Mining LHC Data

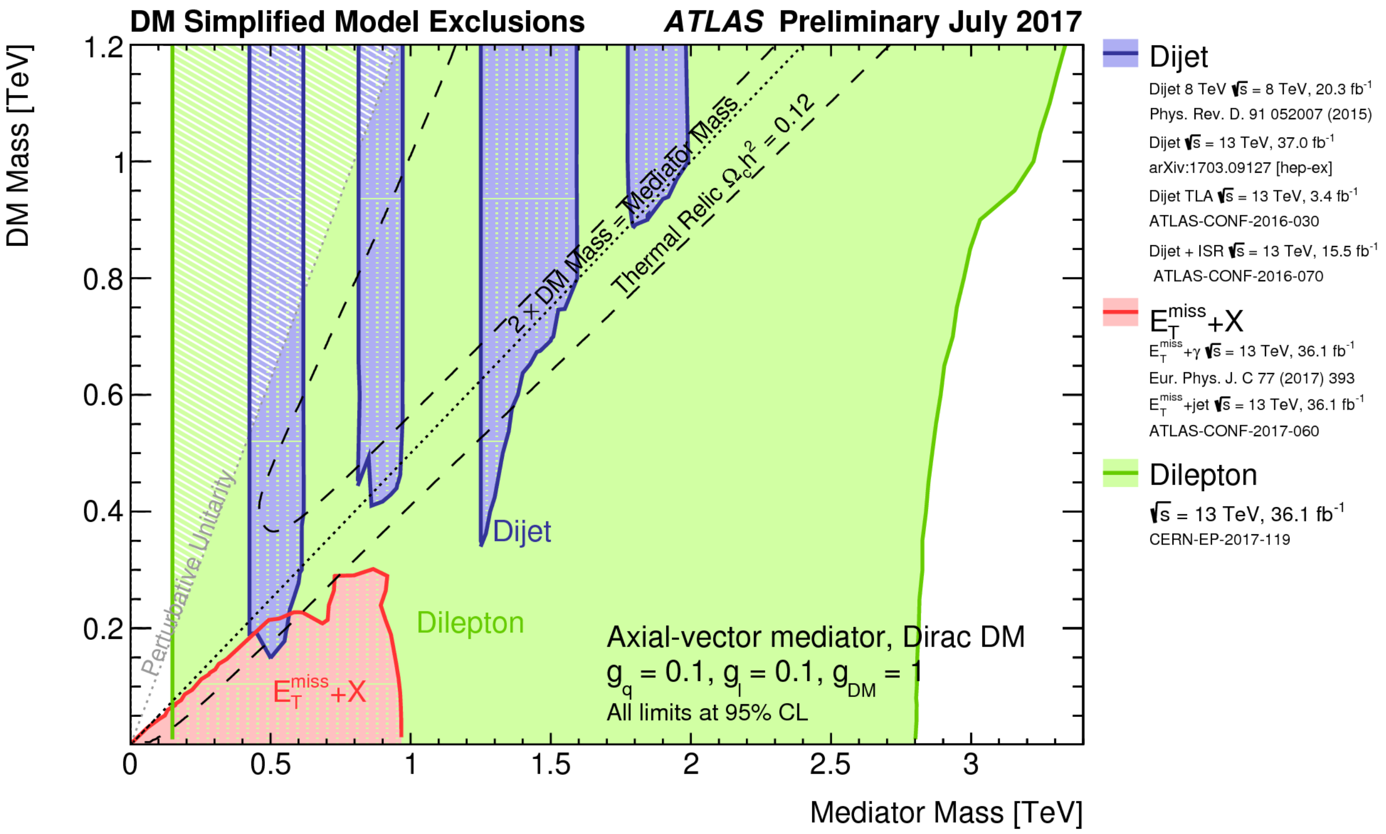
Matthew R Buckley
Rutgers University

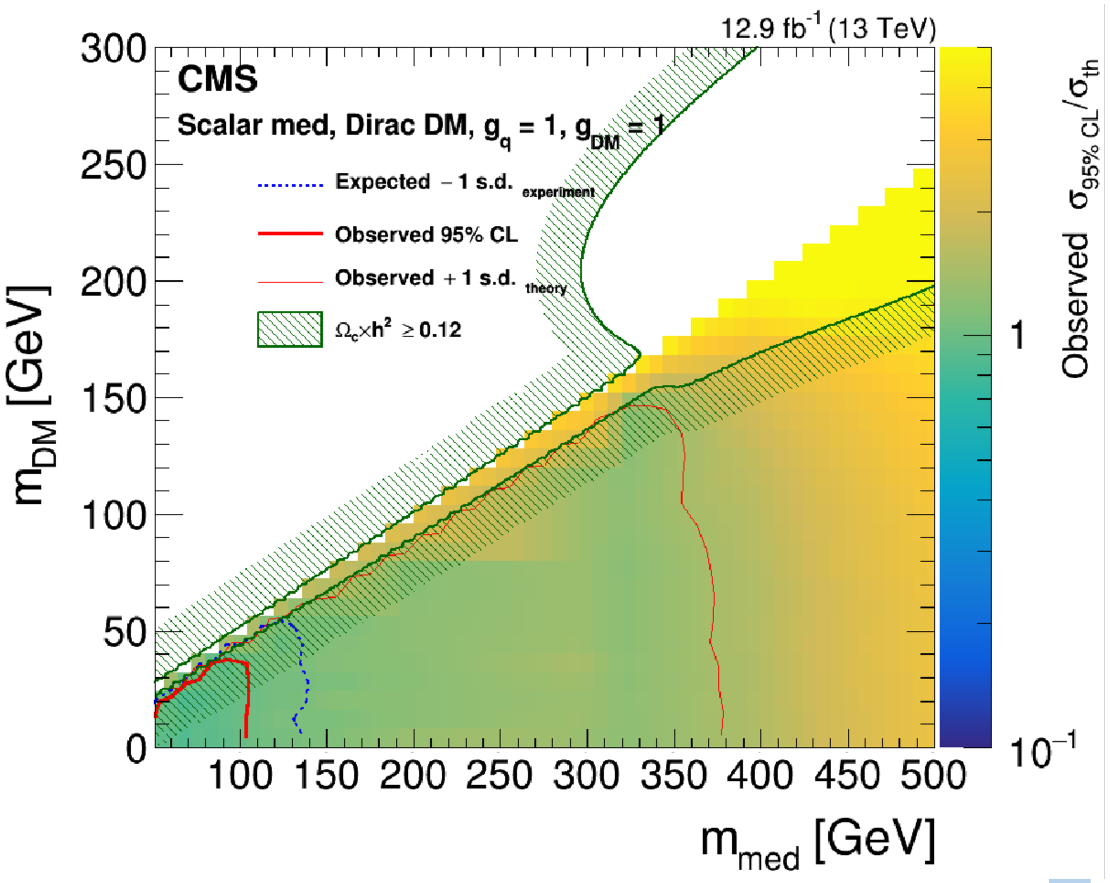
Future of collider searches for Dark Matter
LPC, July 2017

arxiv:1707.05783
P. Asadi, MRB, A. DiFranzo, A. Monteux, D. Shih

New Physics?

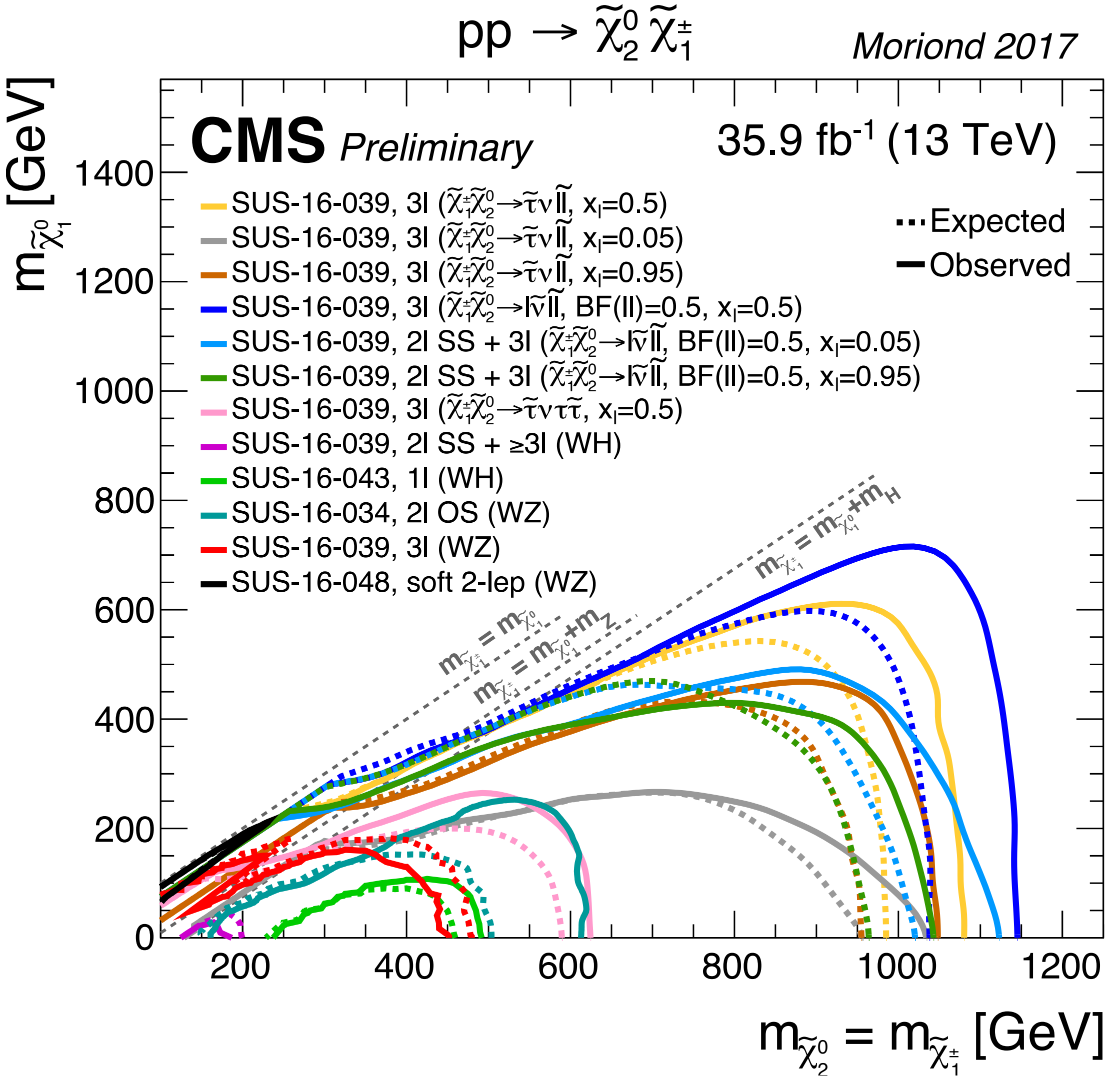
- Where is it?

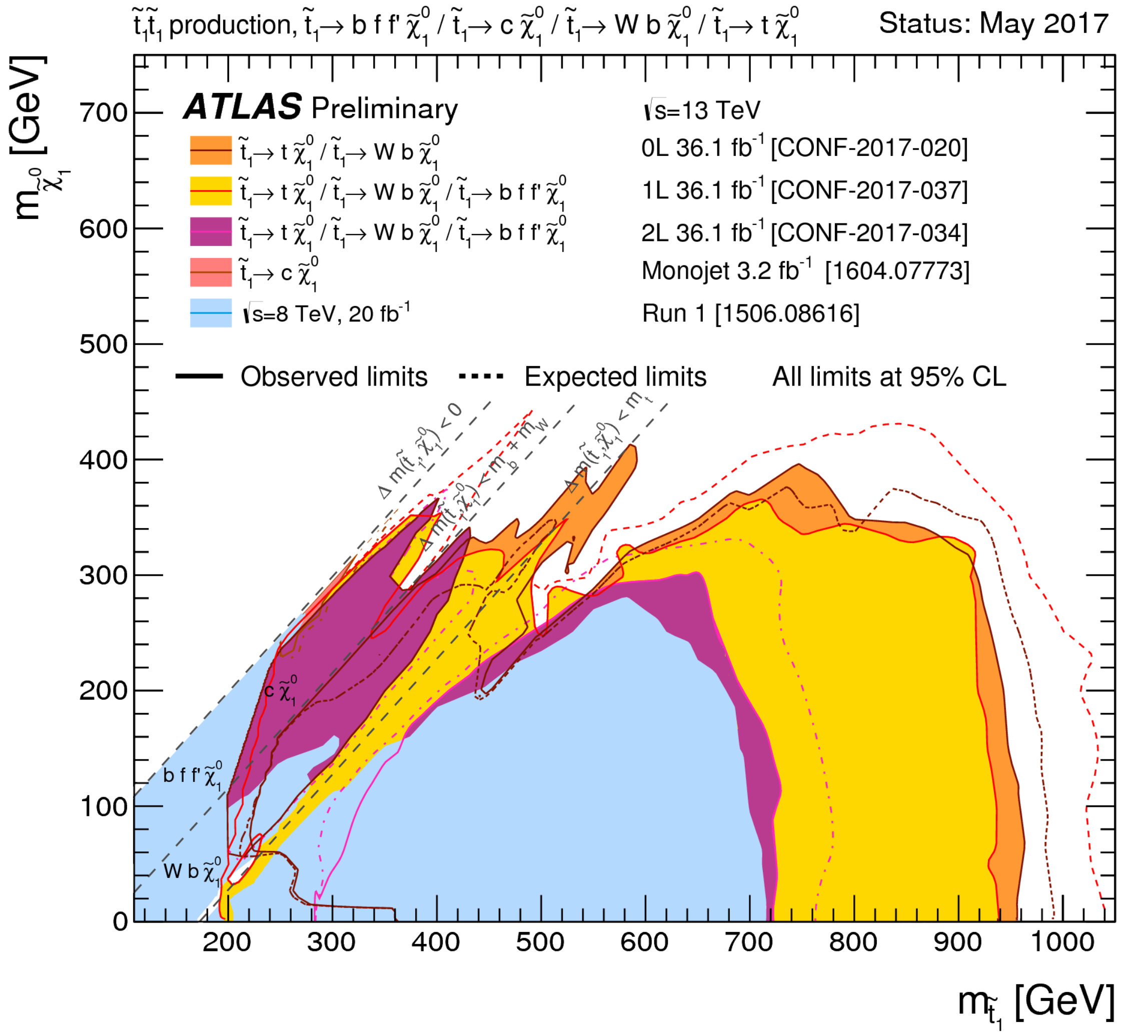




New Physics?

- Where is it?





New Physics?

PARTICLE PHYSICS

What No New Particles Means for Physics

43

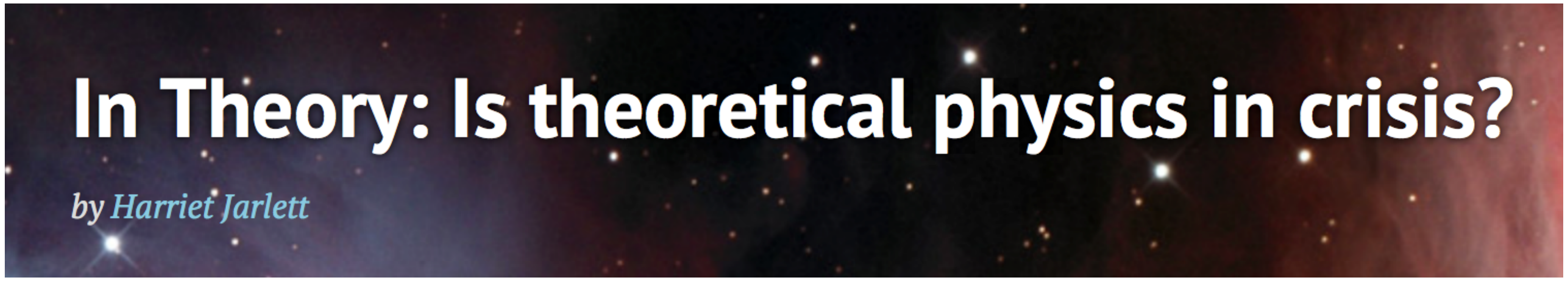
Physicists are confronting their "nightmare scenario." What doe new particles suggest about how nature works?

The New York Times

A Crisis at the Edge of Physics

Gray Matter

By ADAM FRANK and MARCELO GLEISER JUNE 5, 2015



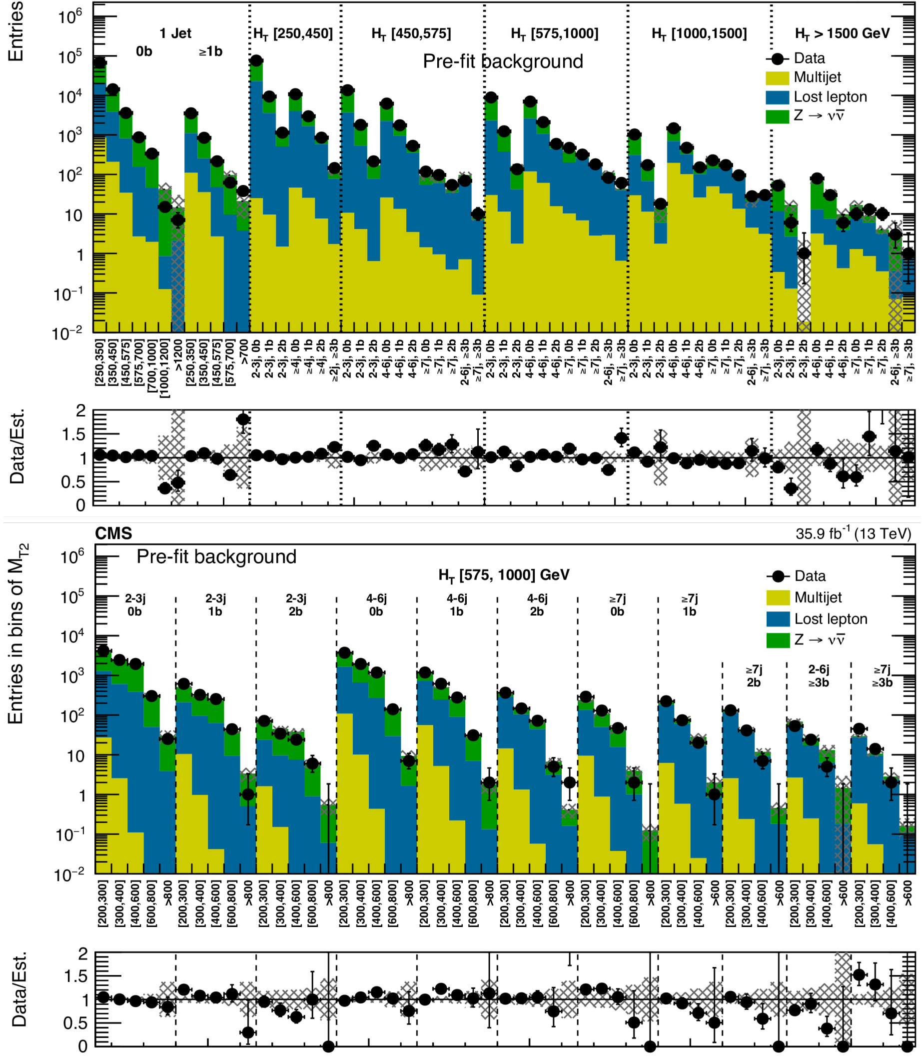
In Theory: Is theoretical physics in crisis?

by *Harriet Jarlett*

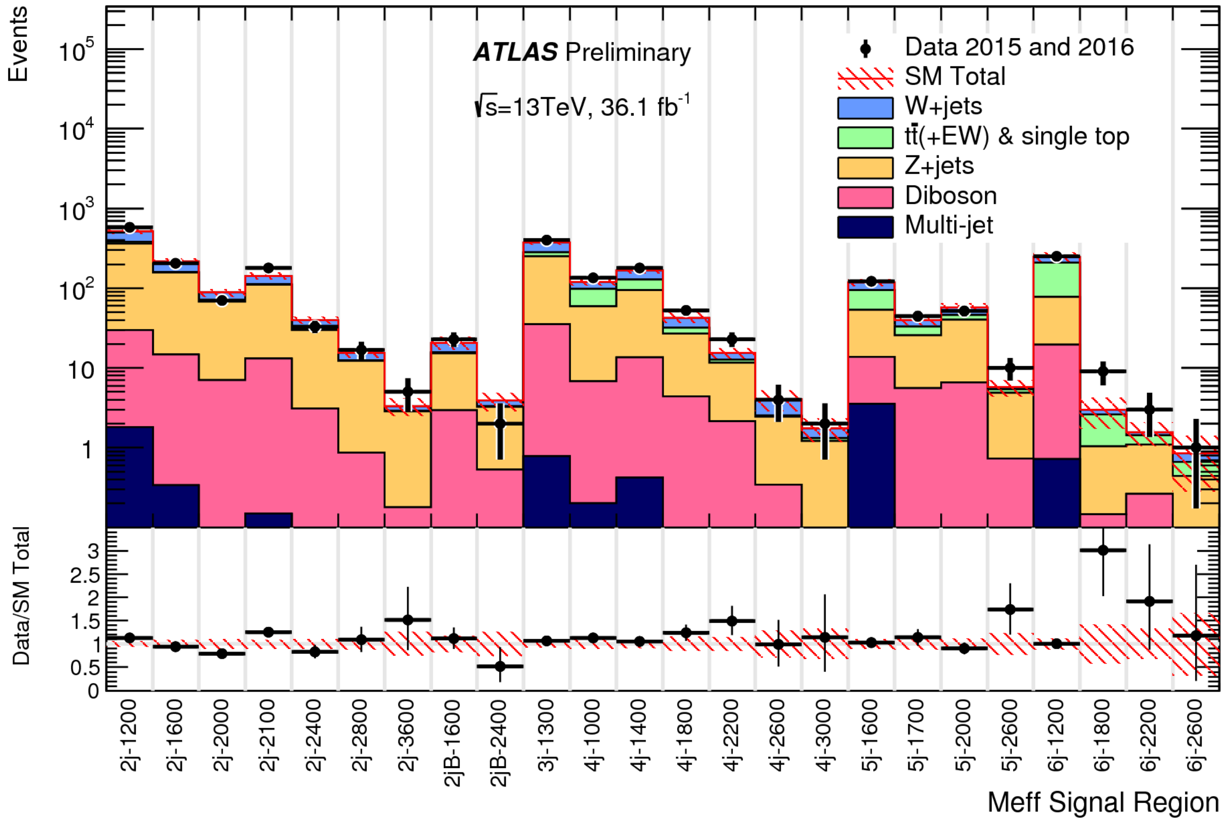
Slicing up Data

- ATLAS and CMS data divided up by topology (number of leptons, fat-jets, *etc.*)
- Then subdivided by kinematics into signal regions

CMS-SUS-16-036



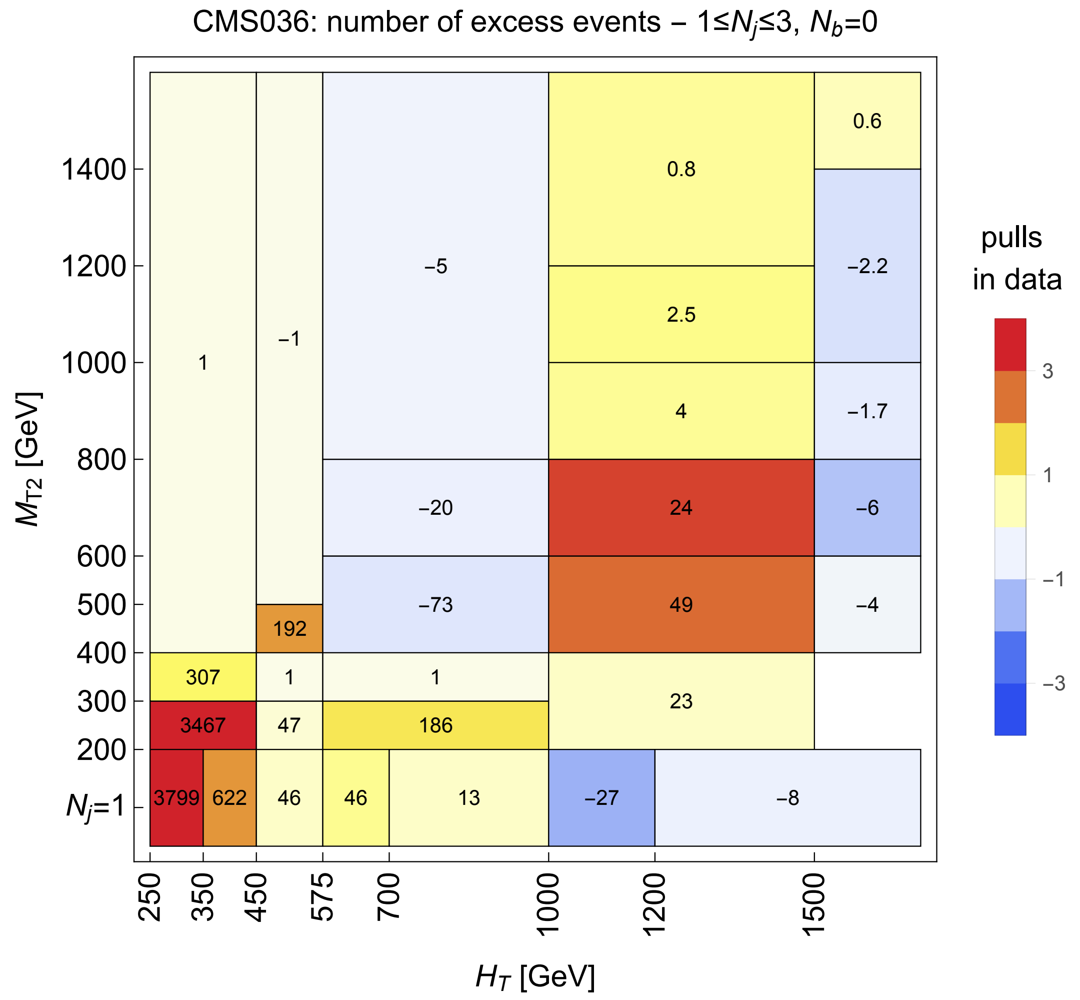
ATLAS-CONF-2017-022



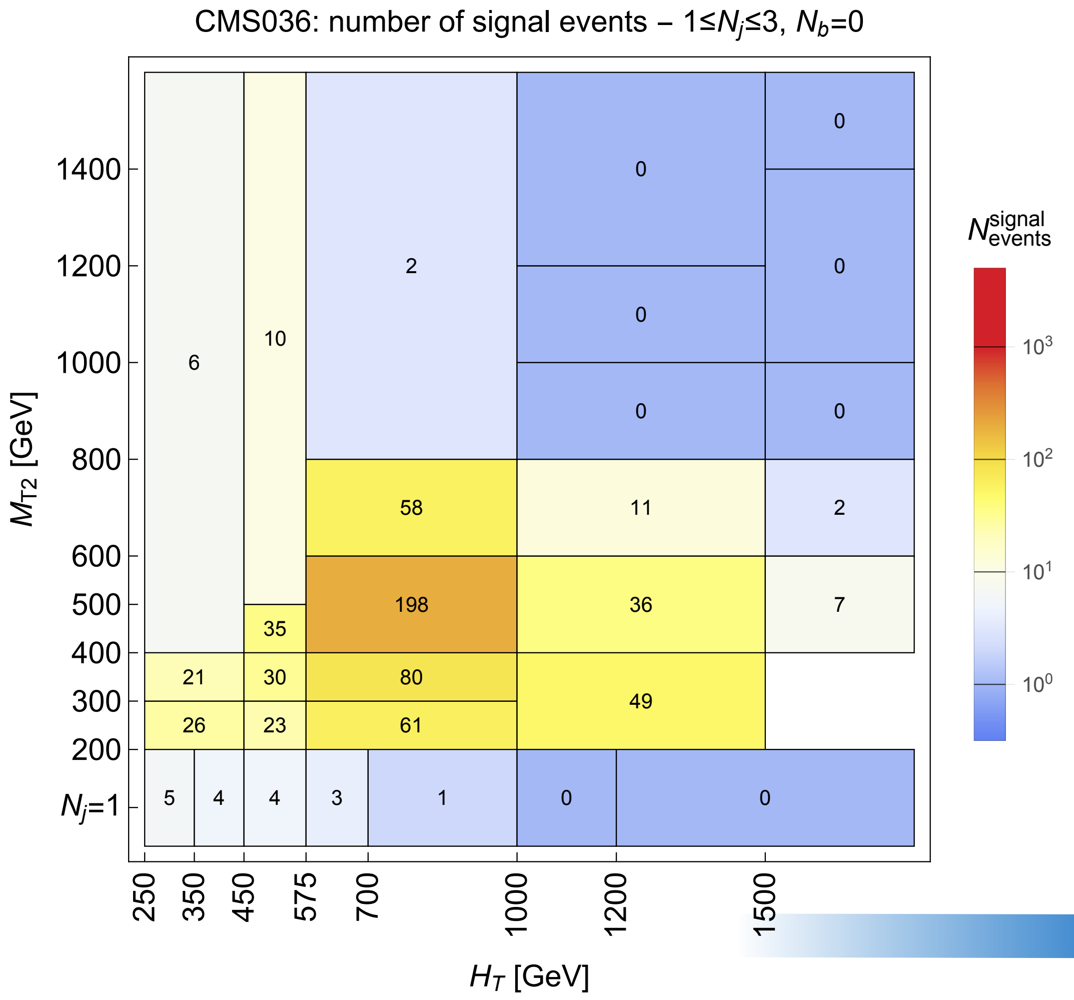
Setting Limits

- Limits are *model-dependent*.
  - Model tells us which how to combine the statistical pull of each signal region.

Residuals in data



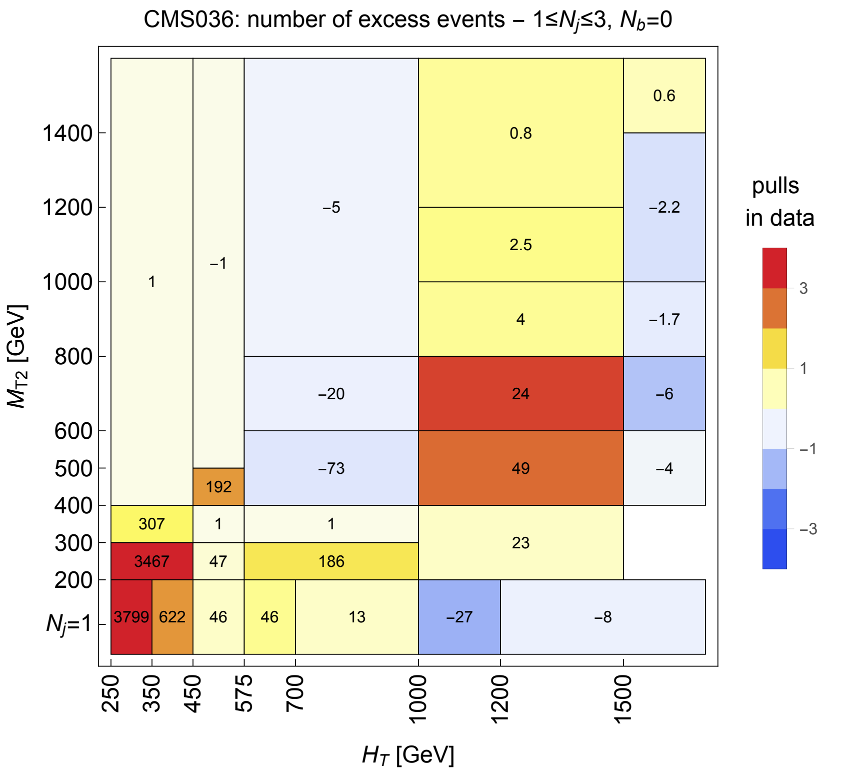
Events predicted in $pp \to \tilde{q}\tilde{q}^* \to q\bar{q}\tilde{\chi}_1^0\tilde{\chi}_1^0$



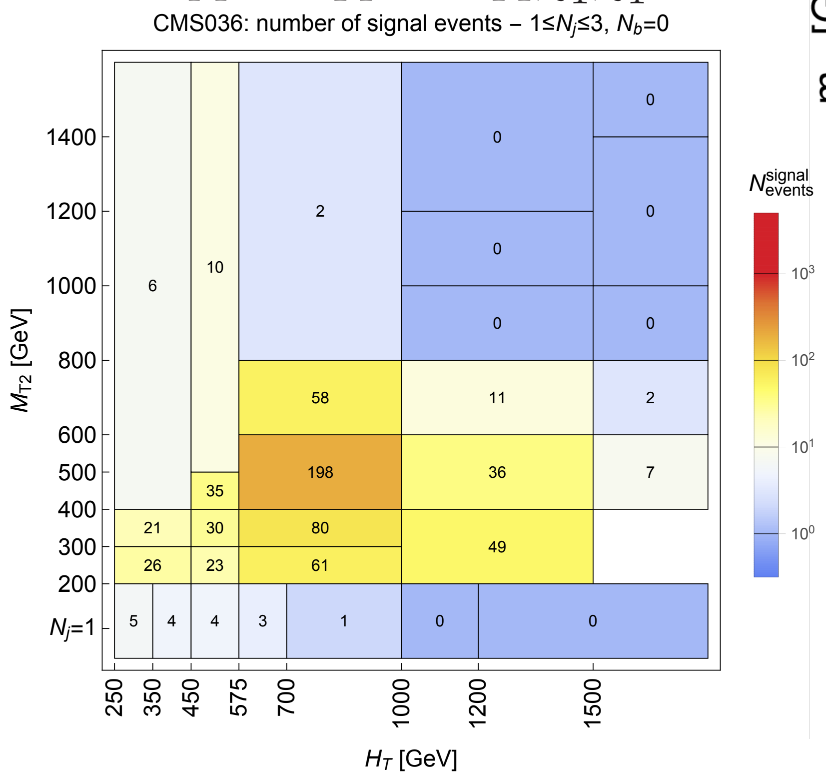
Setting Limits

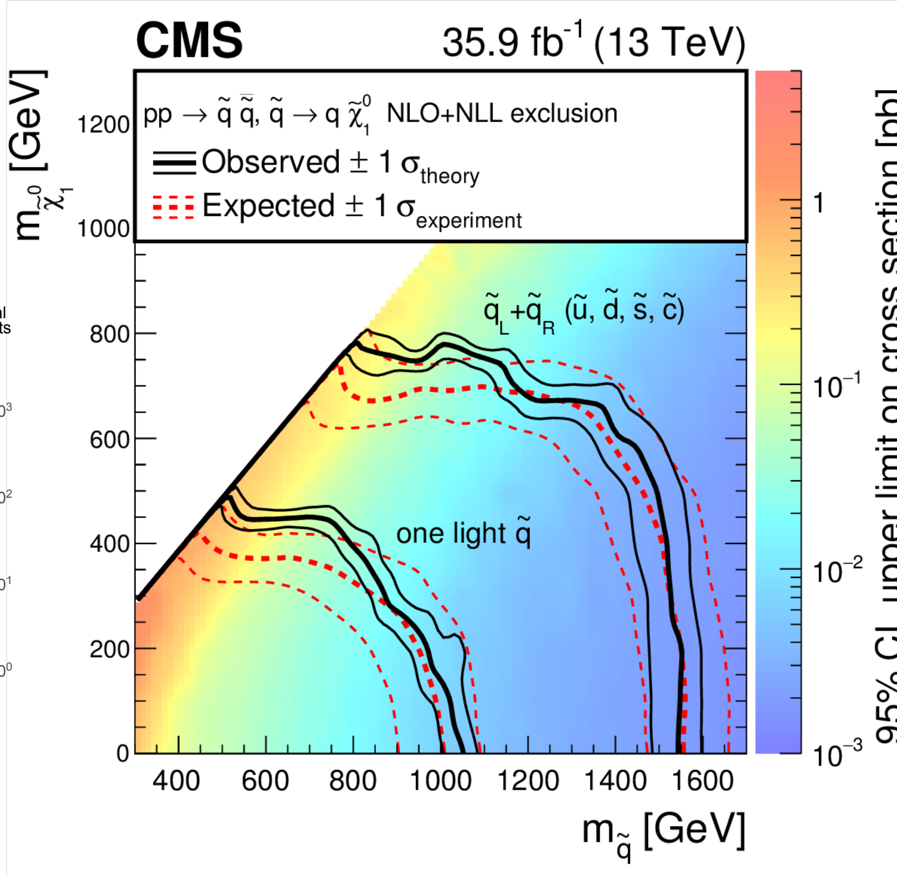
- A search can have many statistically significant excesses over background and still have observed limits equal expected
  - For a particular model
  - Have we looked at all models?

Residuals in data



Events predicted in $pp \to \tilde{q}\tilde{q}^* \to q\bar{q}\tilde{\chi}^0_1\tilde{\chi}^0_1$



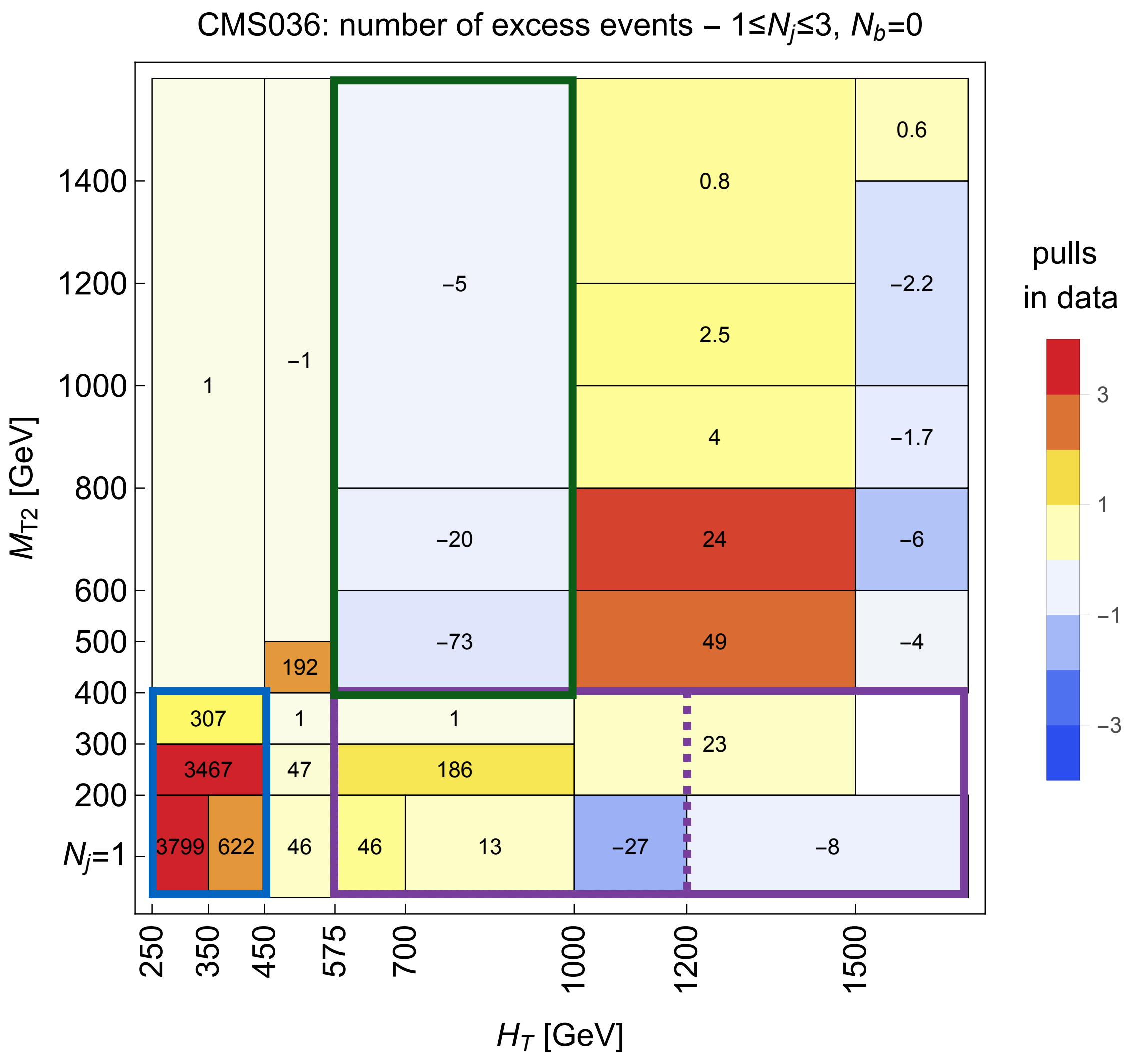


Rectangular Aggregations

- Signal likely to be distributed in “nearby” signal regions
  - Model kinematics, ISR/FSR, detector resolution,…

- Consider all possible “rectangular aggregations” of signal regions to look for signal over background.

- Best for non-overlapping SRs
  - CMS searches



Correlations

$$\mathcal{L}(\mu,\theta) = \prod_i \frac{(\mu s_i + b_i + \theta_i)_i^n e^{-(\mu s_i + b_i + \theta_i)}}{n_i!} \exp\left(-\tfrac{1}{2}\theta^T V^{-1}\theta\right)$$

- We're calculating $\Delta \log \mathcal{L}$, marginalizing over the background uncertainties θ_i (nuisance parameters)
  - Assuming signal populates only one RA at a time.
- CMS now publishing correlation/covariance matrices (thanks, CMS!)
- When we define a rectangular aggregation:

$$V_R = \begin{pmatrix} \sum_{i,j\in R} V_{ij} & \sum_{i\in R} V_{iJ} \\ \sum_{i\in R} V_{iJ} & V_{IJ} \end{pmatrix}$$

Jets + MET

- Concentrate on jets + MET searches as proof-of-principle

CMS-SUS-16-033
$N_j, N_b, \not{E}_T, H_T$
174 SRs, ~7000 RAs

CMS-SUS-16-036
$N_j, N_b, \not{E}_T, M_{T2}$
213 SRs, ~33000 RAs

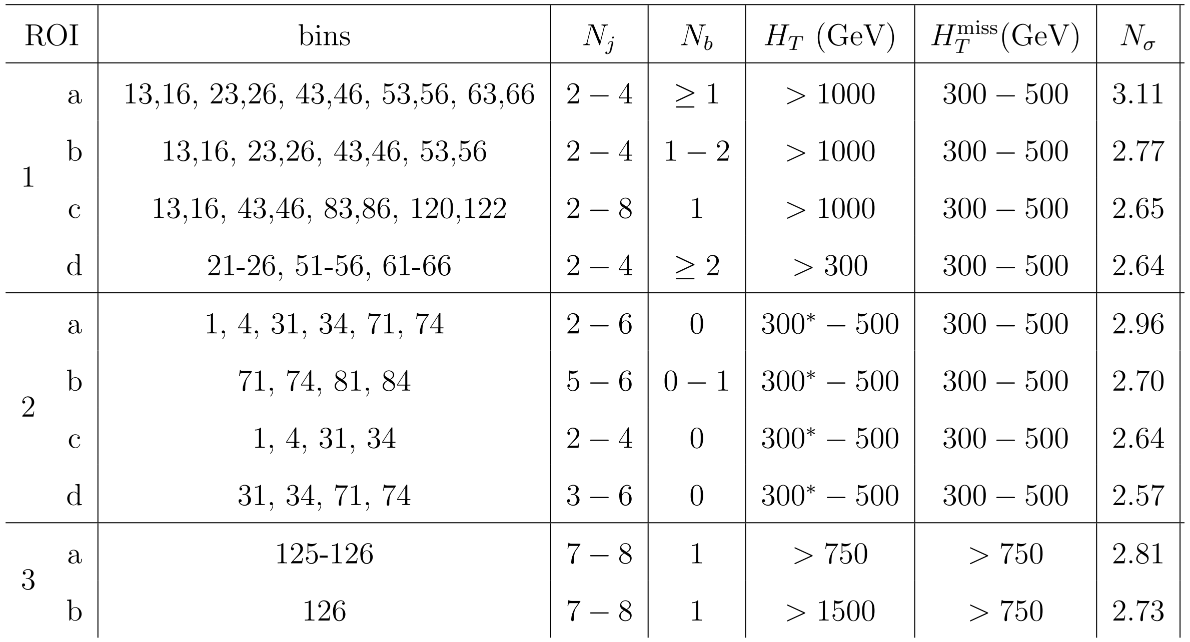
- ATLAS-PAS-17-022 has overlapping SRs
- CMS-EXO-16-048 has 1D SRs ($\not{E}_T$), this technique overkill

- Apply RA technique, assuming signal populates one rectangle and nowhere else.

Aggregating for Anomalies

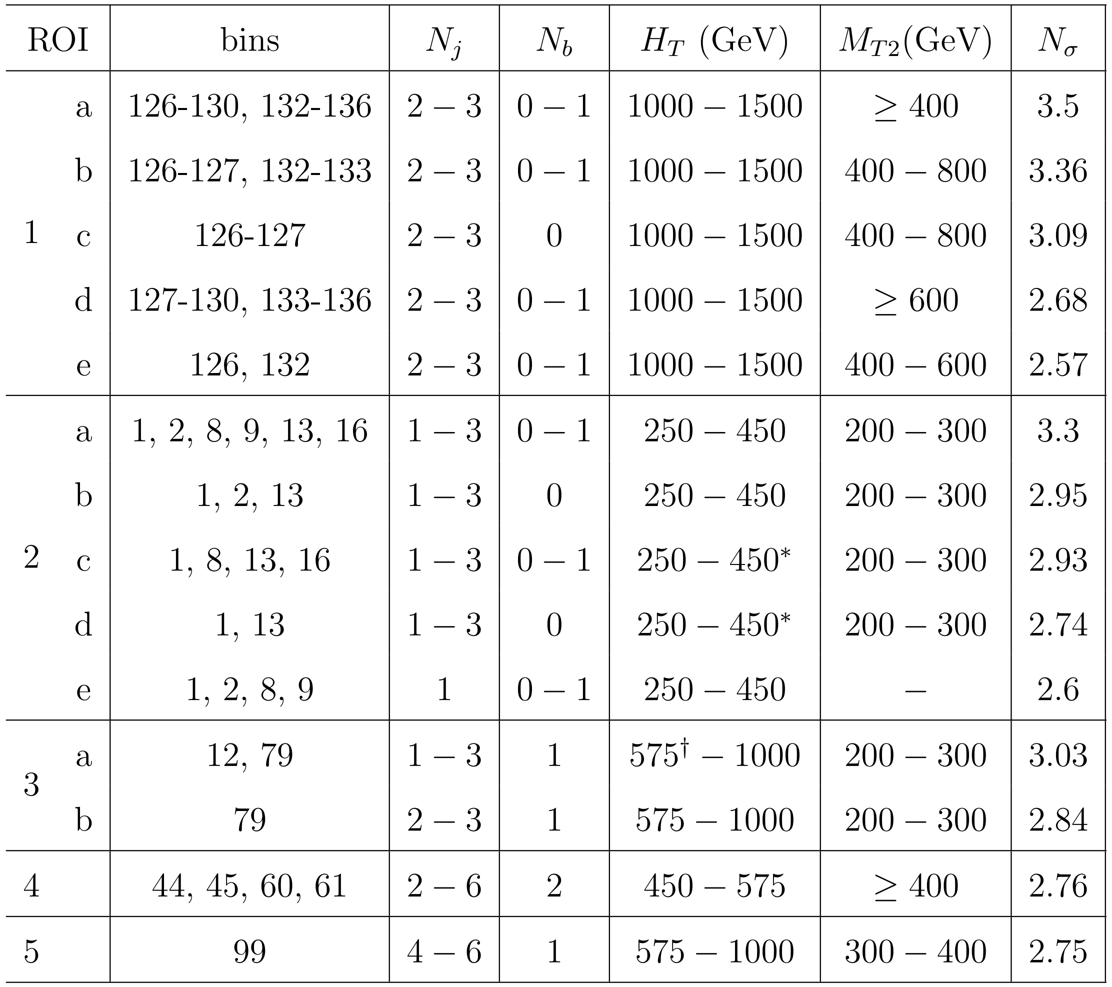
- We’re interested in excesses over background.
  - Keep anything with *p*-value < 1% $N_\sigma > 2.6$

CMS-SUS-16-033



ROI		bins	N_j	N_b	H_T (GeV)	H_T^{miss}(GeV)	N_σ
1	a	13,16, 23,26, 43,46, 53,56, 63,66	$2-4$	≥ 1	> 1000	$300-500$	3.11
	b	13,16, 23,26, 43,46, 53,56	$2-4$	$1-2$	> 1000	$300-500$	2.77
	c	13,16, 43,46, 83,86, 120,122	$2-8$	1	> 1000	$300-500$	2.65
	d	21-26, 51-56, 61-66	$2-4$	≥ 2	> 300	$300-500$	2.64
2	a	1, 4, 31, 34, 71, 74	$2-6$	0	300^*-500	$300-500$	2.96
	b	71, 74, 81, 84	$5-6$	$0-1$	300^*-500	$300-500$	2.70
	c	1, 4, 31, 34	$2-4$	0	300^*-500	$300-500$	2.64
	d	31, 34, 71, 74	$3-6$	0	300^*-500	$300-500$	2.57
3	a	125-126	$7-8$	1	> 750	> 750	2.81
	b	126	$7-8$	1	> 1500	> 750	2.73

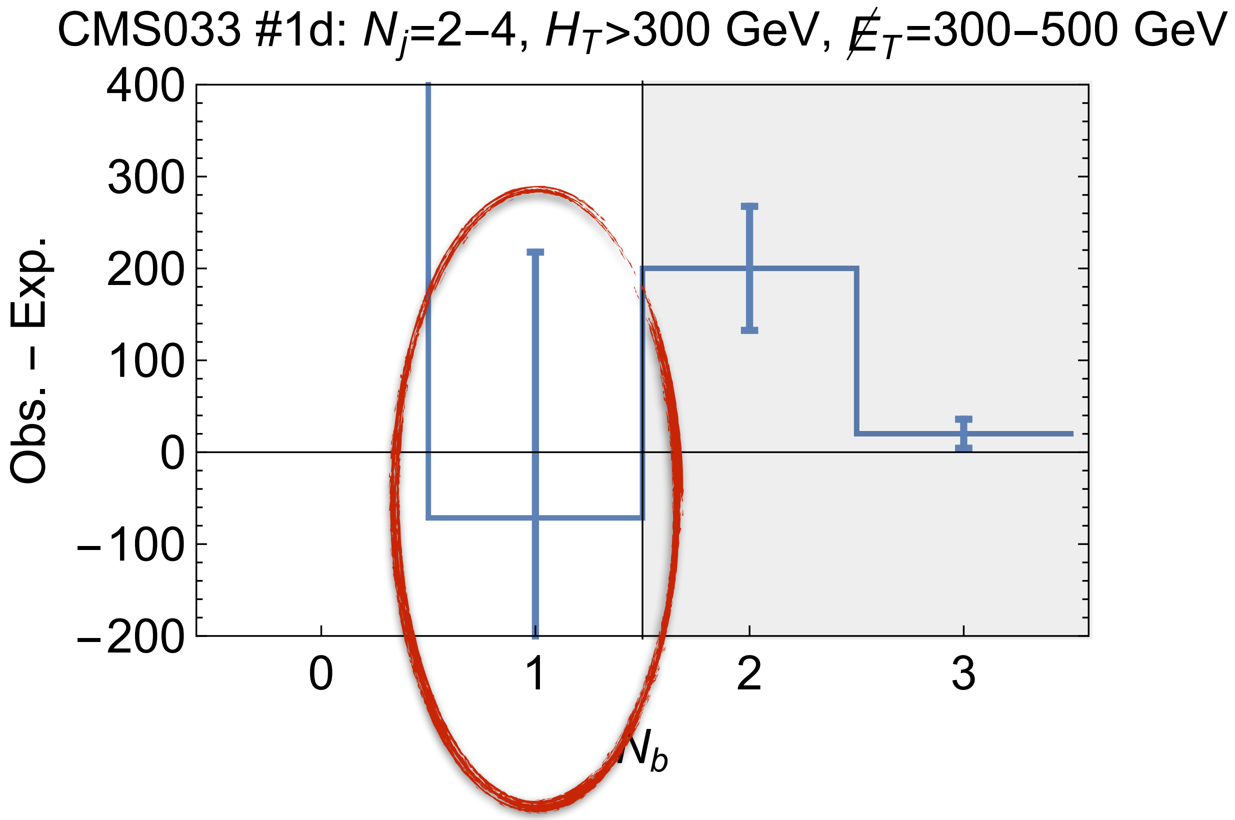
CMS-SUS-16-036

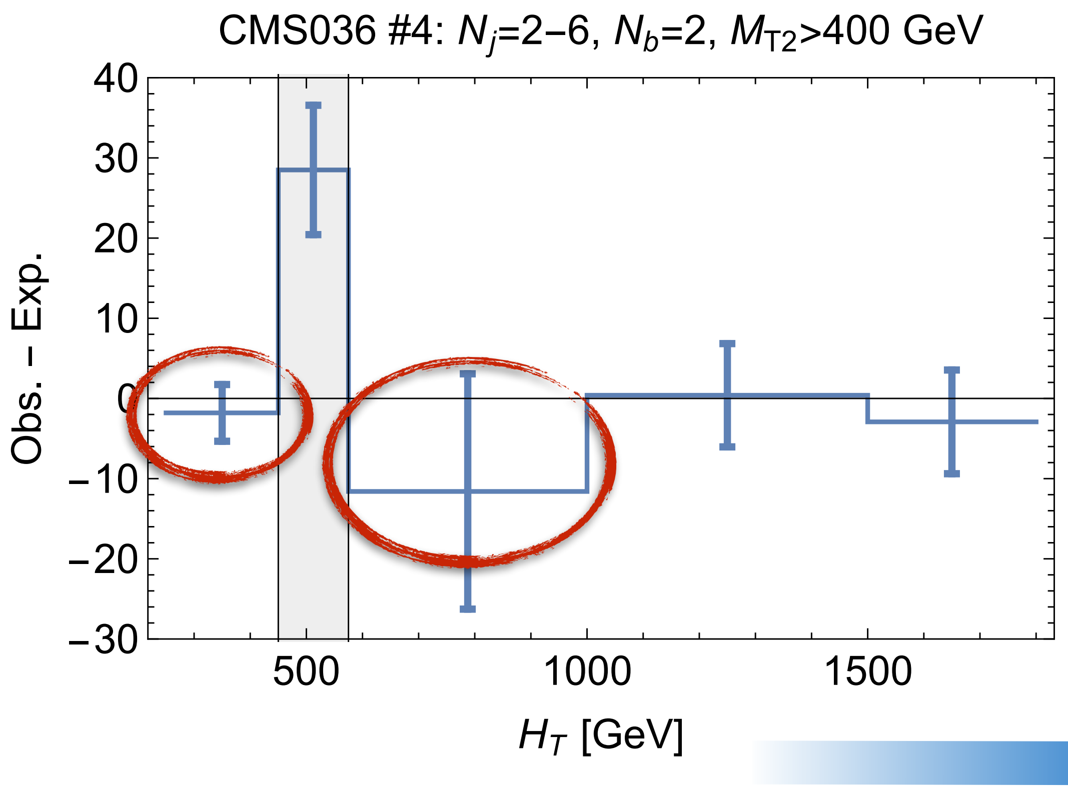


ROI		bins	N_j	N_b	H_T (GeV)	M_{T2}(GeV)	N_σ
1	a	126-130, 132-136	$2-3$	$0-1$	$1000-1500$	≥ 400	3.5
	b	126-127, 132-133	$2-3$	$0-1$	$1000-1500$	$400-800$	3.36
	c	126-127	$2-3$	0	$1000-1500$	$400-800$	3.09
	d	127-130, 133-136	$2-3$	$0-1$	$1000-1500$	≥ 600	2.68
	e	126, 132	$2-3$	$0-1$	$1000-1500$	$400-600$	2.57
2	a	1, 2, 8, 9, 13, 16	$1-3$	$0-1$	$250-450$	$200-300$	3.3
	b	1, 2, 13	$1-3$	0	$250-450$	$200-300$	2.95
	c	1, 8, 13, 16	$1-3$	$0-1$	$250-450^*$	$200-300$	2.93
	d	1, 13	$1-3$	0	$250-450^*$	$200-300$	2.74
	e	1, 2, 8, 9	1	$0-1$	$250-450$	–	2.6
3	a	12, 79	$1-3$	1	$575^\dagger-1000$	$200-300$	3.03
	b	79	$2-3$	1	$575-1000$	$200-300$	2.84
4		44, 45, 60, 61	$2-6$	2	$450-575$	≥ 400	2.76
5		99	$4-6$	1	$575-1000$	$300-400$	2.75

Reality Checks

- Obviously, most of these excesses aren't due to new physics.
- Can eliminate those in tension with equivalent regions in other jet+MET search (033↔036).
- Can further eliminate those that should have excesses in neighboring SRs

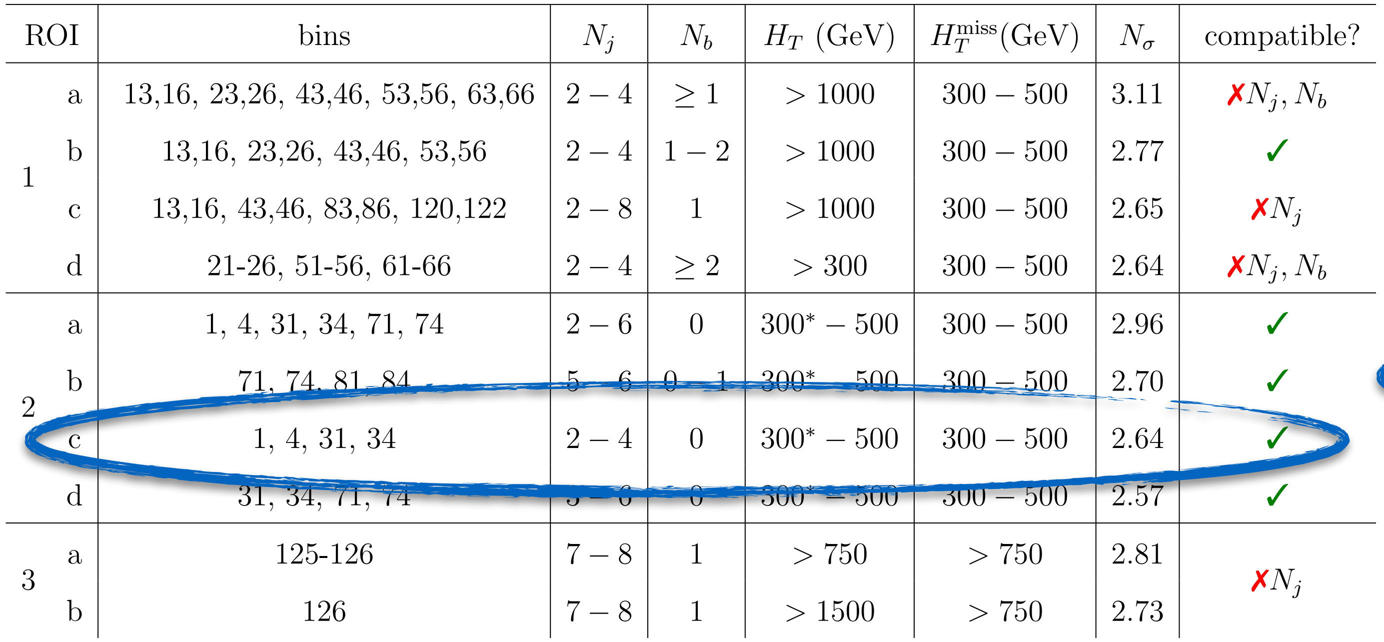




Surviving Anomalies

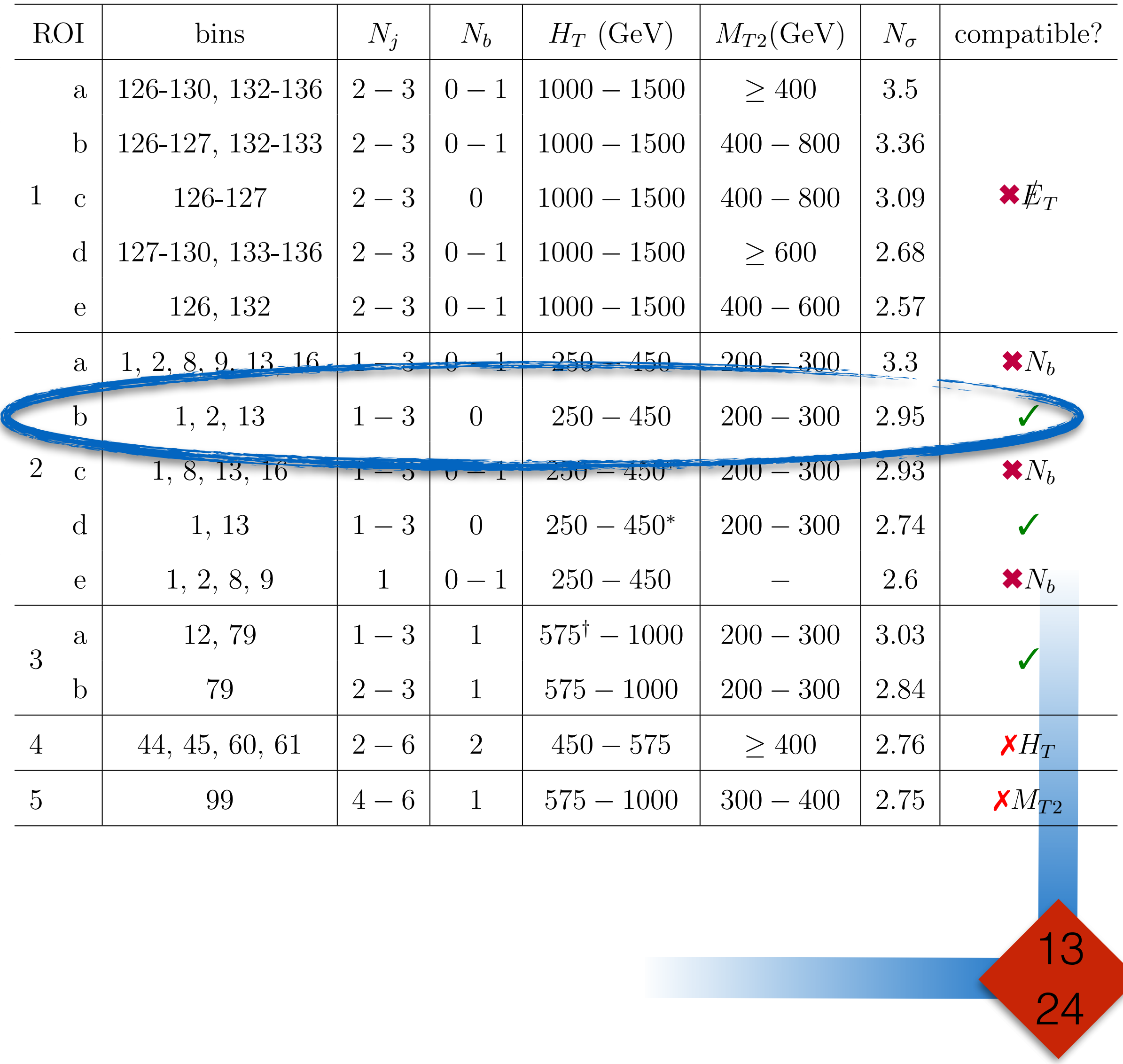
- Two in each search
  - One in each are particularly interesting

CMS-SUS-16-033



ROI		bins	N_j	N_b	H_T (GeV)	$H_T^{\rm miss}$(GeV)	N_σ	compatible?
1	a	13,16, 23,26, 43,46, 53,56, 63,66	$2-4$	≥ 1	> 1000	$300-500$	3.11	✗N_j, N_b
	b	13,16, 23,26, 43,46, 53,56	$2-4$	$1-2$	> 1000	$300-500$	2.77	✓
	c	13,16, 43,46, 83,86, 120,122	$2-8$	1	> 1000	$300-500$	2.65	✗N_j
	d	21-26, 51-56, 61-66	$2-4$	≥ 2	> 300	$300-500$	2.64	✗N_j, N_b
2	a	1, 4, 31, 34, 71, 74	$2-6$	0	300^*-500	$300-500$	2.96	✓
	b	71, 74, 81, 84	$5-6$	$0-1$	300^*-500	$300-500$	2.70	✓
	c	1, 4, 31, 34	$2-4$	0	300^*-500	$300-500$	2.64	✓
	d	31, 34, 71, 74	$5-6$	0	300^*-500	$300-500$	2.57	✓
3	a	125-126	$7-8$	1	> 750	> 750	2.81	✗N_j
	b	126	$7-8$	1	> 1500	> 750	2.73	

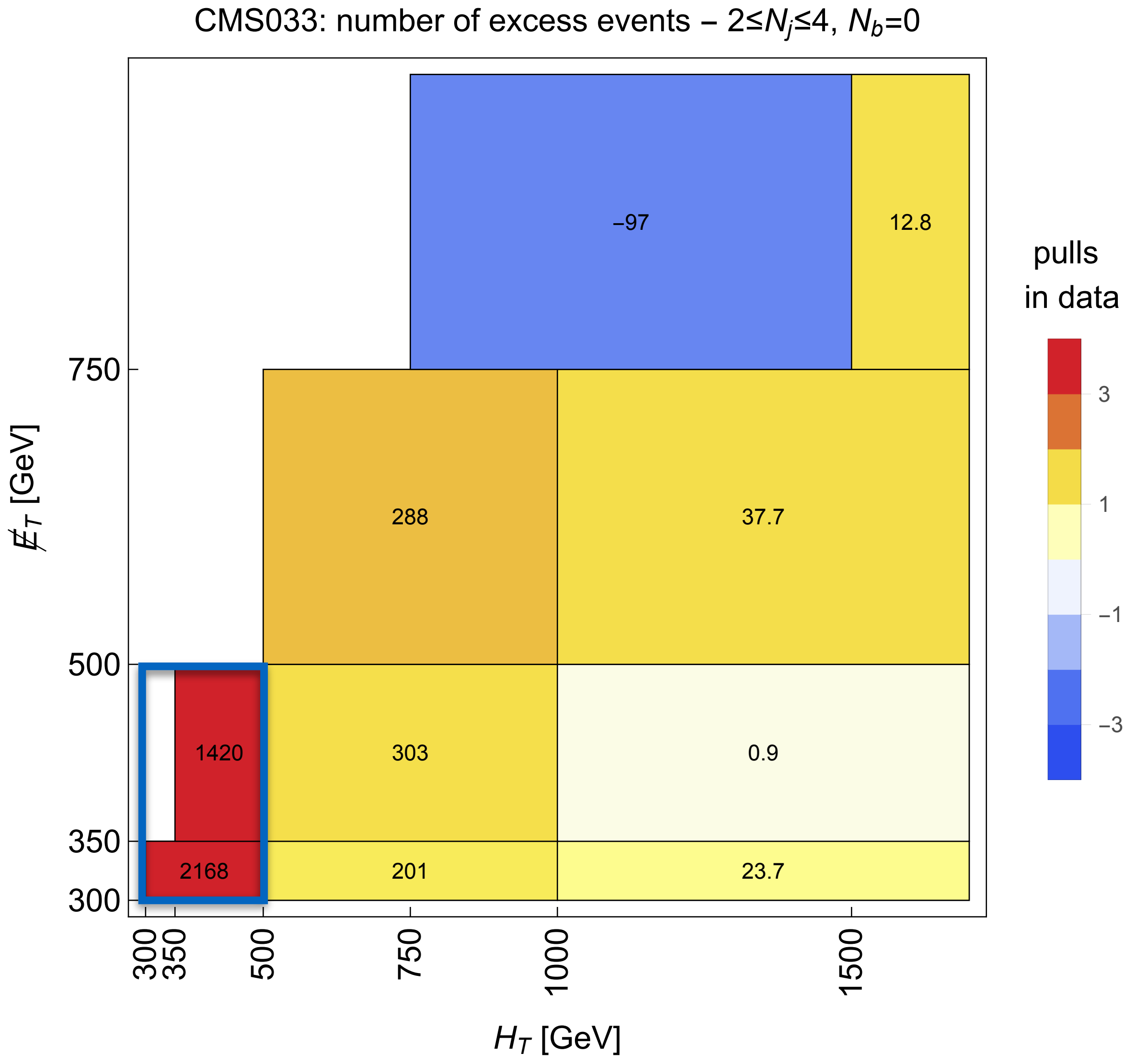
CMS-SUS-16-036

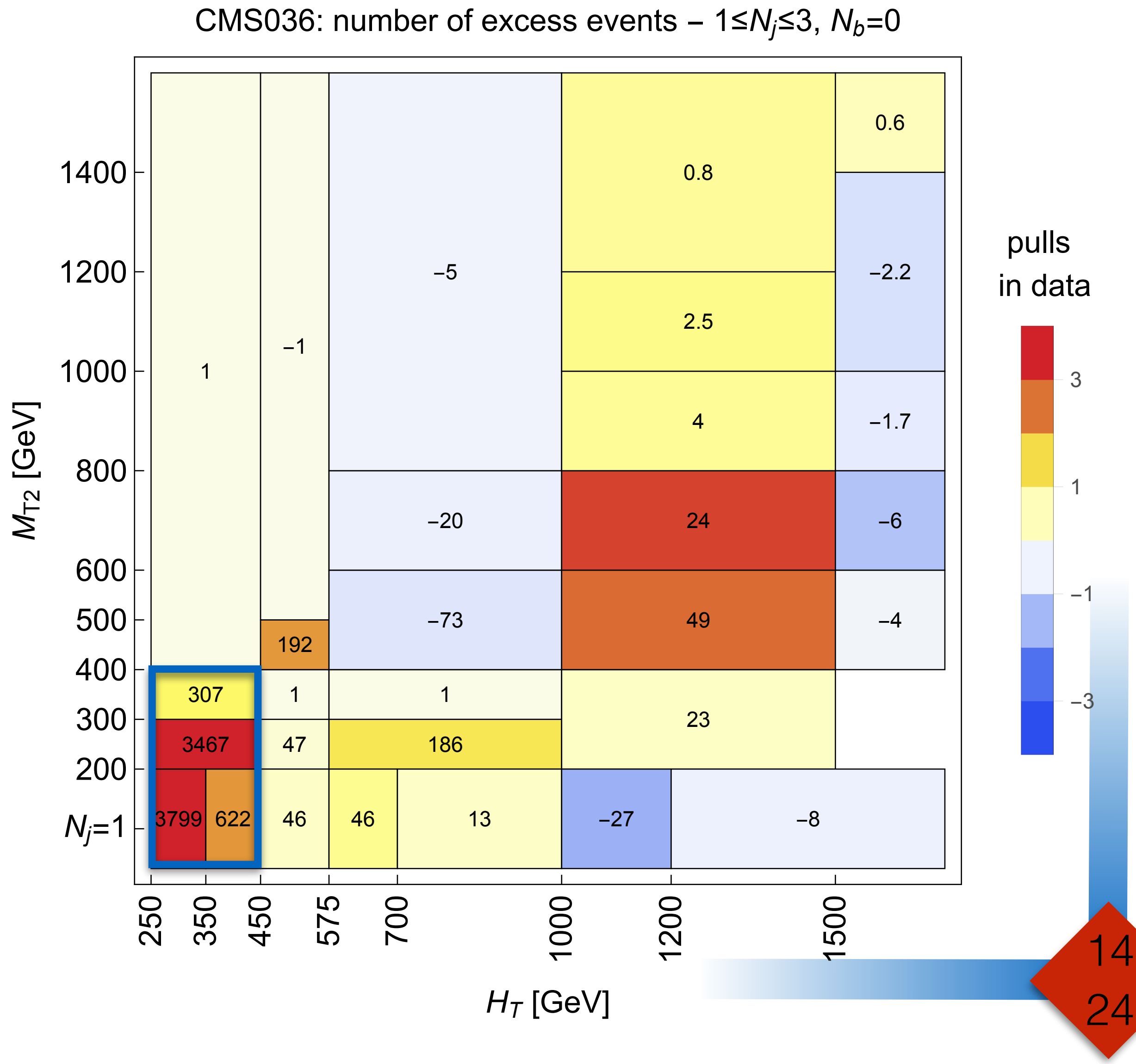


ROI		bins	N_j	N_b	H_T (GeV)	M_{T2}(GeV)	N_σ	compatible?
1	a	126-130, 132-136	$2-3$	$0-1$	$1000-1500$	≥ 400	3.5	✖$\not{E}_T$
	b	126-127, 132-133	$2-3$	$0-1$	$1000-1500$	$400-800$	3.36	
	c	126-127	$2-3$	0	$1000-1500$	$400-800$	3.09	
	d	127-130, 133-136	$2-3$	$0-1$	$1000-1500$	≥ 600	2.68	
	e	126, 132	$2-3$	$0-1$	$1000-1500$	$400-600$	2.57	
2	a	1, 2, 8, 9, 13, 16	$1-3$	$0-1$	$250-450$	$200-300$	3.3	✖N_b
	b	1, 2, 13	$1-3$	0	$250-450$	$200-300$	2.95	✓
	c	1, 8, 13, 16	$1-3$	$0-1$	$250-450^*$	$200-300$	2.93	✖N_b
	d	1, 13	$1-3$	0	$250-450^*$	$200-300$	2.74	✓
	e	1, 2, 8, 9	1	$0-1$	$250-450$	–	2.6	✖N_b
3	a	12, 79	$1-3$	1	$575^\dagger-1000$	$200-300$	3.03	✓
	b	79	$2-3$	1	$575-1000$	$200-300$	2.84	
4		44, 45, 60, 61	$2-6$	2	$450-575$	≥ 400	2.76	✗H_T
5		99	$4-6$	1	$575-1000$	$300-400$	2.75	✗M_{T2}

“Mono-Jet” Excess

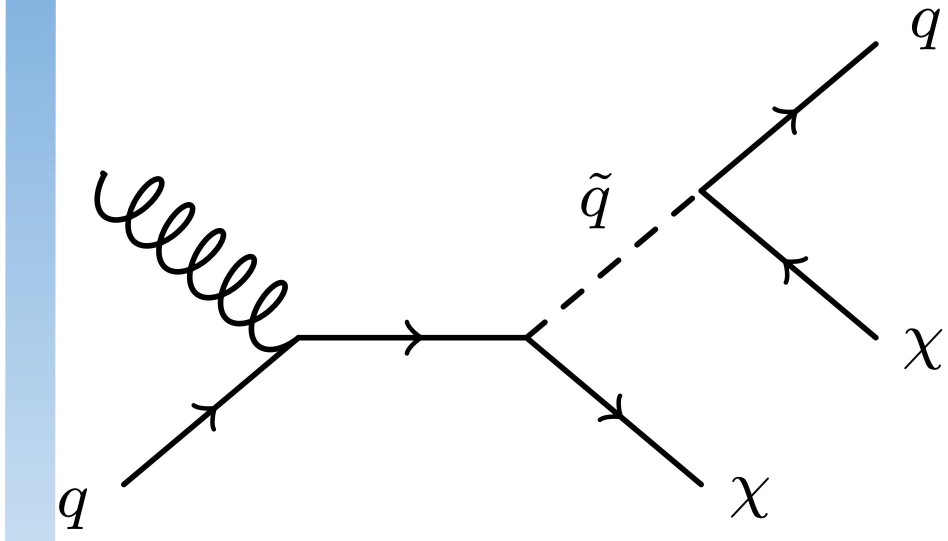
aggregation (significance)	N_j	N_b	$H_T(\text{GeV})$	$M_{T2}, \not{E}_T(\text{GeV})$
CMS036 #2b (2.95σ)	$1-3$	0	$250-450$	$200-300$
CMS033 #2c (2.64σ)	$2-4$	0	$300-500$	$300-500$


CMS033: number of excess events – $2 \leq N_j \leq 4$, $N_b=0$

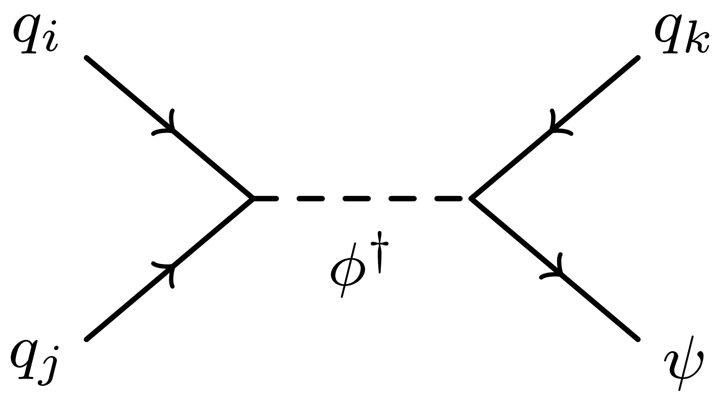

CMS036: number of excess events – $1 \leq N_j \leq 3$, $N_b=0$

“Mono-Jet” Excess

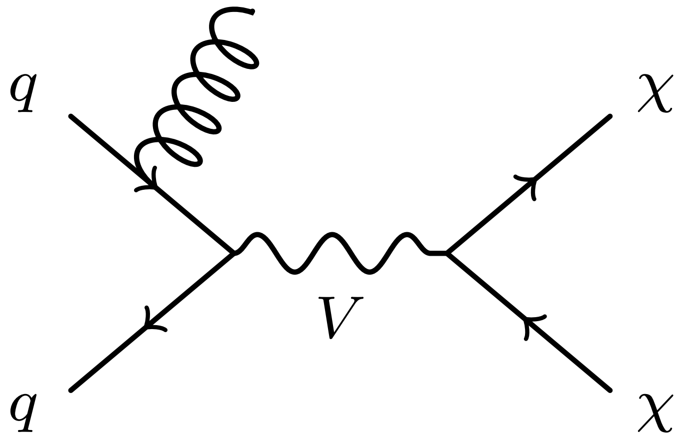
- What models fit this excess?
- Go back to the full analysis, using all data in all SRs
  - MSSM is not a good fit (as expected)
- We considered three models in depth:



squark/neutralino



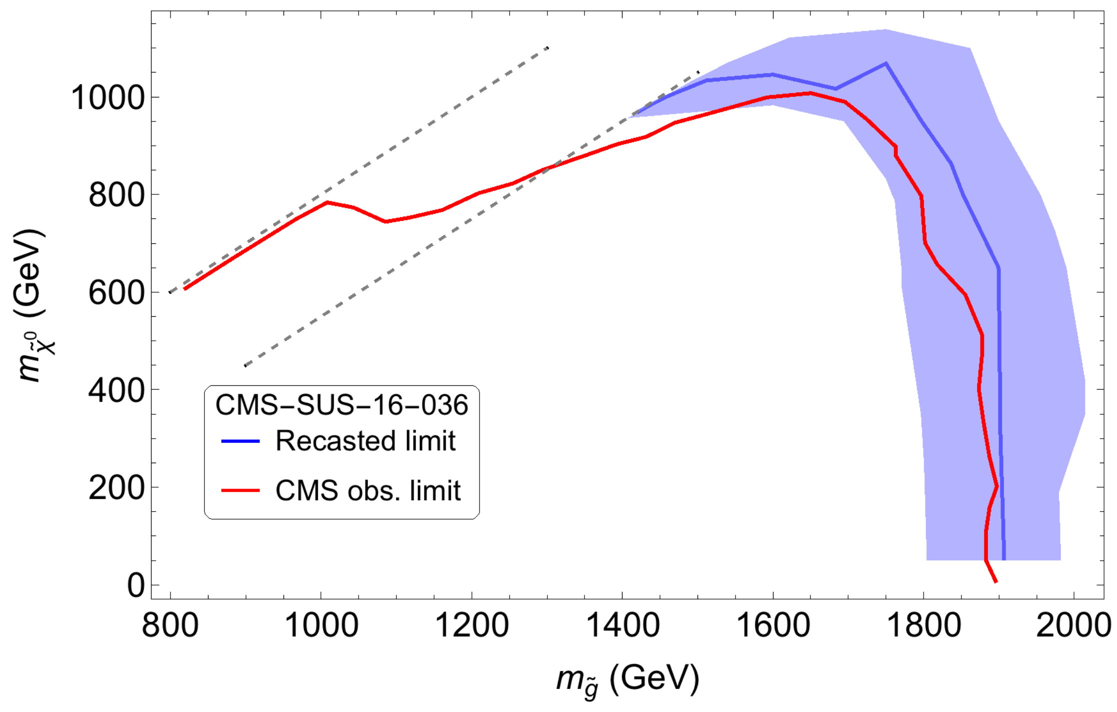
mono-ϕ



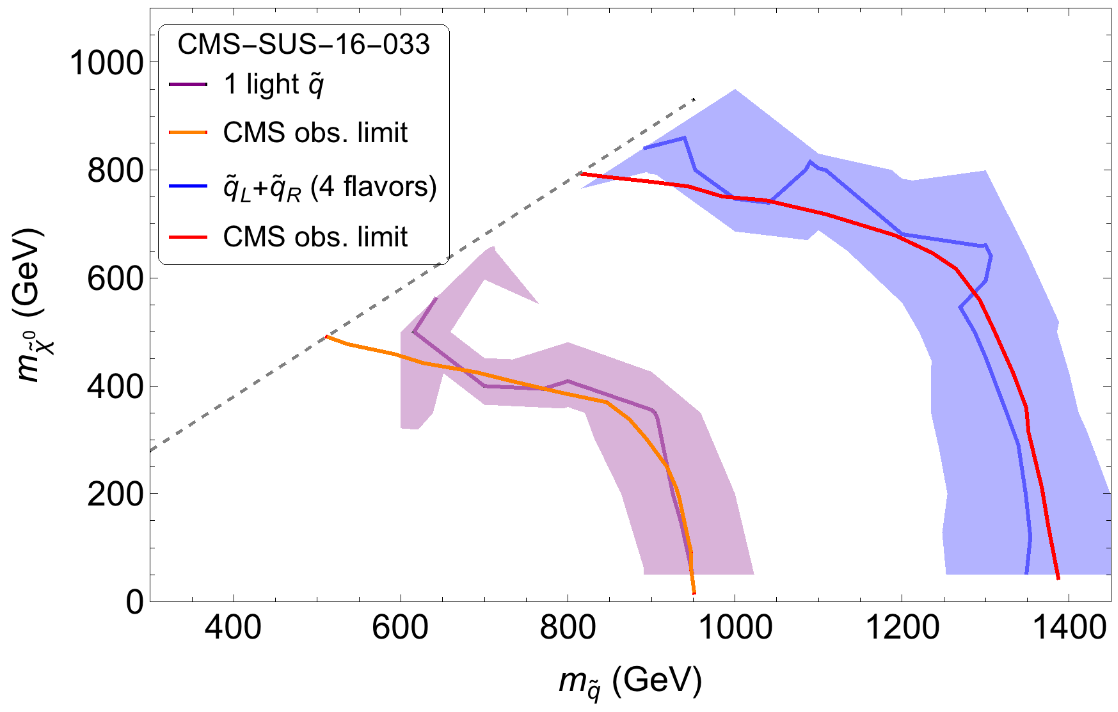
vector-mediated dark matter

Simulation Pipeline

MadGraph5 → Pythia8.2 → Delphes3 (our own tune) → Recast in pyRoot



$p\,p \to \tilde{g}\,\tilde{g},\ \tilde{g} \to t\,t\,\tilde{\chi}^0$



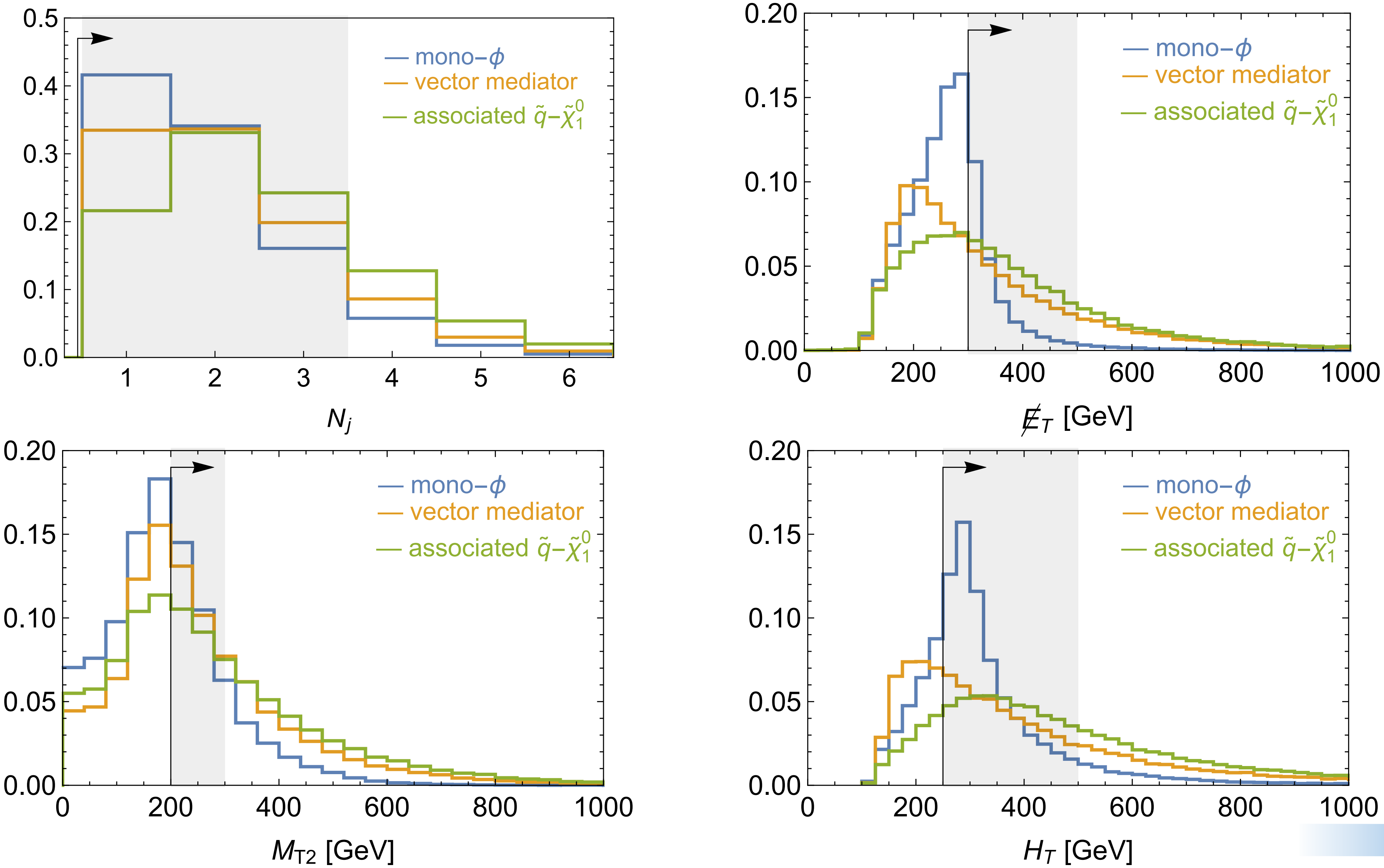
$p\,p \to \tilde{q}\,\tilde{q},\ \tilde{q} \to q\ \tilde{\chi}^0$

etc…

Analysis code available on arxiv for those interested

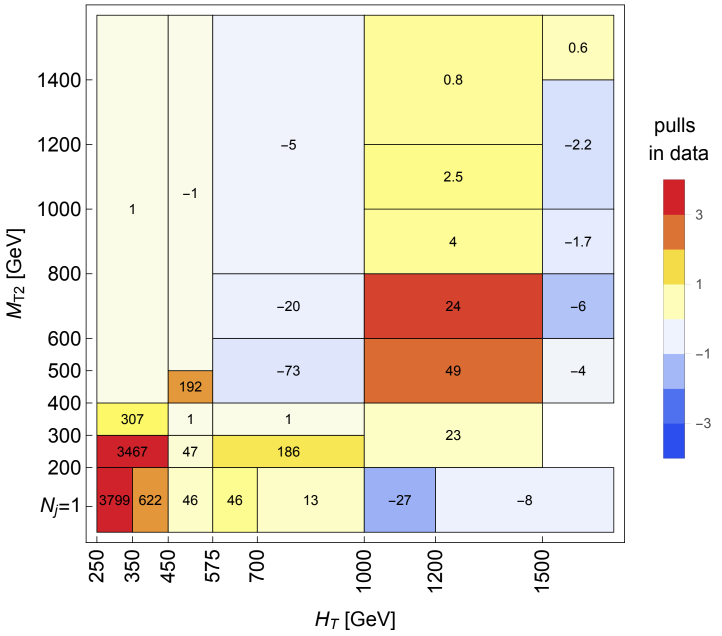
Successful Models

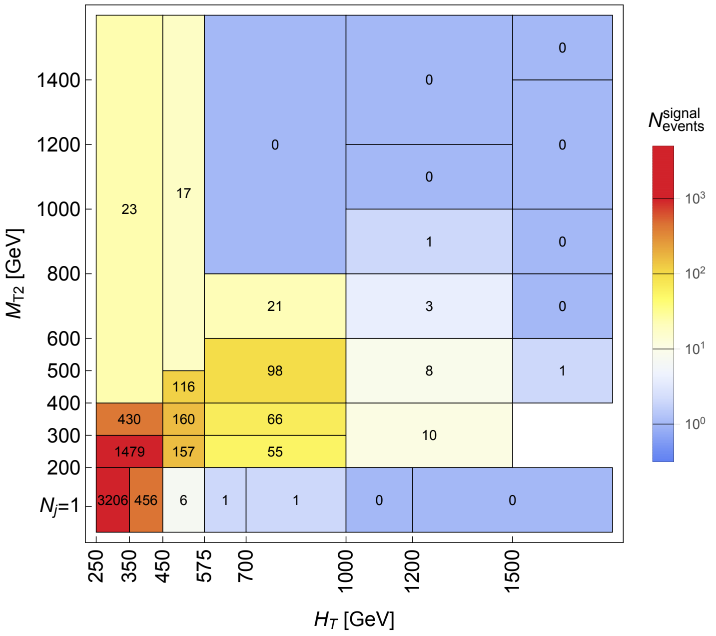
- Vector-mediated dark matter & squark/neutralino spill out into too many other SRs

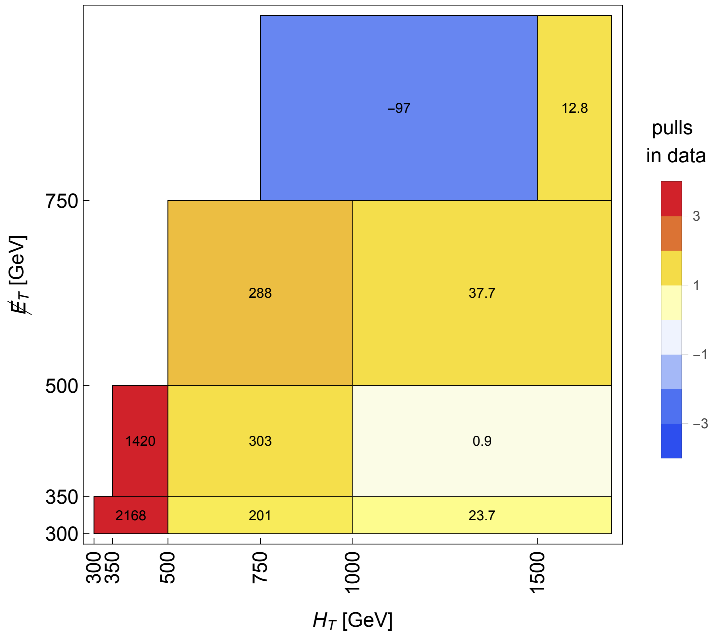


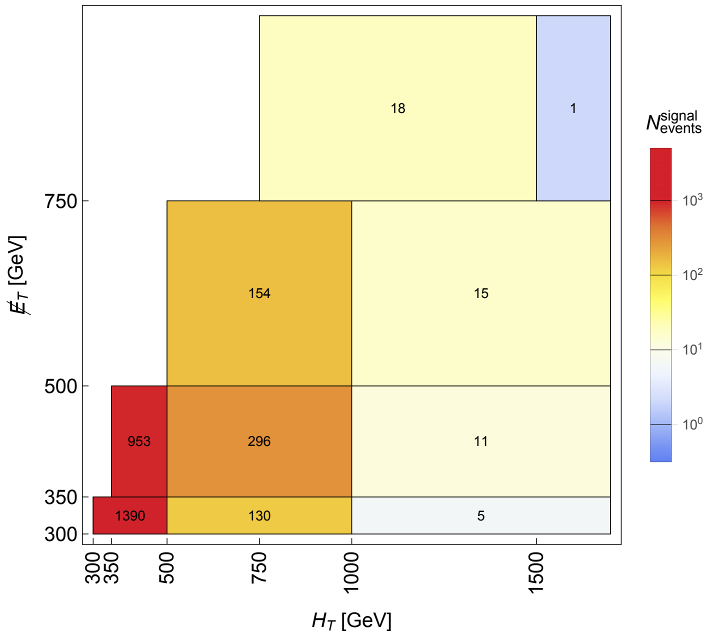
Successful Models

- The mono-ϕ model appears to work well.


CMS036: number of excess events – $1 \leq N_j \leq 3$, $N_b=0$

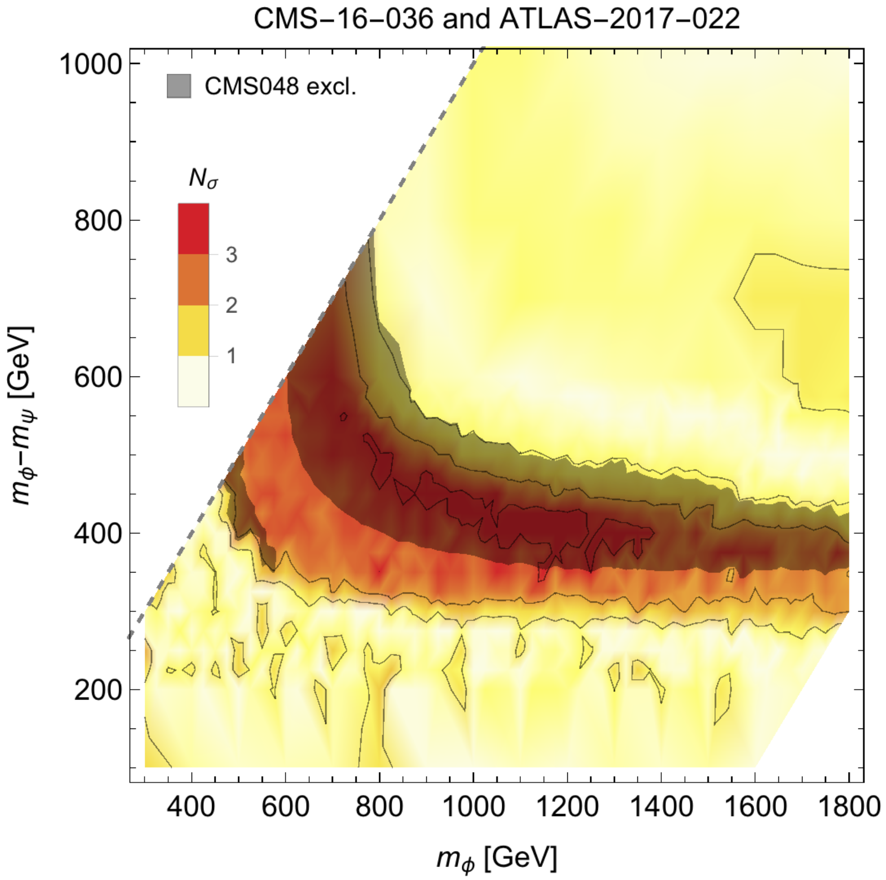

CMS036: number of signal events – $1 \leq N_j \leq 3$, $N_b=0$


CMS033: number of excess events – $2 \leq N_j \leq 4$, $N_b=0$

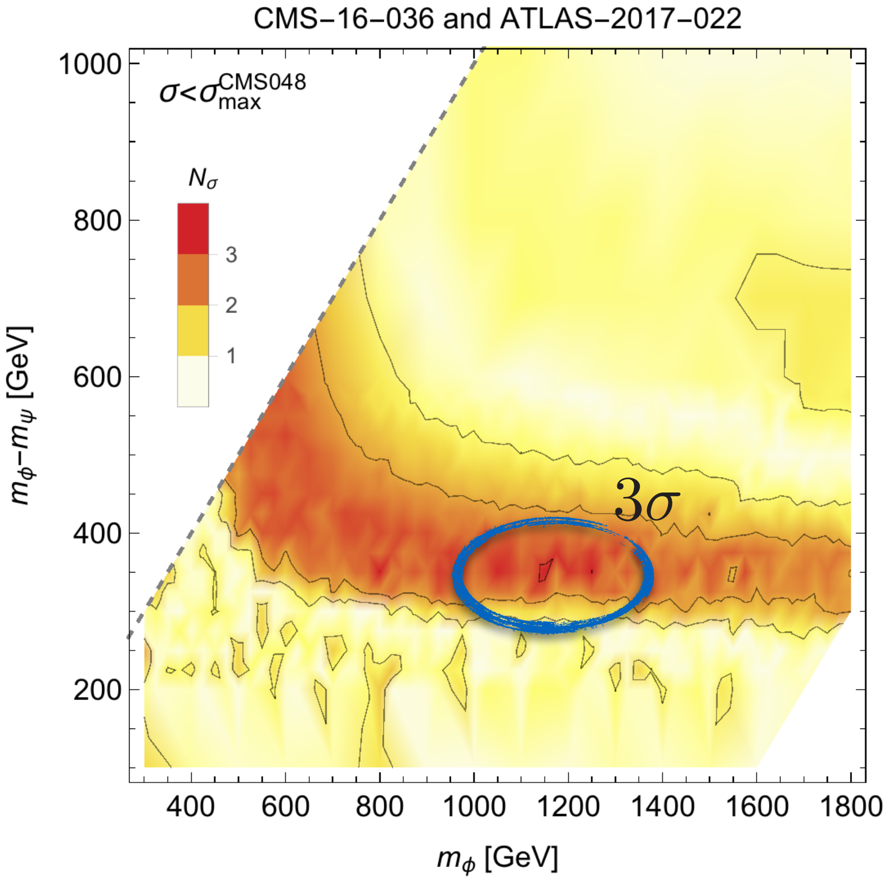

CMS033: number of signal events – $2 \leq N_j \leq 4$, $N_b=0$

Mono-ϕ

- Color triplet, decaying to quark + MET
  - “RPV-MSSM”-ish, but problems with Majorana masses
- This model is preferred at $\sim 3\sigma$ in CMS-036.
  - Tension with CMS-048
- ATLAS-2017-022 has $\sim 1.5\sigma$ preference. $\sim 3.5\sigma$ combined



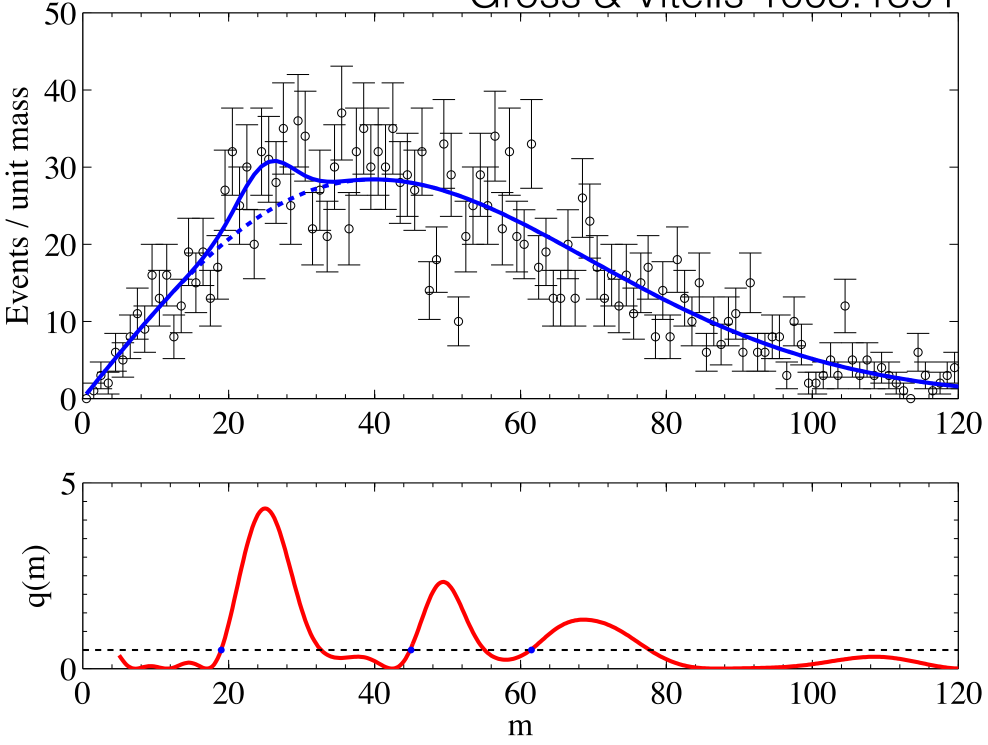
CMS−16−036 and ATLAS−2017−022



CMS−16−036 and ATLAS−2017−022

Look-Elsewhere

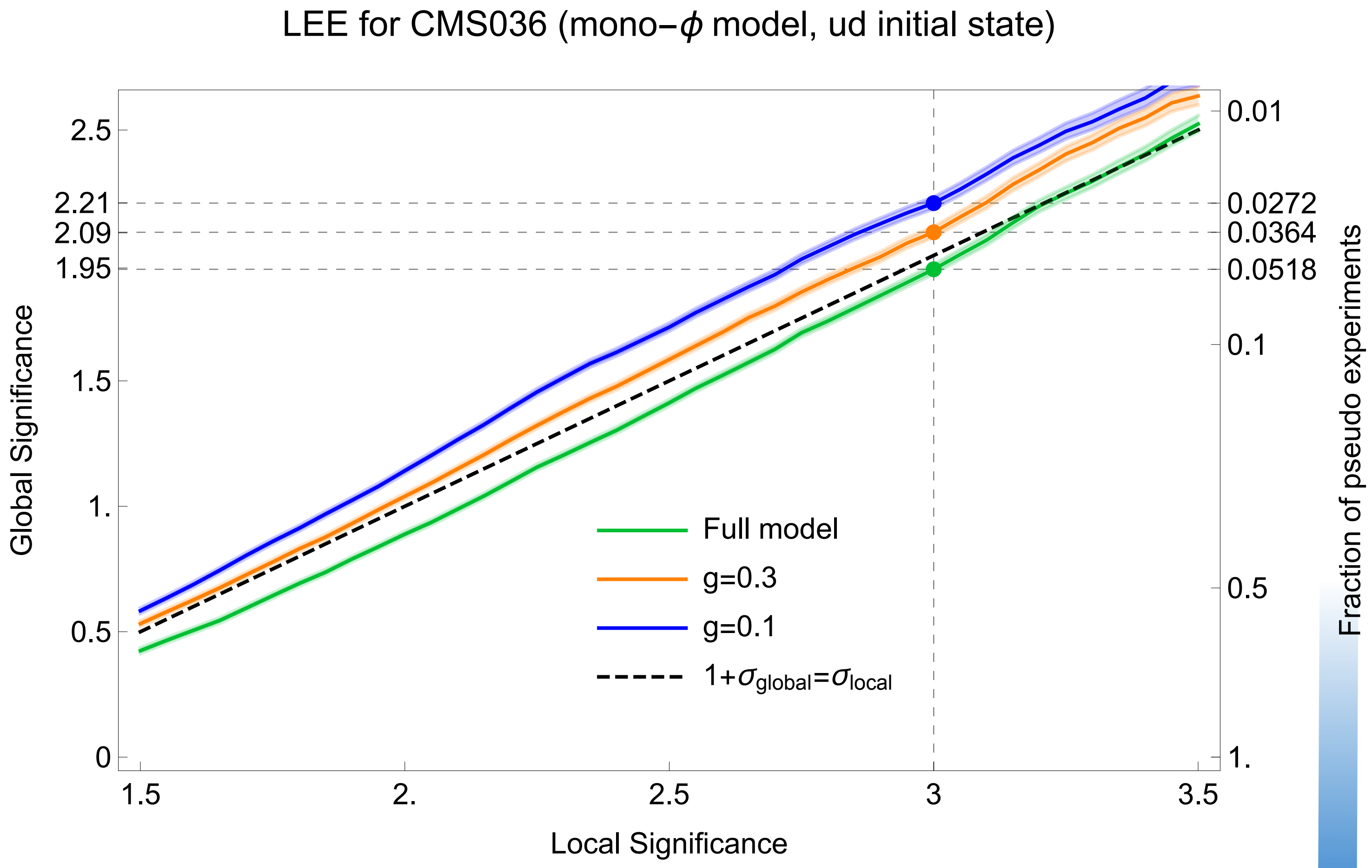
- Having looked in 33,000 rectangles, are we guaranteed to find a $\sim 3\sigma$ excess?
- If we generate 10K pseudoexperiments, we find $\sim 3\sigma$ local anomalies in 15% of them ($\sim 1.5\sigma$ global).
  - But this doesn't account for the reality checks.
- Look-Elsewhere Effect well-defined defined in terms of a model.
  - *e.g.* number of up-crossings in a resonance search.



Look-Elsewhere in a Model

- Work within the mono-ϕ model.
- For 10K pseudoexperiments, fit across the mass plane

- Reduces a $\sim 3\sigma$ local fluctuation to $\sim 2\sigma$
- We couldn't cross-correlate with CMS033/ATLAS022



Outlook

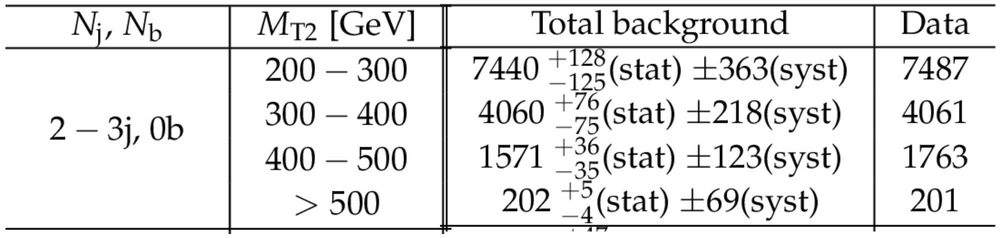
- LHC data contains interesting statistical excesses *now*
  - Can be hidden inside the many high-dimension SRs
- Most (all?) are probably statistical fluctuations, but that takes work to uncover.
- The set of benchmark models used is not a sufficient basis.

- Proof-of-concept: we have identified two <1% anomalies in CMS jets+MET.
  - Can't apply to ATLAS data, because SRs are overlapping.
  - (The ATLAS thresholds seem to be higher as well)

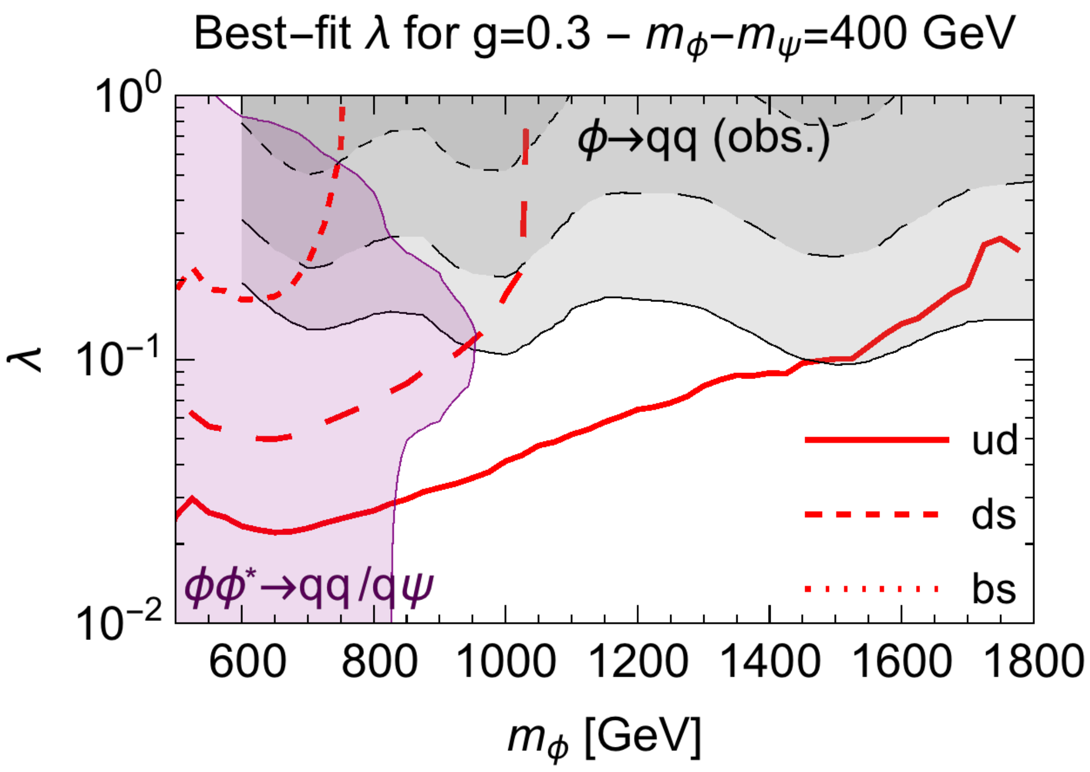
Outlook

- The “mono-jet” anomaly:
    - Well fit by a color-triplet decaying to quarks+MET
    - Associated signatures (dijets, multijets, multijets+MET)
    - Systematics limited
- Identifying these anomalies now gives targets of interest for the future data analysis.
    - Can freeze thresholds to maintain sensitivity.
- Can test signal hypotheses with additional kinematics.

CMS-SUS-16-036

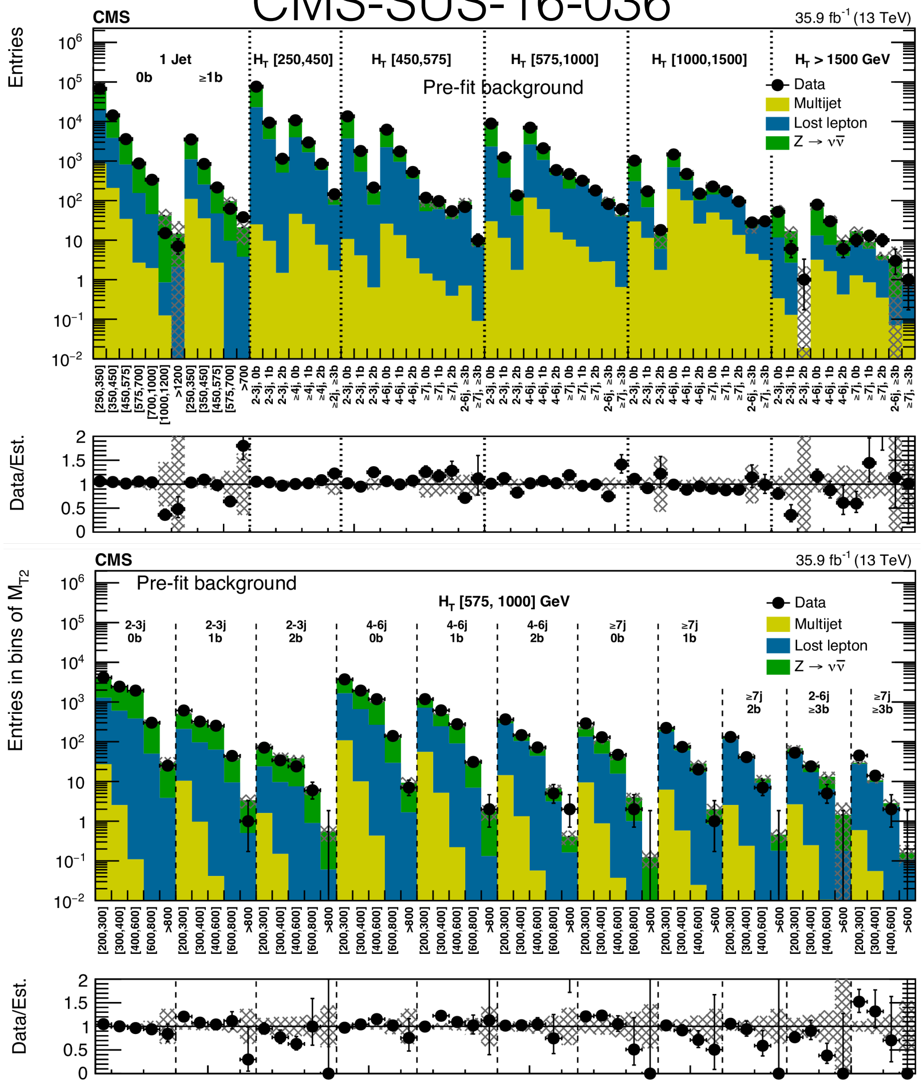


N_j, N_b	M_{T2} [GeV]	Total background	Data
2 – 3j, 0b	200 – 300	$7440\ ^{+128}_{-125}$(stat) ±363(syst)	7487
	300 – 400	$4060\ ^{+76}_{-75}$(stat) ±218(syst)	4061
	400 – 500	$1571\ ^{+36}_{-35}$(stat) ±123(syst)	1763
	> 500	$202\ ^{+5}_{-4}$(stat) ±69(syst)	201

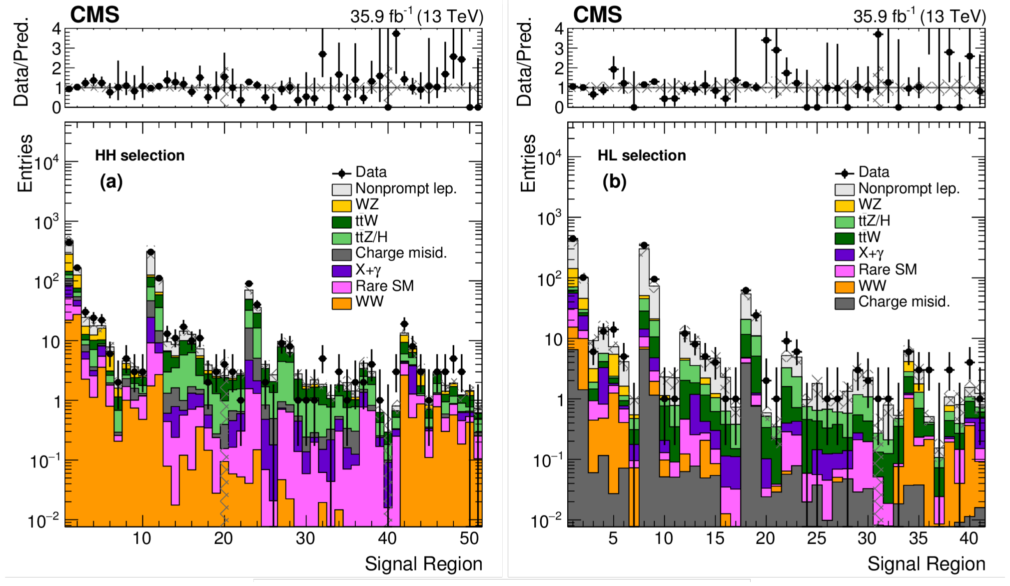


New Physics

CMS-SUS-16-036



CMS-SUS-16-035



ATLAS-CONF-2017-022