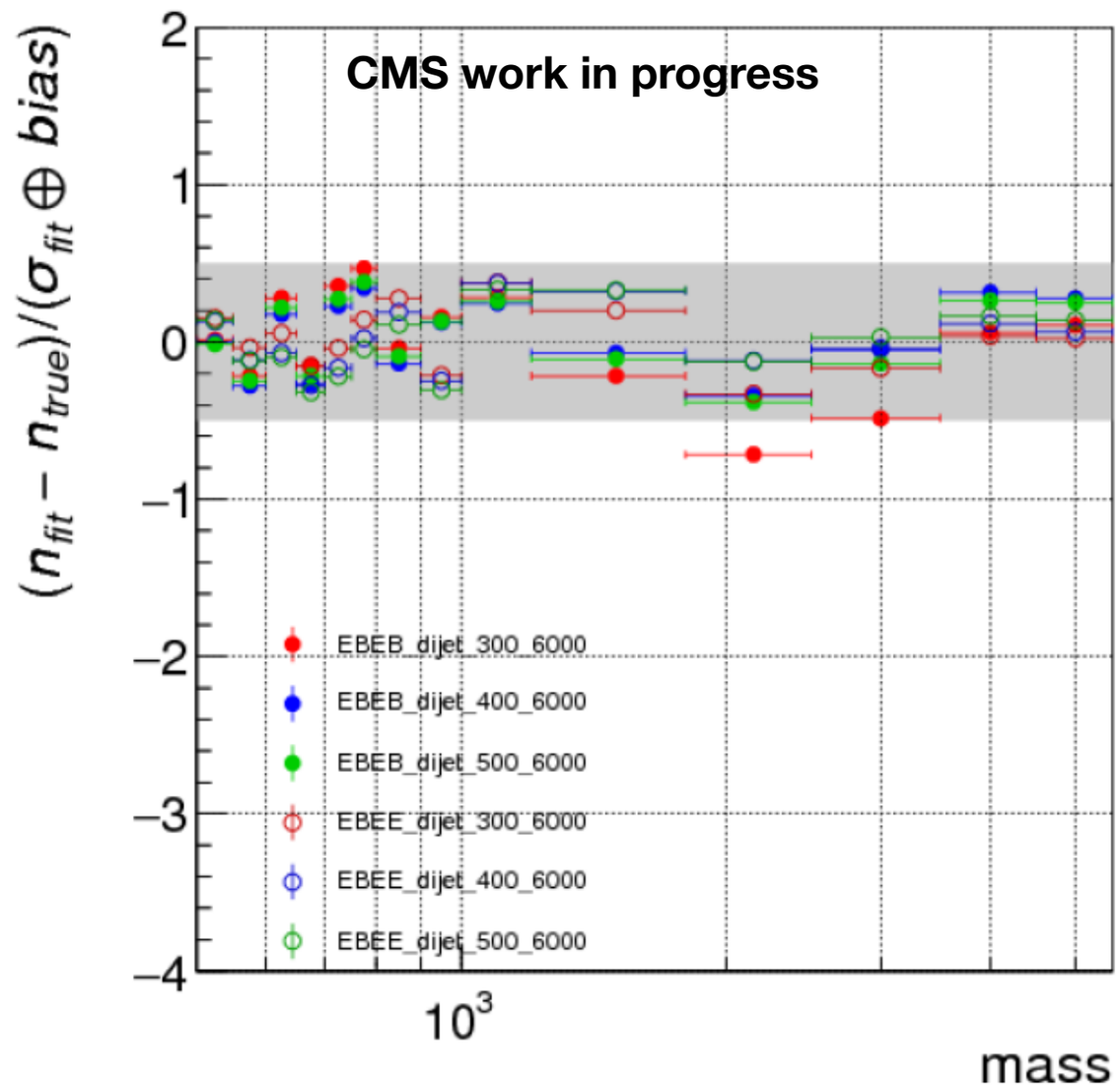
background model : parametric

- baseline parametrisation: $model_{bkg}(m_{\gamma\gamma}) = m_{\gamma\gamma}^{a+b\log(m_{\gamma\gamma})}$
- bias uncertainty: add explicit bias term:

$$n_b^{EBEB}(m_{\gamma\gamma}) = 0.22 \cdot \left(\frac{m_{\gamma\gamma}}{600\,GeV}\right)^{-5} + 10^{-6}\,events/GeV$$

$$n_b^{EBEB}(m_{\gamma\gamma}) = 0.06 \cdot \left(\frac{m_{\gamma\gamma}}{600\,GeV}\right)^{-4} + 10^{-7}\,events/GeV$$





Projection of bkg components