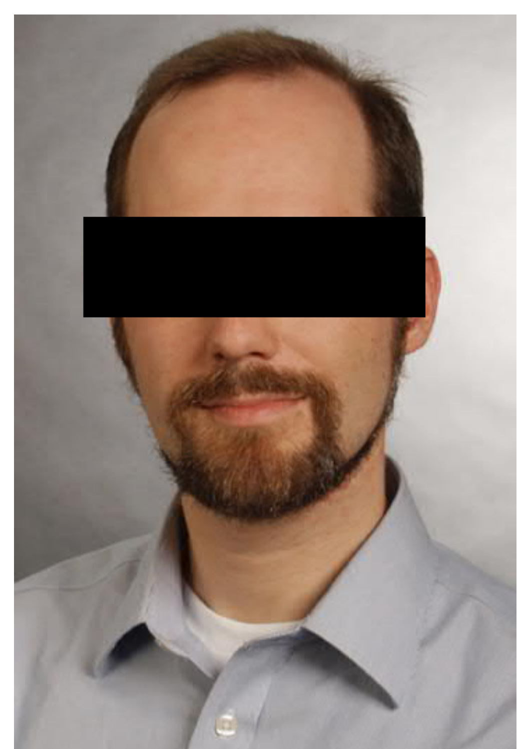
Confessions of Statistics Anonymous

QCHS XIII

[redacted]
Institut für Experimentelle Teilchenphysik, Karlsruher Institut für Technologie



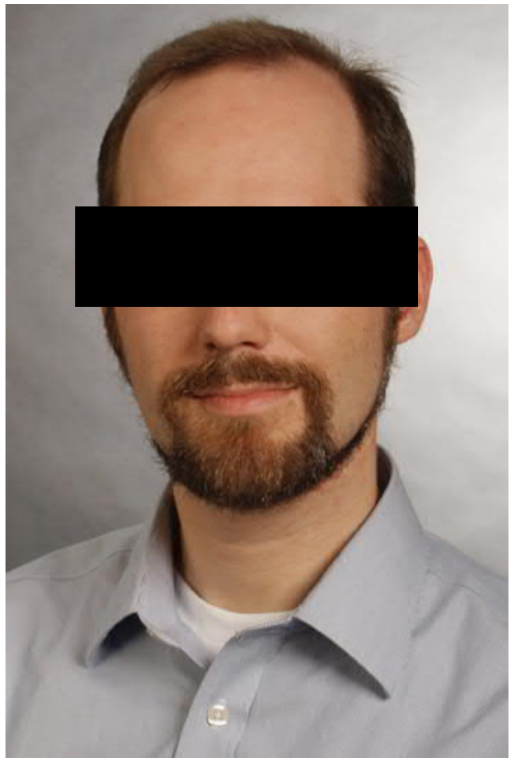
Who am I?

- Professional particle physicist
  => Jefferson Lab (hyperon polarization)
  => HERA (pdf fits)
  => LHC (electroweak and exotica)

- No formal statistics training

- Did some work on look-elsewhere-effect / trial factors for exotic searches

- Member of CMS statistics committee



WHY? HOW?

The 21st century

- Statistics in HEP largely done by physicists, not statisticians

- Traditional methods doing well
  => excellent existing implementations
  => overall reasonable education of physicists
  => progress mostly in scale (e.g. hundreds of nuisance parameters in fits)

- Biggest issue:
  => missing knowledge of bounds of applicability
  => odd corner-cases
  => physics judgement

THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

[redacted]

By applying the principle of maximum likelihood, J. Neyman and E. S. Pearson[2] have suggested a method for obtaining functions of observations for testing what are called *composite statistical hypotheses*, or simply *composite hypotheses*. The procedure is essentially as follows: A population K is assumed in which a variate x (x may be a vector with each component representing a variate) has a distribution function $f(x, \theta_1, \theta_2, \cdots \theta_h)$, which depends on the parameters $\theta_1, \theta_2 \cdots \theta_h$. A *simple hypothesis* is one in which the θ's have specified values. A set Ω of admissible hypotheses is considered which consists of a set of simple hypotheses. Geometrically, Ω may be represented as a region in the h-dimensional space of the θ's. A set ω of simple hypotheses is specified by taking all simple hypotheses of the set Ω for which $\theta_i = \theta_{0i}$, $i = m+1, m+2, \cdots h$.

Training needs to start early

Karlsruher Institut für Technologie

- Hands-on exercises during
  MS-level particle physics lectures

Binomial vs Clopper-Pearson

Übungsblätter

Zugang mit: student/StudentsOnly

Übung 1 Neutraler Strom (Gargamelle)	Übung 2 Symmetrie-Brechung	Übung 3 Trigger Praktikum (Material)	Übung 4 Z-Pol
Übung 5 Statistik Praktikum (Material)	Übung 6 Higgs $\rightarrow \gamma\gamma$	Übung 7 Machine Learning (Material)	

CLs by hand

Neural Networks

Major Effort: Training

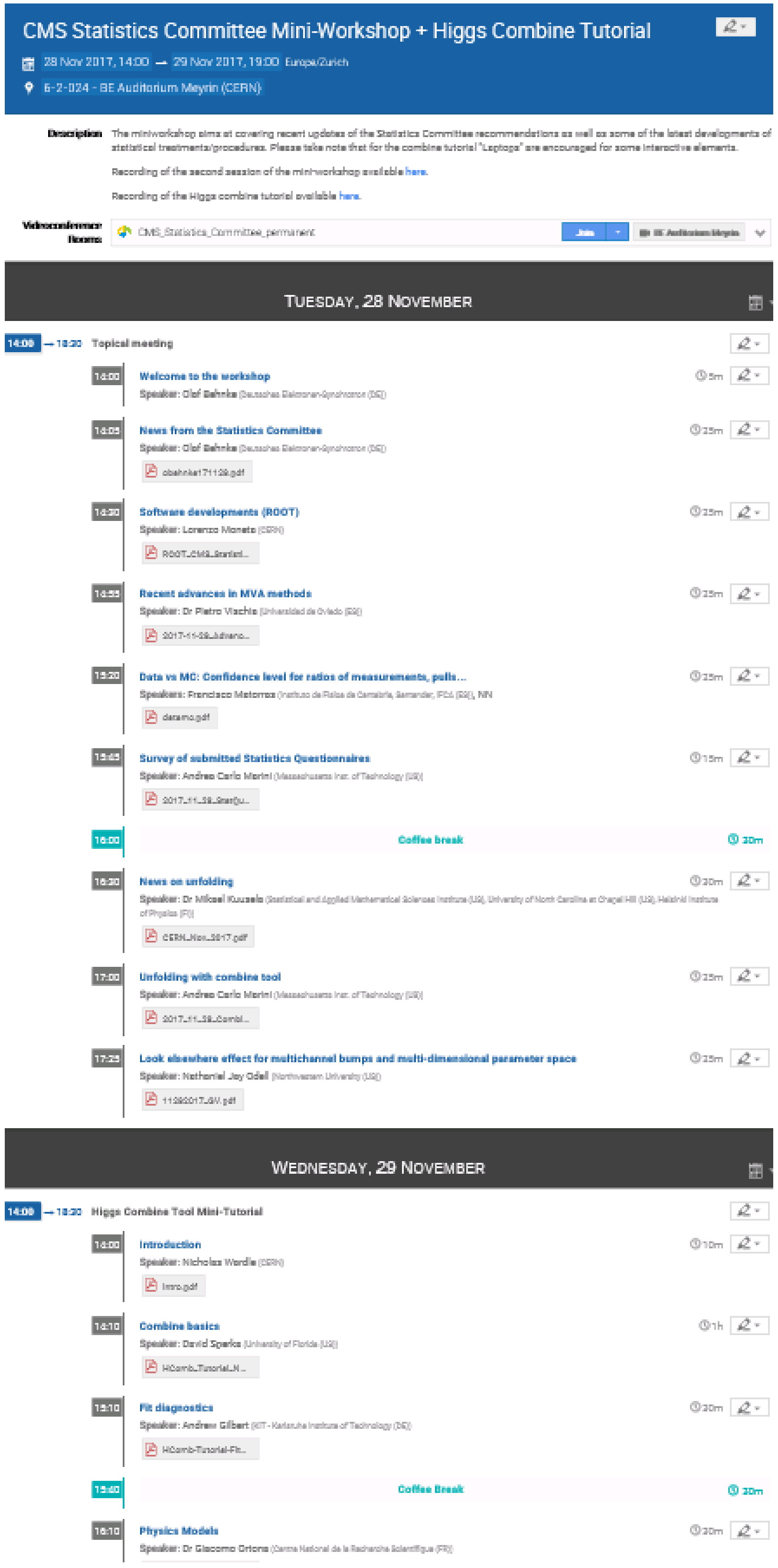
- Doesn’t apparently happen everywhere

  => needs additional training in CMS

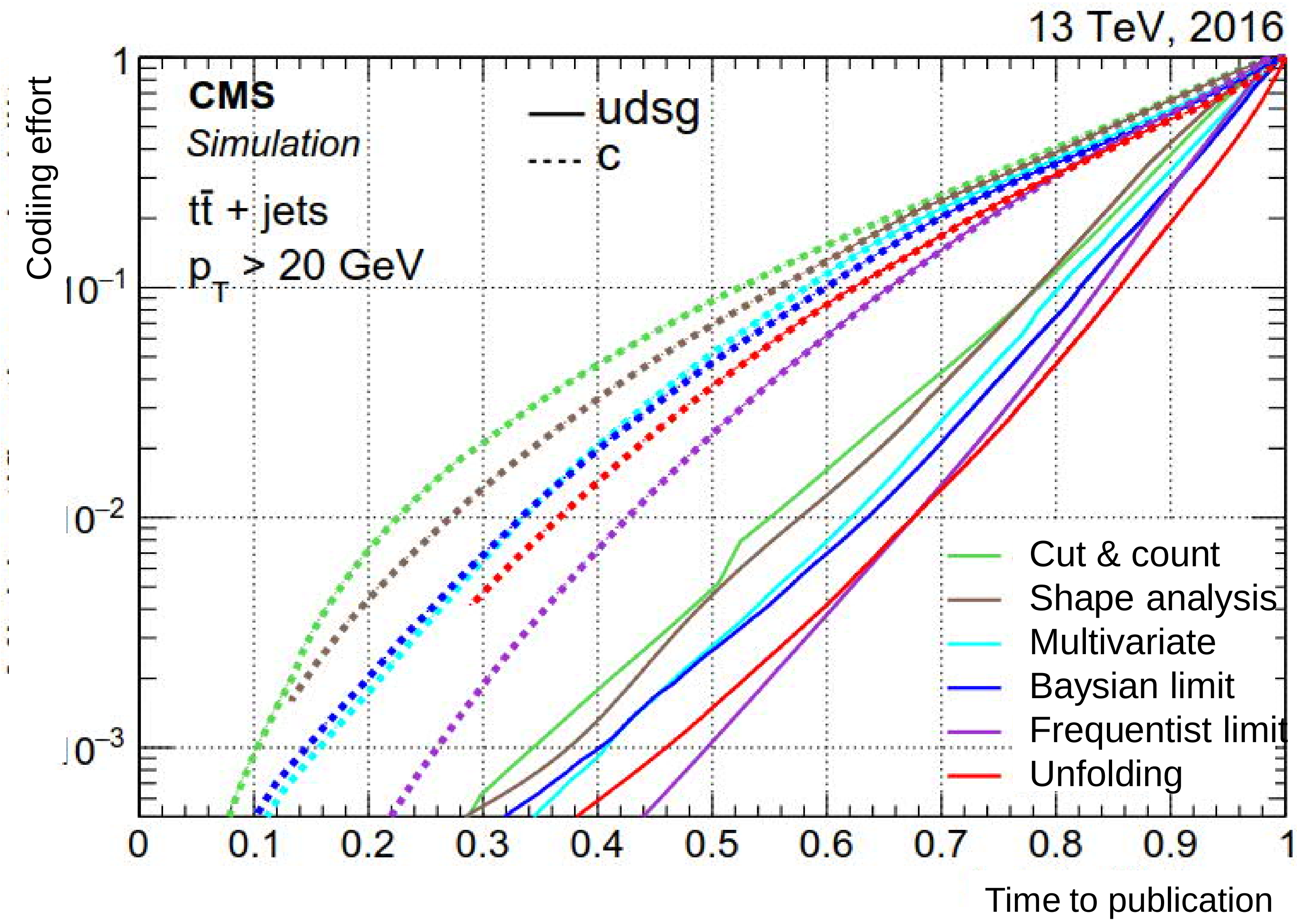
  => improve background knowledge
  - confidence intervals
  - limit setting
  - significance computations

  => practical tool use
  - focused on Higgs combination tool

trying to understand driving
forces behind analyzer decisions



The common perception



- Statistics often seen as black box
- Methods commonly chosen for utilitarian reasons:
  => tool availability
  => speedy publication

Why CMS analysts do CLs

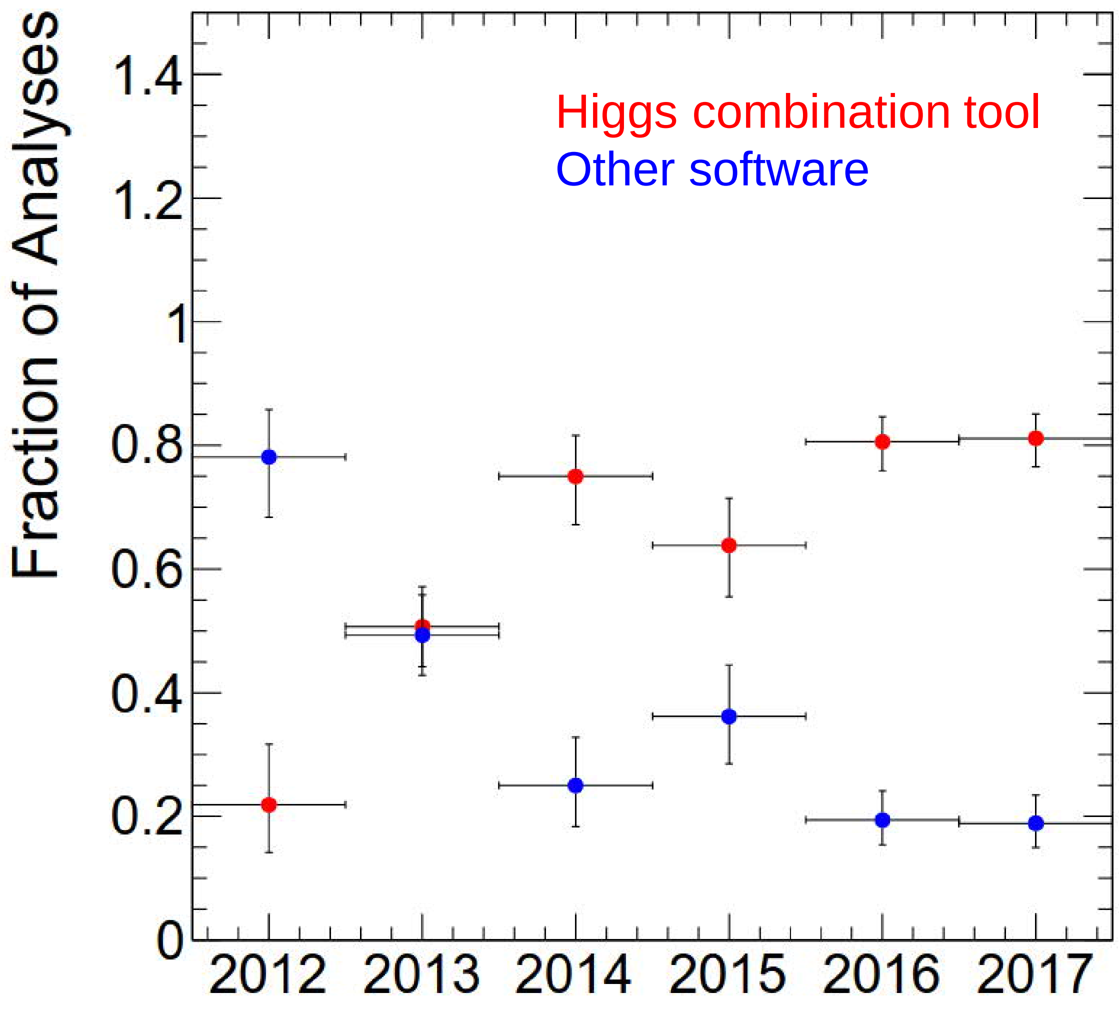
CMS and ATLAS agree to do use CLs for Higgs searches

↓

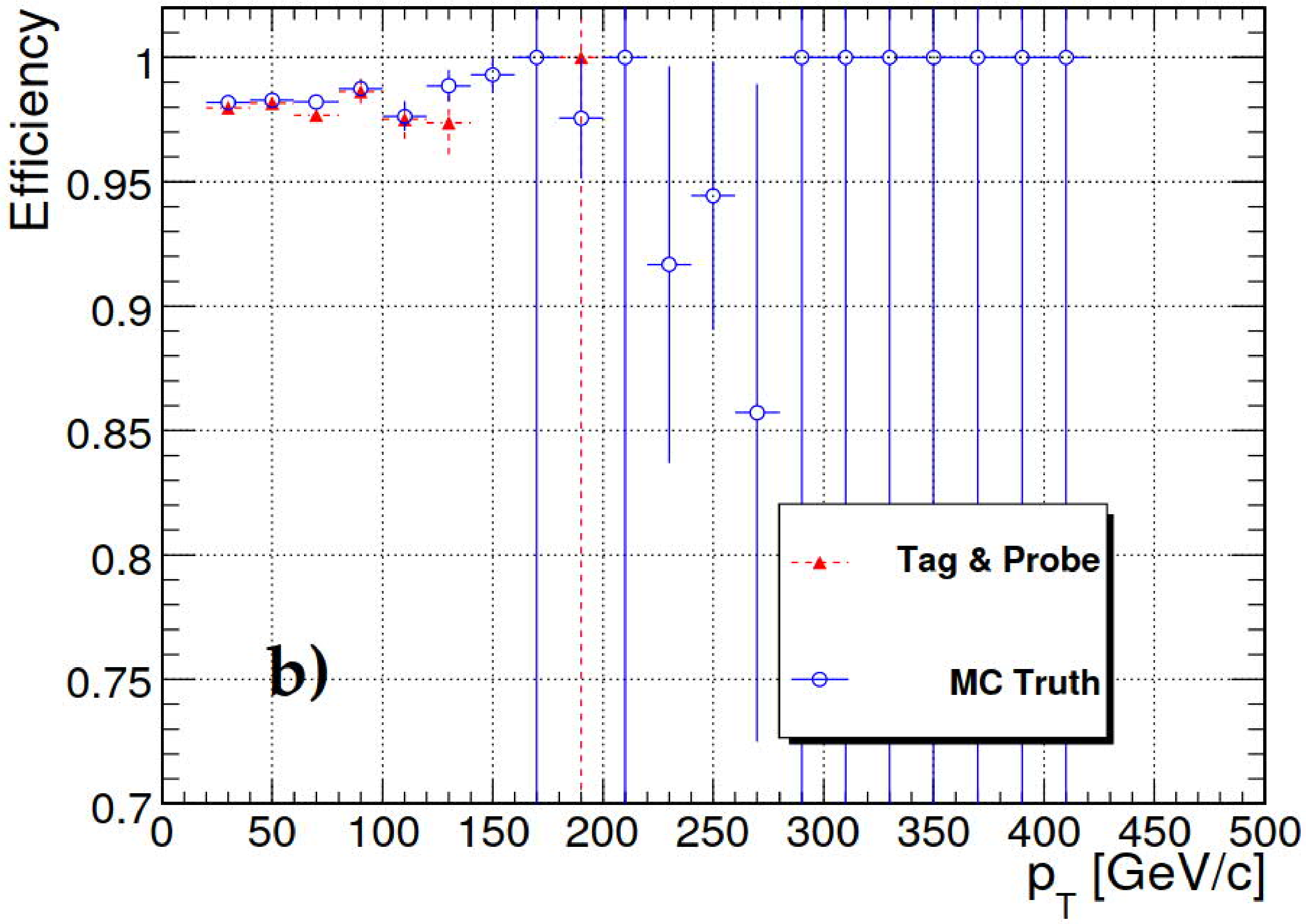
Higgs group develops “Higgs combination tool” to obtain CLs limits for combination of various channels (Nice tool, now public)

↓

Allows (relatively) simple limit calculations
Simplifies approval procedure



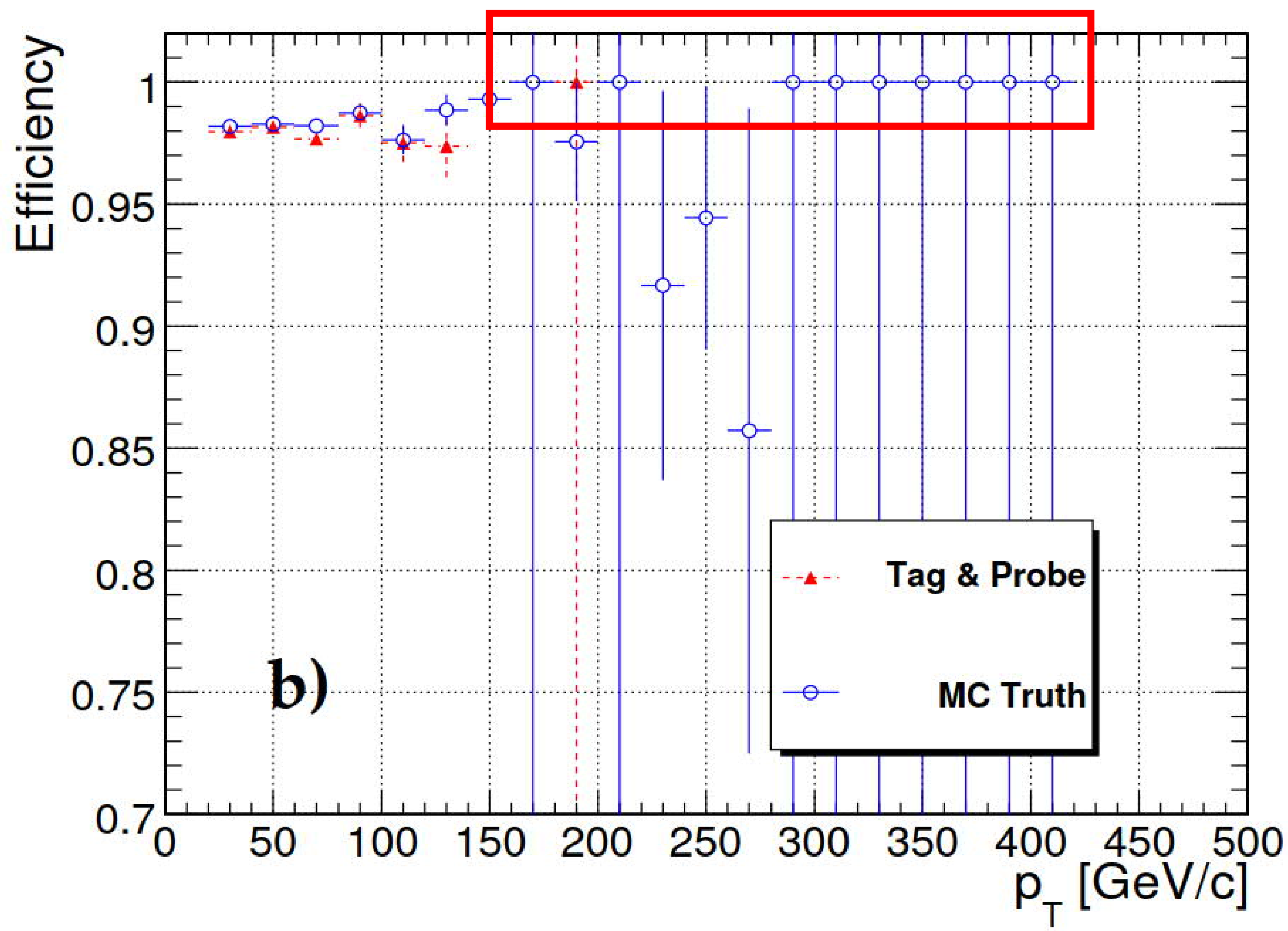
Issue: awareness of limitations



Plot made public as CMS physics analysis summary

Issue: awareness of limitations

- TH1::Divide assumes uncorrelated errors



Plot made public as CMS physics analysis summary

- Got a lot of help from TEfficieny => easy to use interface for all reasonable intervalshistograms

- Succesfully erradicated: poor error estimates for efficiencies

Issue: awareness of limitations

Karlsruher Institut für Technologie

- Example: Shape uncertainties in fits
  => Implemented in combine tool through template morphing
  => Increasingly widespread use through increased expertise
  in combination tool configuration

- Problem: template morphing technique most appropriate when templates have no relative fluctuations
  => same events, different event weights
  => same sample different subsets (e.g. energy scale vs cut value)
  => independent samples

Still an overall improvement in uncertainty treatment

Issue: awareness of limitations

- Asymptotic formulae of CLs criterion known analytically

- Speeds up CLs computation by orders of magnitude compared to toy-MC based evaluation
  => immediate and enthusiastic take-up by LHC community

Eur. Phys. J. C (2011) 71: 1554
DOI 10.1140/epjc/s10052-011-1554-0

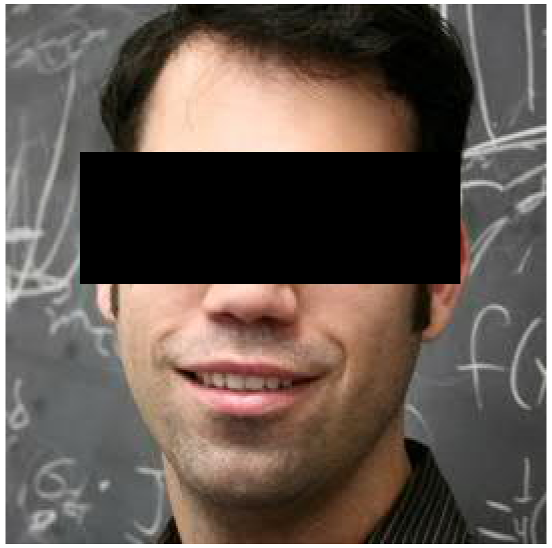
THE EUROPEAN PHYSICAL JOURNAL C

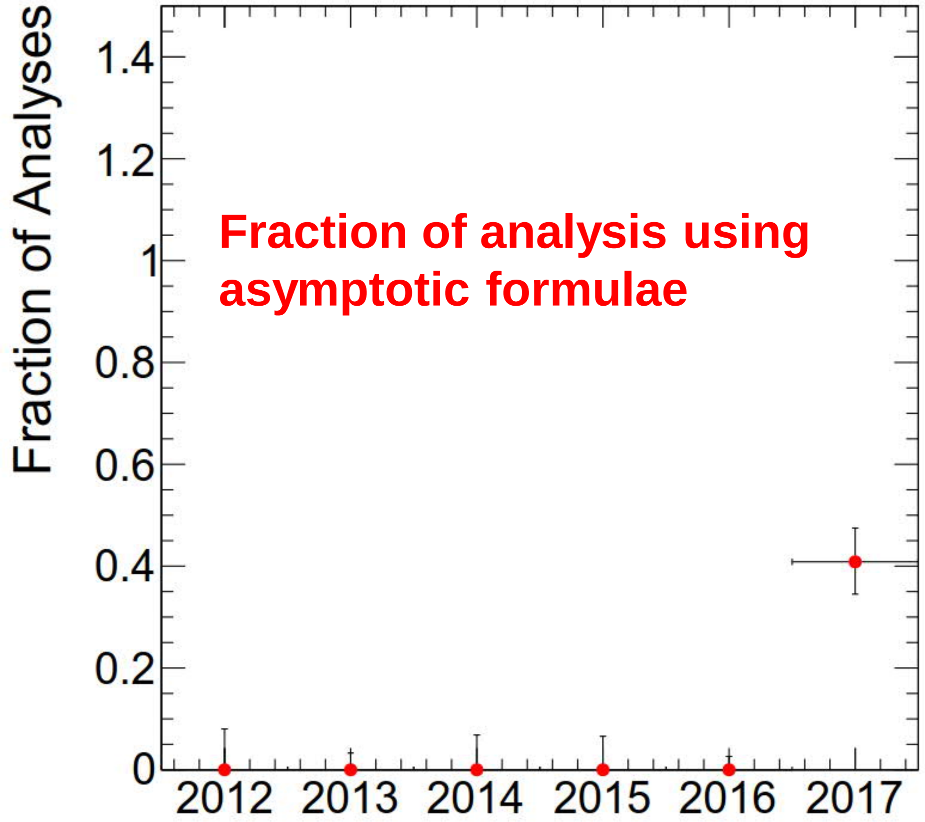
Special Article - Tools for Experiment and Theory

Asymptotic formulae for likelihood-based tests of new physics

[redacted]

[1] Physics Department, Royal Holloway, University of London, Egham TW20 0EX, UK
[2] Physics Department, New York University, New York, NY 10003, USA
[3] Weizmann Institute of Science, Rehovot 76100, Israel





Issue: awareness of limitations

Eur. Phys. J. C (2011) 71: 1554
DOI 10.1140/epjc/s10052-011-1554-0

THE EUROPEAN PHYSICAL JOURNAL C

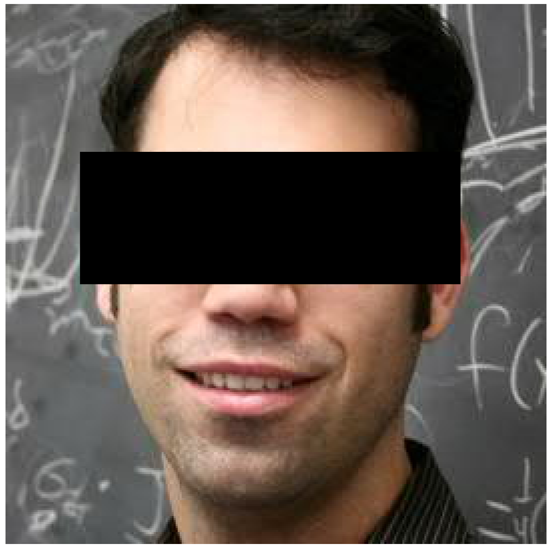
Special Article - Tools for Experiment and Theory

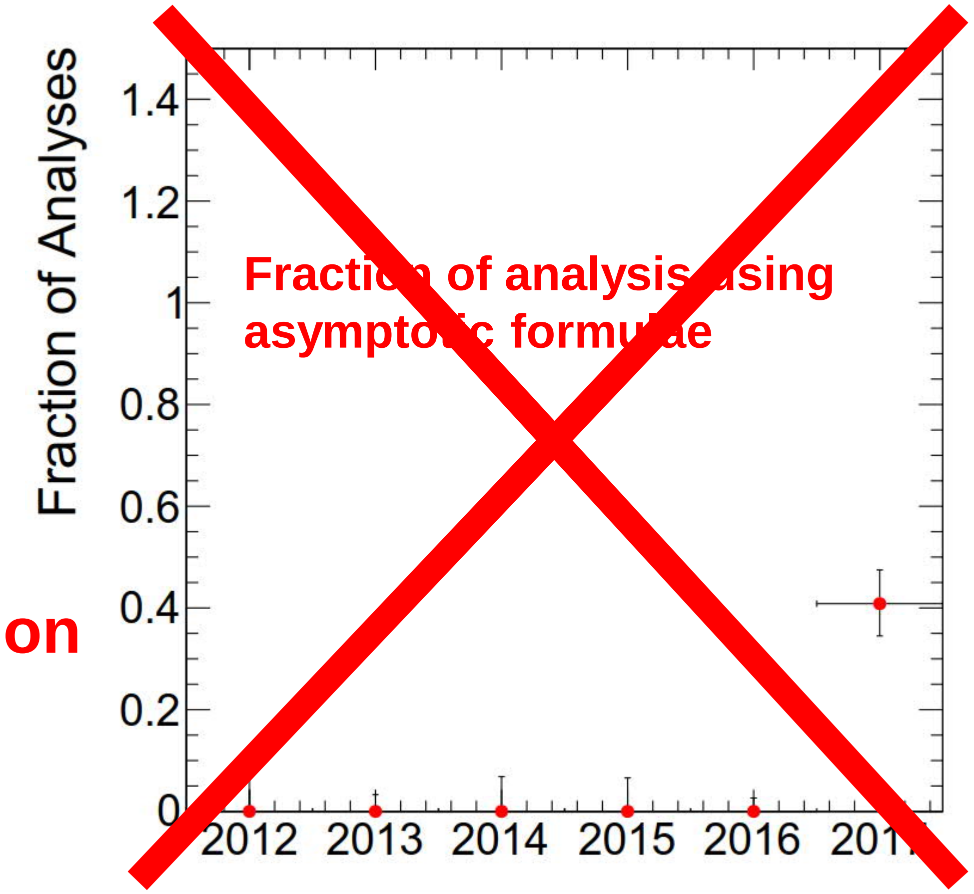
Asymptotic formulae for likelihood-based tests of new physics

[redacted]

[1]Physics Department, Royal Holloway, University of London, Egham TW20 0EX, UK
[2]Physics Department, New York University, New York, NY 10003, USA
[3]Weizmann Institute of Science, Rehovot 76100, Israel

- Asymptotic formulae of CLs criterion known analytically
- Speeds up CLs computation by orders of magnitude by compared to toy-MC based evaluation
  => immediate and enthusiastic take-up by LHC community





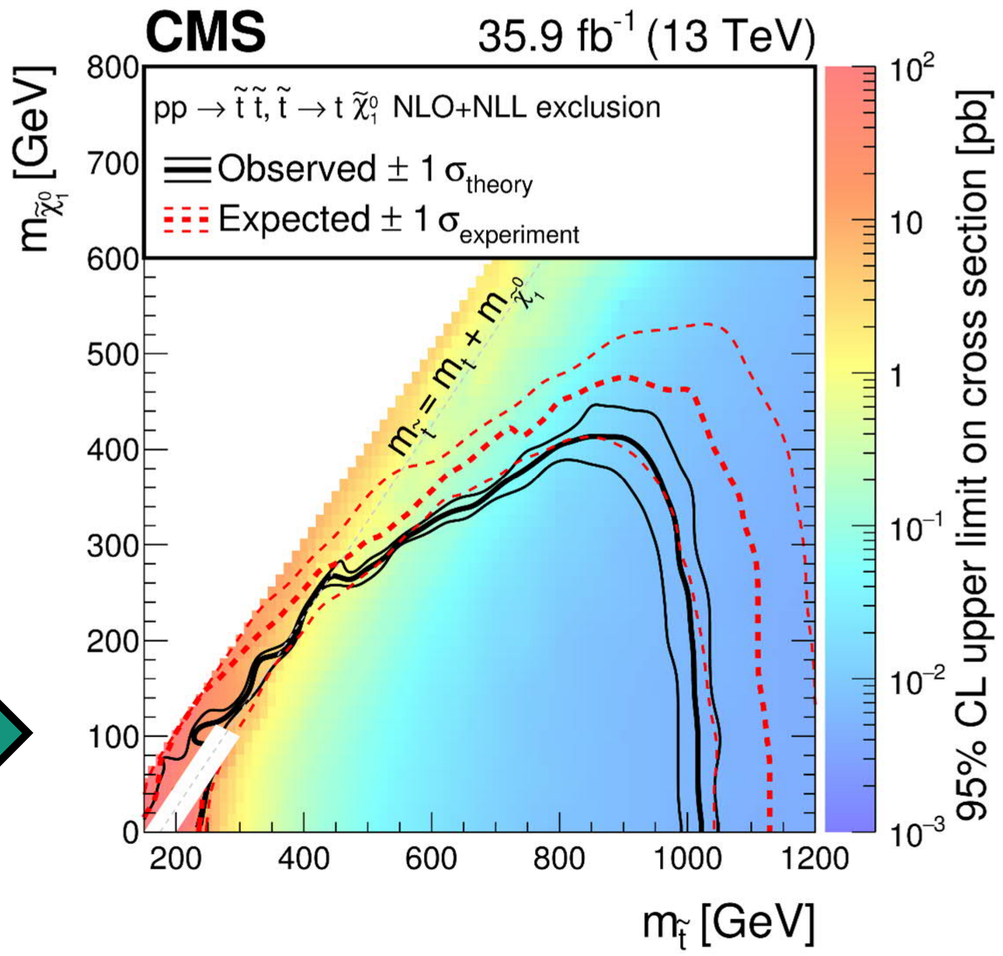
Issues with data-acquisition

Issue: awareness of limitations

- Asymptotic formulae are asymptotic:
  validity relies a „large“ event numbers
- How large is „large“?
  => actually quite small, handfull of events
  commonly enough for accurate result
- Need to evalute deficiencies
  case by case

Postive example:
Analysis evaluates toys on
coarse grid for correction

SUS-16-050



Typical Issue: “Conservative”

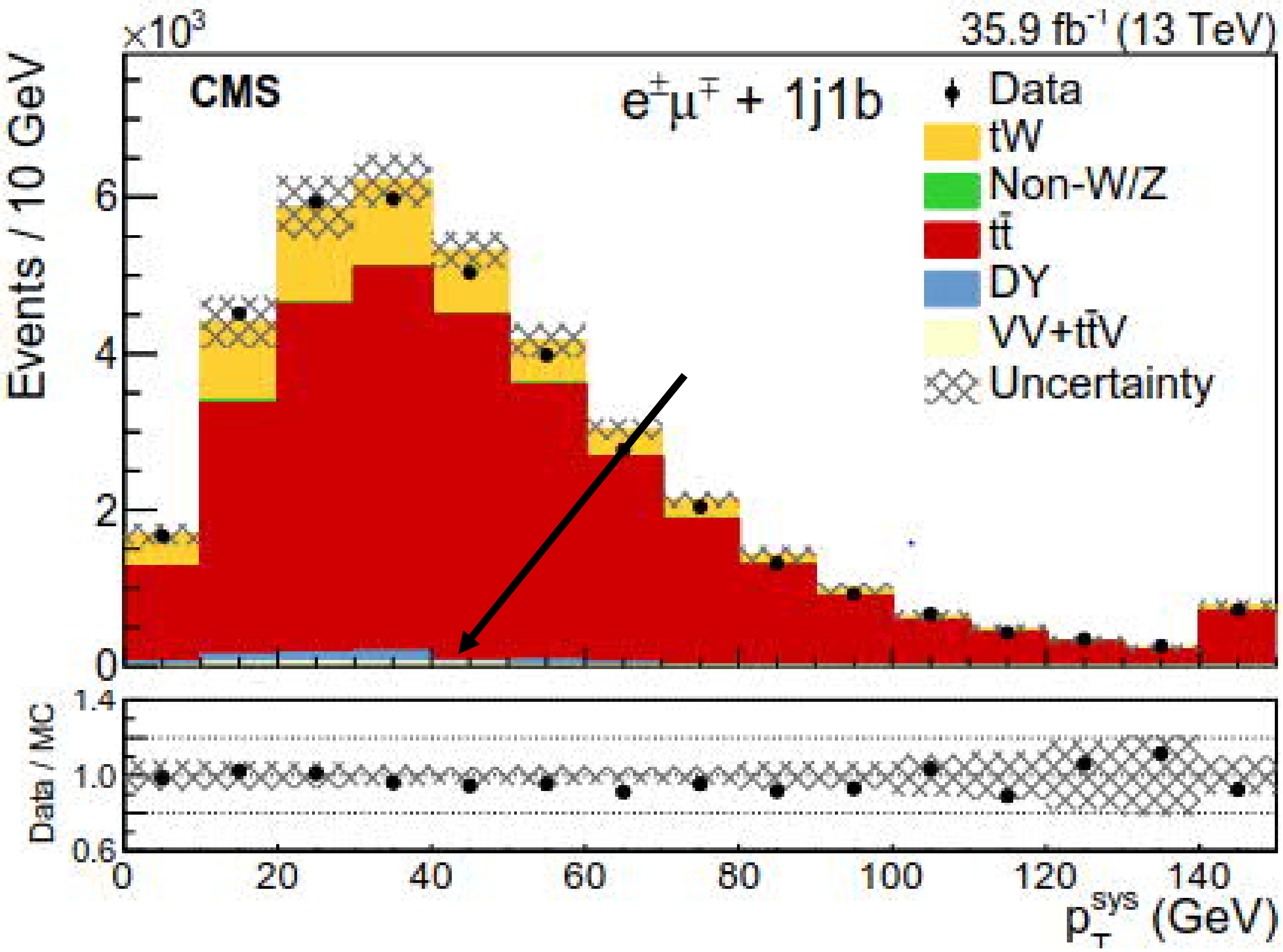
- twofold problem:
  => cross-checks vs uncertainties
  => analysis economy

- Requires serious physics judgement
  => impact on final result

- Requires serious considerations on
  => falsely claiming discovery
  vs missing an important discovery

Example I: No Problem

and μ_F scales, PDFs and α_S in the NNLO calculation [29]. For DY and non-W/Z backgrounds, a normalization uncertainty of $\pm 50\%$ is assumed. This value is motivated by the precision of



Good judgement:
expected improvement from more accurate error estimate is negligible

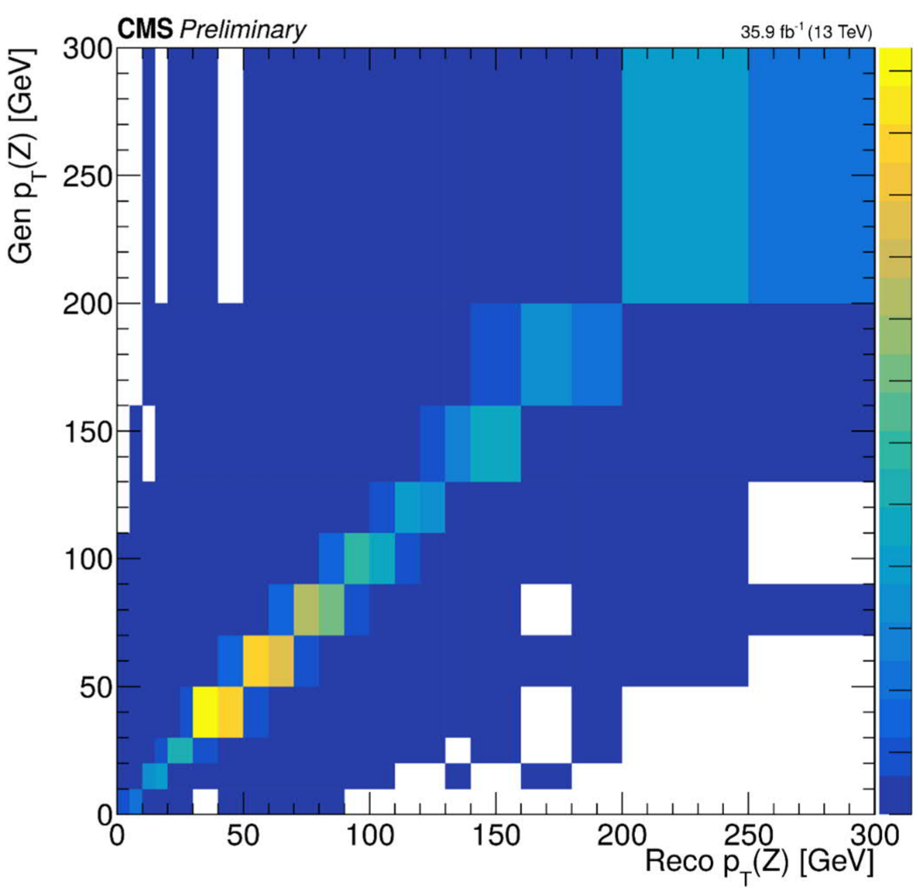
Example II: Tricky

- Background determined from control region
  => but control region gets significant signal contamination
- Originally appraised as conservative!
  => limits still competitive
  => no danger of false discovery!
- Originates with analyzers being entirely driven by producing best limits
- Ultimately driven by journals seeing „best limits“ as driving feature not necessarily best senistivity

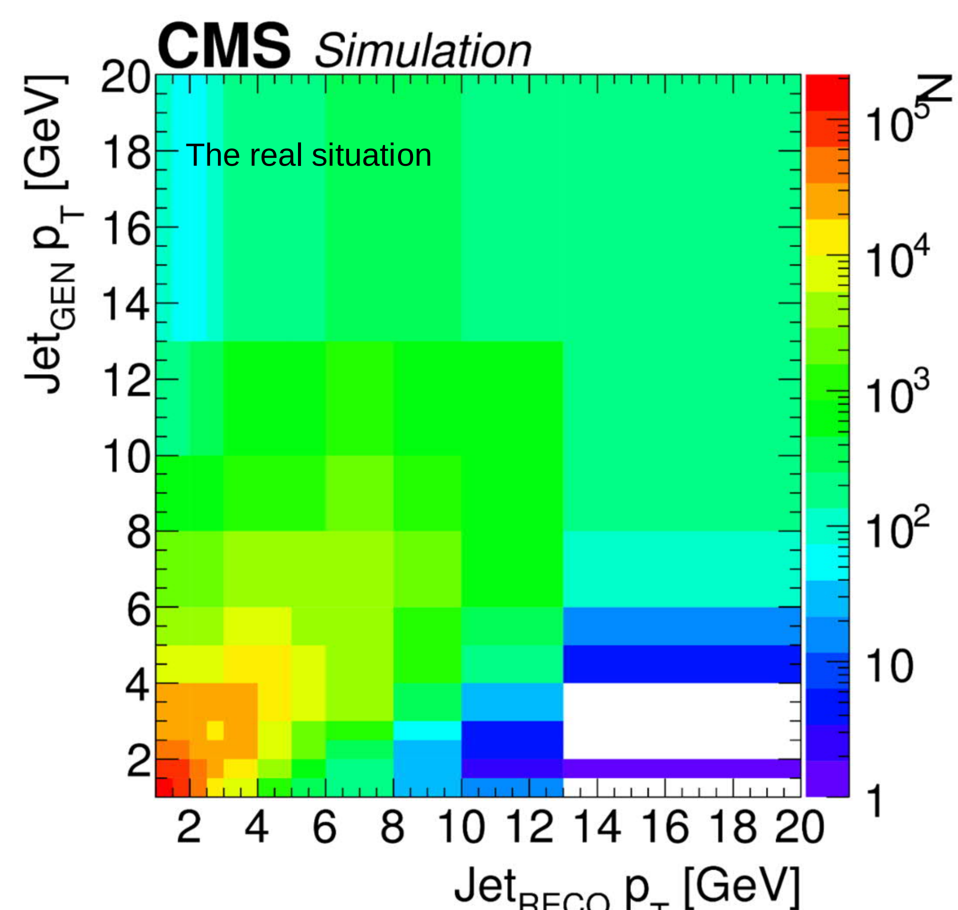
Common Issue: Unfolding

Karlsruher Institut für Technologie

- See talk by [redacted]
- Can be useful, but can it be done?
- Commonly requested by conveners without deeper thought

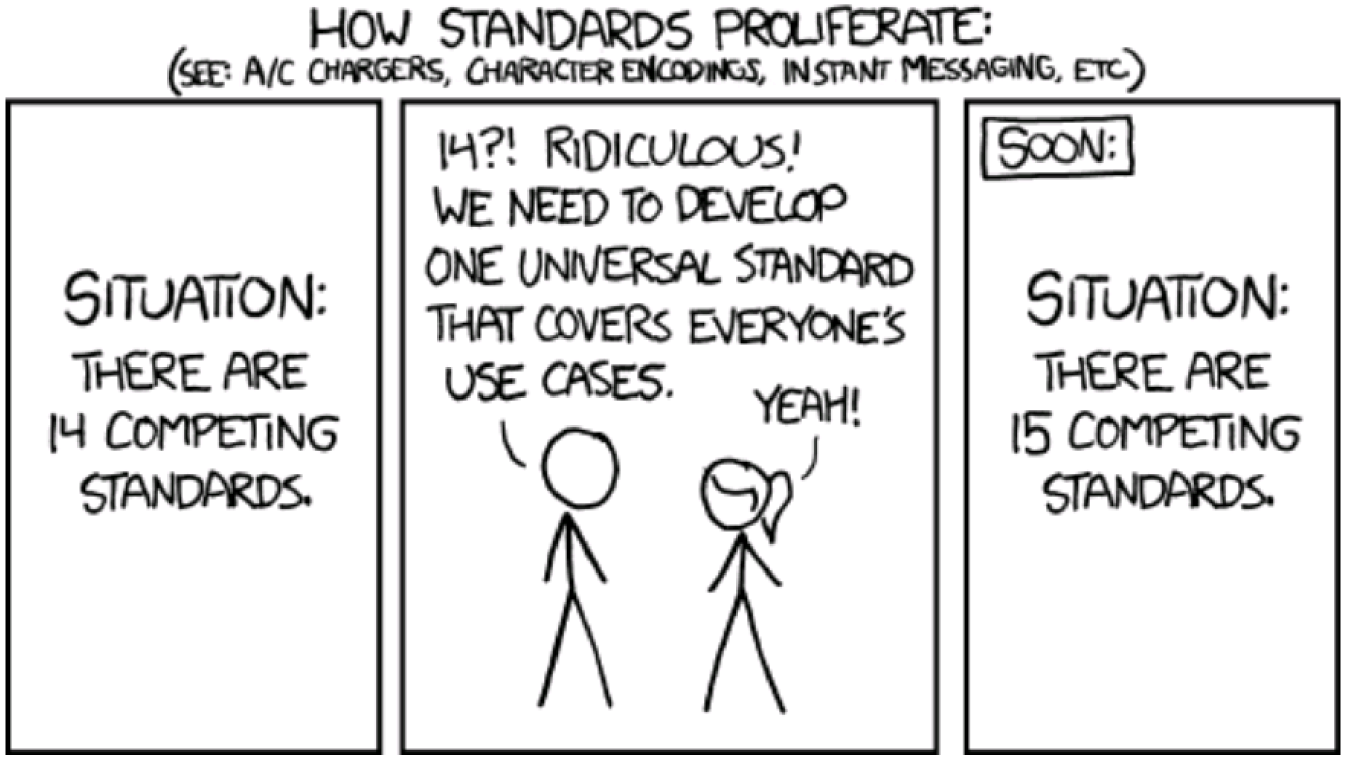


Consider training for management positions, not only analyzers



Common issues: Unfolding

- RooUnfold / Tunfold …
- Confusion about advantages/drawbacks of different methods.
  => SVD / Blobel / D‘Agostini
  => not helped by strong
  and opposing opinions
  of experts
- Provides the illusion of unfolding at the push of a button



- Criteria for choice often not particularly useful
  => „has been used previous iterations of the analysis“
  => „is the default“
  => „I don‘t do SM analysis any more bacause unfolding is so terrible“

Conclusion

- Uptake of reasonable practices for publications in CMS is good
- Requires continued effort of education and training
  => always new students joining
- Remaining issues often related to physics judgement, missing knowledge on details of applicability of methods
- For new ideas / methods and tools, important question should be
  => how easy is it to use?
  => how easy is it to misuse?
- Solutions need to be correct AND practicable,
  only one is not enough

Topics for Discussion

- Ways to improve education:
  => better tool documentations
  => relevant examples, highlighting typical pitfalls

- Incentives:
  => publications / best limits
  => inner-experiment: what gets pushed to be published
  => more explicitly state/discuss tradeoff of false discovery vs overlooking signal

Backup

Why CLs for the Higgs

Why isn't every physicist a Bayesian?

[redacted]

Department of Physics, University of California, Los Angeles, California 90024-1547

(Received 1 June 1994; accepted 3 November 1994)

Physicists embarking on seemingly routine error analyses are finding themselves grappling with major conceptual issues which have divided the statistics community for years. While the philosophical aspects of the debate may be endless, a practicing experimenter must choose a way to report results. The results can depend on which of the two major frameworks, classical or Bayesian, one adopts. This article reviews reasons why most data analysis in particle physics has traditionally been carried out within the classical framework, and why this will probably continue to be the case. However, Bayesian reasoning has recently made significant inroads in some published work in this field, and many other particle physicists may frequently think in a Bayesian manner without realizing it. I illustrate the issues involved with a few simple, commonly encountered examples which reveal how each framework can sometimes lead to unsatisfying results.

Why we're still all Baeysian

Karlsruher Institut für Technologie

6. How many σ for discovery? [redacted]

Apply subconscious Bayes factor p(H1|x)/p(H0|x) * LEE * Worries about Systematics

SEARCH	SURPRISE	IMPACT	LEE	SYSTEMATICS	No. σ
Higgs search	Medium	Very high	M	Medium	5
Single top	No	Low	No	No	3
SUSY	Yes	Very high	Very large	Yes	7
B_s oscillations	Medium/Low	Medium	Δm	No	4
Neutrino osc	Medium	High	$\sin^2 2\vartheta$, Δm^2	No	4
$B_s \rightarrow \mu\,\mu$	No	Low/Medium	No	Medium	3
Pentaquark	Yes	High/V. high	M, decay mode	Medium	7
$(g-2)_\mu$ anom	Yes	High	No	Yes	4
H spin $\neq$ 0	Yes	High	No	Medium	5
4^{th} gen q, l, ν	Yes	High	M, mode	No	6
Dark energy	Yes	Very high	Strength	Yes	5
Grav Waves	No	High	Enormous	Yes	8