Search for additional neutral Higgs bosons through the $\mathbf{H \to ZA \to \ell\ell b\bar{b}}$ process

IRN Terascale meeting
VUB, Brussels

Alessia Saggio[1]

[1]CP3 - Université catholique de Louvain

Louvain-la-neuve, Belgium

16-10-2019

Introduction

- Discovery of h(125) milestone for SM
- But we need BSM physics (to explain hierarchy problem, dark matter,...)

Search for new physics beyond the SM in the Higgs sector

Precise measurements of h(125) properties to:

- Assess whether h(125) is *the* SM Higgs boson
- Constrain BSM models

Search for **extended Higgs sector**

- Additional Higgs bosons predicted by many theoretical models

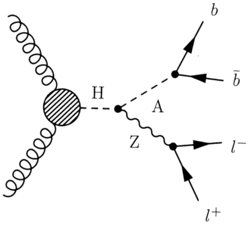
Extending the Higgs sector: the two-Higgs-doublet model

- **2HDM built by adding a second doublet to the SM scalar sector**
- It predicts **5 new Higgs bosons**:
  - 2 neutral CP-even: h, H
  - 1 neutral CP-odd: A
  - 2 charged: $H^{\pm}$
- Other important parameters:
  - $tan\beta = \frac{v1}{v2}$ (doublet vevs ratio)
  - $cos(\beta - \alpha)$, with α mixing angle between h and H
- Alignment limit: $cos(\beta - \alpha) = 0 \rightarrow h_{2HDM} = h_{SM}$

Talk focused on the search for two of these new particles: H and A

Setting the stage

- **Signal**: H $\rightarrow$ Z($\rightarrow \ell^+\ell^-$) A ($\rightarrow$ b$\bar{\text{b}}$)
    - $\ell = e, \mu$
- **Backgrounds**:
    - DY + heavy flavor jets
    - t$\bar{\text{t}}$ fully-leptonic
    - Single top, SM Higgs, VV(V), W+jets

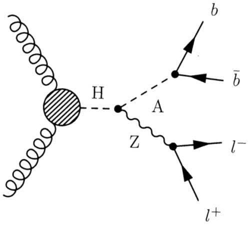


My job today: explain...

- **Where do we search?**
- **How do we search?**
- **What did we find?**

Setting the stage

- **Signal**: H $\rightarrow$ Z($\rightarrow \ell^+\ell^-$) A ($\rightarrow$ b$\bar{\text{b}}$)
  - $\ell = e, \mu$
- **Backgrounds**:
  - DY + heavy flavor jets
  - t$\bar{\text{t}}$ fully-leptonic
  - Single top, SM Higgs, VV(V), W+jets



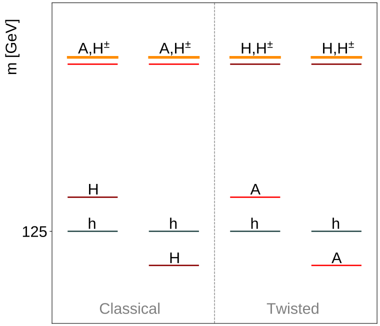
My job today: explain...

- **Where do we search?**
- **How do we search?**
- **What did we find?**

CMS-PAS-HIG-18-012

Where do we search?

The 2HDM mass hierarchy



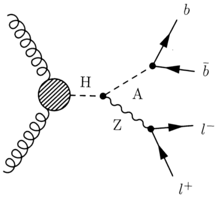
- Search conducted on **H→ZA** and then **mirrored** to extend to A→ZH for theoretical interpretation
- m_A range $= 30 \div 1000$ GeV
- m_H range $= 120 \div 1000$ GeV

▸ CMS-PAS-HIG-18-012

How do we search?

The object reconstruction

- Require 2 isolated opposite-sign leptons and 2 b-tagged jets
- The signal lies in the Z peak $\rightarrow 70 < m_{\ell\ell} < 110$ GeV
- The signal has low MET content $\rightarrow$ MET < 80 GeV
- **Three categories** are built according to the flavor of the di-lepton: $\mu^+\mu^-$, e^+e^-, $\mu^+e^- + \mu^-e^+$
- Reconstruct m_{jj} and $m_{\ell\ell jj}$

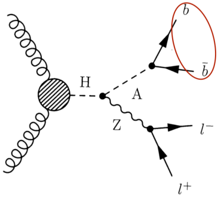


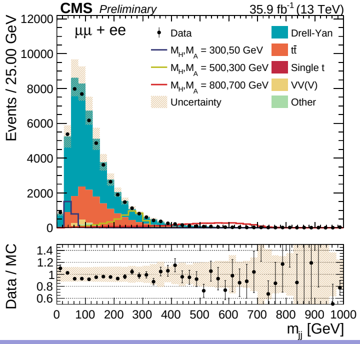
How do we search?

▸ CMS-PAS-HIG-18-012

The object reconstruction

- Require 2 isolated opposite-sign leptons and 2 b-tagged jets
- The signal lies in the Z peak $\rightarrow 70 < m_{\ell\ell} < 110$ GeV
- The signal has low MET content $\rightarrow$ MET < 80 GeV
- **Three categories** are built according to the flavor of the di-lepton:
  $\mu^+\mu^-$, e^+e^-, $\mu^+e^- + \mu^-e^+$
- Reconstruct m_{jj} and $m_{\ell\ell jj}$



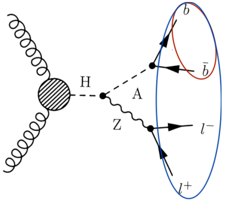


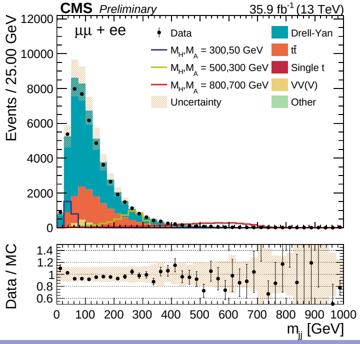
How do we search?

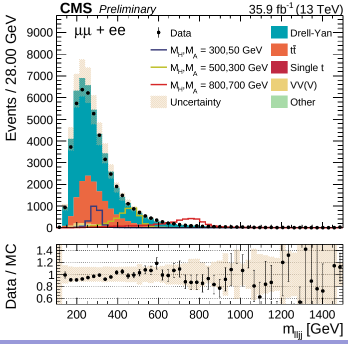
▸ CMS-PAS-HIG-18-012

The object reconstruction

- Require 2 isolated opposite-sign leptons and 2 b-tagged jets
- The signal lies in the Z peak $\rightarrow 70 < m_{\ell\ell} < 110$ GeV
- The signal has low MET content $\rightarrow$ MET < 80 GeV
- **Three categories** are built according to the flavor of the di-lepton:
  $\mu^+\mu^-$, e^+e^-, $\mu^+e^- + \mu^-e^+$
- Reconstruct m_{jj} and $m_{\ell\ell jj}$





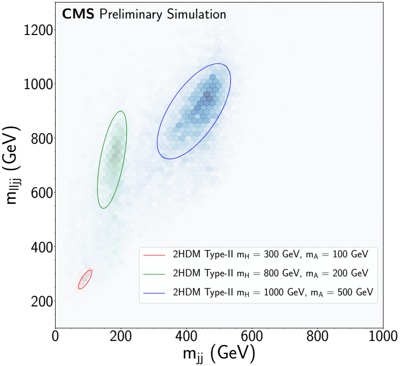


How do we search?

CMS-PAS-HIG-18-012

The signal region definition

- Forget the 1D MC signal mass distributions: **the 2D mass plane contains info about the correlation**
- The signal shape in the $(m_{jj}, m_{\ell\ell jj})$ plane is affected by the experimental resolution of the m_{jj} and $m_{\ell\ell jj}$ distributions
- **Define the SR as an ellipse** in the $(m_{jj}, m_{\ell\ell jj})$ plane

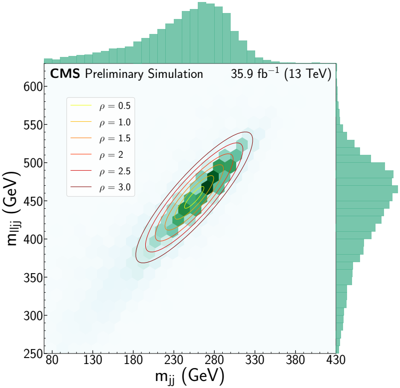


Continuous parametrization of the SR as a function of the mass hypotheses

Optimizing the signal region

▸ CMS-PAS-HIG-18-012

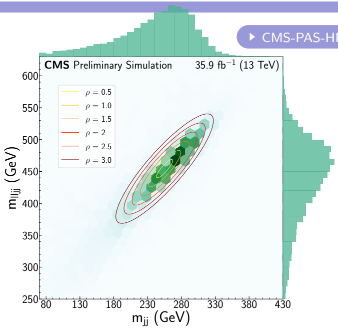
- Ellipses built under the assumption that the signal shape is Gaussian, while it's not!
- **Instead of fixing the size of the ellipse, vary it** to maximize the acceptance
- How? Define a parameter ρ: $\rho = 1$ contains roughly 1 std. dev. of the signal events



CMS-PAS-HIG-18-012

Binning of the m_{jj}-$m_{\ell\ell jj}$ plane

- Vary ρ in range [0.5, 1, 1.5, 2, 2.5, 3]



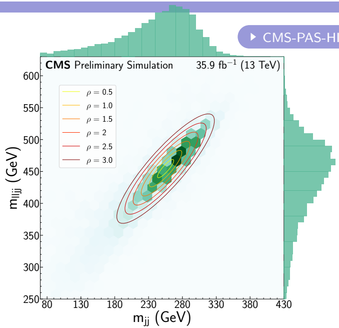
- Then fill histograms with events falling inside each elliptical bin

- Used as templates in the ML fit (e-μ as CR to further constrain $t\bar{t}$)

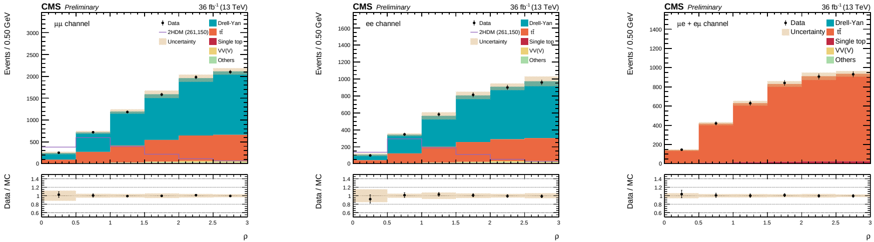
Binning of the m_{jj}-$m_{\ell\ell jj}$ plane

▸ CMS-PAS-HIG-18-012

- Vary ρ in range [0.5, 1, 1.5, 2, 2.5, 3]



  - Then fill histograms with events falling inside each elliptical bin

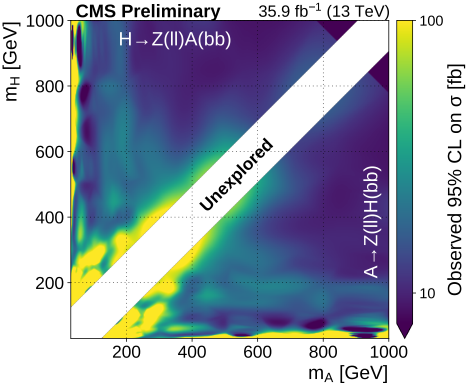


  - Used as templates in the ML fit (e-μ as CR to further constrain $t\bar{t}$)

What did we find?

CMS-PAS-HIG-18-012

Model-independent upper limits

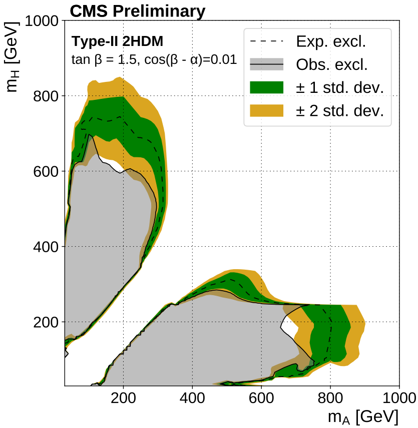


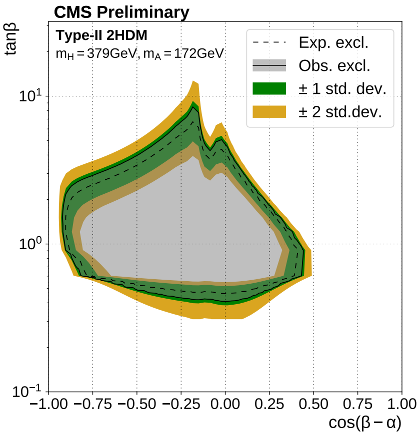
- Highest local significance observed: 3.9σ at $m_H = 627$ GeV, $m_A = 162$ GeV. Don't get too excited... the **look-elsewhere effect** plays a role and needs to be estimated (see next slides)

What did we find?

Interpretation in 2HDM

CMS-PAS-HIG-18-012



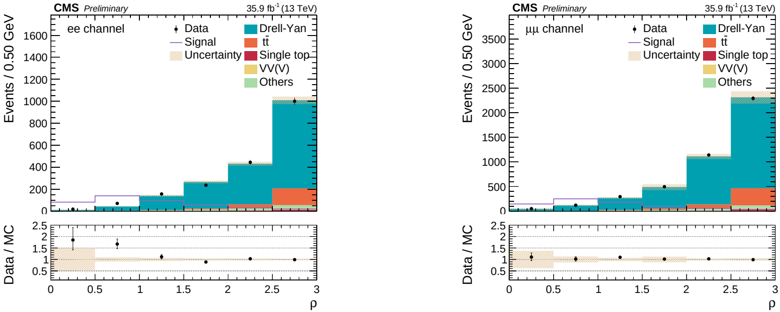


Excess: should we get excited?

Spoiler alert: no, we shouldn't...

CMS-PAS-HIG-18-012

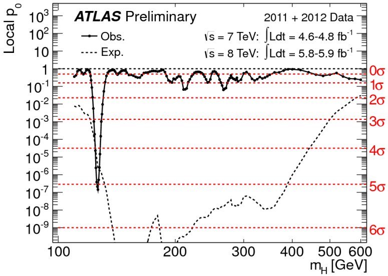
- Local significance of 3.9σ observed at $m_H = 627$ GeV, $m_A = 162$ GeV



- How much is the global significance?

The look-elsewhere effect

- **LEE: possibility for a signal-like fluctuation to appear anywhere within the search range**
- Need to estimate the **trial factor**, then used to correct the local significance
- In **1D (only one parameter defined under the alternative hyp.)**:
  - $p_{global} \simeq p_{local} + e^{\frac{u_0 - Z^2_{max}}{2}} N_{u_0}$
    - Z_{max} = max local significance
    - N_{u_0} = expectation number of up-crossings of test statistics at threshold u_0 (estimated with background toys)

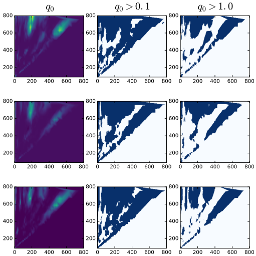


Estimating the look-elsewhere effect in the $H \to ZA$ search

▸ 2D LEE github repo

▸ Paper on multi-dim LEE

- Two parameters defined under the alternative hyp.: m_A and m_H $\to$ **it's a 2D problem**
- Generalization of 1D case (count "holes" instead of up-crossings)



- Toys generated via a **customized Kernel Density Estimation** for smoothing
- **First time this technique was used in CMS!**
- Result: $\sigma_{local} = 3.9 \to \sigma_{global} = 1.3$

More details in backup (or simply ask!)

Conclusions and perspectives

- Search for new Higgs bosons through $H \to ZA \to \ell\ell b\bar{b}$ presented
- Constraints set on 2HDM parameters
- A 3.9σ excess reduced to 1.3σ after estimating the LEE in 2D
- Overall, **no significant deviations from the SM found**
- Looking forward to searches with the full Run II with $\sim$150 fb^{-1}

Thank you!

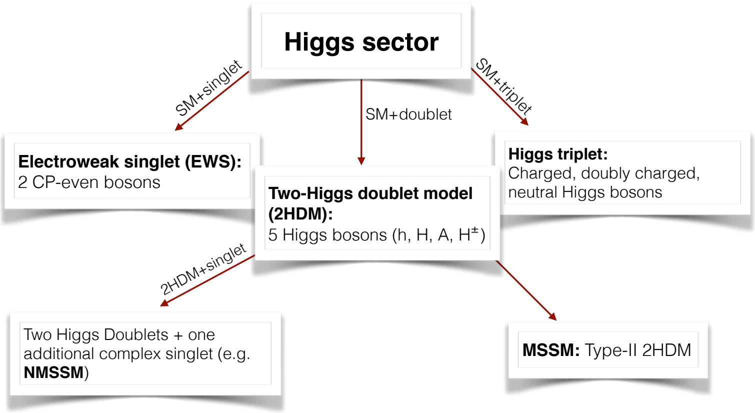
Conclusions and perspectives

- Search for new Higgs bosons through $H \to ZA \to \ell\ell b\bar{b}$ presented
- Constraints set on 2HDM parameters
- A 3.9σ excess reduced to 1.3σ after estimating the LEE in 2D
- Overall, **no significant deviations from the SM found**
- Looking forward to searches with the full Run II with $\sim$150 fb^{-1}

Thank you!

Backup slides

Extending the Higgs sector



2HDM couplings

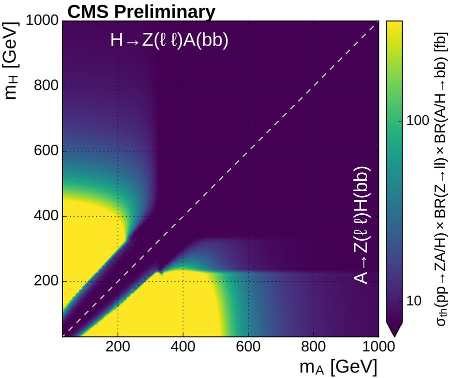
	Type I	Type II	Lepton-specific	Flipped
ξ_h^u	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$
ξ_h^d	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$
ξ_h^ℓ	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$
ξ_H^u	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$
ξ_H^d	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$
ξ_H^ℓ	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$
ξ_A^u	$\cot\beta$	$\cot\beta$	$\cot\beta$	$\cot\beta$
ξ_A^d	$-\cot\beta$	$\tan\beta$	$-\cot\beta$	$\tan\beta$
ξ_A^ℓ	$-\cot\beta$	$\tan\beta$	$\tan\beta$	$-\cot\beta$

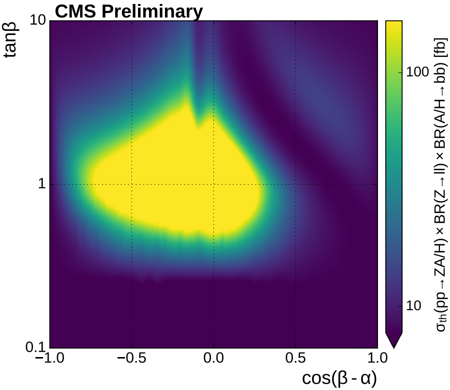
Table 2: Yukawa couplings of u, d, ℓ to the neutral Higgs bosons h, H, A in the four different models. The couplings to the charged Higgs bosons follow Eq. 16.

2HDM types

Two doublets ϕ_1 and ϕ_2.

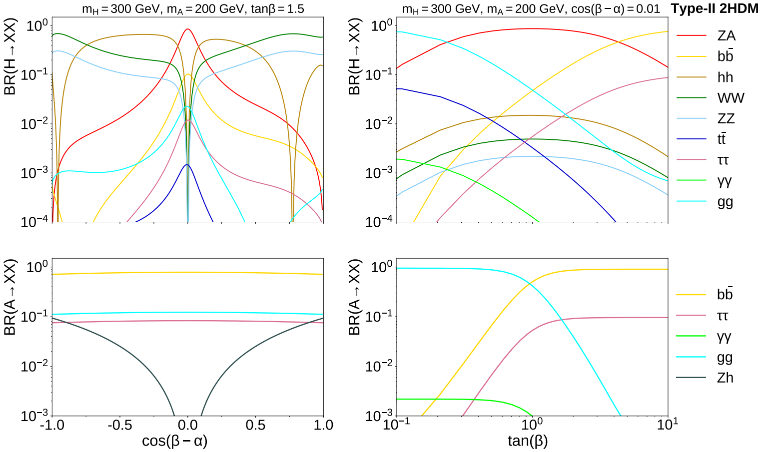
- **Type-I**: all SM particles couple to one doublet only;
- **Type-II**: up-type quarks couple to ϕ_1, down-type quarks and leptons couple to ϕ_2
- **Type-III**: quarks couple to ϕ_1, leptons to ϕ_2
- **Type-IV**: leptons and up-type quarks couple to ϕ_1, down-type quarks couple to ϕ_2


CMS Preliminary
H → Z(ℓ ℓ)A(bb)
A → Z(ℓ ℓ)H(bb)
m_H [GeV]
m_A [GeV]
1000
800
600
400
200
200
400
600
800
1000
100
10
σ_th(pp → ZA/H) × BR(Z → ll) × BR(A/H → bb) [fb]


CMS Preliminary
tanβ
cos(β - α)
10
1
0.1
−1.0
−0.5
0.0
0.5
1.0
100
10
σ_th(pp → ZA/H) × BR(Z → ll) × BR(A/H → bb) [fb]

Why this decay chain?

CMS-PAS-HIG-18-012

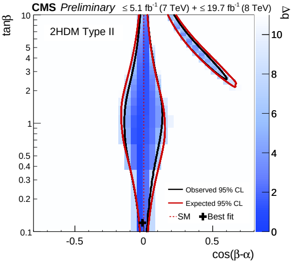


BR(H→ZA) and BR(A→ $b\bar{b}$) maximized in the alignment limit and over a large range of $tan\beta$

BSM searches @RunI

2HDM indirect searches

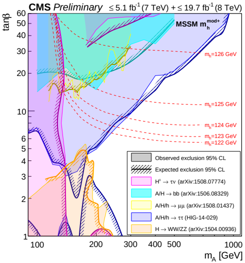
- The exclusion contours are derived from h(125) coupling measurements under the **alignment limit** assumption



$$-0.1 < \cos(\beta - \alpha) < 0.1$$

MSSM direct searches

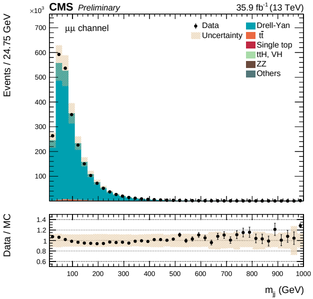
- Several benchmark scenarios for different phase space properties, e.g. MSSM m_h^{mod+}, hMSSM

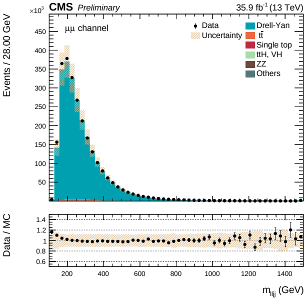


Down-type fermionic channels become interesting

Reweighting the DY background 1/2

- In a **CR** defined by **not requiring the di-jet to be b-tagged**:

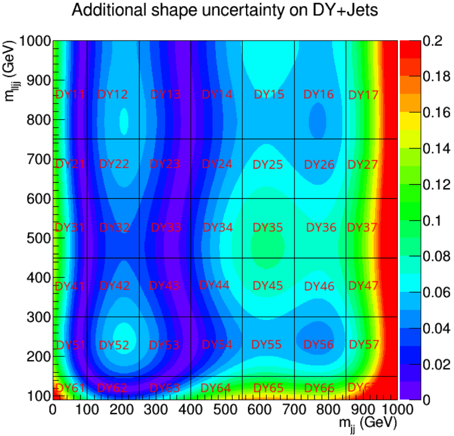




- A **discrepancy up to** $\sim$**10%** is observed is some regions of the parameter space
- To correct for this, **each DY event is reweighted** by the observed discrepancy and **an uncertainty of 100% of this discrepancy** is applied
- In regions where the ellipses are fairly big, we want to avoid that well-modeled regions (very small uncertainty) constrain mismodeled regions (high uncertainty) $\rightarrow$
- **Uncorrelate** this uncertainty across the mass plane by sampling it in 42 regions of approximately 150×150 GeV^2
- Additional 42 uncorrelated shape uncertainties added

Reweighting the DY background 2/2

- Possible complication: when the ellipses are fairly large, regions with different kinematics might constrain each other
- One needs to uncorrelate this shape unc. across the plane
- Mass plane is split in regions of area 150x150GeV^2, yielding 42 new systematics



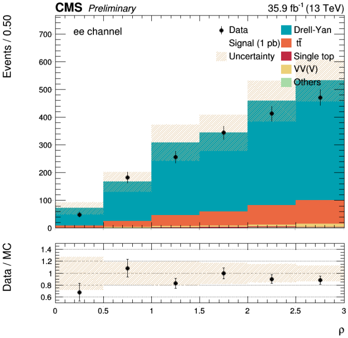
- For reference: x-sec$_{DY}$ uncertainty: 5%

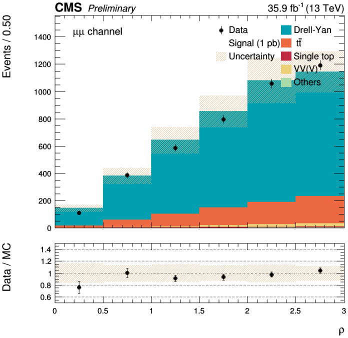
Generating background toys 1/2

- We want to generate N background toys
- This means generating N different mass planes from which to extract the ellipses, or equivalently the ρ histograms
- The toy distribution in the mass plane is drawn from the existing one (DY+ttbar)
- In principle, one could directly generate toys from DY+ttbar in the 2D mass plane
- But stat. is really low in some regions! We want our background distirbutions in the mass plane to be as smooth as possible
- How? Smooth DY+ttbar in the mass plane with a "customized" Kernel Density Estimation
  - The smoothing is elliptic instead of Gaussian, with ellipse parameters interpolated from the existing ones event by event
- With these events, fill a very finely binned TH2 $\rightarrow$ here's DY+ttbar in the mass plane, but smoothed!

Generating background toys 2/2

- Generate N toys from this TH2 with no. events = no. events in data
- Fill ρ histograms for each TH2
- These toys are used as fake data in the LEE estimation





- Specifics:
  - N = 15
  - No. ellipses = 900
- I.e. 900 ρ histograms for each toy

LEE in 1D and 2D

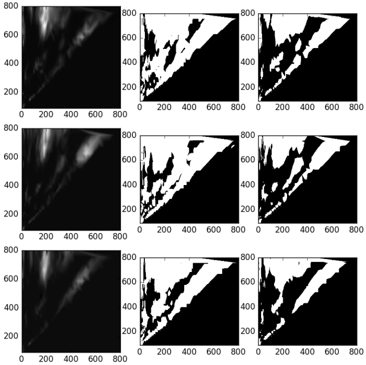
- LEE: possibility for a signal-like fluctuation to appear anywhere within the search range
- Need to estimate the trial factor, then used to correct the local significance (or p-value)
- In 1D:
  - $p_{global} \simeq p_{local} + e^{\frac{u_0 - Z^2_{max}}{2}} N_{u_0}$
  - where: Z_{max} = max local significance, N_{u_0} = expectation number of upcrossings of test statistics at threshold u_0
- In 2D:
  - $p_{global} \simeq p_{local} + e^{-\frac{Z^2_{max}}{2}} (N_1 + N_2 \; Z_{max})$
  - where: Z_{max} = max local significance, N_1, N_2 are two coefficients that we need to estimate

LEE in 2D

- For each toy, compute the likelihood ratio (it's a 2D distribution)
- Set a threshold $\rightarrow$ get holes in the plane (2D generalization of upcrossings in 1D)
- Then get N_1, N_2 from solving the following system:
  - $\mathbb{E}[\phi(A_{u_0})] = \mathbb{P}[\chi^2_1 > u_0] + e^{u_0/2}(N_1 + N_2\ \sqrt{u_0})$
  - $\mathbb{E}[\phi(A_{u_1})] = \mathbb{P}[\chi^2_1 > u_1] + e^{u_1/2}(N_1 + N_2\ \sqrt{u_1})$
- where $\mathbb{E}[\phi(A_{u_0})]$ ($\mathbb{E}[\phi(A_{u_1})]$) is the expectation value of the number of holes in the plane (more precisely Euler characteristic) at threshold u_0 (u_1)
  - How to get $\mathbb{E}[\phi(A_{u_i})]$, in practice? For each toy, get the Euler characteristic at threshold u_i and average over the number of toys
- I set $u_0 = 0.1$ and $u_1 = 1$ (but doesn't matter if number of toys is sufficiently high)
- It's stable with 15 toys already!

Results

- Rows: toys
- Columns:
  - 1) Likelihood ratio distribution in mass plane
  - 2) Likelihood ratio intersected with $u_0 = 0.1$
  - 3) Likelihood ratio intersected with $u_1 = 1$



- $\sigma_{local} = 3.9 \rightarrow \sigma_{global} = 1.3$