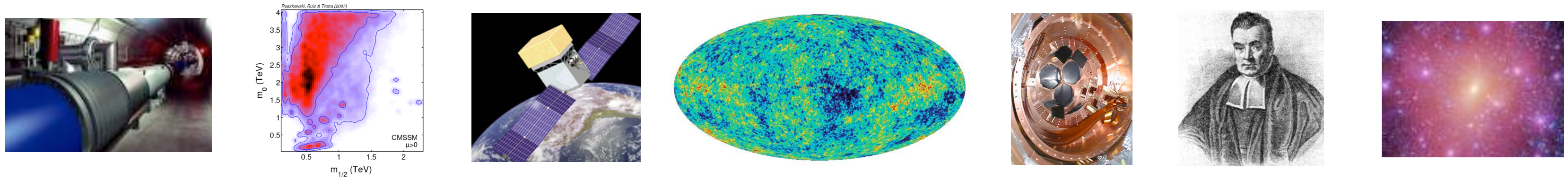
OKC PROSPECTS Workshop - Stockholm, Sept 2010



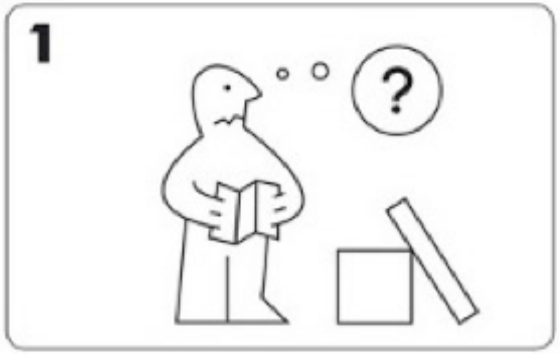
Which statistical approach to SUSY phenomenology?

Roberto Trotta
Imperial College London, Astrophysics

Thanks to Oliver Buchmuller for providing some of the plots

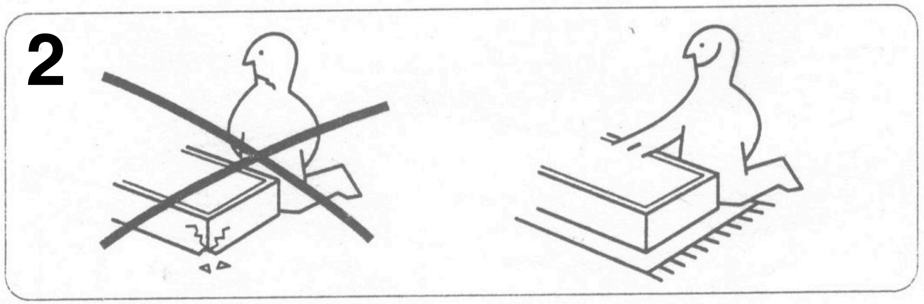
Imperial College
London

PRATSTUND (a flat-packed talk)



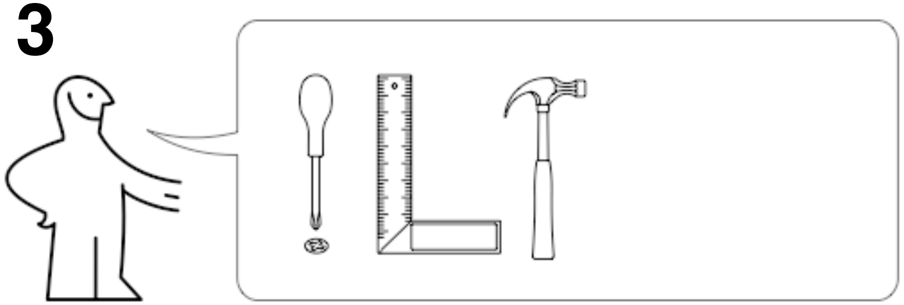
The need for global scans

2



Statistics: do’s and don’ts

3



Tools and results

Why does statistics matter?

- Advanced statistical methods are required for a number of reasons:
    - **Efficiency:** naive approaches become quickly ineffective even for moderate dimensional parameter spaces
    - **Data sets complexity:** upcoming data sets (most notably LHC) require complex modelling for their interpretation
    - **Weak signals:** often discoveries are made in the small sample regime (e.g., direct detection), where statistics does matter
    - **Multi-messenger approach:** cross-correlating probes further increases the complexity of their joint parameter space (incl astro nuisance parameters)

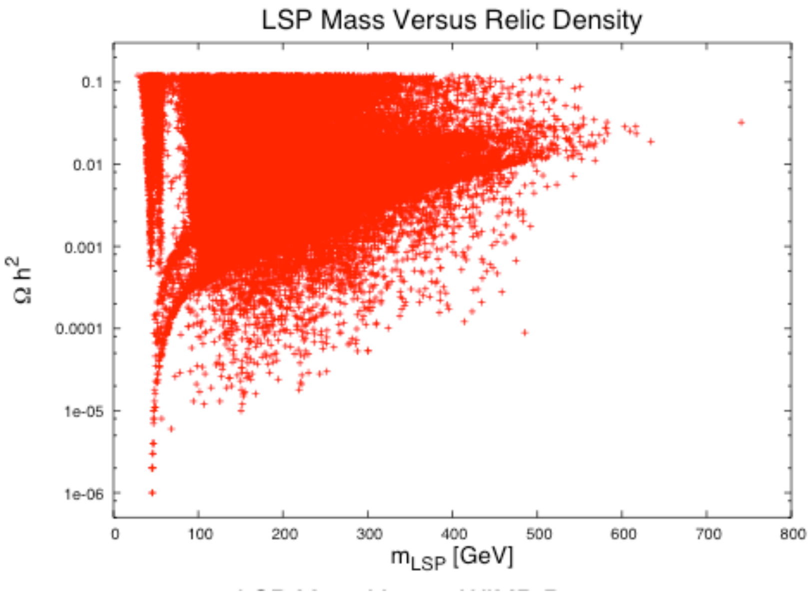
Exploration with “random scans”

C. F. B
Hewett
Supers
Prejudi
[arXiv

check t
pMSSM

- Points accepted/rejected in a in/out fashion (e.g., 2-sigma cuts)
- No statistical measure attached to density of points: no probabilistic interpretation of results possible, although the temptation cannot be resisted...
- Inefficient in high dimensional parameters spaces (D>5)
- **HIDDEN PROBLEM:** Random scan explore only a very limited portion of the parameter space!

One recent example:
Berger et al (0812.0980)
pMSSM scans
(20 dimensions)

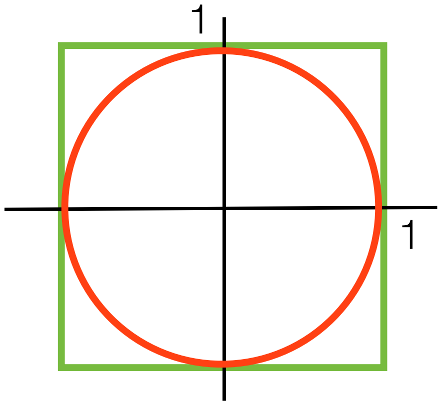


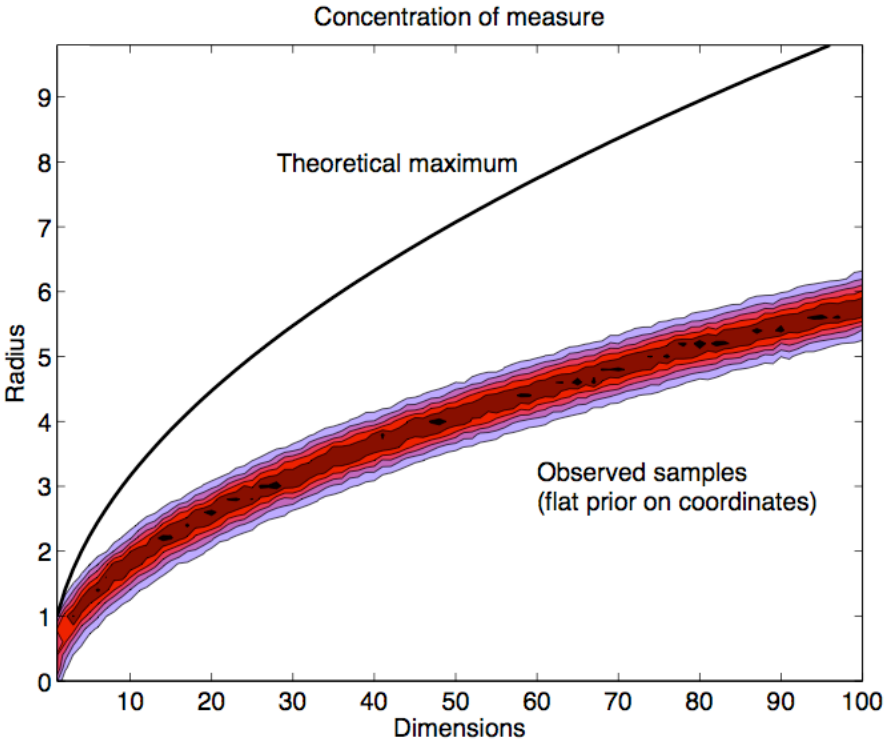
Random scans explore only a small fraction of the parameter space

Imperial College London

- "Random scans" of a high-dimensional parameter space only probe a very limited sub-volume: this is **the concentration of measure phenomenon.**

- **Statistical fact:** the norm of *D* draws from U[0,1] concentrates around $(D/3)^{1/2}$ with constant variance



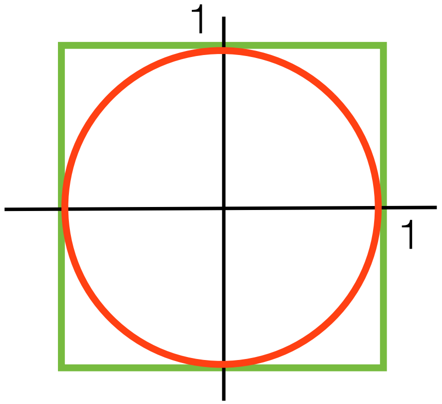


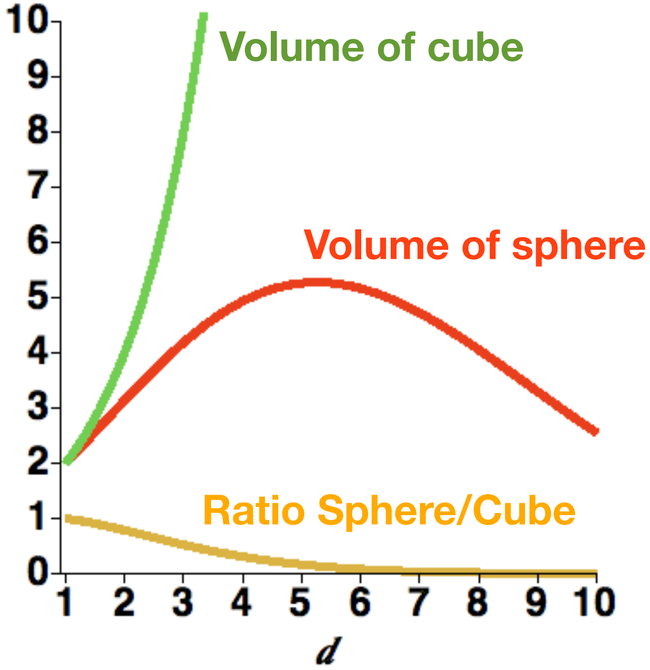
Geometry in high-D spaces

Imperial College London

- **Geometrical fact:** in *D* dimensions, most of the volume is near the boundary. The volume inside the spherical core of *D*-dimensional cube is negligible.

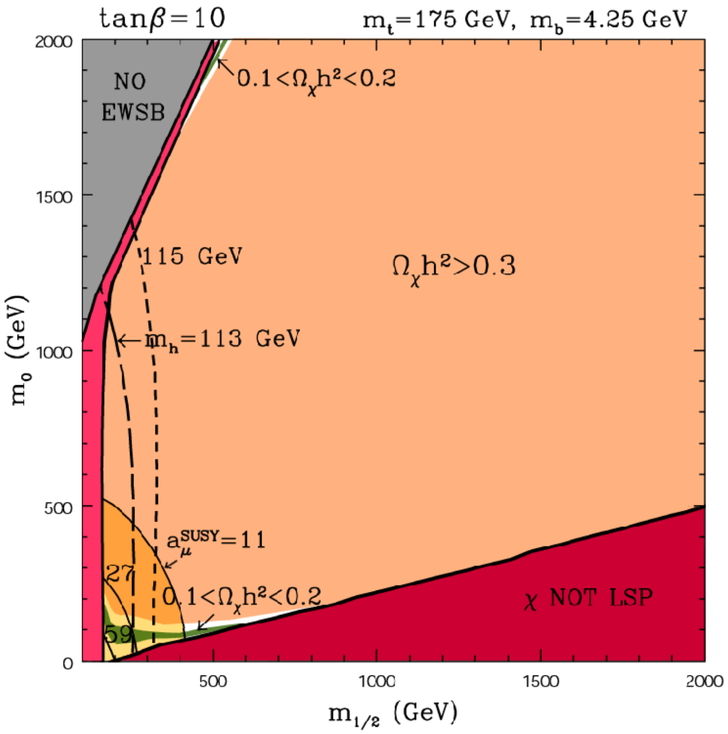
Together, these two facts mean that random scan only explore a very small fraction of the available parameter space in high-dimesional models.

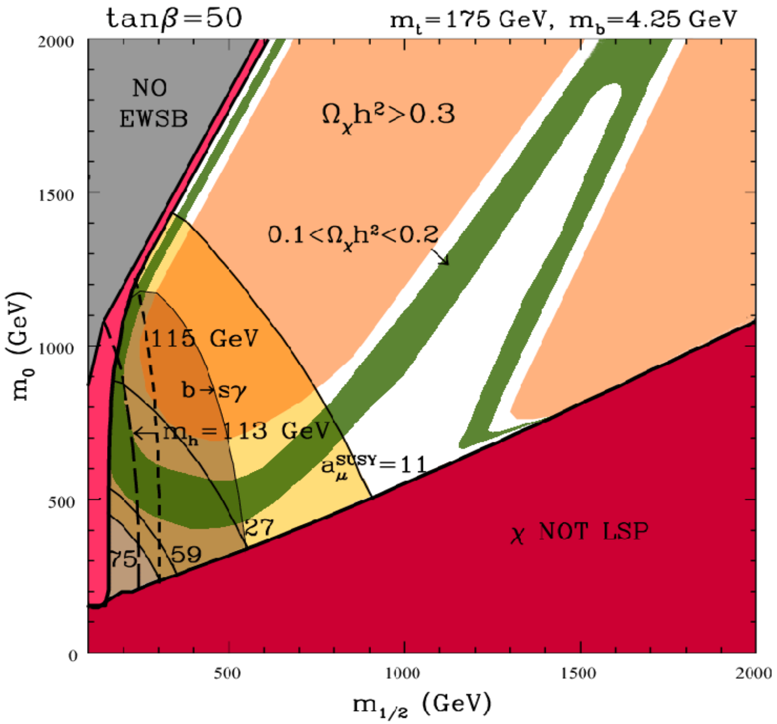




2D scans

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.

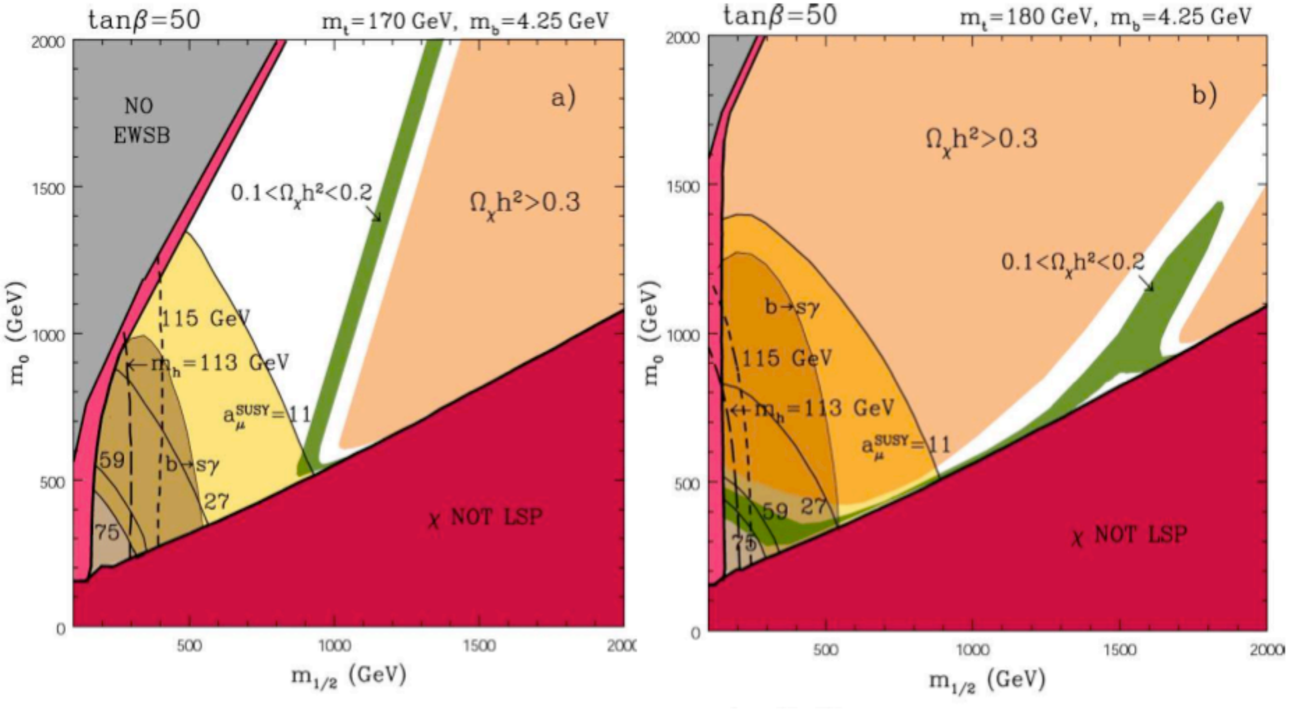




Roszkowski et al (2001)

Dependency on SM (nuisance) parameters

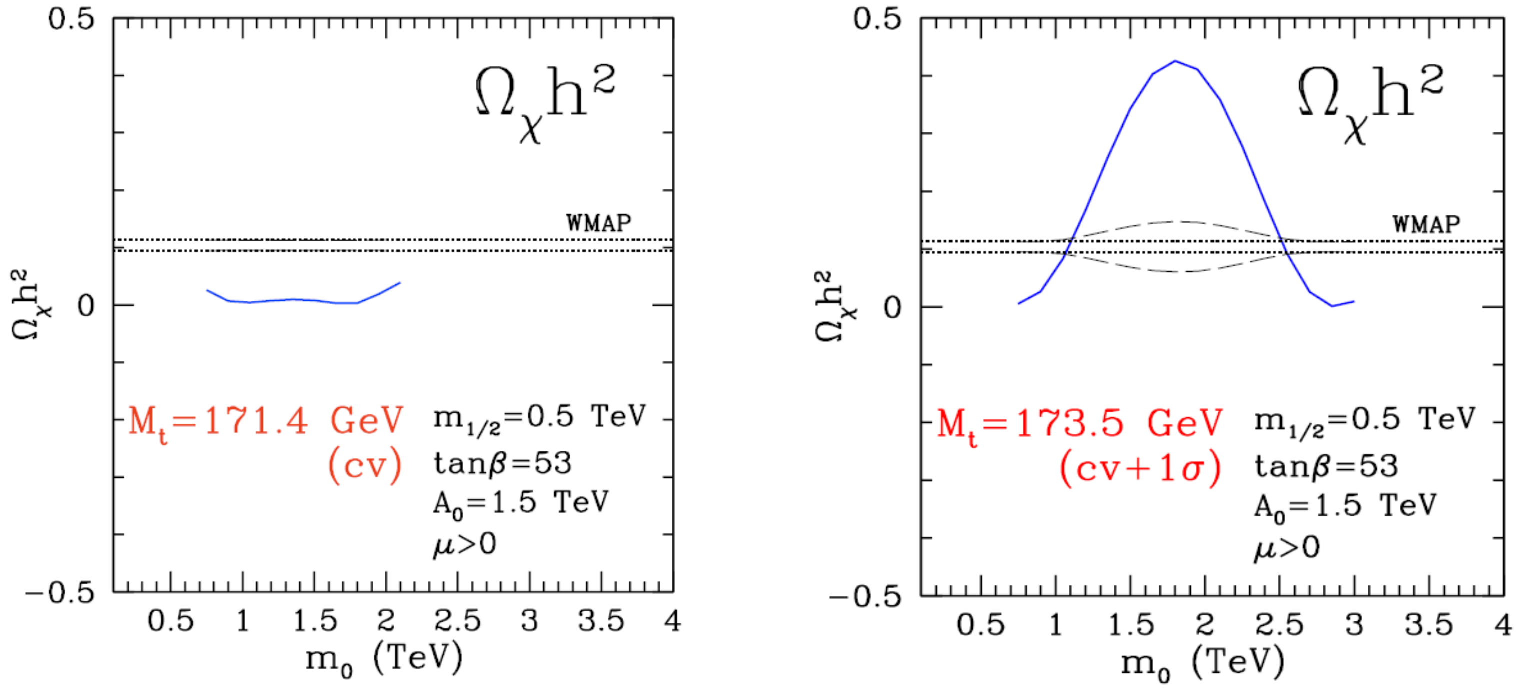
M_{top}=170 GeV M_{top}=180 GeV



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

Impact of top mass on the relic abundance

Changing M_{top} within ±1σ has dramatic consequences for the predicted relic abundance: this parameter cannot be fixed to its central value.



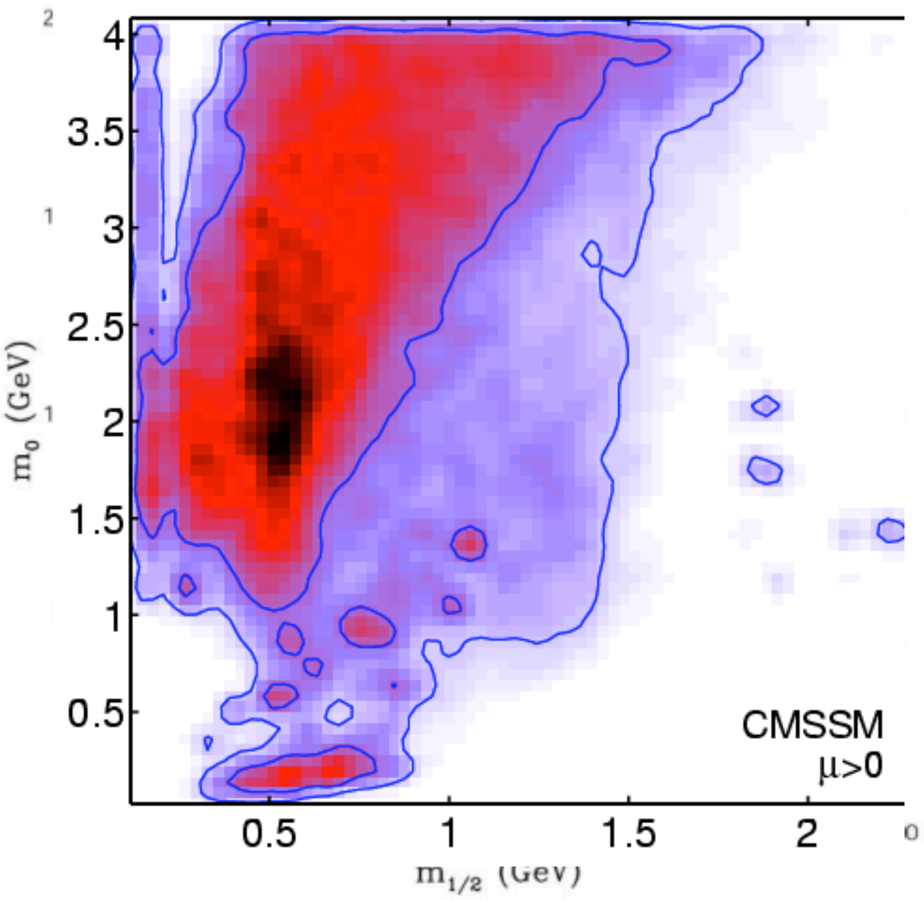
Roszkowski et al (2007)

Solution: global fits

Imperial College London

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.

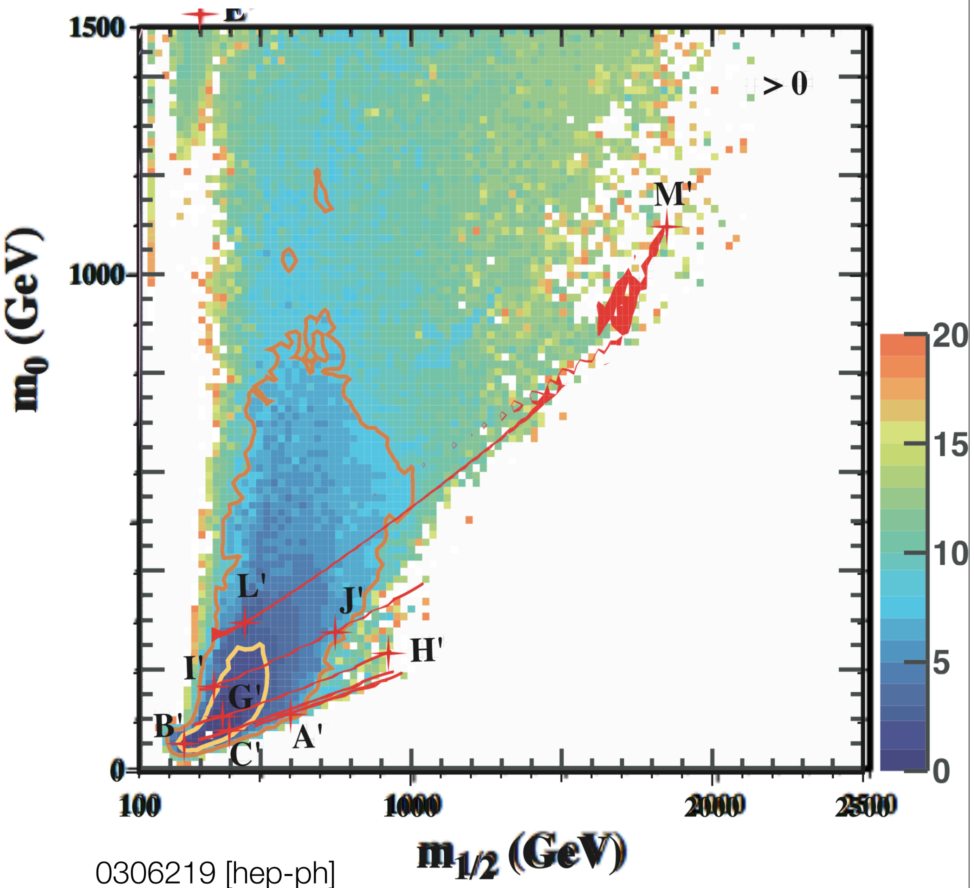


The “WMAP strips”

In 2D scans, enforcing the cosmological relic abundance results in narrow “allowed regions” (the “WMAP strips”), whose location changes with the value of the fixed parameters.

Once fixed parameters are included and hidden dimensions accounted for, WMAP strips widen to become “WMAP blobs”

WMAP strips a few years ago

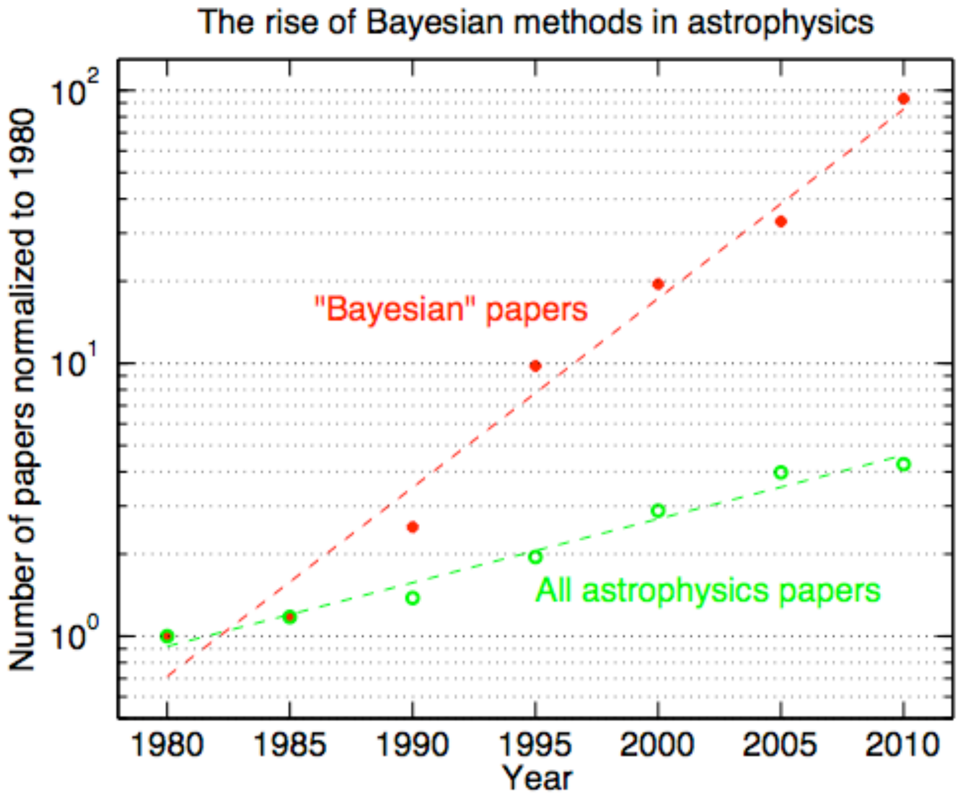


0306219 [hep-ph]

Bayesian methods on the rise

Imperial College London

The frequentist approach (= probability as frequency, based on the likelihood) is naturally suited to particle physics. Bayesian methods are being imported from astrophysics, where they are the norm:



Bayes’ theorem

posterior

likelihood

prior

$$P(\theta|d,I) = \frac{P(d|\theta,I)P(\theta|I)}{P(d|I)}$$

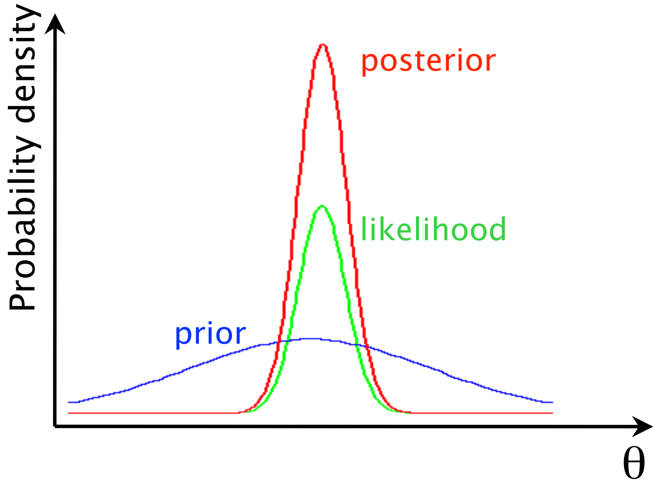
evidence

θ: parameters
d: data
I: any other external information, or the assumed model

For parameter inference it is sufficient to consider

$$P(\theta|d,I) \propto P(d|\theta,I)P(\theta|I)$$

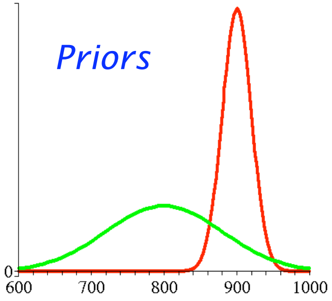
$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$



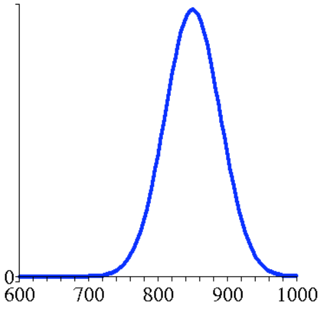
The matter with priors

- In parameter inference, prior dependence will **in principle** vanish for strongly constraining data.

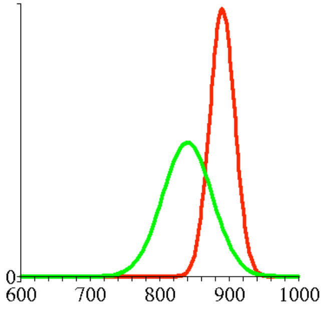
  THIS IS CURRENTLY NOT THE CASE EVEN FOR THE CMSSM!


Priors

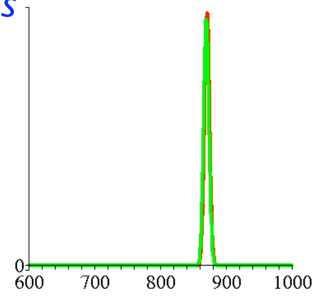
Likelihood (1 datum)

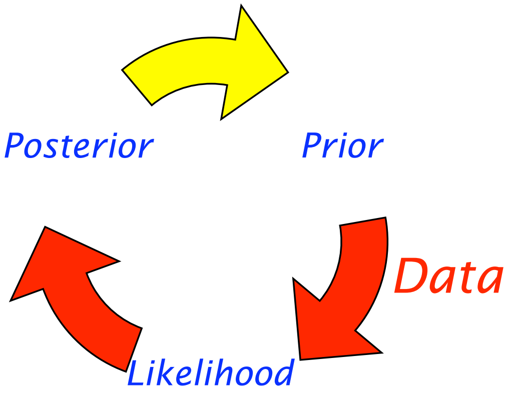


Posterior after 1 datum



Posterior after 100 data points




Posterior

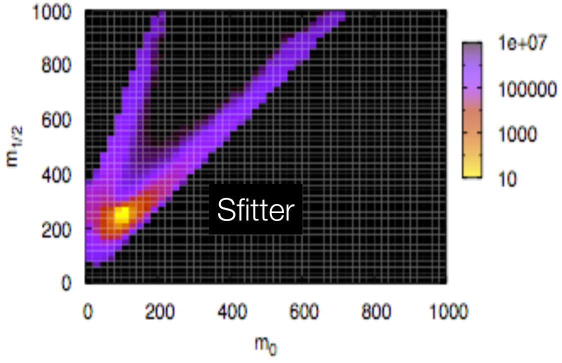
Prior

Data

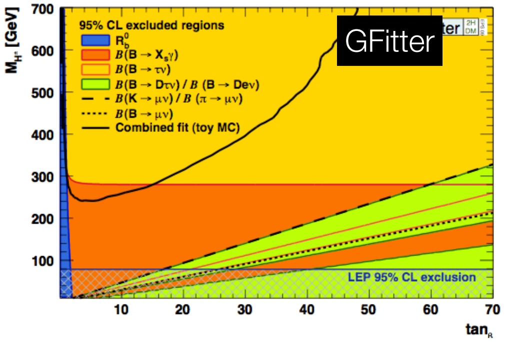
Likelihood

Global fits in the LHC era

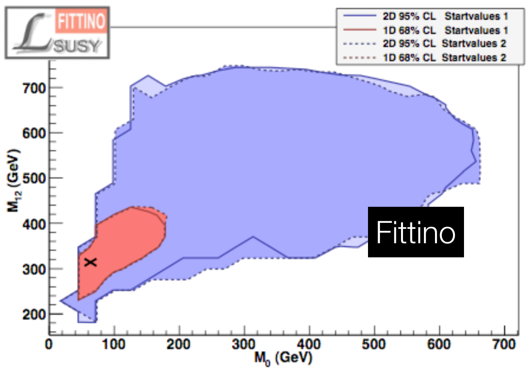
Imperial College London



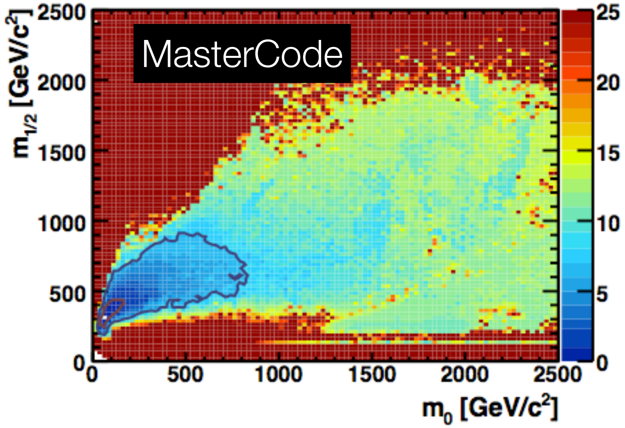
R. Lafaye , M. Rauch, T. Plehn, D. Zerwas



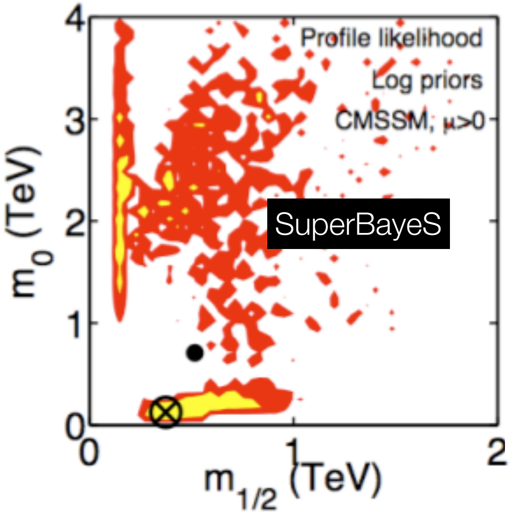
H. Flächer, M. Goebel, J. Haller,
A. Höcker, K. Mönig, J. Stelzer



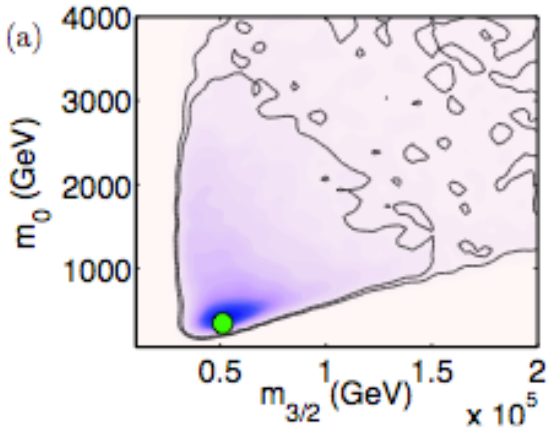
P. Bechtle, K. Desch, M. Uhlenbrock, P. Wienemann



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

Roberto Trotta

The general Bayesian solution

$$P(\theta|d, I) \propto P(d|\theta, I)P(\theta|I)$$

- Once the RHS is defined, how do we evaluate the LHS?
- Analytical solutions exist only for the simplest cases (e.g. Gaussian linear model)
- Cheap computing power means that numerical solutions are often just a few clicks away!
- Workhorse of Bayesian inference: Markov Chain Monte Carlo (MCMC) methods. A procedure to generate a list of samples from the posterior.

MCMC estimation

Imperial College London

$$P(\theta|d,I) \propto P(d|\theta,I)P(\theta|I)$$

- A Markov Chain is a list of samples θ_1, θ_2, θ_3,... whose density reflects the (unnormalized) value of the posterior
- A MC is a sequence of random variables whose *(n+1)*-th elements only depends on the value of the *n*-th element
- **Crucial property:** a Markov Chain converges to a stationary distribution, i.e. one that does not change with time. In our case, the posterior.
- From the chain, expectation values wrt the posterior are obtained very simply:

$$\langle\theta\rangle = \int d\theta P(\theta|d)\theta \approx \frac{1}{N}\sum_i \theta_i$$

$$\langle f(\theta)\rangle = \int d\theta P(\theta|d) f(\theta) \approx \frac{1}{N}\sum_i f(\theta_i)$$

Reporting inferences

- **Once P(θ|d, I) found, we can report inference by:**
    - Summary statistics (best fit point, average, mode)
    - Credible regions (e.g. shortest interval containing 68% of the posterior probability for θ). Warning: this has **not** the same meaning as a frequentist confidence interval! (Although the 2 might be formally identical)
    - Plots of the marginalised distribution, integrating out nuisance parameters (i.e. parameters we are not interested in). This generalizes the propagation of errors:

$$P(\theta|d, I) = \int d\phi P(\theta, \phi|d, I)$$

What does x=1.00±0.01 mean?

Imperial College London

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$\text{Notation}: \quad x \sim N(\mu, \sigma^2)$$

- **Frequentist statistics (Fisher, Neymann, Pearson):**
  E.g., estimation of the mean μ of a Gaussian distribution from a list of observed samples x_1, x_2, x_3...
  The sample mean is the Maximum Likelihood estimator for μ:

  $\mu_{ML} = X_{av} = (x_1 + x_2 + x_3 + \ldots x_N)/N$

- **Key point:**
  in $P(X_{av})$, X_{av} is a random variable, i.e. one that takes on different values across an ensemble of infinite (imaginary) identical experiments. X_{av} is distributed according to $X_{av} \sim N(\mu, \sigma^2/N)$ **for a fixed true μ**
  The distribution applies to imaginary replications of data.

What does x=1.00±0.01 mean?

- **Frequentist statistics (Fisher, Neymann, Pearson):**
  The final result for the confidence interval for the mean

$$P(\mu_{ML} - \sigma/N^{1/2} < \mu < \mu_{ML} + \sigma/N^{1/2}) = 0.683$$

- This means:
  If we were to repeat this measurements many times, and obtain a 1-sigma distribution for the mean, the true value μ would lie inside the so-obtained intervals 68.3% of the time

- This is not the same as saying: “The probability of μ to lie within a given interval is 68.3%”. This statement only follows from using Bayes theorem.

What does x=1.00±0.01 mean?

- **Bayesian statistics (Laplace, Gauss, Bayes, Bernouilli, Jaynes):**

  After applying Bayes therorem $P(\mu \mid X_{av})$ describes the distribution of our degree of belief about the value of μ given the information at hand, i.e. the observed data.

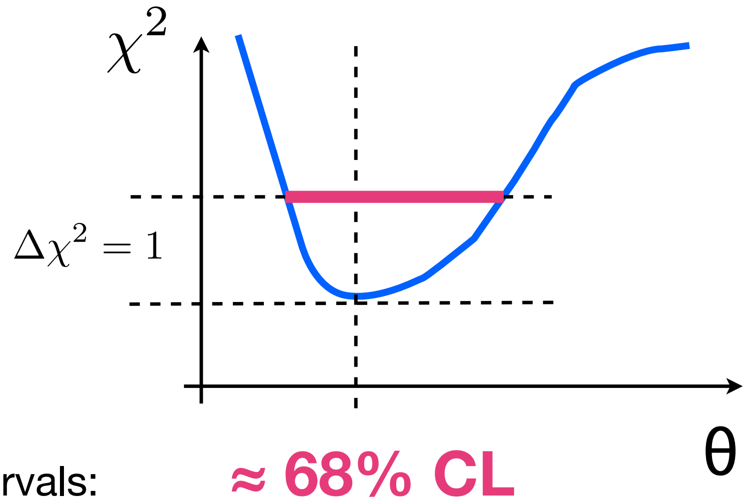
- Inference is conditional only on the observed values of the data.
- There is no concept of repetition of the experiment.

Favoured regions: likelihood-based approach

Imperial College London

- 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) = chi-squared
    - Markov Chain Monte Carlo (MCMC)
    - MCMC and Minuit as “afterburner”
    - Simulated annealing
    - Genetic algorithm



≈ 68% CL

- Determine approximate confidence intervals: Local Δ(chi-squared) method

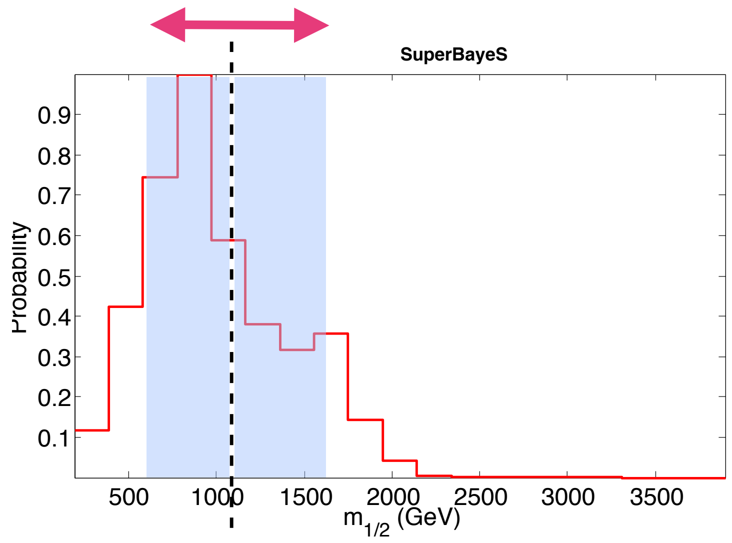
Roberto Trotta

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

Marginalization vs profiling (maximising)

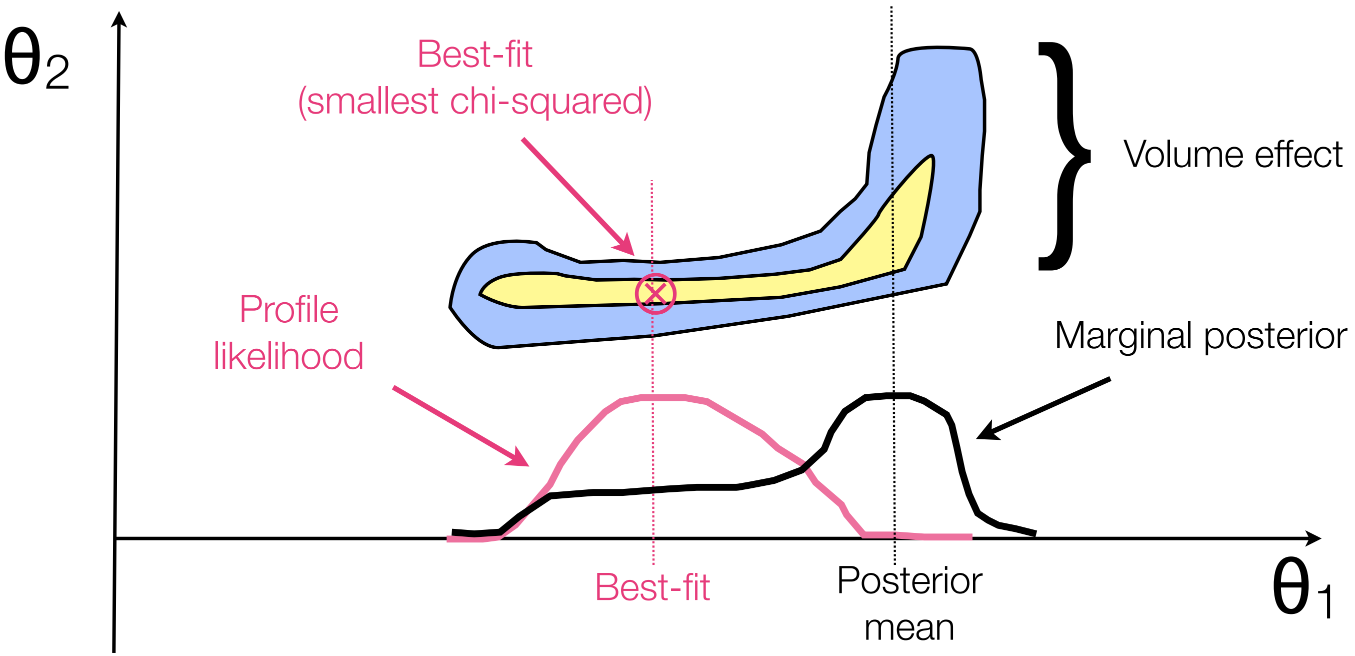
Imperial College London

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)

Roberto Trotta

Marginalization vs profiling (maximising)

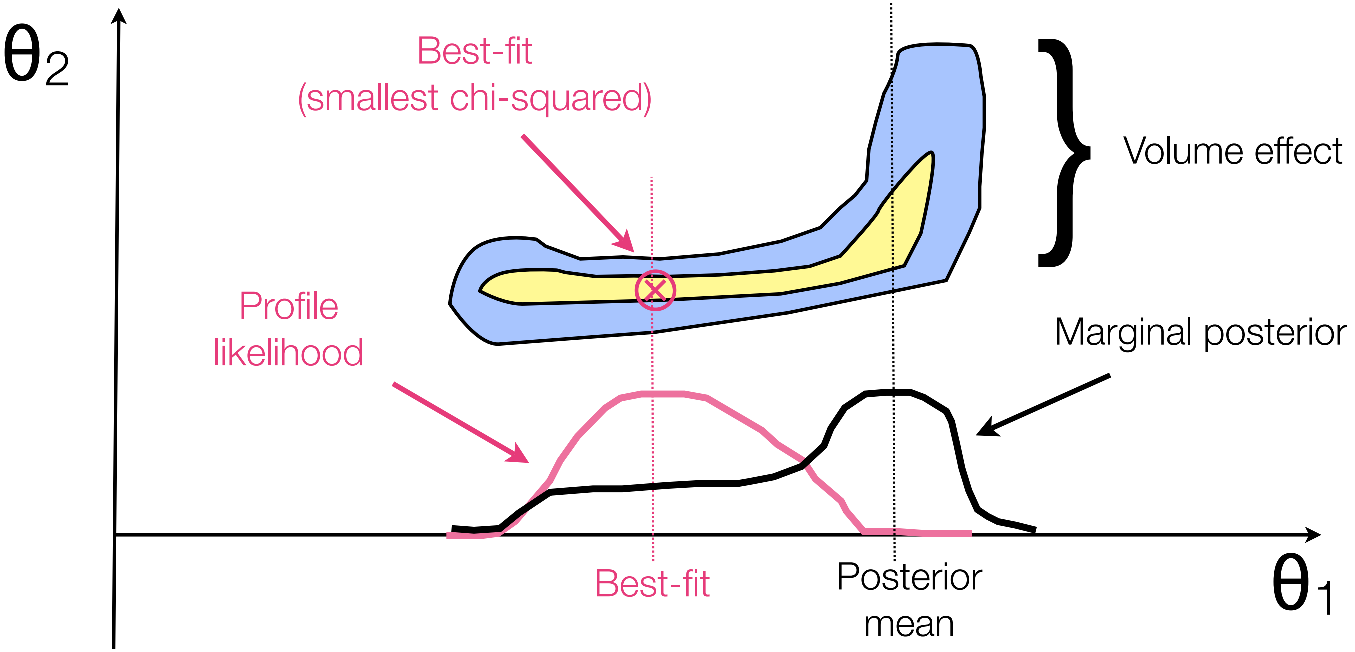
Imperial College London

Physical analogy: (thanks to Tom Loredo)

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

Likelihood = hottest hypothesis
Posterior = hypothesis with most heat

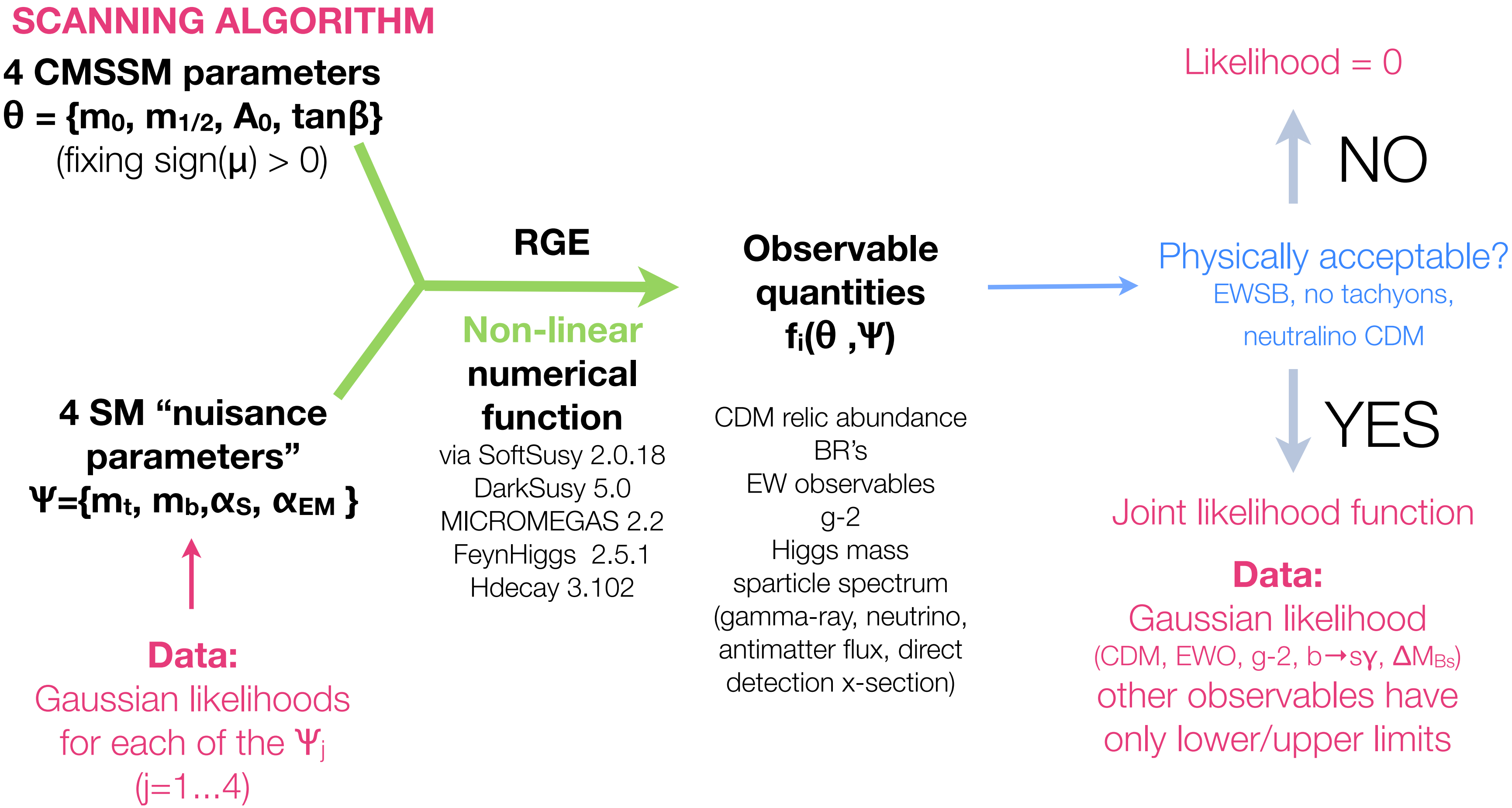
Posterior: $P \propto \int p(\theta) L(\theta) d\theta$



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

Roberto Trotta

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

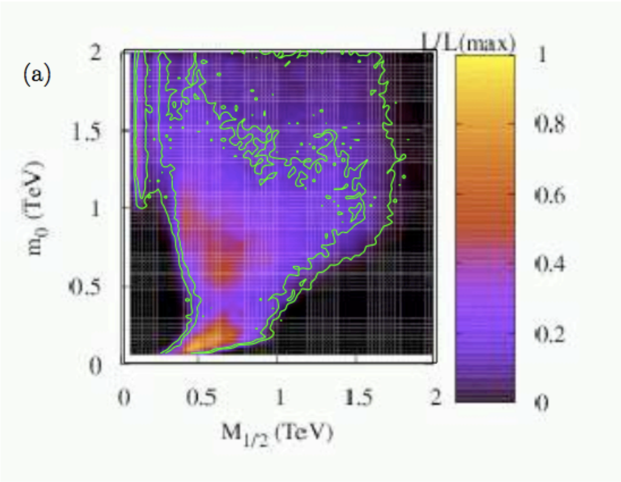
YES

Joint likelihood function

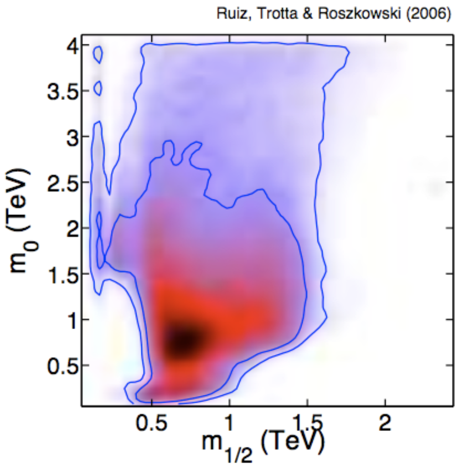
Data:
Gaussian likelihood
(CDM, EWO, g-2, $b\rightarrow s\gamma$, ΔM_{Bs})
other observables have only lower/upper limits

Global CMSSM scans

- Bayesian approach led by two groups (early work by Baltz & Gondolo, 2004):
- Ben Allanach (DAMPT) and collaborators (Allanach & Lester, 2006 onwards)
- Ruiz de Austri, Roszkowski & RT (Ruiz de Austri et al, 2006 onwards)
  + Feroz & Hobson (MultiNest), + Silk (indirect detection), + Strigari (direct detection), + Martinez et al (dwarfs), + de los Heros (IceCube), + Bertone et al (pMSSM), + Cranmer (LHC coverage)
  SuperBayeS public code available from: superbayes.org



Allanach & Lester (2006)



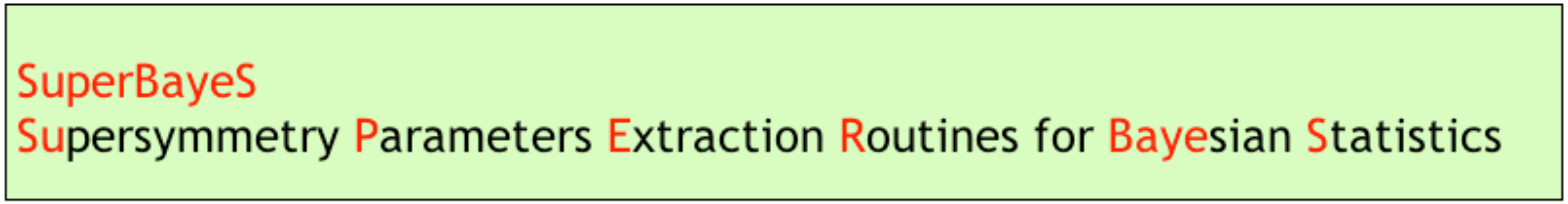
Ruiz de Austri, Roszkowski & RT (2006)

See also: Ellis et al (2004 onwards), Buchmuller et al (2008, 2009), Scott et al (2009), Akrami et al (2009)

Key advantages of the Bayesian approach

- **Efficiency:** computational effort scales ~ N rather than k^N as in grid-scanning methods. Orders of magnitude improvement over grid-scanning.
- **Marginalisation:** integration over hidden dimensions comes for free.
- **Inclusion of nuisance parameters:** simply include them in the scan and marginalise over them.
- **Pdf's for derived quantities**: probabilities distributions can be derived for any function of the input variables (crucial for DD/ID/LHC predictions)

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 aka MultiNest (Feroz & Hobson 2008)
  A full 8D scan now takes less than 2 days on 8 CPUs.

The future: “instantaneous” inference with neural networks

Imperial College London

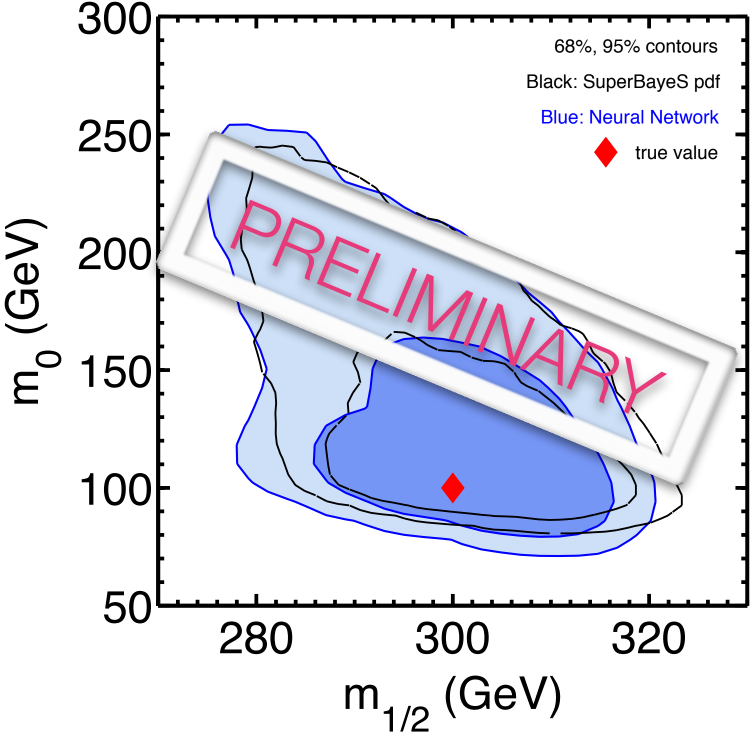
- **Standard MCMC**
  (SuperBayeS v1.23, 2006)
  720 CPU days

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

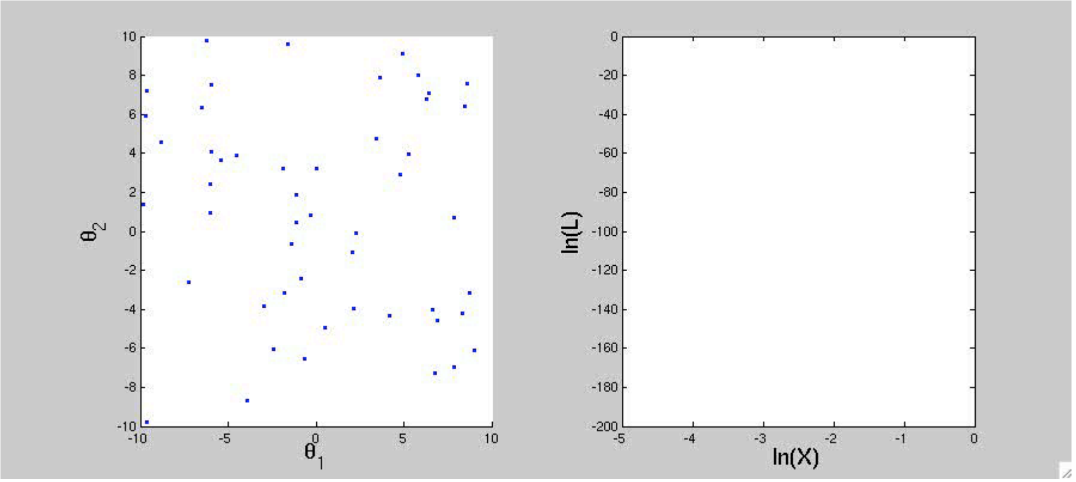
- **SuperBayeS+Neural Networks**
  (Bridges, Cranmer, Feroz, Hobson, Ruiz & RT, in prep)
  15 CPU minutes
  speed-up factor: 70’000

Simulated LHC data

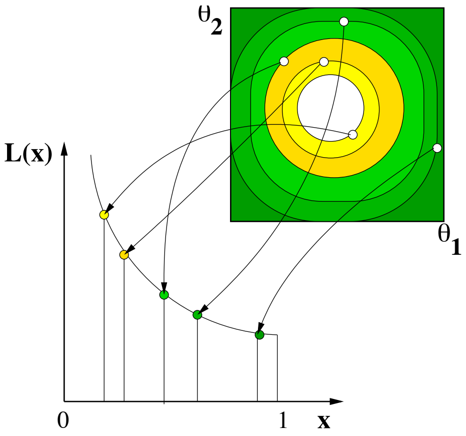
Bridges et al (2009)



Nested sampling



(animation courtesy of David Parkinson)



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

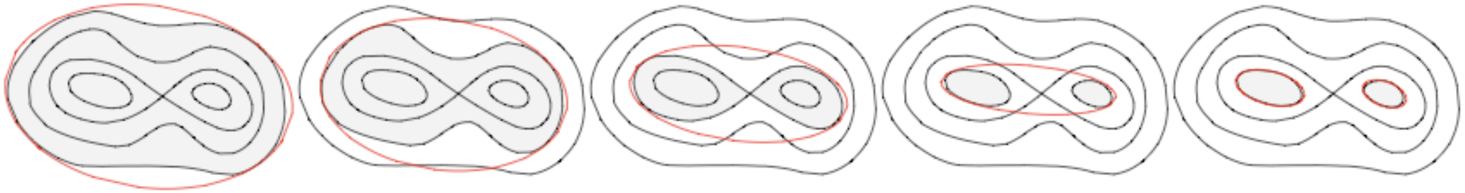
$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta) d\theta$$

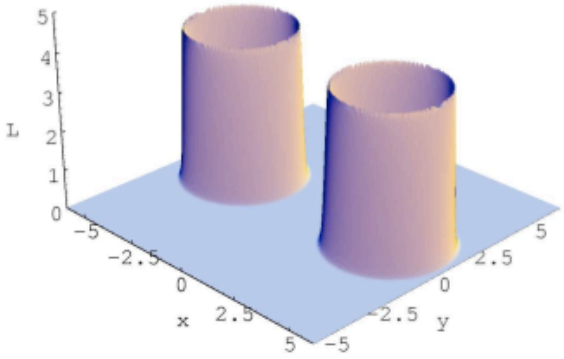
$$P(d) = \int d\theta \mathcal{L}(\theta) P(\theta) = \int_0^1 X(\lambda) d\lambda$$

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

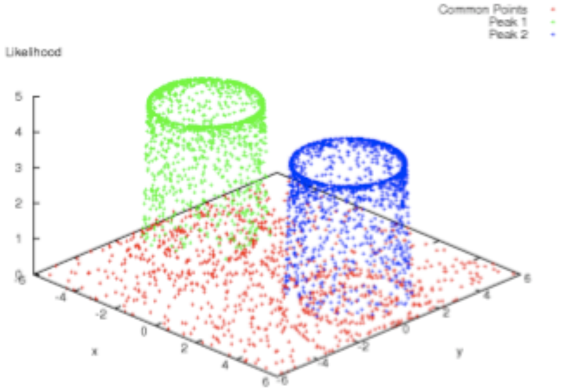
The MultiNest algorithm

- MultiNest: 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: likelihood-based results

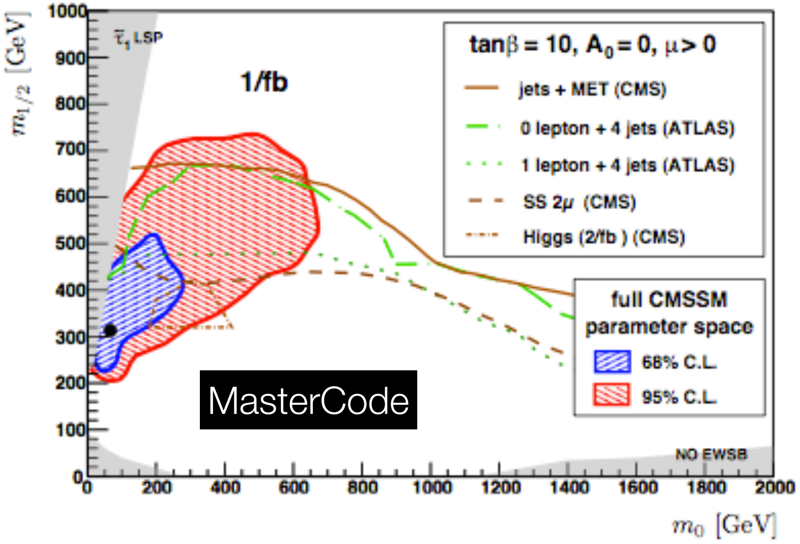
0907.4468 [hep-ph]
0808.4128 [hep-ph]

Best fit points (μ>0)

0907.2589 [hep-ph]

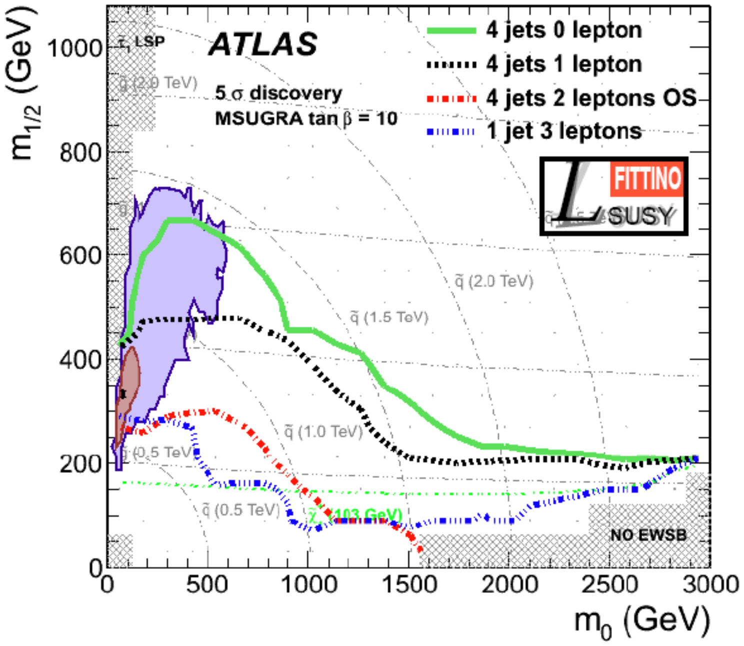
MasterCode

M_0=60 , $M_{1/2}$=310 A_0=130, tan**β**=11



Fittino

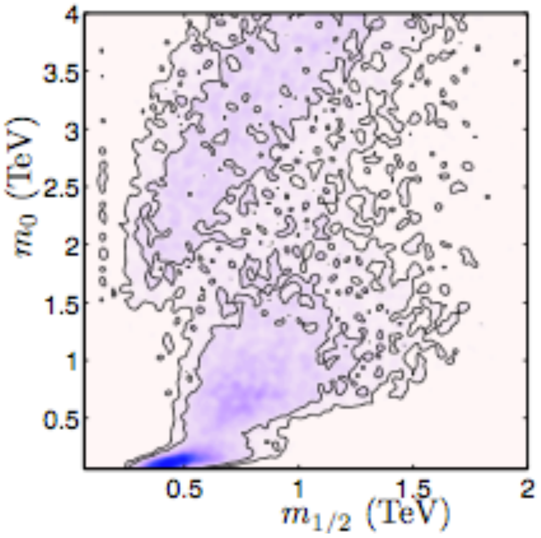
M_0=76 , $M_{1/2}$=332 A_0=383, tan**β**=13



CMSSM today: Bayesian results

“flat prior”

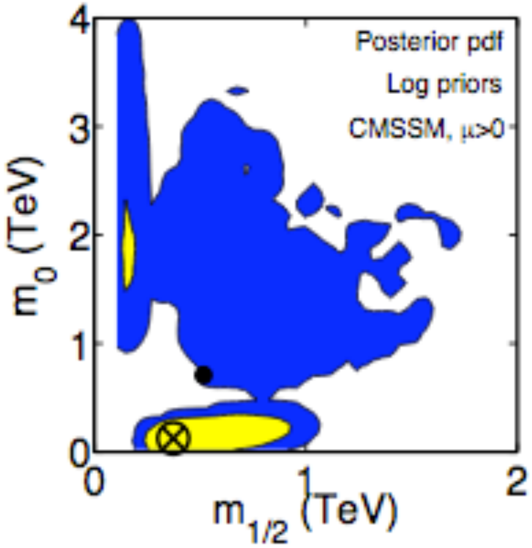
Uniform in $M_0, M_{1/2}, A_0, \tan\beta$



0807.4512 [hep-ph]

“log prior”

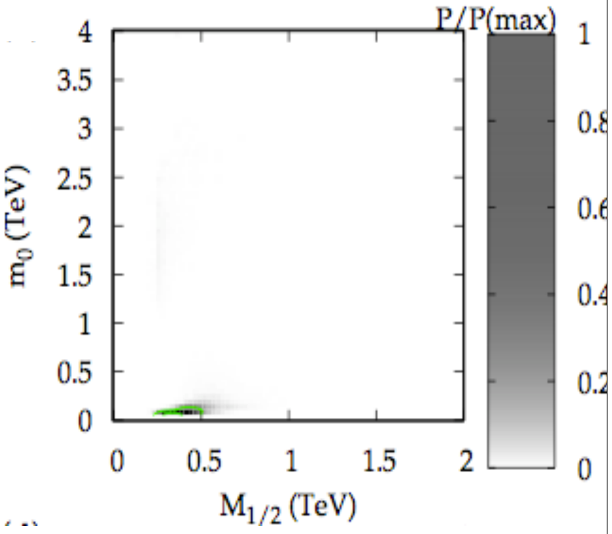
Uniform in $\log(M_0)$, $\log(M_{1/2})$, A_0, $\tan\beta$



0809.3792 [hep-ph]

“naturalness prior”

Penalizes regions of parameter space that are “fine tuned”

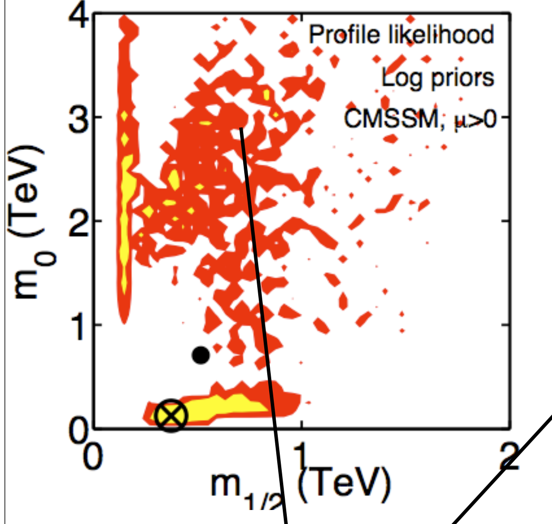


0705.0487 [hep-ph]

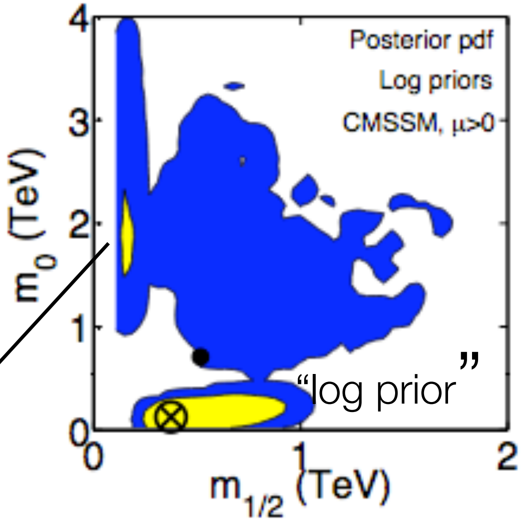
Posterior distributions

CMSSM today: Frequentist vs Bayesian

SuperBayeS: profile likelihood

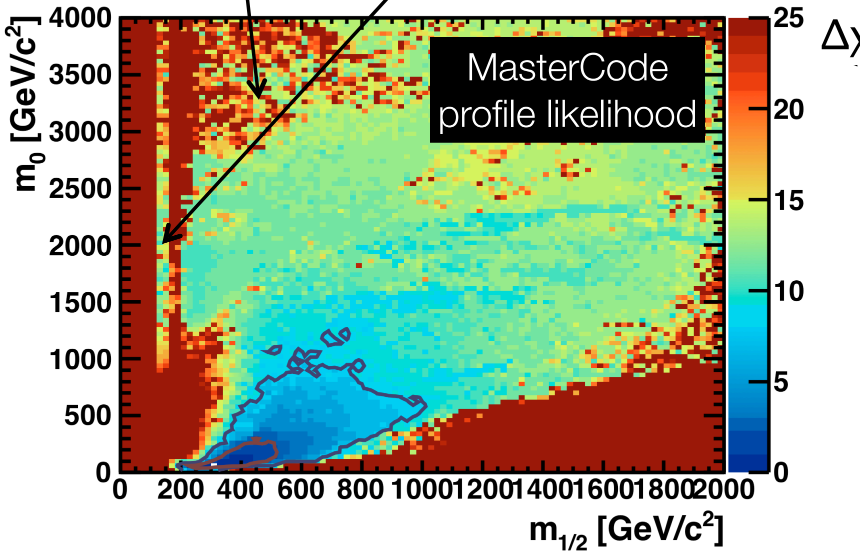


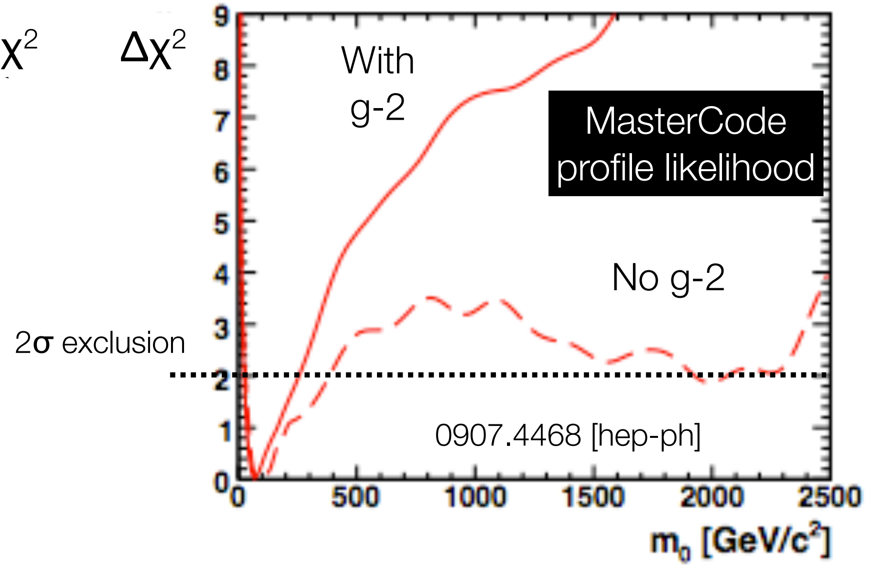
SuperBayeS: posterior



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

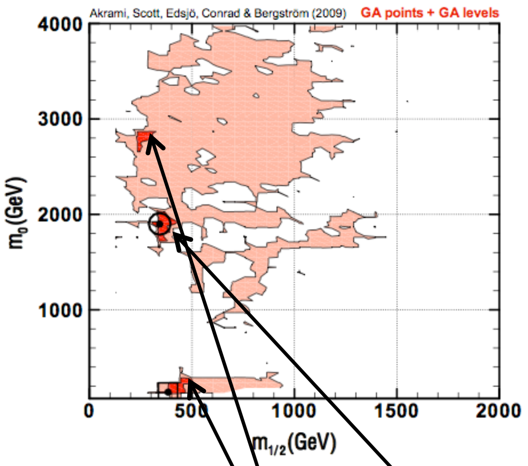




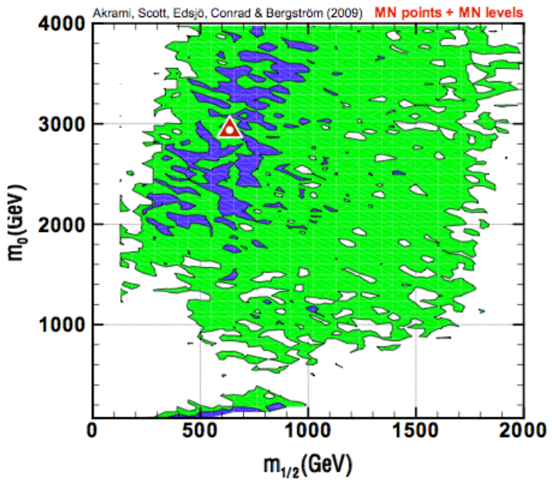
Profile likelihood results: comparison

- Akram 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 high-mass region: **is this something other frequentist fits might have missed?**

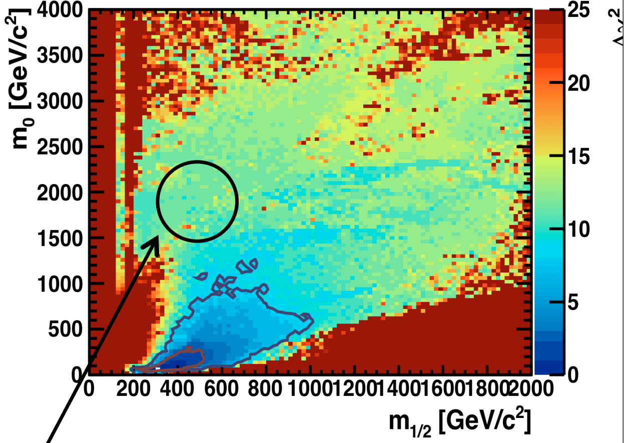
Genetic Algorithm profile likelihood



MultiNest profile likelihood



MasterCode profile likelihood



isolated local maxima

overall best-fit

excluded at ~ 3σ

Caveat: looking for best-fits with MCMC

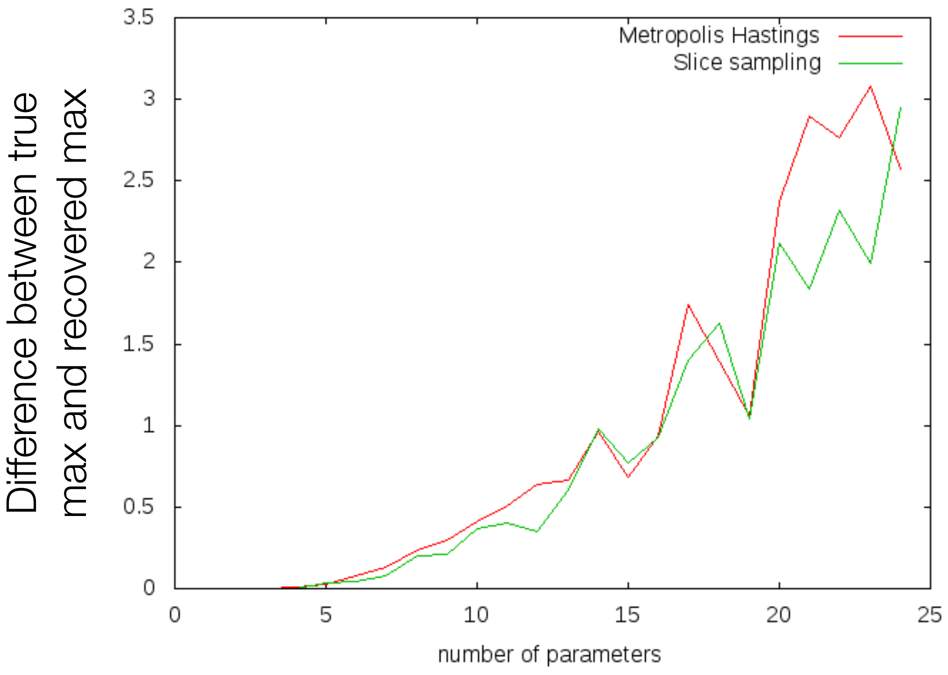
Imperial College London

- MCMC is not geared towards finding the best-fit point. Rather it tries 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 has to be taken with a grain of salt

Toy multinormal likelihood



Roberto Trotta

Which approach is “best”?

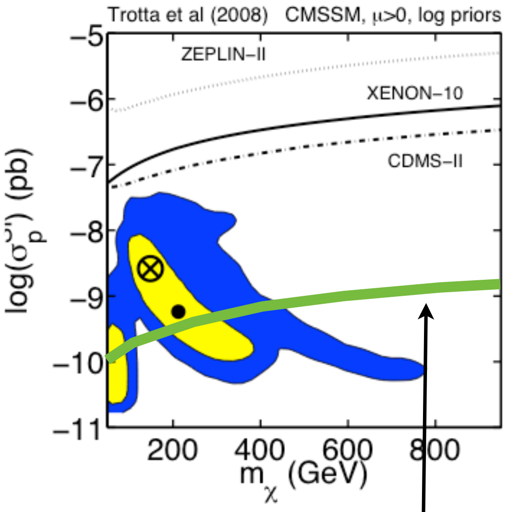
- There are a number of desiderata that any statistical approach shoud meet:
  - **Unbiasedness:** recovery of parameter values should be unbiased
  - **Reasonable efficiency:** limited computational resources mean that sometimes we have to take shortcuts to make the method work
  - **Errobars you can believe in:** here the concept of coverage is the only test I can think of
- Highly timely for the community to test the long-term performance of statistical packages agains realistically simulated data.
- I suggest to initiate a programme of increasingly complex blind reconstructions.
- In the meantime, trust only ‘robust’ results!

Direct detection prospects

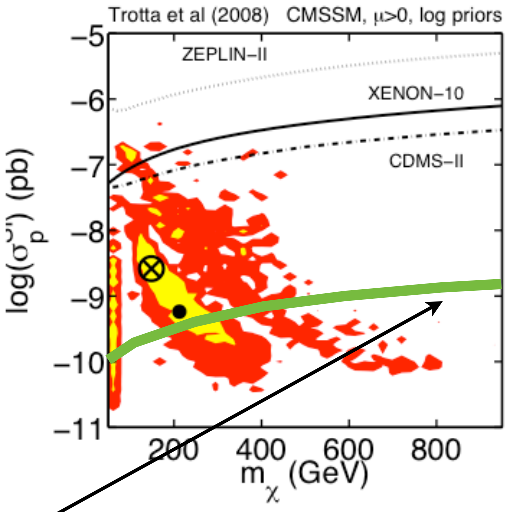
Generally favourable prospects for WIMP discovery in the CMSSM framework for upcoming detectors are **robust**, independently of the choice of statistics.

Notice: canonical local density & velocity dispersion assumed

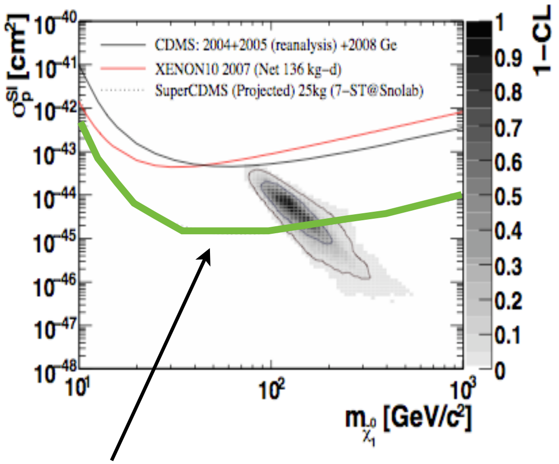
SuperBayeS posterior



SuperBayeS profile likelihood



MasterCode profile likelihood



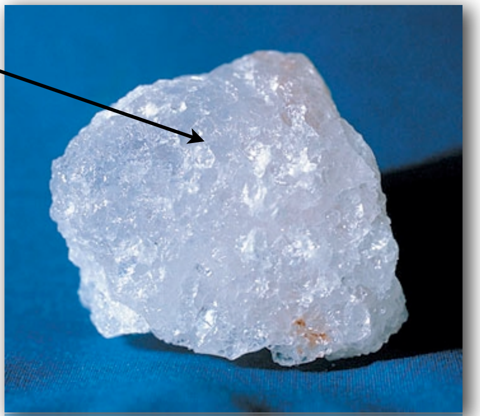
reach of 25 kg Ge

reach of 1t of Xe detector

Statistical conclusions

- Even one of the theoretically most constrained models (the CMSSM) shows signs of ambiguities in the statistical results
- This can be traced back to insufficiently constraining data (at present)
- Low-mass SUSY seems preferred but ~ TeV scale masses cannot be ruled out robustly
- **ALL ENSUING PREDICTIONS HAVE TO BE TAKEN WITH A LARGE GRAIN OF SALT**

grain of salt



ATLAS
(to scale)