SPONSORED BY THE
Federal Ministry of Education and Research

Phenomenology Symposium 2017, Pittsburgh
May 9th 2017

MUSiC - Model Unspecific Search in CMS (CMS PAS EXO-14-016)

Deborah Duchardt
For the CMS Collaboration

Physics Institute IIIA, RWTH Aachen University

Physics Institute III A
RWTH AACHEN UNIVERSITY

Motivation

Challenge at the LHC

Many competing models!
Difficult to have dedicated analyses for all.
Are there models no one has thought of yet?

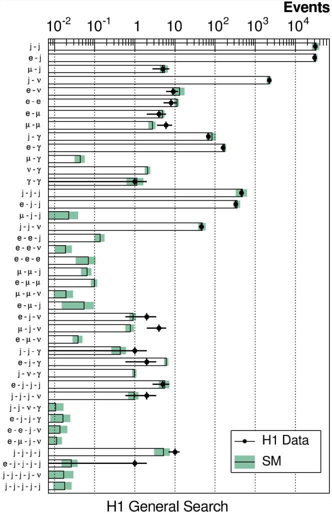
Alternative Approach

Model independent search!

- Assume only: **The Standard Model**.
- Search for deviations in many final states.

Previous Model Independent Searches

Performed at: L3, H1, D0 (SLEUTH), CDF (VISTA), **ATLAS**, **CMS**, ...



arXiv:hep-ex/0408044

Scope of MUSiC

What is MUSiC?

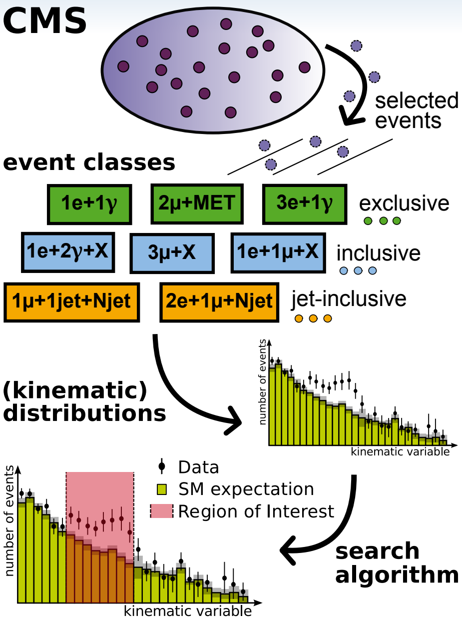
- A physics and detector monitor.
- An alarm system.
- Complementary to dedicated analyses.

What if MUSiC finds something?

- New physics?
- Inadequate detector simulation?
- Inadequate Monte-Carlo: tuning, missing (rare) processes, etc.?

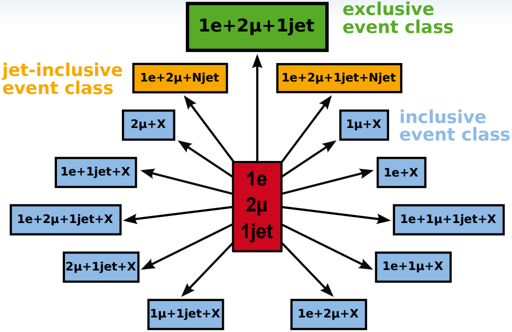
$\Longrightarrow$ Careful investigation is essential!

MUSiC Workflow



- Sort data and MC into **event classes** by physics object content of final state: e, μ, γ, jet, MET
- **Three distributions** per class: $\sum |\vec{p}_\mathrm{T}|$, $M_{(\mathrm{T})}$, MET
- Most significant region in each distribution: **Region of Interest**
- Determine look-elsewhere effect corrected ***p*-value** ($\tilde{p}$) for each event class distribution
- Compare distribution of $\tilde{p}$-values from data with SM MC

Event Classes



After Applying Selection Criteria

- Each **event** ends up in exactly one **exclusive** and in general in several **jet-inclusive** as well as **inclusive** event classes.
- Event classes with at least 6 jets are combined, e.g. $1\mu+\geq 6$jets.
- At 8 TeV: 337 (341, 321) exclusive (inclusive, jet-inclusive) event classes. Around 50% of event classes contain data.

Selection at 8 TeV

Event Selection

Use inclusive lepton triggers to select events with at least one lepton:

≥ 1 **electron** with $p_T > 100\,\text{GeV}$
OR
≥ 2 **electrons** with $p_T > 25\,\text{GeV}$

≥ 1 **muon** with $p_T > 30\,\text{GeV}$
OR
≥ 2 **muons** with $p_T > 25\,\text{GeV}$

Object Selection

For all subsequent objects use well studied identification criteria:

Object	p_T/GeV	Identification Summary
e	> 25	track quality, isolation, calo signature
μ	> 25	track quality, isolation
γ	> 25	isolation, veto against e
jet	> 50	anti-k_t algorithm $(R = 0.5)$
MET	> 50	-

Systematic Uncertainties at 8 TeV

Apply to Simulated Background Events

Contribution	Value	Remarks
Luminosity	2.6 %	-
Parton density functions	varies	PDF4LHC recipe
Energy/momentum scales	1 % to 10 %	per object
MC Reconstruction efficiencies	1 % to 3 %	per object
MC Misidentification rates	50 % to 100 %	per object
SM cross sections	5 % to 50 %	process dependent
Number of MC events	varies	relative value $< 60\%$

MUSiC's p-value

Use a **Gaussian prior** to model the systematic uncertainties on the mean of a *Poisson counting experiment*.

$$p = \begin{cases} \underbrace{\sum_{i=N_{\text{obs}}}^{\infty} C}_{\text{normalization}} \cdot \int_0^\infty \mathrm{d}\theta \underbrace{\exp\left(-\frac{(\theta - N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right)}_{\text{systematics}} \cdot \underbrace{\frac{\theta^i\,\mathrm{e}^{-\theta}}{i!}}_{\text{statistics}} & \text{if } N_{\text{obs}} \geq N_{\text{SM}} \\ \sum_{i=0}^{N_{\text{obs}}} C \cdot \int_0^\infty \mathrm{d}\theta \exp\left(-\frac{(\theta - N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right) \cdot \frac{\theta^i\,\mathrm{e}^{-\theta}}{i!} & \text{if } N_{\text{obs}} < N_{\text{SM}} \end{cases}$$

with:

N_{obs} : Observed events in data

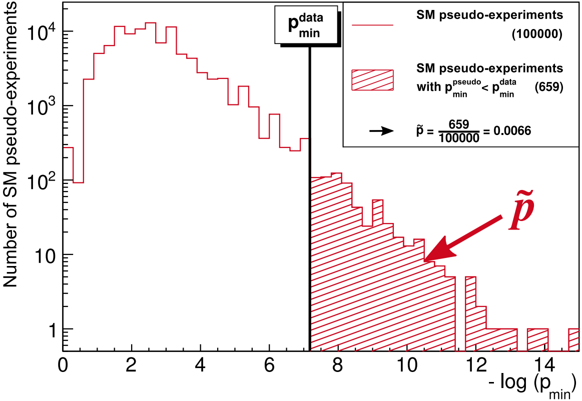
N_{SM} : Pure SM (Monte Carlo) expectation in region

σ_{SM} : Uncertainty on the SM MC prediction

Region with the smallest p-value: **Region of Interest**

"Look-Elsewhere Effect"

Considering **many regions** in each distribution makes it more probable to see a large deviation somewhere by chance due to statistical fluctuations.

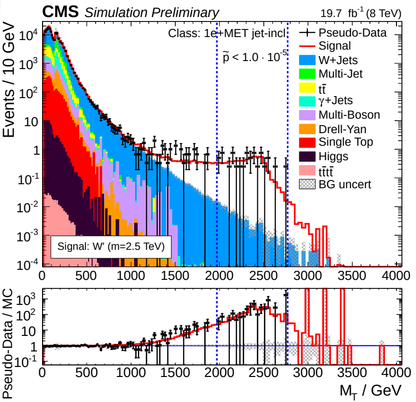


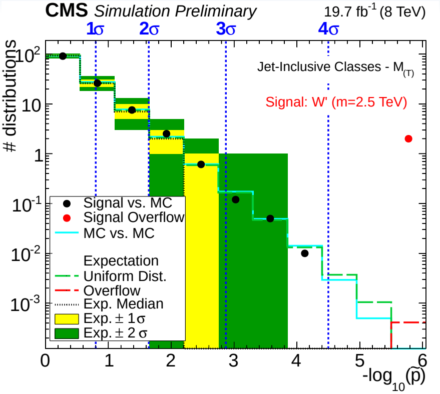
- Randomize SM MC expectation bin by bin, accounting for known uncertainties.
- Determine most significant region in pseudo data.

New (post-trial) significance:

$$\tilde{p} = \frac{\text{number of pseudo experiments with } p^{\text{pseudo}}_{\text{min}} < p^{\text{data}}_{\text{min}}}{\text{total number of pseudo experiments}}$$

Sensitivity to BSM Signals

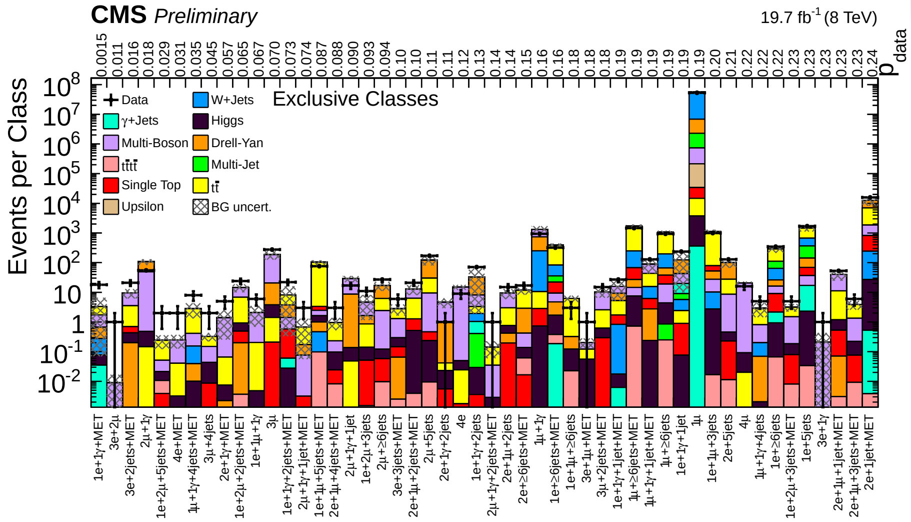




M_T of Jet-Inclusive Event Class

- Assume hypothetial leptonically decaying SSM W' boson and generate pseudo-data according to SM+W', compare with SM.
- On average 2 very significant event classes found for m(W') = 2.5 TeV.

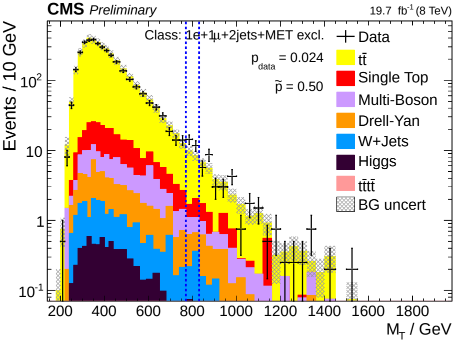
Total Event Yield Per Event Class

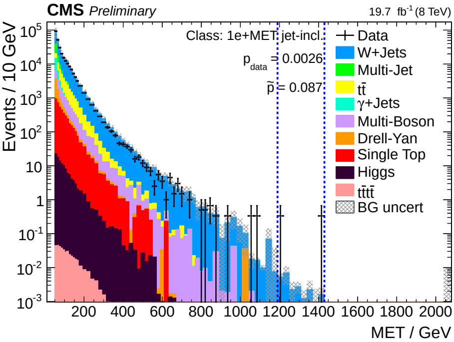


Exclusive Event Classes

All events condensed into a single bin, sorted here by significance.

Example Distributions





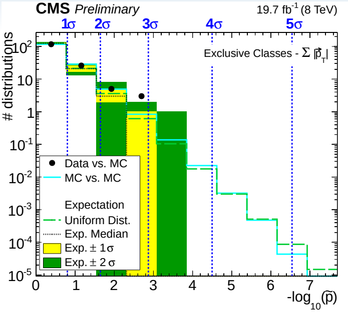
Exclusive Event Class

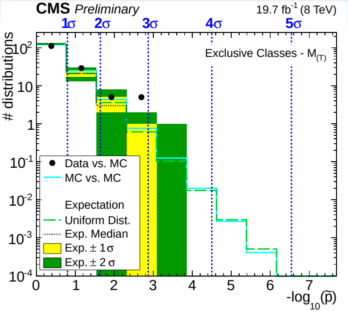
- Dominated by $t\bar{t}$ production
- Insignificant excess found in bulk

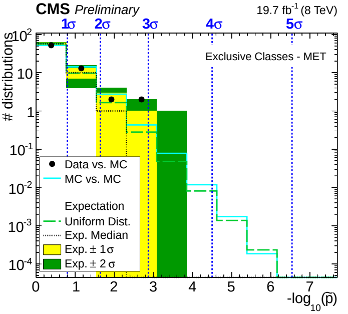
Jet-Inclusive Event Class

- Dominated by W+jets production
- Insignificant excess found in tail

Distribution of Deviations



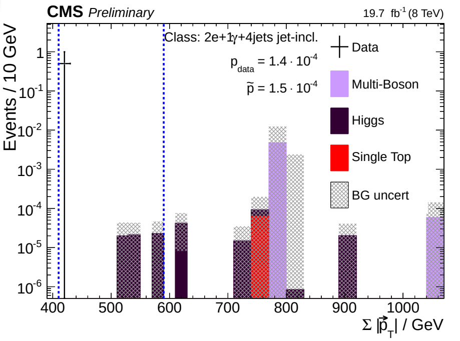


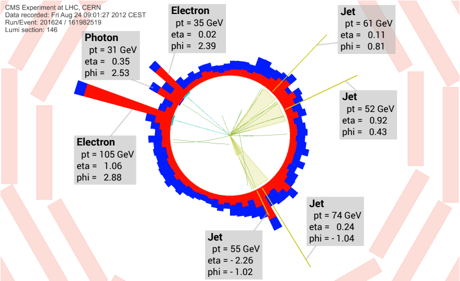


Global Overview

- Determined individually for each variable and event class type.
- Overall good agreement.
- Observed significances are compatible with SM fluctuations.

Most Significant Deviation





Jet-Inclusive Event Class

- Single data event and low SM MC event yield.
- $\sum |\vec{p}_{\mathrm{T}}| = 414\,\mathrm{GeV}$, $M_{(\mathrm{T})} = 718\,\mathrm{GeV}$, MET $= 27\,\mathrm{GeV}$
- Many vertices (30) $\rightarrow$ pileup
- $M_{\mathrm{inv}}(\mathrm{ee}\gamma) = 85\,\mathrm{GeV}$

Summary

MUSiC (CMS PAS EXO-14-016)

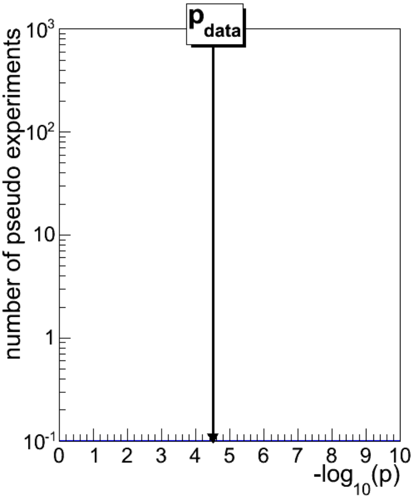
- Based on 8 TeV data in electron- and muon-triggered events, MUSiC has found no significant deviations beyond expected fluctuations.
- Good agreement between data and SM simulation over a wide range of final states.
- Sensitivity studies provide proof of MUSiC's discovery potential for specific signatures.

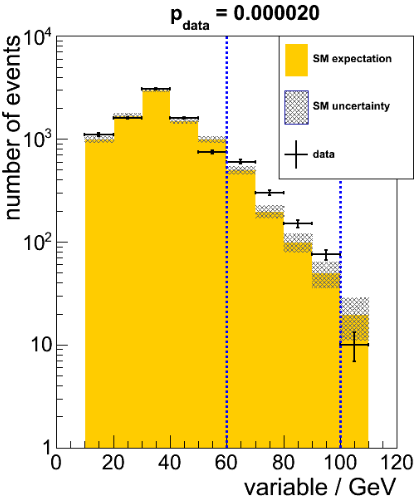
Outlook

- Expand analysis scope: photon and jet triggered events, b-tagged jets.
- Analysis of data collected at 13 TeV is ongoing.
- Stay tuned for results and updates on most significant deviations in 8 TeV data.

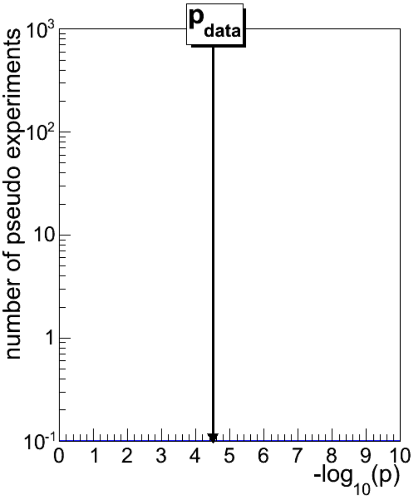
Additional Information

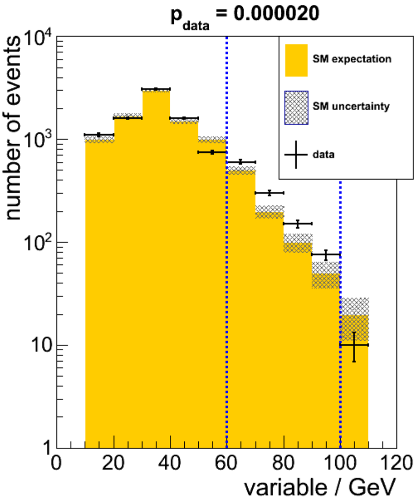
Dicing of Pseudo-Experiments



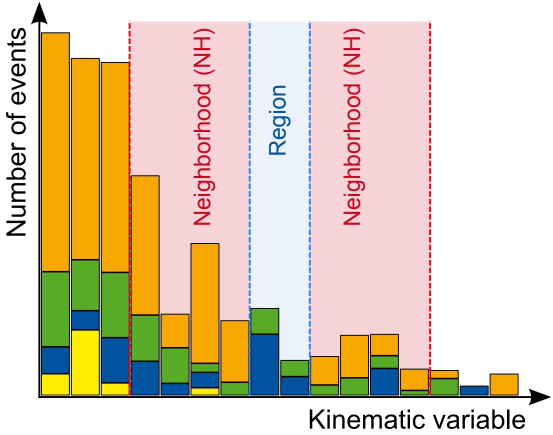


Dicing of Pseudo-Experiments





Region-Veto Method



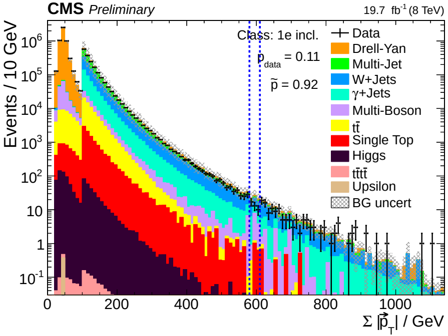
SM MC Data sets

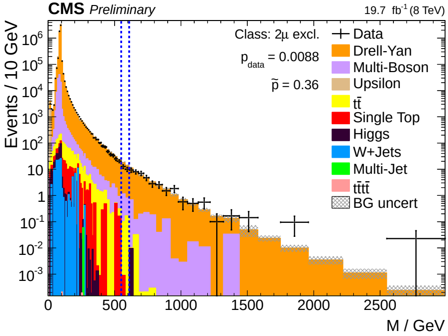
Process	Details	Generator	k-factor (order)	# Events
γ+Jets		PYTHIA6	1.3 (NLO)	$1.6 \cdot 10^6$
W+Jets		MADGRAPH	1.21 (NNLO)	$6.0 \cdot 10^7$
	$M_{l^\pm\nu} > 200$ GeV	MADGRAPH	1.21 (NNLO)	$1.2 \cdot 10^6$
	$H_T > 150$ GeV	MADGRAPH	1.21 (NNLO)	$4.3 \cdot 10^7$
	Wbb $\rightarrow l^\pm \nu$bb	MADGRAPH	1.79 (NLO)	$2.0 \cdot 10^7$
Drell-Yan	$M_{l^\pm l^\pm} > 10$ GeV	MADGRAPH	1.2 (NNLO)	$6.0 \cdot 10^7$
	$p_T(Z) > 50$ GeV	MADGRAPH	1.0 (LO)	$3.8 \cdot 10^8$
	$M_{l^\pm l^\pm} > 200$ GeV	POWHEG	1.02 (NNLO)	$2.2 \cdot 10^6$
	Z+Jets $\rightarrow \nu\nu$	MADGRAPH	1.0 (LO)	$3.0 \cdot 10^7$
Multi-Boson	$\gamma\gamma$	PYTHIA6	1.0 (LO)	$7.7 \cdot 10^6$
	WW	PYTHIA6	1.63 (NLO)	$1.0 \cdot 10^7$
	$W^\pm W^\pm$	MADGRAPH	1.0 (LO)	$1.8 \cdot 10^5$
	WW via DPS	PYTHIA6	1.0 (LO)	$8.0 \cdot 10^5$
	gg $\rightarrow$ WW	gg2VV	1.4 (NLO)	$1.0 \cdot 10^5$
	ZZ $\rightarrow 4l^\pm$	POWHEG	1.0 (NLO)	$1.5 \cdot 10^7$
	ZZ $\rightarrow 2l2$q	MADGRAPH	1.0 (LO)	$2.7 \cdot 10^6$
	gg $\rightarrow$ ZZ	gg2VV	1.0 (NLO)	$9.0 \cdot 10^5$
	WZ	MADGRAPH	1.0 (LO)	$7.9 \cdot 10^6$
	Vγ	MADGRAPH	1.2 (NLO)	$1.0 \cdot 10^7$
	WWW	MADGRAPH	0.98 (NLO)	$2.0 \cdot 10^5$
	ZZZ	MADGRAPH	1.2 (NLO)	$2.0 \cdot 10^5$
	WWZ	MADGRAPH	0.92 (NLO)	$2.0 \cdot 10^5$
	WZZ	MADGRAPH	1.02 (NLO)	$2.0 \cdot 10^5$
	WWγ	MADGRAPH	1.0 (LO)	$5.0 \cdot 10^5$
	W$\gamma\gamma$	MADGRAPH	1.0 (LO)	$2.0 \cdot 10^6$

SM MC Data sets

Multi-Jet		PYTHIA6	1.0 (LO)	$4.5 \cdot 10^8$
Single Top		POWHEG	1.04-1.34 (NNLL)	$5.2 \cdot 10^6$
$t\bar{t}$		POWHEG	1.0 (NNLO)	$2.6 \cdot 10^7$
	$t\bar{t}$ ($M_{tt} > 700$ GeV)	POWHEG	1.11 (NNLO)	$4.0 \cdot 10^6$
	$t\bar{t}\gamma$	MADGRAPH	1.0 (LO)	$1.0 \cdot 10^6$
	$t\bar{t}W$	MADGRAPH	1.08 (NLO)	$1.0 \cdot 10^5$
	$t\bar{t}Z$	MADGRAPH	1.2 (NLO)	$2.0 \cdot 10^5$
	$t\bar{t}WW$	MADGRAPH	1.0 (LO)	$2.0 \cdot 10^5$
$t\bar{t}t\bar{t}$		MADGRAPH	1.28 (NLO)	$9.0 \cdot 10^4$
Higgs	$gg \rightarrow ZZ$	MiNLO	1.0 (NLO)	$8.0 \cdot 10^5$
	$gg \rightarrow bb, \gamma\gamma, \tau\tau, Z\gamma, WW$	POWHEG	1.0 (NLO)	$2.6 \cdot 10^6$
	$VBF \rightarrow bb, \gamma\gamma, \tau\tau, Z\gamma, VV$	POWHEG	1.0 (NLO)	$2.5 \cdot 10^6$
	$WH \rightarrow \mu\mu, bb$	POWHEG	1.0 (NLO)	$1.1 \cdot 10^6$
	$ZH \rightarrow \mu\mu, bb$	POWHEG	1.0 (NLO)	$2.9 \cdot 10^6$
	$TTH \rightarrow bb, \gamma\gamma, Z\gamma$	PYTHIA6	1.0 (NLO)	$1.2 \cdot 10^6$
	$WH, ZH \rightarrow \gamma\gamma, Z\gamma$	PYTHIA6	1.0 (NLO)	$1.1 \cdot 10^6$
	$WH, ZH, TTH \rightarrow \tau\tau, VV$	PYTHIA6	1.0 (NLO)	$9.0 \cdot 10^5$
Upsilon	$\Upsilon(1S)/\Upsilon(2S)/\Upsilon(3S) \rightarrow \mu\mu$	EVTGEN	1.0 (LO)	$5.0 \cdot 10^6$

Example Distributions





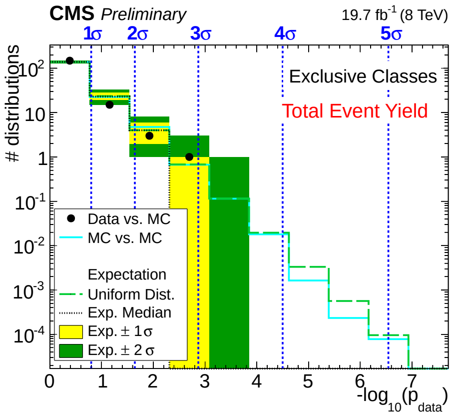
Jet-Inclusive Event Class

- Effect of different muon trigger thresholds visible
- Insignificant deficit identified in RoI

Exclusive Event Class

- Dominated by Drell-Yan production
- Insignificant deficit identified in RoI

Scan of Total Event Yield

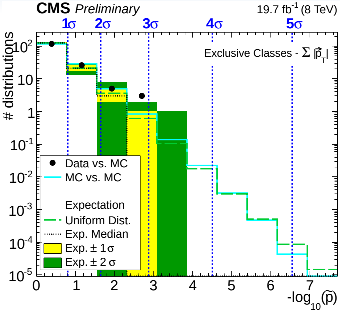


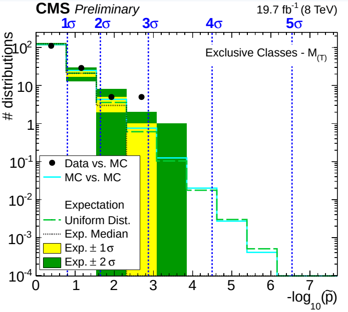
Yield in Exclusive Event Classes

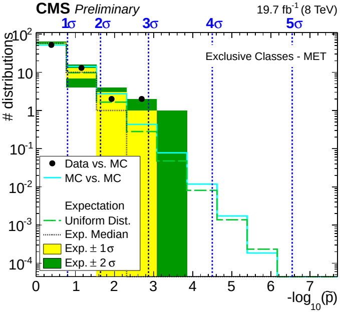
Global overview of all event classes:

- Determined individually for each kinematic variable and event class type.
- $\tilde{p}$-values from data vs. SM MC comparison (black markers).
- $\tilde{p}$-values from SM pseudo-data vs. SM MC comparison from LEE correction (blue line).
- Expectation assuming a uniform distribution given along with 68% and 95% interval.

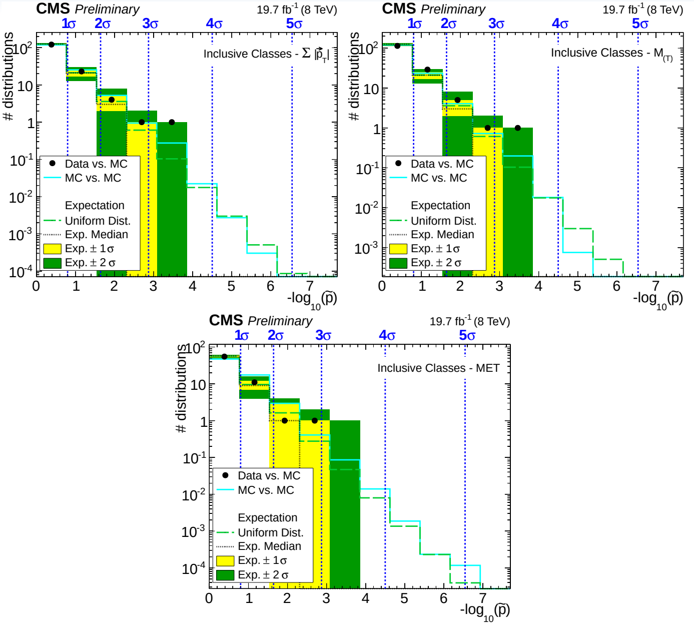
Distribution of $\tilde{p}$-values



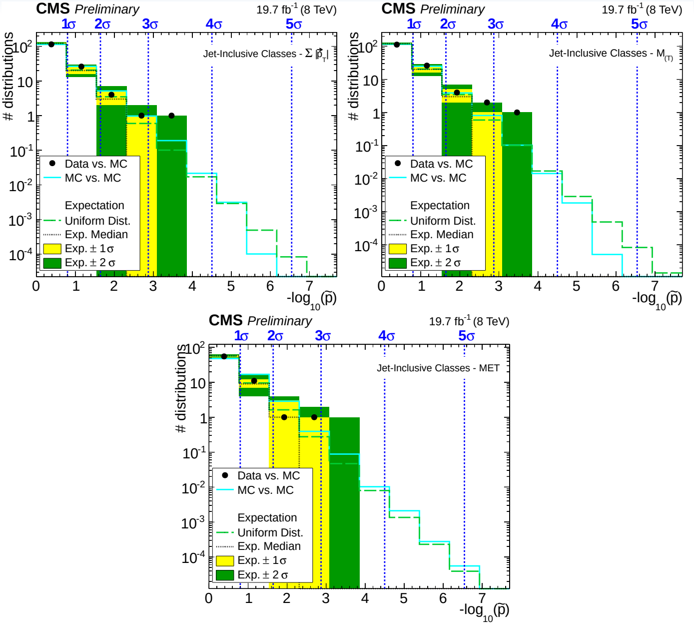




Distribution of $\tilde{p}$-values



Distribution of $\tilde{p}$-values

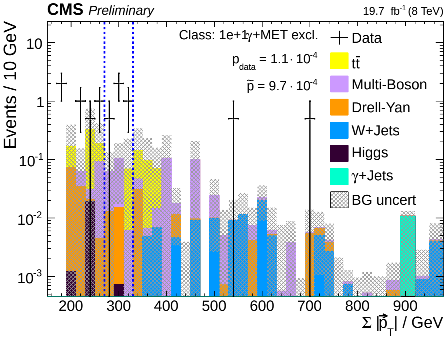


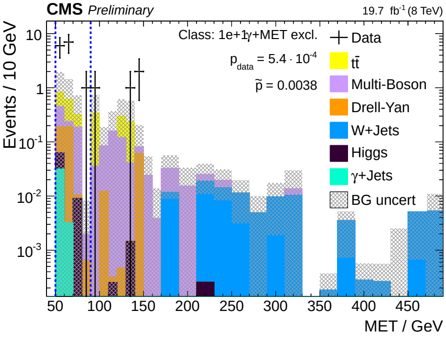
Most Significant Deviations

Table 3: Overview of the two most significant event classes from each scan type. Corresponding significance values are given in each case, where the p-value represents the full significance in the scan of total event yield. The $\tilde{p}$-value is stated for scans of kinematic distributions.

Scan type		Most significant		Second most significant	
		Event class	$\tilde{p}$ (p)	Event class	$\tilde{p}$ (p)
Total Event Yield	excl.	1e+1γ+MET	(0.0015)	3e+2μ	(0.011)
	incl.	3e+2jets+MET+X	(0.014)	3e+2μ+X	(0.019)
	jet-incl.	3e+2jets+MET+Njet	(0.012)	2μ+1γ+Njet	(0.015)
$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	excl.	1e+1γ+MET	0.00097	3e+2jets+MET	0.0027
	incl.	2e+1γ+4jets+X	0.00069	3e+2μ+X	0.0041
	jet-incl.	2e+1γ+4jets+Njet	0.00015	3e+2μ+Njet	0.0040
$M_{(T)}$	excl.	1e+1γ+MET	0.0020	1μ+1γ	0.0021
	incl.	2e+1γ+4jets+X	0.00071	2μ+1jet+X	0.0016
	jet-incl.	2e+1γ+4jets+Njet	0.00017	2μ+1jet+Njet	0.0014
MET	excl.	1e+1γ+MET	0.0038	2μ+1γ+1jet+MET	0.0039
	incl.	2μ+1γ+1jet+MET+X	0.0013	3e+2jets+MET+X	0.013
	jet-incl.	2μ+1γ+1jet+MET+Njet	0.0013	3e+2jets+MET+Njet	0.0095

Most Significant Deviations

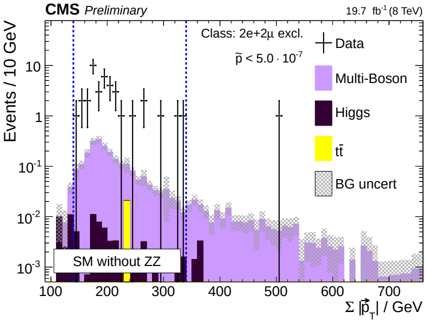


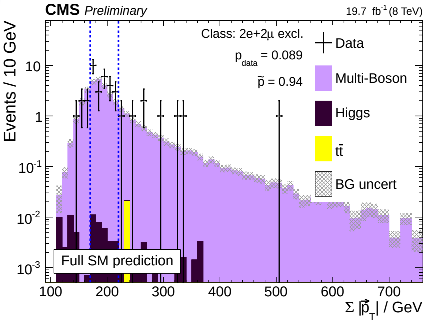


Exclusive Event Class

- Most significant exclusive event class.

Sensitivity to SM Processes

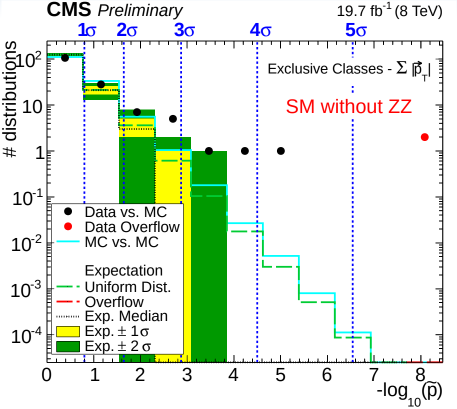




$\sum |\tilde{p}_t|$ of Exclusive Event Class

- Remove ZZ production from SM description.
- Perform comparison with measured 8 TeV data.
- Most significant deviation observed in 2e+2μ exclusive event class.
- Similar results for $M_{(T)}$, whereas MET (>50 GeV) is not sensitive.

Sensitivity to SM Processes



$\sum |\tilde{p}_t|$ of Exclusive Event Class

- Remove ZZ production from SM description.
- Perform comparison with measured 8 TeV data.
- Event classes with 4 leptons show strongest deviations.