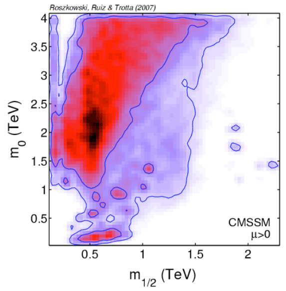
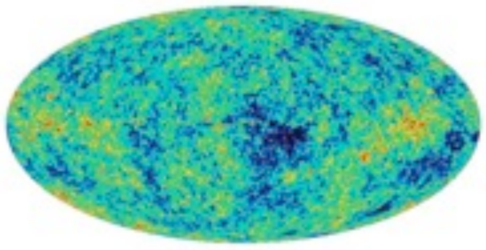
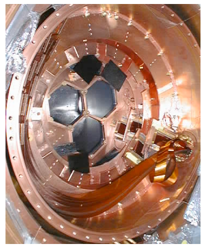
PHYSTAT2011 - CERN, Jan 2011

Statistical challenges of global SUSY fits

Kyle Cranmer
NYU

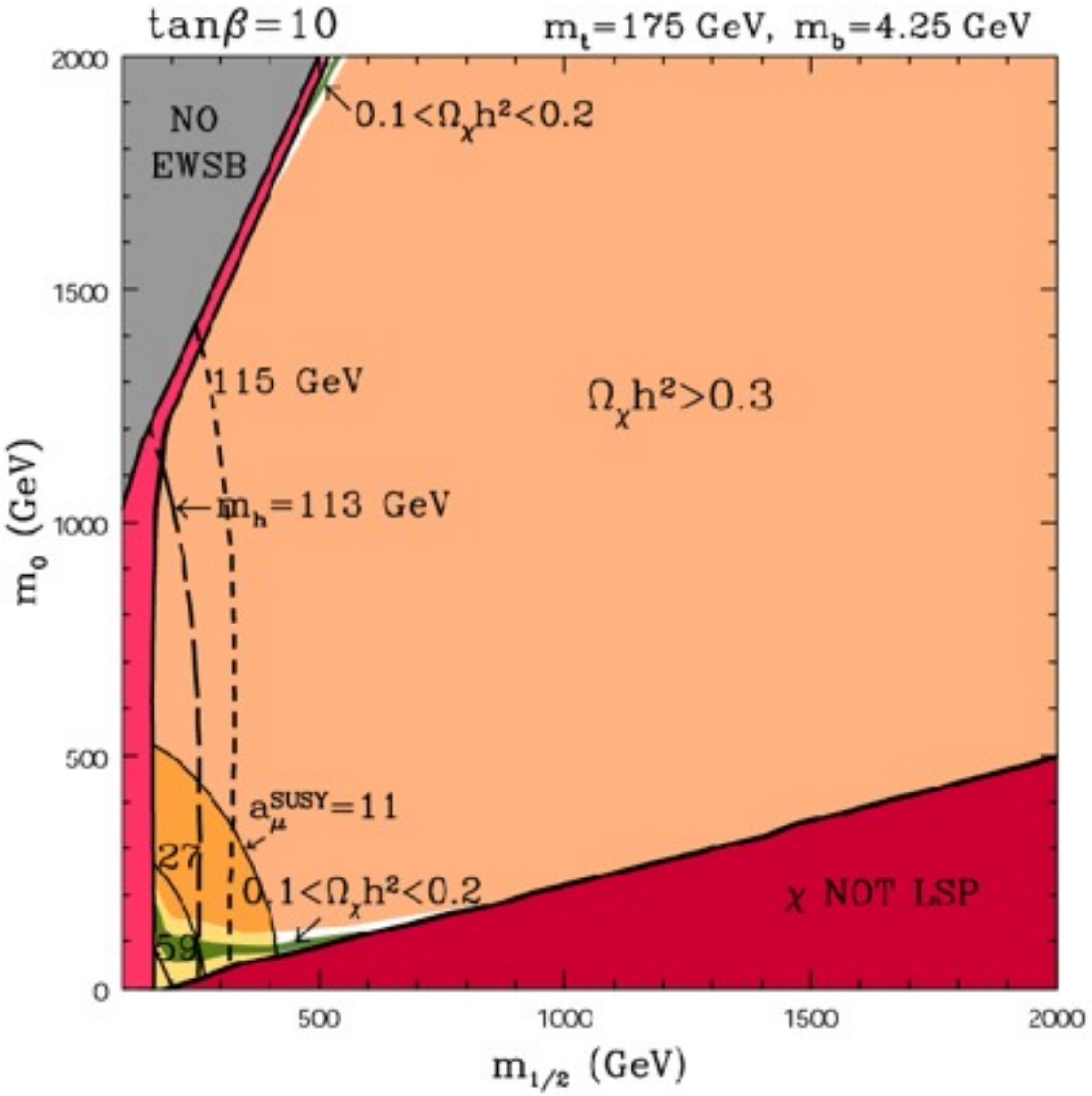
Roberto Trotta
Imperial College London

1. Status of global SUSY fits
2. Challenges of profile likelihood evaluation
3. Coverage of CMSSM intervals

Recent developments:
Bridges, KC, RT et al [arXiv: 1011.4306]
Akrami et al [arXiv: 1011.4297]
Feroz, KC, RT et al [arXiv:1101.3296]

2D scans

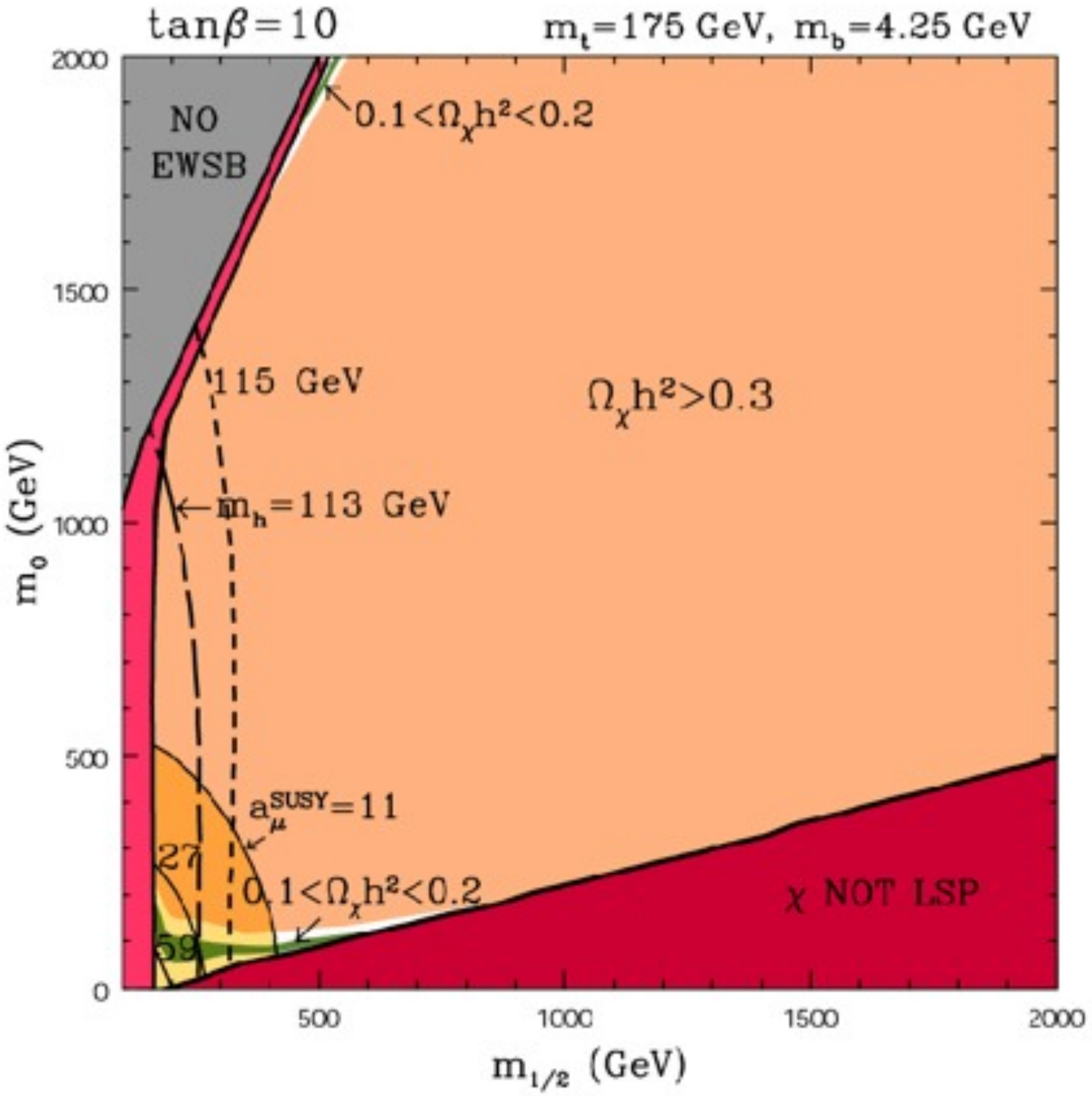
Determining constraints on SUSY models is a multi-dimensional problem. Even in one of the simplest cases, the CMSSM, there are four 4 parameters (M_0, $M_{1/2}$, A_0, $\tan\beta$) as well as SM parameters (e.g. M_{top}, M_b) The traditional strategy in the field was to carry out "2D scans" by fixing the other relevant parameters to certain values.

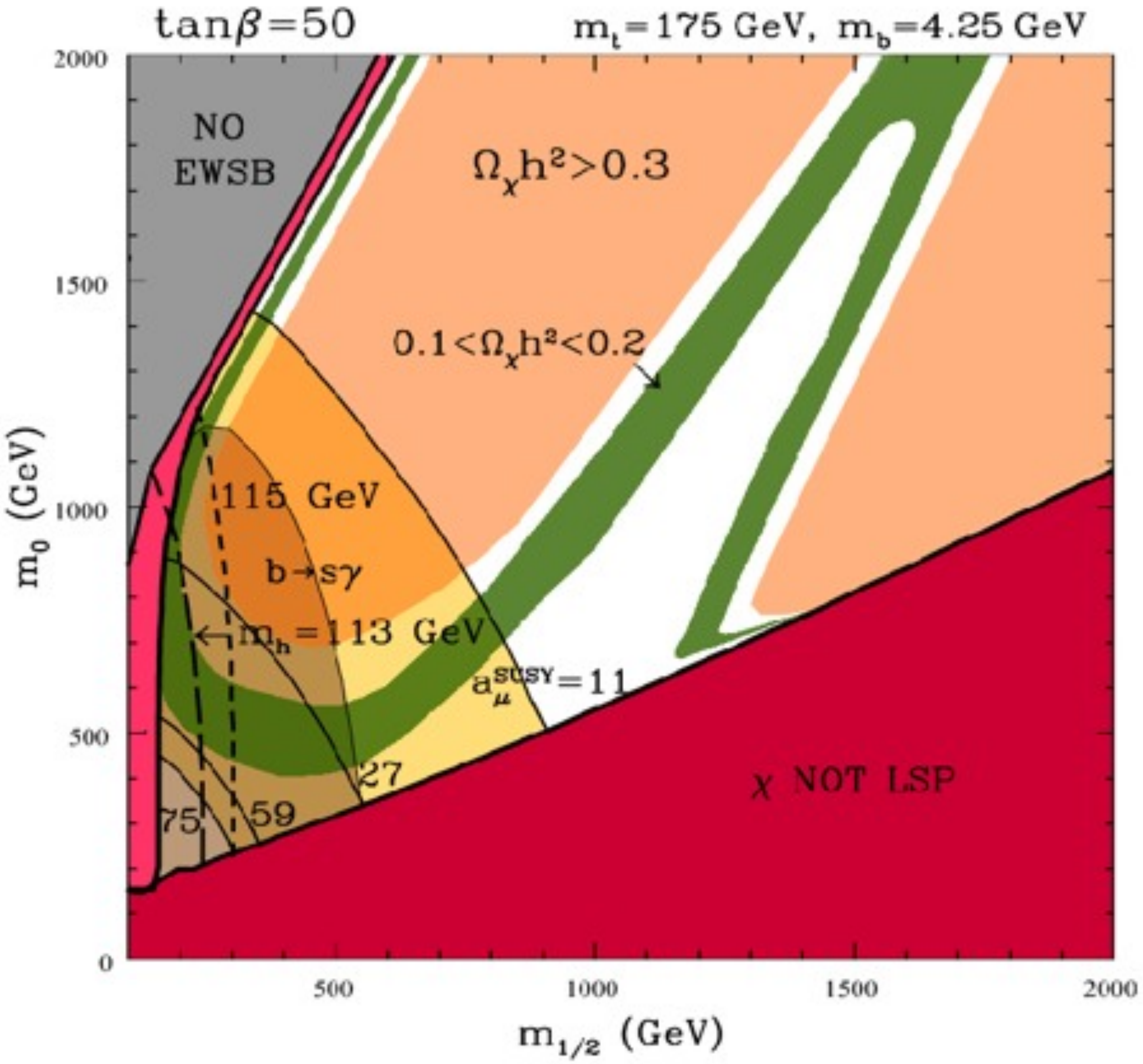


Roszkowski et al (2001)

2D scans

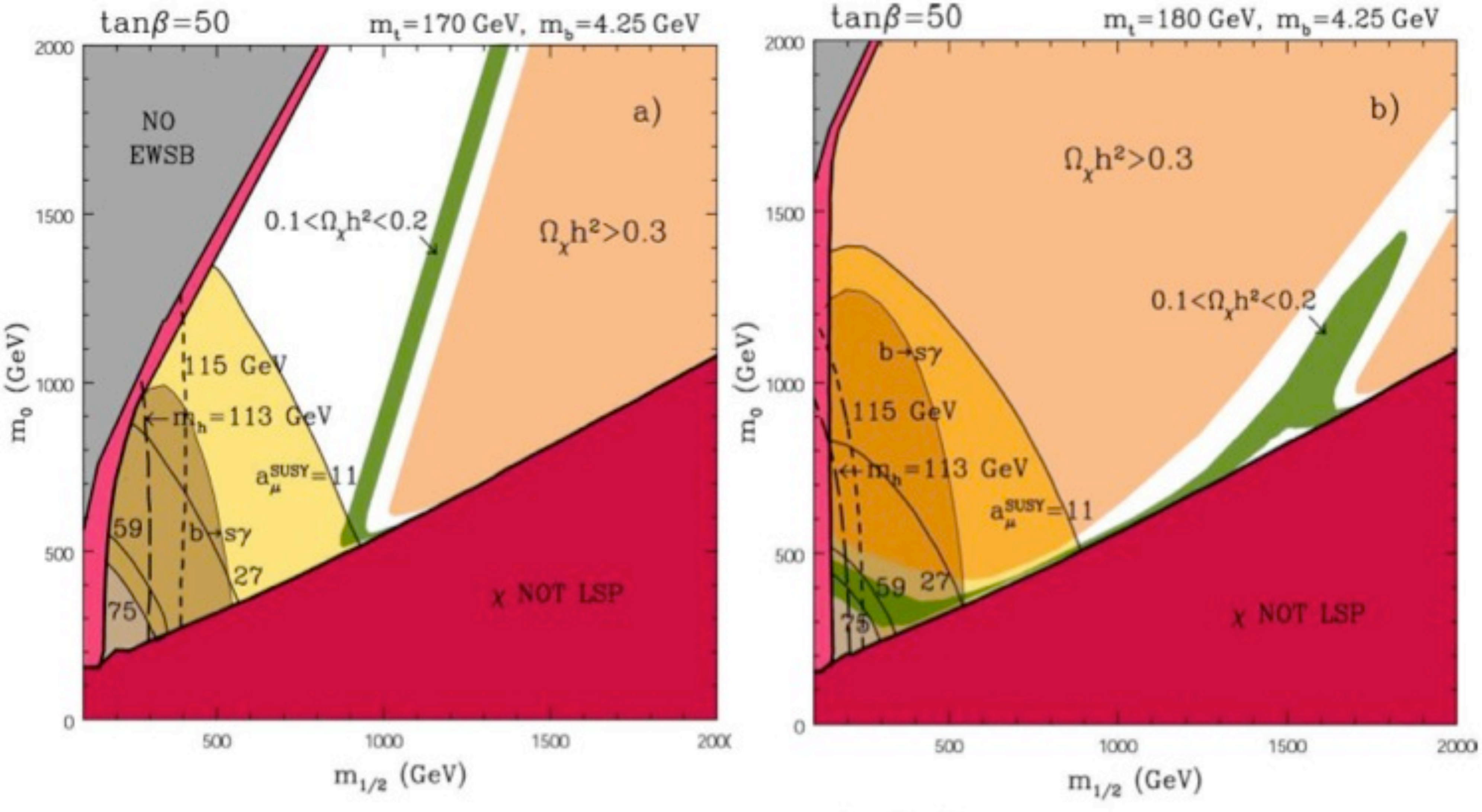
Determining constraints on SUSY models is a multi-dimensional problem. Even in one of the simplest cases, the CMSSM, there are four 4 parameters (M_0, $M_{1/2}$, A_0, tanβ) as well as SM parameters (e.g. M_{top}, M_b) The traditional strategy in the field was to carry out "2D scans" by fixing the other relevant parameters to certain values.





Roszkowski et al (2001)

Dependency on SM (nuisance) parameters



M_{top}=170 GeV

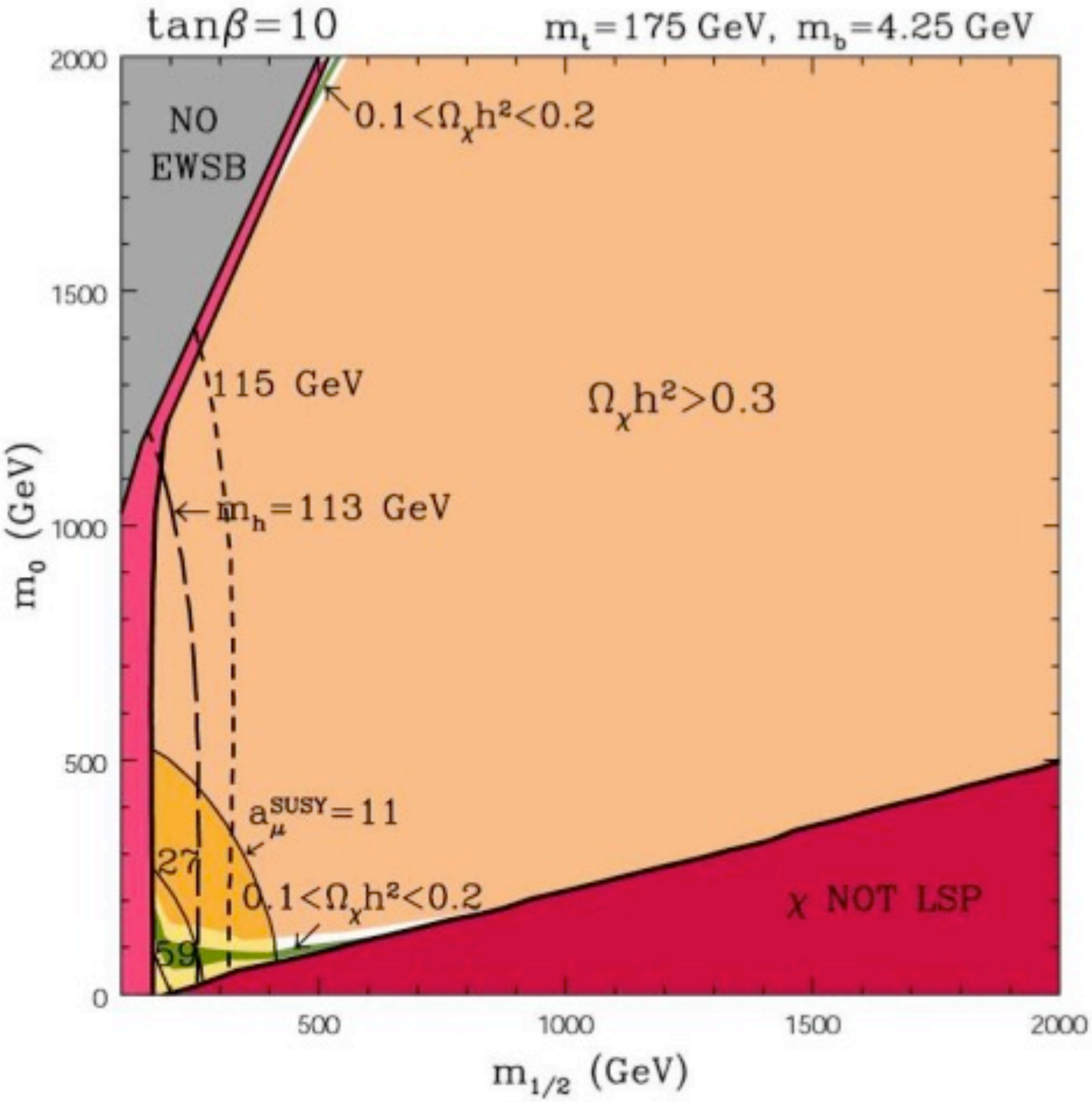
M_{top}=180 GeV

There is also a strong dependence on the important SM parameters!
(which are known only with limited accuracy)

Solution: global fits

Carry out a **simultaneous fit** of all relevant SUSY and SM parameter to the experimental data/constraints.

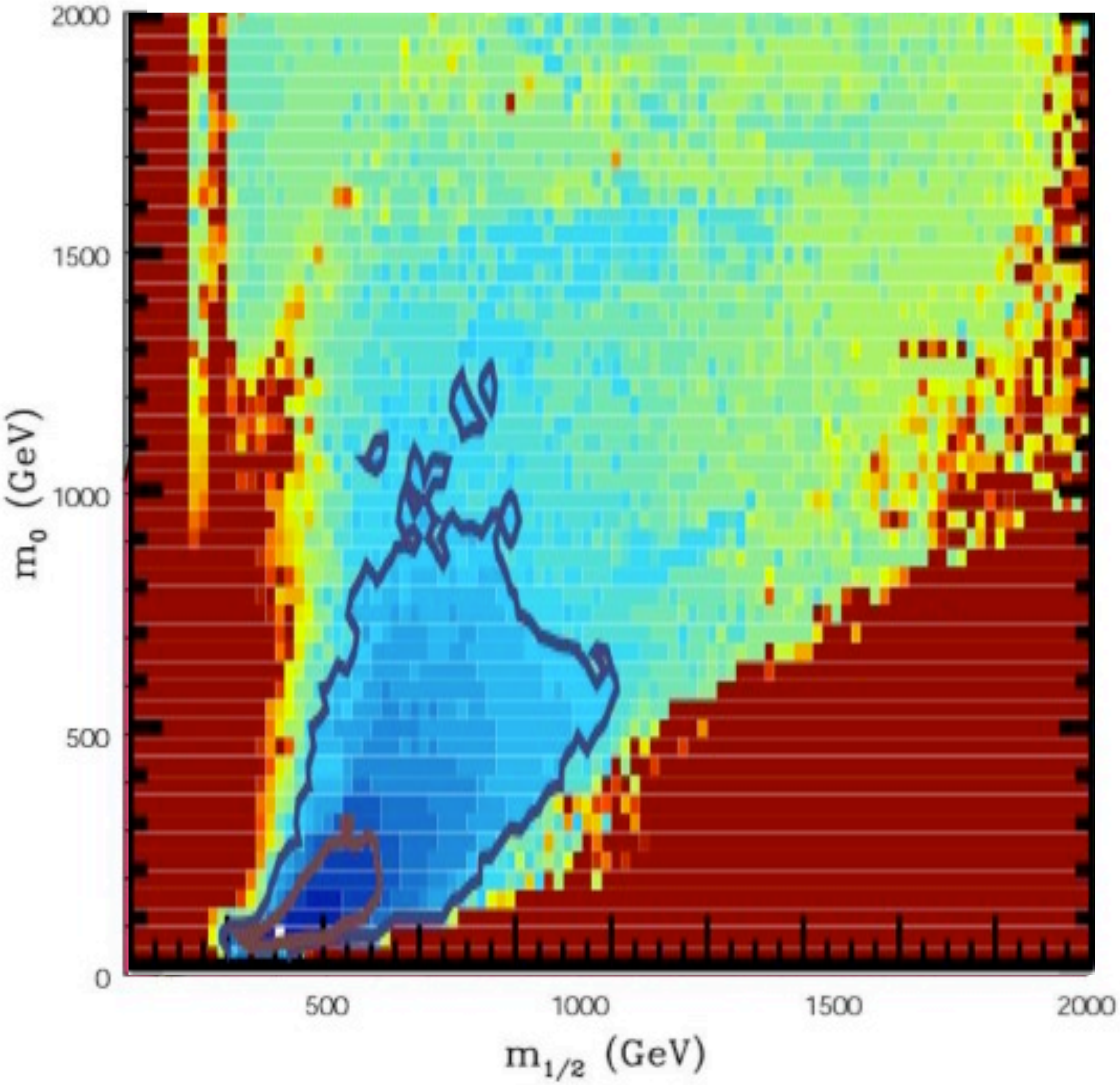
Marginalize (= integrate) or maximise along the hidden dimensions to obtain results that account for the multi-dimensional nature of the problem.



Solution: global fits

Carry out a **simultaneous fit** of all relevant SUSY and SM parameter to the experimental data/constraints.

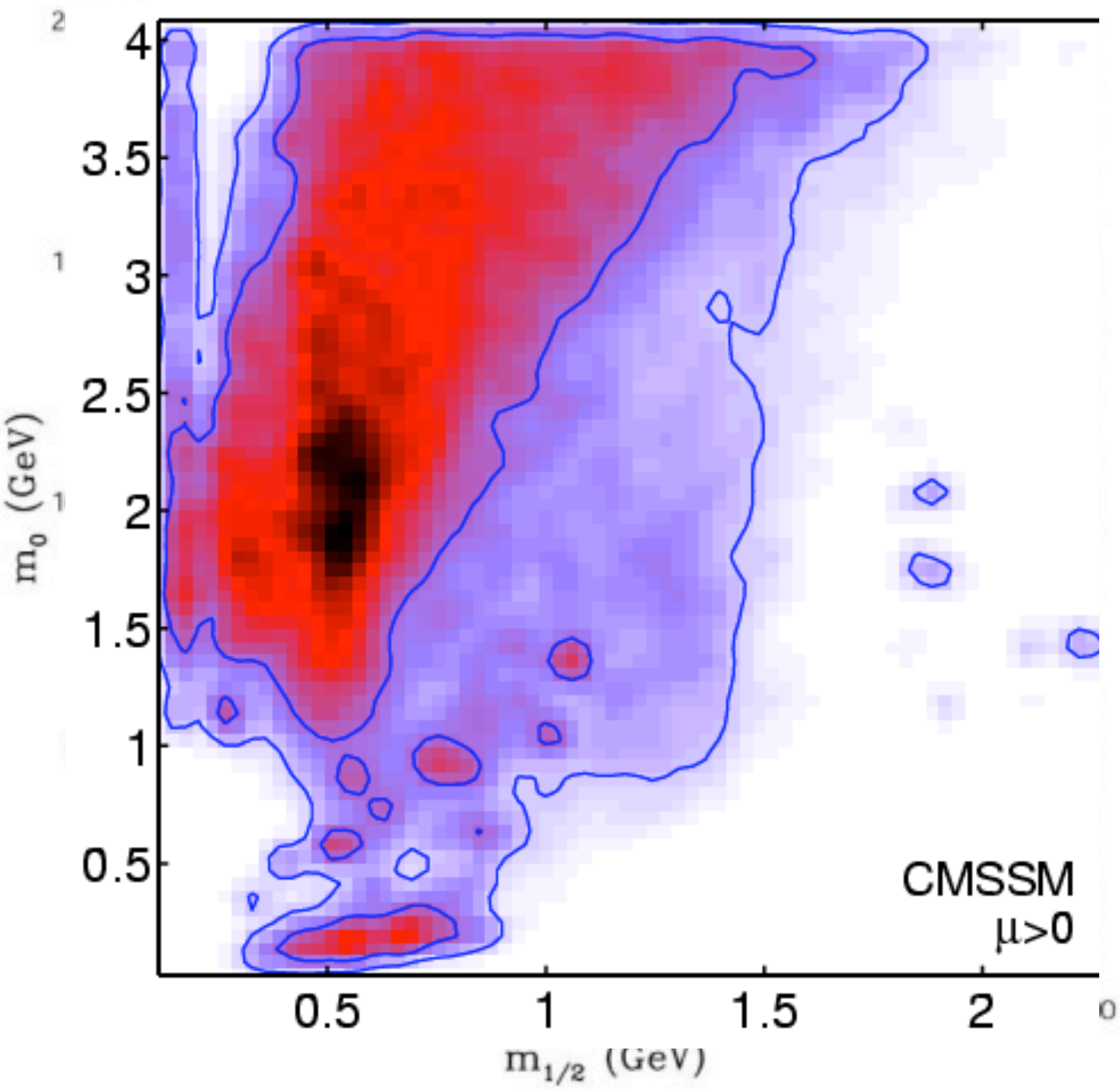
Marginalize (= integrate) or maximise along the hidden dimensions to obtain results that account for the multi-dimensional nature of the problem.



Solution: global fits

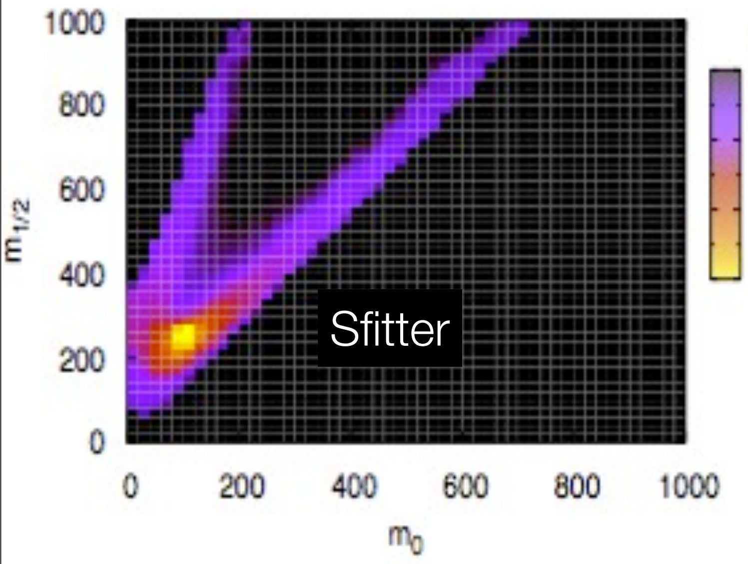
Carry out a **simultaneous fit** of all relevant SUSY and SM parameter to the experimental data/constraints.

Marginalize (= integrate) or maximise along the hidden dimensions to obtain results that account for the multi-dimensional nature of the problem.

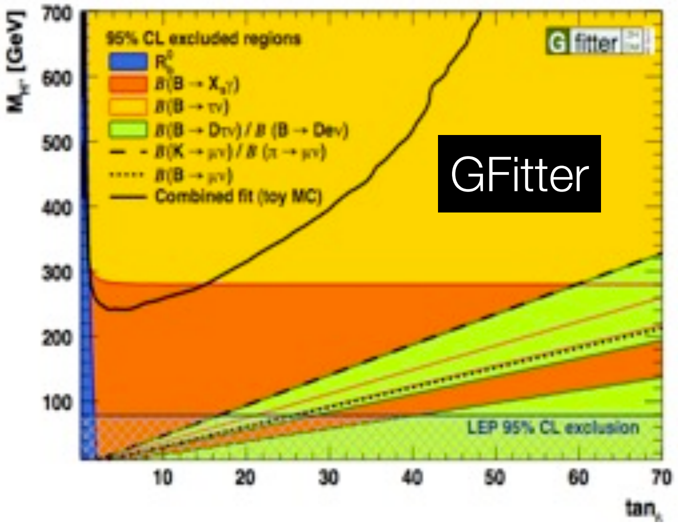


Global fits in the LHC era

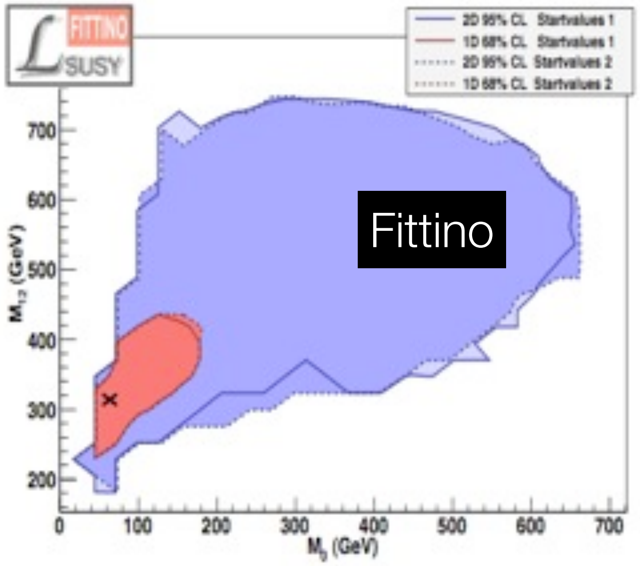
R. Lafaye , M. Rauch, T. Plehn, D. Zerwas



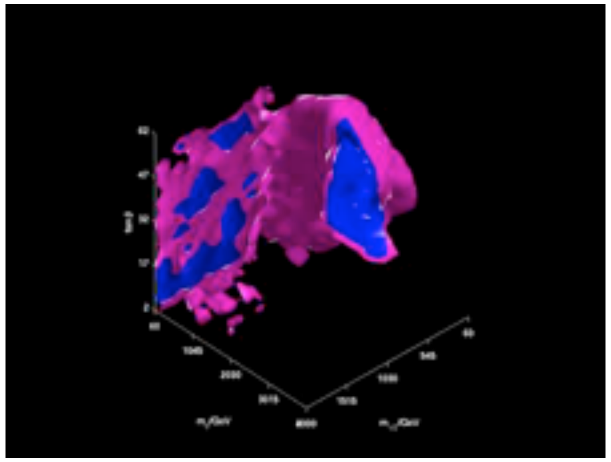
H. Flächer, M. Goebel, J. Haller,
A. Höcker, K. Mönig, J. Stelzer



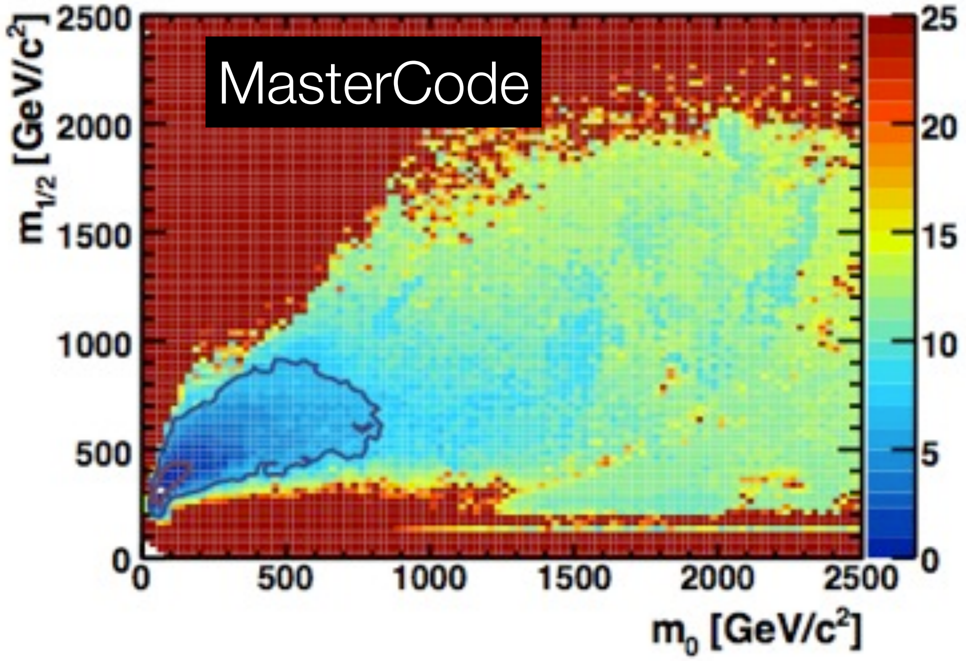
P. Bechtle, K. Desch, M. Uhlenbrock, P.
Wienemann



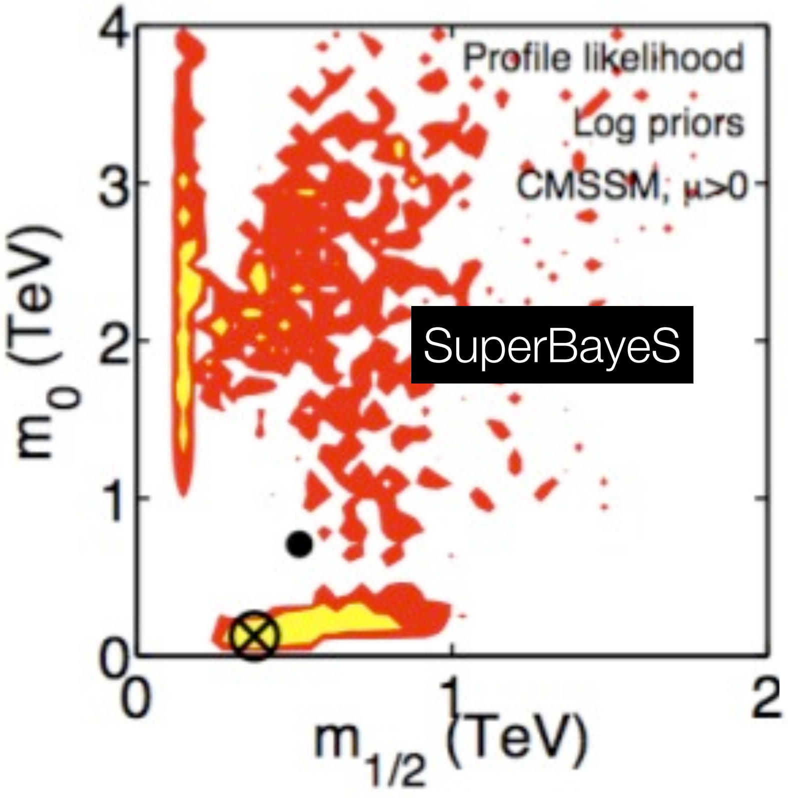
B.C. Allanach, K.Cranmer,
C.G. Lester, A.M. Weber



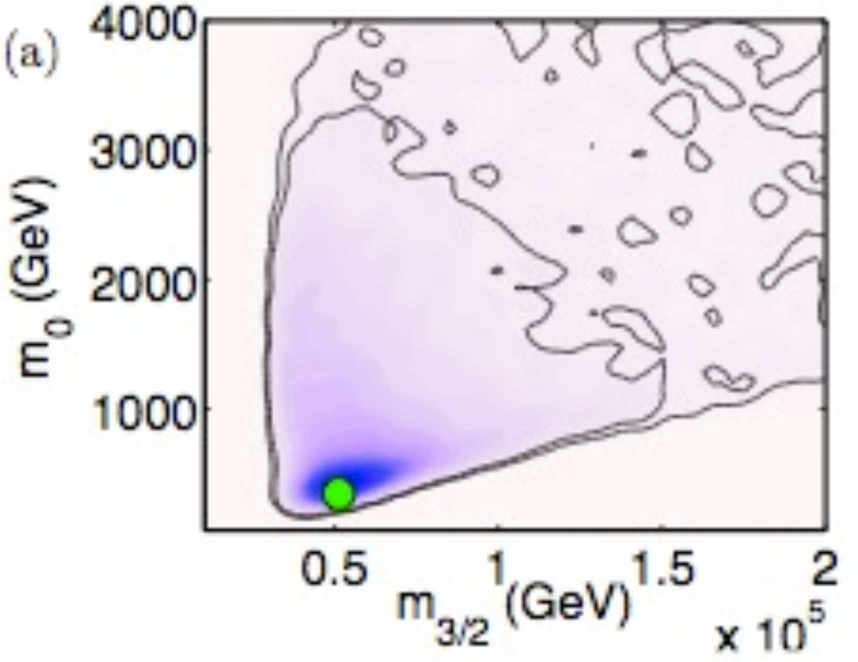
O. Buchmueller, R. Cavanaugh, A. De Roeck, J.R. Ellis,
H.Flacher, S. Heinemeyer, G. Isidori, K.A. Olive,
F.J. Ronga, G. Weiglein



F. Feroz, L. Roszkowski, R. Ruiz de
Austri, R. Trotta

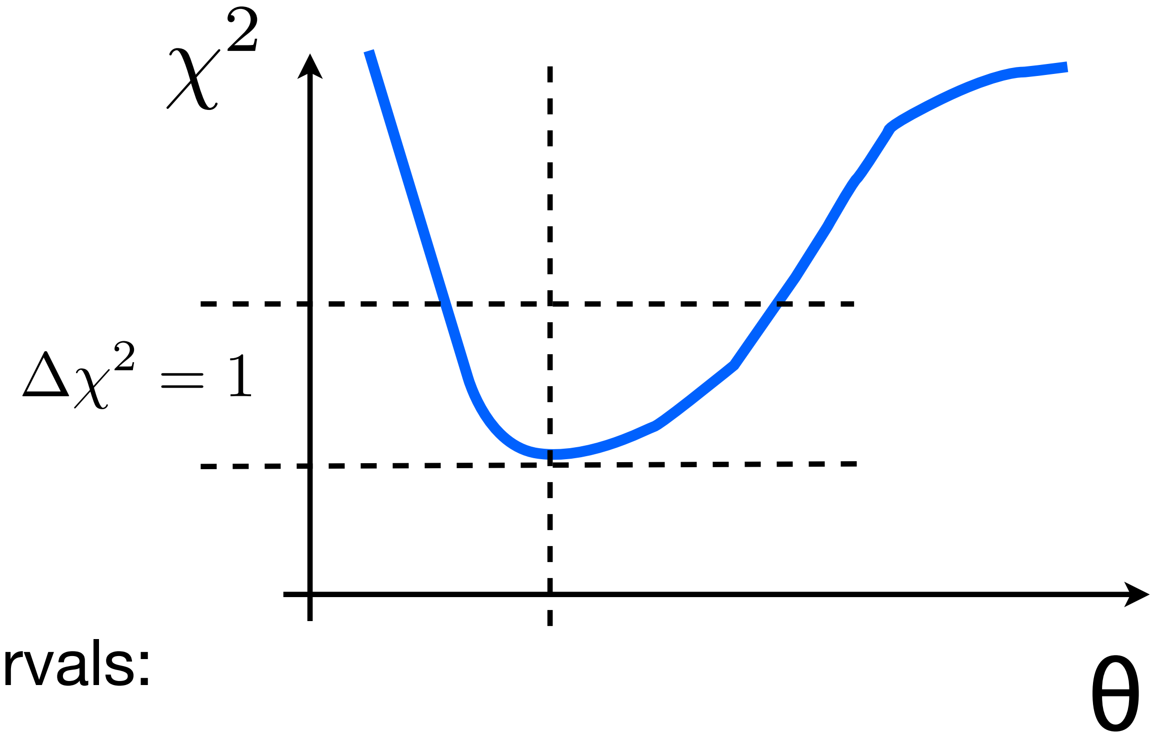


S.S. AbdusSalam, B.C. Allanach, M.J. Dolan,
F. Feroz, M.P. Hobson



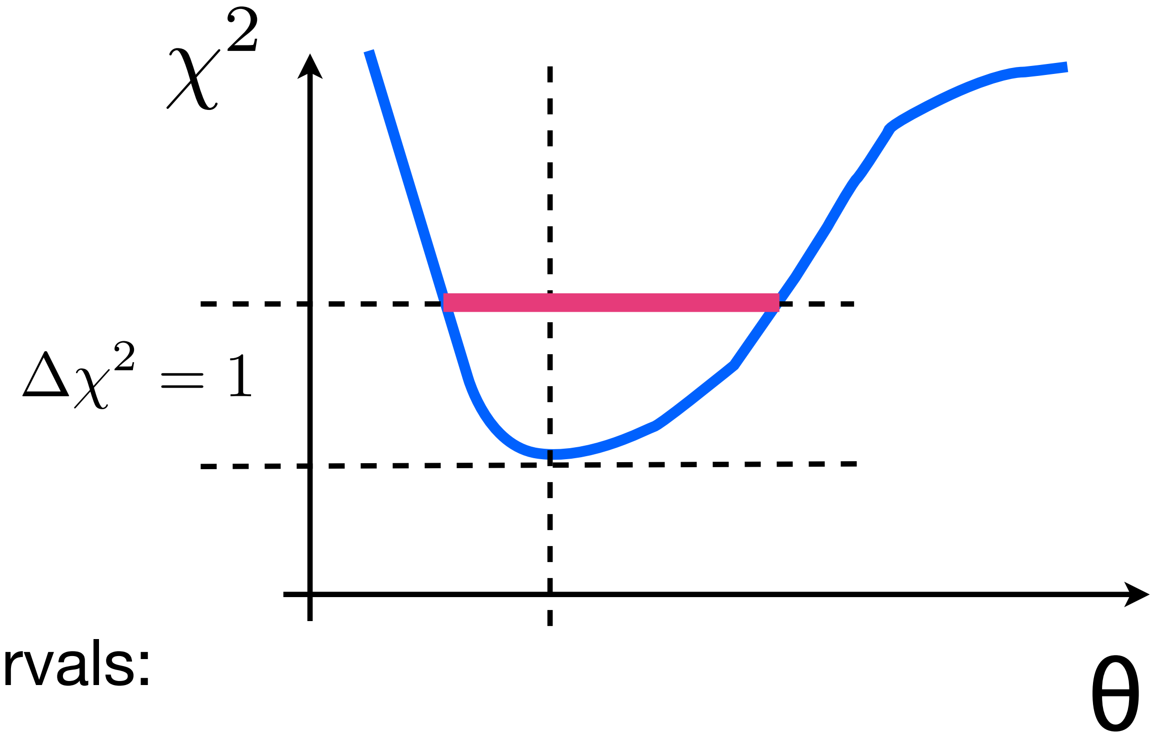
Favoured regions: likelihood-based approach

- Due to the weak nature of constraints, different scanning techniques and statistical methods will generally give different answers
- **Likelihood-based methods:** determine the best fit parameters by finding the minimum of -2Log(Likelihood) = χ^2
  - Markov Chain Monte Carlo and Minuit as "afterburner"
  - Simulated annealing
  - Genetic algorithm
- Determine approximate confidence intervals: Local $\Delta(\chi^2)$ method



Favoured regions: likelihood-based approach

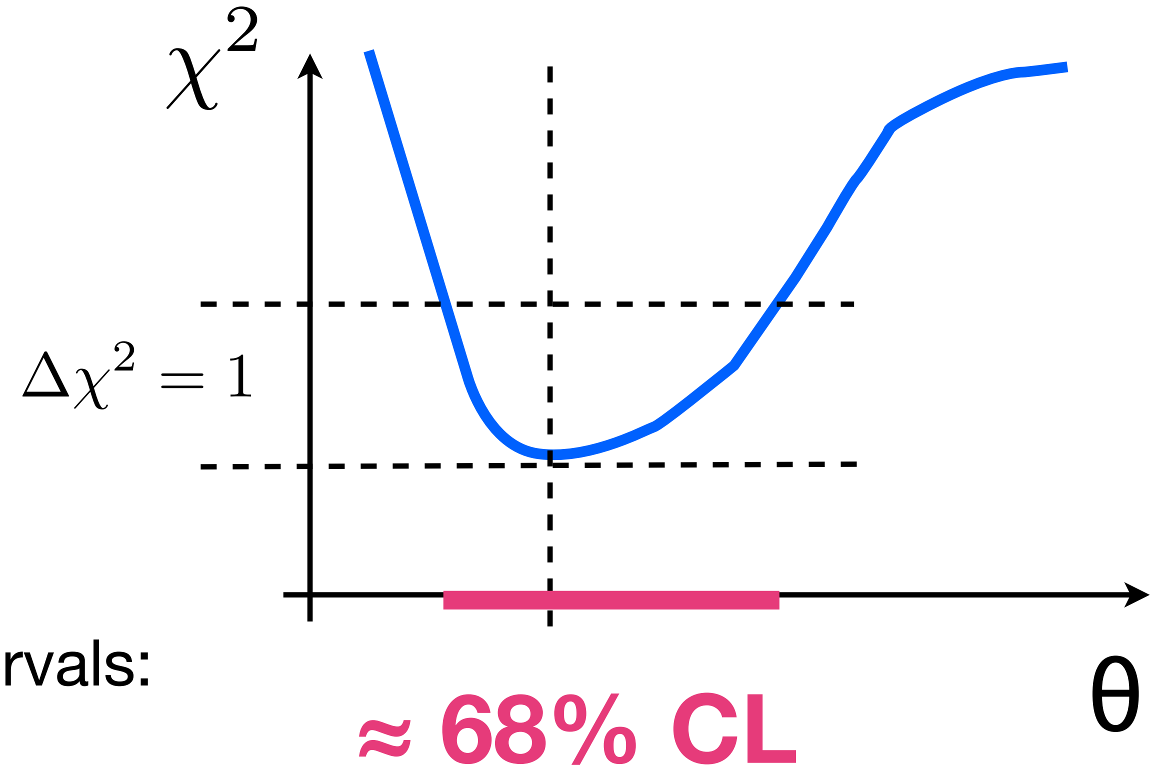
- Due to the weak nature of constraints, different scanning techniques and statistical methods will generally give different answers
- **Likelihood-based methods:** determine the best fit parameters by finding the minimum of -2Log(Likelihood) = χ^2
    - Markov Chain Monte Carlo and Minuit as "afterburner"
    - Simulated annealing
    - Genetic algorithm
- Determine approximate confidence intervals: Local $\Delta(\chi^2)$ method



Favoured regions: likelihood-based approach

- Due to the weak nature of constraints, different scanning techniques and statistical methods will generally give different answers

- **Likelihood-based methods:** determine the best fit parameters by finding the minimum of -2Log(Likelihood) = χ^2
  - Markov Chain Monte Carlo and Minuit as "afterburner"
  - Simulated annealing
  - Genetic algorithm

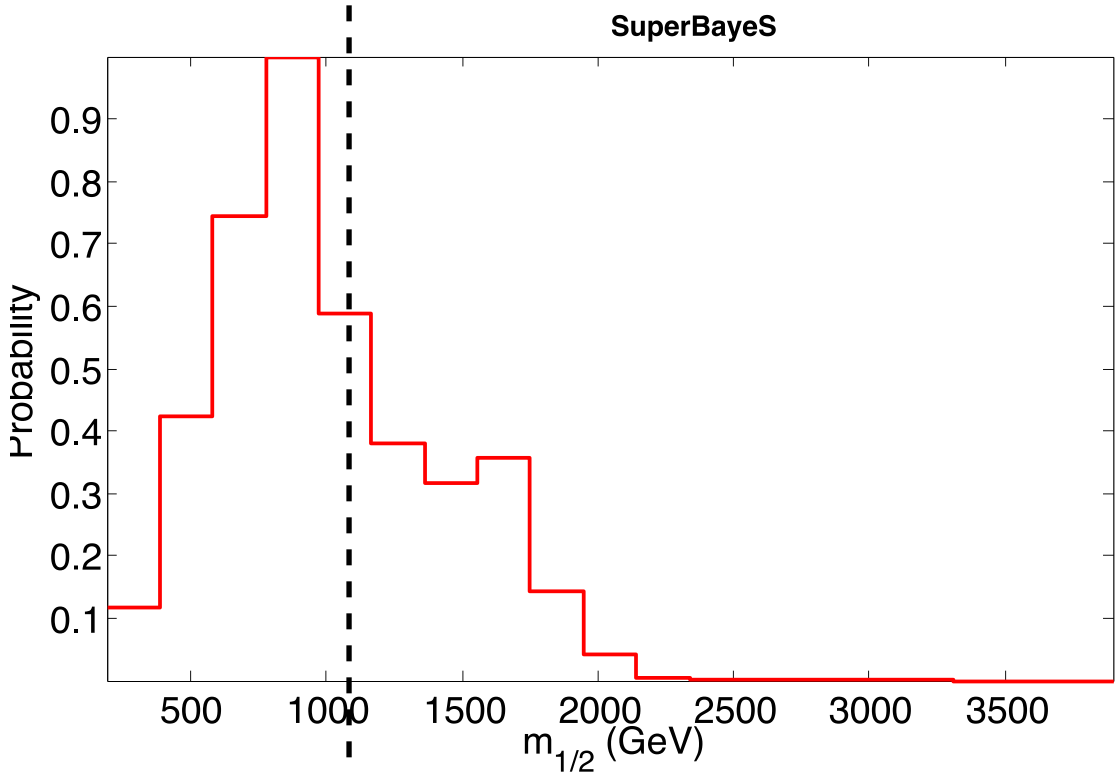


≈ 68% CL

- Determine approximate confidence intervals: Local $\Delta(\chi^2)$ method

Favoured regions: Bayesian approach

- Use the prior to define a metric on parameter space.
- **Bayesian methods:** the best-fit has no special status. Focus on region of large posterior probability mass instead.
  - Markov Chain Monte Carlo (MCMC)
  - Nested sampling
  - Hamiltonian MC
- Determine posterior credible regions: e.g. symmetric interval around the mean containing 68% of samples



Favoured regions: Bayesian approach

- Use the prior to define a metric on parameter space.

- **Bayesian methods:** the best-fit has no special status. Focus on region of large posterior probability mass instead.

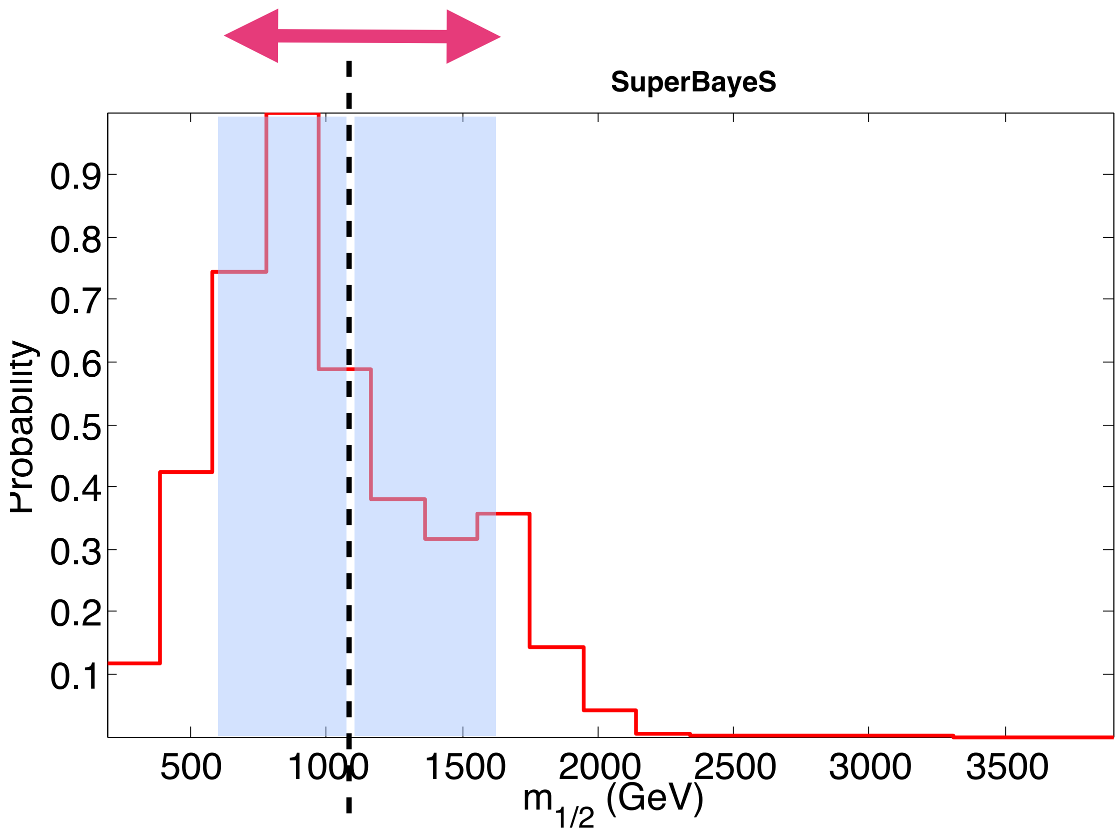
    - Markov Chain Monte Carlo (MCMC)

    - Nested sampling

    - Hamiltonian MC

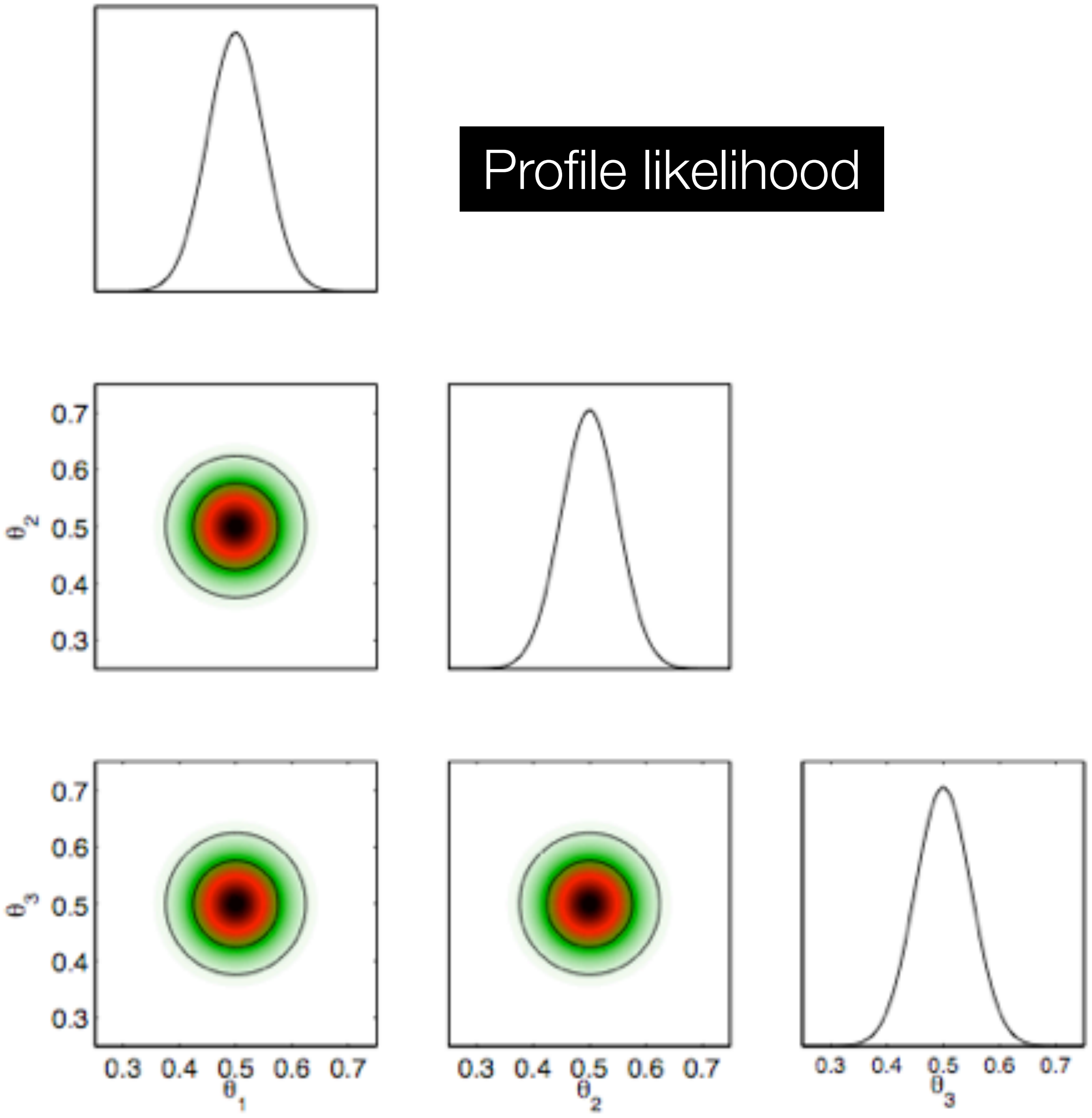
- Determine posterior credible regions: e.g. symmetric interval around the mean containing 68% of samples

68% CREDIBLE REGION

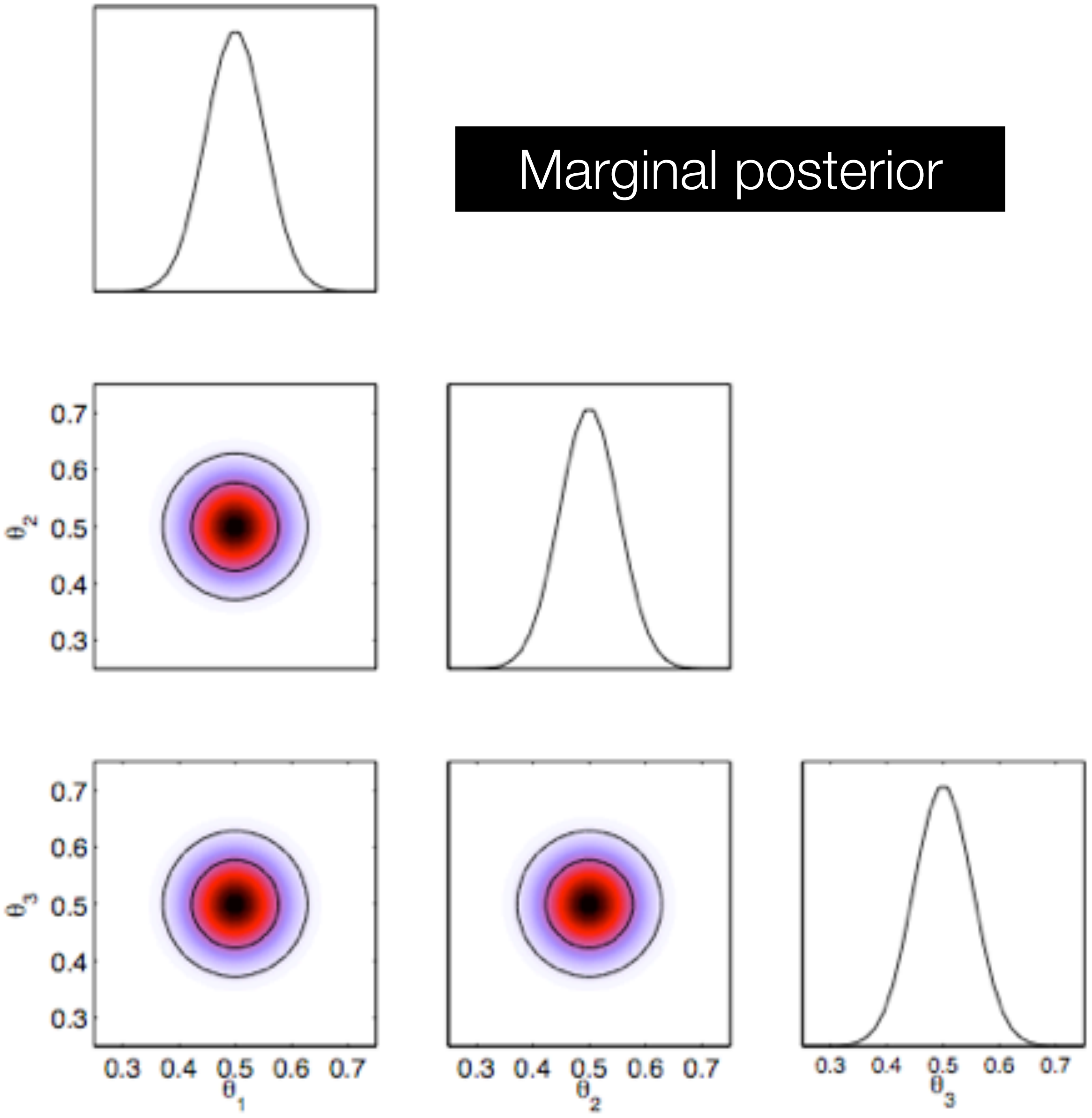


The Gaussian case

- Life is easy (and boring) in Gaussianland:



Profile likelihood



Marginal posterior

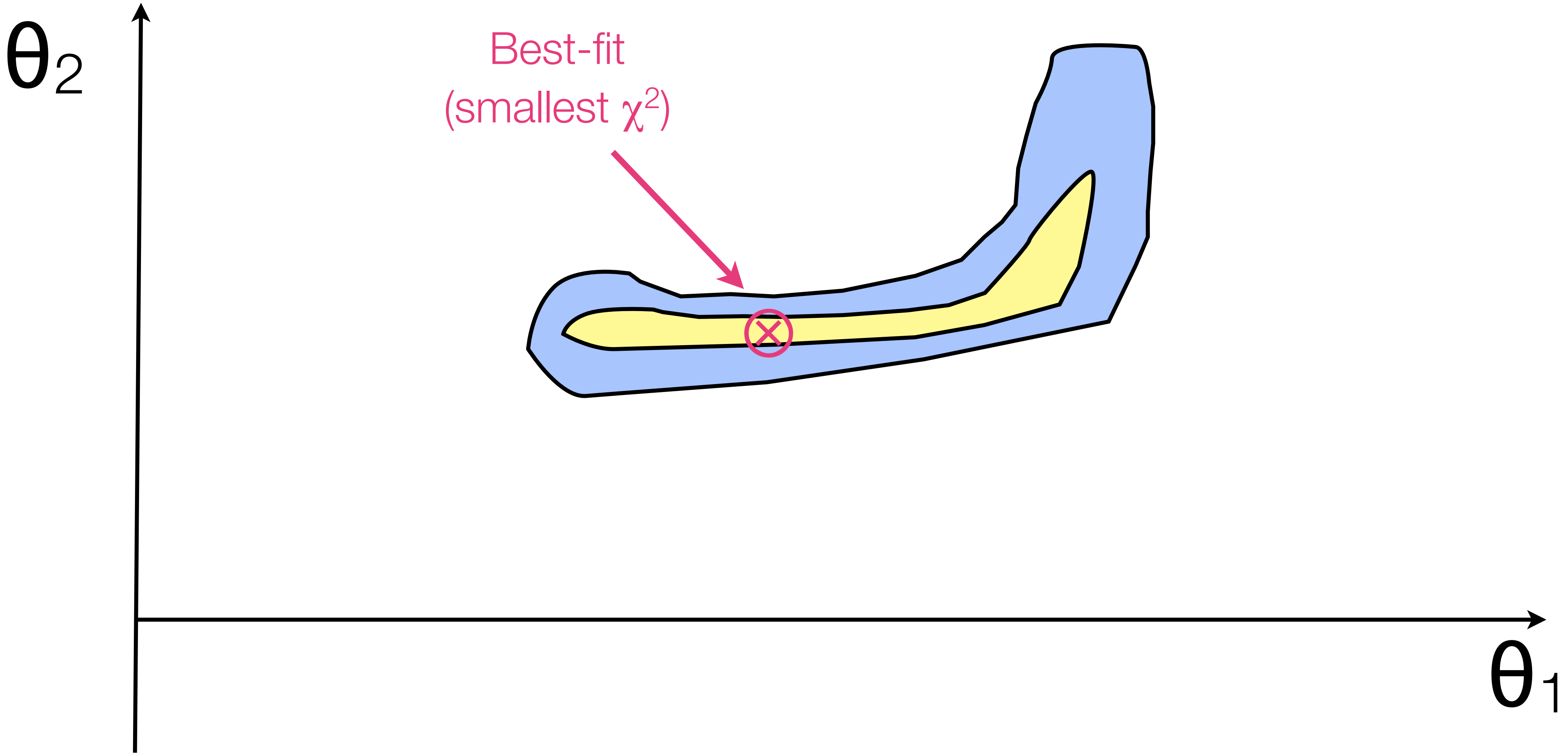
Marginalization vs profiling (maximising)

Marginal posterior:

$$P(\theta_1|D) = \int L(\theta_1,\theta_2)p(\theta_1,\theta_2)d\theta_2$$

Profile likelihood:

$$L(\theta_1) = max_{\theta_2} L(\theta_1,\theta_2)$$



(2D plot depicts likelihood contours - prior assumed flat over wide range)

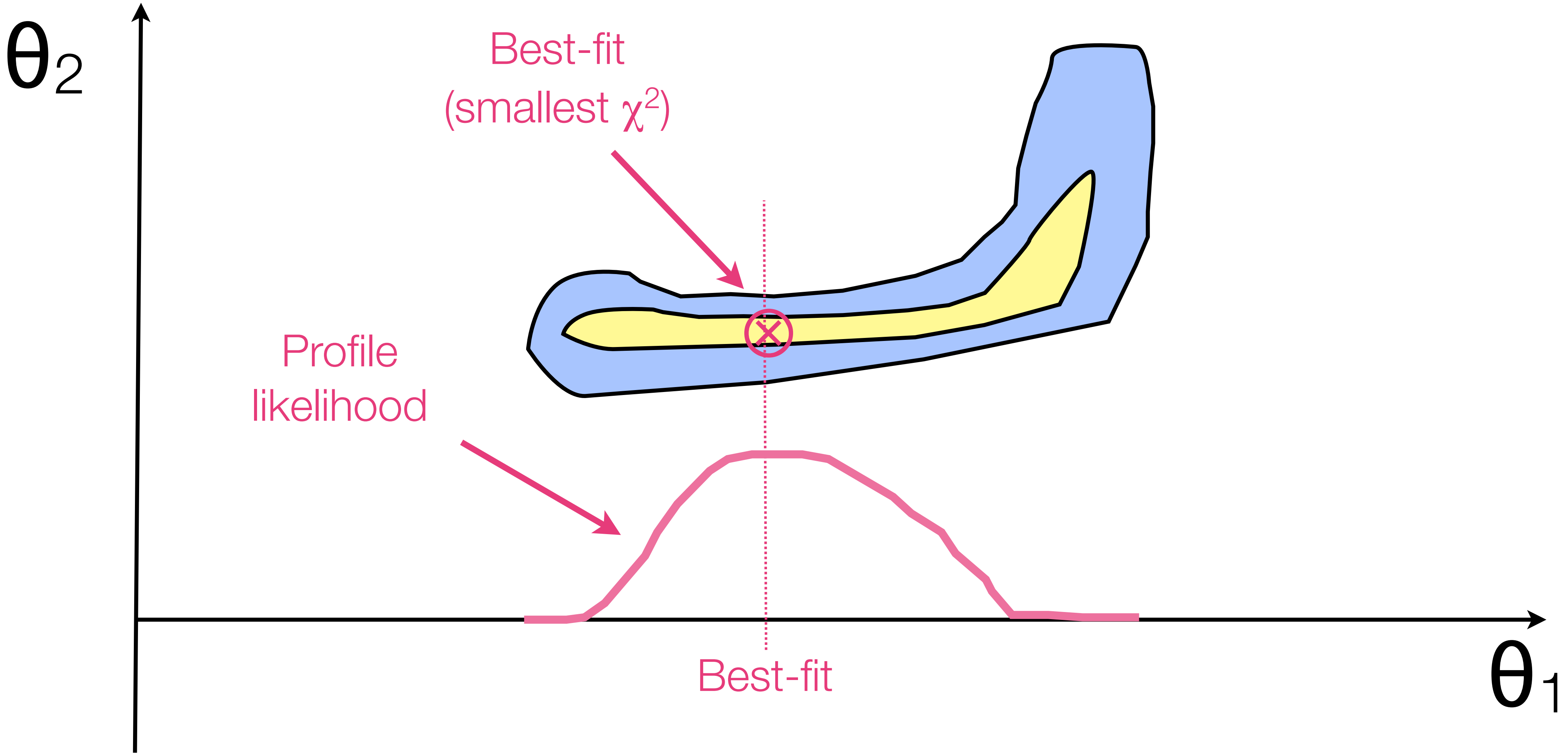
Marginalization vs profiling (maximising)

Marginal posterior:

$$P(\theta_1|D) = \int L(\theta_1, \theta_2) p(\theta_1, \theta_2) d\theta_2$$

Profile likelihood:

$$L(\theta_1) = max_{\theta_2} L(\theta_1, \theta_2)$$



(2D plot depicts likelihood contours - prior assumed flat over wide range)

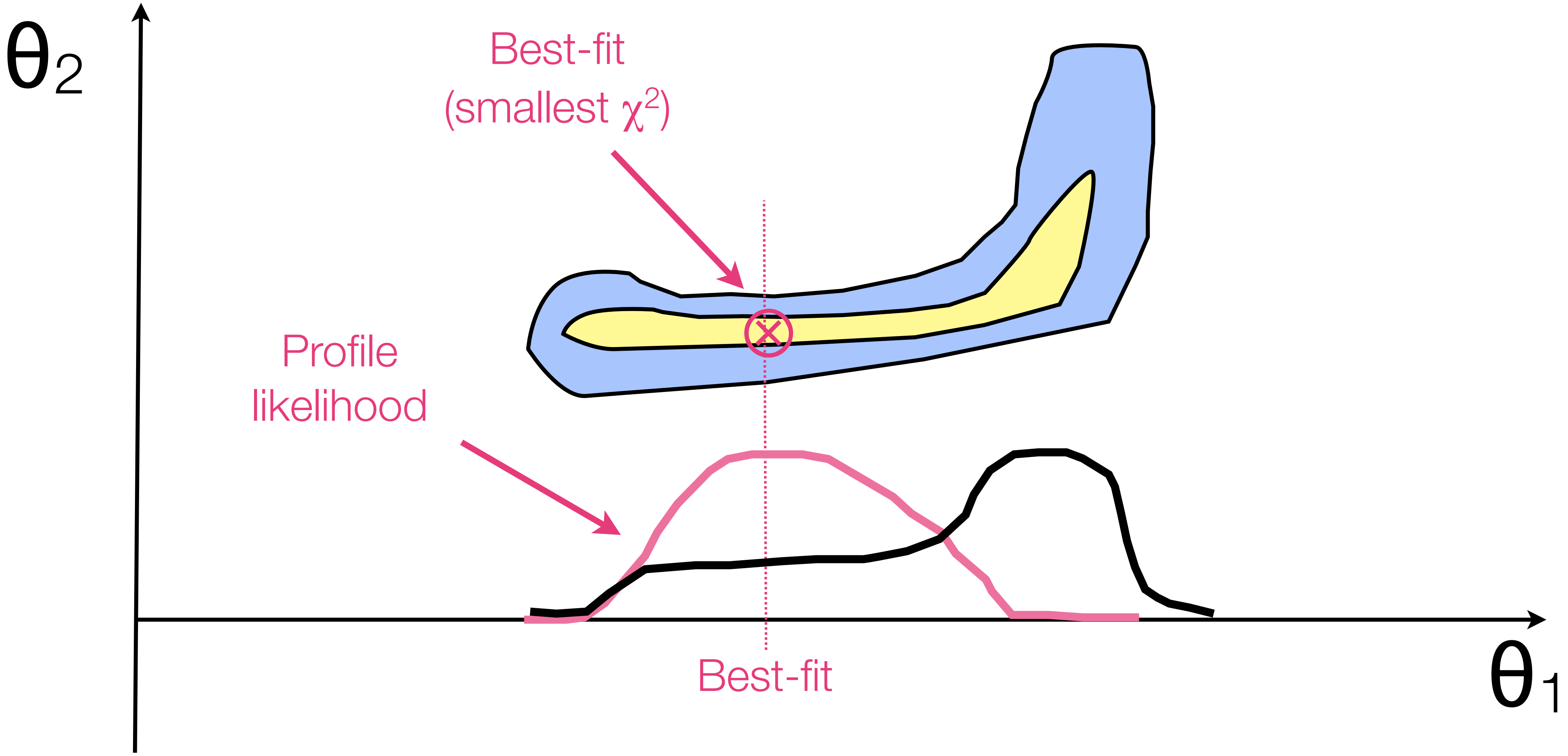
Marginalization vs profiling (maximising)

Marginal posterior:

$$P(\theta_1|D) = \int L(\theta_1, \theta_2)p(\theta_1, \theta_2)d\theta_2$$

Profile likelihood:

$$L(\theta_1) = max_{\theta_2} L(\theta_1, \theta_2)$$



(2D plot depicts likelihood contours - prior assumed flat over wide range)

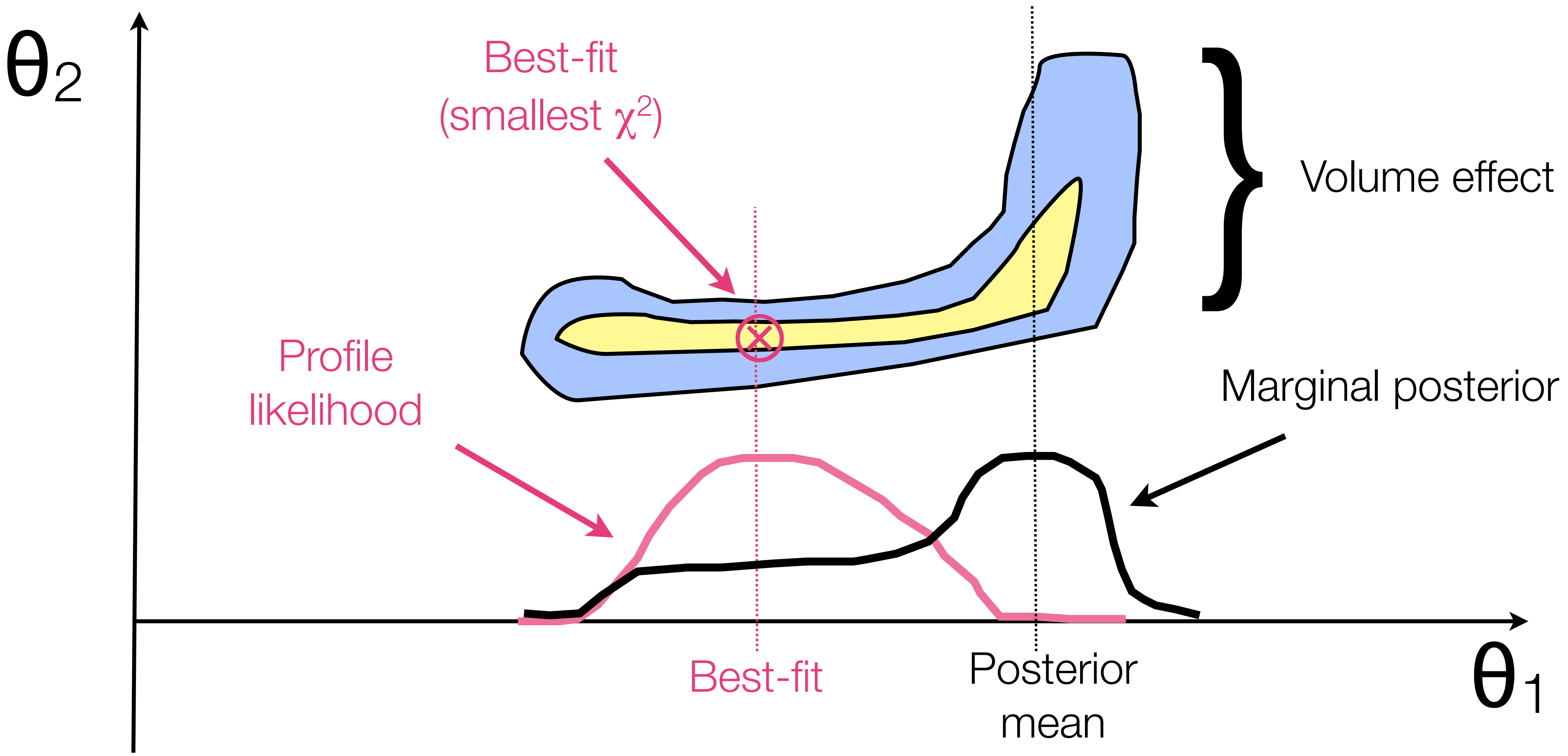
Marginalization vs profiling (maximising)

Marginal posterior:

$$P(\theta_1|D) = \int L(\theta_1, \theta_2) p(\theta_1, \theta_2) d\theta_2$$

Profile likelihood:

$$L(\theta_1) = max_{\theta_2} L(\theta_1, \theta_2)$$



(2D plot depicts likelihood contours - prior assumed flat over wide range)

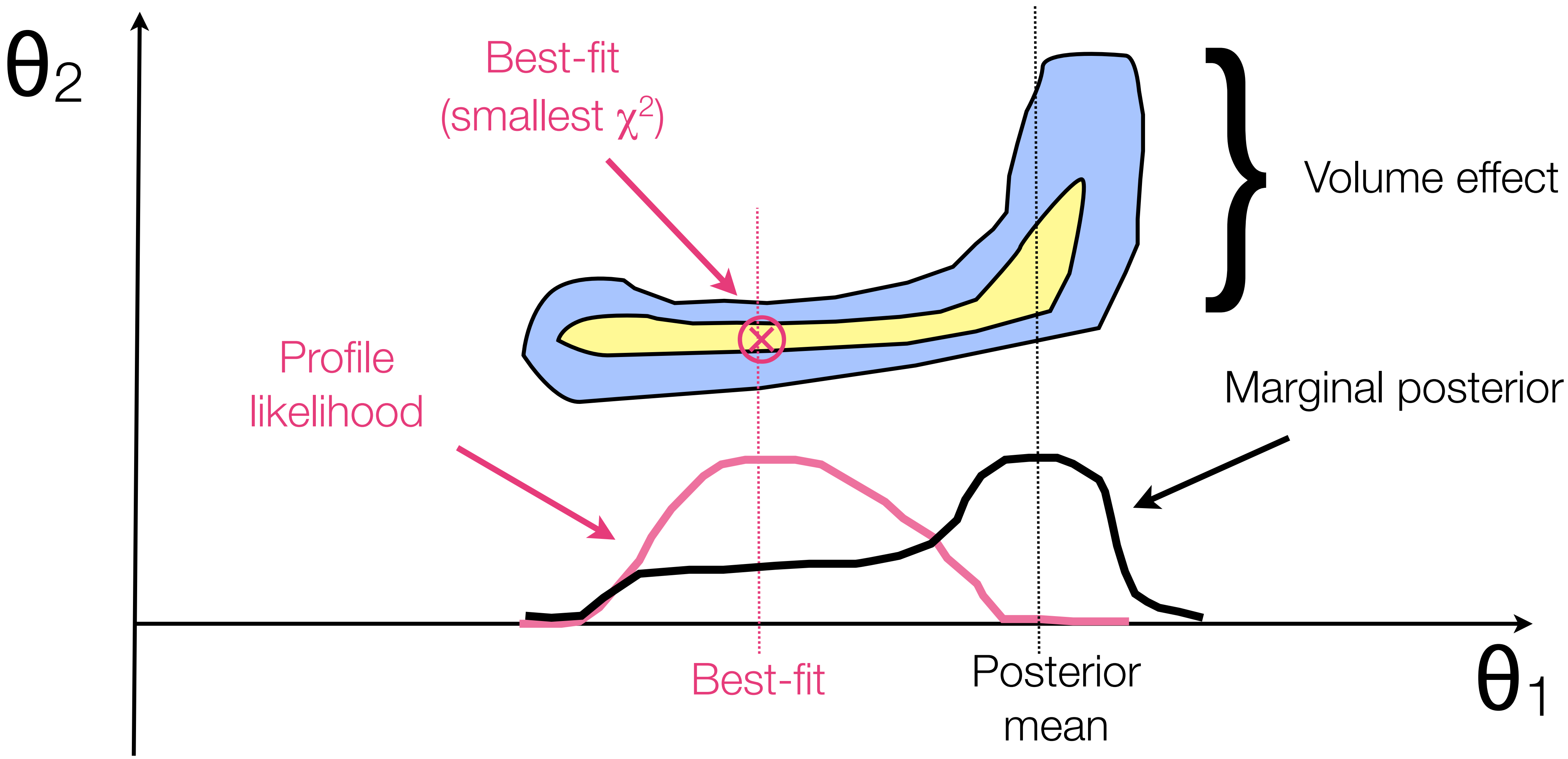
Marginalization vs profiling (maximising)

Physical analogy: (thanks to Tom Loredo)

Likelihood = hottest hypothesis
Posterior = hypothesis with most heat

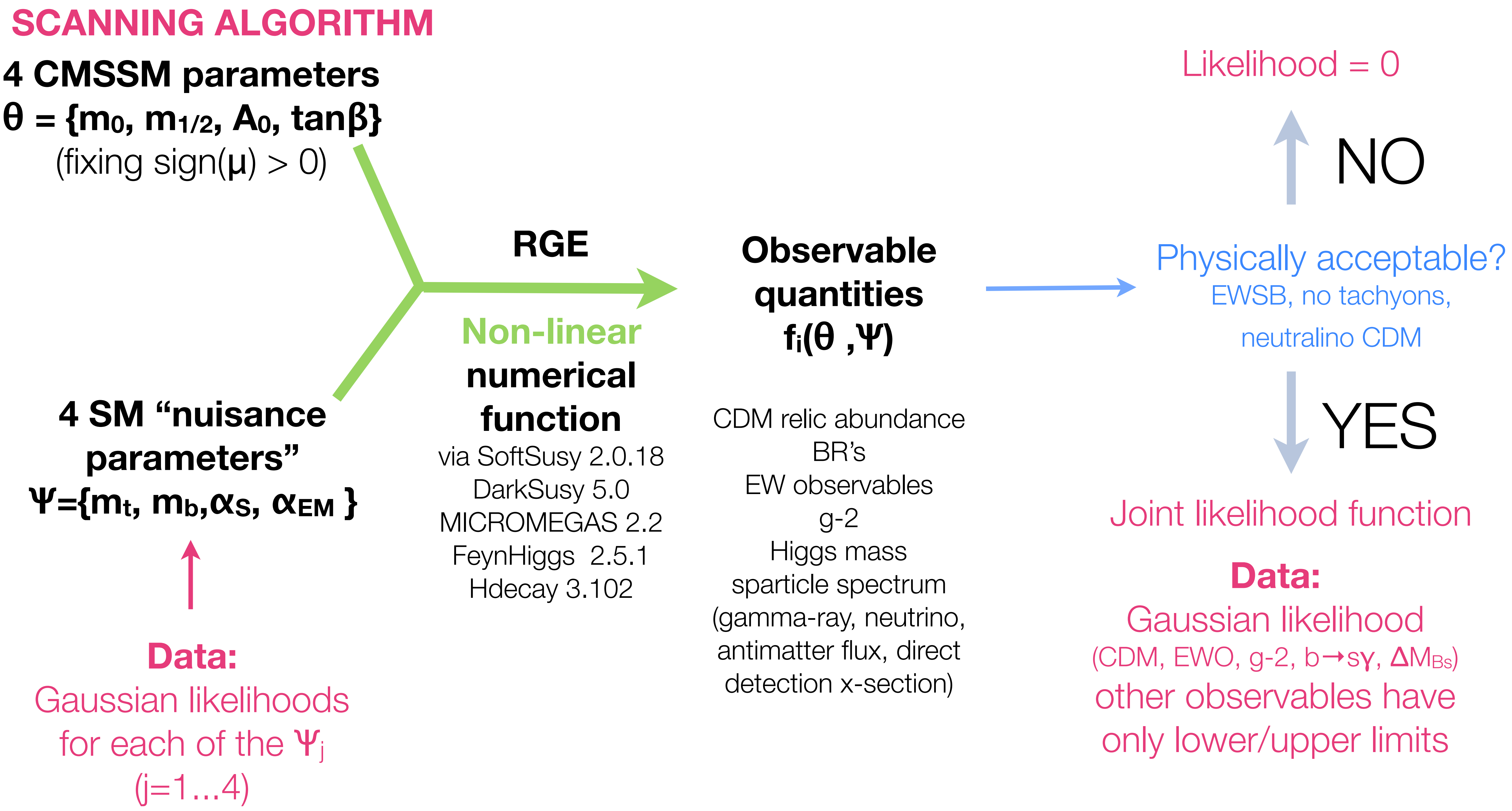
Heat: $Q = \int c_V(x) T(x) dV$

Posterior: $P \propto \int p(\theta) L(\theta) d\theta$



(2D plot depicts likelihood contours - prior assumed flat over wide range)

Constrained MSSM analysis pipeline



SCANNING ALGORITHM

4 CMSSM parameters

$\theta = \{m_0, m_{1/2}, A_0, \tan\beta\}$

(fixing sign(μ) > 0)

4 SM “nuisance parameters”

$\Psi=\{m_t, m_b, \alpha_S, \alpha_{EM}\}$

Data:

Gaussian likelihoods for each of the Ψ_j (j=1...4)

RGE

Non-linear numerical function

via SoftSusy 2.0.18
DarkSusy 5.0
MICROMEGAS 2.2
FeynHiggs 2.5.1
Hdecay 3.102

Observable quantities

$f_i(\theta, \Psi)$

CDM relic abundance
BR’s
EW observables
g-2
Higgs mass
sparticle spectrum
(gamma-ray, neutrino, antimatter flux, direct detection x-section)

Physically acceptable?

EWSB, no tachyons, neutralino CDM

NO

Likelihood = 0

YES

Joint likelihood function

Data:

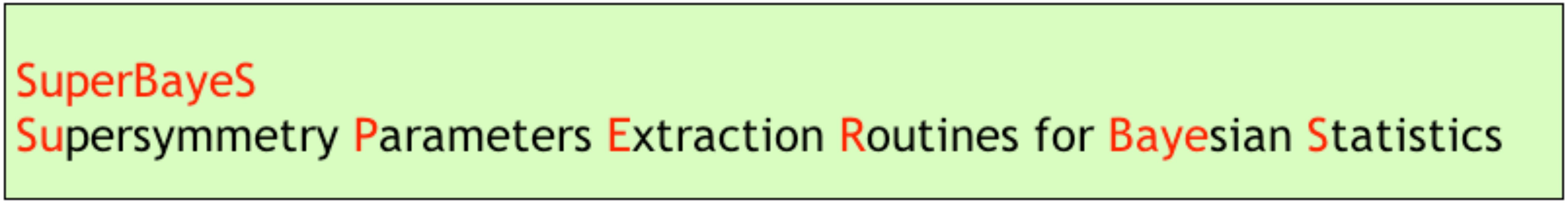
Gaussian likelihood

(CDM, EWO, g-2, b→sγ, ΔM_{B_s})

other observables have only lower/upper limits

Kyle Cranmer & Roberto Trotta

www.superbayes.org

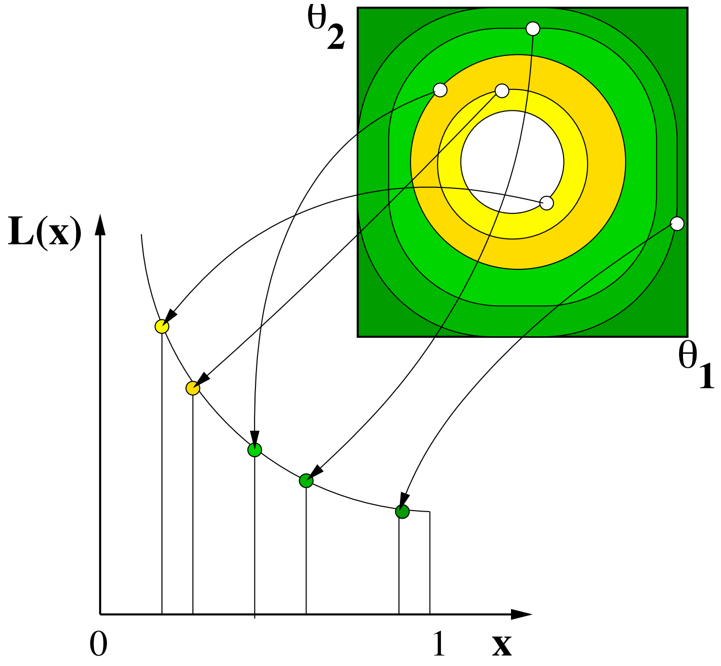


SuperBayeS

Supersymmetry Parameters Extraction Routines for Bayesian Statistics

- Implements the CMSSM, but can be easily extended to the general MSSM
- **New release (v 1.50) in June 2010:** linked to SoftSusy 2.0.18, DarkSusy 5.0, MICROMEGAS 2.2, FeynHiggs 2.5.1, Hdecay 3.102.
- Includes up-to-date constraints from all observables, plotting routines, statistical analysis tools, posterior and profile likelihood plots. Fully parallelized, MPI-ready, user-friendly interface
- MCMC engine (Metropolis-Hastings, bank sampler), grid scan mode, multi-modal nested sampling MultiNest algorithm (Feroz & Hobson 2008)
  A full 8D scan now takes less than 2 days on 8 CPUs.

The nested sampling algorithm



(animation courtesy of David Parkinson)

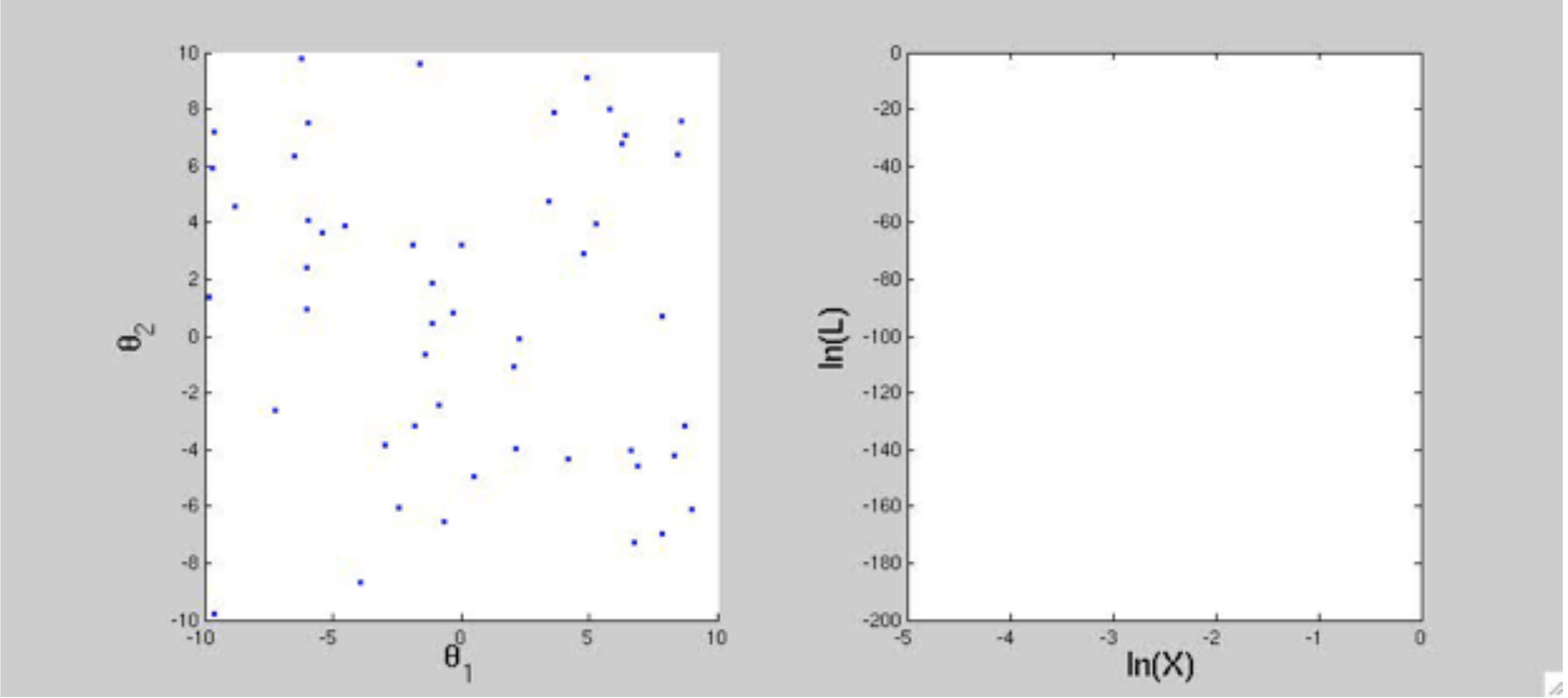
An algorithm originally aimed primarily at the Bayesian evidence computation (Skilling, 2006):

$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta)d\theta$$

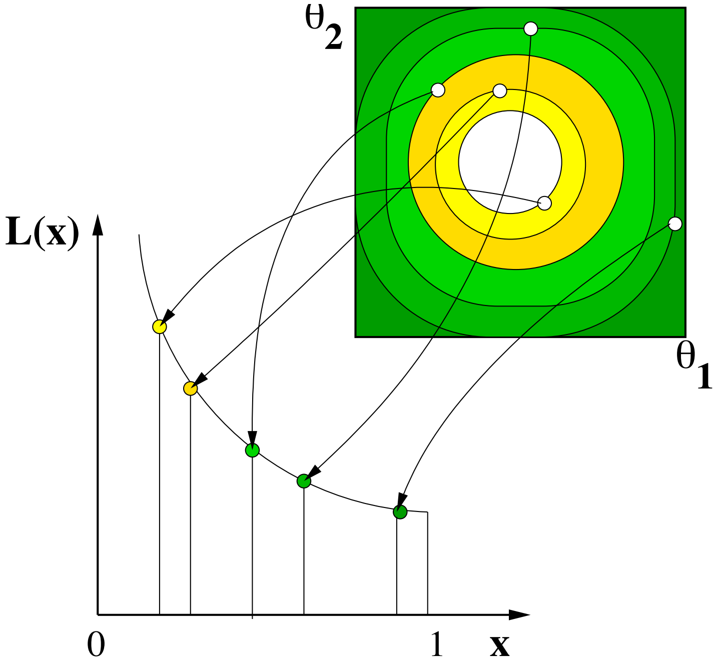
$$P(d) = \int d\theta \mathcal{L}(\theta) P(\theta) = \int_0^1 X(\lambda) d\lambda$$

Feroz et al (2008), *arxiv: 0807.4512*, Trotta et al (2008), *arxiv: 0809.3792*

The nested sampling algorithm



(animation courtesy of David Parkinson)



An algorithm originally aimed primarily at the Bayesian evidence computation (Skilling, 2006):

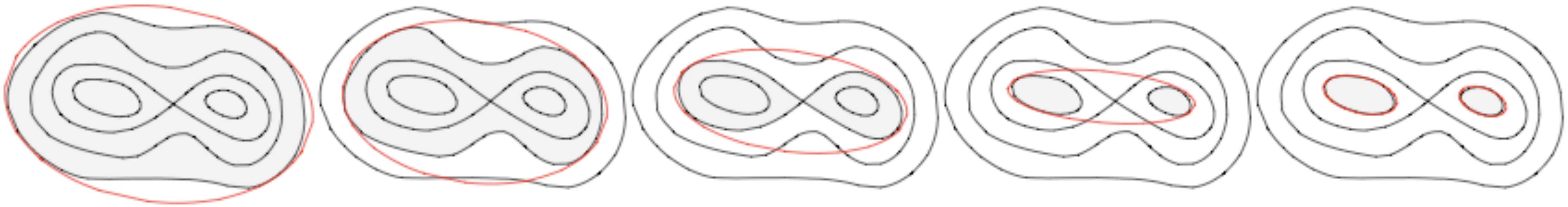
$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta) d\theta$$

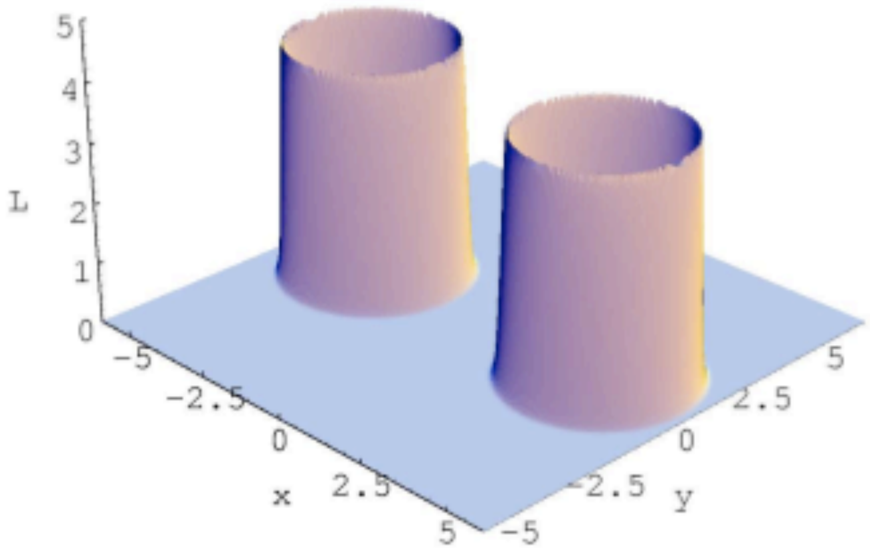
$$P(d) = \int d\theta \mathcal{L}(\theta) P(\theta) = \int_0^1 X(\lambda) d\lambda$$

Feroz et al (2008), *arxiv: 0807.4512*, Trotta et al (2008), *arxiv: 0809.3792*

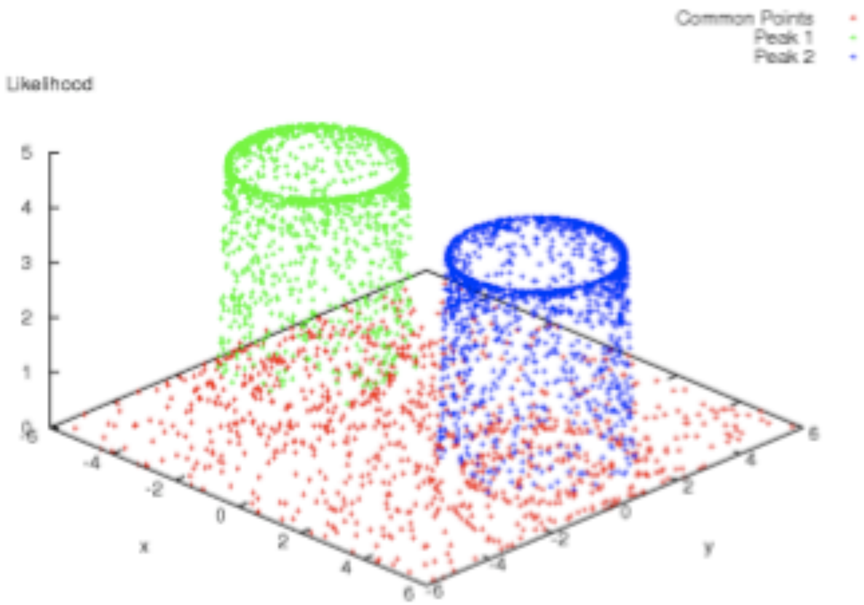
The MultiNest algorithm

- MultiNest: a multi-modal implementation of nested sampling. Also an extremely efficient sampler for multi-modal likelihoods
  Feroz & Hobson (2007), RT et al (2008), Feroz et al (2008)





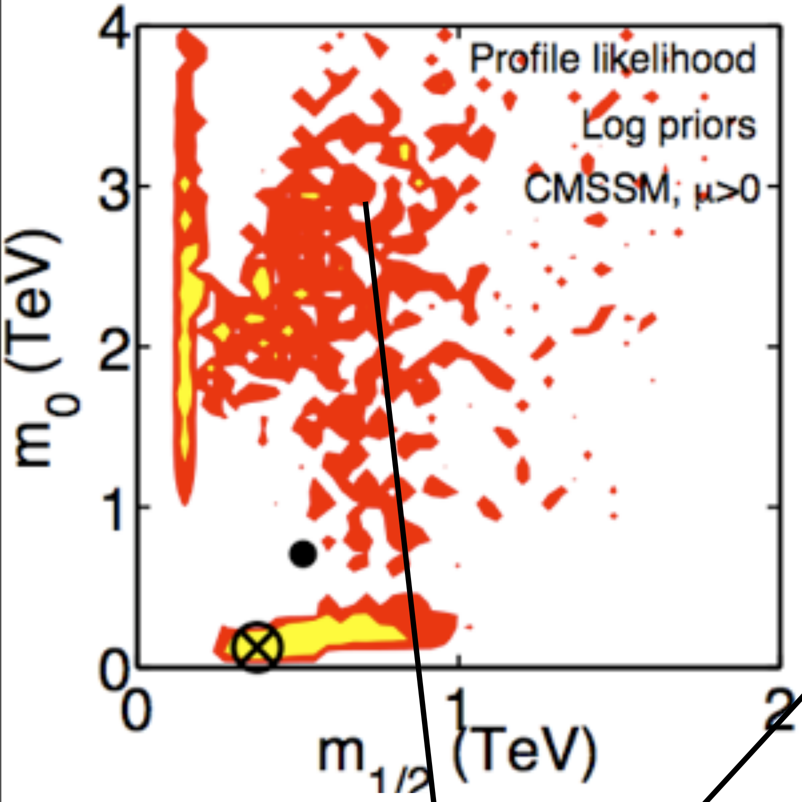
Target Likelihood



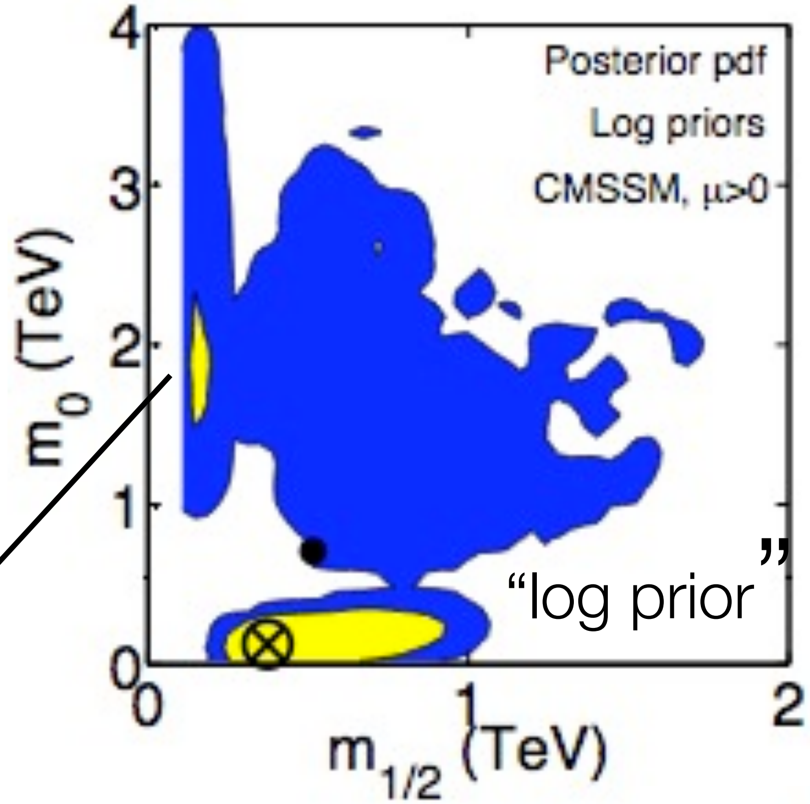
Sampled Likelihood

CMSSM today: Frequentist vs Bayesian

SuperBayeS: profile likelihood



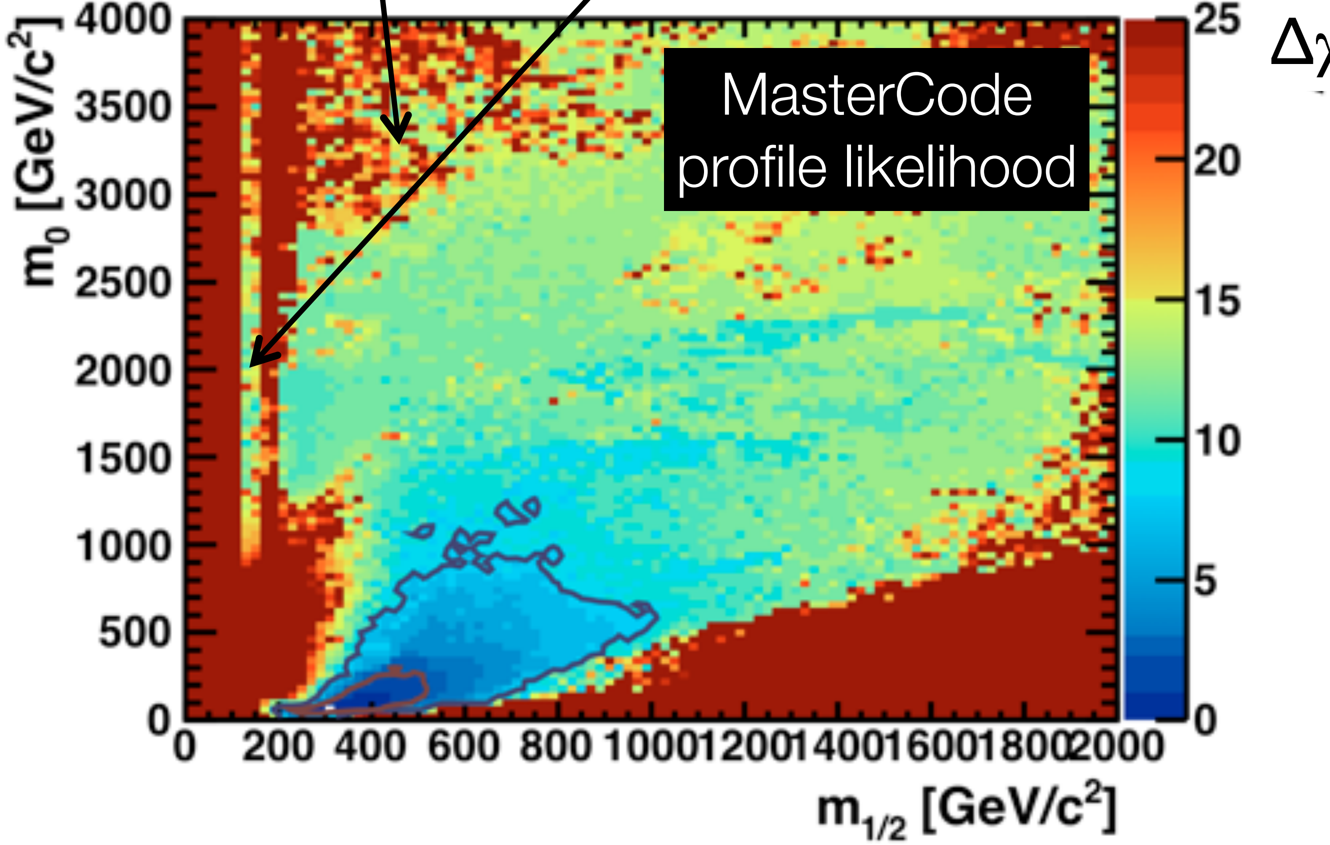
SuperBayeS: posterior



“log prior”

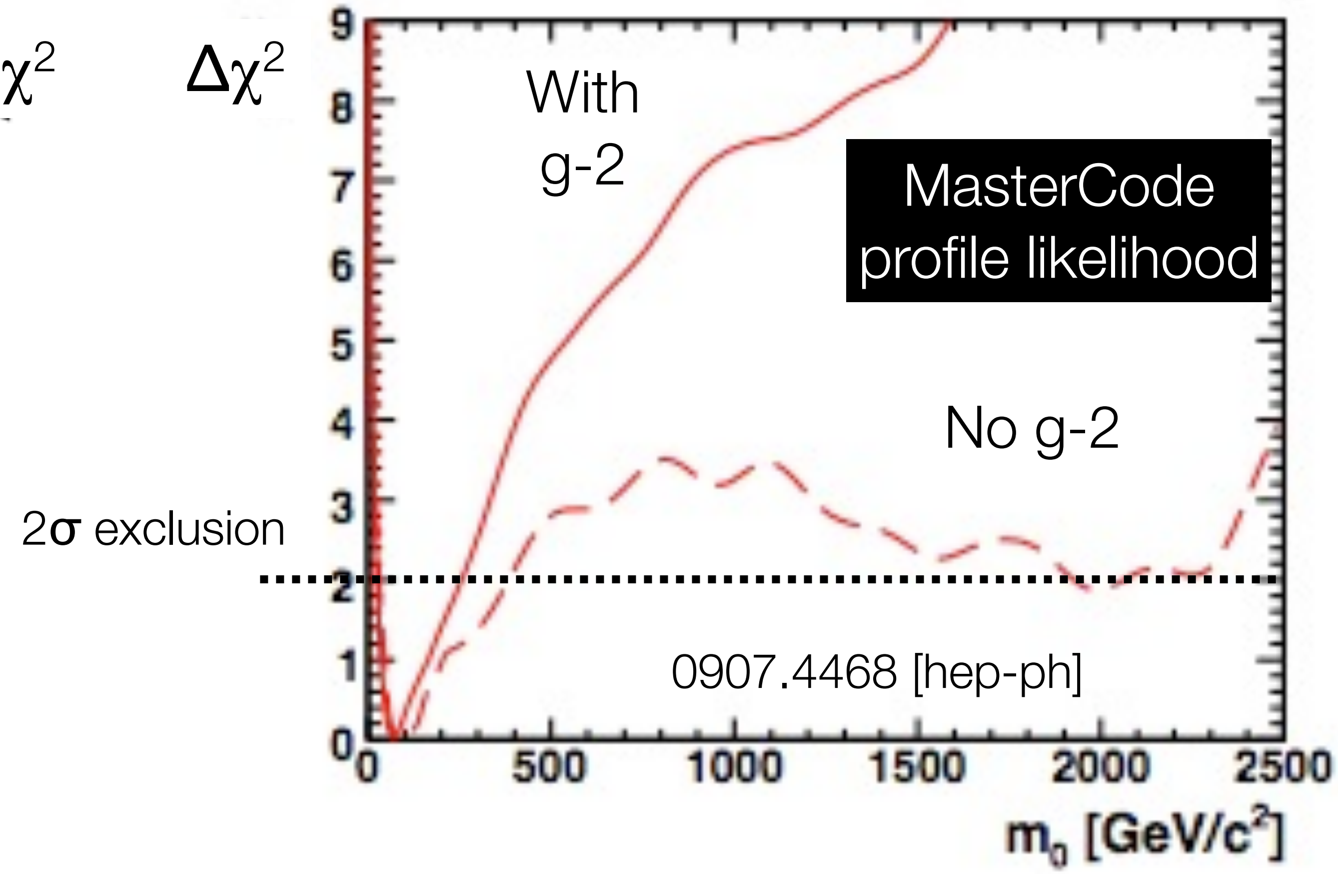
Both methods find a favoured low mass SUSY region: how constrained is it?

The g-2 constraint is critical in robustly excluding TeV-scale masses in the frequentist approach



$\Delta\chi^2$

MasterCode profile likelihood



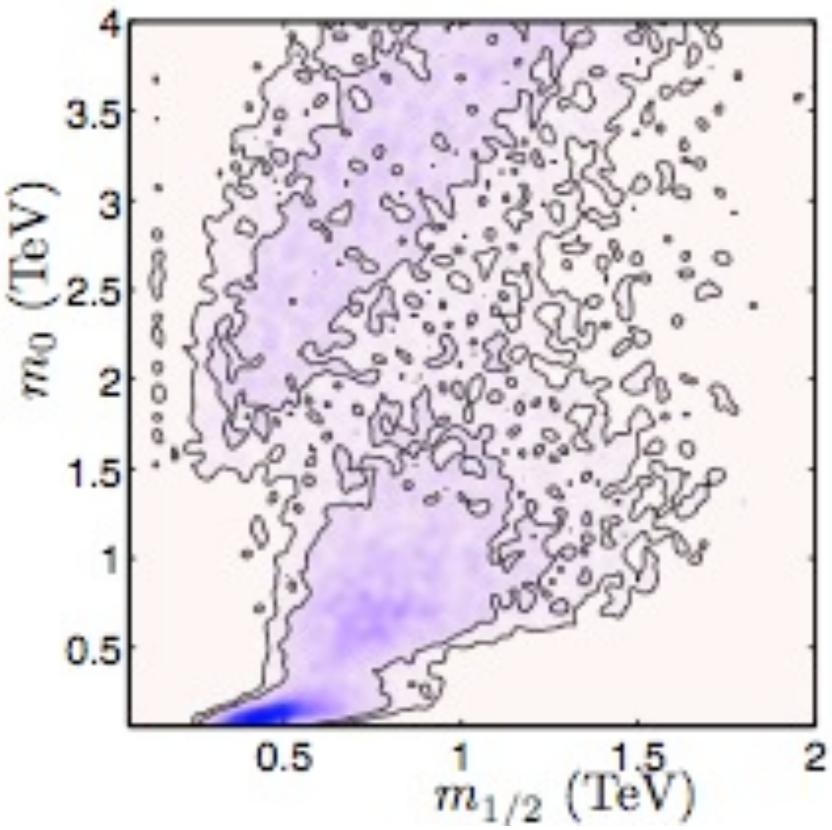
MasterCode profile likelihood

Bayesian results: prior dependency with present-day constraints

“flat prior”

Uniform in $M_0, M_{1/2}, A_0, \tan\beta$

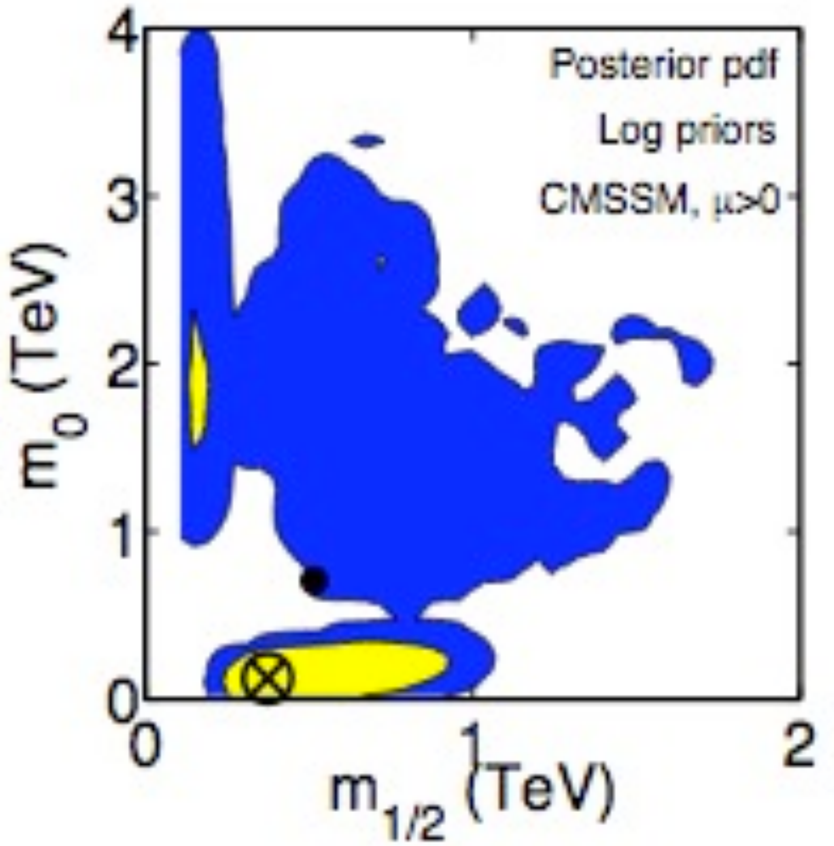
0807.4512 [hep-ph]



“log prior”

Uniform in $\log(M_0)$, $\log(M_{1/2})$, A_0, $\tan\beta$

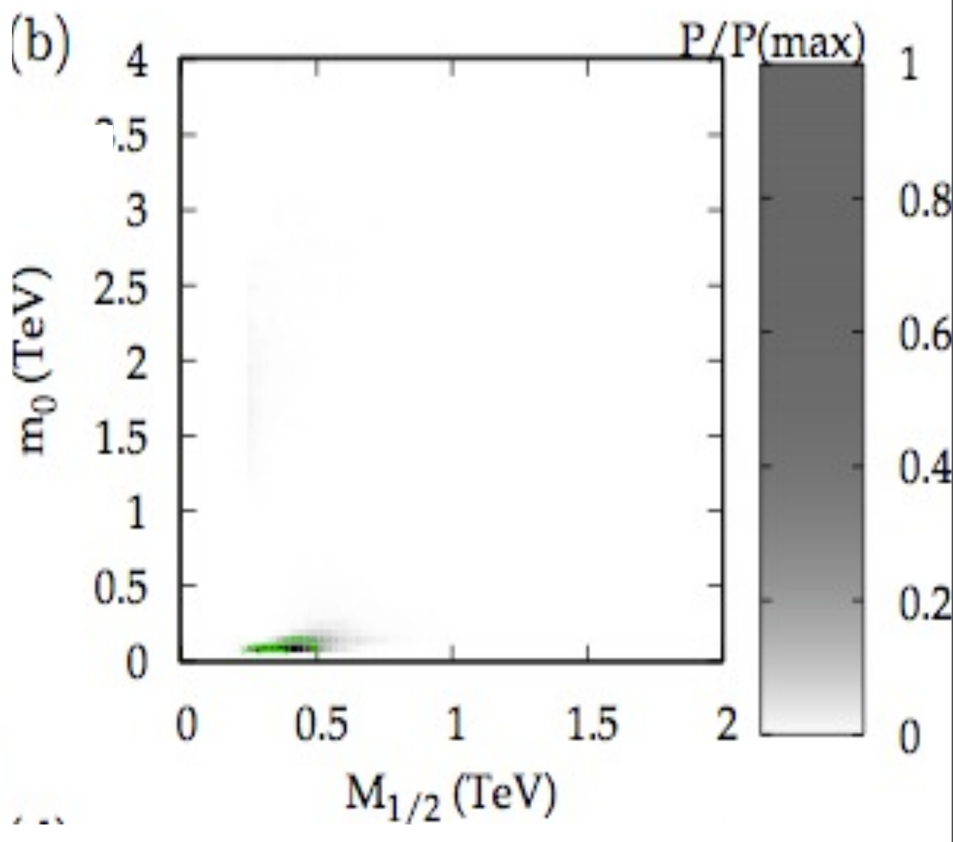
0809.3792 [hep-ph]



“naturalness prior”

Penalizes regions of parameter space that are “fine tuned”

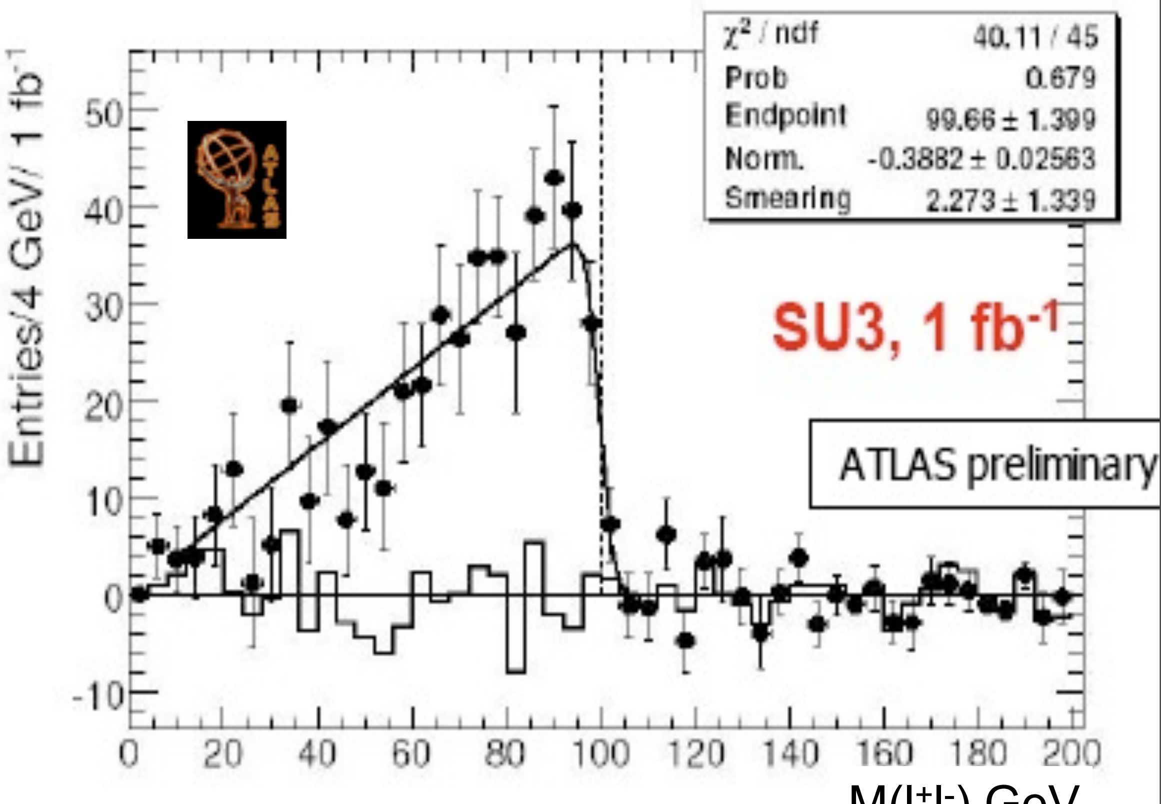
0705.0487 [hep-ph]



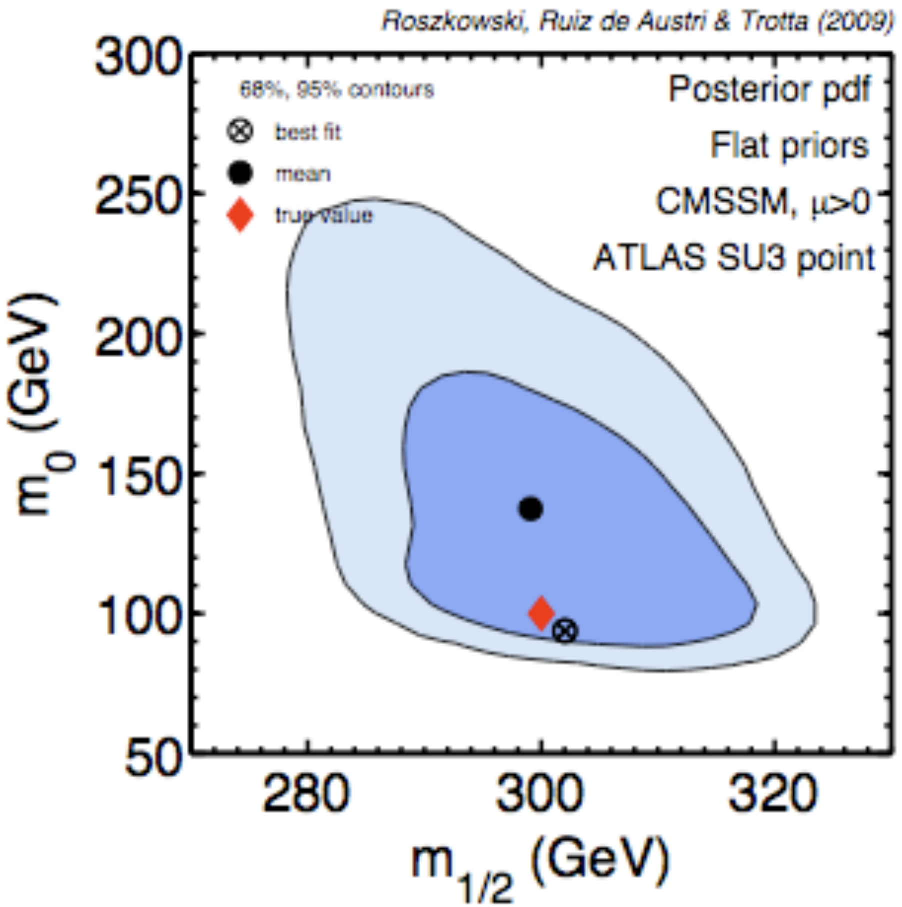
Prior dependence of the Bayesian fits results from weak constraints on parameter space. Stronger assumptions (e.g. naturalness priors) lead to posteriors dominated by prior information (rather than data).

ATLAS data will mitigate the prior dependency for the CMSSM

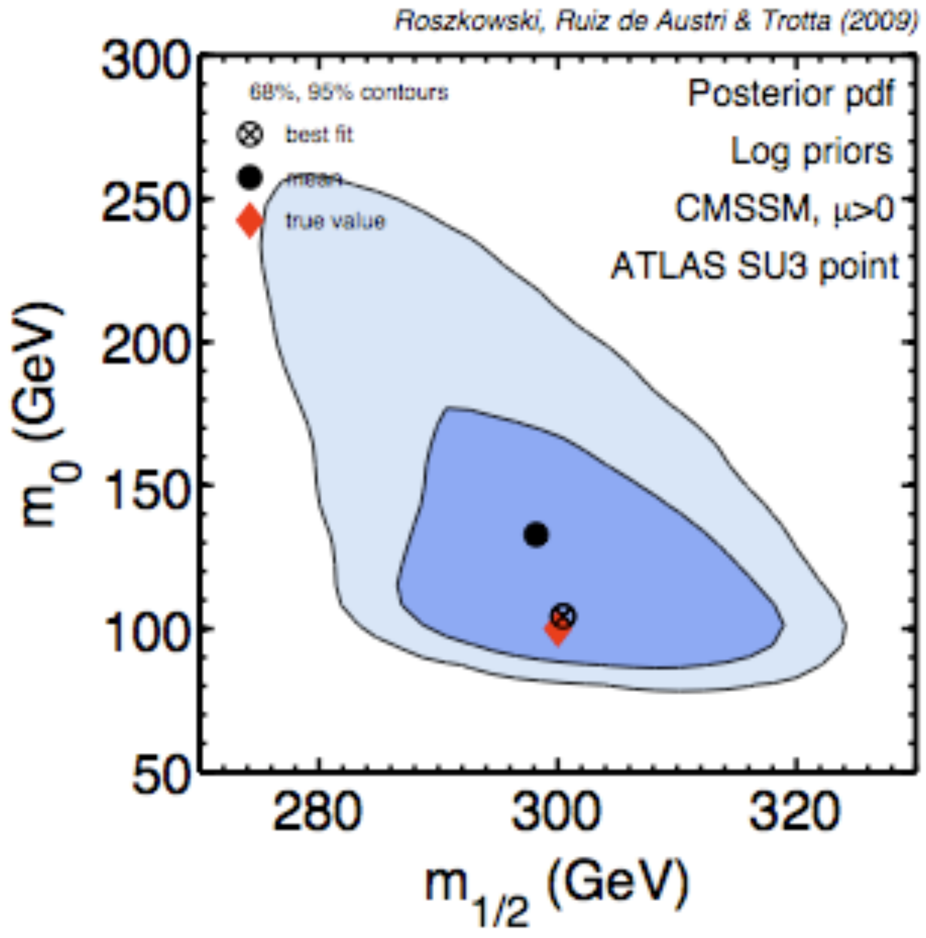
- Projected constraints from ATLAS, (dilepton and lepton+jets edges, 1 fb^{-1} luminosity)



SuperBayeS
Flat prior



SuperBayeS
Log prior

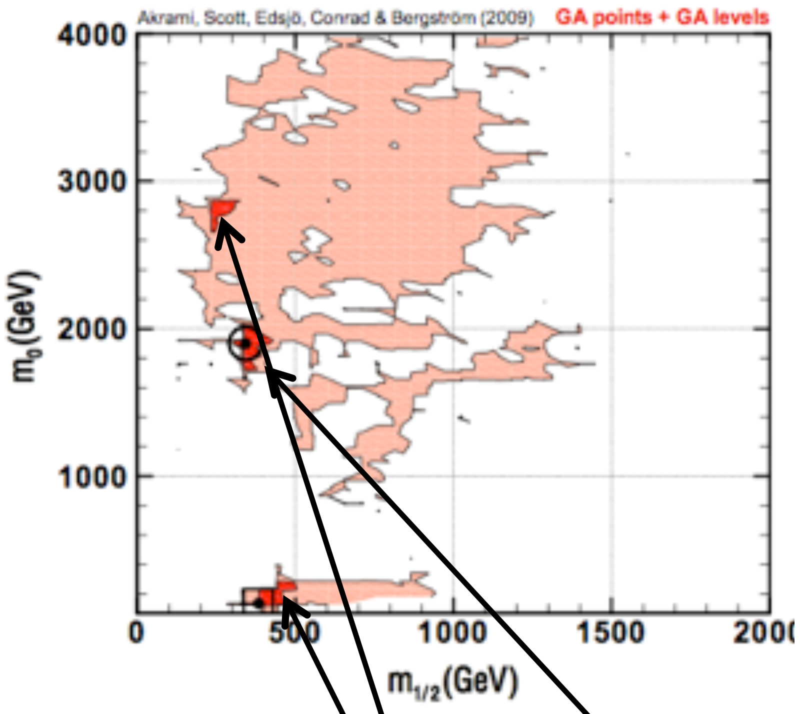


Roszkowski, Ruiz & RT (2010, 0907.0594)

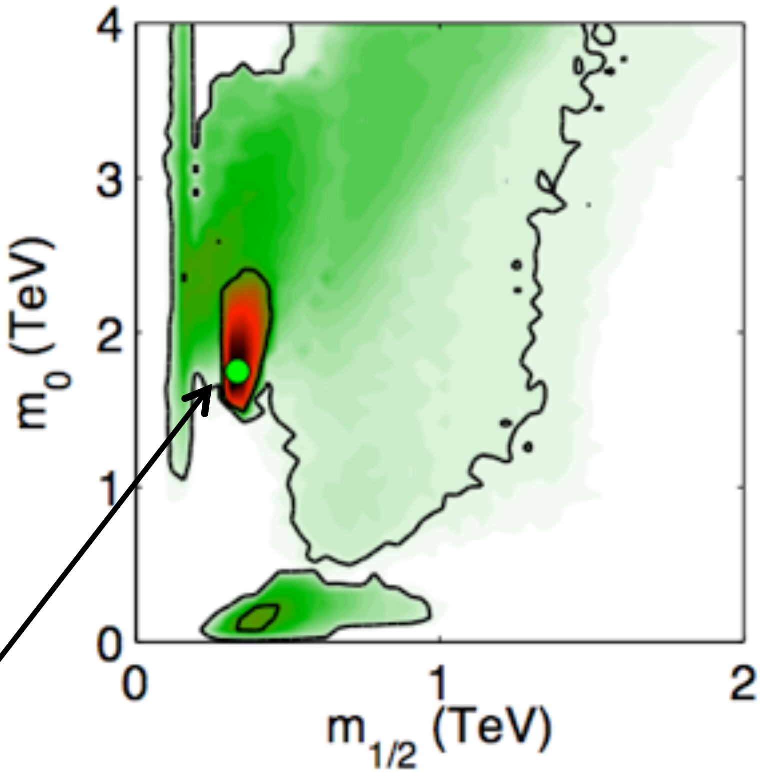
Profile likelihood results: comparison

- Akrami et al (0910.3950) adopted a genetic algorithm (GA) to map out the profile likelihood.
- This allows to find isolated spikes in the likelihood in the focus point region: **is this something other frequentist fits might have missed?**

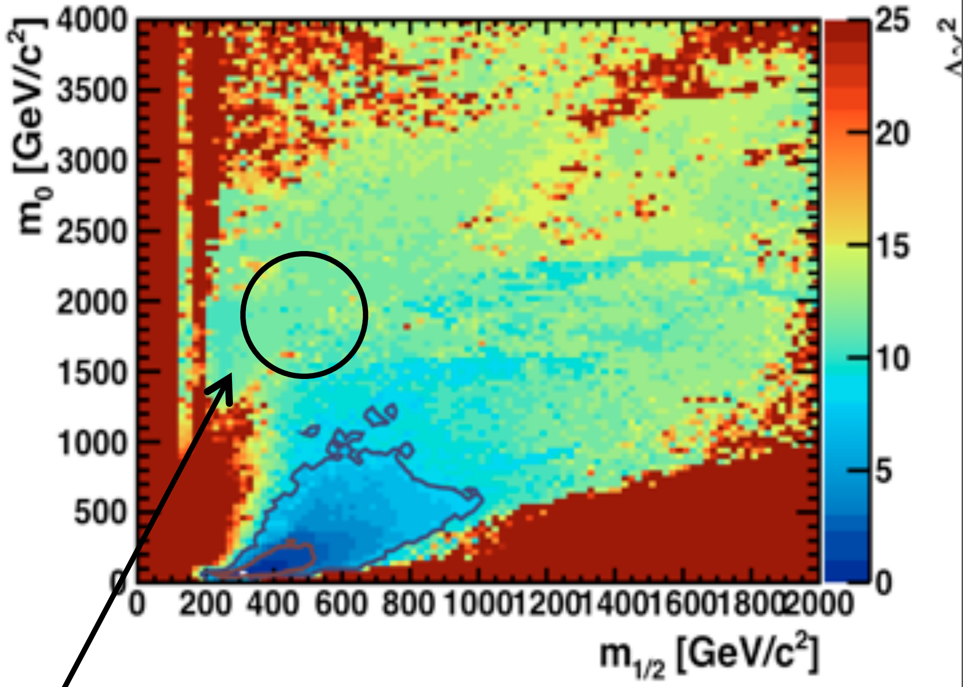
Genetic Algorithm profile likelihood



MultiNest profile likelihood **- see later!**



MasterCode profile likelihood



isolated local maxima

overall best-fit

excluded at ~ 3σ

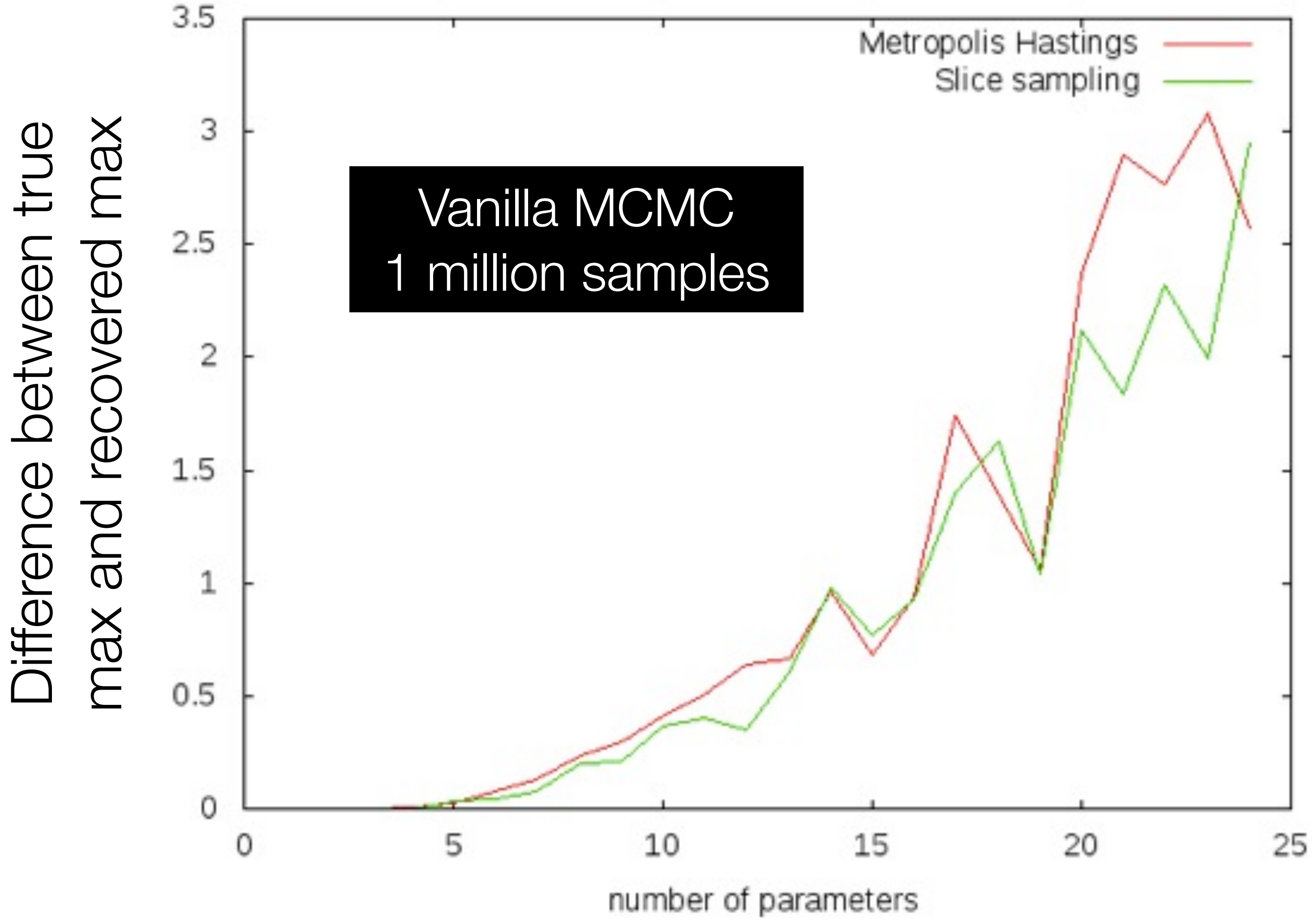
Challenges of profile likelihood evaluation

- MCMC/MultiNest are not designed to find the best-fit point. Bayesian algorithms are designed to map out regions of significant posterior probability mass

- Even for a simple Gaussian toy model, this becomes difficult to do as the number of dimensions of the parameter space increases

- **Profiling with vanilla MCMC or MultiNest scans has to be done with caution!**

Toy multinormal likelihood



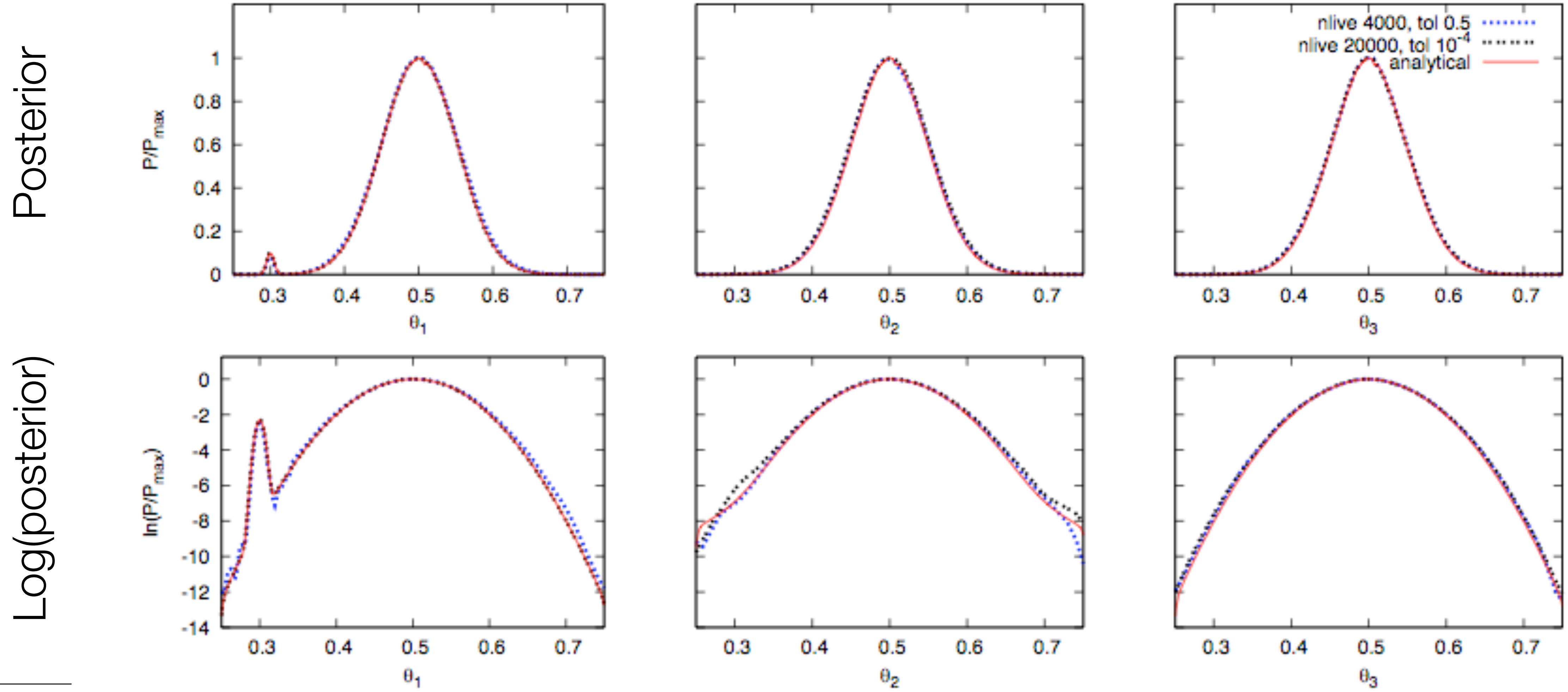
Vanilla MCMC
1 million samples

Posterior pdf from MultiNest scans

- MultiNest is primarily aimed at evaluation of the posterior pdf. It does an excellent job even for multi-modal problems. 8D toy case (Feroz, KC, RT et al, [arXiv:1101.3296])

- The tolerance parameter (**tol**) determines the stopping criterium (based on the incremental change of the value of the local evidence). Lower tol gives a finer exploration around the peak, important for profile likelihood evaluation

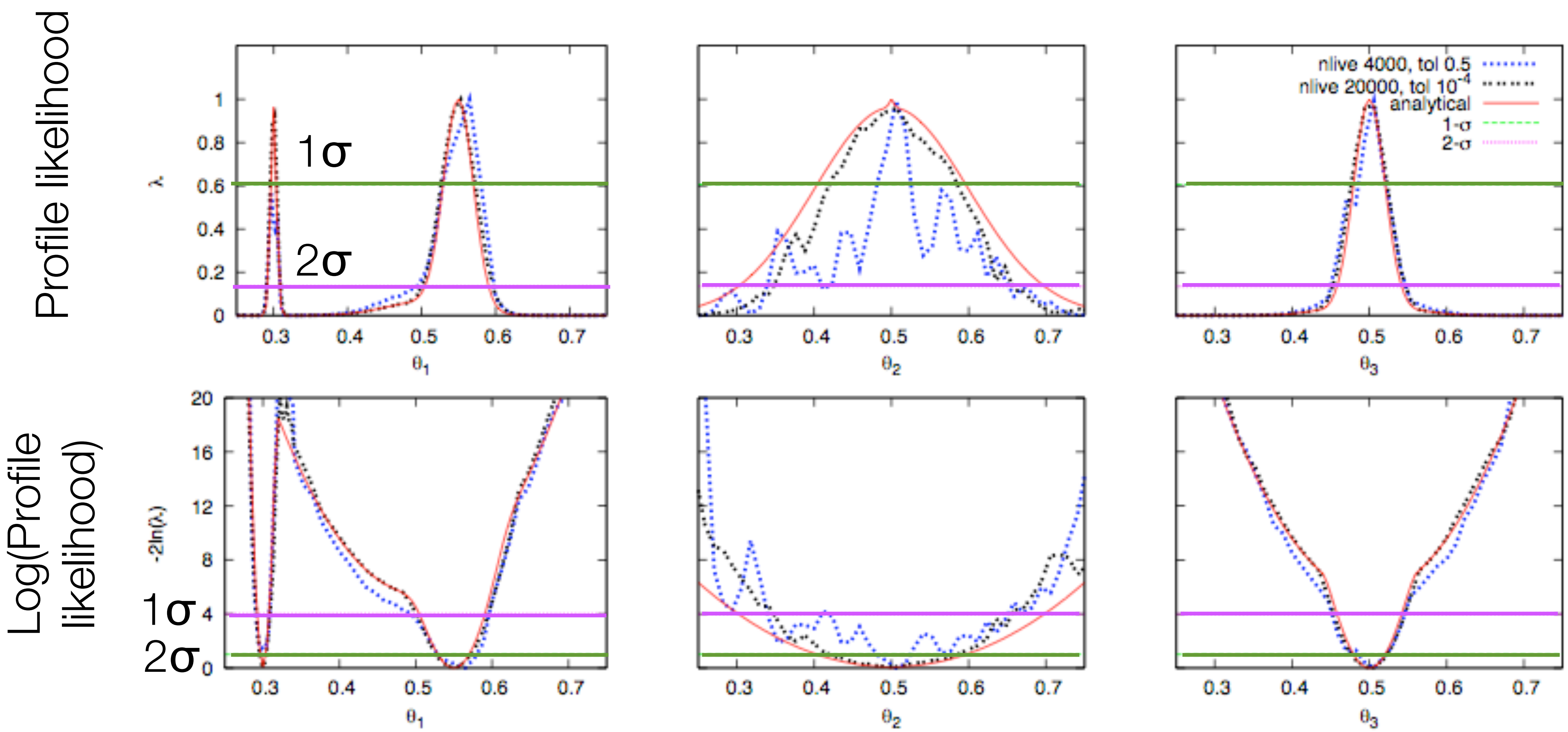
red: analytical **blue: MN, tol=0.5** **black: MN, tol=0.0001**



Profile likelihood from MultiNest scans

- A fairly accurate profile likelihood can be obtained with MultiNest by tuning the tolerance (lower, **tol=0.0001**) and the number of live points (higher, **n_{live}=20,000**) (Feroz, KC, RT et al, [arXiv:1101.3296]), even for highly multi-modal distributions. 8D toy:

red: analytical **blue: MN, tol=0.5** **black: MN, tol=0.0001**

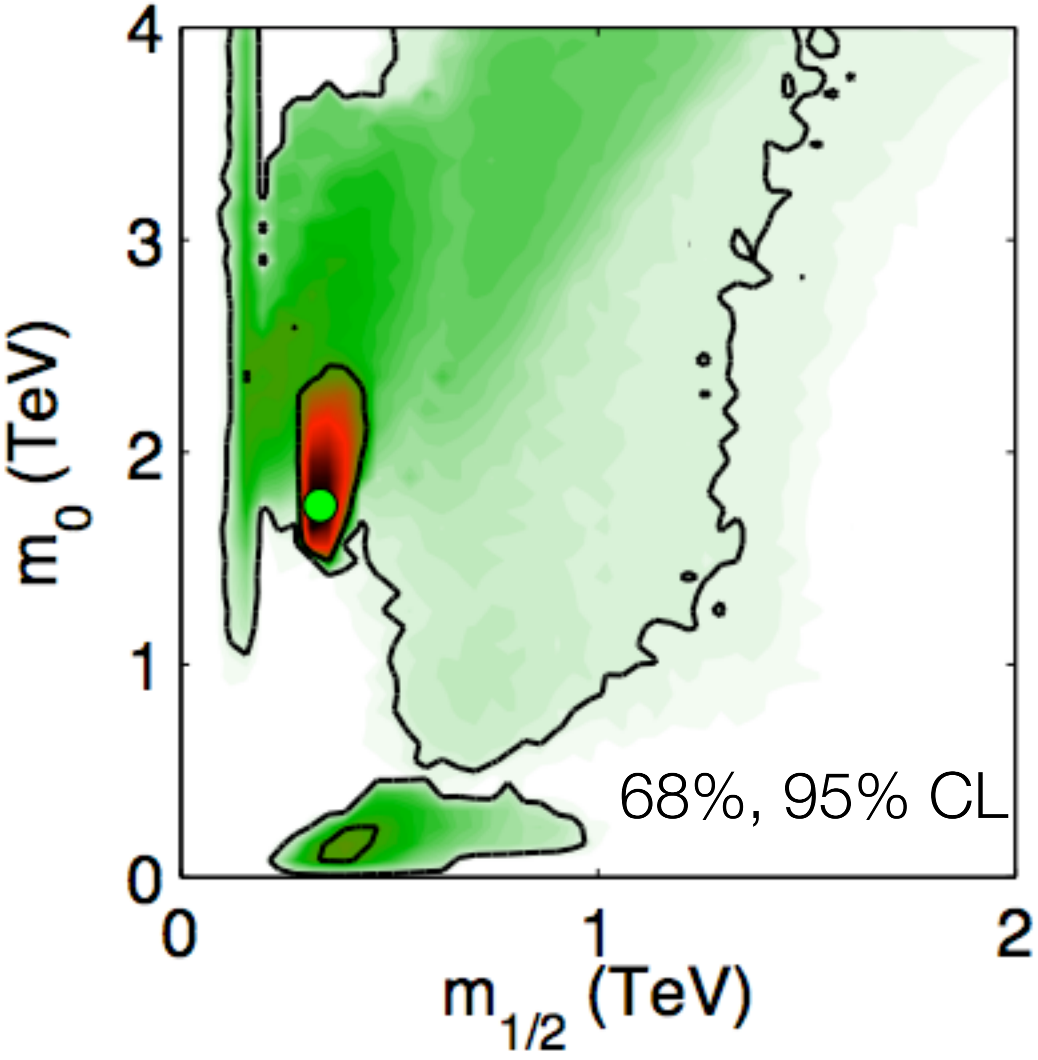


Profile likelihood from MultiNest

- MultiNest scan with 20,000 live points (usually: 4,000) and tolerance 0.0001 (usually: 0.5) results in 5.5 million likelihood evaluations (Akrami et al, GA: 3 million), and best-fit χ-square = 9.26 (Akrami et al, GA: 9.35).

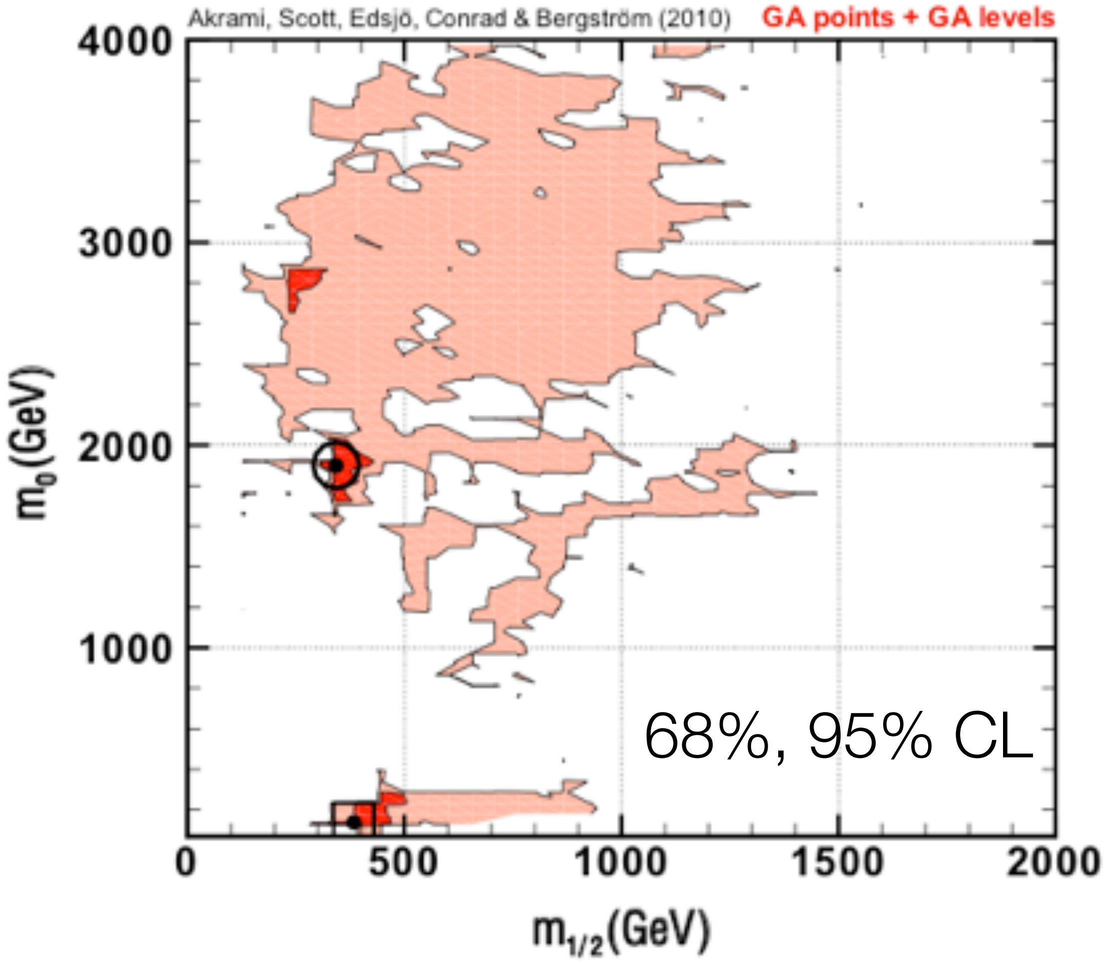
  MultiNest finds a better best-fit + smoother contours than GA.

Profile likelihood MultiNest, tol=10^{-4}
Merged log and flat priors scans



Feroz, KC, RT et al (2011)

Profile likelihood
Genetic algorithm



Akrami et al (2010)

Coverage properties of intervals

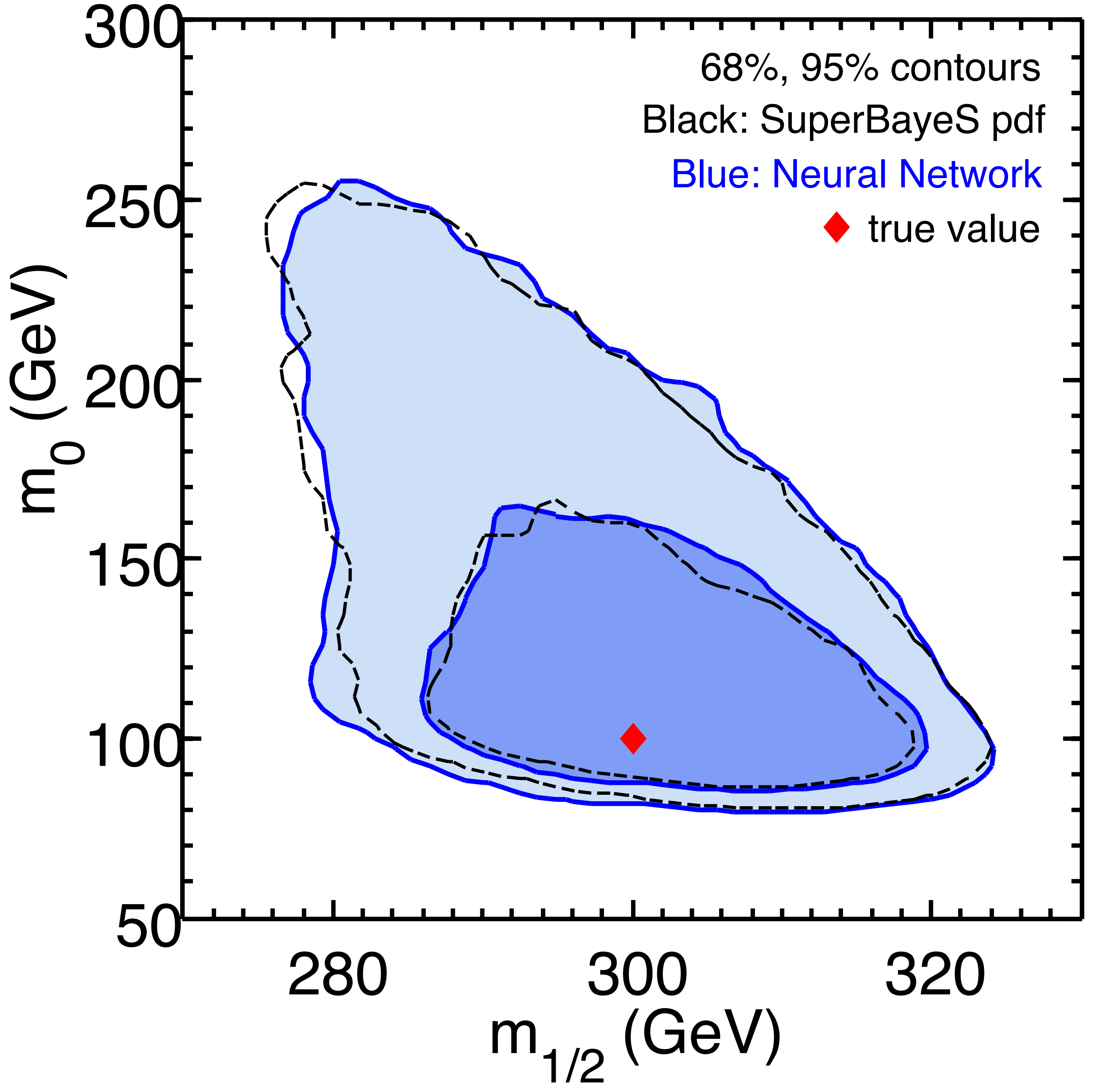
- Coverage of intervals is a frequentist property.
- Guaranteed when using e.g. Feldman-Cousins procedure to build intervals.
- Approximate confidence intervals are obtained via the Neyman construction with profile likelihood ratio as a test statistic.
- From a Bayesian perspective, coverage properties of credible intervals (if desired) can be used to calibrate priors.
- **Coverage studies are computationally expensive:**
  (a) choose fiducial point in parameter space
  (b) generate pseudo-data
  (c) reconstruct credible/confidence interval
  (d) check whether fiducial point within/without interval.

“Instantaneous” inference with neural networks

- **Standard MCMC**
  (SuperBayeS v1.23, 2006 release)
  720 CPU days
- **MultiNest**
  (SuperBayeS v1.5, 2010 release)
  16 CPU days
  speed-up factor: ~ 50

Simulated ATLAS data

Bridges et al (2010)



“Instantaneous” inference with neural networks

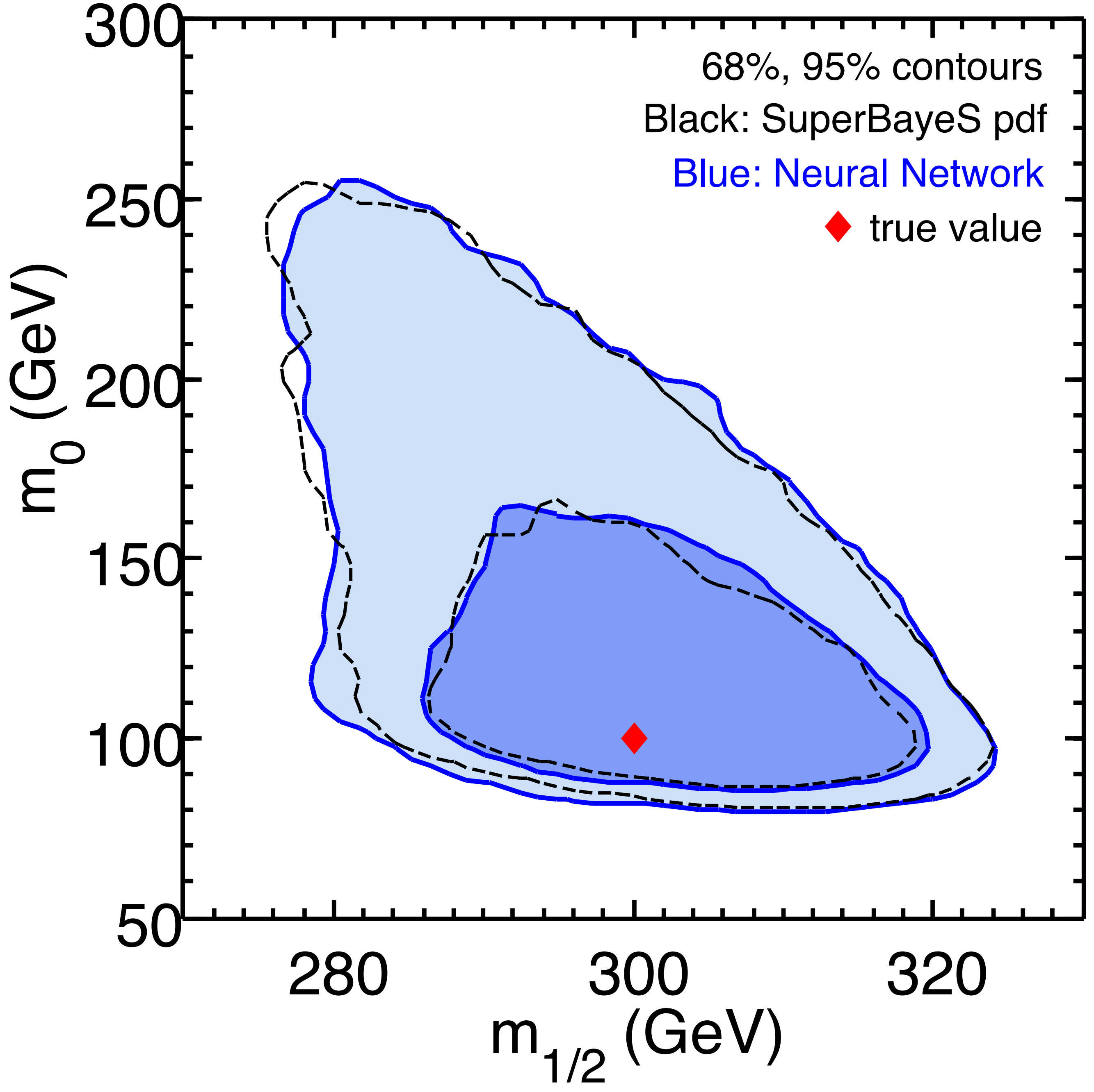
- **Standard MCMC**
  (SuperBayeS v1.23, 2006 release)
  720 CPU days

- **MultiNest**
  (SuperBayeS v1.5, 2010 release)
  16 CPU days
  speed-up factor: ~ 50

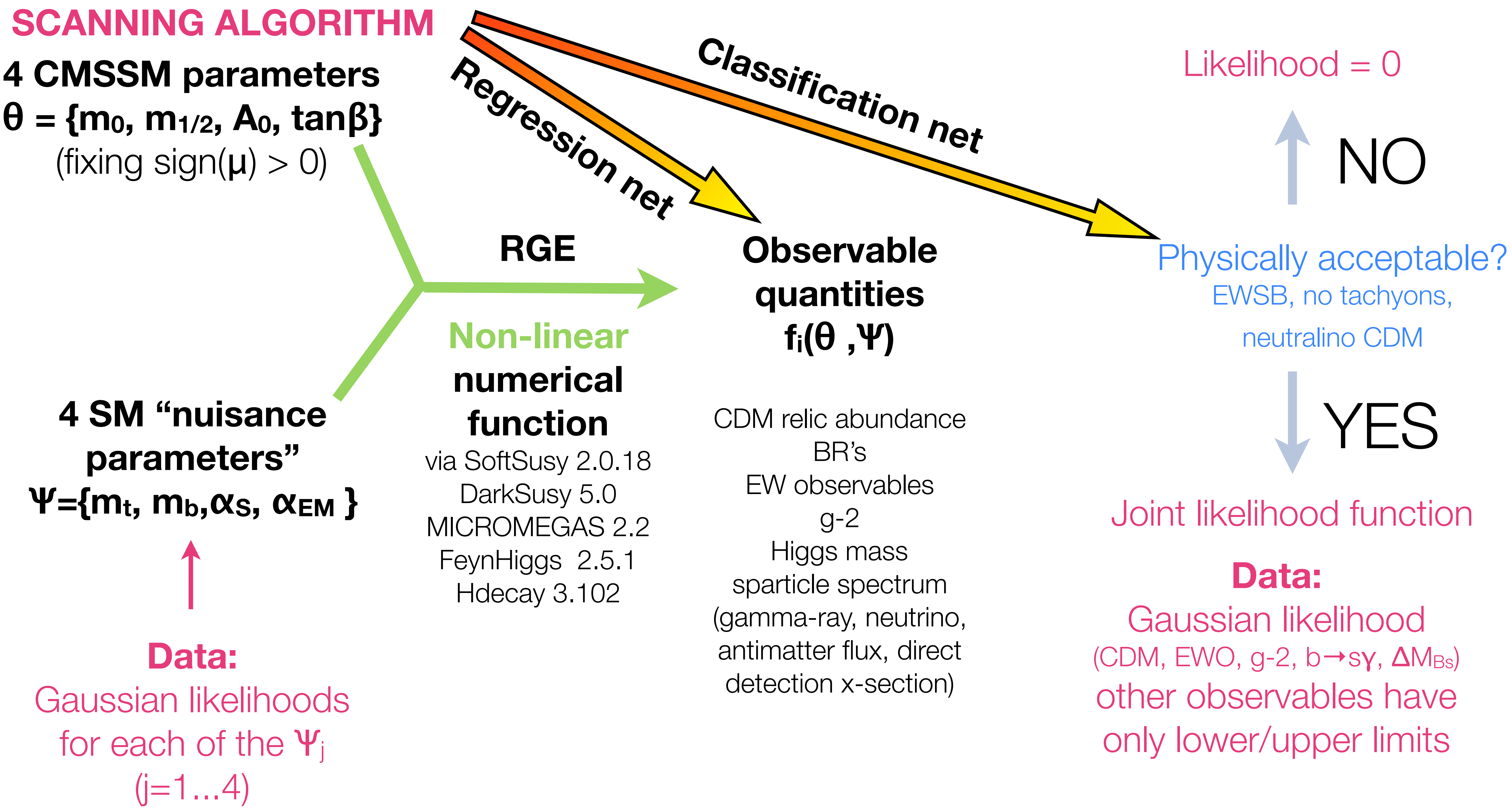
- **SuperBayeS+Neural Networks**
  (Bridges, Cranmer, Feroz, Hobson, Ruiz & RT, 1011.4306)
  less than 1 CPU minute
  speed-up factor: 30’000

Simulated ATLAS data

Bridges et al (2010)



Neural nets shortcuts



SCANNING ALGORITHM

4 CMSSM parameters
$\theta = \{m_0, m_{1/2}, A_0, \tan\beta\}$
(fixing sign(μ) > 0)

Regression net

Classification net

RGE

Non-linear numerical function
via SoftSusy 2.0.18
DarkSusy 5.0
MICROMEGAS 2.2
FeynHiggs 2.5.1
Hdecay 3.102

4 SM "nuisance parameters"
$\Psi=\{m_t, m_b, \alpha_S, \alpha_{EM}\}$

Data:
Gaussian likelihoods for each of the Ψ_j ($j=1\ldots4$)

Observable quantities
$f_i(\theta, \Psi)$

CDM relic abundance
BR's
EW observables
g-2
Higgs mass
sparticle spectrum
(gamma-ray, neutrino, antimatter flux, direct detection x-section)

Likelihood = 0

NO

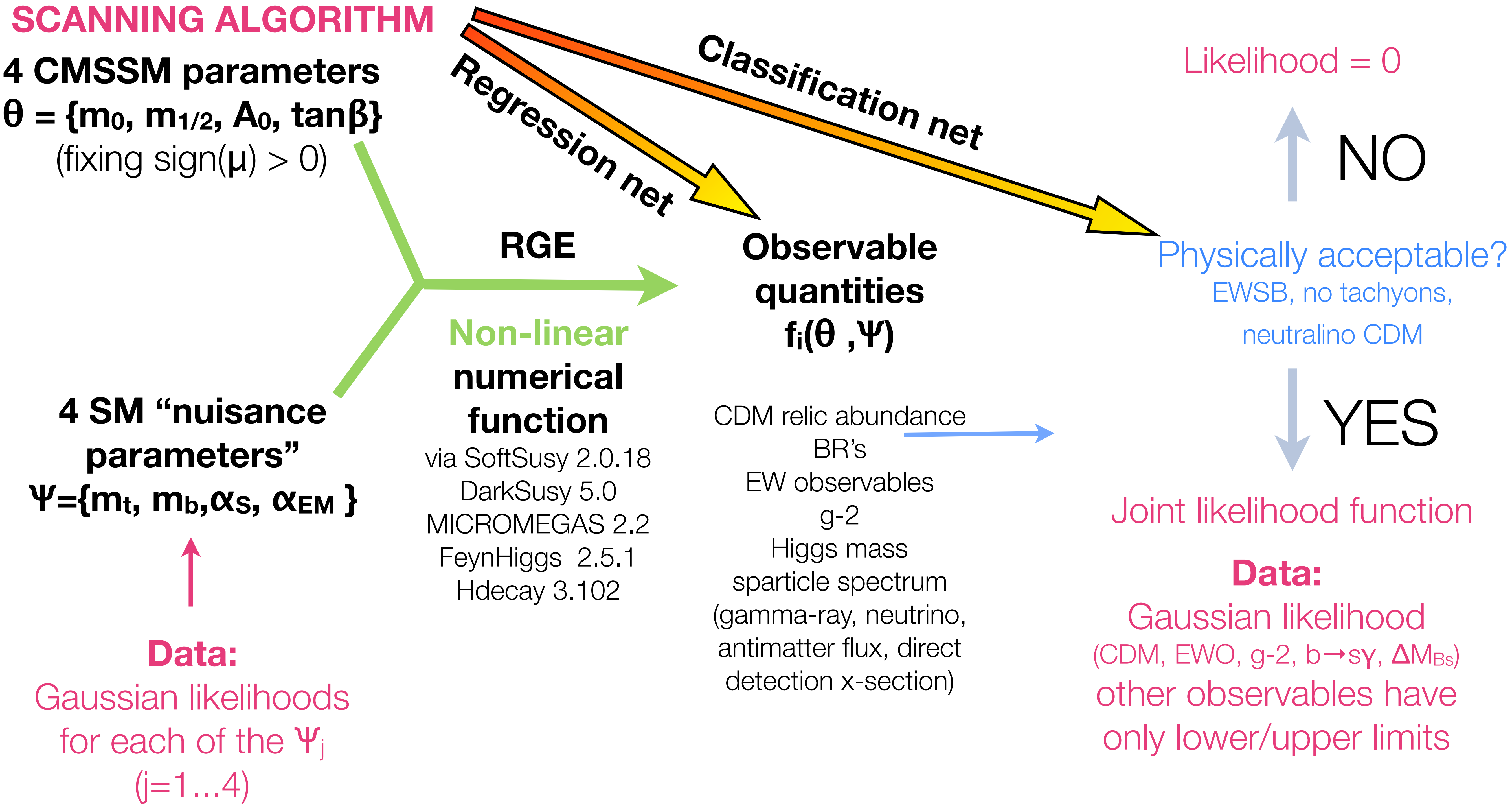
Physically acceptable?
EWSB, no tachyons,
neutralino CDM

YES

Joint likelihood function

Data:
Gaussian likelihood
(CDM, EWO, g-2, $b \to s\gamma$, ΔM_{Bs})
other observables have only lower/upper limits

Neural nets shortcuts



SCANNING ALGORITHM

4 CMSSM parameters
$\theta = \{m_0, m_{1/2}, A_0, \tan\beta\}$
(fixing sign(μ) > 0)

Regression net

Classification net

4 SM "nuisance parameters"
$\Psi=\{m_t, m_b, \alpha_S, \alpha_{EM}\}$

Data:
Gaussian likelihoods
for each of the Ψ_j
($j=1\ldots4$)

RGE

Non-linear
numerical
function
via SoftSusy 2.0.18
DarkSusy 5.0
MICROMEGAS 2.2
FeynHiggs 2.5.1
Hdecay 3.102

Observable
quantities
$f_i(\theta, \Psi)$

CDM relic abundance
BR's
EW observables
g-2
Higgs mass
sparticle spectrum
(gamma-ray, neutrino,
antimatter flux, direct
detection x-section)

Likelihood = 0

NO

Physically acceptable?
EWSB, no tachyons,
neutralino CDM

YES

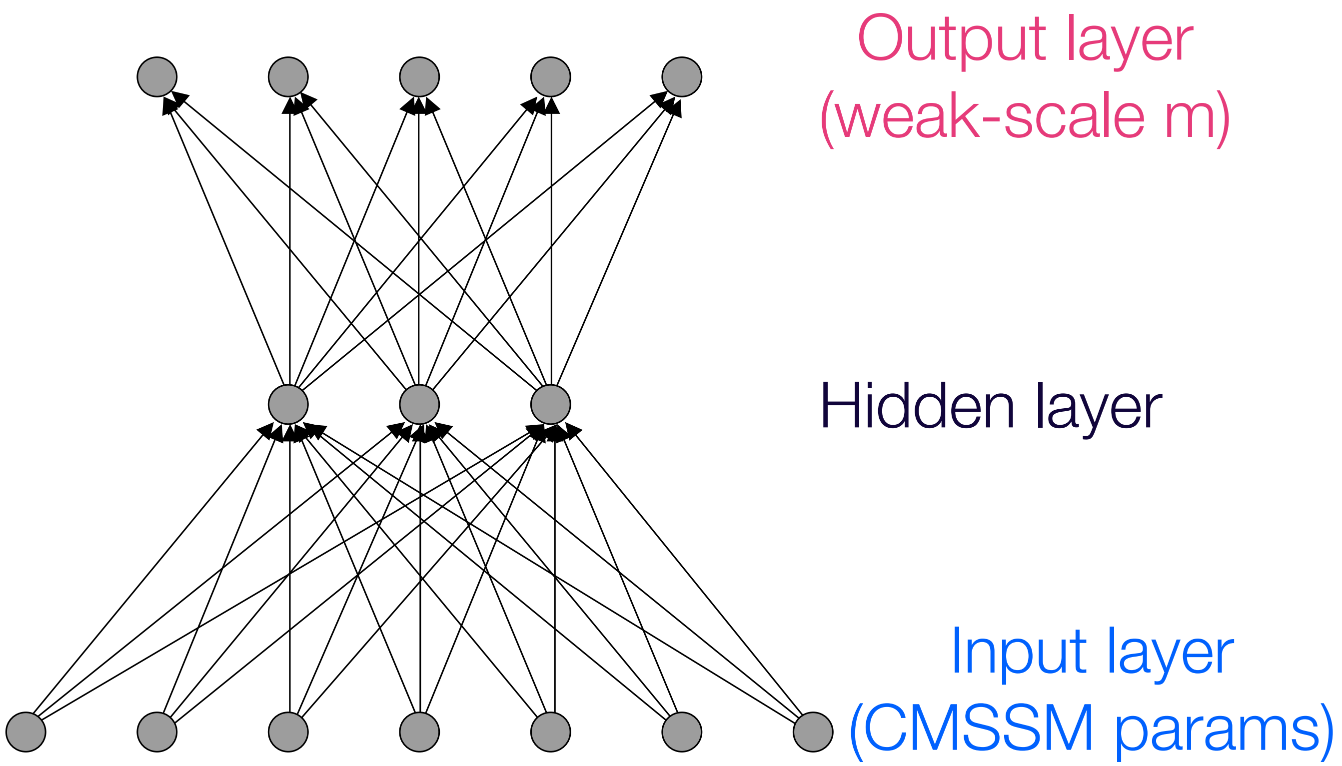
Joint likelihood function

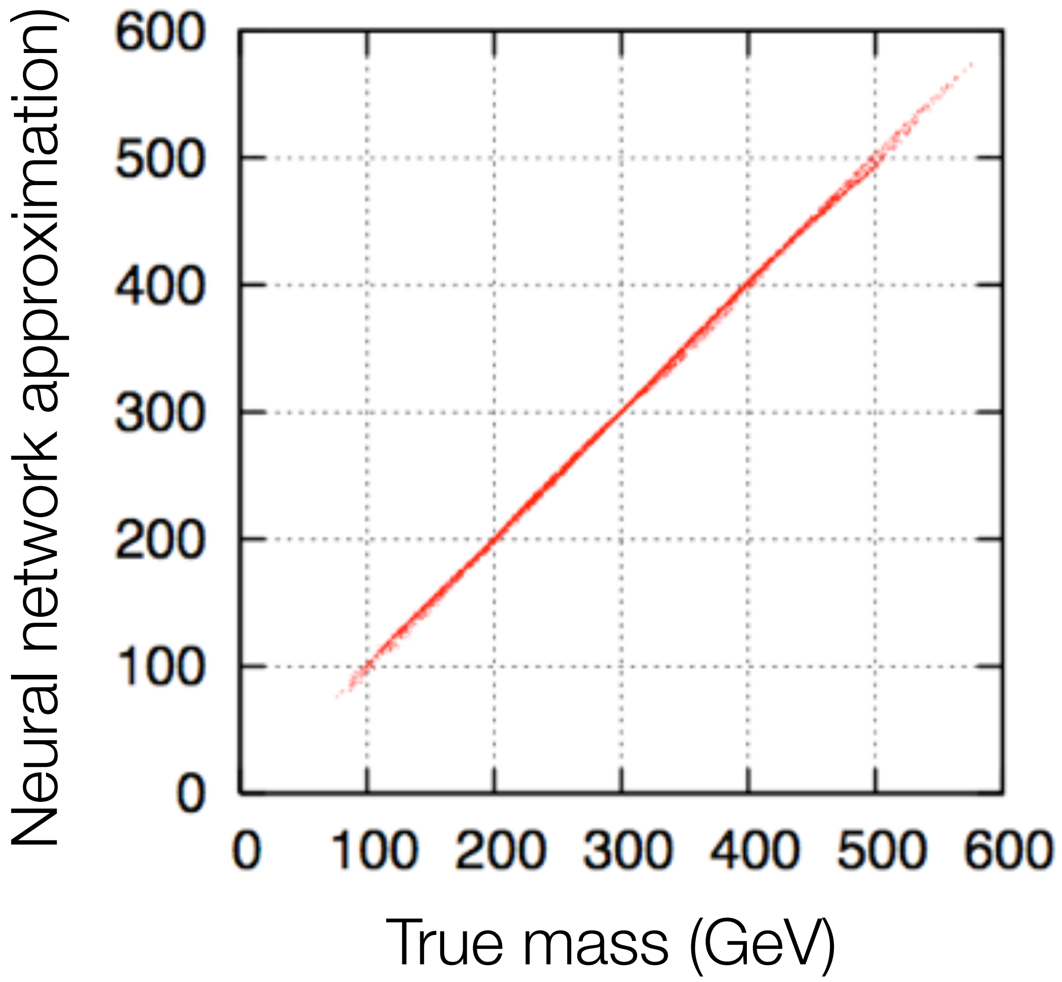
Data:
Gaussian likelihood
(CDM, EWO, g-2, $b \to s\gamma$, ΔM_{Bs})
other observables have
only lower/upper limits

Neural networks technology

Bridges, KC,
RT et al (1011.4306)

- We used a feed-forward multi-layer perceptron to “replace” SoftSusy in predicting the weak-scale masses from the CMSSM input parameters
- After training with a few 1000’s samples, the neural net achieved a correlation > 99.99%

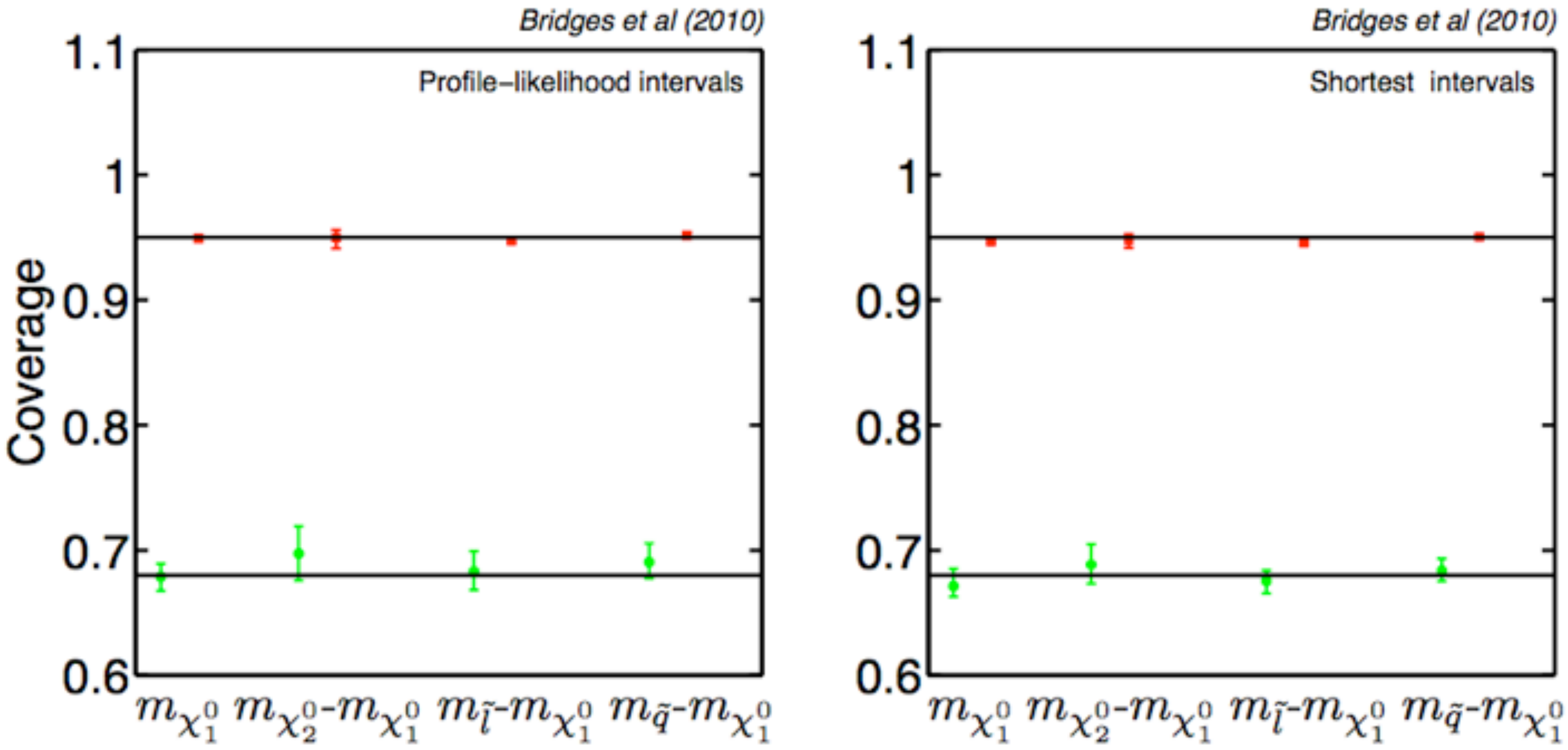




Coverage: are intervals what they say?

- We did 10,000 reconstructions, each with 1 million samples. This would have taken 1,100 CPU yrs using standard methods. **Neural network speed-up is dramatic, of order 10^4.**

- Test case: use weak-scale masses as input, with Gaussian likelihood. Coverage is exact (within noise), as expected:

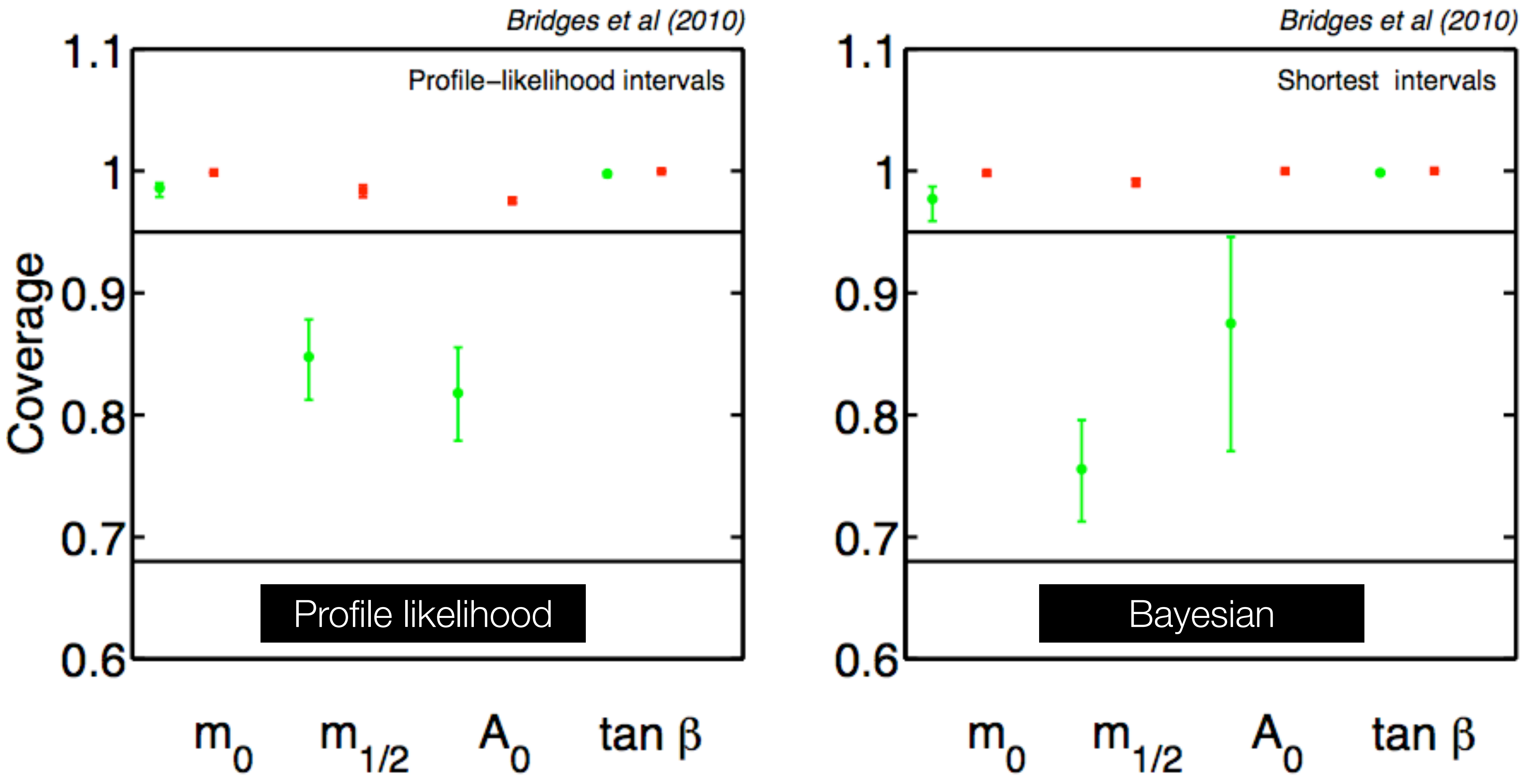


Profile likelihood

Bayesian

Coverage: are intervals what they say?

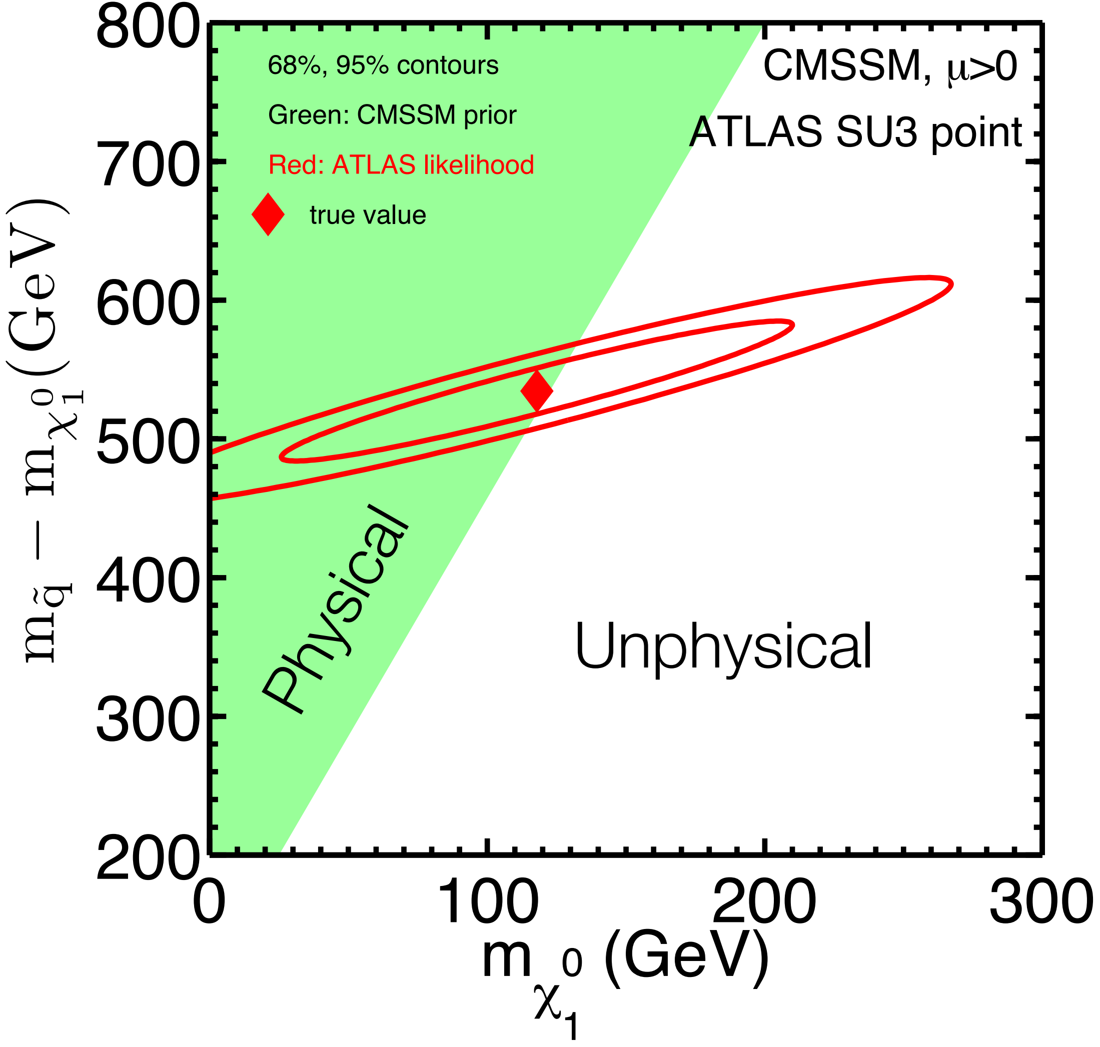
- Mapping back constraints to the CMSSM parameters, we find substantial over-coverage for both Bayesian and profile likelihood intervals:



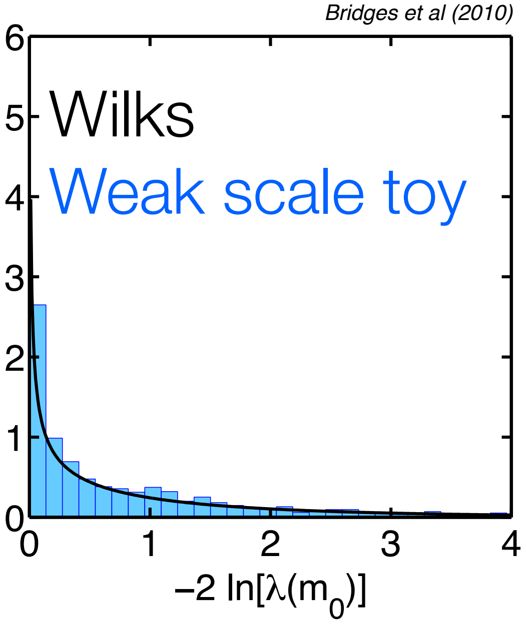
Origin of over-coverage in the CMSSM

- The CMSSM prior introduces “physicality” boundaries in the weak-scale masses space. As a consequence, the distribution of -2 ln(λ) is not well approximated by χ^2 and Wilks’ theorem does not apply.

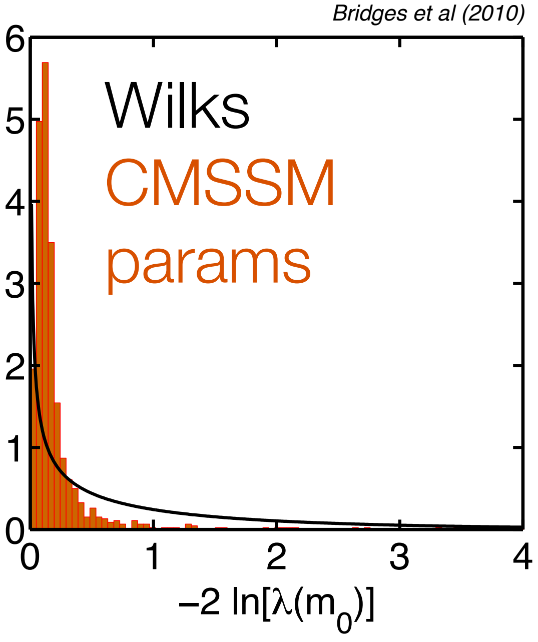
Bridges et al (2010)



Bridges et al (2010)



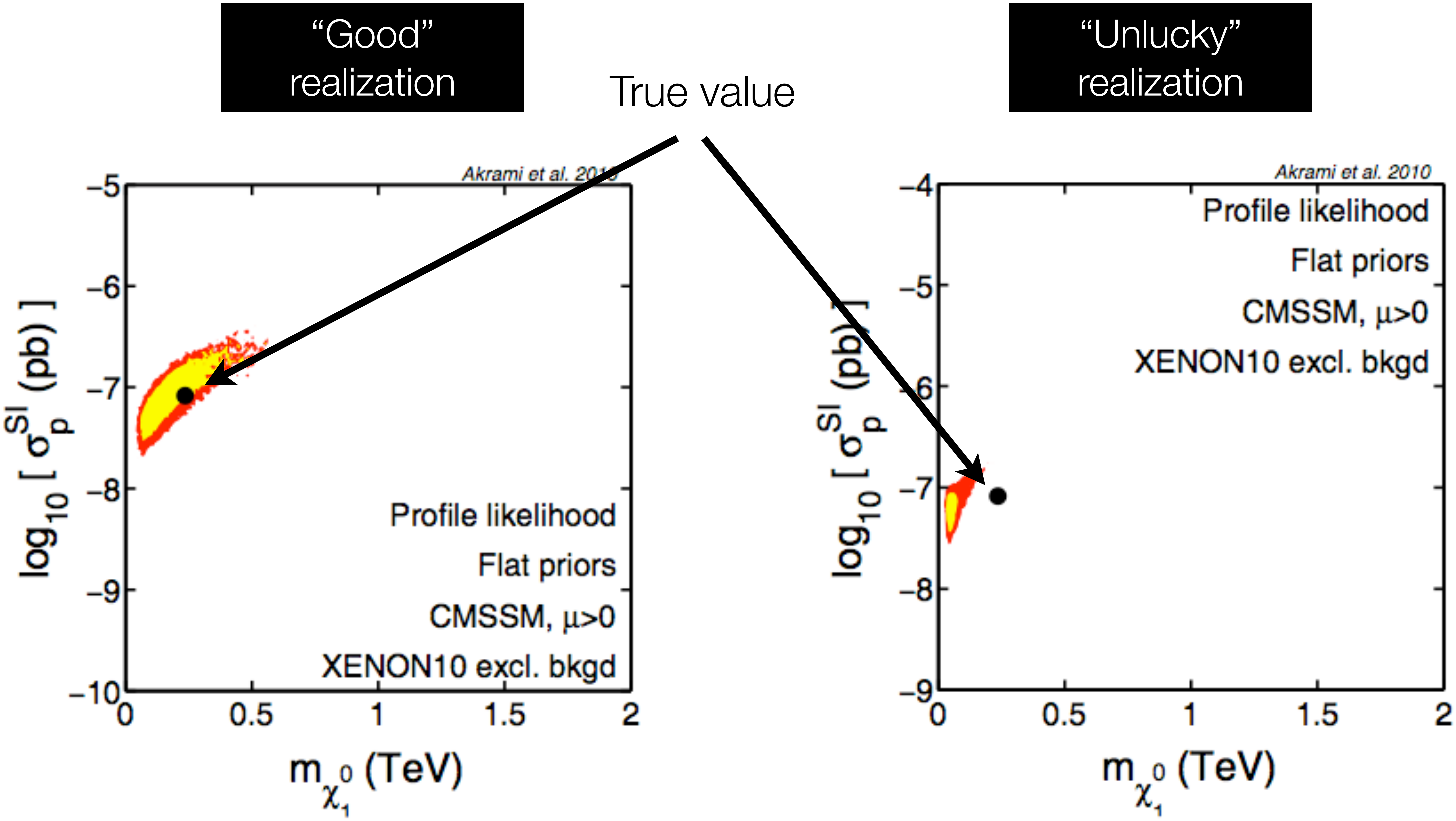
Bridges et al (2010)



Bridges, KC, RT et al (1011.4306)

Coverage from future direct detection ton-scale experiments

- Similar study by Akrami et al (1011.4297), investigating the CMSSM coverage from future ton-scale direct detection experiments (100 realizations):



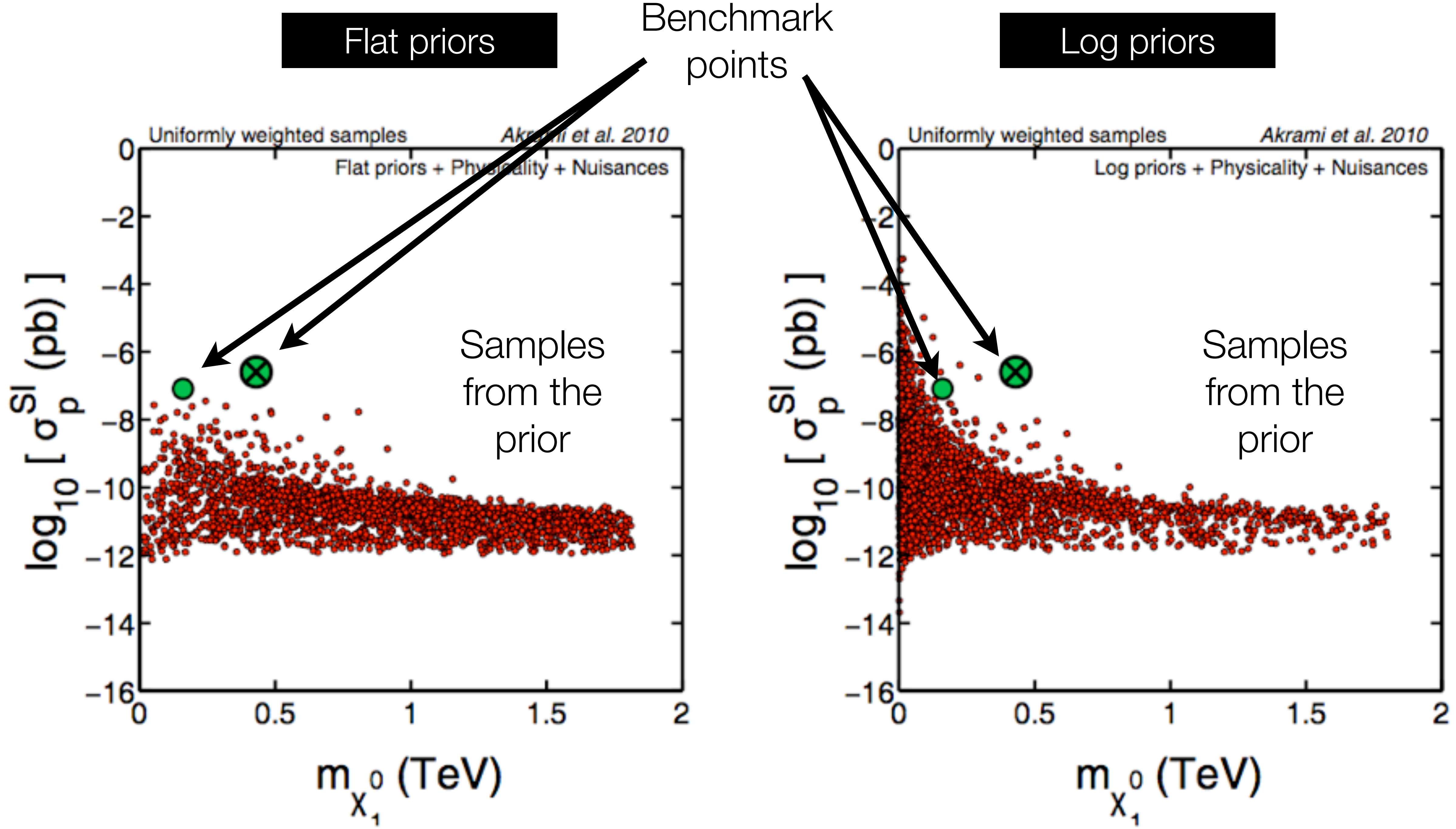
“Good” realization

True value

“Unlucky” realization

Under-coverage from future direct detection ton-scale experiments

- Akrami et al (1011.4297) find significant under-coverage (i.e., too short intervals) in many cases. They claim this to be an effect due to effective priors + insufficient sampling of MultiNest.


Flat priors

Benchmark points

Log priors

Samples from the prior

Samples from the prior

Conclusions

- A rigorous assessment of the statistical performance of various approaches/ algorithms is only just beginning.

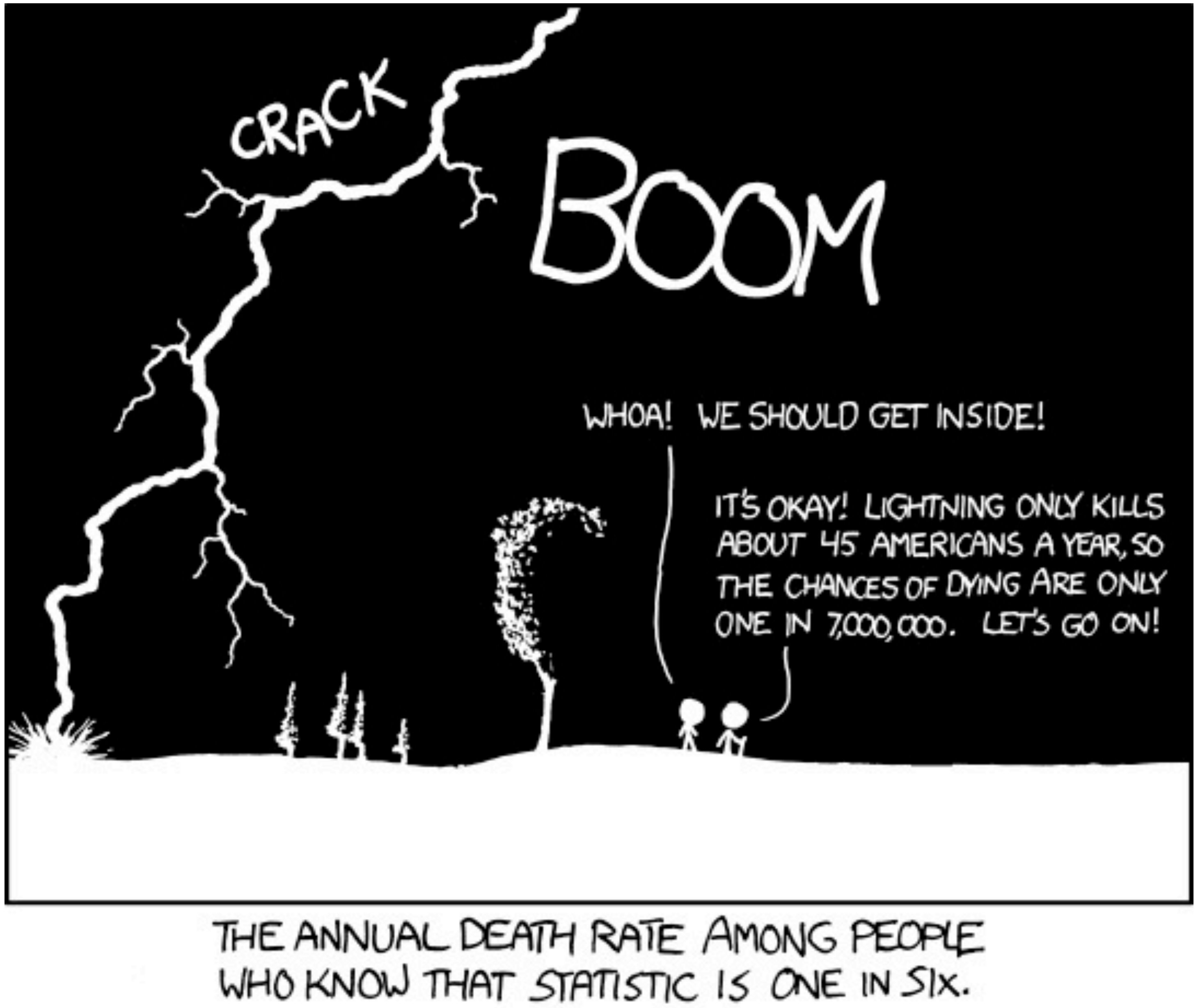
- Coverage studies (LHC, direct detection) find significant over/under-coverage of both Frequentist and Bayesian intervals. Origin: over-simplified likelihood functions (LHC) and ‘hole’ structure of parameter space.

- MultiNest can be tuned to perform profile likelihood scans (at a higher computational cost). “Vanilla” MultiNest scanning parameters (perfectly adequate for Bayesian inference) are **NOT** sufficient for robust profile likelihood estimation. This requires lower tolerance and higher number of samples (~ 10 times more).

- Future work: MultiNest to identify local modes + e.g. Minuit as afterburner to polish the quality of the maximum likelihood point.

Thank you!

Coda:
is statistics good for your health?



xkcd.com