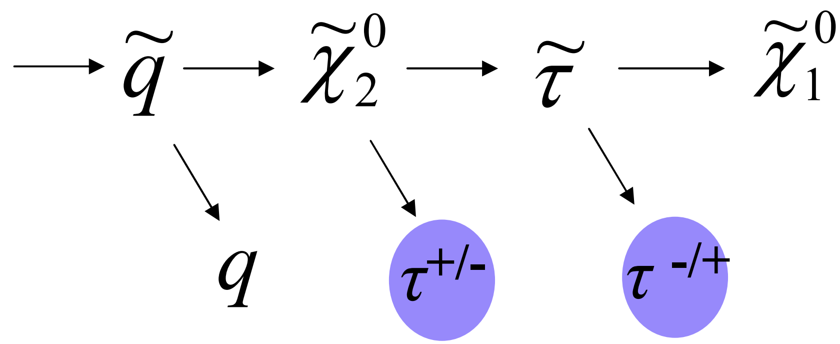
April 29 2008

Inclusive Search for mSUGRA events using MET + hadronic jets + di-τ's signature with the CMS detector

HOUCHU Ludovic, GOERLACH Ulrich (IPHC, Strasbourg)

- Observing SUSY by the appearance of an excess of tau cand. pairs as a function of a kinematical variable ; 2 methods are described.

- Motivation for this type of signature:



- 2 test points considered in mSUGRA parameter space:

 - "*<u>L</u>ow <u>M</u>ass <u>2</u>*" ($m_{1/2}$=350GeV/c^2, m_0=185GeV/c^2, A_0=0, tanβ=35, sign(μ)=+),
 BR($\tilde{\chi}_2^0 \to \tilde{\tau}\tau$) ≈ 0.96,
 24.6% of the mSUGRA events contain at least 1 above cascade,

 - "*<u>L</u>ow <u>M</u>ass <u>1</u>*" ($m_{1/2}$=250GeV/c^2, m_0=60GeV/c^2, A_0=0, tanβ=10, sign(μ)=+),
 BR($\tilde{\chi}_2^0 \to \tilde{\tau}\tau$) ≈ 0.46,
 15.0% of the mSUGRA events contain at least 1 above cascade.

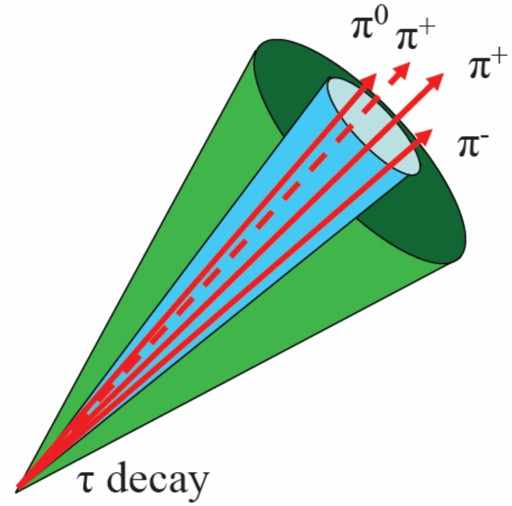
- A previous study of these events by D. J. Mangeol and U. Goerlach in CMS NOTE 2006/096

Events considered (simul'ed with OSCAR and reco'ed with ORCA)

physics process (+3.5 OR 5 on average pile-up events per bunch crossing)		expected LO σ (pb)	# events used	normalization factor for $\int L dt = 10 fb^{-1}$
mSUGRA *LM2*	all	7.38	74K	**1.00**
	containing $\tilde{q}$ cascade	1.82	18K	**1.00**
mSUGRA *LM1*	all	49.00	110K	4.43
	containing $\tilde{q}$ cascade	6.77	15K	4.43
QCD	80GeV/c < p_T < 120GeV/c	$2.96\ 10^6$	111K	$265.89\ 10^3$
small samples → high event weights	120GeV/c < p_T < 170GeV/c	$497.50\ 10^3$	93K	$53.46\ 10^3$
	170GeV/c < p_T < 230GeV/c	$100.20\ 10^3$	213K	$4.70\ 10^3$
	230GeV/c < p_T < 300GeV/c	$23.80\ 10^3$	242K	983.82
	300GeV/c < p_T < 380GeV/c	$6.39\ 10^3$	171K	374.53
	380GeV/c < p_T < 470GeV/c	$1.89\ 10^3$	142K	133.48
	470GeV/c < p_T < 600GeV/c	690.00	140K	49.17
	600GeV/c < p_T < 800GeV/c	202.00	60K	33.79
	800GeV/c < p_T < 1000GeV/c	35.70	64K	5.55
ttbar incl.		492.00	581K	8.46
single *t* incl.		259.00	78K	33.29
W+jet(s)	75GeV/c < p_T < 125GeV/c	945.00	55K	170.79
	125GeV/c < p_T < 200GeV/c	215.00	78K	27.44
	200GeV/c < p_T < 350GeV/c	43.80	80K	5.44
	350GeV/c < p_T < 2200GeV/c	4.90	110K	0.44
WW+jet(s)		188.00	235K	7.99
Z+jet(s)	75GeV/c < p_T < 125GeV/c	125.00	53K	23.42
	125GeV/c < p_T < 200GeV/c	27.00	82K	3.29
	200GeV/c < p_T < 350GeV/c	5.40	52K	1.04
	350GeV/c < p_T < 2200GeV/c	0.70	52K	0.13

●hadr. τ-jet particularities which are used to distinguish tau-jet from a q/g-jet :

- narrowness,
- low # charged particles,
- low # neutral clusters visible in ECAL,
- low proportion of neutral E relative to tracks ,
- non-negligible τ flight path.



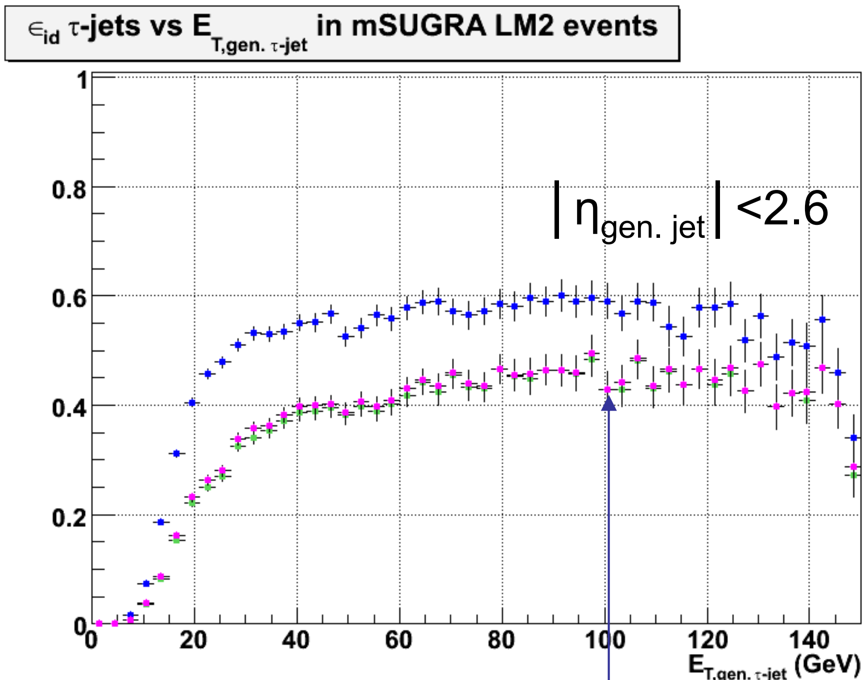
●hadr. τ-jet / q/g-jet discrimination developed within ORCA (used here) and modified slightly within CMSSW :

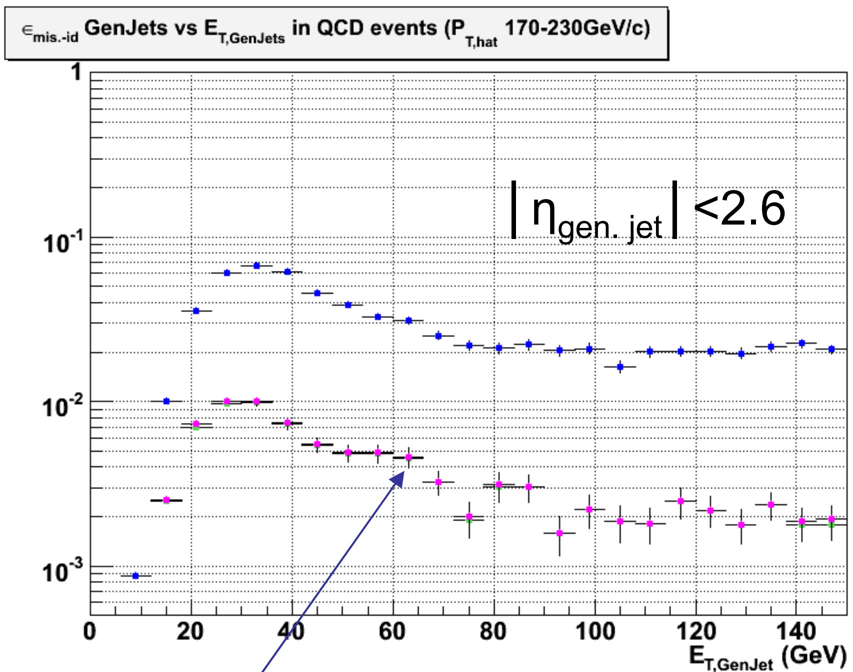
- track isolation
- $\gamma - \pi^0$ reconstruction inside jet,
- use the resulting $\gamma - \pi^0$ candidates in a likelihood ratio.

} **developed these 2 items**

Tau reco + id

- jet reco. + tracker sel.
- + -no neutral ECAL activity- sel / LR sel if –neutral ECAL activity- → discriminator > 0.8
- + -not e/μ- tagged (based on E_{elm},E_{had},$p_{T,track}$





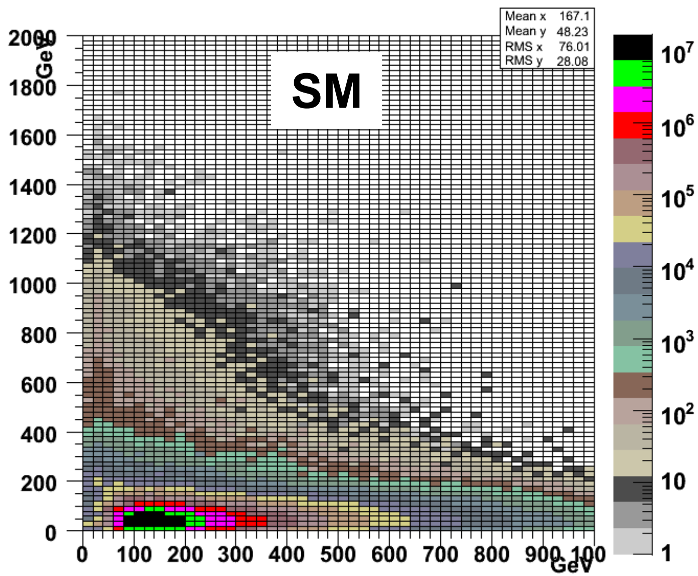
around 40% tau-jet tagging efficiency and few ‰ QCD-jet mis-tagging. efficiency

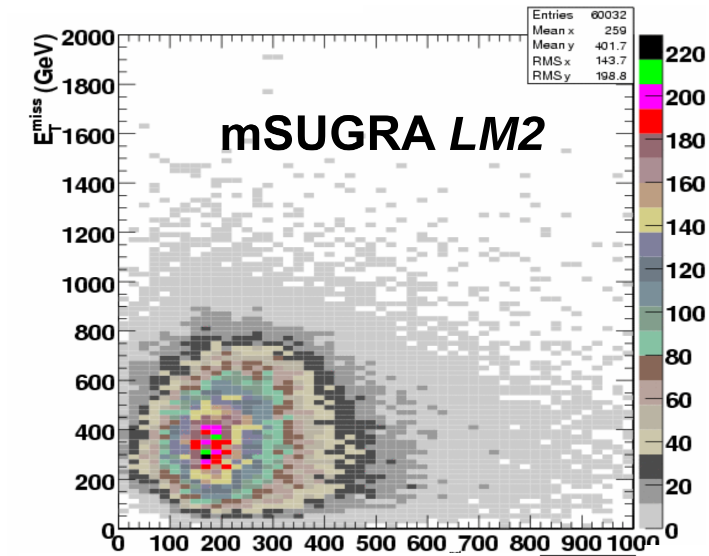
SM / mSUGRA *LM2* discrimination

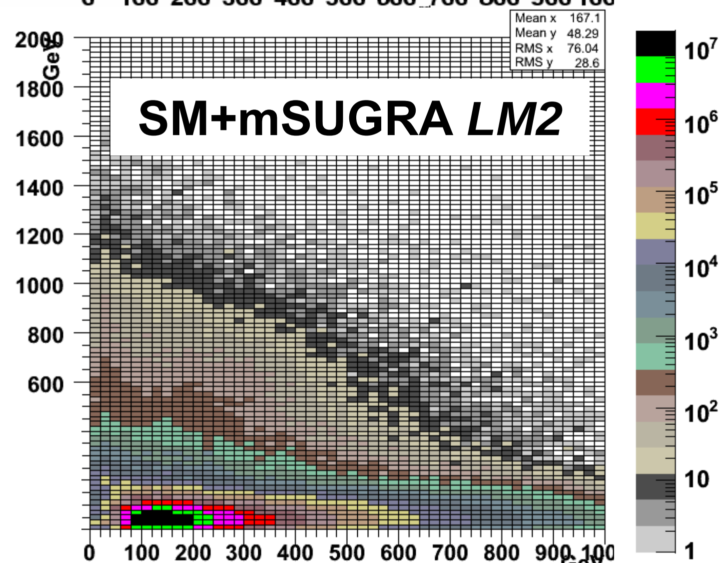
$\int L dt = 10 fb^{-1}$

- Trigger pre-selection (ORCA) :
  - –L1 #28 bit (1 central jet with ET > 88 GeV + ETmiss > 46 GeV)
  - –.AND. HLT #125 bit (1 single jet with ET > 180 GeV)

E_T^{miss} vs $E_T^{2nd\ q/g\text{-}jet\ cand.*}$





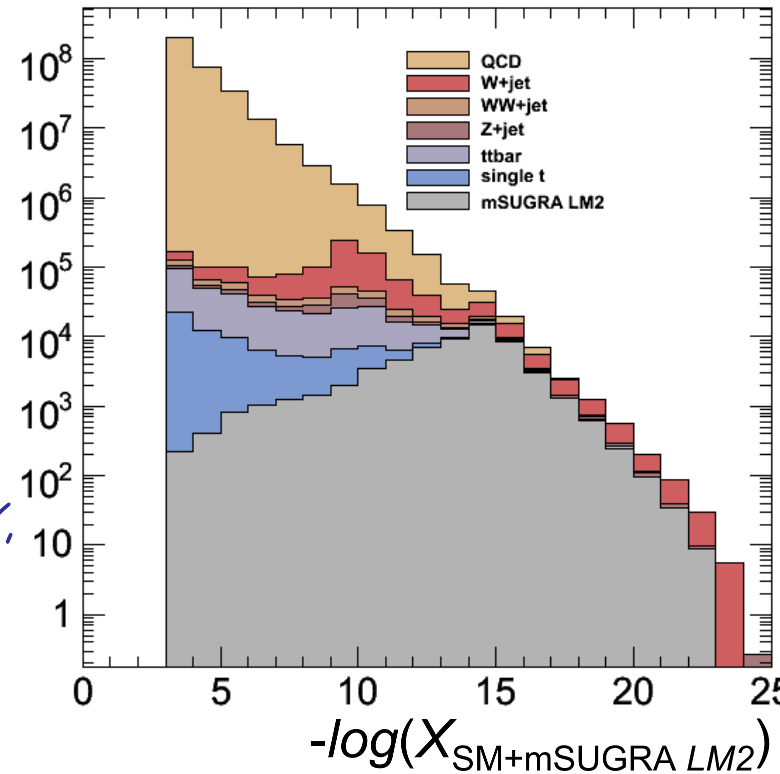


*$E_T^{2nd\ q/g\text{-}jet\ cand.}$ = E_T of the 2nd highest E_T calo. jet “not lepton” tagged

Method A: looking for SUSY as function of the event density

Kinematical variable considered :

- $-log(X)$ where X is defined through the following scheme :
  - a 2D ($E_T^{2nd\ q/g\text{-}jet\ cand.}$, E_T^{miss}) histogram would be filled with a sample of real events, and then would be normalized to 1;



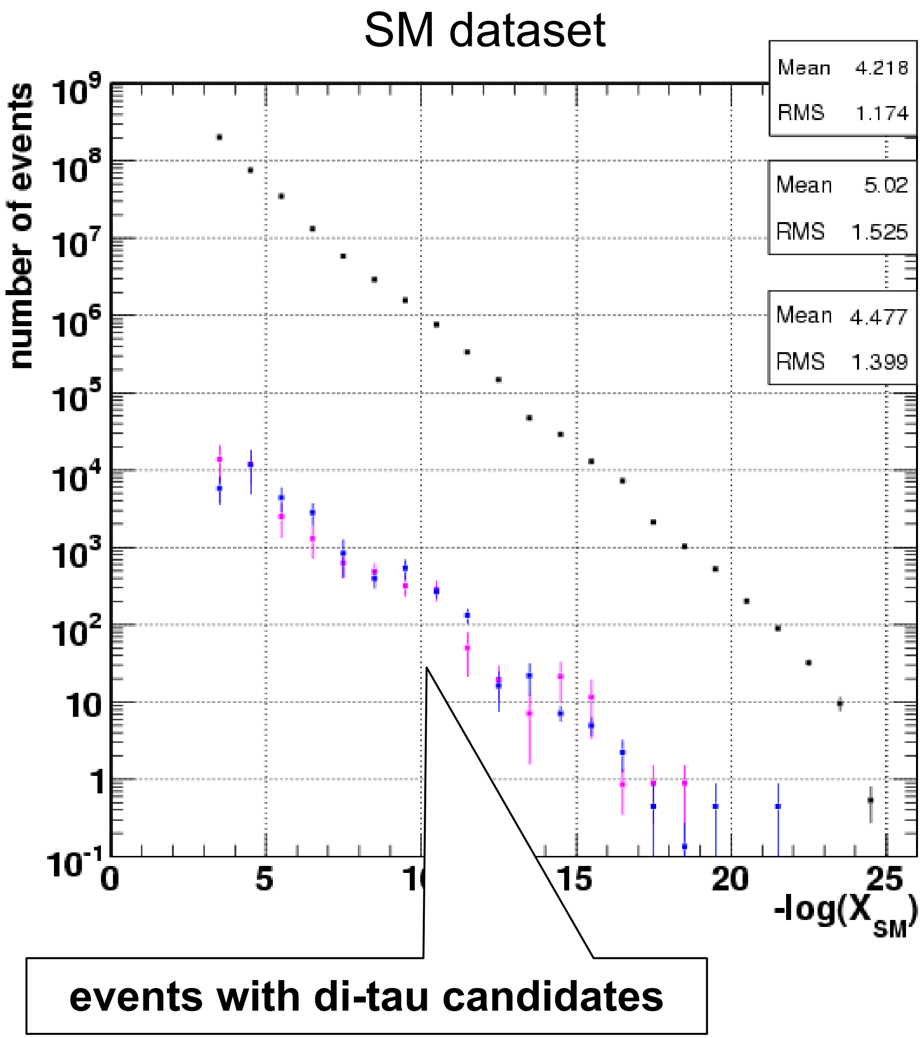
for each event with variables ($E_T^{2nd\ q/g\text{-}jet\ cand.}$, E_T^{miss}), the content of the corresponding bin in the histo. , X, would then describe approximately the density of events in its neighbourhood ;

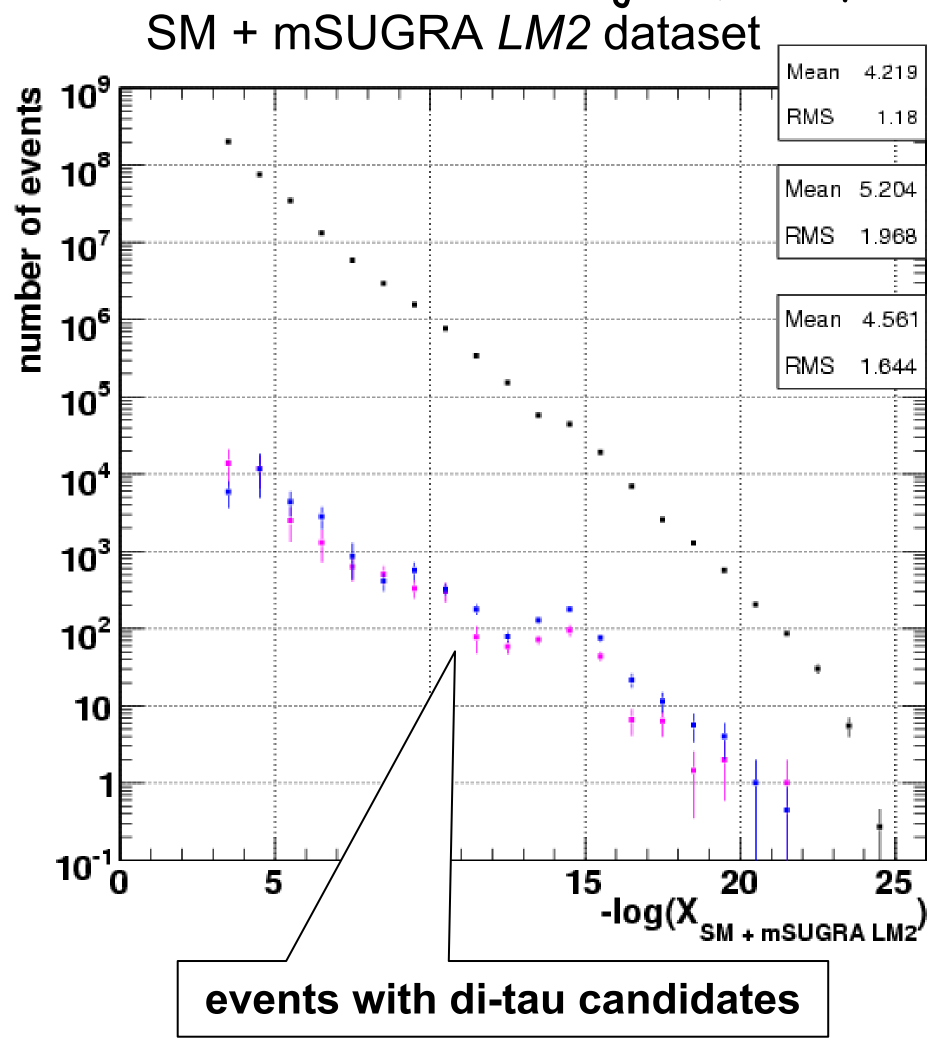
In the following slides two different event samples are studied :
- SM processes*,
- SM +mSUGRA LM2 processes*.

* weighted following the expected cross-section of the process

Method A: looking for SUSY as function of the event density

$\int L dt = 10 fb^{-1}$



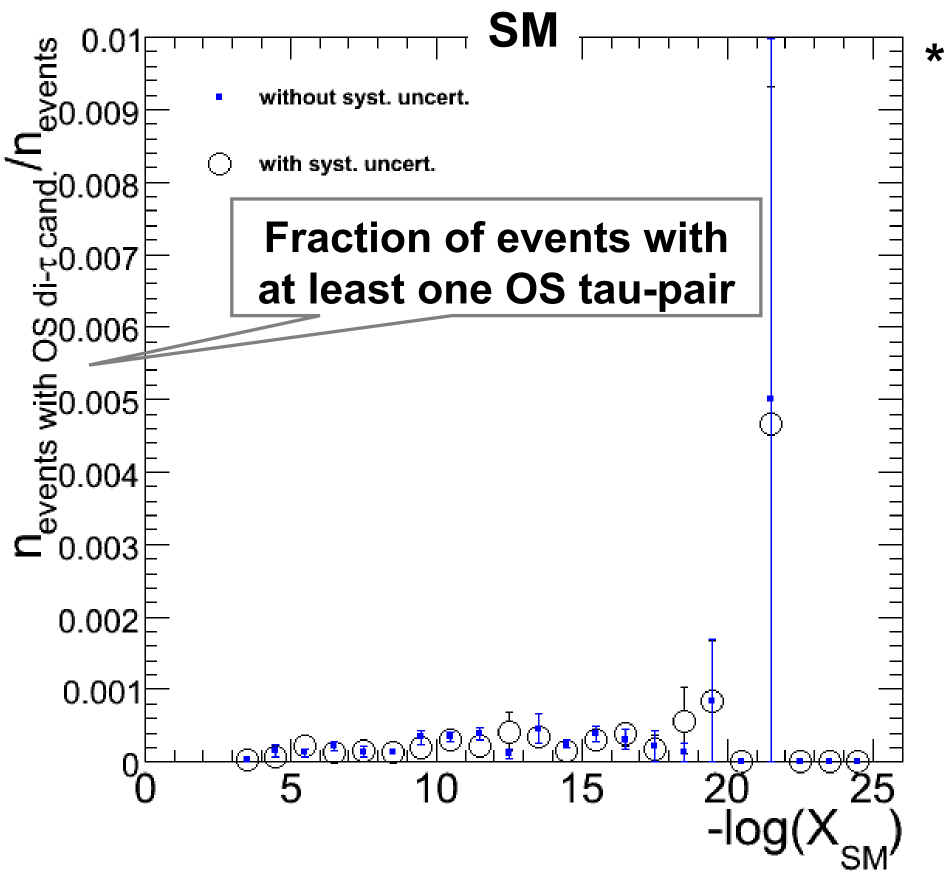


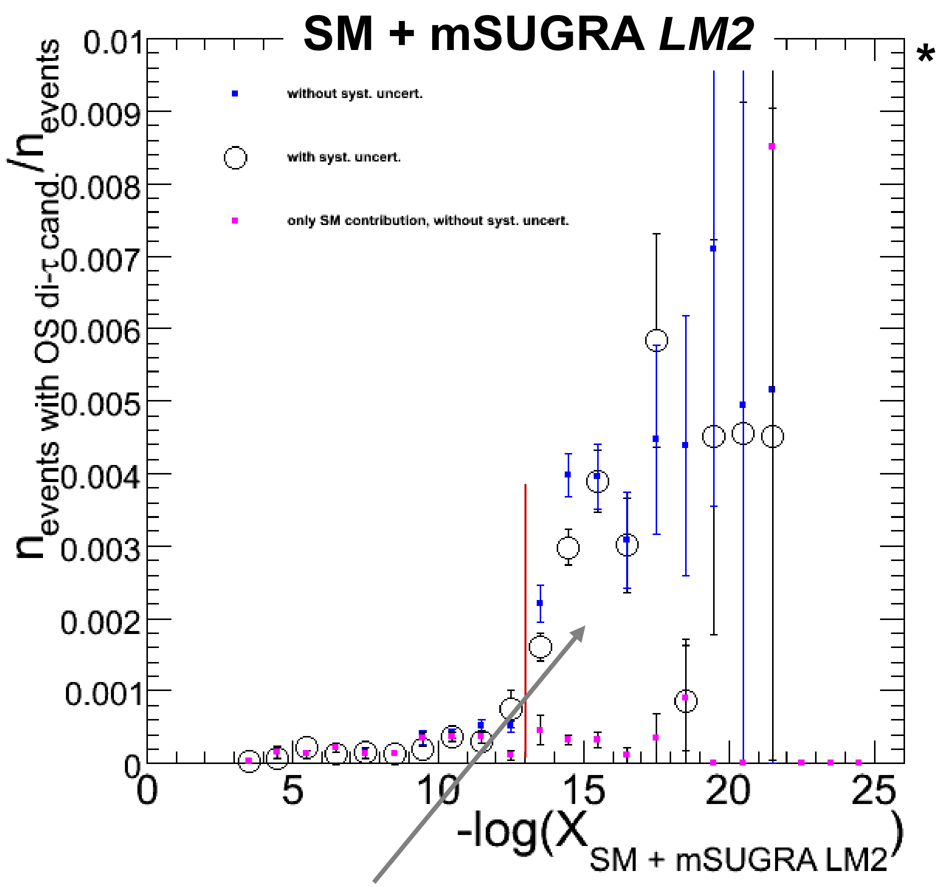
- all events
- events with OS di-tau candidates
- events with SS di-tau candidates

Method A for observing mSUGRA events in data

$\int L dt = 10 fb^{-1}$

-*log*(*X*) histogram obtained by the division of
-*log*(*X*) histogram for the events containing at least 1 rec. OS charge hadr. τ-jet cand. pair
by -*log*(*X*) histogram for all the events





$N_{\text{events with OS di-}\tau\text{ cand.}}$ **obs. = 424**

$N_{\text{events with OS di-}\tau\text{ cand.}}$ **exp. SM = 112**

see slides 32-33

* with binomial errors

✓ A natural approach, by answering to the question:

Are the kinematically particular events
particular by their multiplicity in (tau-)tagged reconstructed objects too?

✓ The simple observation of a bump in a distribution
allows to sign the presence of non-SM events,
<u>without preselection except at trigger level</u>.

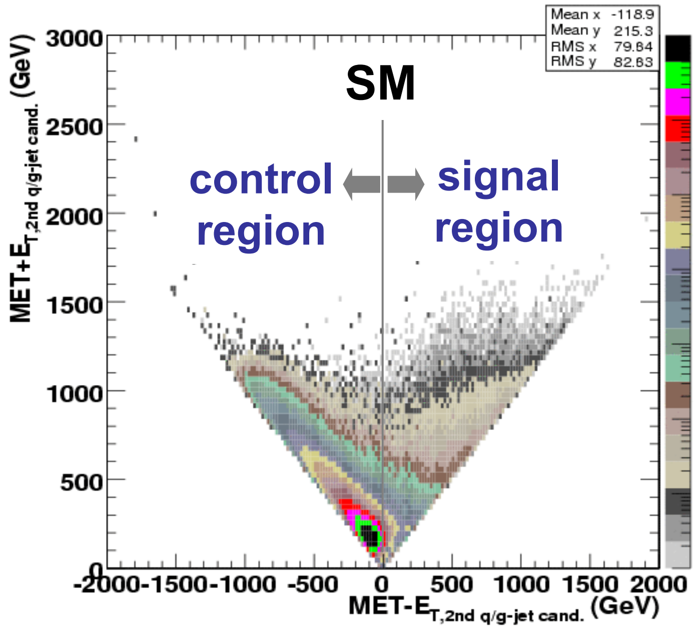
✓ The null hypothesis -*data contain only SM events*- is tested, not by the comparison between expected and observed numbers of events in the tail of a distribution (ex:MET), but by the comparison between fractions of events with >0 tau cand. pair in such a tail
→ not sensitive to kinematically badly reconstructed events which would populate the tail.

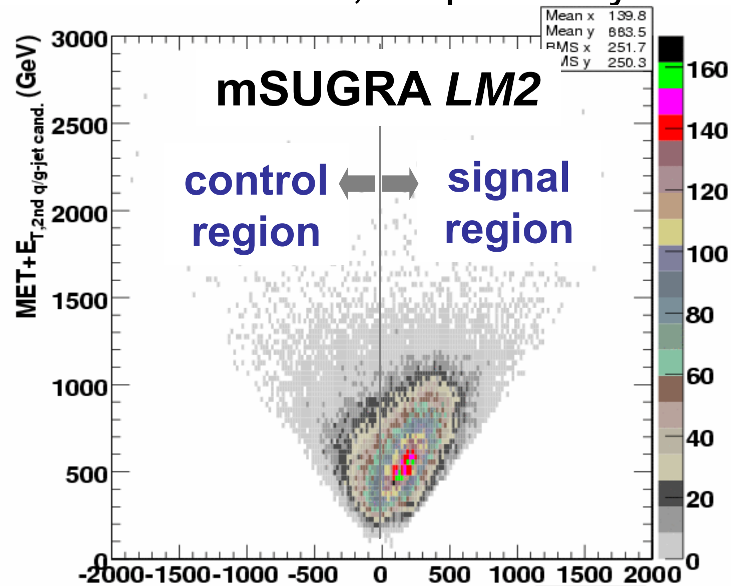
✗ -Problem pointed out by the CMS Statistics Committee- for each sample, X is different :
the same event present in two different samples A and B gets not equal X_A and X_B values,
→ difficult comparison between X_A and X_B distributions from A and B samples respectively.

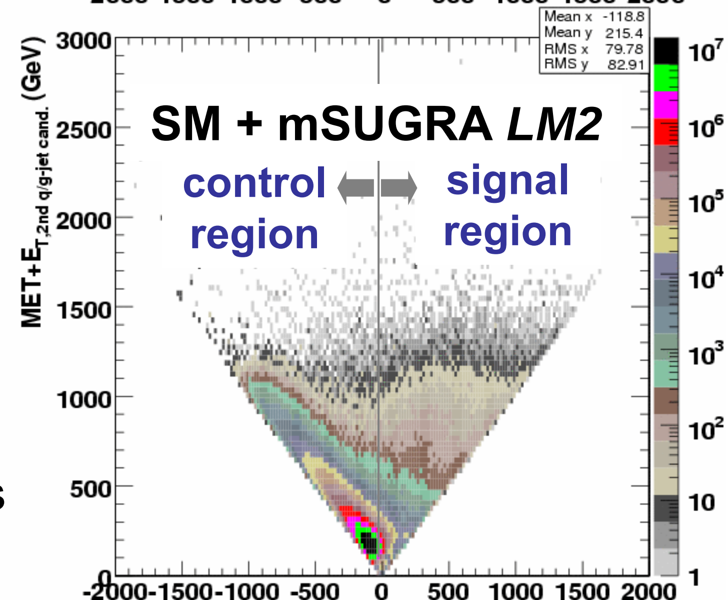
Method B Using a control region to estimate SM contribution

$E_T^{miss} + E_T^{2nd\ q/g\text{-}jet\ cand.}$ vs $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet\ cand.}$ $\int L dt = 10 fb^{-1}$

•I define 2 regions (control region, dominated by SM events, and signal region, influenced by SUSY events) using $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet} < 0$ or > 0, respectively





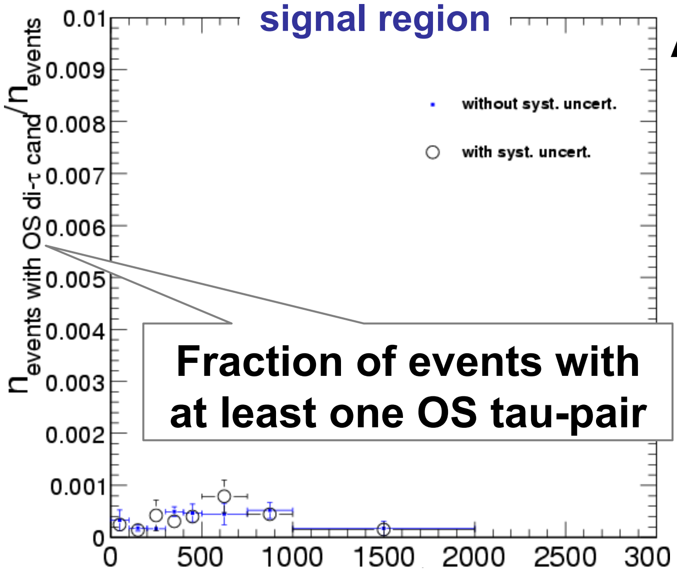


•The SM contribution in the signal region is estimated from the control region using the knowledge (MC and data) of how the SM events are distributed between the two regions.

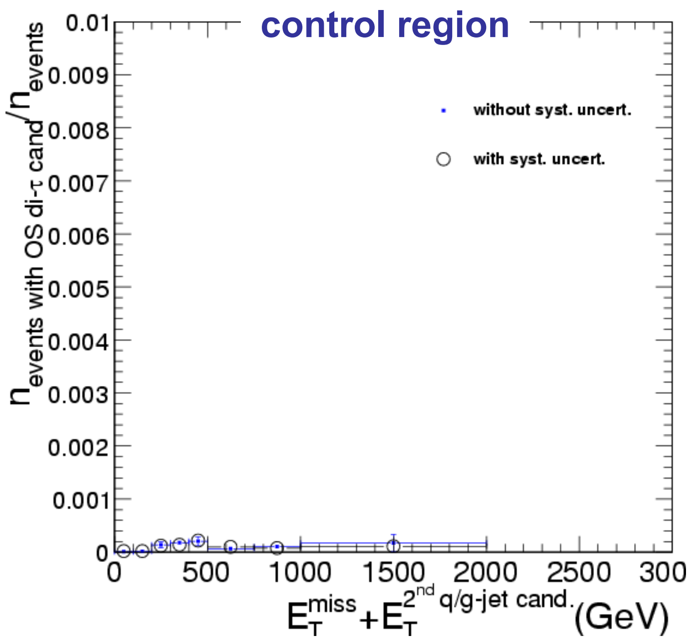
Method B Using a control region to estimate SM contribution

$\int L dt = 10 fb^{-1}$

SM MC

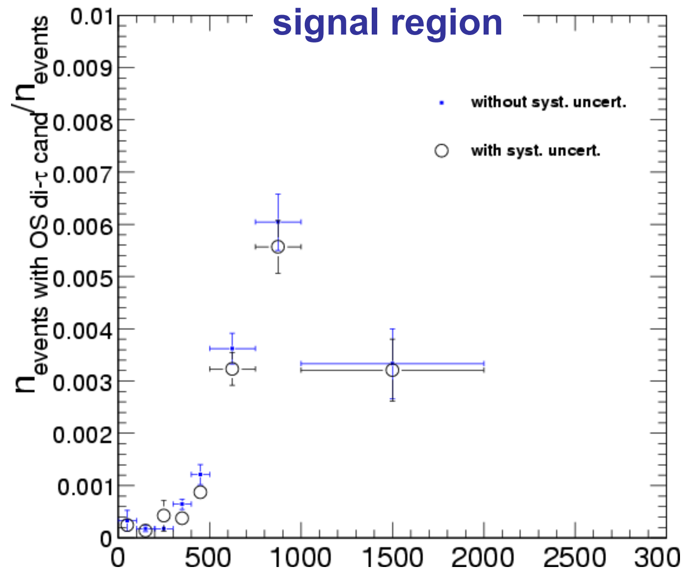


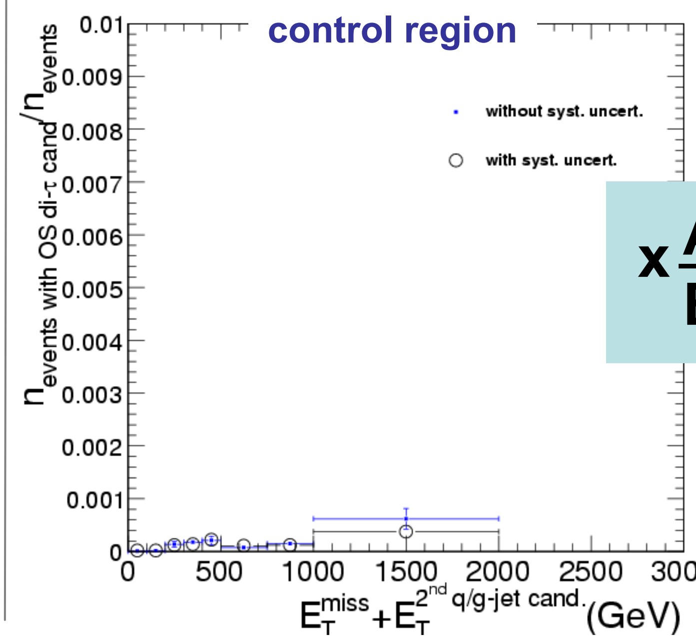
A



B

data which contain SM + mSUGRA *LM2*





$\times \frac{A}{B}$

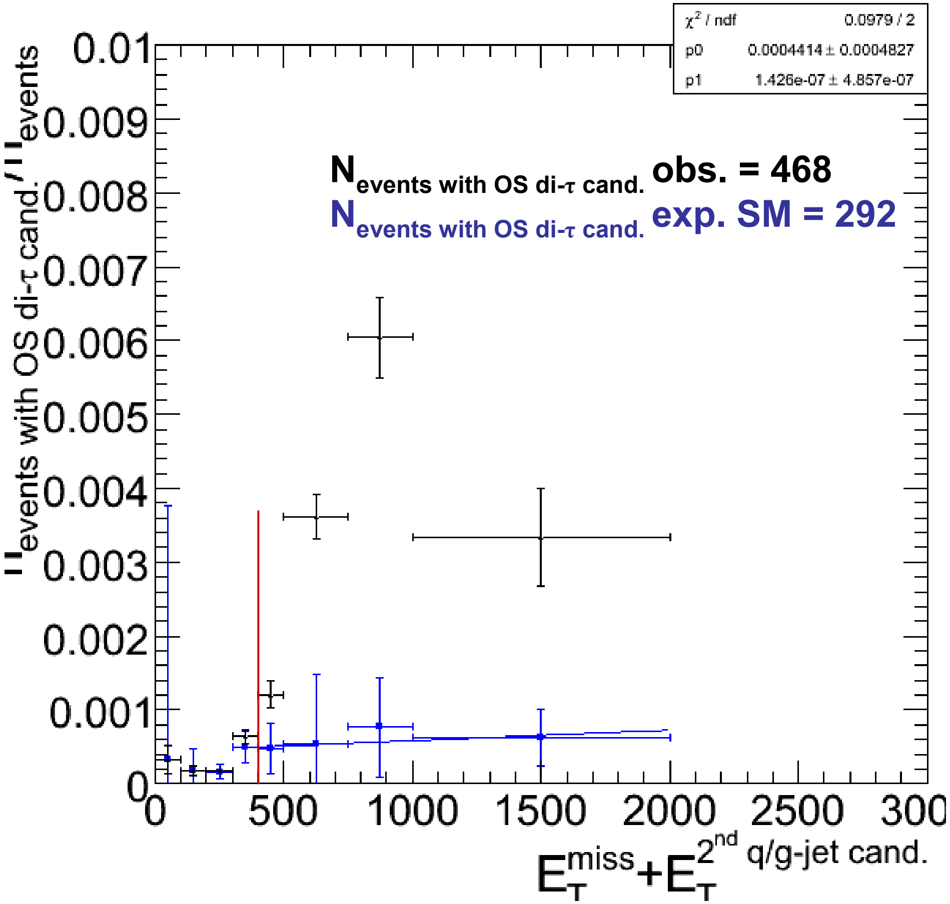
give the estimate of the SM contribution in signal region

Method B for observing mSUGRA events in data

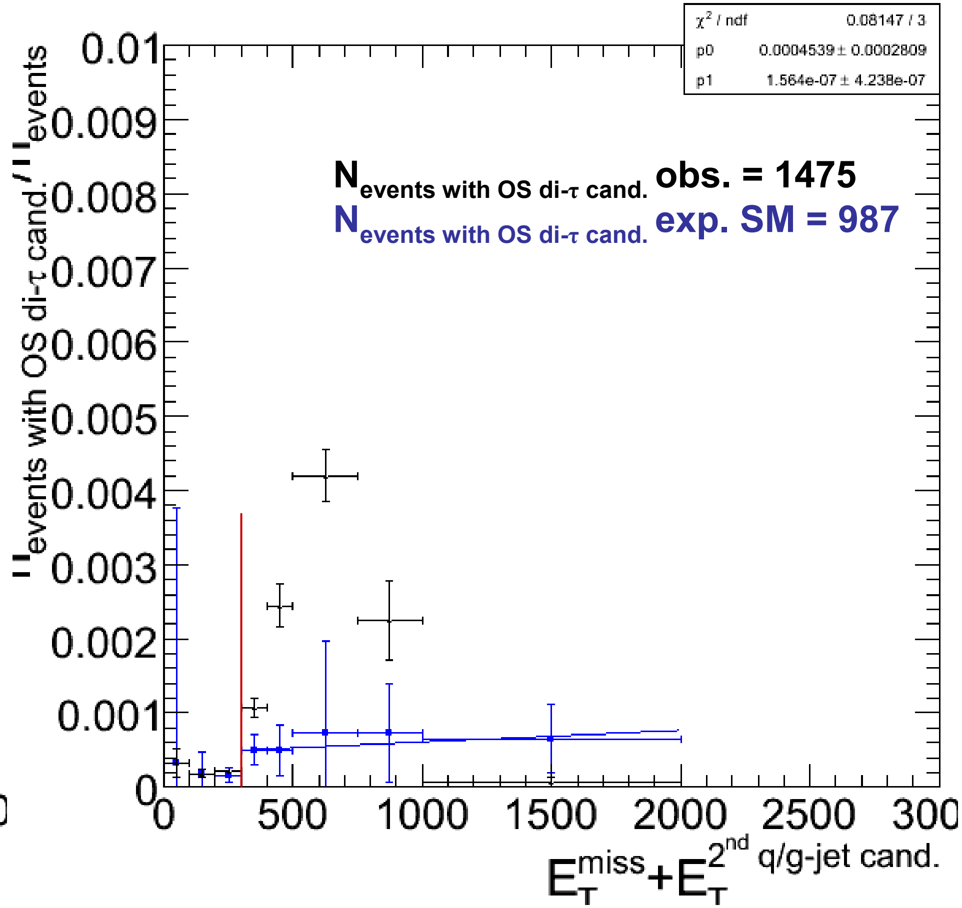
$\int L dt = 10 fb^{-1}$

$$\frac{\text{number of events with OS di}\tau}{\text{number of events}}$$ signal region

for LM2



for LM1



- ● observed
- ● expected (extrapolated from control region)

Conclusions

- I have developed 2 methods to detect SUSY events in data, both based on the appearance of an excess of tau-pairs as a function of a kinematical variable

  - 1st method uses the pseudo-density of events in a kinematical space,

  - 2nd method divides a kinematical plane into a control region (where SM events dominate) and a signal region (where SUSY events dominate). The SM contribution in the signal region is estimated from the data in the control region.

  (other variables were tried ; they did not change the picture significantly.)

Backup slides

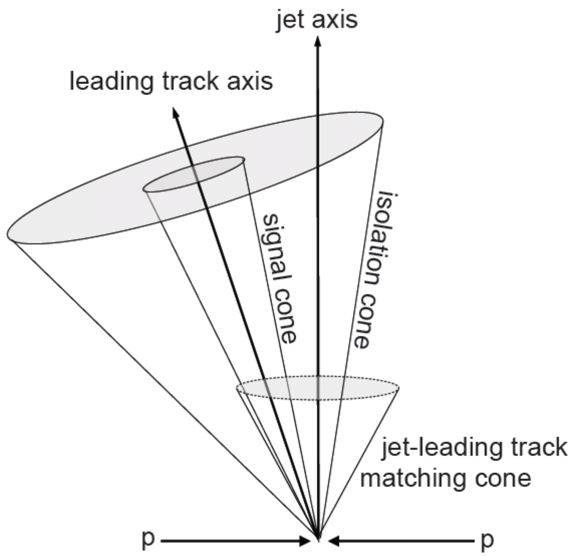
Tau reco + id

• The hadr. tau-jet candidates selection scheme used in the following study :
(underwent modifications afterwards)

- step 0 : a rec. ECAL+HCAL $\Delta R<0.50$ (*iter cone* algo.) jet with $10\text{GeV} < E_{T,rec.jet} < 150\text{GeV}$*
and $|\eta_{rec.jet}| < 2.6$,

- step 1(tracker) : 0 rec. tk (with $P_T>1.5\text{GeV/c}$) in an $x<\Delta R<0.40$ isolation annulus,
1 or 3 rec. tk(s) (with $P_T>1.5\text{GeV/c}$) in a $\Delta R \le x$ signal cone
around a rec. leading ($P_T>5\text{GeV/c}$ if 1 signal tk, $>2.5\text{GeV/c}$ if 3 signal tks) tk
found in a $\Delta R<0.17$ matching cone around jet axis

with $x = 0.17$ when $E_{T,rec.jet} < 21\text{GeV}$,
$x = 3.5/E_{T,rec.jet}$ when $21\text{GeV} \le E_{T,rec.jet} < 70\text{GeV}$,
$x = 0.05$ when $E_{T,rec.jet} \ge 70\text{GeV}$.



- →next slide

* ORCA version of the likelihood ratio algo. (in step 2') not usable above the upper limit

Tau reco + id

- step 2(ECAL+tracker) : no rec. *neutral*** ECAL clus. (with E>1GeV) inside jet,

- step 2'(ECAL+tracker) : if rec. *neutral* ECAL clus. inside jet, then minimal value of a likelihood ratio which combines the following discriminant variables :

case 1 signal tk :
- # rec. *neutral* ECAL clus.,
- mean $\Delta R_{\text{tk-} \textit{neutr.} \text{ ECAL clus.}}$,
- $E_{\textit{neutr.} \text{ ECAL clus. in isol. strip}} / (E_{\textit{neutr.} \text{ ECAL clus}} + P_{tk})$,
- tk transverse impact parameter

case 3 signal tks:
- # rec. neutral ECAL clus.,
- $E_{\text{neutr. ECAL clus. in isol. strip}} / (E_{\text{neutr.ECAL clus}} + \Sigma P_{tk})$,
- signed flight path significance (secondary vtx reco)

-step 3 : not *e*/*μ* cand. - tagged, using D. J. Mangeol technique (see CMS AN 2006/015) slightly modified.

** specific to an ECAL cluster whose direction is not inside a ΔR<0.015 cone around the direction of contact point between any propagated track and the ECAL surface

The events were simulated with OSCAR and reconstructed with ORCA(8_13_3).

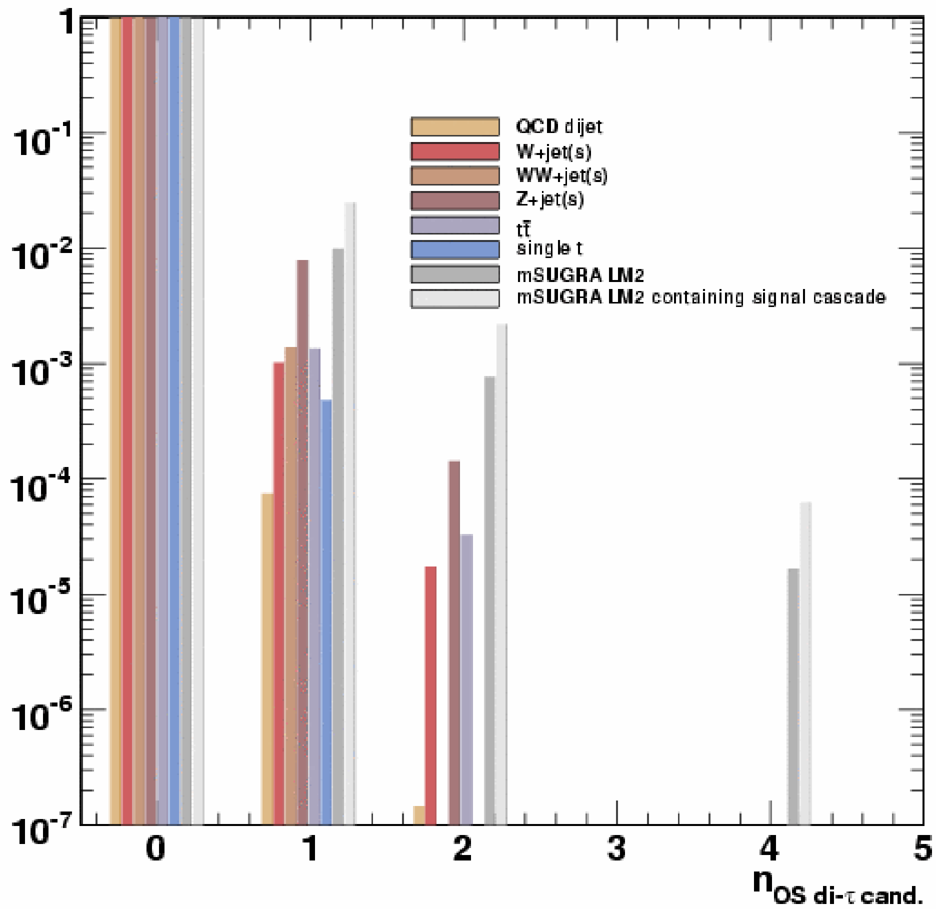
Event selection

- Trigger:
  - L1 #28 bit (1 central jet with ET > 88 GeV + ETmiss > 46 GeV)
  - .AND. HLT #125 bit (1 single jet with ET > 180 GeV)
- Tau
  - |Eta-tau| < 2.6
  - ET-tau > 10 GeV
- Jets (gamma jet calibred when *not lepton*-tagged)
  - |Eta-jet| > 2.6
  - ET-jet >5 GeV
- MET (based on calo-jets)
  - No clean-up
  - No muon correction

histogram normalized to unit area.

- Tau-related variable

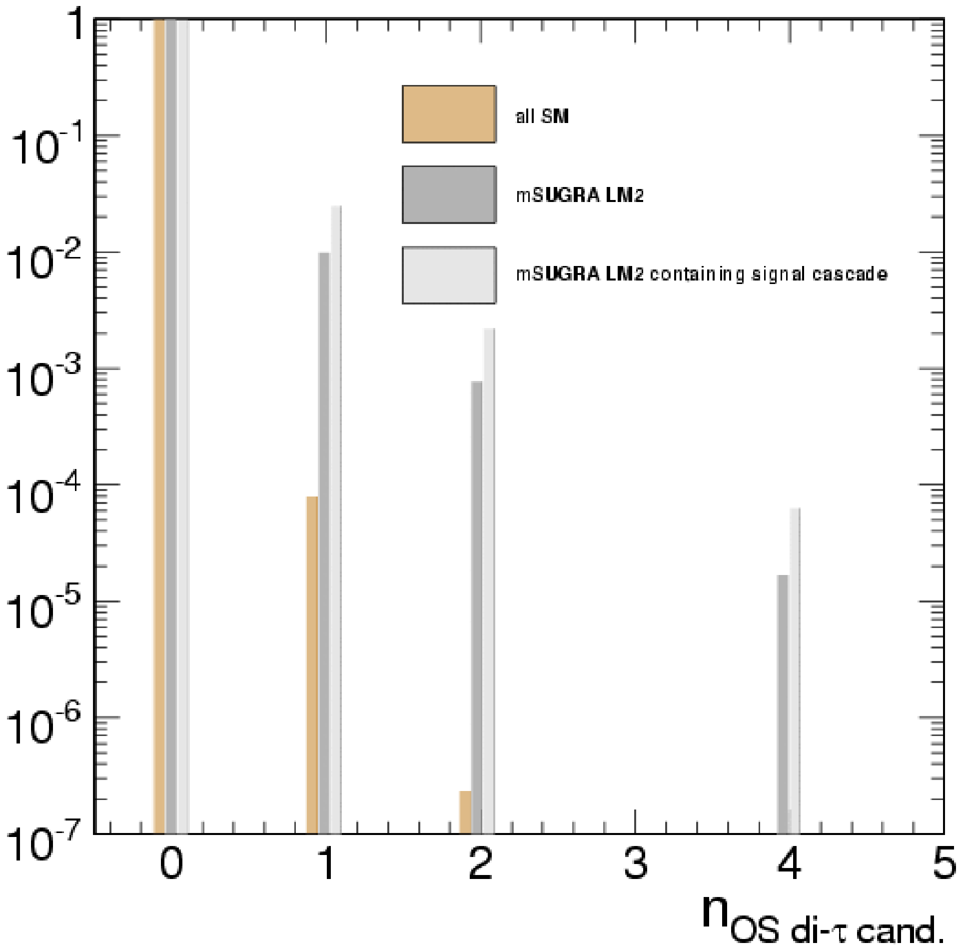
rec. hadr. τ-jet cand. pairs
of OScharge



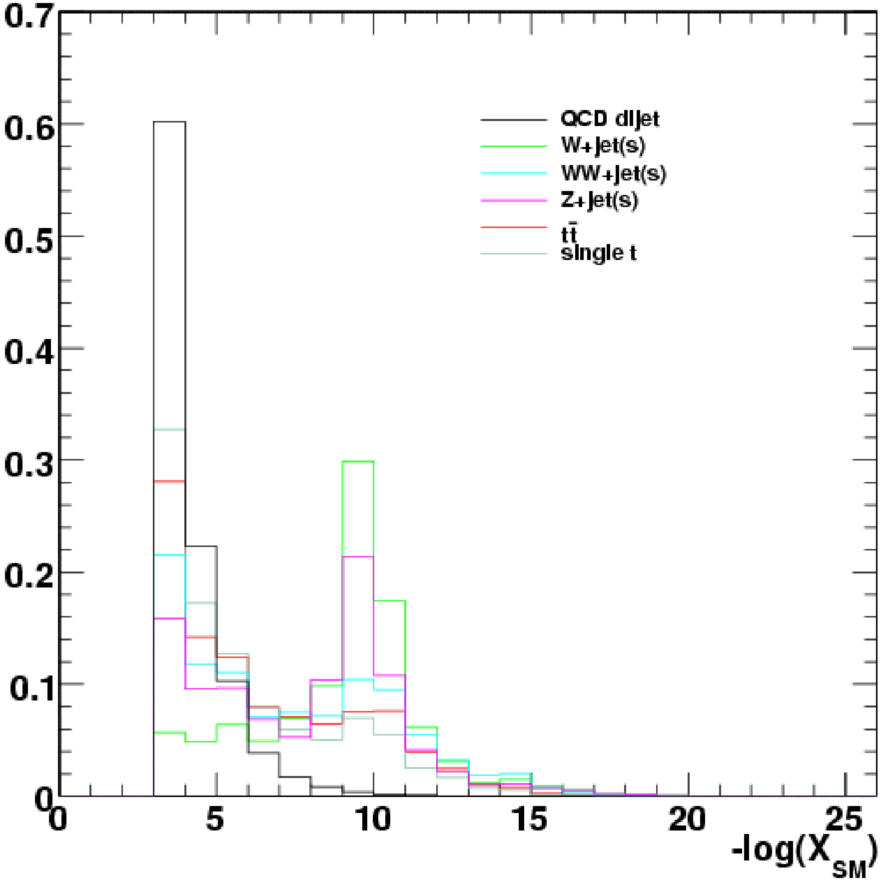
histogram normalized to unit area.

- Tau-related variable

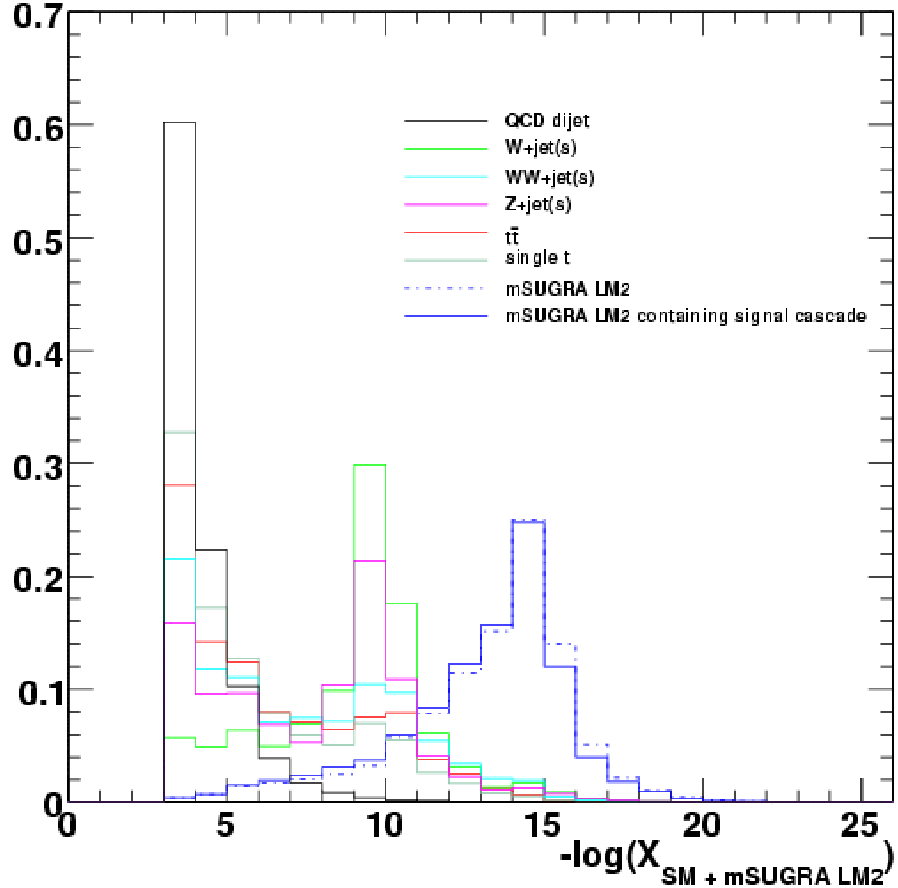
rec. hadr. τ-jet cand. pairs of OScharge



only SM processes



SM and mSUGRA *LM2* processes

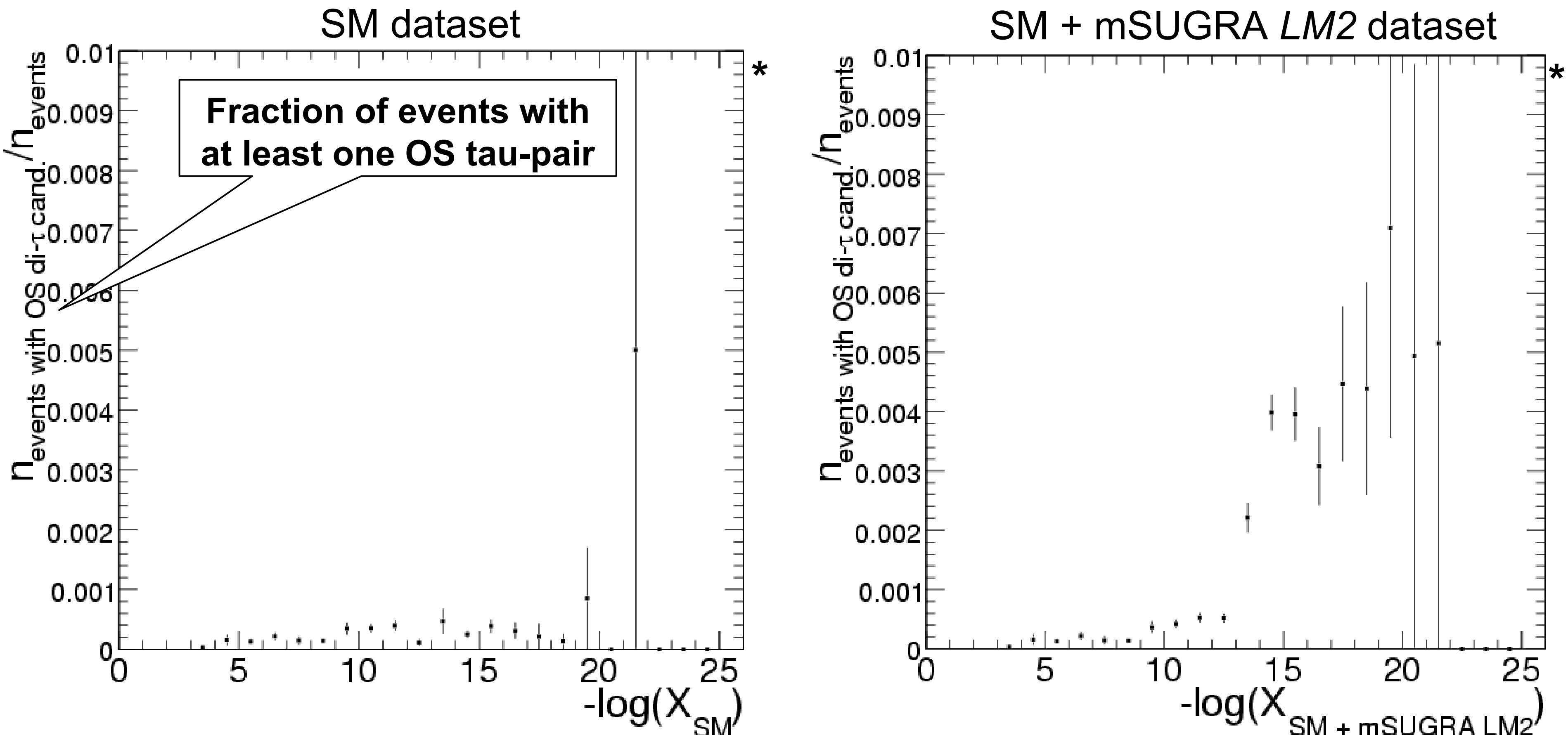


* Distributions normalized to 1

Method A for observing mSUGRA events in data

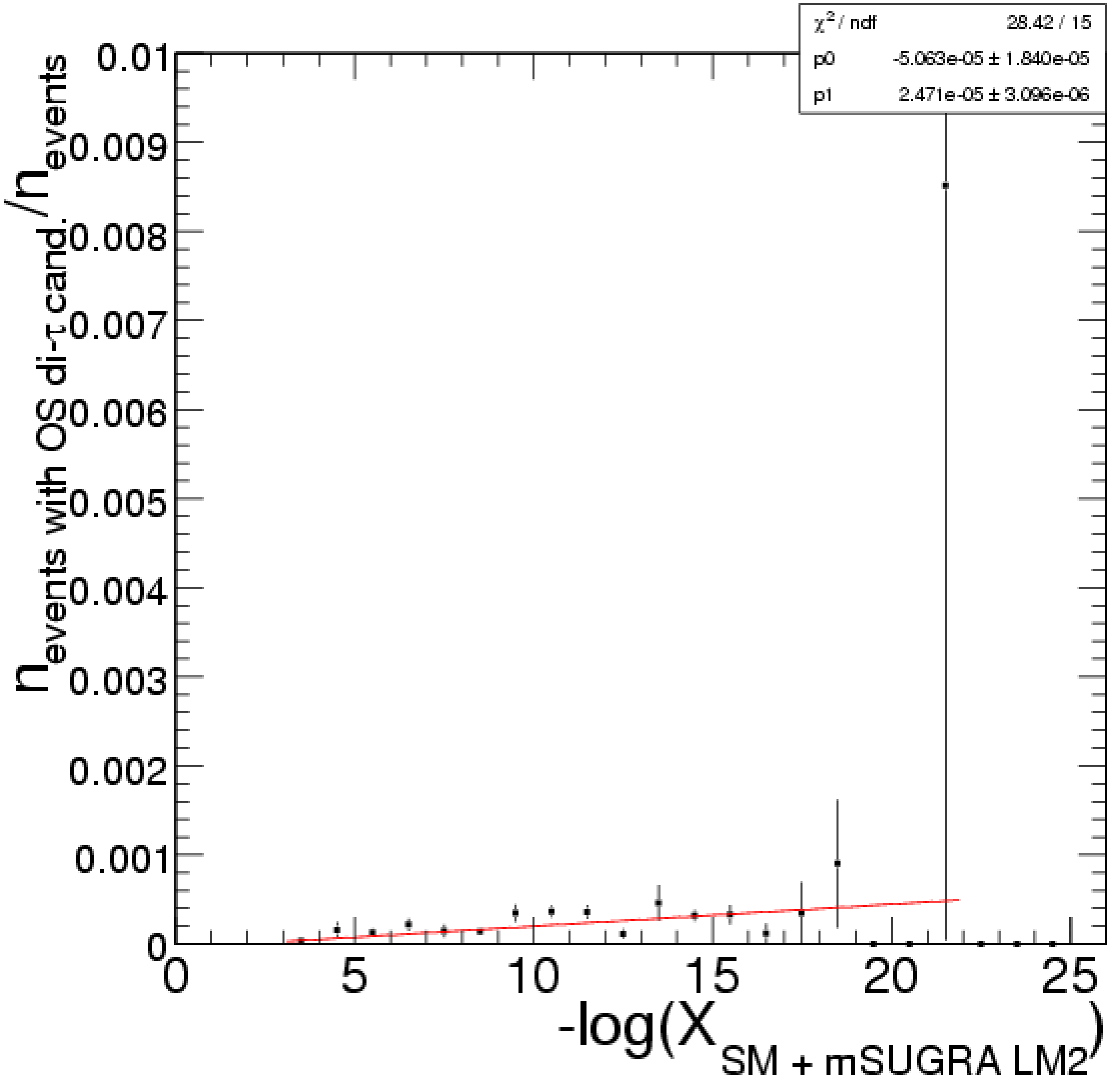
$\int L dt = 10 fb^{-1}$

-*log*(*X*) histogram obtained by the division of
-*log*(*X*) histogram for the events containing at least 1 rec. OS charge hadr. τ-jet cand. pair
by -*log*(*X*) histogram for all the events



* with binomial errors

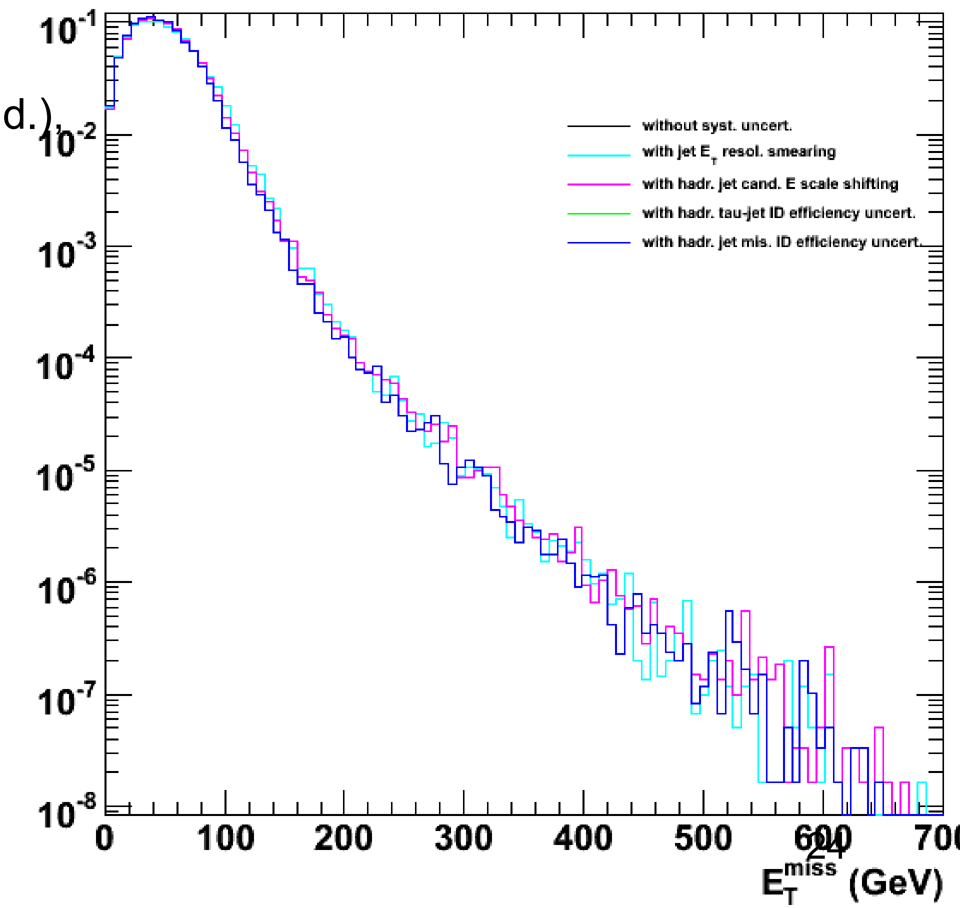
SM process events in SM + mSUGRA *LM2* dataset



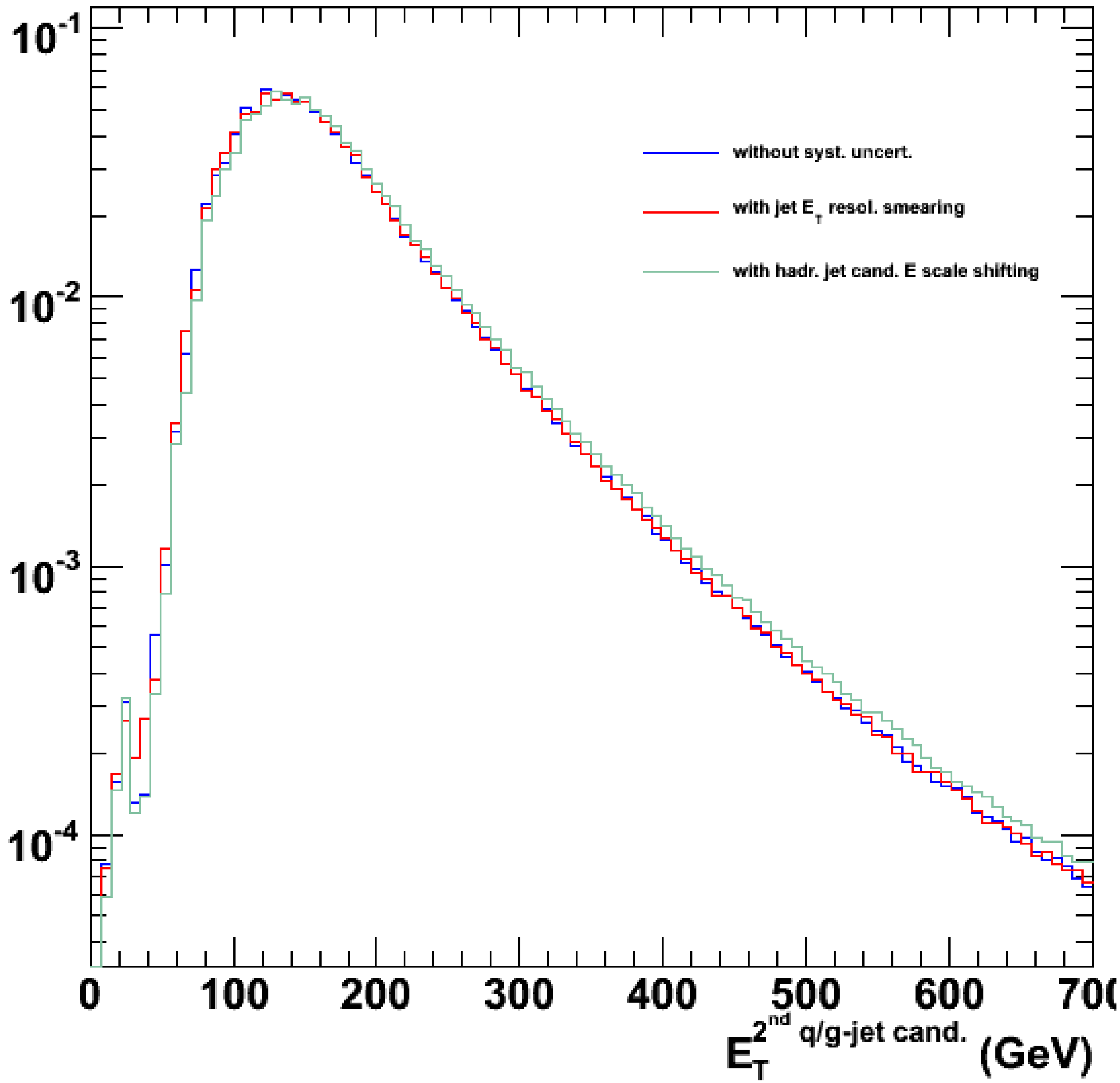
Systematic uncertainties

- **uncertainty on the jet E_T resolution**
  (smeared it by 10%),
- **uncertainty on hadr. Tau-jet ID efficiency**
  (removed 9% of the truth matched hadr. tau-jet cand.),
- **uncertainty on q/g-jet mis. ID efficiency**
  (added 10% of q/g-jet matched hadr. tau-jet cand.),
- **uncertainty on q/g-jet E scale**
  (increased/decreased rec. q/g-jet cand. E by a fraction dependent on its P_T).

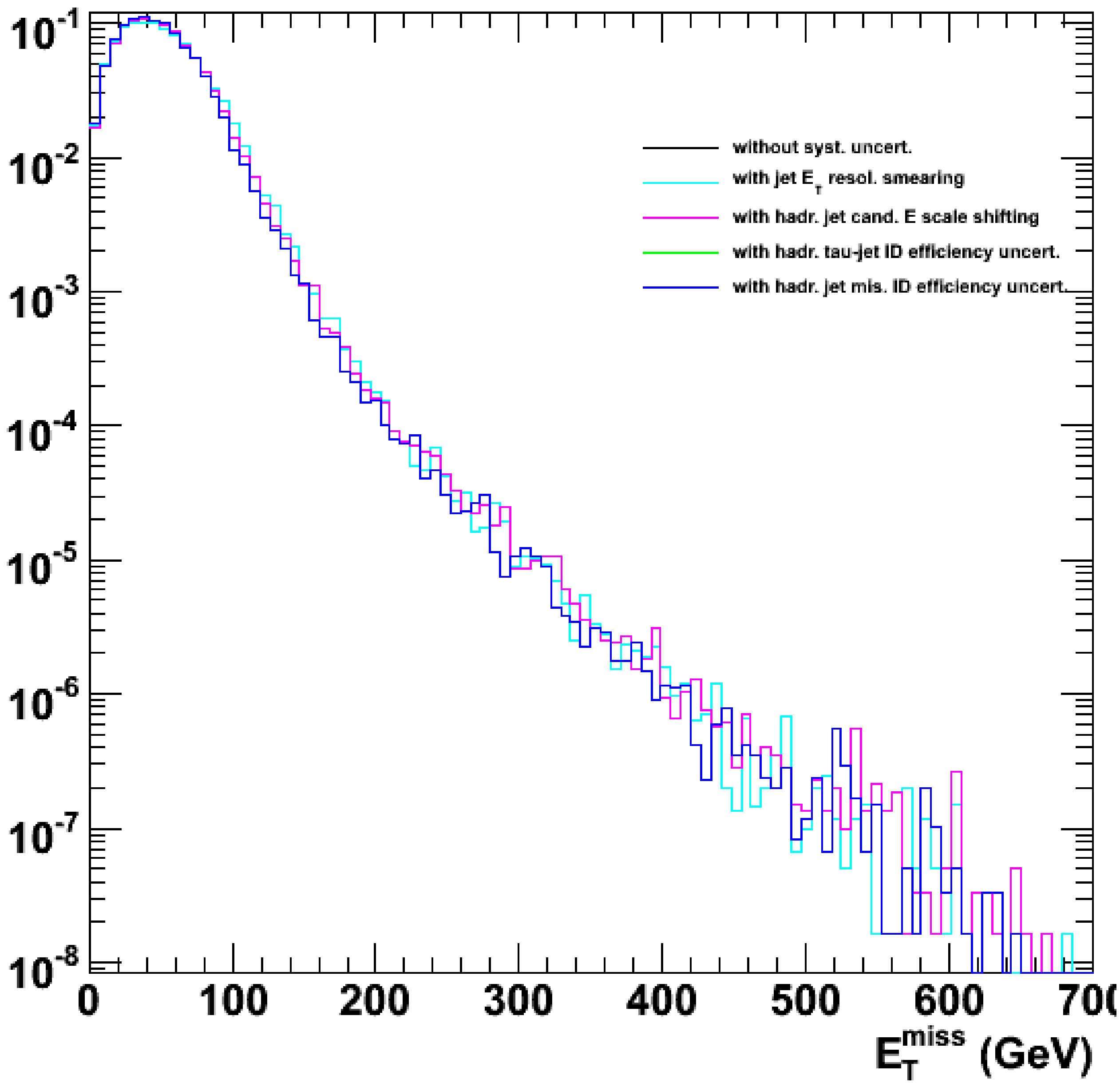
Example in QCD dijet processes



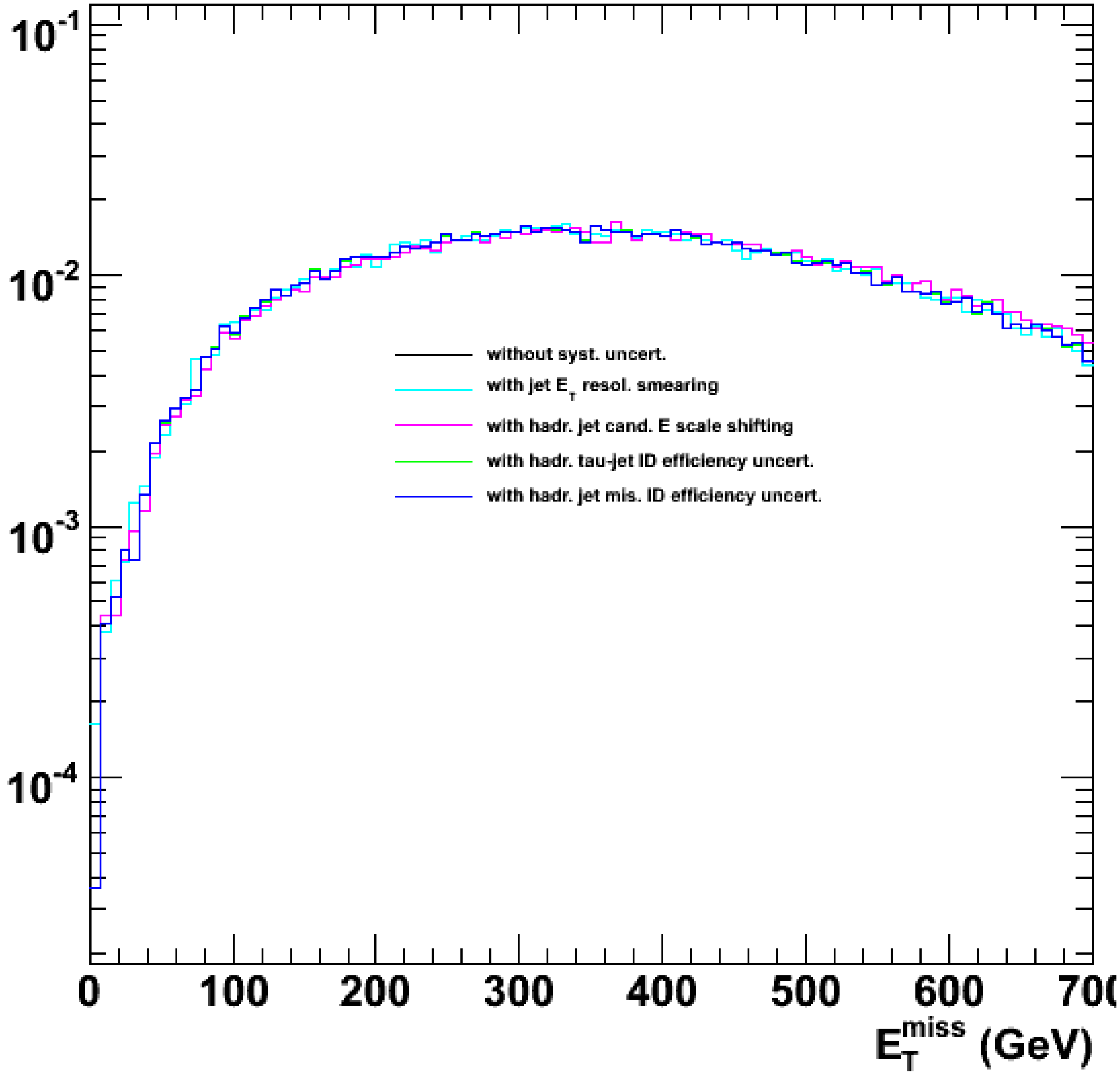
QCD dijet processes



QCD dijet processes



mSUGRA *LM2* processes

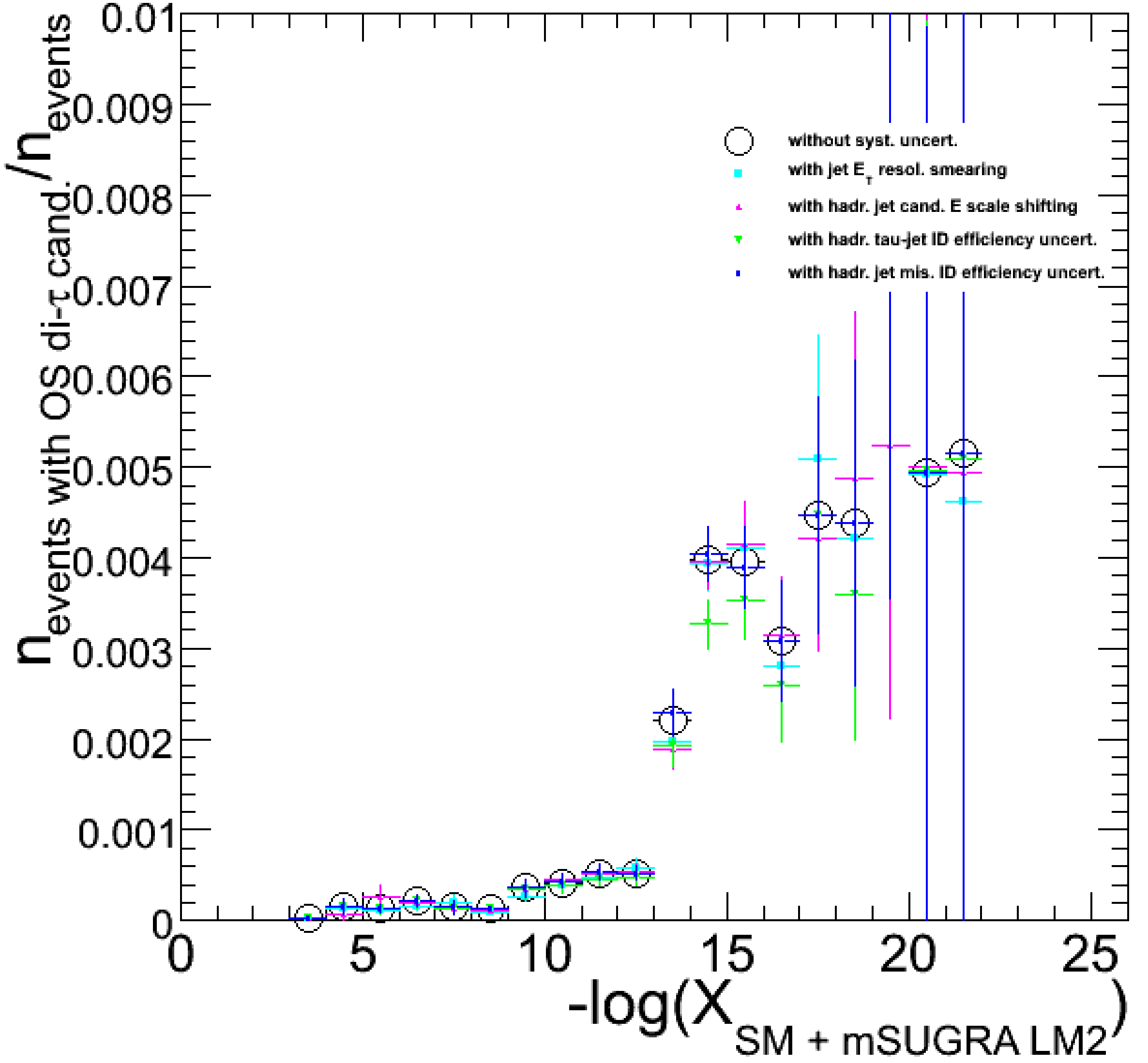


Method A for pointing out the presence of mSUGRA *LM2* events in data

(kinematic reference variable : -*log*(*X*))

$\int L dt$ = 10fb^{-1}

SM + mSUGRA *LM2* dataset



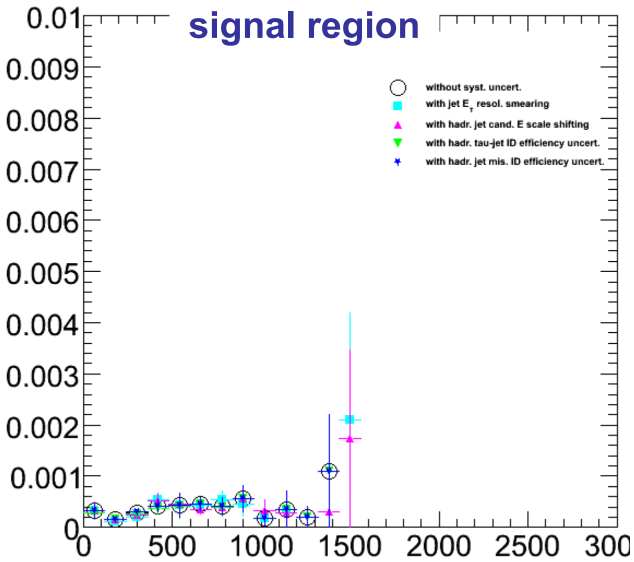
*

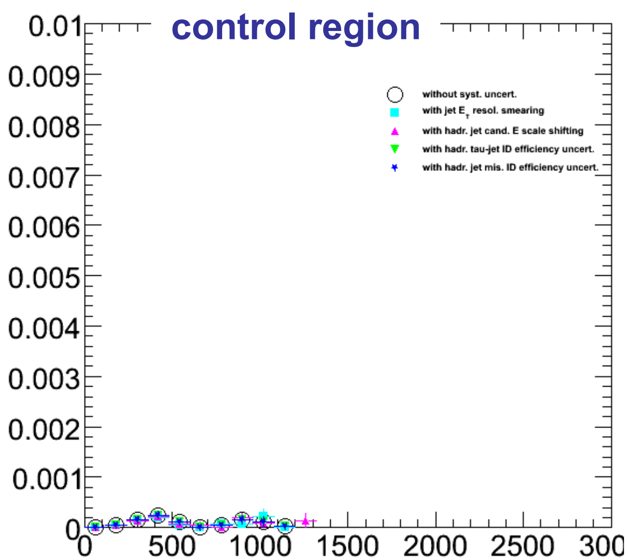
* with binomial errors

Method A' for pointing out the presence of mSUGRA *LM2* events in data

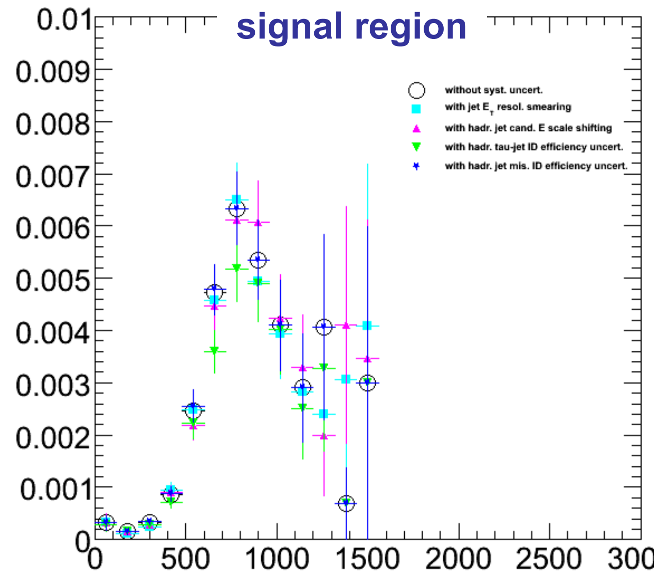
$\int L dt = 10 fb^{-1}$

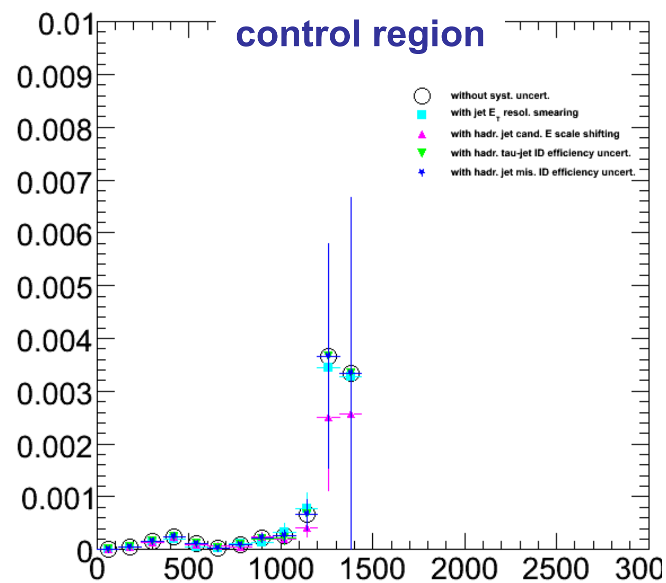
SM dataset





SM + mSUGRA *LM2* dataset





Method A for observing mSUGRA events in data

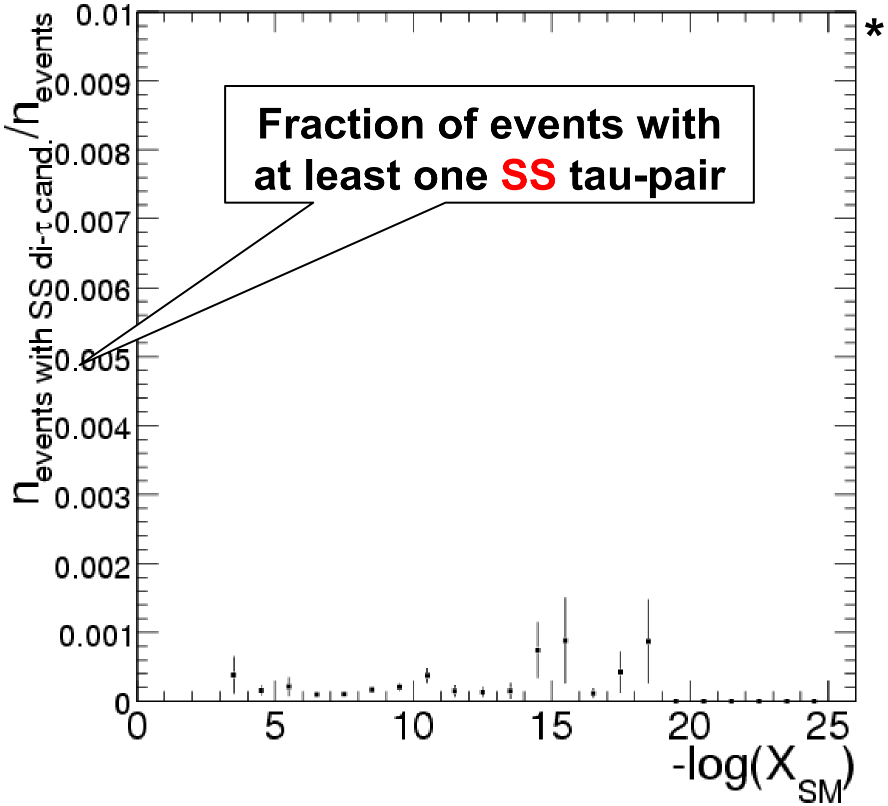
$\int L dt = 10 fb^{-1}$

-log(*X*) histogram obtained by the division of

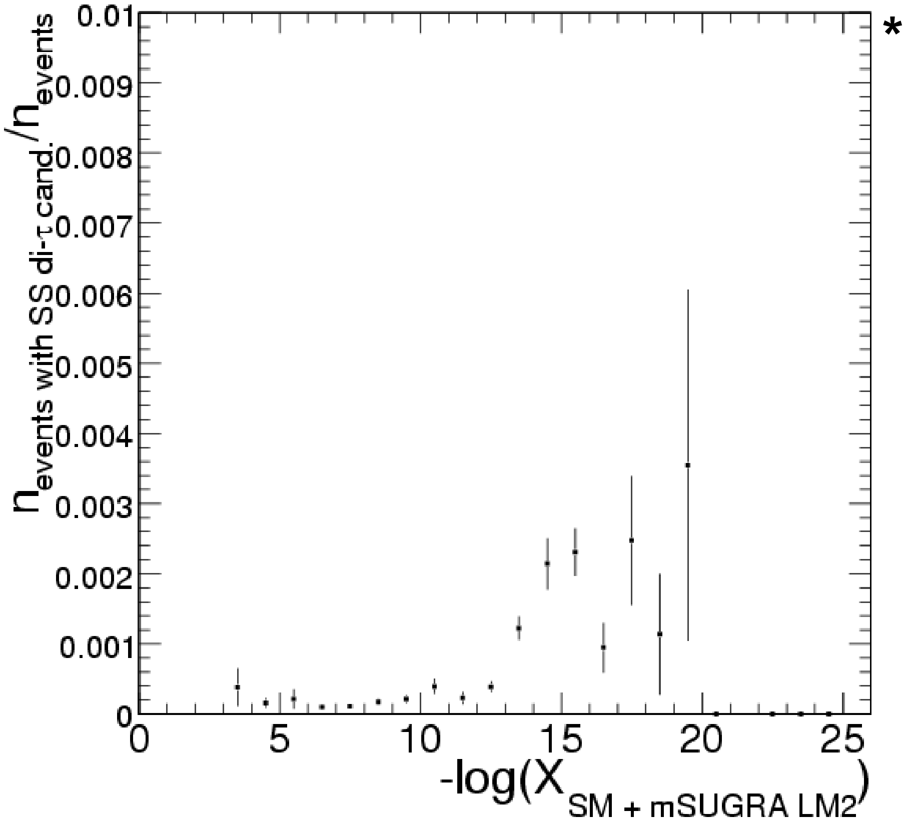
-log(*X*) histogram for the events containing at least 1 rec. SS charge hadr. τ-jet cand. pair

by *-log*(*X*) histogram for all the events

SM dataset



SM + mSUGRA *LM2* dataset



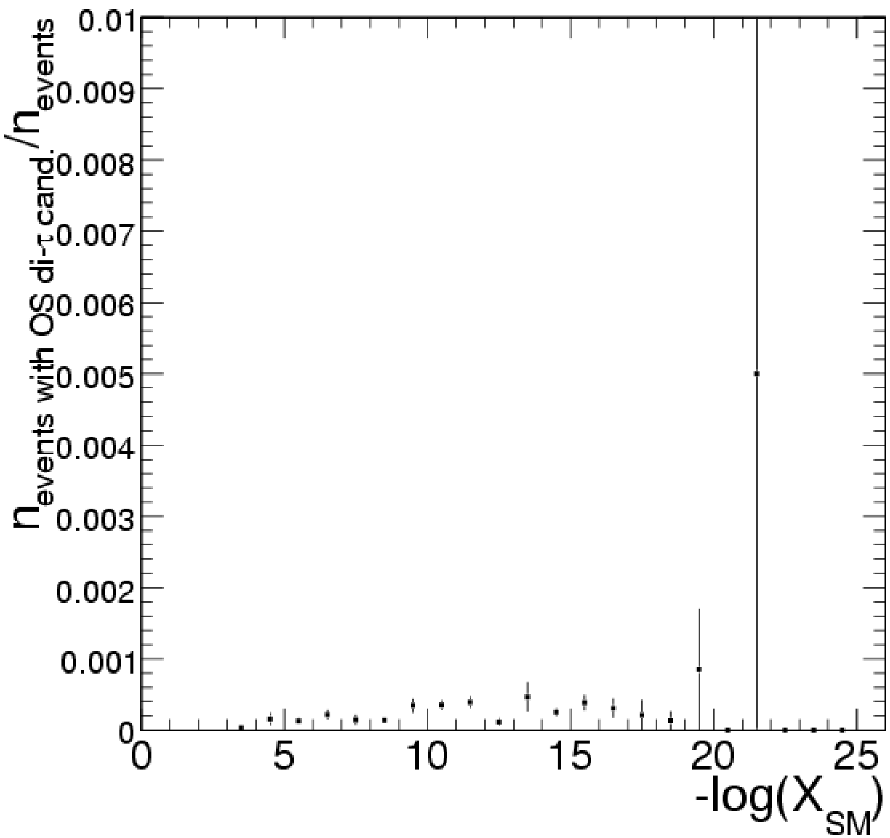
* with binomial errors

Same histograms as on slide 16 except the mSUGRA test point used, now *LM1*

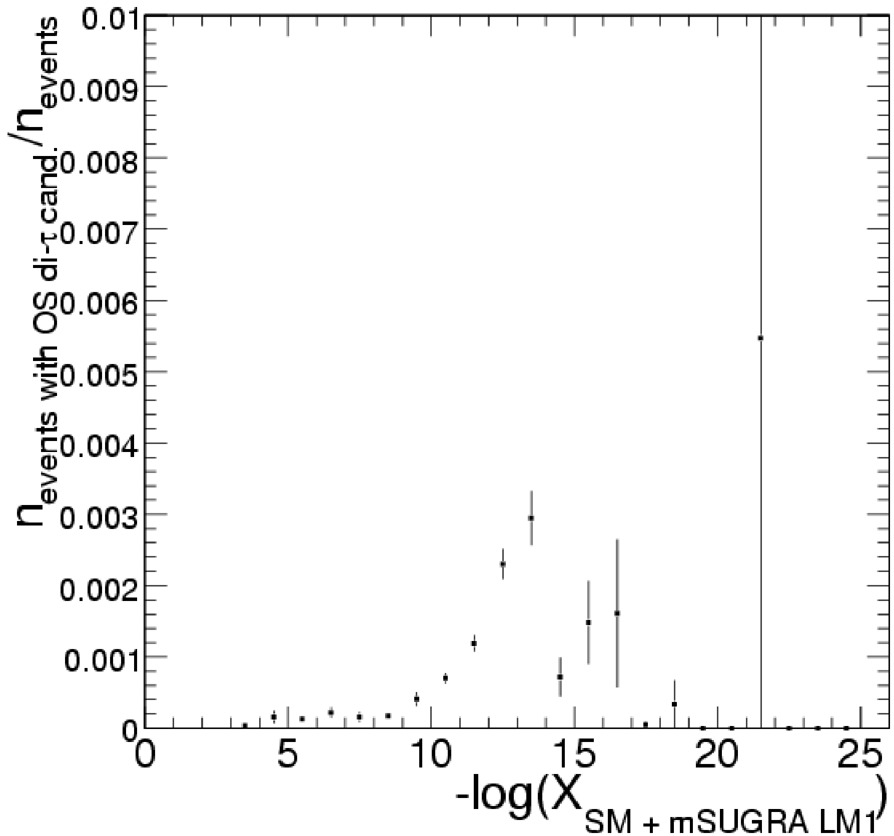
$\int L dt$ = 10fb^{-1}

- at *LM1* (test point defined by $m_{1/2}$=250GeV/c², m_0=60GeV/c², A_0=0, tanβ=10, sign(μ)=+), expected LO σ = 49.00 pb , BR($\tilde{\chi}_2^0 \rightarrow \tilde{\tau}\tau$) ≈ 0.46,
  15.0% of the mSUGRA events contain at least 1 signal cascade.

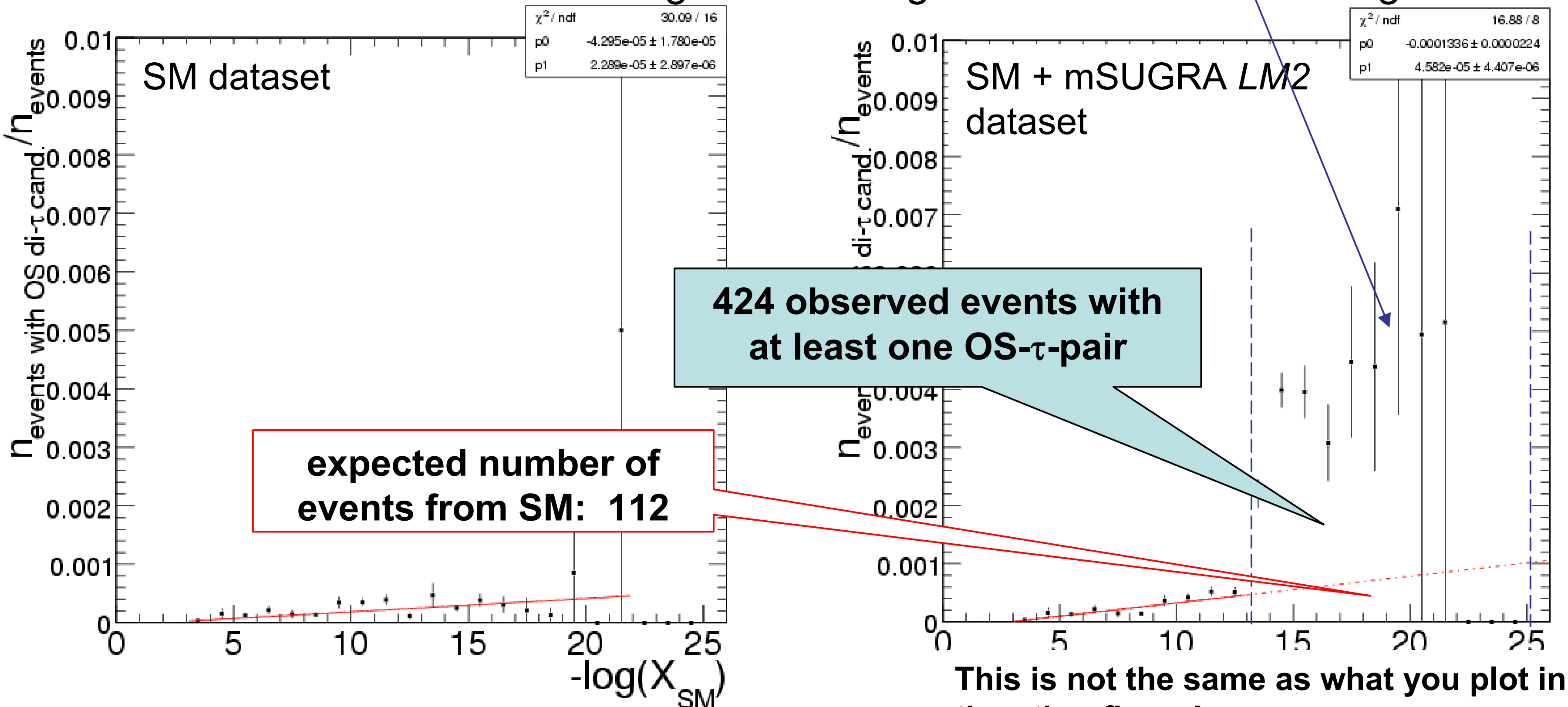
SM dataset



SM + mSUGRA *LM1* dataset



An estimator of the significance of the observed peak in the SM+mSUGRA *LM2* histogram resulting of the division of 2 histograms



This is not the same as what you plot in the other figure!

-in the kinematically most particular region (in a ($E_T^{2nd\ q/g\text{-}jet\ cand.}$, E_T^{miss})) space),
no excess of events containing more rec. OS charge hadr. τ-jet cand. pairs than the mass of the events,
-small dependency between the kinematic variable X_{SM} and the fraction of events with ≥1 OS di-tau cand.-related variable.

From the upper-right plot, -in the SM+mSUGRA LM2 sample- :
-we define a $-log(X_{SM+mSUGRA\ LM2})$ signal region : $13 \leq -log(X_{SM+mSUGRA\ LM2}) < 25$
-we estimate SM fraction of events with ≥1 OS di-tau cand.-related variable in signal region : the value given by the fit at $-log(X_{SM+mSUGRA\ LM2})=19$ (middle of the region) + its error : 8.41×10^{-4}

… an estimator of the significance of the observed peak
in the SM+mSUGRA *LM2* histogram resulting of the division of 2 histograms

- The observed number of SM+mSUGRA *LM2* events with $13 \leq -log(X_{\text{SM+mSUGRA } LM2}) < 25$ is equal to 133284.

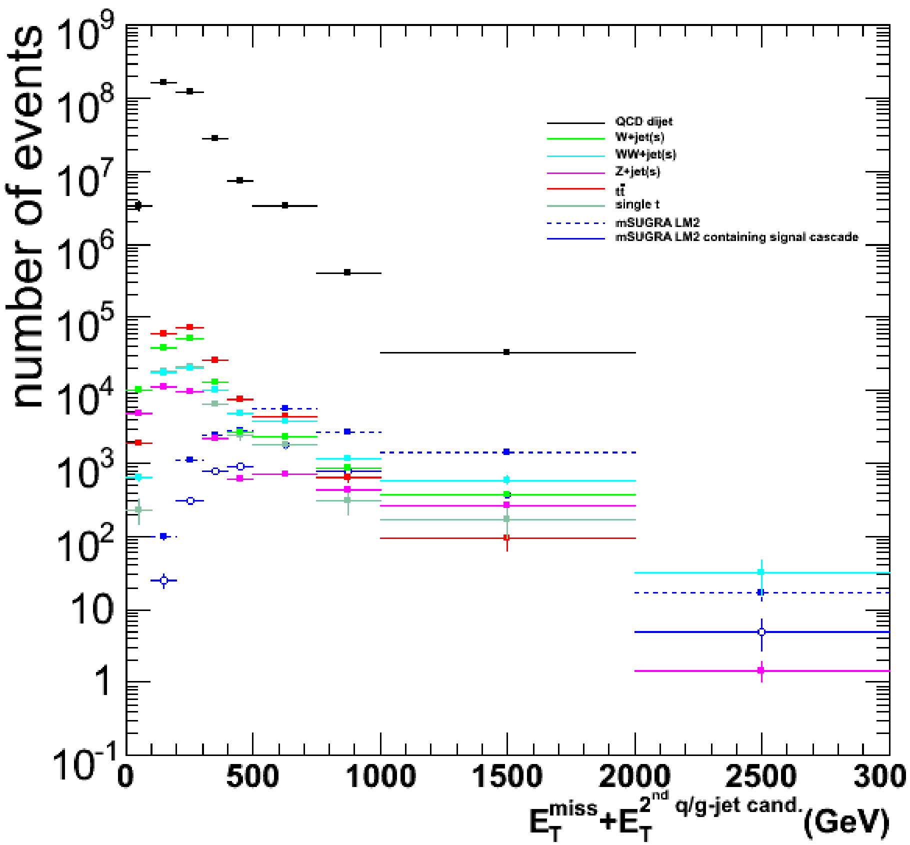
- The expected SM number of events with $13 \leq -log(X_{\text{SM+mSUGRA } LM2}) < 25$ and ≥1 rec. OS charge hadr. τ-jet cand. pair(s) is set equal to $8.41 \times 10^{-4} \times 133284 \approx 112$.
  The observed SM+mSUGRA *LM2* number of events with $13 \leq -log(X_{\text{SM+mSUGRA } LM2}) < 25$ and ≥1 rec. OS charge hadr. τ-jet cand. pair(s), is equal to 424.

Method B Using a control region to estimate SM contribution

$\int L dt = 10 fb^{-1}$

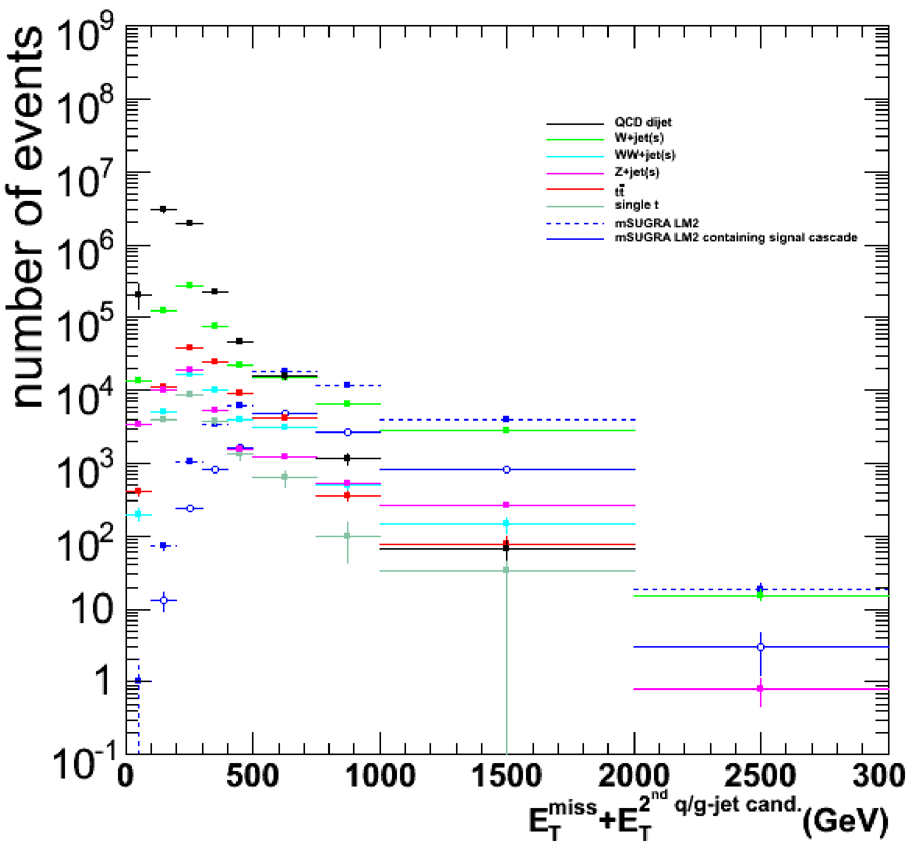
control region

i.e. $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet\ cand.} < 0GeV$



signal region

i.e. $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet\ cand.} > 0GeV$

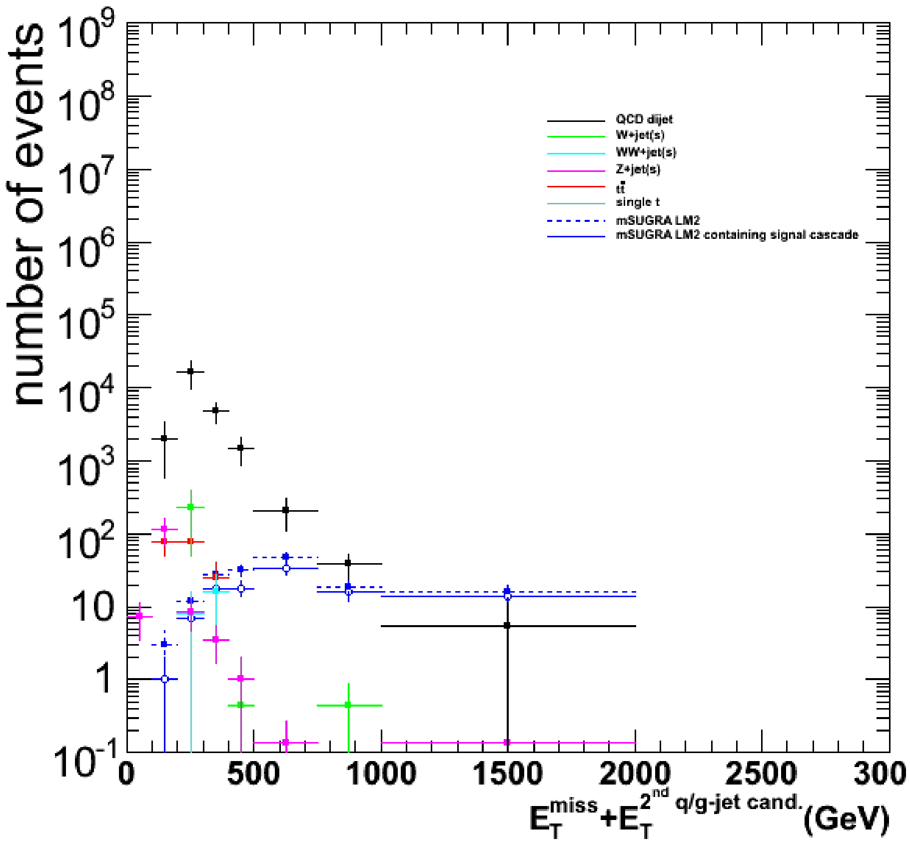


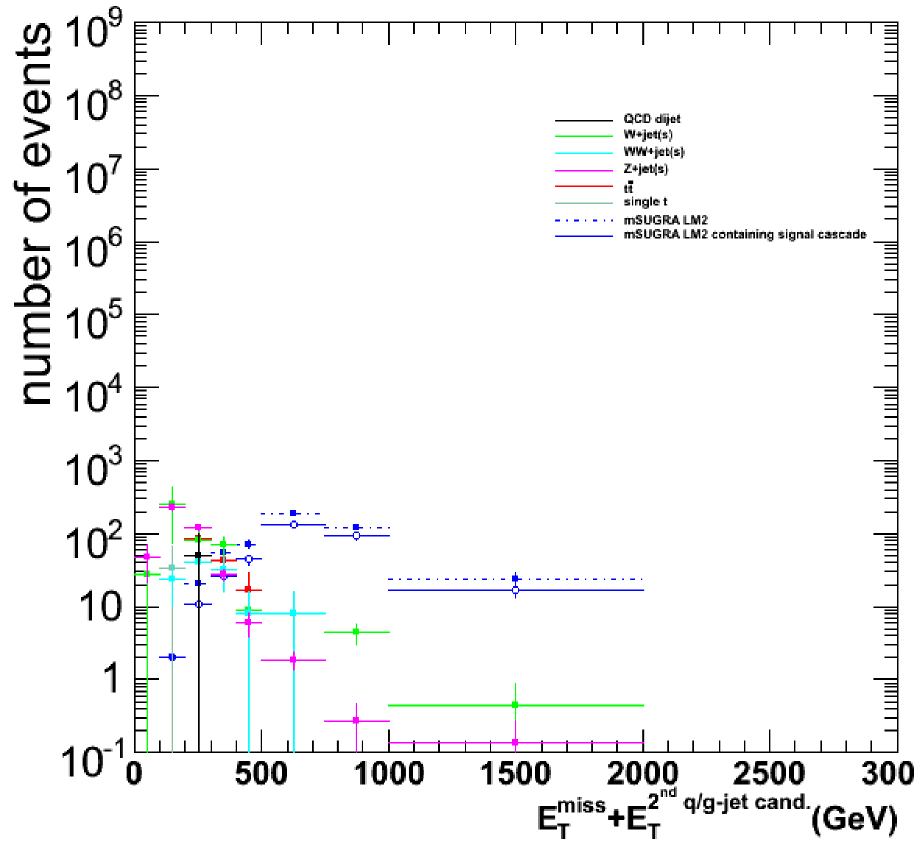
Method B Using a control region to estimate SM contribution

$\int L dt = 10fb^{-1}$

events with $n_{OS\ di\tau\ cand's} > 0$

control region	**signal region**
i.e. $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet\ cand.} < 0GeV$	*i.e.* $E_T^{miss} - E_T^{2nd\ q/g\text{-}jet\ cand.} > 0GeV$

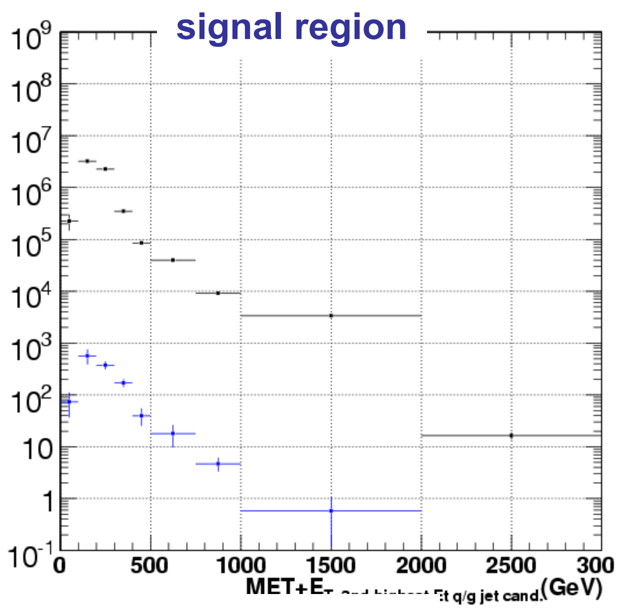


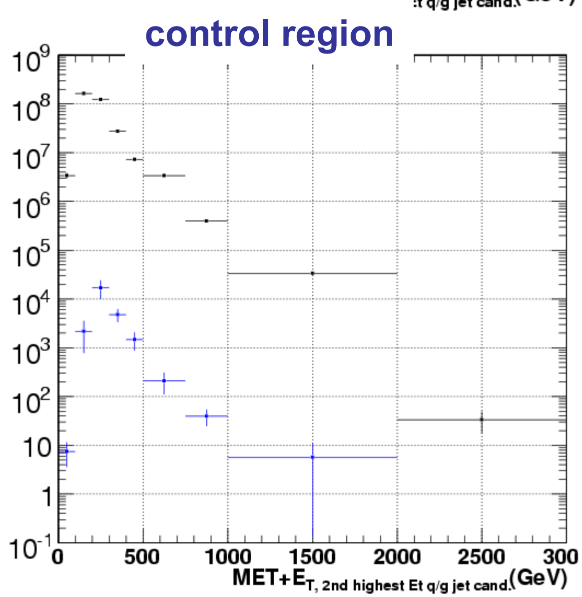


- We observe again the number of events with at least 1 di-tau as a function of a the kinematical variable $E_T^{miss} + E_T^{2nd\ q/g\text{-}jet}$,

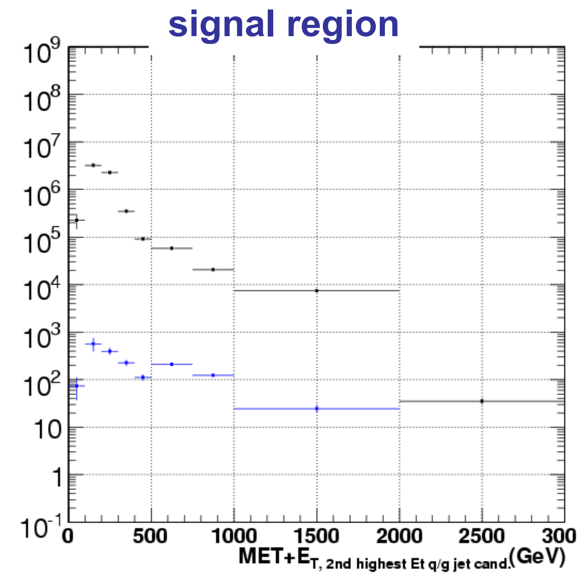
Method B Using a control region to estimate SM contribution

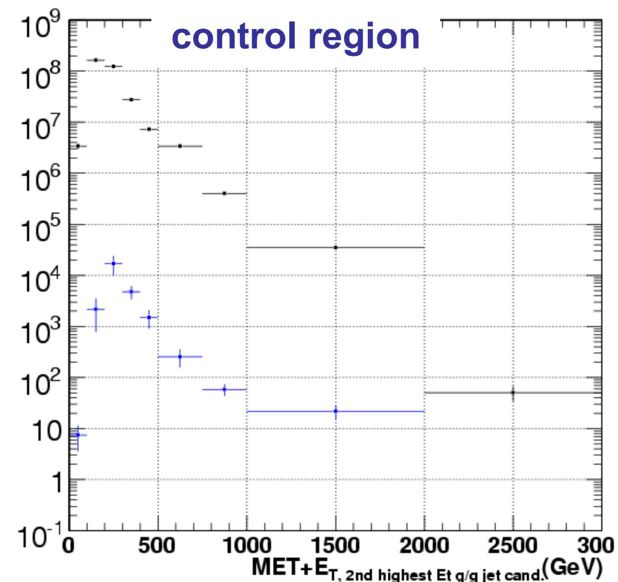
SM dataset





SM + mSUGRA *LM2* dataset $\int \mathcal{L}dt = 10fb^{-1}$



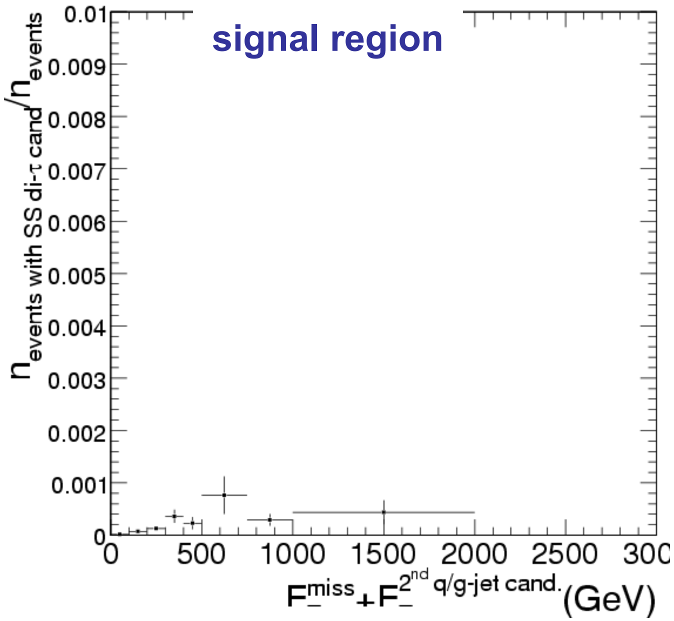


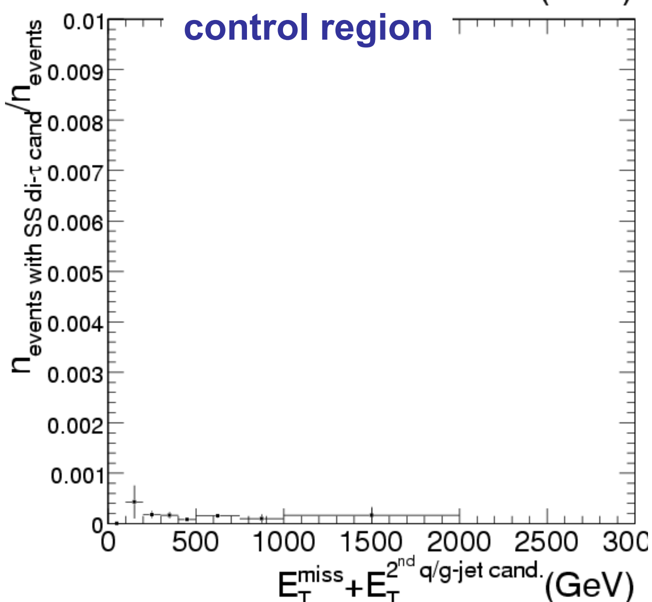
- ● all events
- ● events with $n_{OS\ di\tau\ cand's} > 0$

Method B for observing mSUGRA events in data

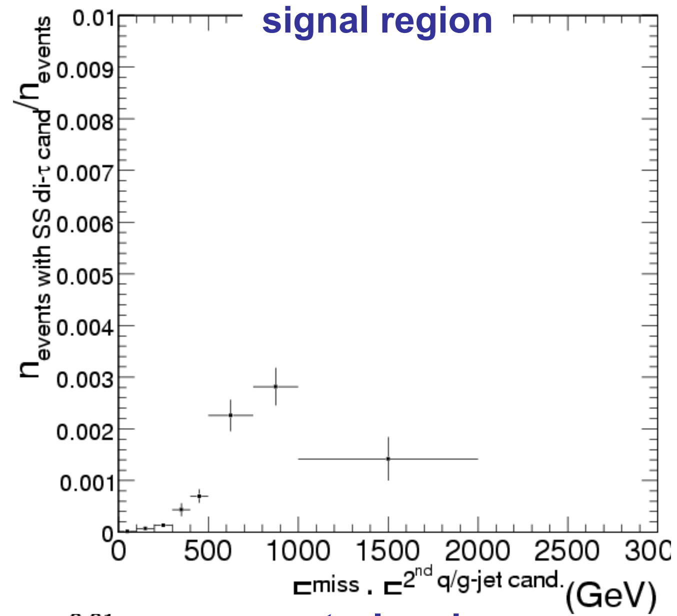
When considering SS charge rec. hadr. τ-jet cand. pairs, $\int \mathcal{L} dt = 10 fb^{-1}$

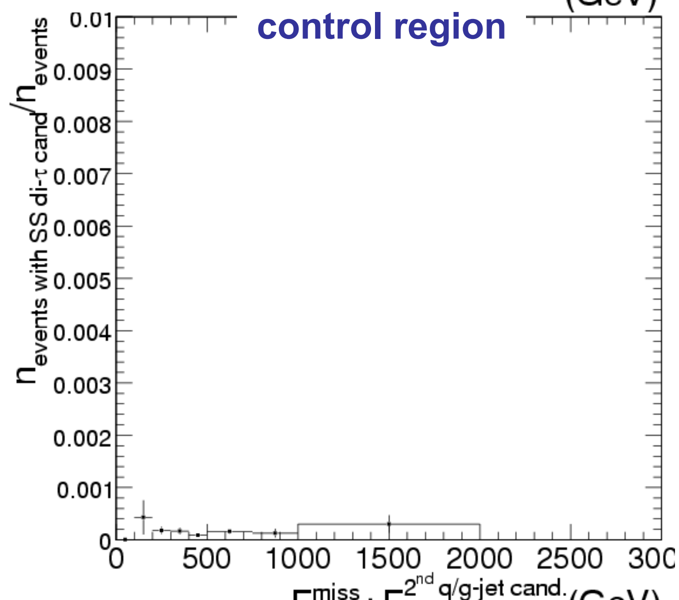
SM dataset





SM + mSUGRA *LM2* dataset

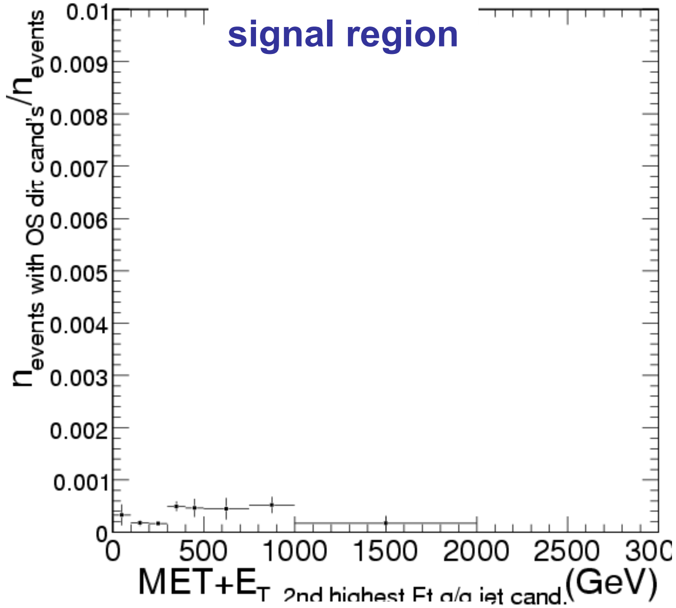


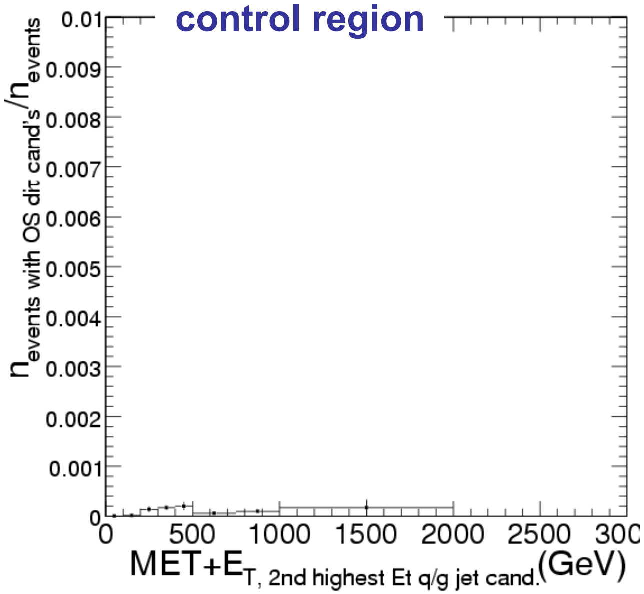


Method B for observing mSUGRA events in data

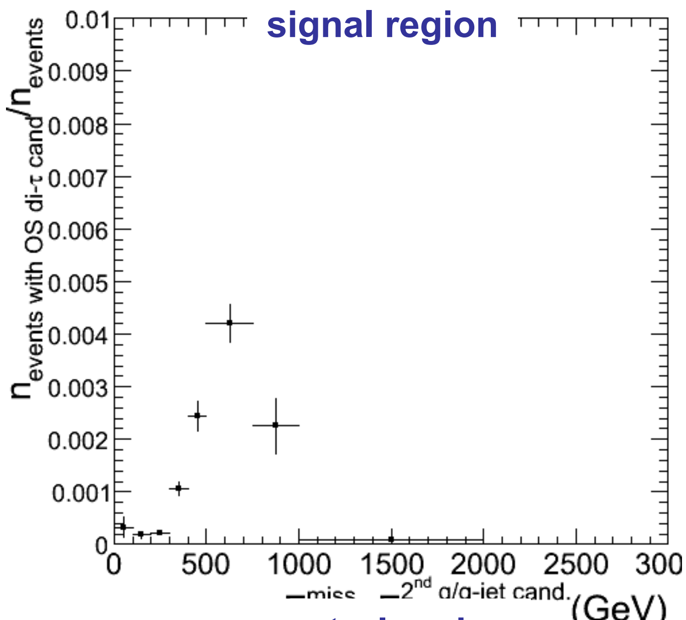
$\int L dt = 10 fb^{-1}$

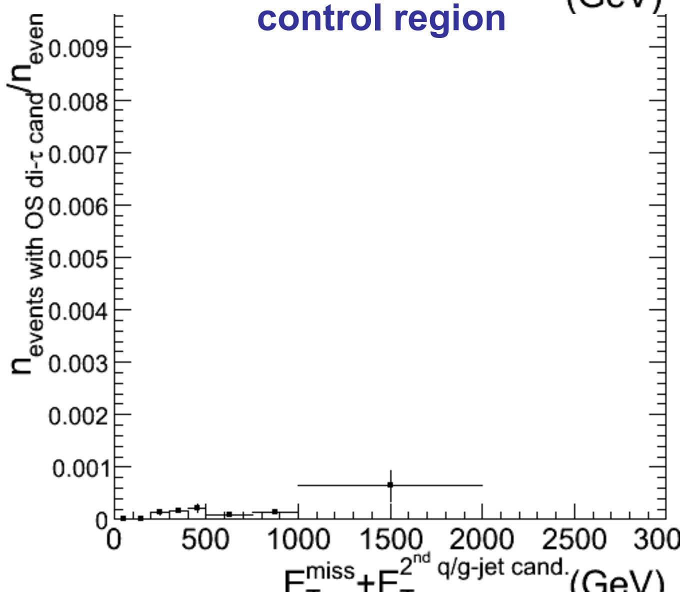
SM dataset





SM + mSUGRA *LM1* dataset





In case the SM+mSUGRA *LM2* event sample is the one observed,

we estimate an expected fraction of SM process events with OS di-τ as a function of $E_T^{miss} + E_T^{2nd\ q/g\text{-}jet\ cand.}$ in signal region :

per bin in $E_T^{miss} + E_T^{2nd\ q/g\text{-}jet\ cand.}$:

observed $\dfrac{\textbf{number of SM + LM2 OS di-}\tau}{\textbf{number of SM + LM2 events}}$ | control region

X

expected (through MC) $\dfrac{\textbf{number of SM OS di-}\tau}{\textbf{number of SM events}}$ | signal region

/

expected (through MC) $\dfrac{\textbf{number of SM OS di-}\tau}{\textbf{number of SM events}}$ | control region

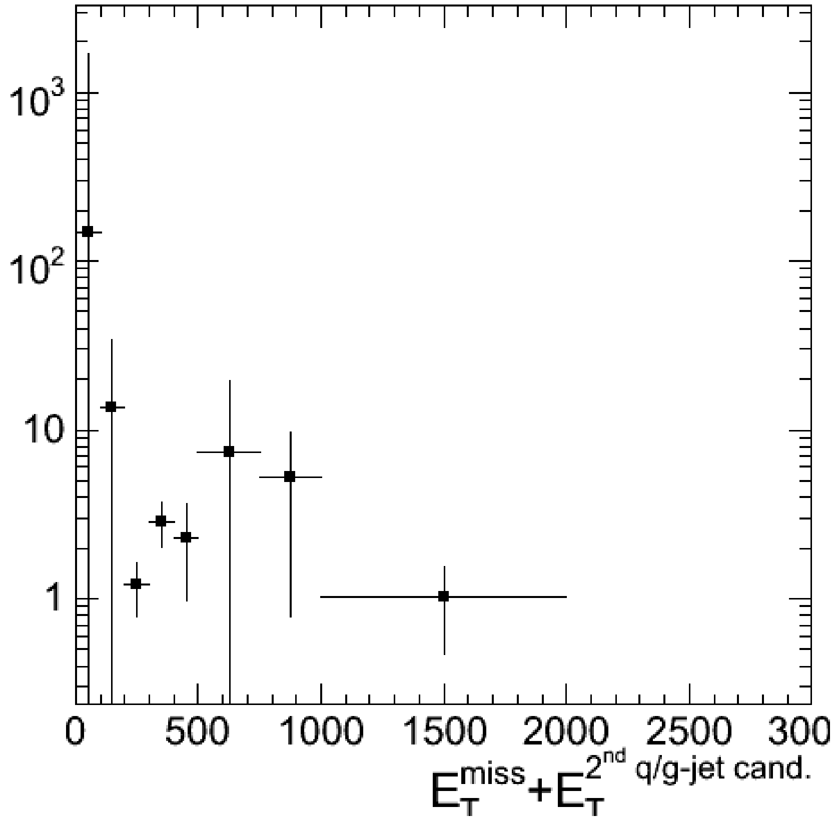
SM correction function based on MC ratio of signal to control regions

=

expected (through MC+DATA) $\dfrac{\textbf{number of SM OS di-}\tau}{\textbf{number of SM events}}$ | signal region

$\int L dt = 10 fb^{-1}$

The correction function



We define 2 regions (control region, dominated by SM events and signal region, dominated by SUSY events) using $\Delta\Phi_{\text{MET-1st highest ET calo. jet}} + \Delta\Phi_{\text{MET-2nd highest ET calo. jet}}$ <3.5 or >3.5, respectively