A new approach to treat non-resonant BSM physics at LHC applied to GF hh (targeting discovery)

Preliminary results

Tommaso Dorigo, Alexandra Oliveira*, Mia Tosi, Martino Dall'Osso
INFN and University of Padova

Higgs pair productions at colliders
27-30 April 2015
Mainz Institute for Theoretical Physics
Johannes Gutenberg University

Many thanks to Maxime Gouzevich and Olivier Bondu for nice synergy

Fast introduction

Different types of new physics can induce both big cross section variations AND big signal topology variations of GF hh production wrt SM

How to choose benchmarks to searches?

↔

How to have as best as possible experimental results in non-resonant hh useful to interpretation in a big plethora of models?

Fast introduction

Different types of new physics can induce both big cross section variations AND big signal topology variations of GF hh production wrt SM

How to choose benchmarks to searches?

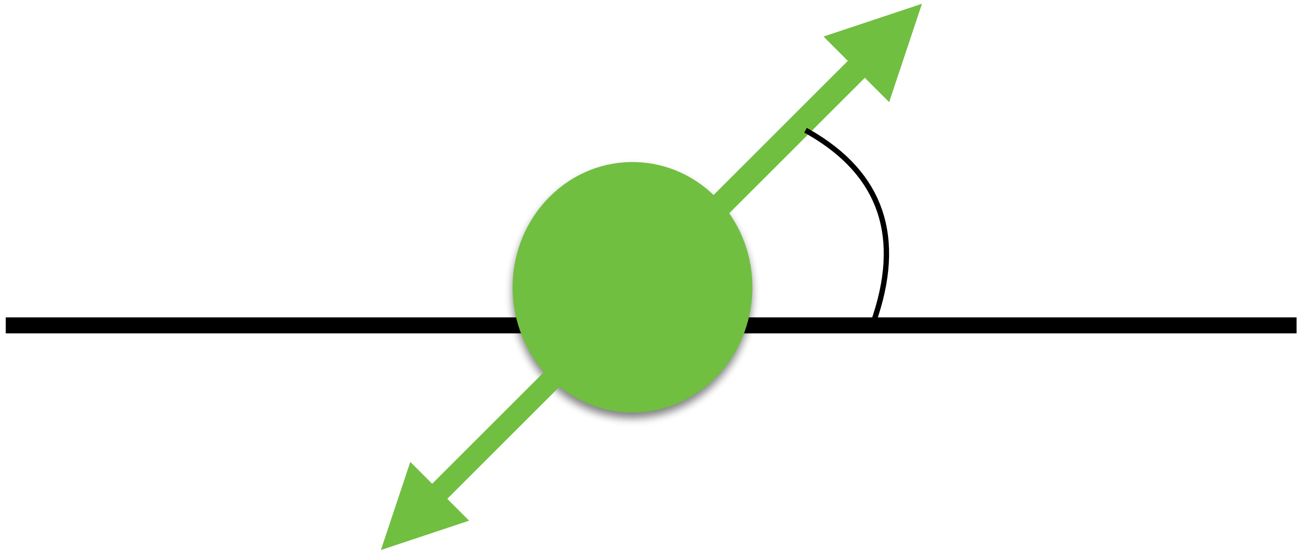
↔

How to have as best as possible experimental results in non-resonant hh useful to interpretation in a big plethora of models?

Paraphrasing Jose we can start pretending to be dumb

Conceptually EXP put cross sections limits in full shapes hypotheses

It is a 2 -> 2 process at leading order



CM energy (mhh)
+ angles (ME properties)

Theory wise: very generically physics models needs to know only mhh in order to interpret results.

Experimental wise: acceptance, object reconstruction and analysis methods need to know all the process information to be best.

pt of a higgs candidate
pz of higgses (PDF)

Couplings basis for di-higgs experimental results

Worse than parametrize our ignorance in EFT coefficients we parametrize it in terms of the physical125 GeV boson (h) couplings

If we neglect enhancement of h-bb coupling

the parameters that define production are five

$$\Delta L = \partial_\mu h \partial^\mu h - m_H h^2 + \kappa_\lambda \lambda_{SM} v h^3 - \frac{\sqrt{2} m_t}{v}\left(v + \kappa_t h + \frac{\sqrt{2}}{v} c_2 h h\right)(\bar{t}_L t_R + h.c.)$$

$$+ \frac{\alpha_s(m_Z)}{v}\left(c_g h - \frac{c_{2g}}{2v} h h\right) G^{mu,nu} G_{mu,nu}$$

where $v = 246$ GeV , $\lambda_{SM} \equiv m_H^2/(2v^2) = 0.129$, and $/v = 0.995\,(\sim 1)$.

Approach: A model to construct benchmarks and understand sensitivity in shapes to different processes interferences

Cross section written as 15 terms polynomial in the parameters:

$$\sigma_{hh} = k_{fac}\,\sigma_{hh}^{LO} \times \Big(A_1\,\kappa_t^4 + A_2\,c_2^2 + (A_3\,\kappa_t^2 + A_4\,c_g)\,\kappa_\lambda^2 + A_5\,c_{2g}^2 + (A_6\,c_2 + A_7\,\kappa_t\kappa_\lambda)\kappa_t^2$$
$$+ (A_8\,\kappa_t\kappa_\lambda + A_9\,c_g\kappa_\lambda)c_2 + A_{10}\,c_2c_{2g} + (A_{11}\,c_g\kappa_\lambda + A_1 2\,c_{2g})\,\kappa_t^2$$
$$+ (A_{13}\,\kappa_\lambda c_g + A_{14}\,c_{2g} + A_{15}\,c_{2g})\,\kappa_t\kappa_\lambda \Big)$$

LO coefficients can be derived by simulations and fits

e.g. @ 14 TeV - from Panico et all'15

$A_1 = 2.13$
$A_2 = 10.1$
$A_3 = 0.3$
$A_4 = 21.8$
$A_5 = 188$
$A_6 = -8.62$
$A_7 = -1.43$
$A_8 = 2.93$
$A_9 = 21$
$A_{10} = 59.8$
$A_{11} = -9.93$
$A_{12} = -23.1$
$A_{13} = 4.87$
$A_{14} = 10.5$
$A_{15} = 96.6$

Subjected to PDF and fit uncertainties that should be quantified

Overall cross section normalization is approximate by the SM NNLO QCD value

$$@14\ \text{TeV}\ k_{fac}\sigma_{hh}^{LO} = 44.1\ \text{fb}$$

HH-x recommendations to GF HH CX

As reference, individual channels absorb higgs BR uncertainties

Which benchmarks could allow us most general search coverage?

We can change the question to two other questions

1) How different can be the kinematics of samples within parameter space of higgs couplings?

arxiv citations and big names

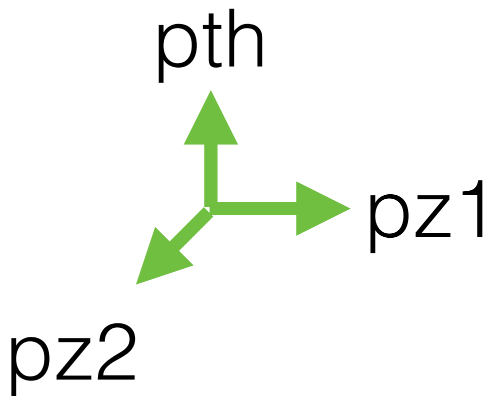
Criteria

kinematics behavior

2) how well LHC results could catch difference?

How different are the kinematics of samples within parameter space of higgs couplings?

Automat road to answer => statistics



- We construct a two-sample Test Statistics (TS) based in likelihood shape of 3D <u>normalized</u> binned histograms

(full kinematical information at LO)

$$TS = -\log L = -\log\left[\prod_{bins}\left(\exp^{-\frac{n_1+n_2}{2}}\right)^2\left(\frac{n_1+n_2}{2}\right)^{-n_1-n_2}(n_1!n_2!)^{-1}\right] =$$

$$= \sum_{bins}\left[n_1+n_2+(n_1+n_2)\log\left(\frac{n_1+n_2}{2}\right)+\sum_{i=1}^{n_1}\log i+\sum_{i=1}^{n_2}\log i\right]$$

20 bins for pt, 8 bins in each pz.
20k events are enough to populate O(1k) bins and do a meaningful test.

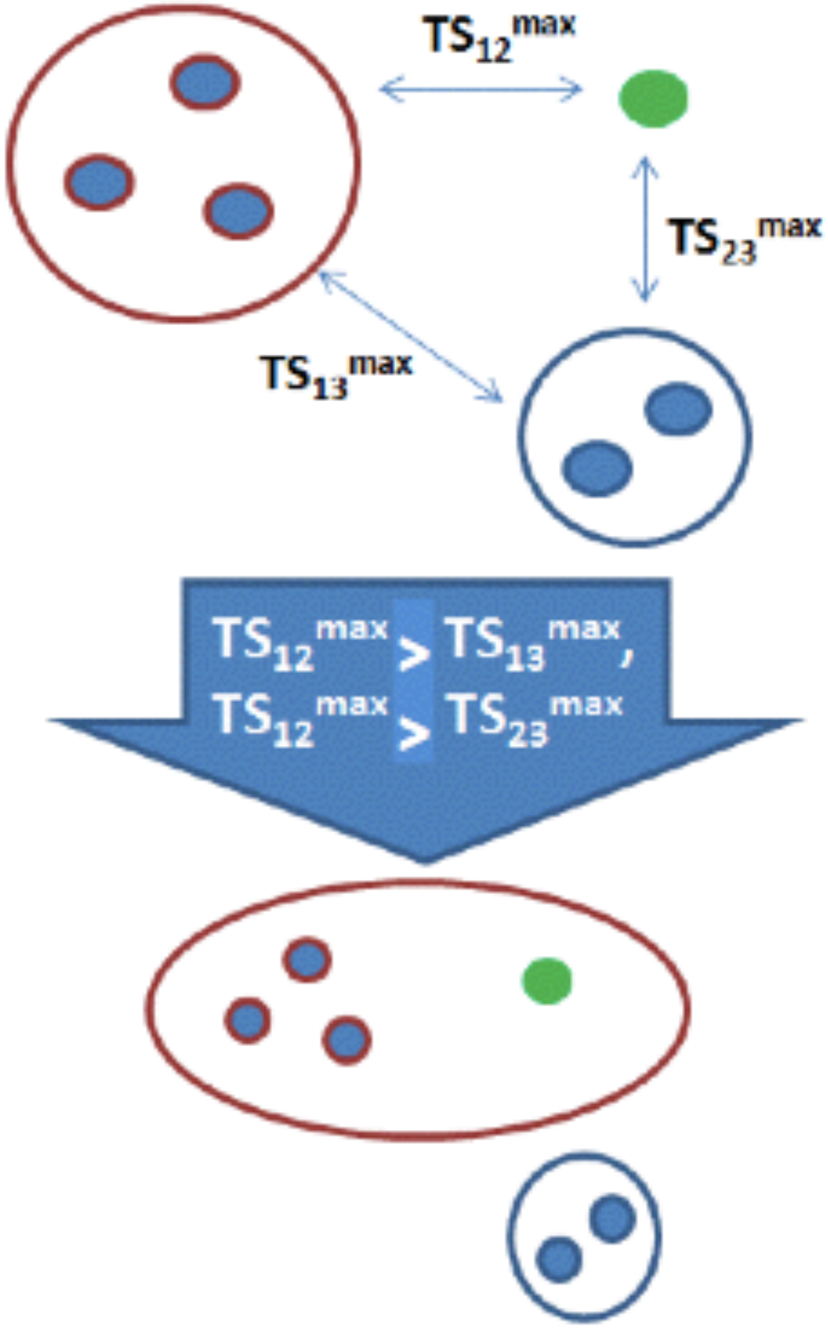
Computationally feasible to perform ME level big parameter scan

Clustering benchmarks based in the two sample TS

- We take ***TS_{ij}*** **as ordering parameter.**
- At the start all samples are considered "clusters of one element".
- We define the **distance between two clusters** as:

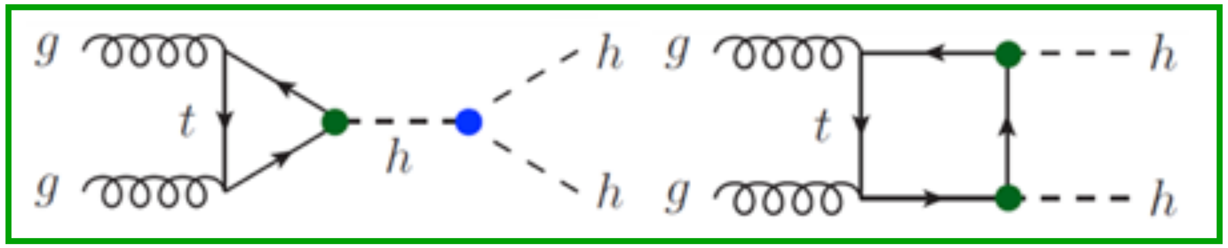
$$CD_{ij} = \max_{m=1\ldots N_i, n=1\ldots N_j}[TS_{mn}]$$

- We look for the pair of clusters *i,j* which have **maximum value of CD_{ij}** among all the possible combinations, and merge them together in a single cluster.
- This is iterated as desired. **We repeat until we are left with a pre-defined number of clusters (*totClus*).**
- In each cluster a **benchmark** is defined as: the sample which has smallest value of the maximum distance from all other elements of a group.

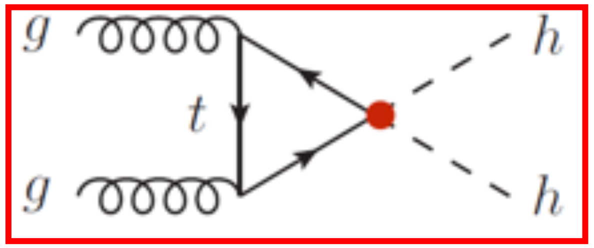


Relative kinematical distance between parameter space points

Illustrate the method in three parameter dimension



SM-like diagrams



ttHH interaction

The Procedure:

1. Produce **ME simulation** for N points of the parameters space (with N >> n° of clusters).
2. Perform a **two-sample test** between each pair of samples to measure their proximity in kinematics.
3. By using the value of the test statistic, perform a **Cluster Analysis** to group samples together and to identify a benchmark for each cluster.
   - The benchmark of each cluster is defined as the one closer to all the other cluster members

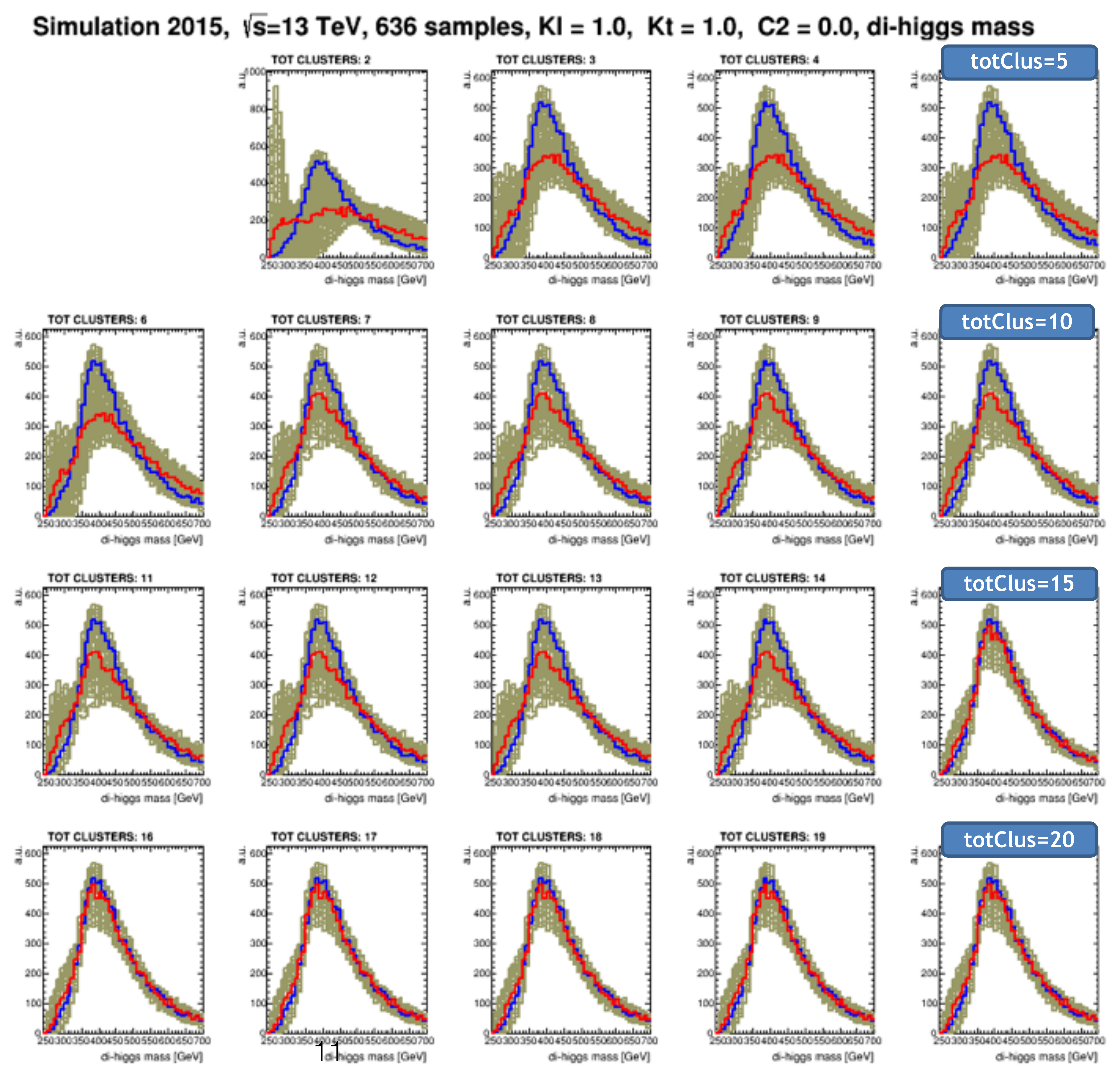
Parameter scan with 636 simulations based in gradients of cross section and current experimental limits from single and SM-like double higgs production
(wait a bit to know where are the test points)

Quality of node clusters in the di-higgs invariant mass

Follow the evolution of the cluster node that contains the SM point in comparison with the benchmark node

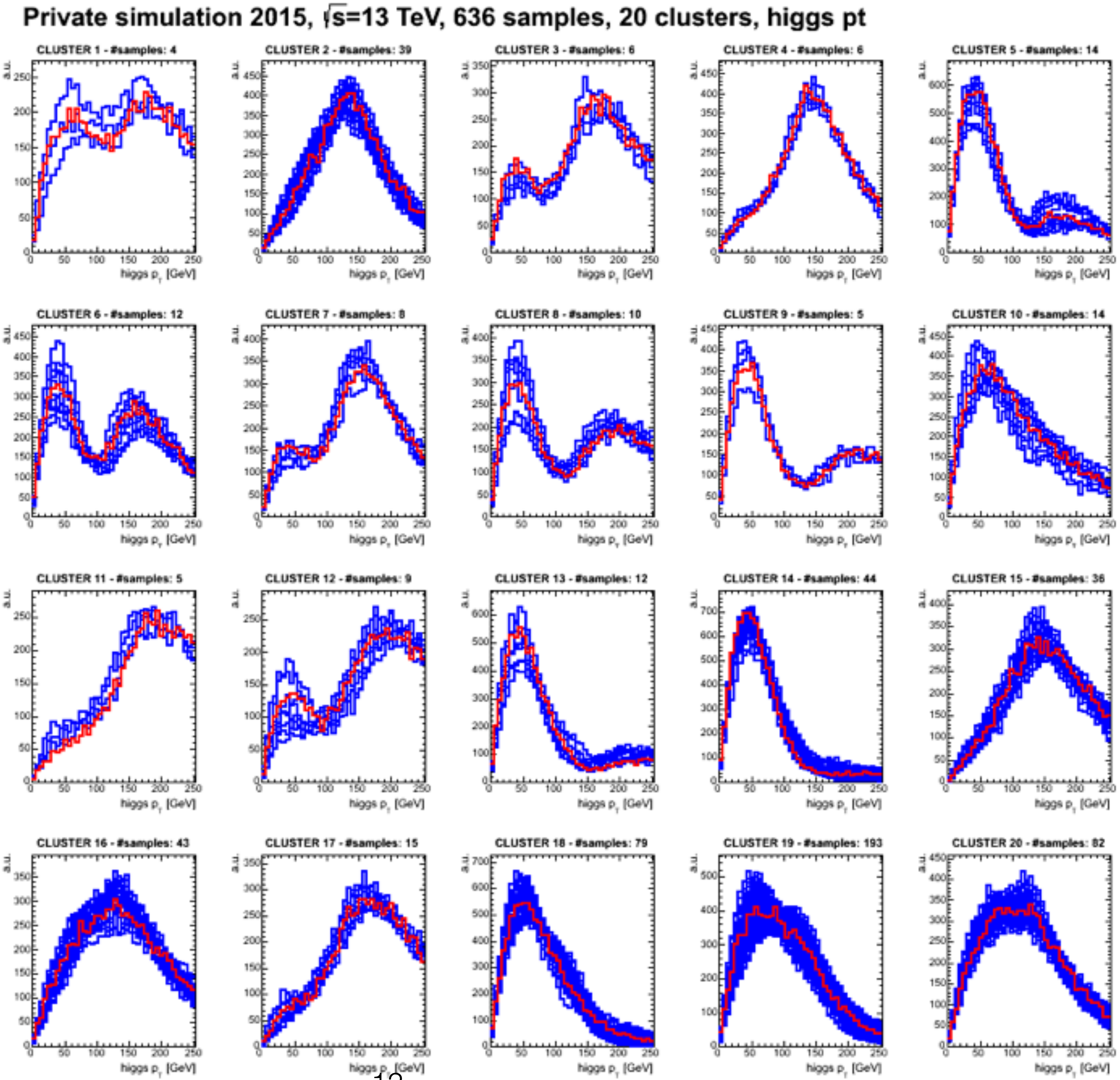
Visible clustering homogeneity with toClus > 15

The longitudinal higgses momentum are very homogeneous



20 cluster sampling in the individual higgs pt

Different strategies for higgses reconstruction need/can be exploited to cover BSM hh different topologies

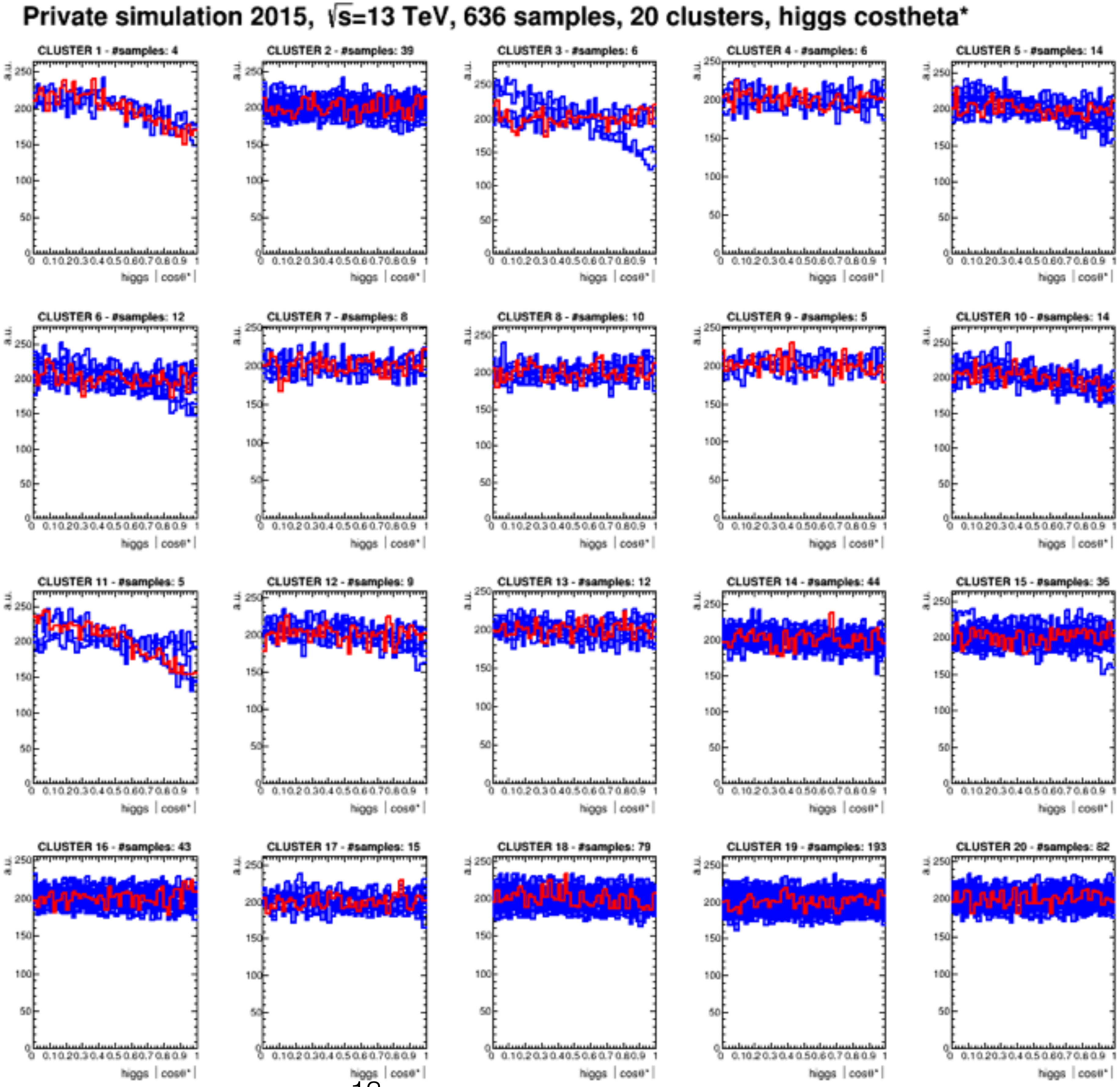


20 cluster sampling in angular differeces

As expected the higgs pt leads the clustering definition.

The longitudinal higgs momenta is mostly - but not completely - homogeneous

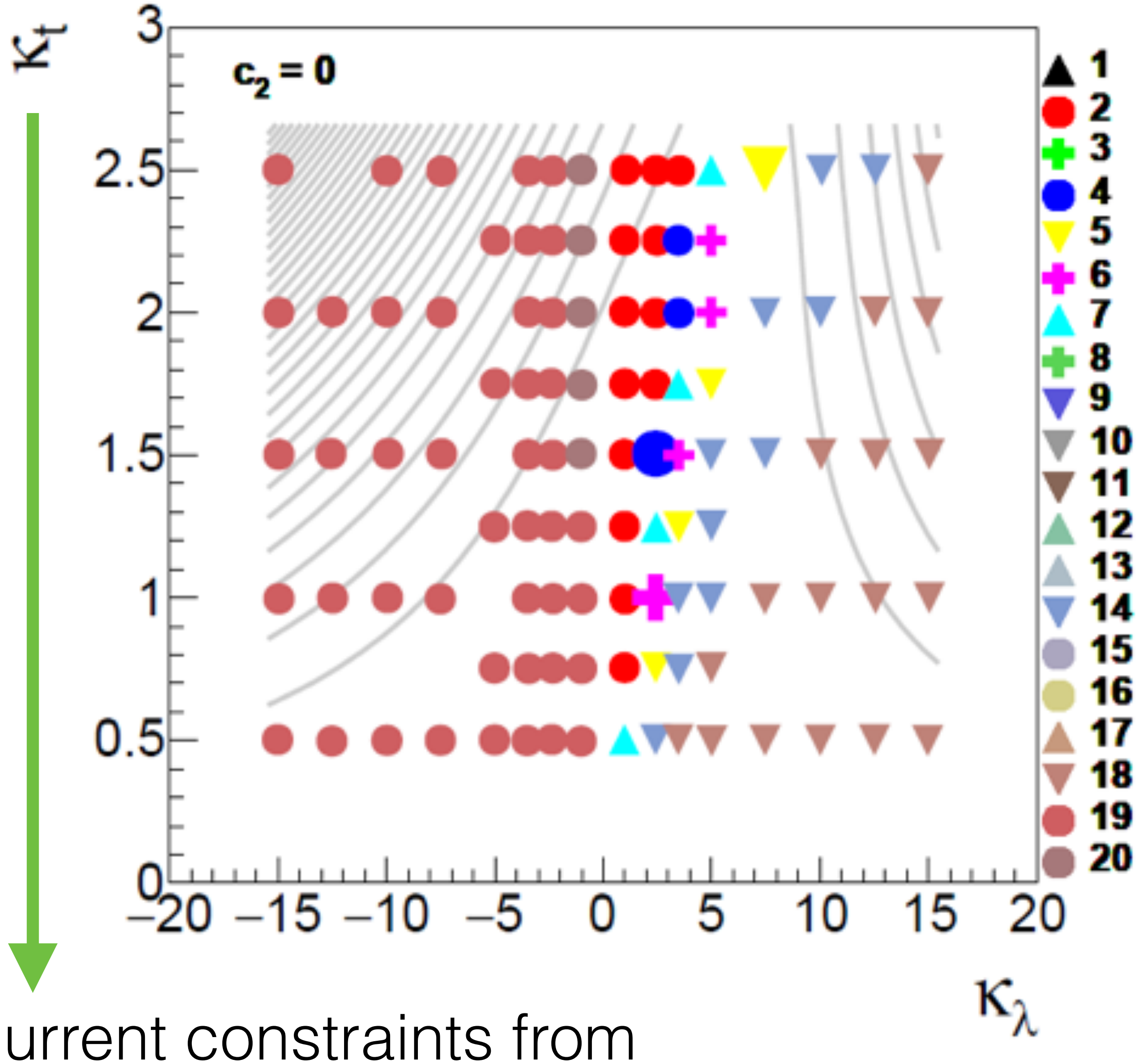
What is exemplified in the homogeneity of CS cos_theta* angle between the higgses



Mapping of the kinematic nodes in slices of parameter space

Temperature plot in the SM-like parameters plane

Gray lines = isolines of cross section



Current constraints from single higgs marginalizing

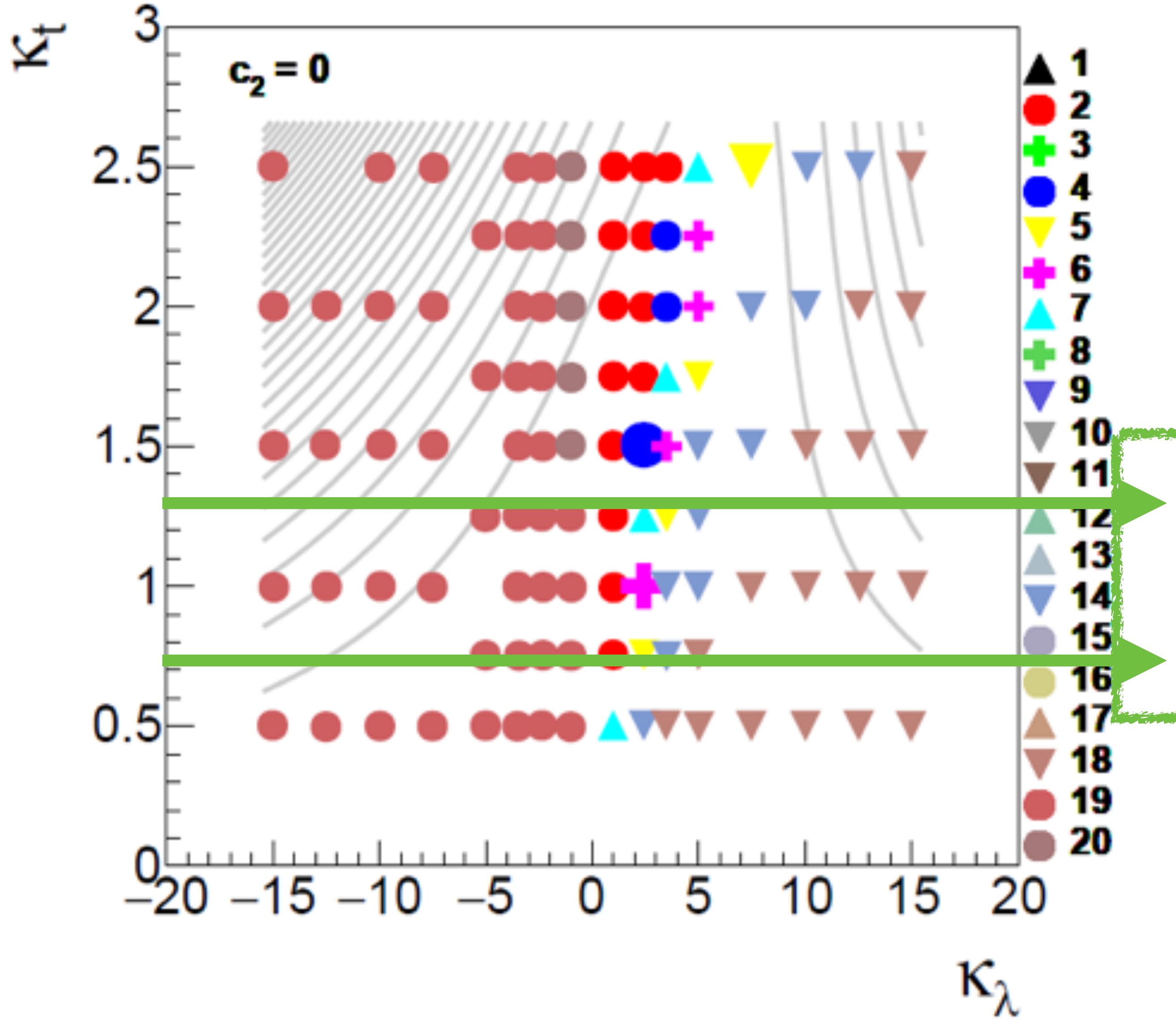
Geometric code:

- ▲ higgs pt peaks ~ >150 GeV
- ● higgs pt averages ~ 100 GeV
- ▼ higgs pt bellow ~ < 50 GeV
- ✚ higgs pt have two peaks

Bigger points are the benchmarks

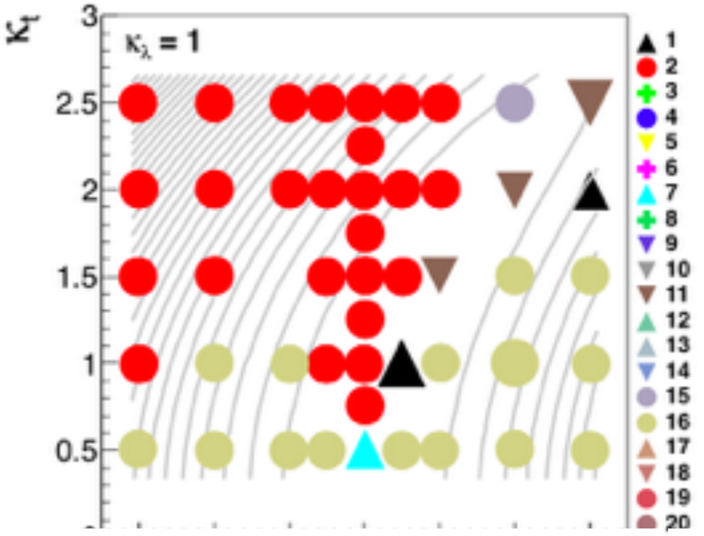
Map back clusters in slices of parameter space SM-like parameters

Regions of minimal cross section are related with bigger kinematic changes, the contrary however is not true

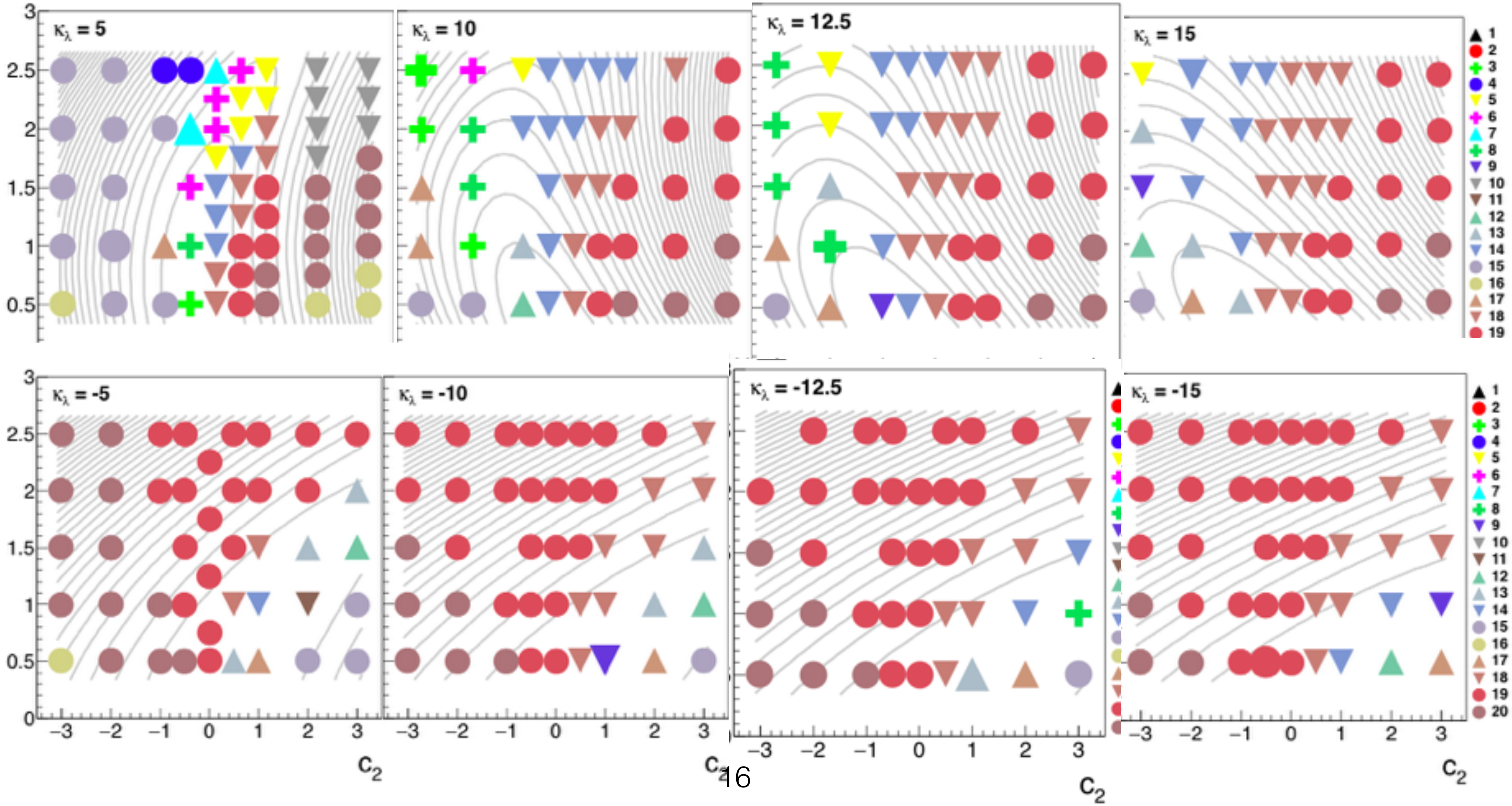


30% change in kt near to the SM hypothesis shifts the peak of the higses pt in 50 GeV - 100 GeV

Distribution of nodes in slices of kt X c2 parameters



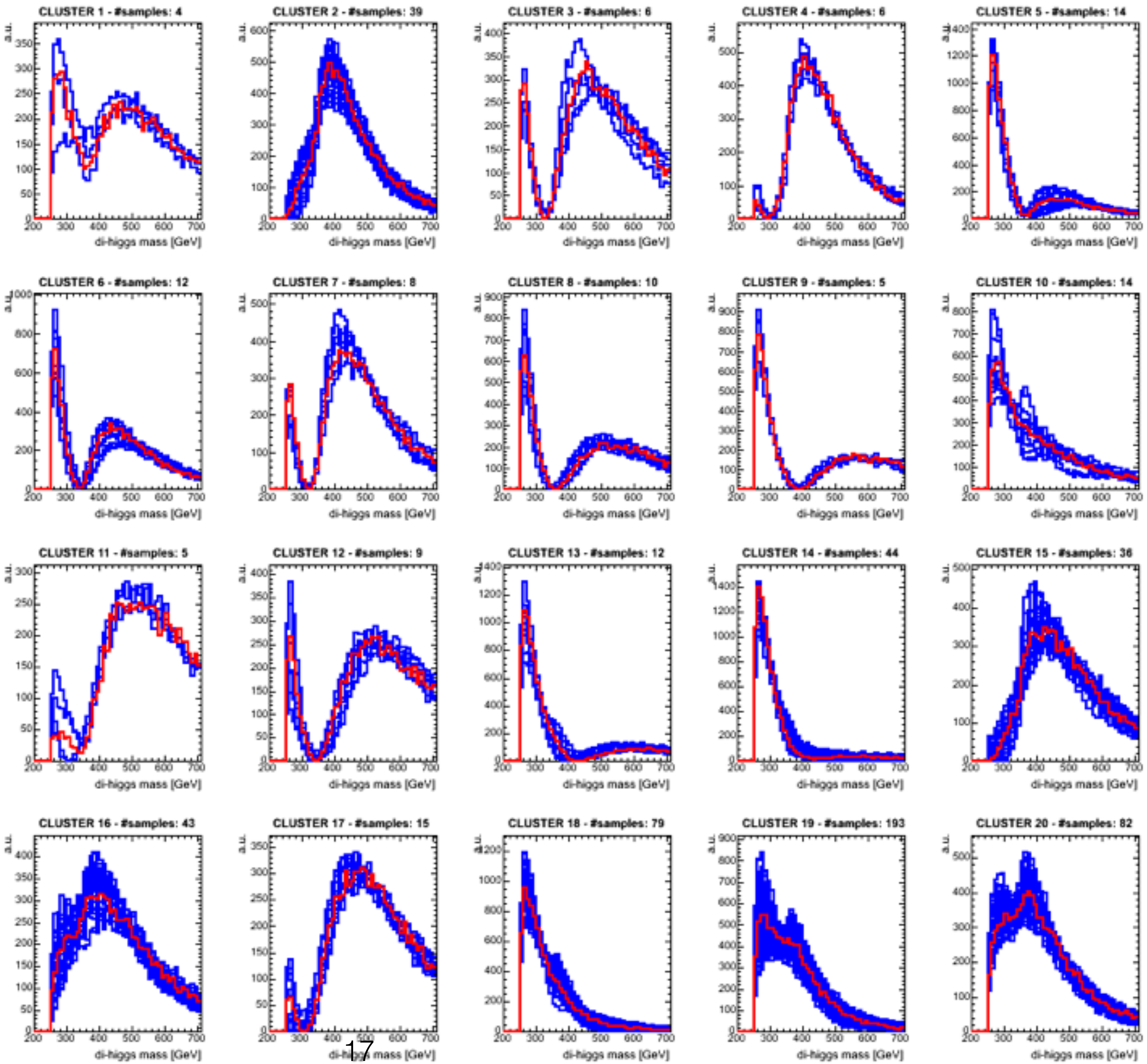
No clear asymptotic behavior in kinematics at LO when big trilinear coupling deviation



The experimental results however are still in kinematics

20 cluster sampling in the di-higgs invariant mass in the considered parameter space

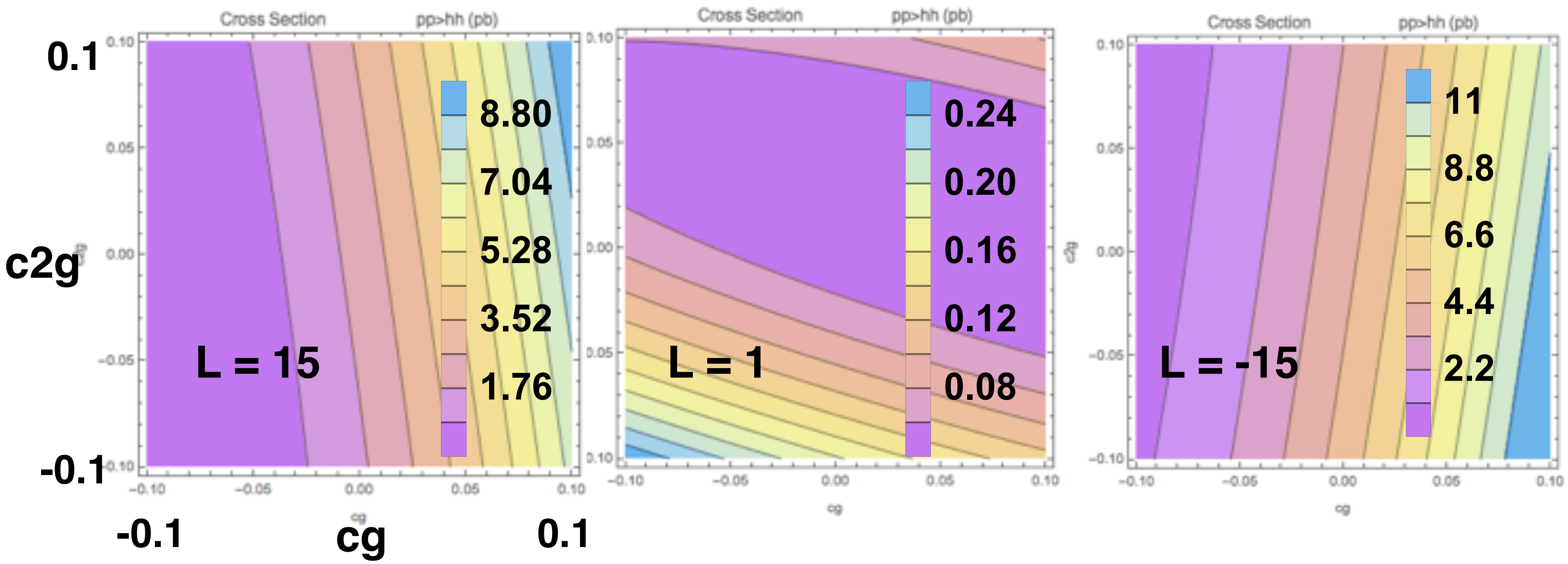
Good sampling of threshold and asymptotic mhh limits


Private simulation 2015, $\sqrt{s}$=13 TeV, 636 samples, 20 clusters, di-higgs mass

CLUSTER 1 - #samples: 4
CLUSTER 2 - #samples: 39
CLUSTER 3 - #samples: 6
CLUSTER 4 - #samples: 6
CLUSTER 5 - #samples: 14
CLUSTER 6 - #samples: 12
CLUSTER 7 - #samples: 8
CLUSTER 8 - #samples: 10
CLUSTER 9 - #samples: 5
CLUSTER 10 - #samples: 14
CLUSTER 11 - #samples: 5
CLUSTER 12 - #samples: 9
CLUSTER 13 - #samples: 12
CLUSTER 14 - #samples: 44
CLUSTER 15 - #samples: 36
CLUSTER 16 - #samples: 43
CLUSTER 17 - #samples: 15
CLUSTER 18 - #samples: 79
CLUSTER 19 - #samples: 193
CLUSTER 20 - #samples: 82

Work ongoing in extension to 5D parameters

Cross section hints why we should keep cg and c2g independently in benchmarks construction



Kt = 1, c2 = 0

14 TeV

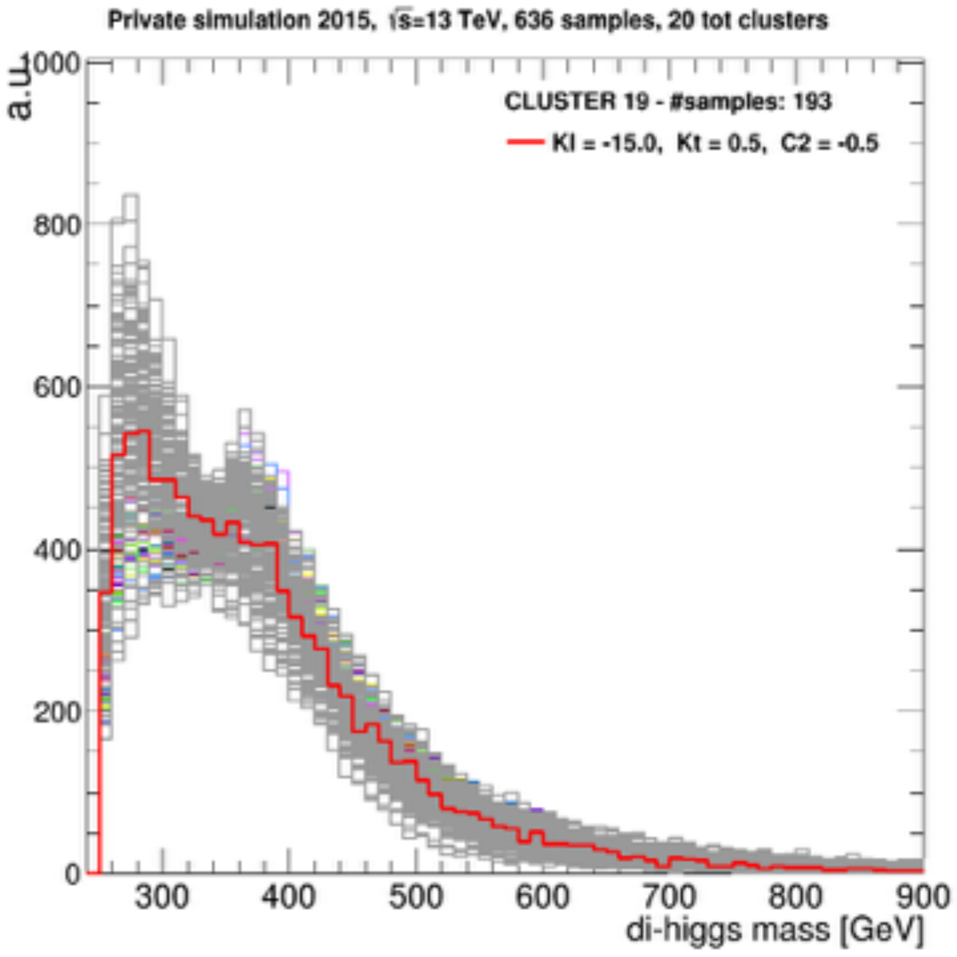
Cross section valleys are related with different interference patterns. There is no cross section flat direction in the cg = c2g hypothesis we shall expect different interference patterns to small parameter difference

Maximal kinematic sampling when keeping the parametrization general
Being more smart using gradients of cross sections a reasonable sampling can be done
O(1k) simulations points are still computationally feasible

How well LHC results could catch kinematic differences?

Only full analysis strategy can resolve it
===> Maxime's talk

Expected analysis sensitivity to small scale shape differences defines smaller quantity of clusters such that experimental limits derived based in the cluster benchmark are valid to all the elements of that node



Usages of the method / On the results presentation

It is not clear yet how QCD virtual corrections and real emissions would affect interference patterns and processes kinematics.

Experimental results based in cluster analysis will still be robust in the following cases:

- **Heaven**: QCD corrections are flat among different processes
  - Impact in shapes from loop induced corrections would make truth to be inside one cluster definition, the experimental results are out of the box in terms of wilson coefficients

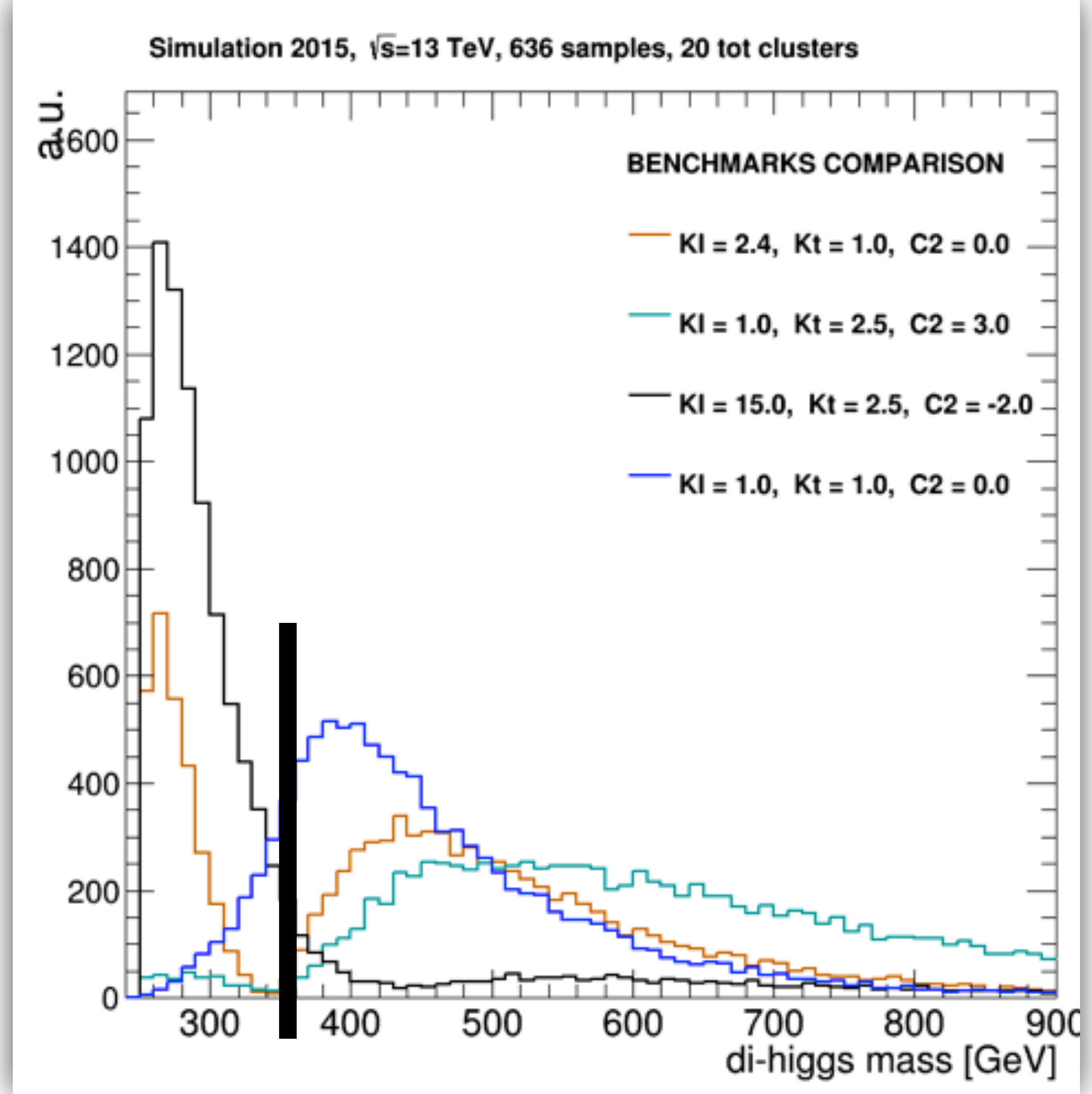
(see study ISR-only effects in the clustering on backup)

Usages of the method / On the results presentation

- **Earth**: QCD corrections are flat within different processes
  => Truth it will change interference pattern (virtual corrections) + induce global shape smearing (real emissions)
  => equivalent to navigate in the LO parameter space
  => true shape will be mapped in one of the of the other clusters

Hell: If QCD NLO completely changes shapes and sometimes it will not fit in any cluster

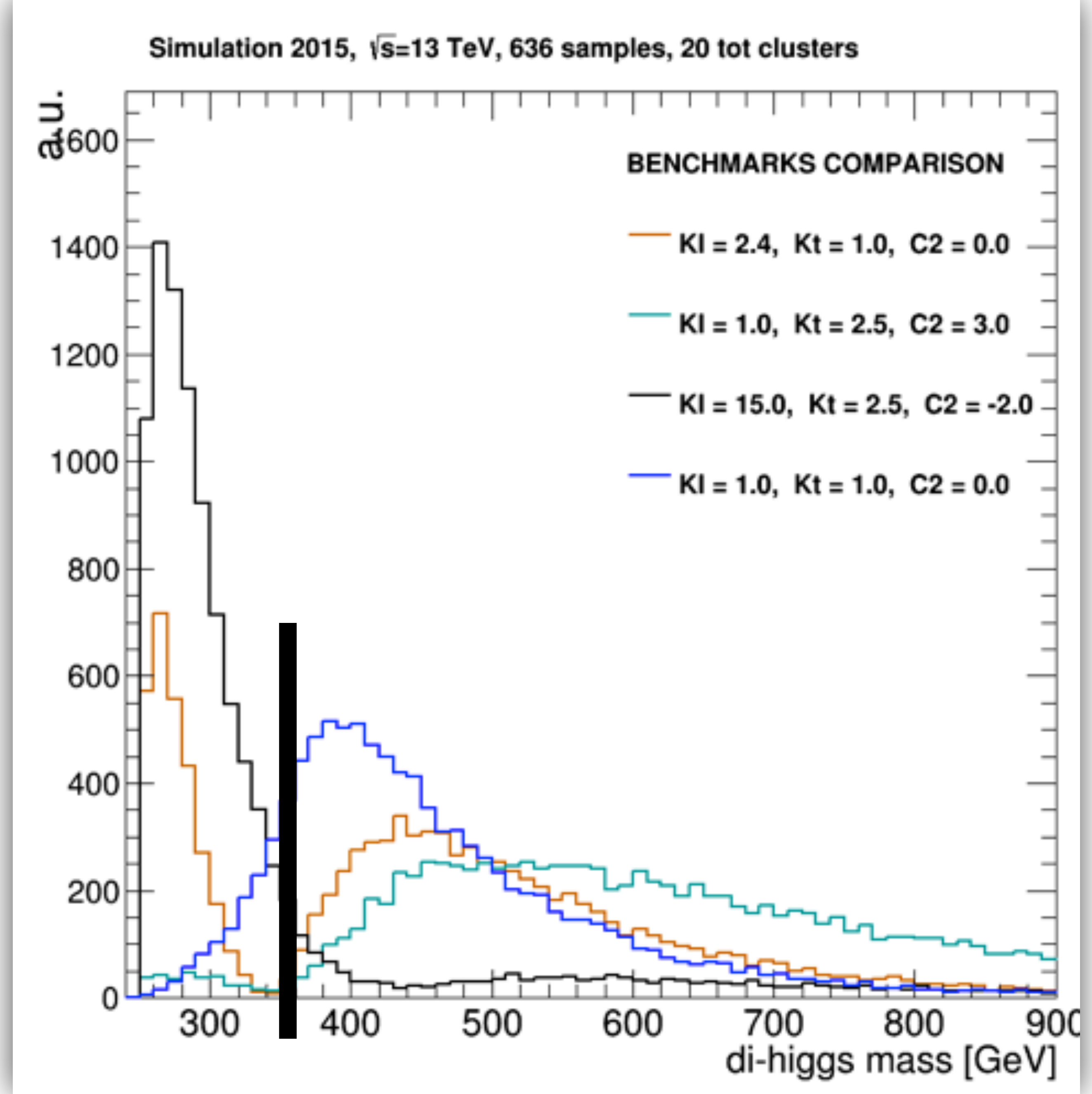
Negative search results based in kinematic classification can still be generally used if limits are also quoted in bins of Mhh



It is not out of the box to combine bins to a limit in the hole range, however it will be possible to interpret the limit from the most sensitive bin

Hell: If QCD NLO completely changes shapes and sometimes it will not fit in any cluster

Negative search results based in kinematic classification can still be generally used if limits are also quoted in bins of Mhh



We can hope low tan_beta MSSM/ NMSSM to fit or in earth, or in hell

Model builders: Does other information from experimental side would be necessary for model limit setting?
(eg correlation mhh with hh angles)

Conclusions

BSM GF hh cross sections ARE accessible to LHC13 (even LHC8)

Signal kinematic behavior is the best criteria to resume experimental results.

The feasibility to be fully differential depend in the considered final state. We would still have the problem of define the model benchmarks and its correct signal shape.

Full QCD corrected distributions to GF hh BSM physics are still in future.
(tricky business, full details can evolve in parallel with actually find the BSM-like excess)

Sampling benchmarks funded in well based theory assumptions will help experiments to take the maximum of data keeping results interpretable to a plethora of model building hypotheses

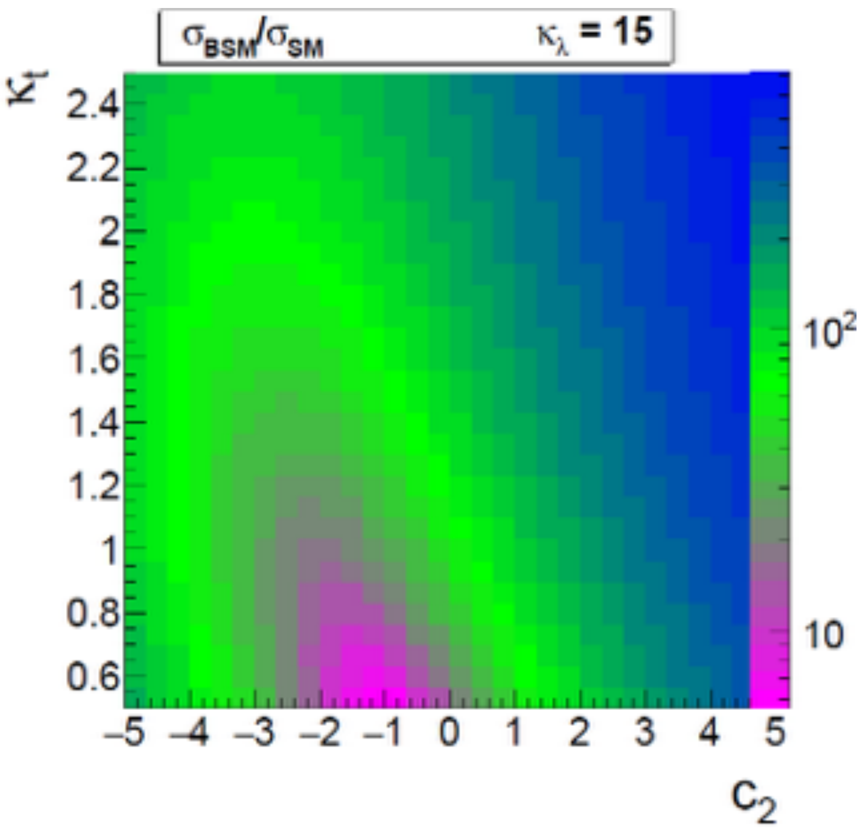
Of course synergy with the industry of QCD computations is necessary.
And feedback very welcome

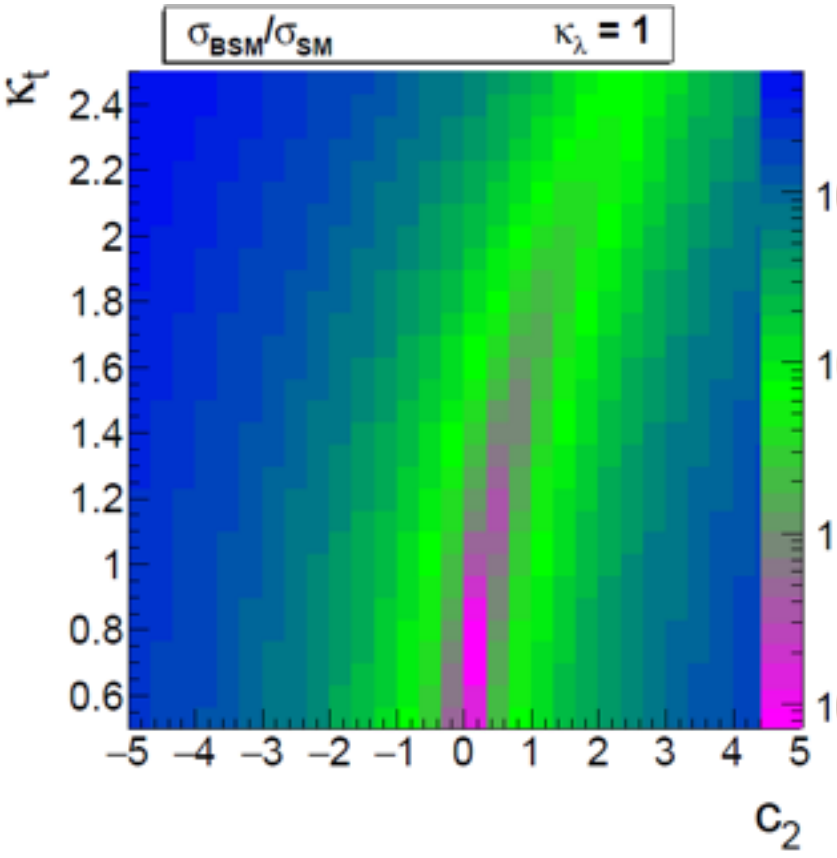
NOTE: To correspond to theoretical expectations the TS can be constructed based in mhh, cos_theta* and longitudinal boost.
The philosophy of the method and qualitative conclusions will remain the same.

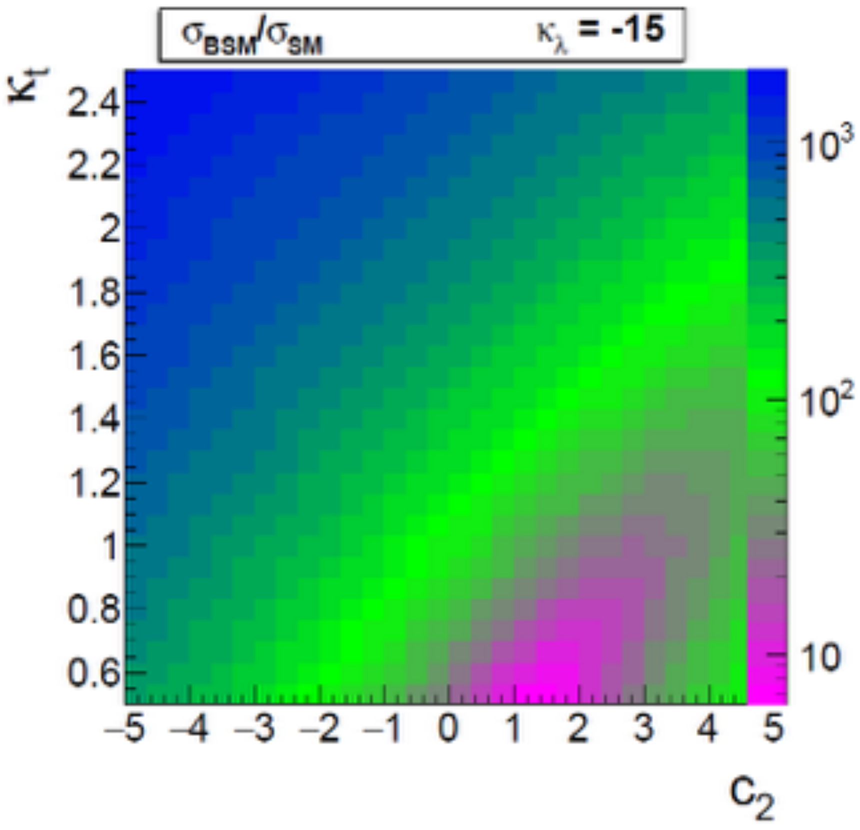
Thank you for attention

feedback very welcome!

Cross section enhancements from 3D parameter space







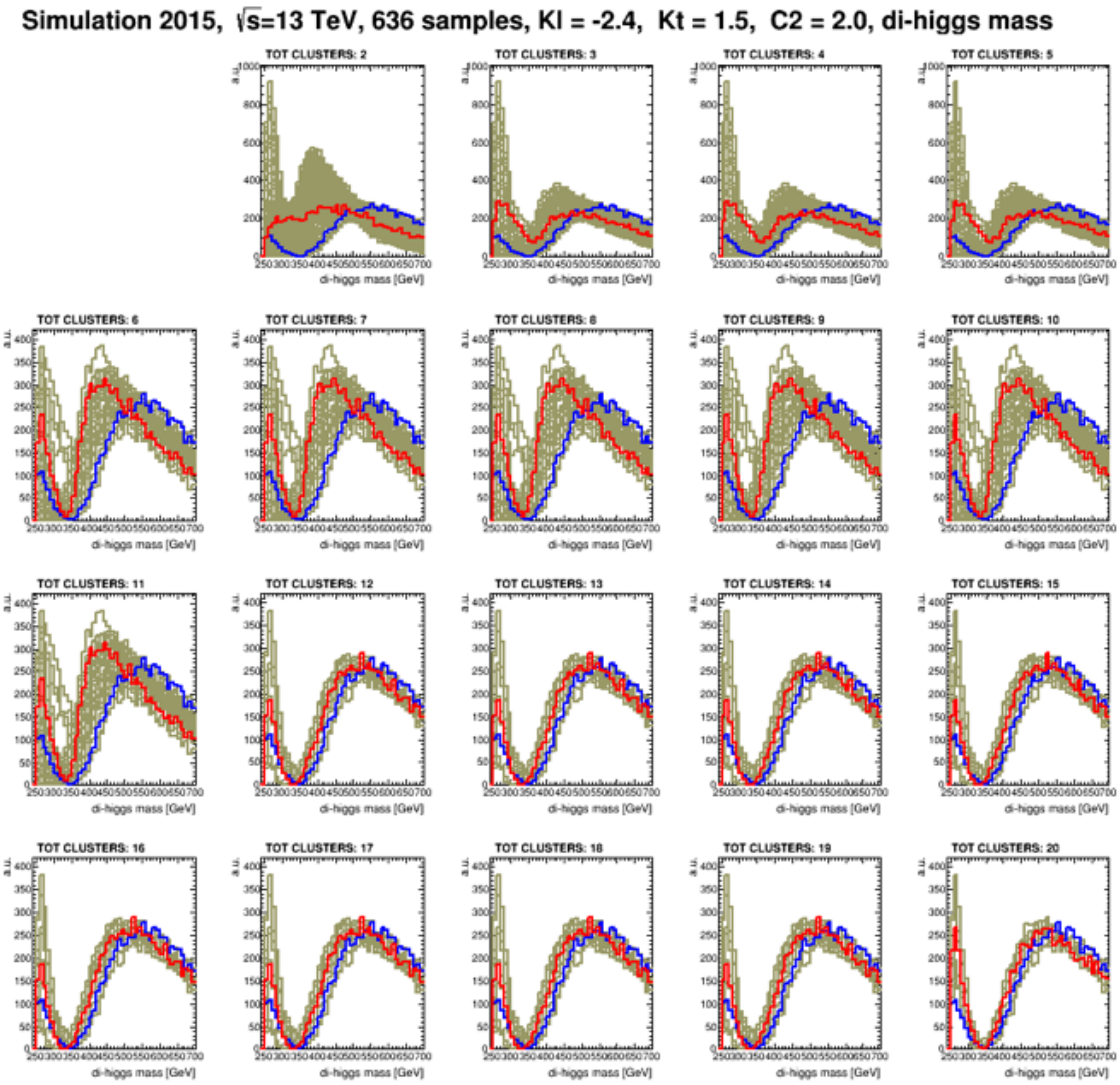
Following evolution of clusters with more weird shapes

Cluster Analysis: 13 TeV, 3 param, 636 samples.

Clusters evolution wrt selected final number of cluster.

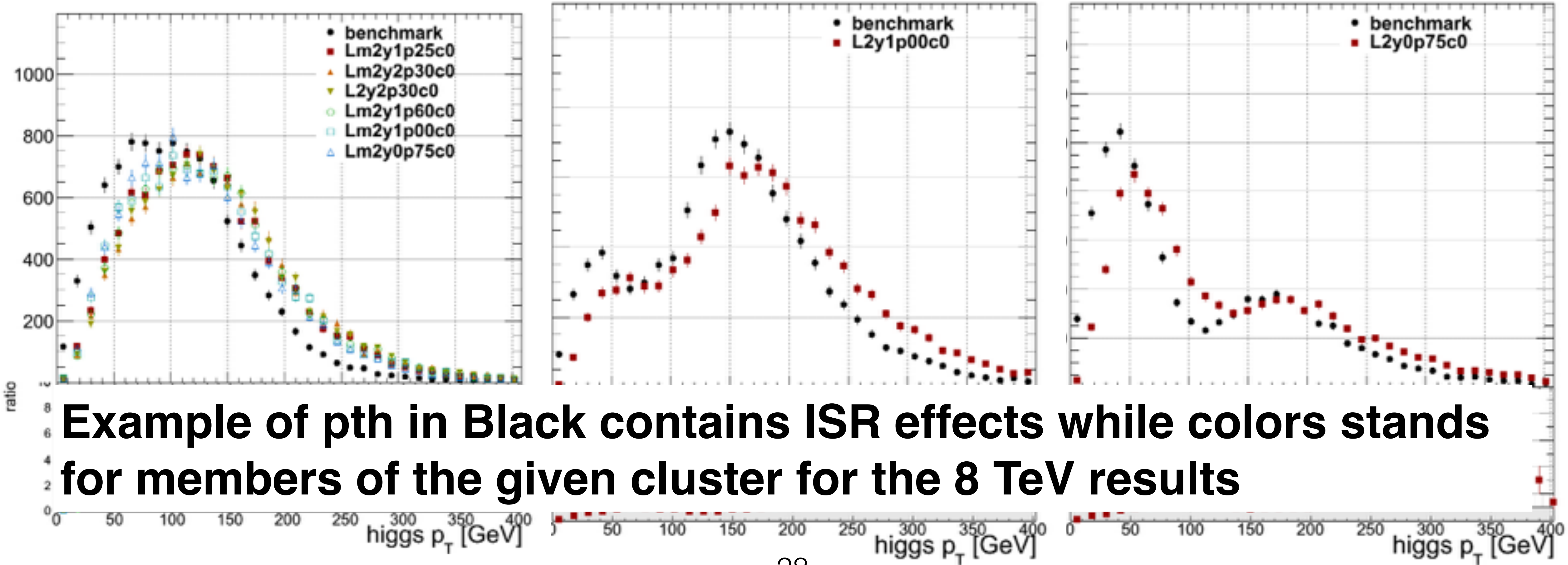
Blue → the sample: Kl = -2.4, Kt = 1.5, C2 = 2.0

Red → the benchmark



Modeling shower: ISR from LO simulation

- We test 44 random scattered points in the parameter space by comparing the kinematics before/after radiation in the clustering variables
  - The longitudinal higgs momenta holds negligible ISR impact
  - The pt of the higgs instead have visible effect.
    - We compare the lye-level with the leading pth after showering
    - The ISR impact only arrive from the initial gluons
      - Systematic effect => it could be modeled by a global function



Example of pth in Black contains ISR effects while colors stands for members of the given cluster for the 8 TeV results

Modeling shower: ISR from LO simulation

We expect a systematic effect in shape from gluons PDF, not depending on the physics inside

- If F(x) is the p_T before ISR effects, and I(x) is the ISR contribution, one can write for the after-ISR p_T the distribution:

$$G(x) = \int_0^x F(t) I(x-t) dt$$

where logNormal is an appropriate functional form to I(x)
(the logNormal describes the product of several small factors):

$$I(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2} \right)$$

The studies were done with the 8 TeV simulations

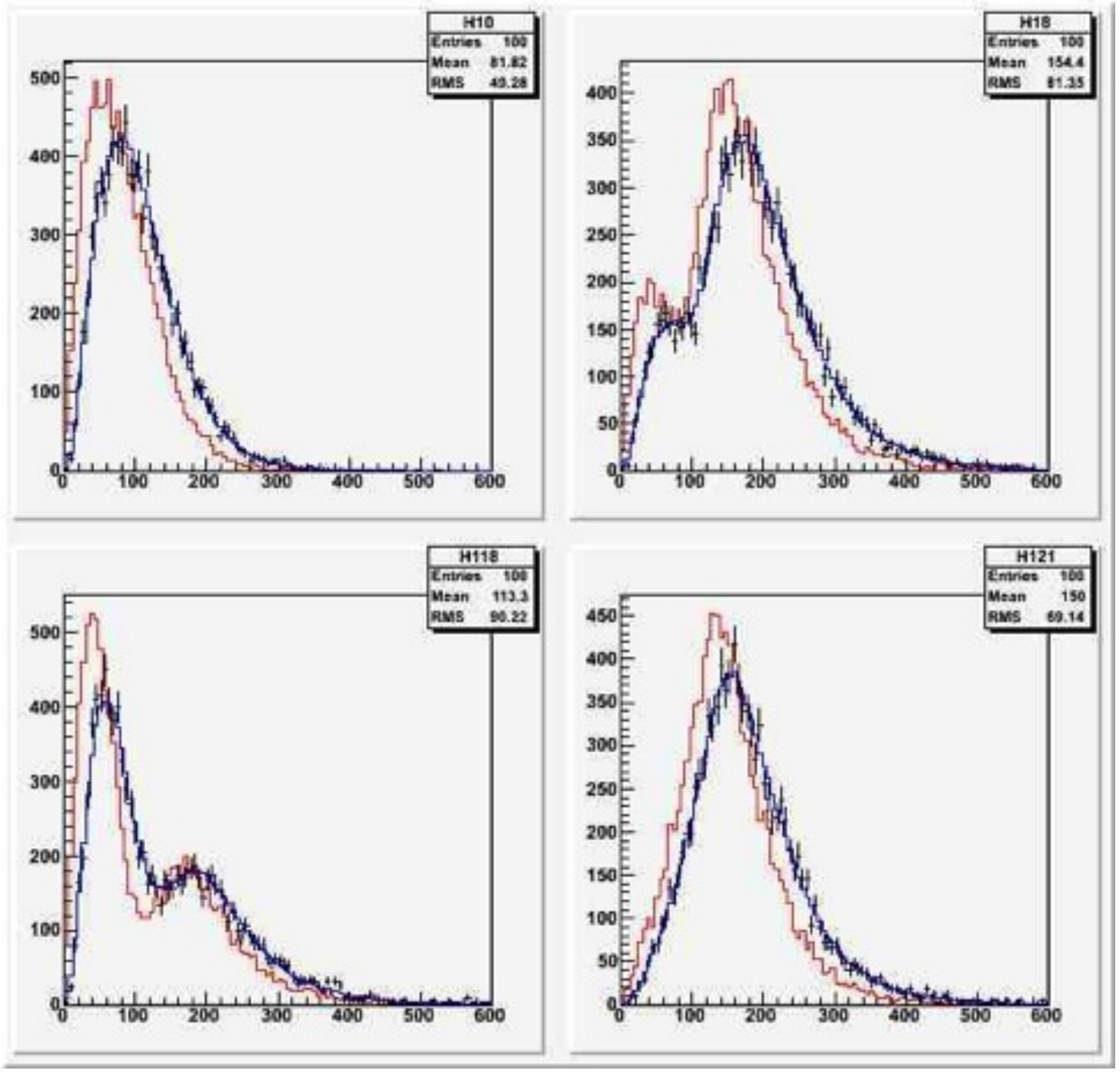
Modeling shower: ISR from LO simulation

Leading higgs pt to 4 investigated points

Red: LHE level ; **Black:** with ISR ;

Blue: result of global fi

- The fit now needs to handle 44*100 data points with three parameters
- The distributions are correctly modeled, with a chi2 of 5626 for 4397 degrees of freedom
- Next slide we show the global fit to all the 44 pth spectra

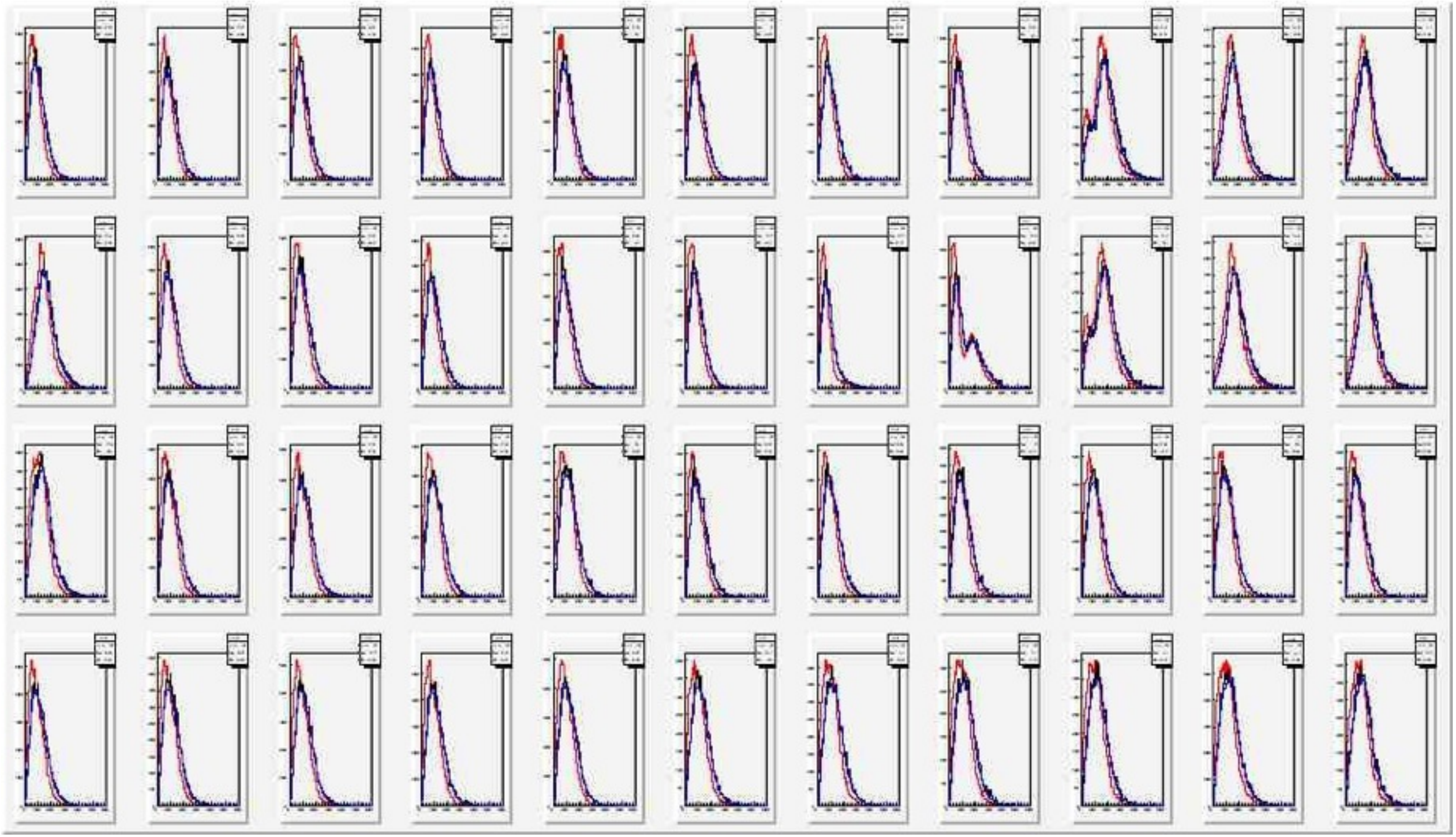


- A single logNormal distribution models reasonably well the ISR effects on the leading higgs boson p_T distribution, for different values of the anomalous coupling parameters
- This implies that the clustering can indeed be operated at LHE level

Modeling shower: ISR from LO simulation

A glimpse in the 44 points investigated

Red: LHE level ; **Black:** with ISR ; **Blue:** result of global fit

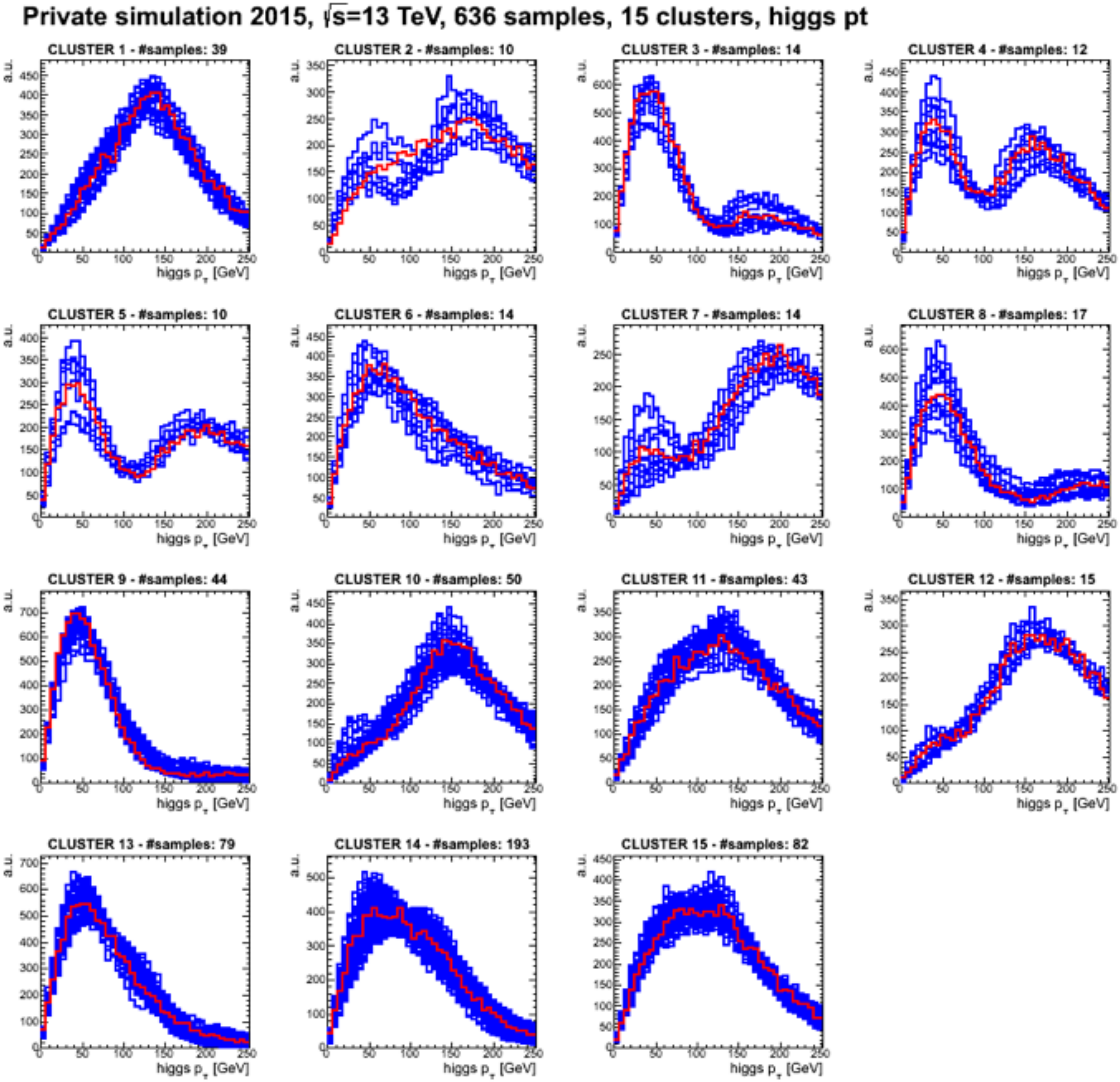


Cluster Analysis: 13 TeV, 3 param, 636 samples.

***totClus* = 15**

di-higgs mass

Red --> the benchmark



Private simulation 2015, √s=13 TeV, 636 samples, 15 clusters, higgs pt

Cluster Analysis: 13 TeV, 3 param, 636 samples.

***totClus* = 10**

di-higgs mass

Red --> the benchmark