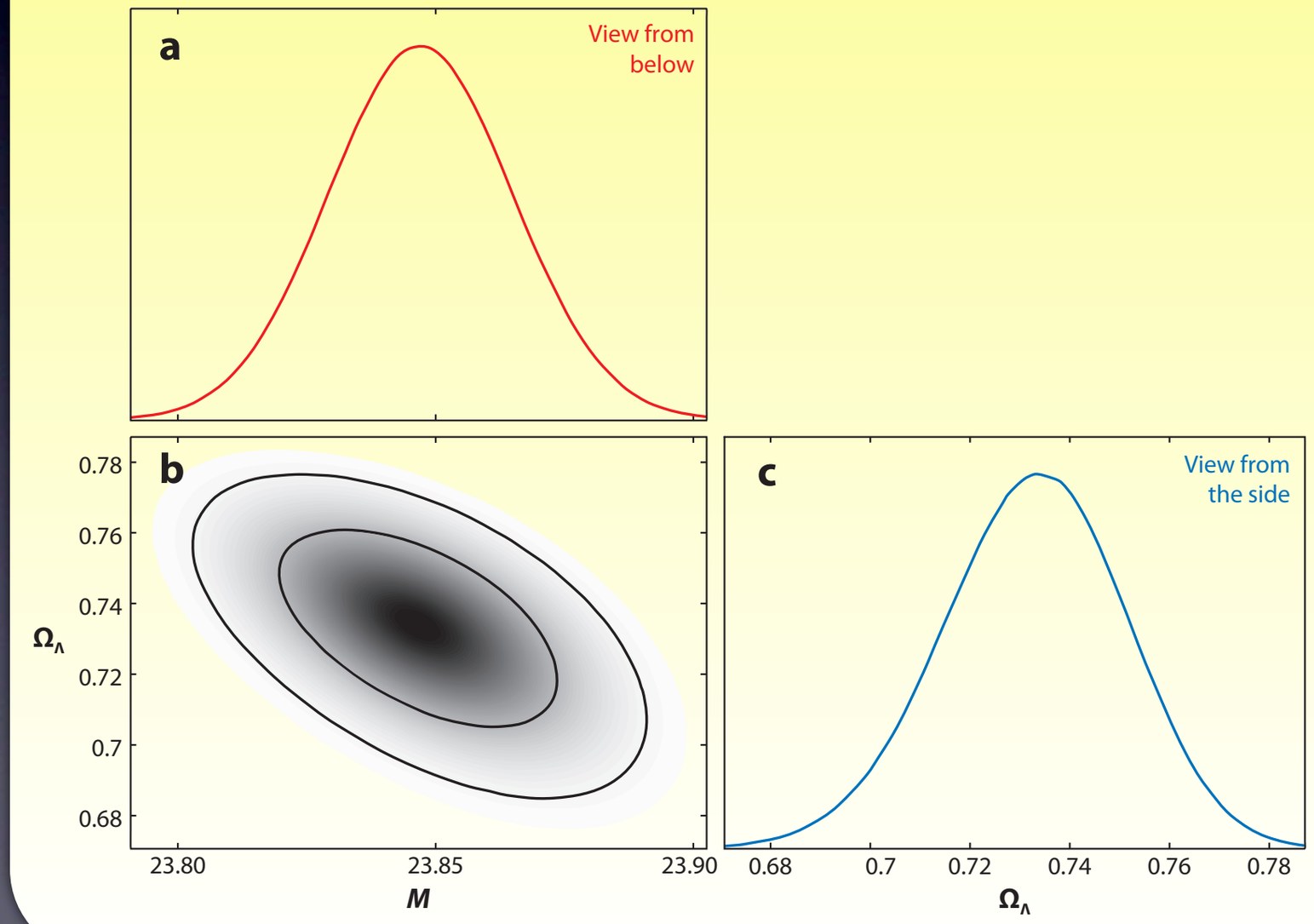
University of Sussex

Parameter Estimation



Microsoft-free
presentation

Levels of Bayesian inference

Parameter Estimation

I've decided what the correct model is.

Now I want to know what values of the parameters are consistent with the data.

I can do this using e.g. Markov Chain Monte Carlo.

Model Selection

Now I think about it, I don't actually know what the correct model is. It could be one of several.

Now I want to know what the best model is.

I can do this by computing the Bayesian Evidence. I can then do parameter estimation using the best model.

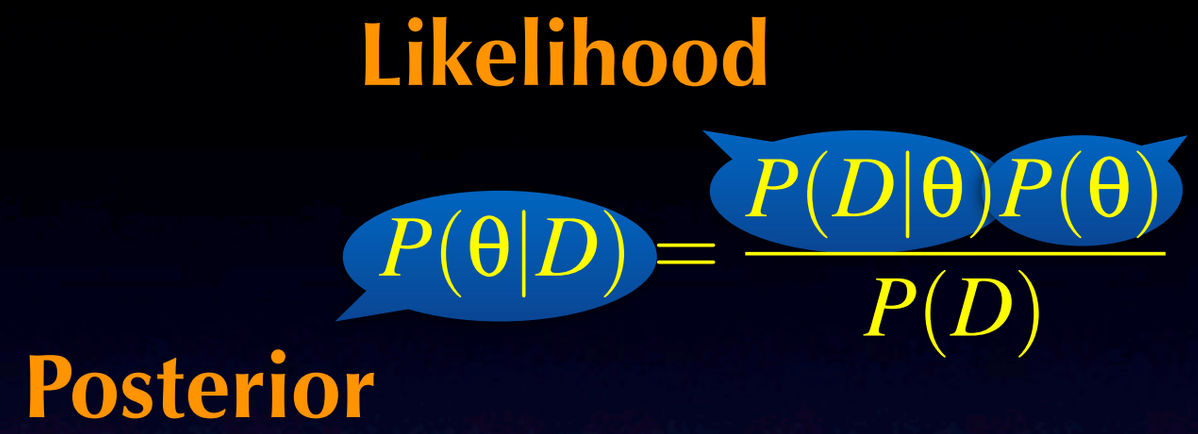
Multi-model Inference

Mmm, I did the model selection thing, but there wasn't a single best model.

But I still want to know how probable the parameter values are.

I can do this by combining the parameter likelihoods using Bayesian Model Averaging, adding them together weighted by the model probabilities.

Parameters from Bayes' Theorem



θ = parameter value
D = data

We assume that we have a dataset D, plus a model with parameter vector θ which which we can extract predictions for the data, in the form of the likelihood $L(\theta) \equiv P(D|\theta)$.

Our aim is to apply Bayes' Theorem to update our prior knowledge $P(\theta)$ to the posterior $P(\theta|D)$.

The prior knowledge must be specified, and could represent purely theoretical expectation, or the outcome from some previous experiment, or both.

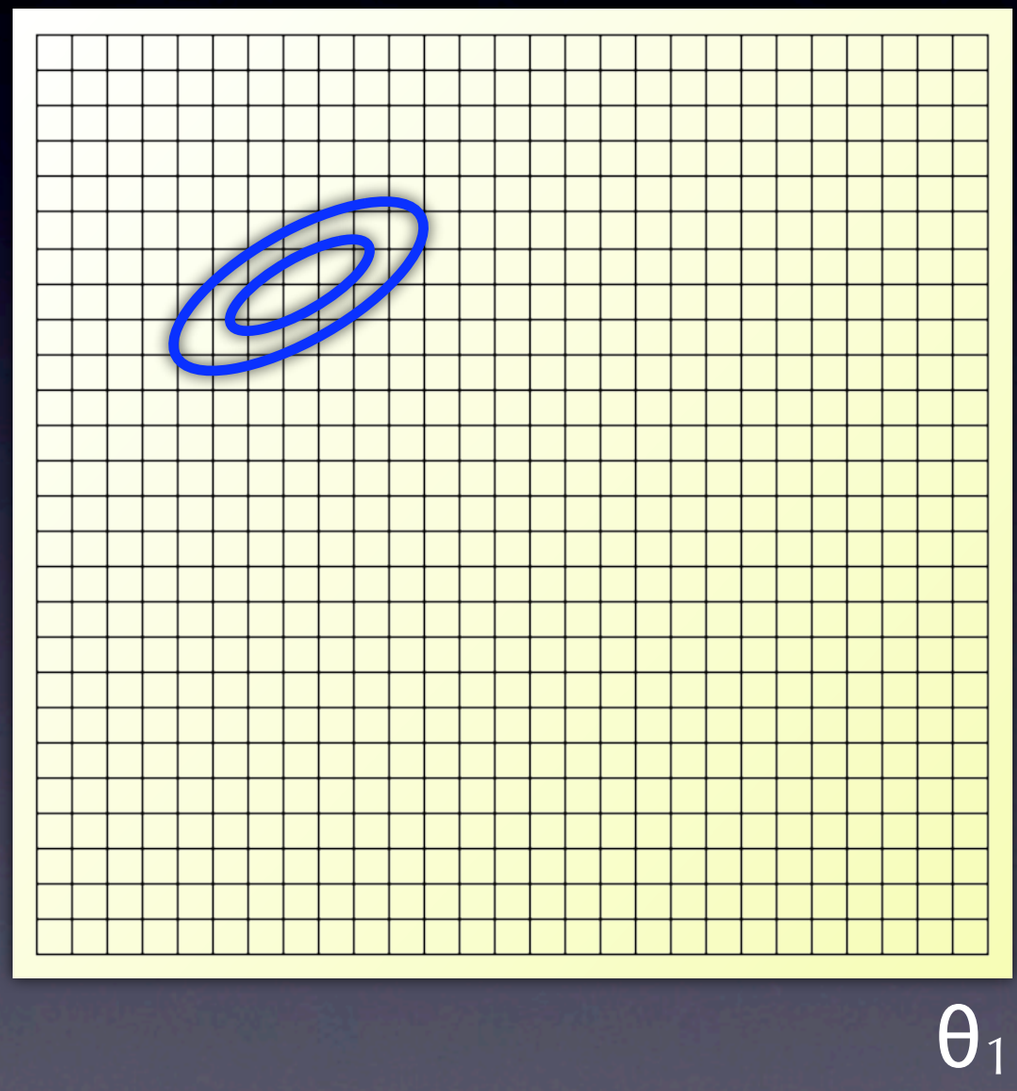
Obstacles to parameter estimation

So, we just have to find the region(s) in parameter space where the posterior probability is high. However

- The parameter space may have a high dimensionality (cosmological examples typically have 6 to 10 parameters simultaneously varying).
- Individual evaluations of the likelihood function may be computationally time-consuming (a few seconds each in typical cosmology examples, ie one CPU-month per million calculations).
- The likelihood function may be sharply peaked, at an unknown location, and it may have several maxima masquerading as the true maximum.
- There may be parameter degeneracies, where the likelihood varies only weakly, or not at all, along some direction in parameter space.

Grid evaluation

The simplest approach is evaluation of the posterior on a grid within parameter space.

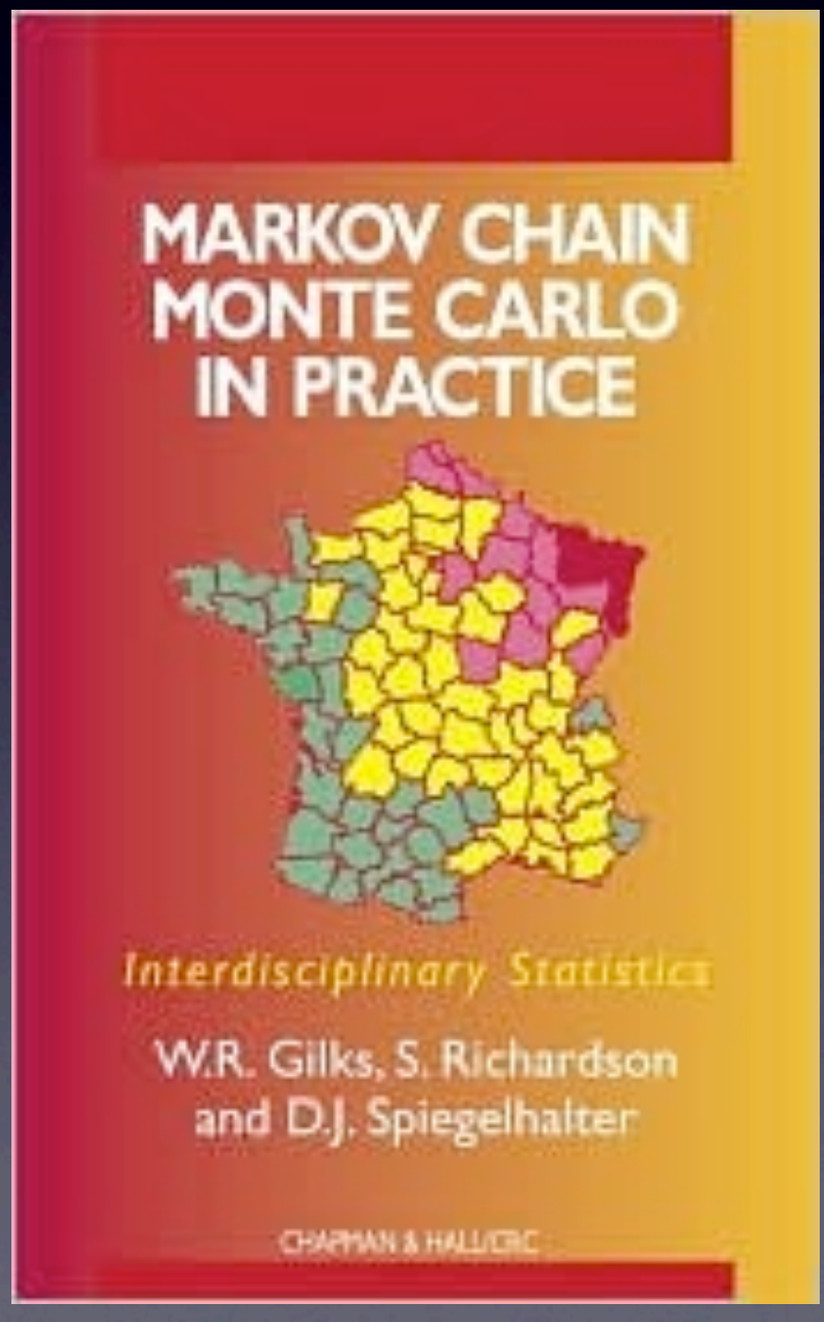


This will be fine if the dimensionality is small, but becomes inefficient if it is not, since 10 or 20 evaluations per direction is the almost certainly necessary.

In 6 dimensions at 10 evaluations per dimension, and a few seconds per evaluation, we need a CPU-month to get our result.

Instead ... Monte Carlo methods

Instead of grid-based methods, Monte Carlo is the method of choice for parameter estimation, particularly Markov Chain Monte Carlo.

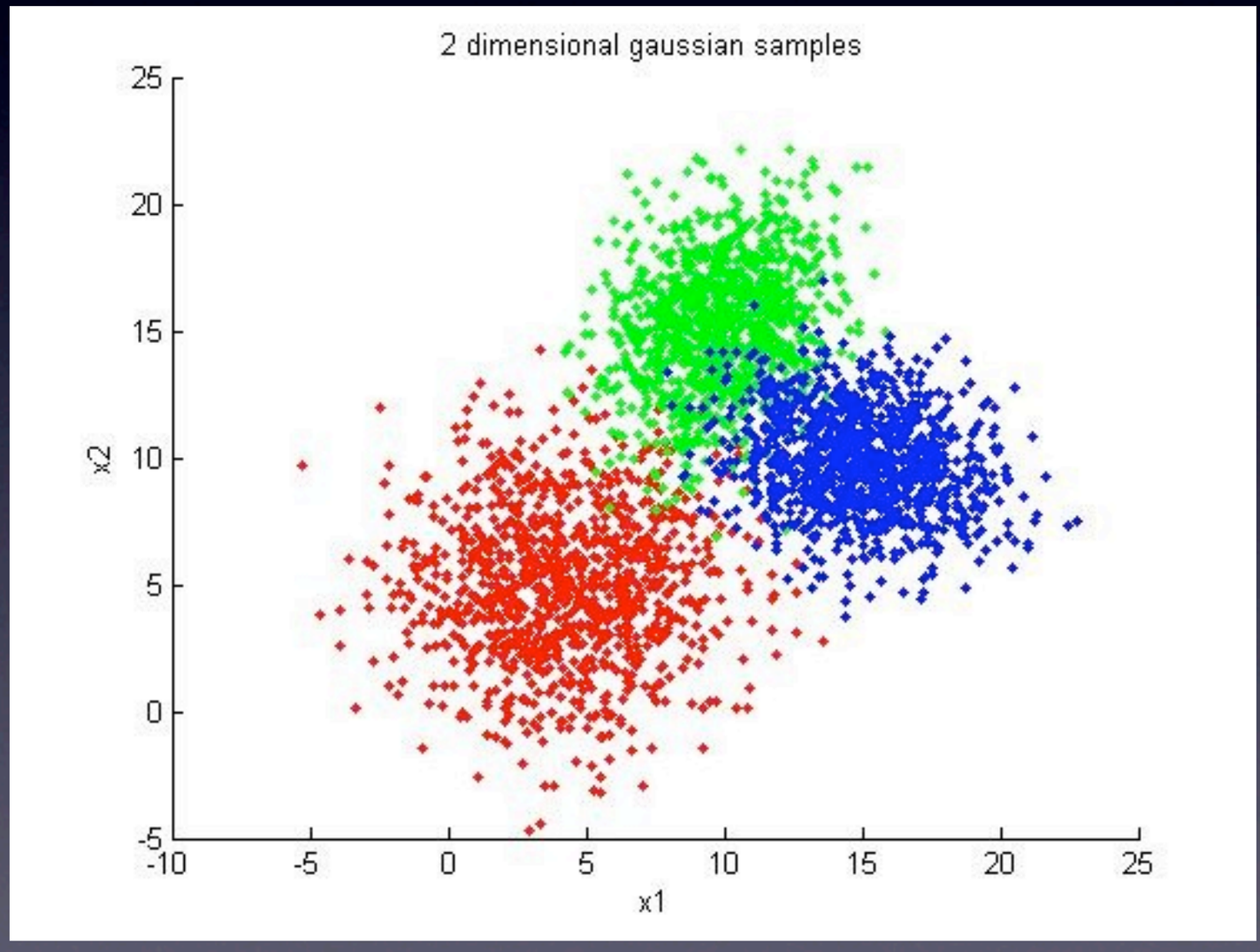


Monte Carlo methods

Rather than attempting to map out the shape of the posterior probability distribution directly, we instead seek to obtain a large set of points sampled from the distribution.

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results.

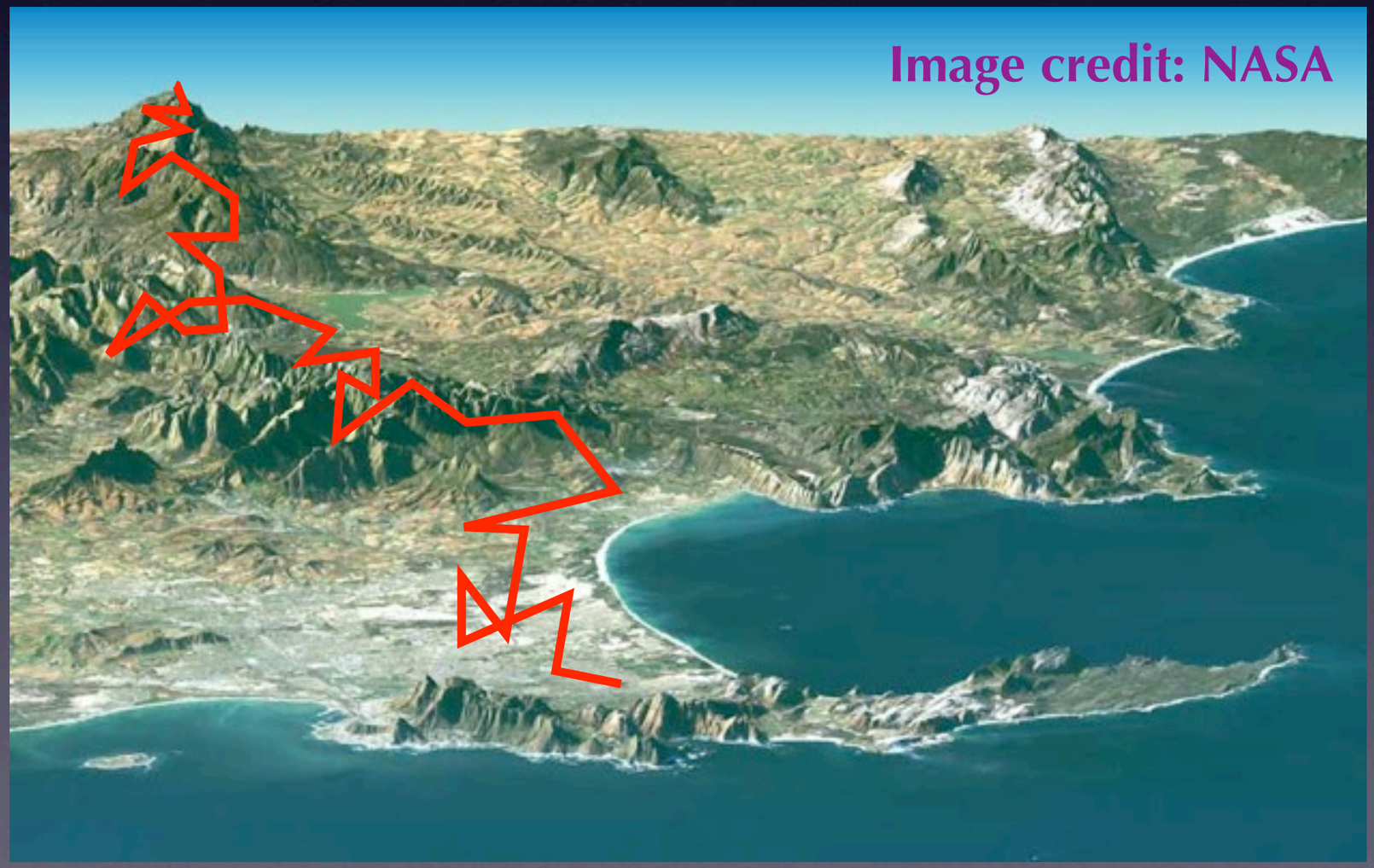
Wikipedia



Markov Chain Monte Carlo (MCMC) methods

Markov Chain Monte Carlo methods are a subset of Monte Carlo methods which rely on constructing a Markov chain.

Markov Chain:
A sequence of random variables in which the distribution of each random variable depends only on the value of its predecessor.



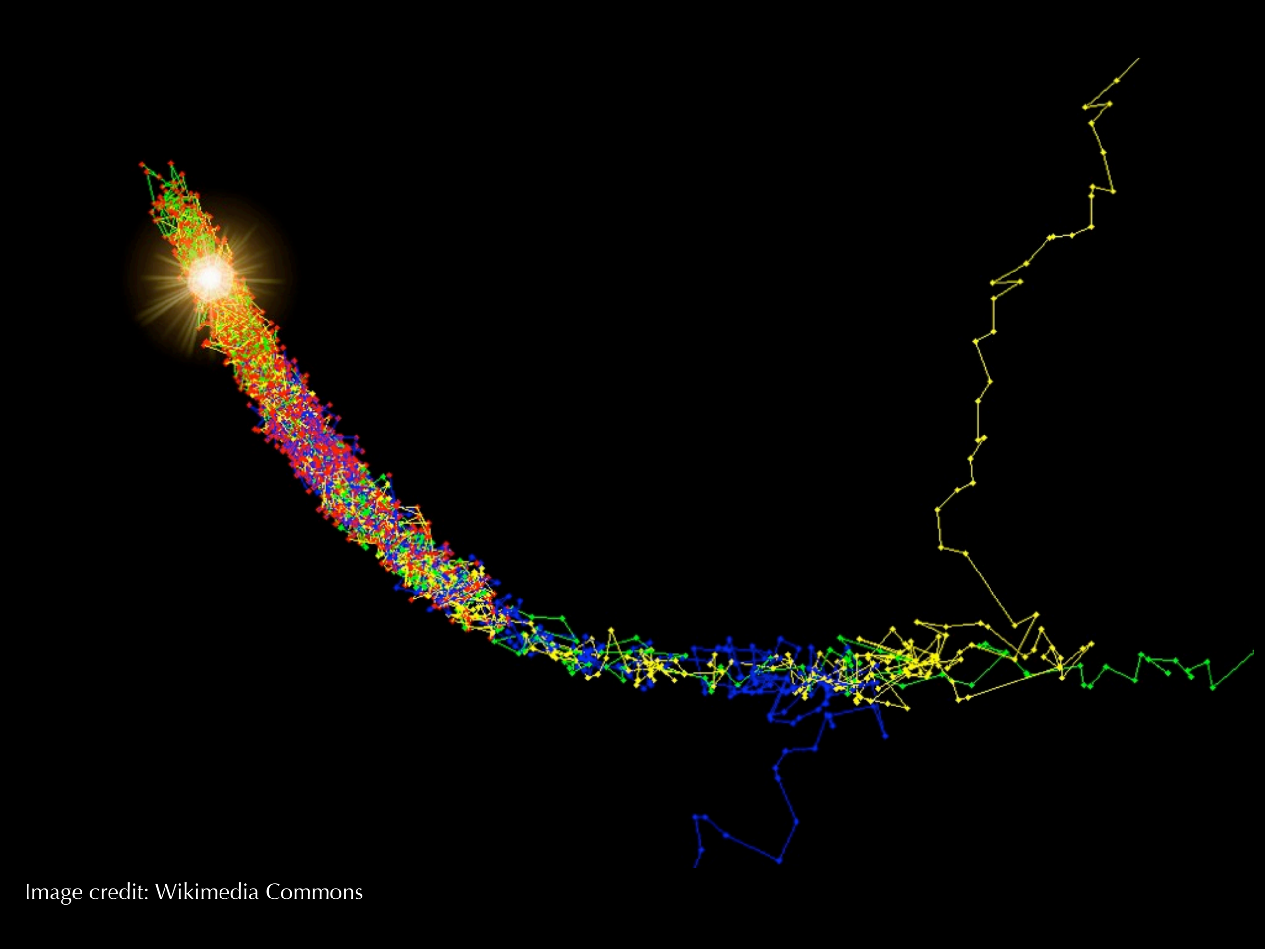
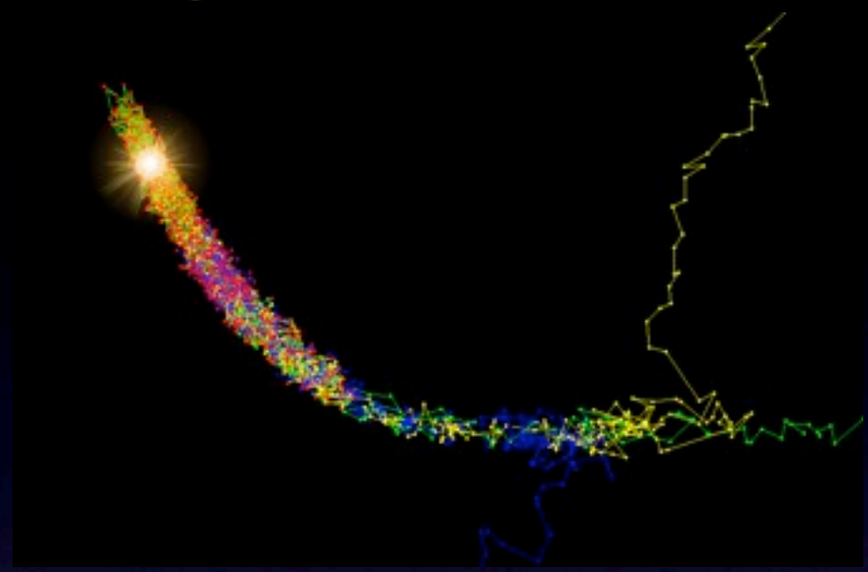


Image credit: Wikimedia Commons

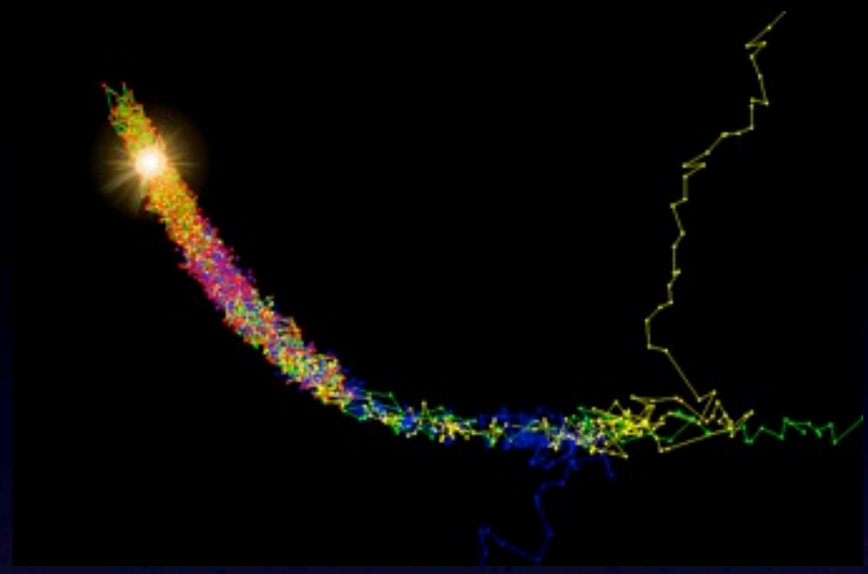
The Metropolis-Hastings algorithm

The simplest, and often most effective, algorithm for constructing a Markov Chain is the Metropolis-Hastings algorithm.



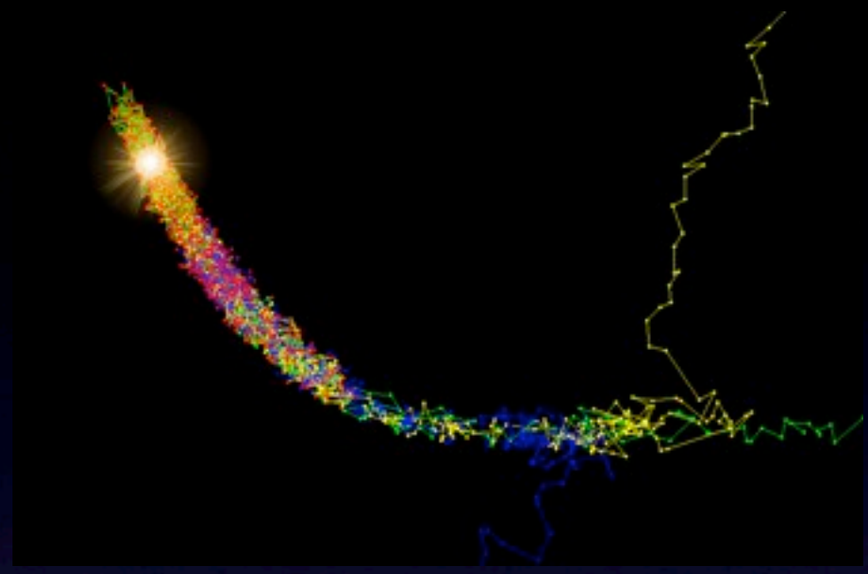
1) Choose a starting point in parameter space θ_{old}, and compute its posterior probability P_{old}.
2) Make a random jump to a new point θ_{new}, and compute its posterior P_{new}.
3) If $P_{new} > P_{old}$, then accept the jump. Otherwise, accept the jump with probability P_{new}/P_{old}. Add the new point to the chain if accepted, otherwise add a duplicate of the old point.
4) Go back to step 2 and repeat until satisfied.

The Metropolis-Hastings algorithm



The Metropolis-Hastings algorithm is designed so that the points in the chain correspond to samples from the posterior probability distribution.

The Metropolis-Hastings algorithm



Considerations:

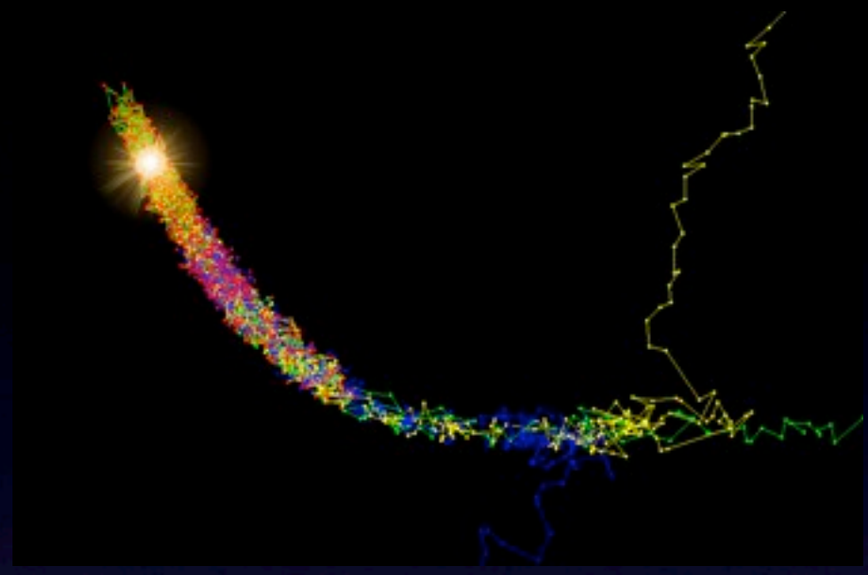
1) How to jump?

   This is done by a proposal function, from which a jump is randomly selected. Eg it might be a Gaussian centred at the current point. A proposal function must satisfy detailed balance, meaning we are as likely to jump from point A to B as from point B to A.

2) How to tune the proposal function?

   Jumps should be neither too big (too hard to find the maximum) or too small (takes too long to reach the maximum). A test run may be used to optimize the size and orientation of the proposal function.

The Metropolis-Hastings algorithm



Considerations:

3) When to start?

The initial phase of finding the high-likelihood region depends on the start point. This phase, the burn-in, should be discarded as unrepresentative of the posterior.

4) When to stop?

Convergence tests on the chain can be carried out, eg by splitting into separate sections and comparing. The Gelman-Rubin test is a popular such diagnostic. It is usually a good idea to run several chains from different start points (eg on a parallel computing platform) in case one gets hung up in a false maximum.

A bit of a chain (from WMAP)

```
0.2000000E+01  0.5851608E+04  0.2015806E-01  0.1056880E+00  0.1052353E+01  0.1572130E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9654508E+00  0.0000000E+00  0.0000000E+00  0.3046525E+01  0.0000000E+00  0.0000000E+00  0.7791765E+00  0.1357033E+02  0.2208235E+00  0.7946533E+00  0.1775950E+02  0.0000000E+00  0.7549133E+02
0.1000000E+01  0.5847794E+04  0.2015806E-01  0.1056880E+00  0.1052353E+01  0.1572130E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9742470E+00  0.0000000E+00  0.0000000E+00  0.3056430E+01  0.0000000E+00  0.0000000E+00  0.7791765E+00  0.1357033E+02  0.2208235E+00  0.8014988E+00  0.1775950E+02  0.0000000E+00  0.7549133E+02
0.1000000E+01  0.5843263E+04  0.2010433E-01  0.1014225E+00  0.1047609E+01  0.1655797E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9702829E+00  0.0000000E+00  0.0000000E+00  0.3045888E+01  0.0000000E+00  0.0000000E+00  0.7877861E+00  0.1368099E+02  0.2122139E+00  0.7744952E+00  0.1822450E+02  0.0000000E+00  0.7567444E+02
0.1000000E+01  0.5827815E+04  0.2010433E-01  0.1014225E+00  0.1047609E+01  0.1655797E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9612351E+00  0.0000000E+00  0.0000000E+00  0.3043113E+01  0.0000000E+00  0.0000000E+00  0.7877861E+00  0.1368099E+02  0.2122139E+00  0.7705776E+00  0.1822450E+02  0.0000000E+00  0.7567444E+02
0.2000000E+01  0.5827671E+04  0.2005007E-01  0.9967086E-01  0.1047475E+01  0.1679666E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9610596E+00  0.0000000E+00  0.0000000E+00  0.3040069E+01  0.0000000E+00  0.0000000E+00  0.7947711E+00  0.1367524E+02  0.2052289E+00  0.7615068E+00  0.1835000E+02  0.0000000E+00  0.7637756E+02
0.1000000E+01  0.5823237E+04  0.2062331E-01  0.9755769E-01  0.1048439E+01  0.1589623E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9725500E+00  0.0000000E+00  0.0000000E+00  0.3020521E+01  0.0000000E+00  0.0000000E+00  0.8064781E+00  0.1357294E+02  0.1935219E+00  0.7447448E+00  0.1726400E+02  0.0000000E+00  0.7814636E+02
0.1000000E+01  0.5810297E+04  0.2062331E-01  0.9755769E-01  0.1048439E+01  0.1589623E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9778811E+00  0.0000000E+00  0.0000000E+00  0.3030873E+01  0.0000000E+00  0.0000000E+00  0.8064781E+00  0.1357294E+02  0.1935219E+00  0.7502713E+00  0.1726400E+02  0.0000000E+00  0.7814636E+02
0.1000000E+01  0.5804656E+04  0.2062801E-01  0.8928808E-01  0.1047195E+01  0.1529961E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9788681E+00  0.0000000E+00  0.0000000E+00  0.2981949E+01  0.0000000E+00  0.0000000E+00  0.8358389E+00  0.1352269E+02  0.1641611E+00  0.6901900E+00  0.1643250E+02  0.0000000E+00  0.8182678E+02
0.1000000E+01  0.5749039E+04  0.2062801E-01  0.8928808E-01  0.1047195E+01  0.1529961E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9413199E+00  0.0000000E+00  0.0000000E+00  0.2973182E+01  0.0000000E+00  0.0000000E+00  0.8358389E+00  0.1352269E+02  0.1641611E+00  0.6764783E+00  0.1643250E+02  0.0000000E+00  0.8182678E+02
0.2000000E+01  0.5749209E+04  0.2061104E-01  0.8939423E-01  0.1047172E+01  0.1515630E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9407399E+00  0.0000000E+00  0.0000000E+00  0.2970405E+01  0.0000000E+00  0.0000000E+00  0.8353964E+00  0.1352616E+02  0.1646036E+00  0.6760311E+00  0.1634000E+02  0.0000000E+00  0.8174988E+02
0.1000000E+01  0.5748409E+04  0.2072671E-01  0.8786918E-01  0.1048379E+01  0.1491505E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9446303E+00  0.0000000E+00  0.0000000E+00  0.2962835E+01  0.0000000E+00  0.0000000E+00  0.8426777E+00  0.1346236E+02  0.1573223E+00  0.6660739E+00  0.1603800E+02  0.0000000E+00  0.8308289E+02
0.3000000E+01  0.5710965E+04  0.2072671E-01  0.8786918E-01  0.1048379E+01  0.1491505E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9447833E+00  0.0000000E+00  0.0000000E+00  0.2983899E+01  0.0000000E+00  0.0000000E+00  0.8426777E+00  0.1346236E+02  0.1573223E+00  0.6731693E+00  0.1603800E+02  0.0000000E+00  0.8308289E+02
0.1000000E+01  0.5702064E+04  0.2162730E-01  0.7827552E-01  0.1049630E+01  0.1234248E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9633905E+00  0.0000000E+00  0.0000000E+00  0.2902750E+01  0.0000000E+00  0.0000000E+00  0.8760756E+00  0.1321413E+02  0.1239244E+00  0.5933253E+00  0.1335250E+02  0.0000000E+00  0.8978638E+02
0.1000000E+01  0.5686054E+04  0.2162730E-01  0.7827552E-01  0.1049630E+01  0.1234248E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9805061E+00  0.0000000E+00  0.0000000E+00  0.2928784E+01  0.0000000E+00  0.0000000E+00  0.8760756E+00  0.1321413E+02  0.1239244E+00  0.6056091E+00  0.1335250E+02  0.0000000E+00  0.8978638E+02
0.2000000E+01  0.5680245E+04  0.2151910E-01  0.7507140E-01  0.1046657E+01  0.1286617E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9768334E+00  0.0000000E+00  0.0000000E+00  0.2919222E+01  0.0000000E+00  0.0000000E+00  0.8824201E+00  0.1325722E+02  0.1175799E+00  0.5828394E+00  0.1363850E+02  0.0000000E+00  0.9063599E+02
0.7000000E+01  0.5676884E+04  0.2012698E-01  0.6586596E-01  0.1044043E+01  0.9536330E-01  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9467331E+00  0.0000000E+00  0.0000000E+00  0.2792258E+01  0.0000000E+00  0.0000000E+00  0.9052024E+00  0.1327366E+02  0.9479761E-01  0.4926153E+00  0.1125100E+02  0.0000000E+00  0.9524292E+02
0.1000000E+01  0.5670625E+04  0.2012698E-01  0.6586596E-01  0.1044043E+01  0.9536330E-01  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9425158E+00  0.0000000E+00  0.0000000E+00  0.2797389E+01  0.0000000E+00  0.0000000E+00  0.9052024E+00  0.1327366E+02  0.9479761E-01  0.4929855E+00  0.1125100E+02  0.0000000E+00  0.9524292E+02
0.3000000E+01  0.5669518E+04  0.2020259E-01  0.6691369E-01  0.1043785E+01  0.1044381E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9444320E+00  0.0000000E+00  0.0000000E+00  0.2820847E+01  0.0000000E+00  0.0000000E+00  0.9022123E+00  0.1329783E+02  0.9778772E-01  0.5054021E+00  0.1197700E+02  0.0000000E+00  0.9438599E+02
0.2000000E+01  0.5669907E+04  0.2020259E-01  0.6691369E-01  0.1043785E+01  0.1044381E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9403662E+00  0.0000000E+00  0.0000000E+00  0.2818038E+01  0.0000000E+00  0.0000000E+00  0.9022123E+00  0.1329783E+02  0.9778772E-01  0.5038149E+00  0.1197700E+02  0.0000000E+00  0.9438599E+02
0.1000000E+01  0.5656928E+04  0.2020259E-01  0.6691369E-01  0.1043785E+01  0.1044381E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9468315E+00  0.0000000E+00  0.0000000E+00  0.2849296E+01  0.0000000E+00  0.0000000E+00  0.9022123E+00  0.1329783E+02  0.9778772E-01  0.5131708E+00  0.1197700E+02  0.0000000E+00  0.9438599E+02
0.2000000E+01  0.5655742E+04  0.2015999E-01  0.6764988E-01  0.1043429E+01  0.1221102E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9476981E+00  0.0000000E+00  0.0000000E+00  0.2887990E+01  0.0000000E+00  0.0000000E+00  0.8999174E+00  0.1332733E+02  0.1000827E+00  0.5283710E+00  0.1336300E+02  0.0000000E+00  0.9366821E+02
0.2000000E+01  0.5652584E+04  0.2060524E-01  0.6537398E-01  0.1042378E+01  0.1267854E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9562625E+00  0.0000000E+00  0.0000000E+00  0.2890769E+01  0.0000000E+00  0.0000000E+00  0.9053195E+00  0.1327033E+02  0.9468052E-01  0.5137574E+00  0.1342500E+02  0.0000000E+00  0.9529419E+02
0.1000000E+01  0.5647807E+04  0.2047238E-01  0.8164196E-01  0.1041798E+01  0.8982697E-01  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9417917E+00  0.0000000E+00  0.0000000E+00  0.2877844E+01  0.0000000E+00  0.0000000E+00  0.8553504E+00  0.1359470E+02  0.1446496E+00  0.6052601E+00  0.1120800E+02  0.0000000E+00  0.8402039E+02
```

Part of a chain containing tens of thousands of elements.

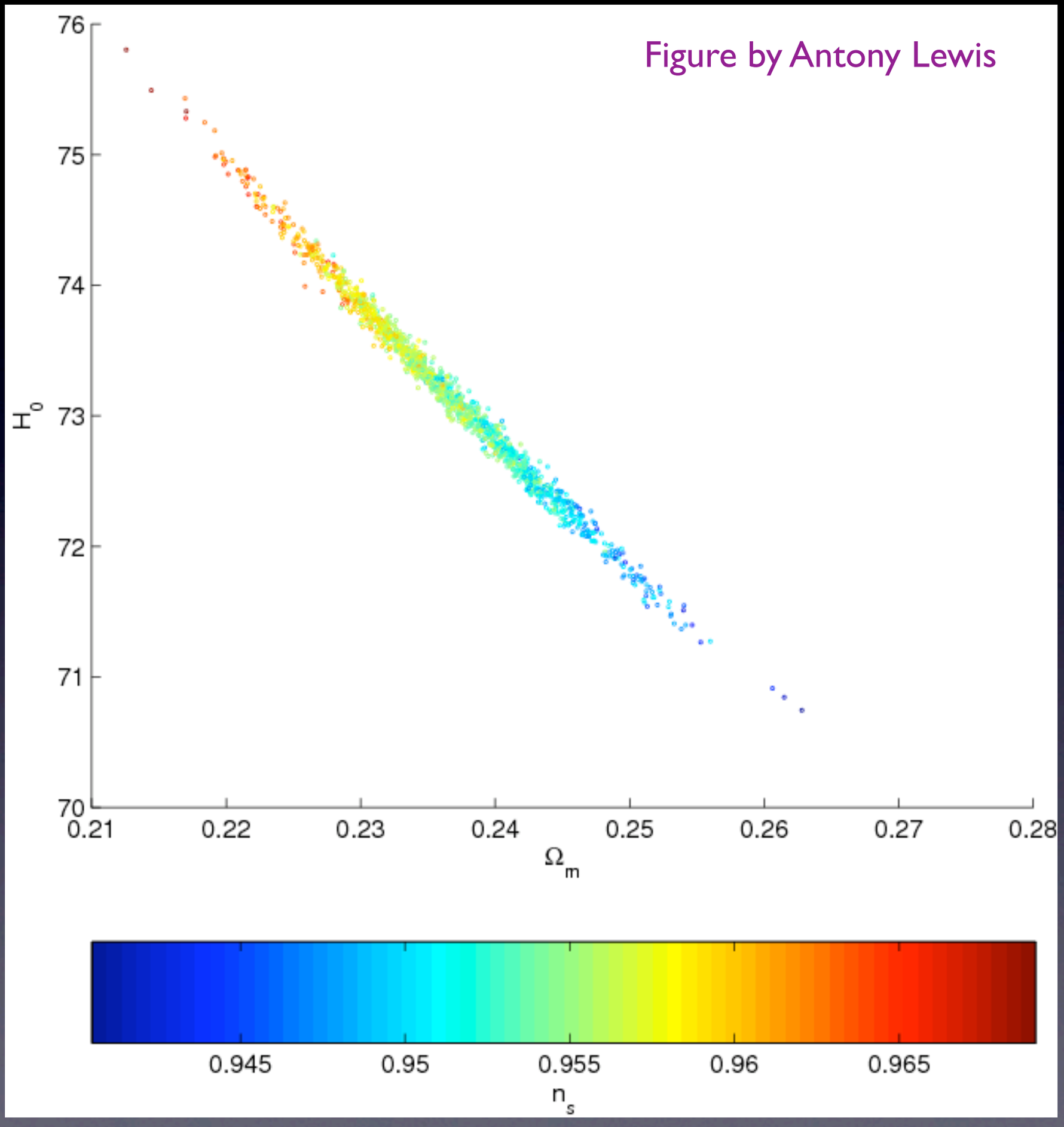
```
 0.2000000E+01  0.5851608E+04  0.2015806E-01  0.1056880E+00  0.1052353E+01  0.1572130E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9654508E+00  0.0000000E+00
 0.0000000E+00  0.3046525E+01  0.0000000E+00  0.0000000E+00  0.7791765E+00  0.1357033E+02  0.2208235E+00  0.7946533E+00  0.1775950E+02  0.0000000E+00  0.7549133E+02
 0.1000000E+01  0.5847794E+04  0.2015806E-01  0.1056880E+00  0.1052353E+01  0.1572130E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9742470E+00  0.0000000E+00
 0.0000000E+00  0.3056430E+01  0.0000000E+00  0.0000000E+00  0.7791765E+00  0.1357033E+02  0.2208235E+00  0.8014988E+00  0.1775950E+02  0.0000000E+00  0.7549133E+02
 0.1000000E+01  0.5843263E+04  0.2010433E-01  0.1014225E+00  0.1047609E+01  0.1655797E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9702829E+00  0.0000000E+00
 0.0000000E+00  0.3045888E+01  0.0000000E+00  0.0000000E+00  0.7877861E+00  0.1368099E+02  0.2122139E+00  0.7744952E+00  0.1822450E+02  0.0000000E+00  0.7567444E+02
 0.1000000E+01  0.5827815E+04  0.2010433E-01  0.1014225E+00  0.1047609E+01  0.1655797E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9612351E+00  0.0000000E+00
 0.0000000E+00  0.3043113E+01  0.0000000E+00  0.0000000E+00  0.7877861E+00  0.1368099E+02  0.2122139E+00  0.7705776E+00  0.1822450E+02  0.0000000E+00  0.7567444E+02
 0.2000000E+01  0.5827671E+04  0.2005007E-01  0.9967086E-01  0.1047475E+01  0.1679666E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9610596E+00  0.0000000E+00
 0.0000000E+00  0.3040069E+01  0.0000000E+00  0.0000000E+00  0.7947711E+00  0.1367524E+02  0.2052289E+00  0.7615068E+00  0.1835000E+02  0.0000000E+00  0.7637756E+02
 0.1000000E+01  0.5823237E+04  0.2062331E-01  0.9755769E-01  0.1048439E+01  0.1589623E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9725500E+00  0.0000000E+00
 0.0000000E+00  0.3020521E+01  0.0000000E+00  0.0000000E+00  0.8064781E+00  0.1357294E+02  0.1935219E+00  0.7447448E+00  0.1726400E+02  0.0000000E+00  0.7814636E+02
 0.1000000E+01  0.5810297E+04  0.2062331E-01  0.9755769E-01  0.1048439E+01  0.1589623E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9778811E+00  0.0000000E+00
 0.0000000E+00  0.3030873E+01  0.0000000E+00  0.0000000E+00  0.8064781E+00  0.1357294E+02  0.1935219E+00  0.7502713E+00  0.1726400E+02  0.0000000E+00  0.7814636E+02
 0.1000000E+01  0.5804656E+04  0.2062801E-01  0.8928808E-01  0.1047195E+01  0.1529961E+00  0.0000000E+00  0.0000000E+00 -0.1000000E+01  0.9788681E+00  0.0000000E+00
 0.0000000E+00  0.2981949E+01  0.0000000E+00  0.0000000E+00  0.8358389E+00  0.1352269E+02  0.1641611E+00  0.6901900E+00  0.1643250E+02  0.0000000E+00  0.8182678E+02
```

What to do with a chain

With a chain of samples, many things can be done.

1) Find the best fit, by identifying the highest-likelihood point.
2) Find confidence ranges for parameters simply by discarding information on all the other parameters and ordering the points in parameter value.
3) Make two-dimensional parameter constraint plots by plotting points in 2D (possibly with extra information conveyed by colour) or by plotting the point density onto a grid.

Figure by Antony Lewis



What to do with a chain

With a chain of samples, many things can be done.

1) Find the best fit, by identifying the highest-likelihood point.
2) Find confidence ranges for parameters simply by discarding information on all the other parameters and ordering the points in parameter value.
3) Make two-dimensional parameter constraint plots by plotting points in 2D (possibly with extra information conveyed by colour) or by plotting the point density onto a grid.
4) Importance sample: if you change your mind about your prior, or if new data comes along of known likelihood, you can reweight your samples without needing to recompute a chain.
5) Bayesian complexity: this quantity estimates how many parameters are actually constrained by your data, and can be computed from a chain.

Marginalisation

Marginalization is the process of averaging over uninteresting (nuisance) parameters. This is something traditionally difficult in frequentist methods, and relatively easy in Bayesian ones.

As an integral:

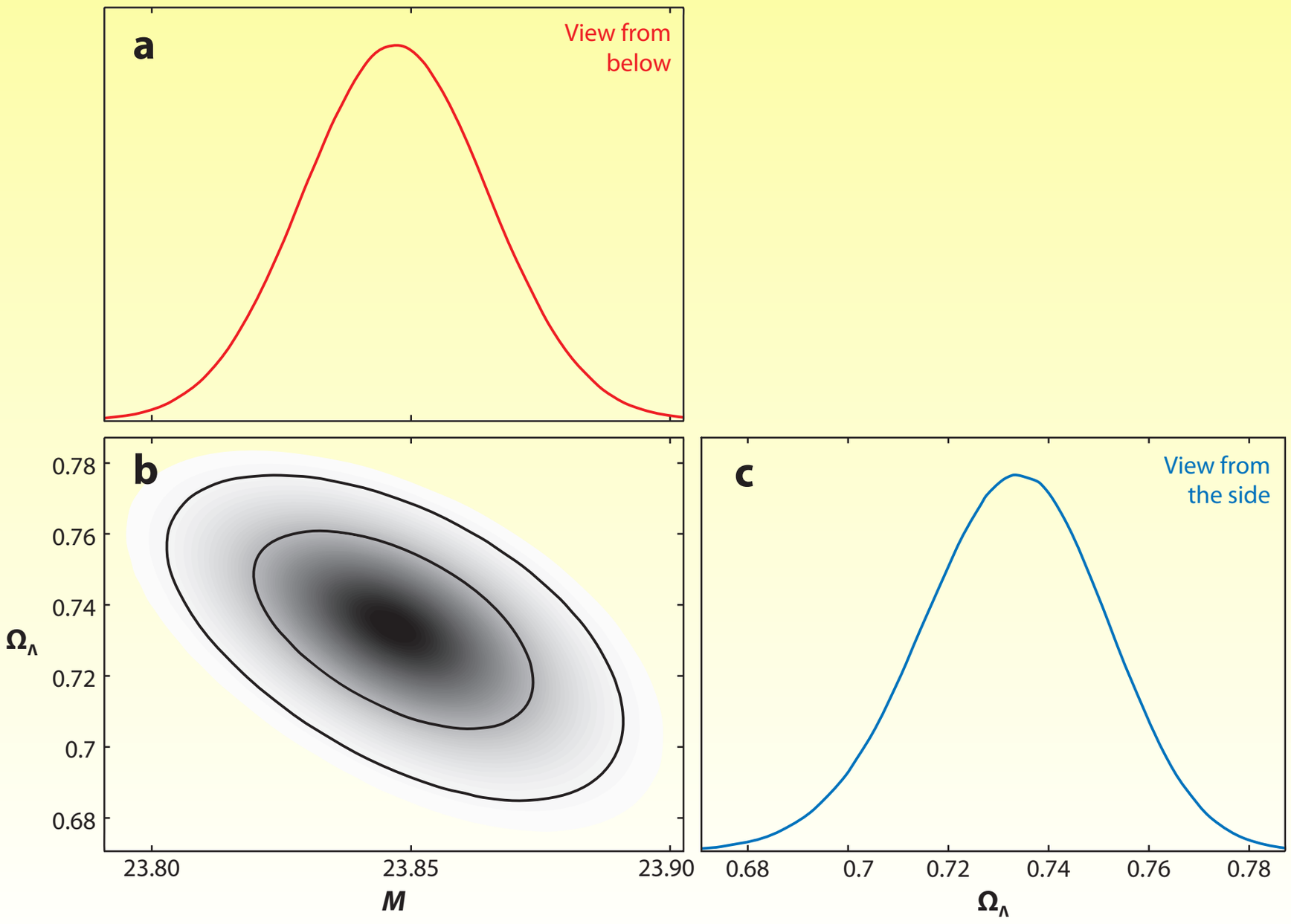
$$P(\theta_1) = \int_{\text{Domain}} P(\theta_1, \theta_2) d\theta_2$$

With samples:

Just ignore the θ_2 column in the chains and histogram the θ_1 column!

A simple example: cosmological parameters from supernova data

From Sahlen-Liddle-Parkinson



From WMAP5: Dunkley et al

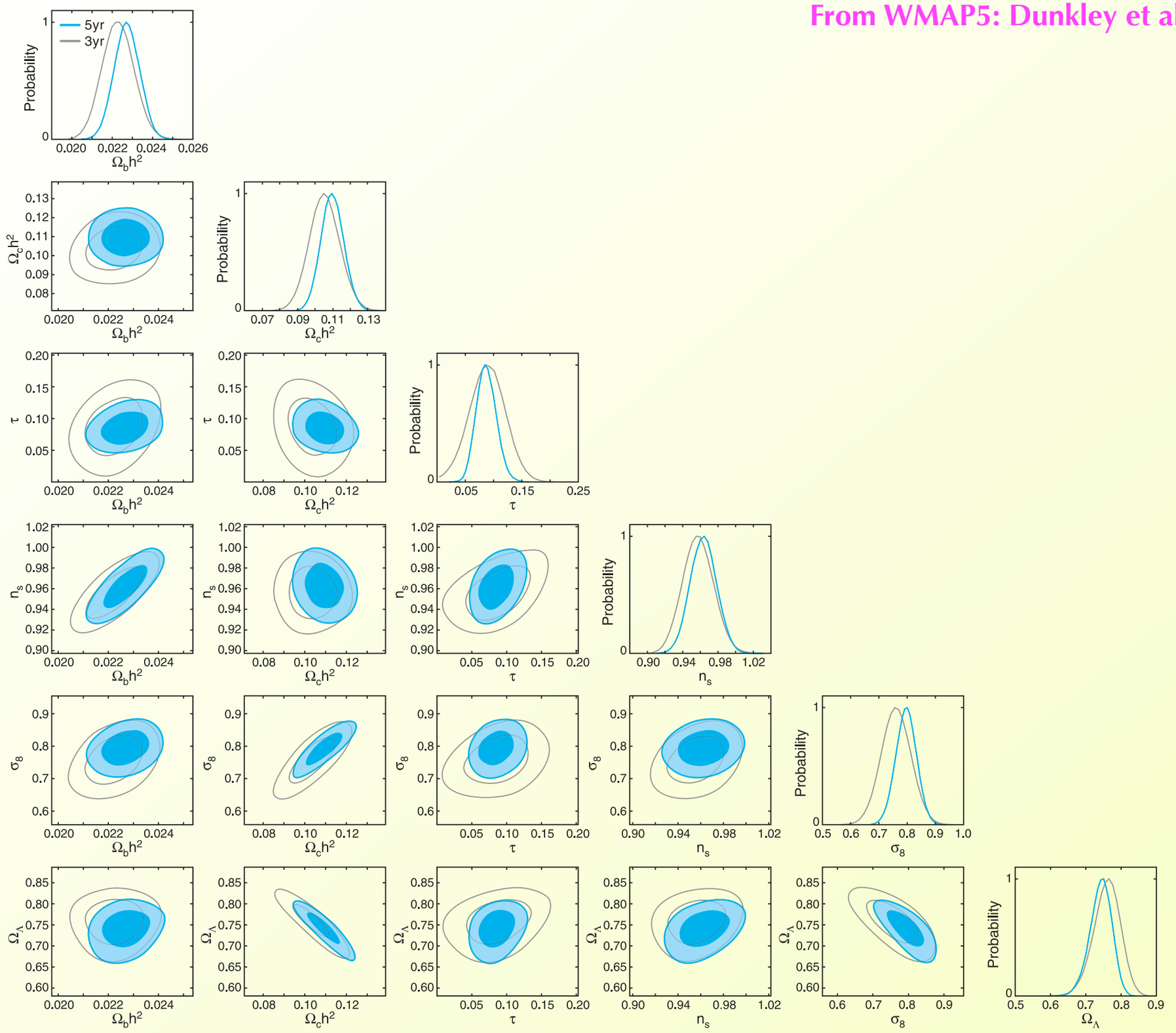


Fig. 4.— Constraints from the five-year *WMAP* data on ΛCDM parameters (blue), showing marginalized one-dimensional distributions and two-dimensional 68% and 95% limits. Parameters are consistent with the three-year limits (grey) from Spergel et al. (2007), and are now better constrained.

Parameters of the standard cosmological model

	WMAP5 alone	WMAP5 + BAO + SN
Baryon density $\Omega_{\rm b}h^2$	0.0227 ± 0.0006	0.0227 ± 0.0006
Dark matter density $\Omega_{\rm cdm}h^2$	0.110 ± 0.006	0.113 ± 0.003
Cosmological constant Ω_Λ	0.74 ± 0.03	0.726 ± 0.015
Spectral index n	$0.963^{+0.014}_{-0.015}$	0.960 ± 0.013
Optical depth τ	0.087 ± 0.017	0.084 ± 0.016
Perturbation amplitude $\Delta^2_{\mathcal{R}} \times 10^9$	2.41 ± 0.11	2.44 ± 0.10

The currently-favoured cosmology is a ΛCDM model, in a spatially-flat Universe, with initial conditions of the form expected from simple inflation models.

Other sampling algorithms

Although Metropolis-Hastings is the most popular option at present, other more sophisticated choices exist that may be superior in the right circumstances.

- Slice sampling

  This method allow the proposal function to change during the calculation, tuning to an appropriate scale, by making steps in one parameter direction at a time.

- Gibbs sampling

  This obtains a proposal step by sampling from conditional probabilities $P(\theta_1|\theta_2)$ etc.

- Hamiltonian sampling

  This uses an analogy with Hamiltonian dynamics to define a momentum from derivatives of the likelihood. This enables large proposal steps to be taken along trajectories of constant `energy'.

Machine learning/classification

A somewhat related topic to parameter estimation is machine learning, which for instance seeks to classify objects in a large dataset by training on a subset.

- Bayesian classifiers

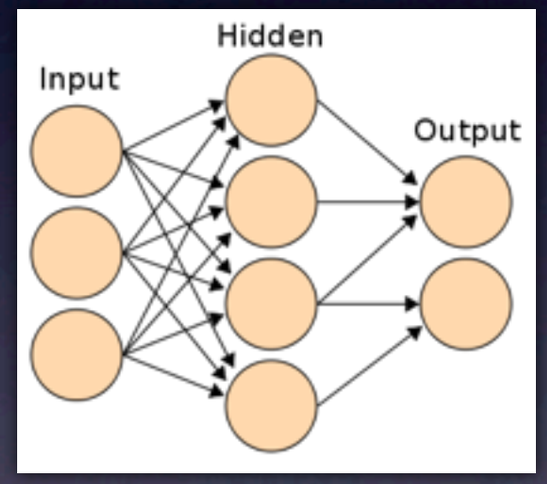
  Define outcome probabilities using training set data in a probabilistic framework.

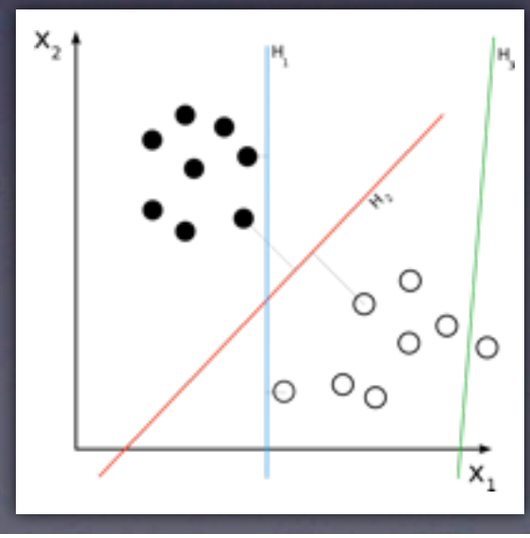
- Artificial neural networks

  These connect a set of nodes and train the coefficients.

- Support vector machines

  Hyperplanes are defined within the parameter space to segregate points with different classifications.





Classification of genetic damage in fish

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part C

journal homepage: www.elsevier.com/locate/cbpc

A novel population health approach: Using fish *retinoblastoma* gene profiles as a surrogate for humans☆

Jeanette M. Rotchell [a,d,*], Frances A. du Corbier [a], Grant D. Stentiford [b], Brett P. Lyons [b], Andrew R. Liddle [c], Gary K. Ostrander [d]

- Predictive classification of cancer incidence in marine organisms based on Rb tumour-suppressor gene mutation status.
  [Here with the goal of developing early-warning indicators of pollution-induced tumour incidence.]

Classification of genetic damage in fish

Table 1
Summary of *Rb* gene status (within exons 11–23) in histologically-normal, adenoma, and carcinoma liver tissues from *L. limanda*

Position within dab *Rb* cDNA →	1065 bp	1088 bp	1101 bp	1119 bp	1514 bp	1592 bp	1650 bp
Putative change in encoded protein →	P▸H	Polymorphism	N▸D	E▸K	Polymorphism	Polymorphism	L▸M
Sample no./tissue type /sample site							
1 Histologically normal IOM	C	T/C	G	G	T	T	A/T
2 Histologically normal IOM	C	T/C	G	G	G	G	A/T
3 Histologically normal OH	C	C	G	G/A	G/T	G/T	A/T
4 Histologically normal OH	C	C	G	G	G/T	G/T	A/T
5 Histologically normal IOM	C	T/C	G	G/A	G	G	A/T
6 Histologically normal OH	C	T/C	G	G	G	G	A
7 Histologically normal OH	C	T/C	G	G	G/T	T	A/T
8 Histologically normal IOM	C	T/C	G	G/A	G	G/T	A
9 Histologically normal IOM	C	T/C	G	G	G	G	A/T
10 Histologically normal OH	C	T/C	G	G	G/T	G/T	A/T
Sample no./tissue type /sample site							
1N Adenoma LB	C	T	G	G/A	G	G	A
1A Adenoma panc LB	C	T	G	G	G	G	A
2N Adenoma ICB	C	T/C	G	G	G/T	G/T	A
2A Adenoma ICB	C	T/C	G	G	G/T	G/T	A
3N Adenoma LB	C	T/C	G	G	G/T	G/T	A
3A Adenoma LB	C	T/C	G	G/A	G/T	G/T	A
4N Adenoma LB	C/A	T/C	G	G	G/T	G/T	A
4A Adenoma LB	C	T/C	G	G/A	G/T	G/T	A
5N Adenoma LB	C	T/C	G	G	G/T	G/T	A
5A Adenoma LB	C	T/C	G	G	G/T	G/T	A
Sample no./tissue type /sample site							
1N Carcinoma SCB	C	T/C	G	G/A	G/T	G/T	A
1C Carcinoma SCB	C	T/C	G/A	G	G/T	G/T	A
2N Carcinoma SCB	C	C	G/A	G/A	T	T	A
2C Carcinoma SCB	C	C	G	G/A	T	T	A
3N Carcinoma Rye	C	T	G	G/A	G	G	A
3C Carcinoma Rye	C	T/C	G	G/A	G/T	G/T	A
4N Carcinoma ICB	C	T/C	G	G/A	G	G	A
4C Carcinoma ICB	C	T	G/A	G	G	G	A
5N Carcinoma ICB	C	T/C	G	G/A	G/T	G/T	A
5C Carcinoma ICB	C	T/C	G/A	G/A	G/T	G/T	A
6N Carcinoma ICB	C	T/C	G/A	G/A	G/T	G/T	A
6C Carcinoma ICB	C	T	G/A	G	G	G	A

As described in Section 2.3, the table contains two categories of 'normal' tissue. The first category is from fish that exhibited no evidence of histopathological liver lesions and, in the absence of laboratory-reared *L. limanda*, these serve as the control group. The second category, those labelled as 'XN', are histopathologically-determined, laser micro-dissected, 'normal' liver tissue taken from the vicinity of a tumor, in the same fish bearing that tumor. Sample site abbreviations: IOM—Isle of Man, OH—Off Humber, LB—Liverpool Bay, ICB—Inner Cardigan Bay, SCB—South Cardigan Bay.

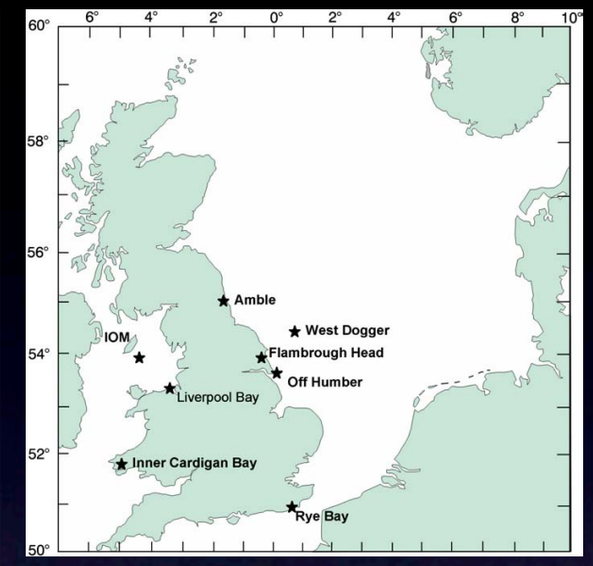


Fig. 1. Map of sampling sites.

Conclusions

- Markov chain Monte Carlo is the method of choice for parameter estimation.
- Its advantages include computational efficiency and easy marginalization over uninteresting parameters.
- A variety of sampling methods can be brought into play. Metropolis-Hastings is typically sufficient but some circumstances may require specialist methods. In particular Hamiltonian sampling may be suited to deploying in models of very high dimensionality (thousands or even millions of parameters).