Higgs Pair Production:
Choosing Benchmarks with Cluster Analysis

M. Dall'Osso, T. Dorigo, F. Goertz, C. A. Gottardo
A. Oliveira, M. Tosi

INFN - Padova University - CERN

101° Congresso SIF, 21/9/2015, Roma

Higgs Pair Production

Looking at the Higgs Lagrangian

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2}m_h^2 h^2 - \lambda v h^3 - \frac{\lambda}{4}h^4$$

we spot a parameter predicted by the SM but not experimentally constrained: the trilinear coupling λ.
To probe λ we need to study double-Higgs production, realized in hadronic colliders through ggF:

Very low production cross section according to SM:

$$\sigma_{hh}(8TeV)^{*} = 9.96 fb \pm 10\%$$

$$\sigma_{hh}(13TeV)^{*} = 34.3 fb \pm 10\%$$

* √s in pp collisons

Beyond Standard Model

New physics could constribute to the double Higgs production.

BSM effective theory:

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2} m_h^2 h^2 - \kappa_\lambda \lambda_{SM} v h^3 - \frac{m_t}{v}(v + \kappa_t h + \frac{c_2}{v} h h)(\bar{t_L} t_R + h.c.)$$
$$+ \frac{1}{4}\frac{\alpha_s}{3\pi v}(c_g h - \frac{c_{2g}}{2v} h h) G^{\mu\nu} G_{\mu\nu}$$

κ_λ	anomalous trilinear	$\lvert\kappa_\lambda\rvert \sim 15$ (κ_λ only variation)(2)
κ_t	anomalous top Yukawa	$\kappa_t \in [0.5, 2.5]$(1)
c_2	tthh interaction	$\lvert c_2\rvert < 5$ if $\kappa_\lambda = 1$ and $\kappa_t \in [0.5, 2.5]$(2)
c_g	h-gluon contact int.	$c_g \sim O(1)$ theoretical assumption
c_{2g}	hh-gg contact int.	$c_{2g} \sim O(1)$ theoretical assumption

5D parameter space

(1) from single h RUN I study (CMS-PAS-HIG-14-009, ATLAS-CONF-2015-007)

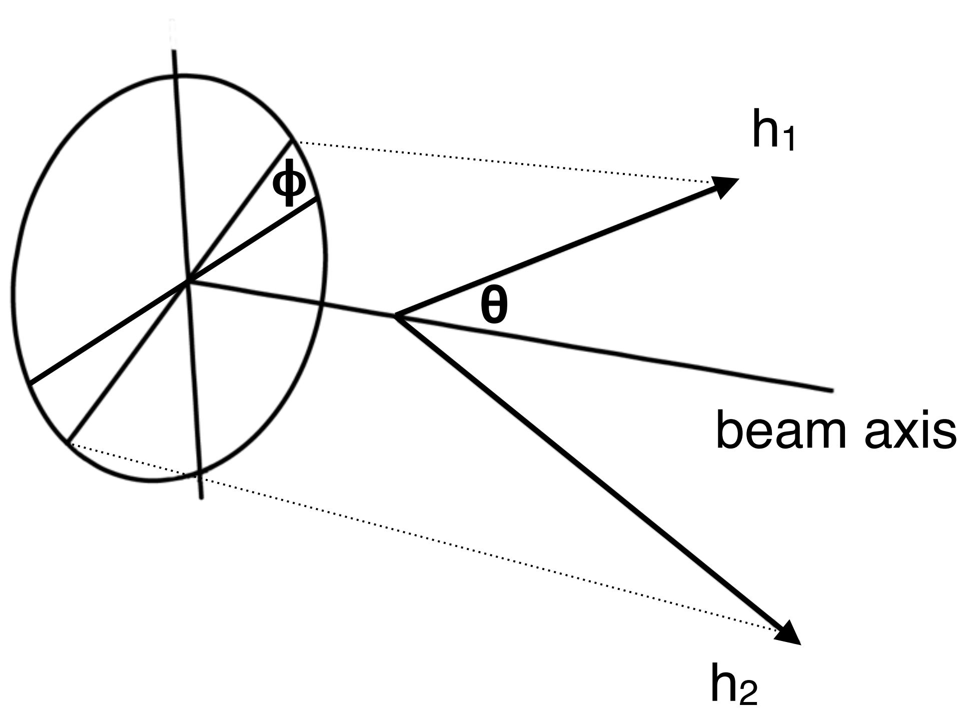
(2) from hh → γγbb 8 TeV analysis by ATLAS and CMS, and hh → bbbb 8 TeV ATLAS

Deviations from SM values or new couplings could **enhance the cross section** up to >100 times but also **change drastically the kinematics** requiring a custom analysis for each set of values.

Setup

Aim Cluster together parameter space regions that share the same kinematics, making the probing of the whole parameter space possible with a few analyses.

Variables choice



The bosons are back-to-back in ϕ (no ISR), so, disregarding of the particular azimuthal angle, we just need 3 variables to describe the system

$$p_T, \quad p_{z,1}, \quad p_{z,2}$$

Actually the boost along the z axis comes from the parton distribution functions we do not want to account for. So we study the process in the center of mass frame with just two variables

$$m_{hh}, \quad \cos\theta^*$$

Parameter space point → Monte Carlo sample → 2D shape

Binning sufficiently populated 50 (m_{hh}) x 5 ($|\cos\theta^*|$) bins.

m_{hh} [0, 1500 GeV] 30 GeV wide-bin
$|\cos\theta^*|$ [0,1] 0.2 wide-bin.

Test Statistic I

Test Statistic (TS)

Several possible choices: Kolmorov-Smirnov, Anderson-Darling, Zach-Aslan energy test….
Final choice: log likelihood function based on Poisson counts.

The maximum likelihood of sample 1 and 2 sharing the same parent distribution is given, for the i-th bin by

$$p\left(n_{i,1}, n_{i,2} \mid \hat{\mu}_i = \frac{n_{i,1}+n_{i,2}}{2}\right) = e^{-2\mu_i}\ \frac{\mu_i^{n_{i,1}+n_{i,2}}}{n_{i,1}!\ n_{i,2}!}$$

$\hat{\mu}_i$ is chosen as the minimum-variance unbiased estimator

The TS is defined as the log-likelihood:

$$TS = log(\mathcal{L}_{12}) = \sum_{i=1}^{N_{bin}^{tot}} [-2\mu_i + (n_{i,1}+n_{i,2}) log\mu_i - log(n_{i,1}!) - log(n_{i,2}!)]$$

Compared to Z.A. test it was proved to be more sensible to small scale features of the distributions under test.

Test Statistic II

Steps:

1) Identify each sample as one element cluster

2) define cluster-to-cluster similarity as $TS^{min} = min(TS_{ij})$ where i runs on first cluster elements and j on the second one

3) merge the pair of clusters with highest TS^{min}

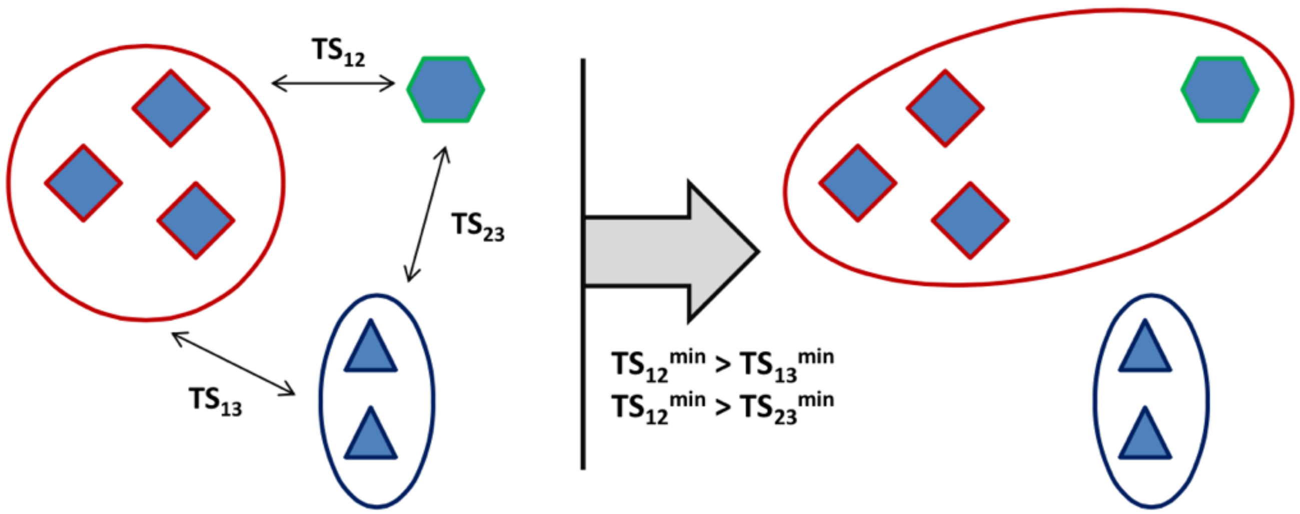
4) repeat until the desired numer of clusters $\mathbf{N_{clus}}$ is reached

5) identify the benchmark k of each cluster as the one with the highest

$$TS_k^{min} = \min_i(TS_{ki})$$

where i runs on the cluster elements.

$TS_{ij} > TS_{kl}$ ⟶ i and j are more similar to each other then kl

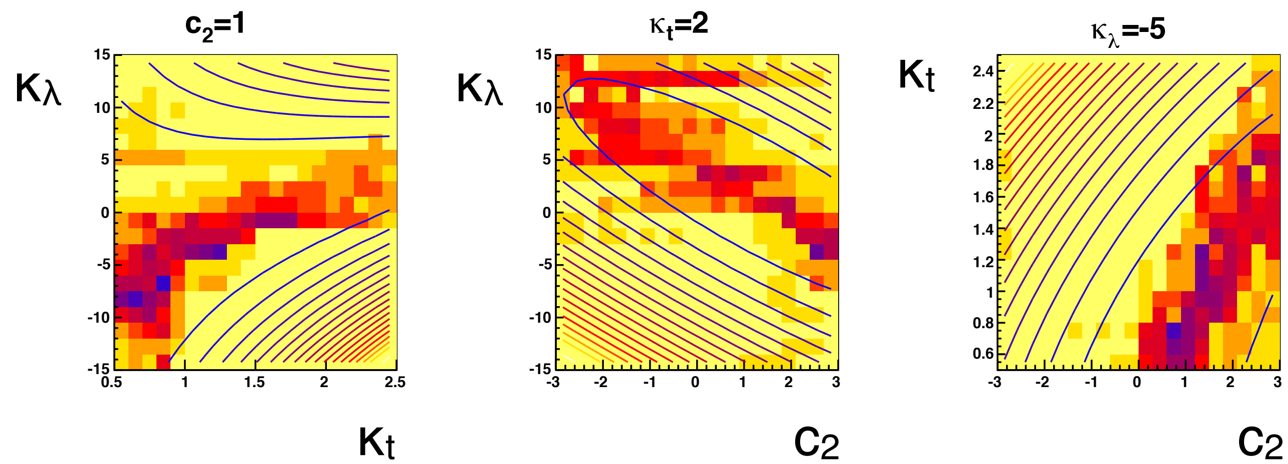


<u>$\mathbf{N_{clus}}$ is the only free parameter, fixed a posteriori.</u>

Application of the algorithm

1483 Monte Carlo initial samples (20k events LHE, 13 TeV, MadGraph5 + aMC@NLO)

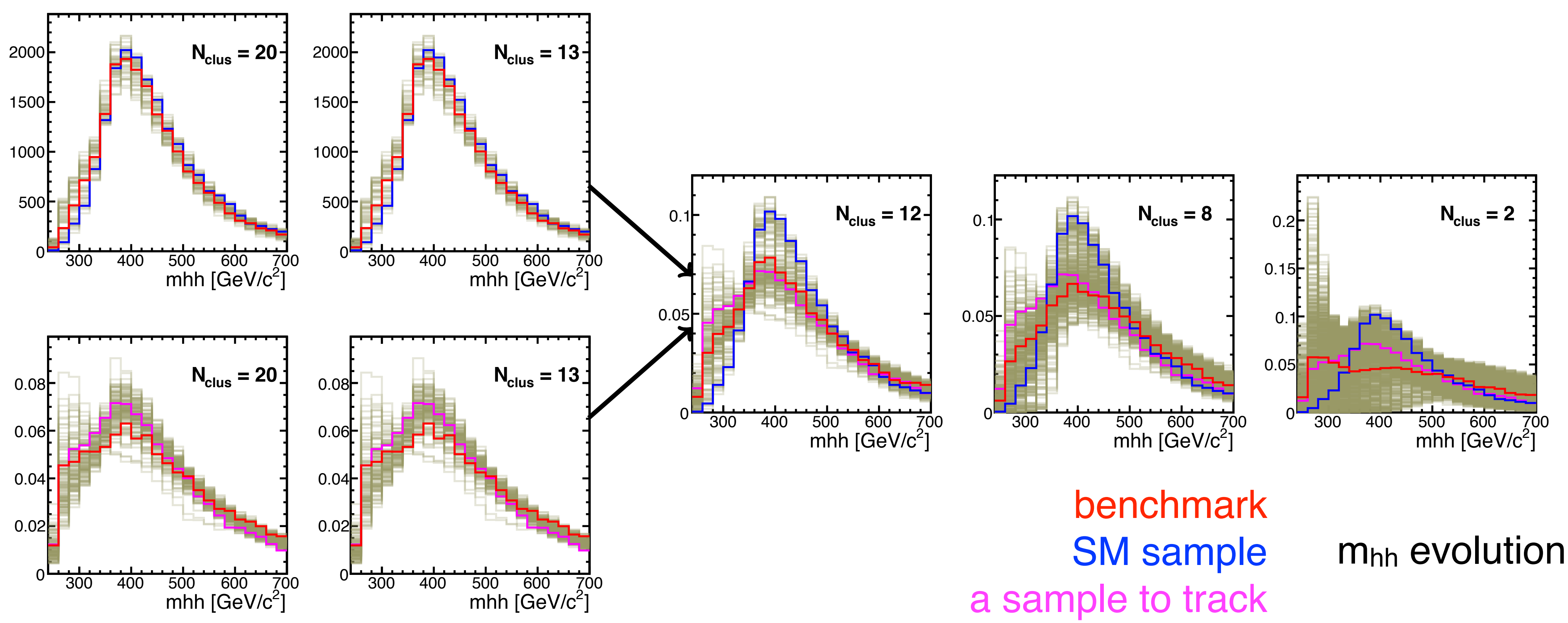
- corresponding to different points in the par. space,
- near local minima of cross section where maximum kinematical variability shows up
- no generation cut



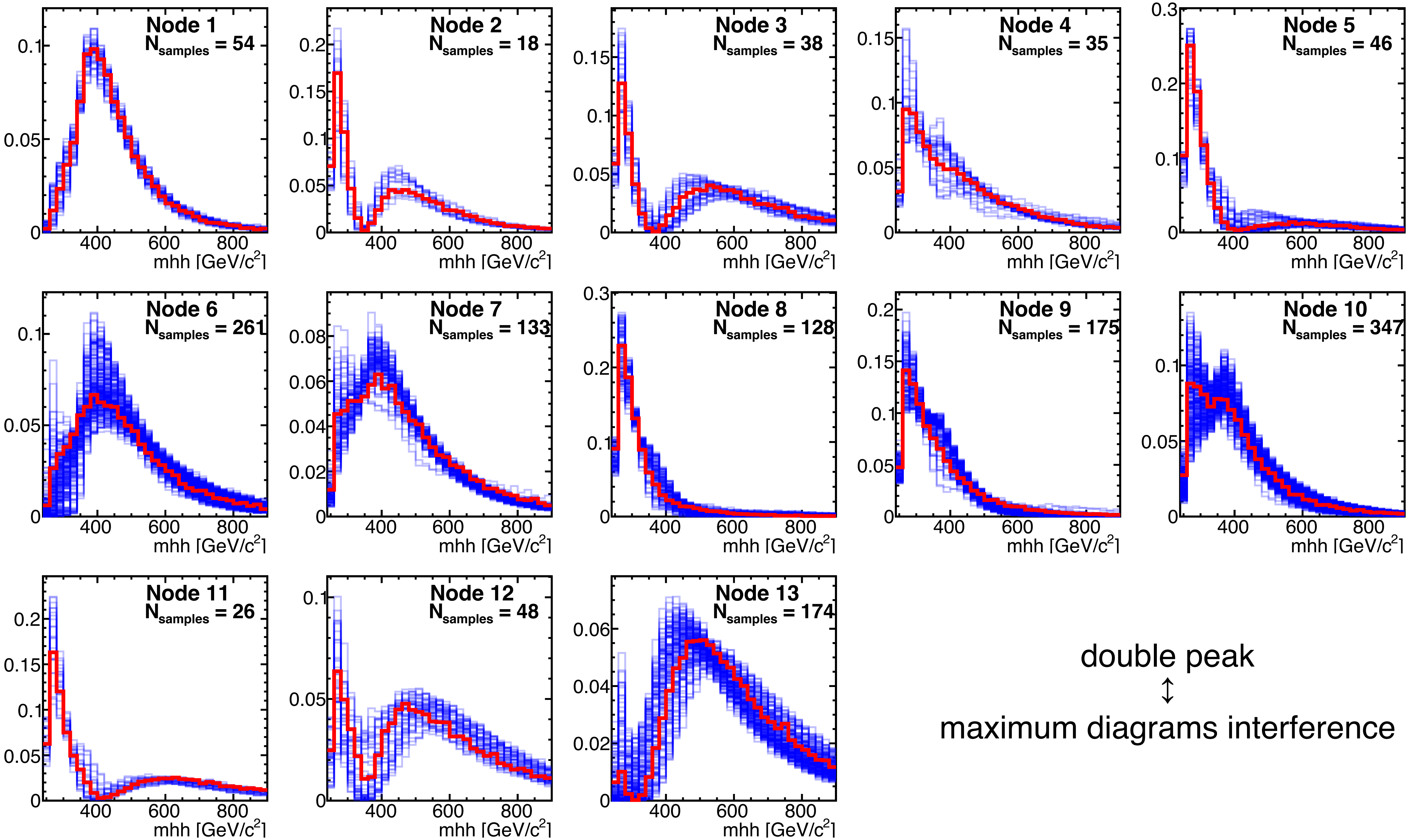
cross section isolines and TS variation speed (interpolation)

13 benchmarks

Best N_{clus} evaluated a posteriori equals 13 ↔ homogeneity - numerosity trade-off

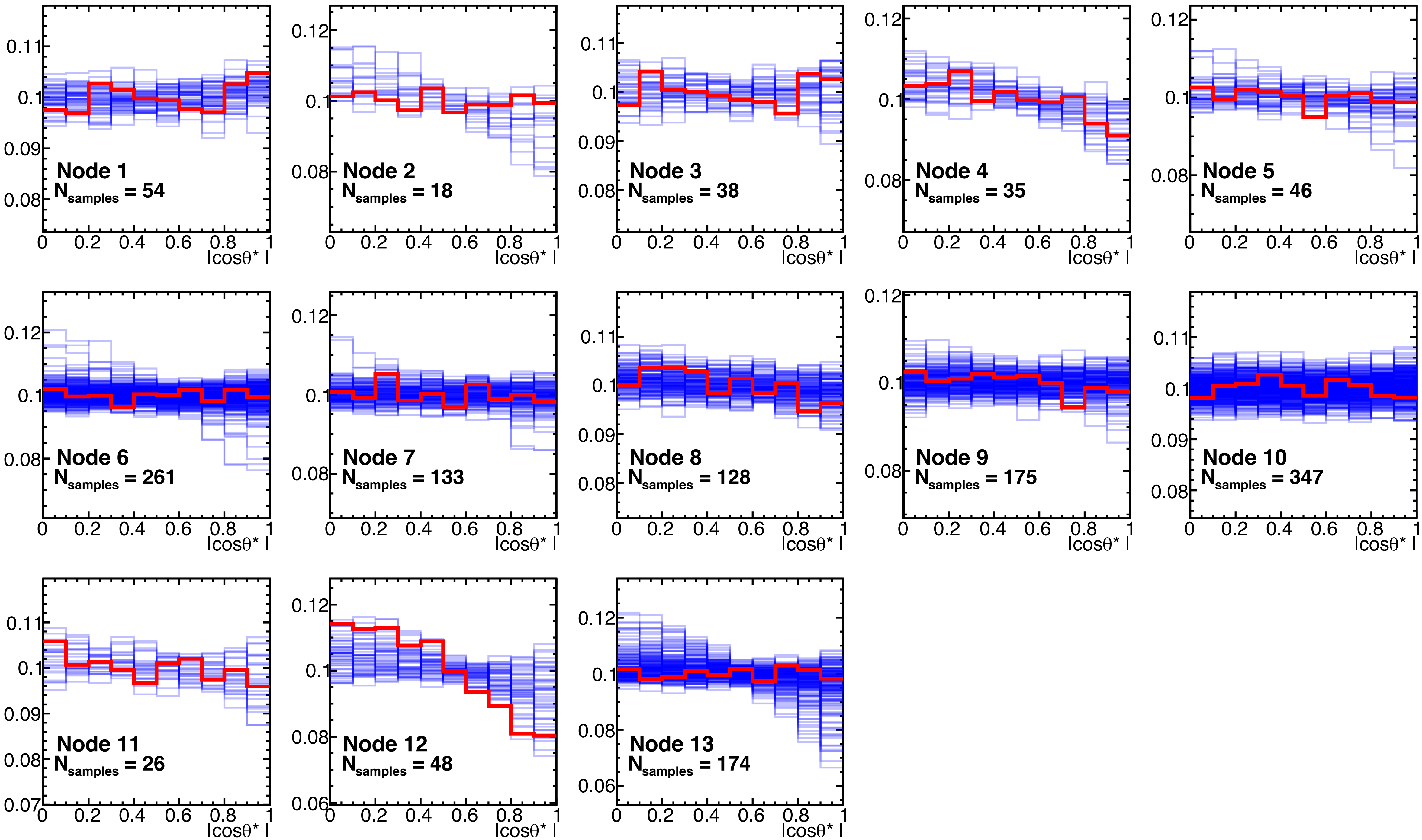


Clusters m_{hh} distributions

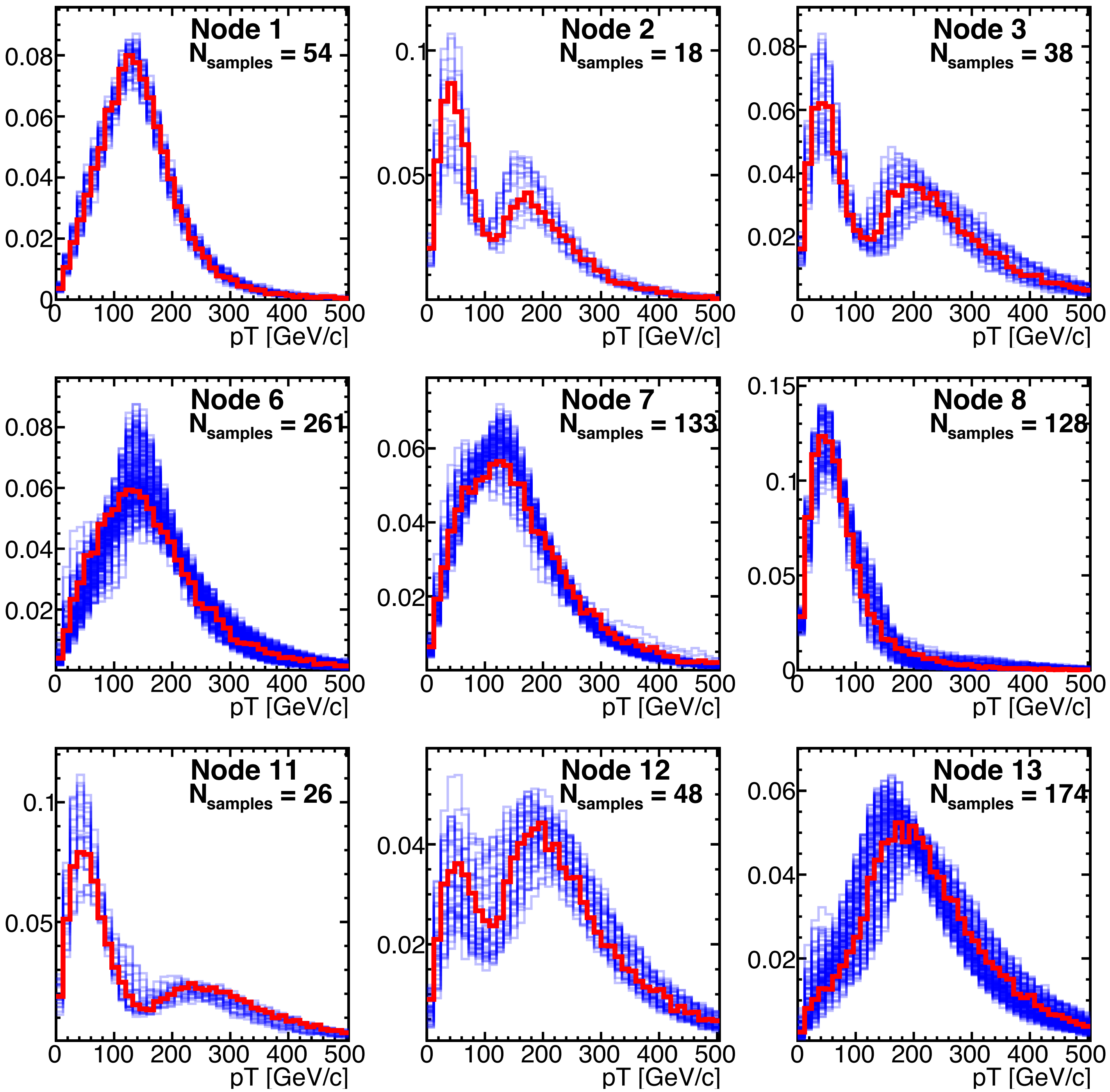


double peak
↕
maximum diagrams interference

Clusters |cosθ*| distributions

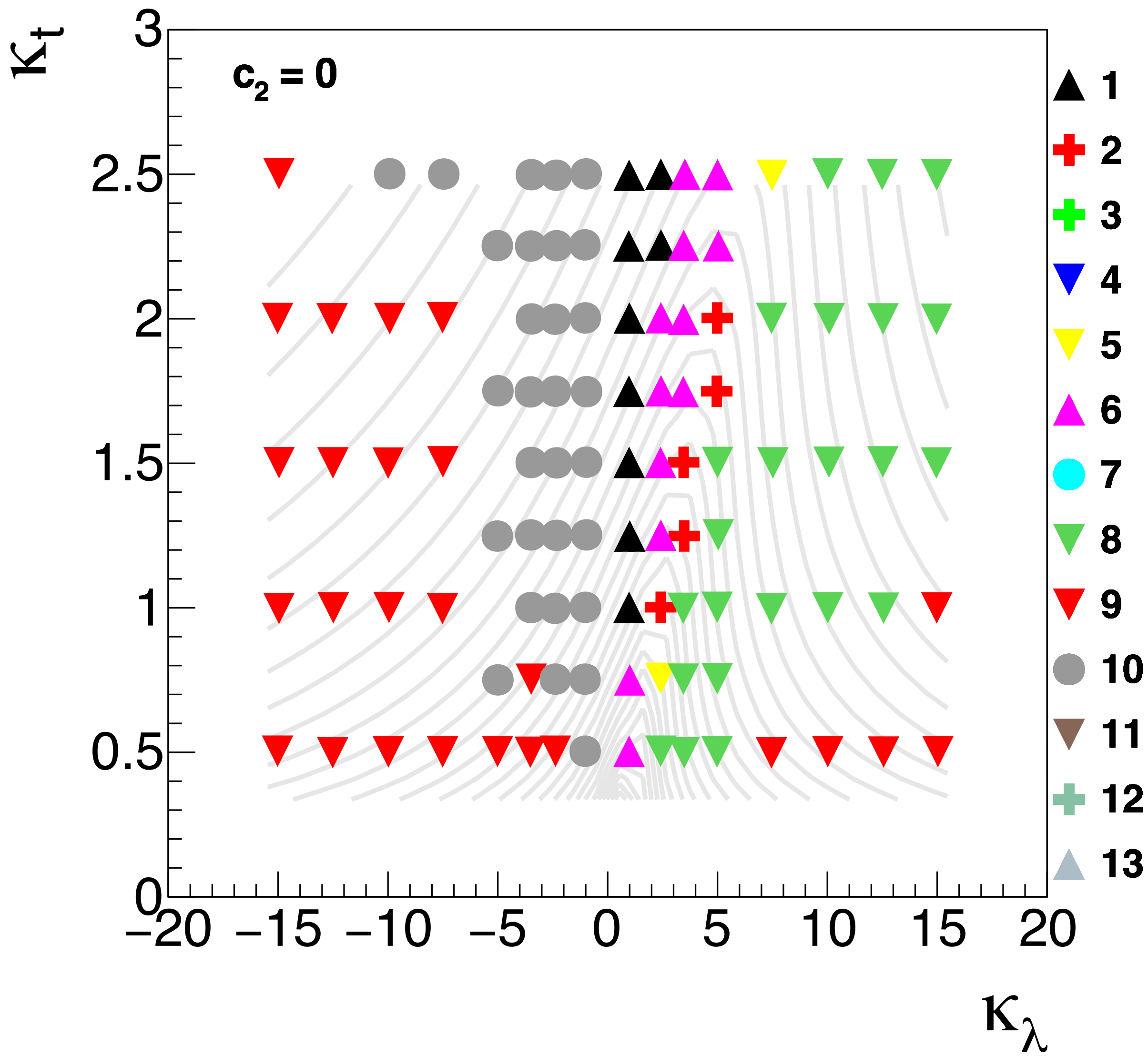


Clusters p_T distribution



Validation variable,
not used by clustering
similar shapes are clustered together.

Clusters map in $\kappa_t \times \kappa_\lambda$ plane



samples in κ_t and κ_λ plane, one color per cluster.

- ▼ p_T peak < 50 GeV
- ● p_T peak ~ 100 GeV
- ▲ p_T peak > 150 GeV
- ✚ double peak in m_{hh}

great variability around SM point which is a cross section minimum

Conclusions

- Higgs pair production is under study by ATLAS and CMS collaboration [CMS][ATLAS]
- An EFT parametrization of the gg→hh process has been provided and brings to a five dimensional parameter space
- The parameter space is wide and only a limited number of analyses can be performed
- **A clustering tecnique has been developed and lead to a subdivision of the parameter space into 13 regions**
- Good uniformity of kinematical distributions in each cluster validates the method
- This uniformity has also been checked at the reco level at least for the γγbb final state at 8 TeV (link)

Work documented in arXiv:1507.02245