Bayesian parameter estimation with weak data and when combining evidence: the case of climate sensitivity

Nicholas Lewis

Independent climate scientist

CliMathNet 5 July 2016

Main areas to be discussed

- Why Objective not Subjective Bayes is needed for parameter estimation
- How correctly to combine evidence in the Objective Bayesian case

2

My talk concerns estimating a fixed but unknown, continuously valued parameter, linked to data by a statistical model.

I focus on contrasting Subjective and Objective Bayesian parameter estimation methods and on their use in estimating equilibrium climate sensitivity or ECS.

I will restrict discussion of combining evidence to the case where it is independent.

Probability is not a settled field

- Probabilistic inference has a troubled history
- Bitter disputes and personality clashes in past
- Deep disagreements over fundamental issues
- A few main belief sets and multiple variants
- In parts more like philosophy than mathematics
- Bayesian inference disdained for much of 20th C
- Subjective Bayes now strong: suits computers

3

Efron 1998: the "250-year search for a dependable objective Bayesian theory".

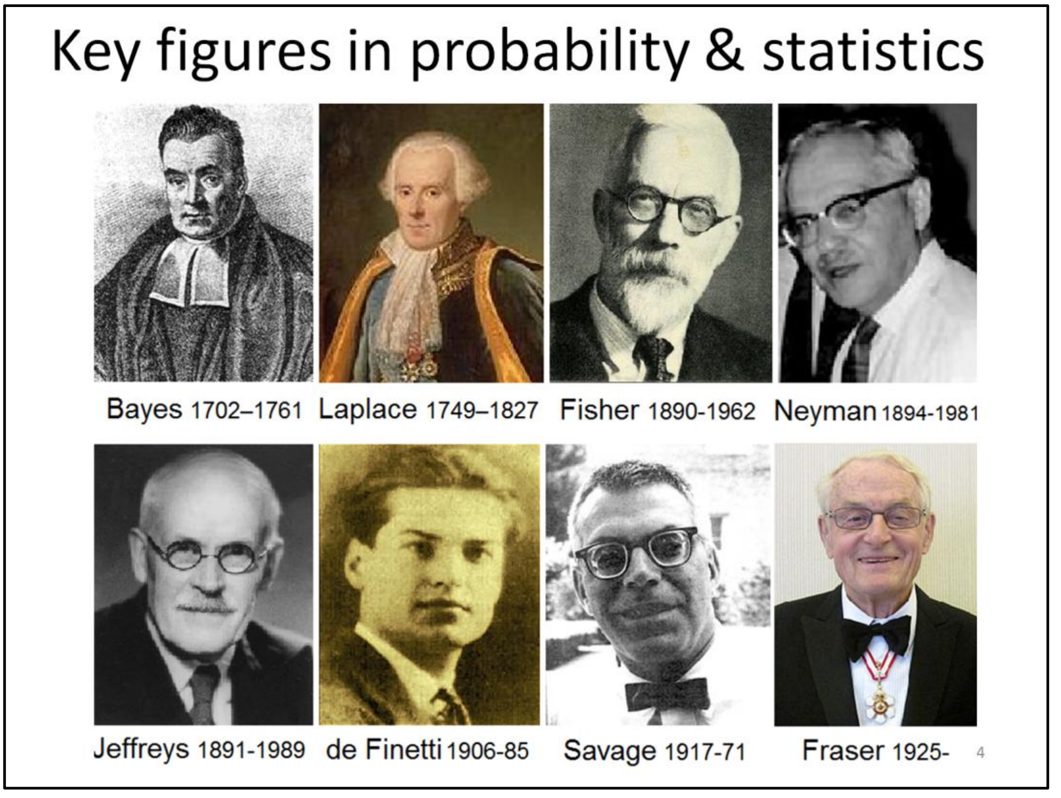
Still disagreement over fundamentals of probability and doubts over supposedly proven theories.

Main belief sets are frequentist and Bayesian (subjective dominant, also objective – and others)

Philosophers still involved in probability theory, its foundations and its development.

Bayes theorem oscillated between boom and bust; still controversial.

Computing power has made it easy to apply powerful Bayesian methods, so have become popular.



Bayes – developed theorem for a randomly-generated parameter with known prior distribution

Laplace – applied Bayes theorem for scientific parameter inference, using uniform priors

Fisher – dominant: significance testing, likelihood; tried to replace Bayes (fiducial inference)

Neyman – developed frequentist paradigm: devised confidence intervals; hypothesis testing

Jeffreys – father of objective Bayesianism: first noninformative prior, based on invariance

de Finetti – developed subjective Bayes: probability as a purely personal, coherent, belief

Savage – another famous developer of the personalistic, subjective Bayesian, approach

Fraser – Excellent grasp of problems with Bayesian inference and how to improve its and likelihood methods’ probability matching.

Bayesian estimation: continuous case

- Usual, Subjective method: apply Bayes theorem
- Likelihood: probability density for data at observed value *y*, expressed as a function of the parameter θ
- Prior: estimated PDF for θ given current evidence
- Posterior PDF= Likelihood x Prior PDF, normalised
- $p(\theta|y) = c\ p(y|\theta)\ p(\theta)$; c set so that $\int p(\theta\,|y)d\theta = 1$
- x%–y% range: posterior CDF credible interval; CrI

5

Bayes theorem involves multiplication by a prior to convert Likelihood (a probability density for the observed data, as a function of the parameter) into a density for the parameter – the posterior PDF.

In the Subjective Bayesian view, the prior is a probability density function or PDF that reflects personal beliefs about the likely parameter value, before updating by the data likelihood function.

Posterior PDF is normalised so that its $\int$, the cumulative distribution function (CDF) reaches 1; CDF gives CrI ranges.

Does Bayes theorem apply here?

- Bayes theorem restates conditional probability $p(\theta|y)\ p(y) = p(y,\theta) = p(y|\theta)\ p(\theta)$; divide by $p(y)$
- Mathematically valid in the continuous case iff y and θ are Kolmogorov random variables (KRV)
- Bayes theorem gives mathematical probability iff parameter value is random with known prior PDF
- A fixed but unknown parameter is not a KRV
- Subjective Bayes provides coherent personal beliefs about parameters, **not** scientific validity

6

Fundamental difference exists between a known probability distribution that governs generation of a RV (=> conditional probability works) and an estimated PDF for a fixed but unknown parameter.

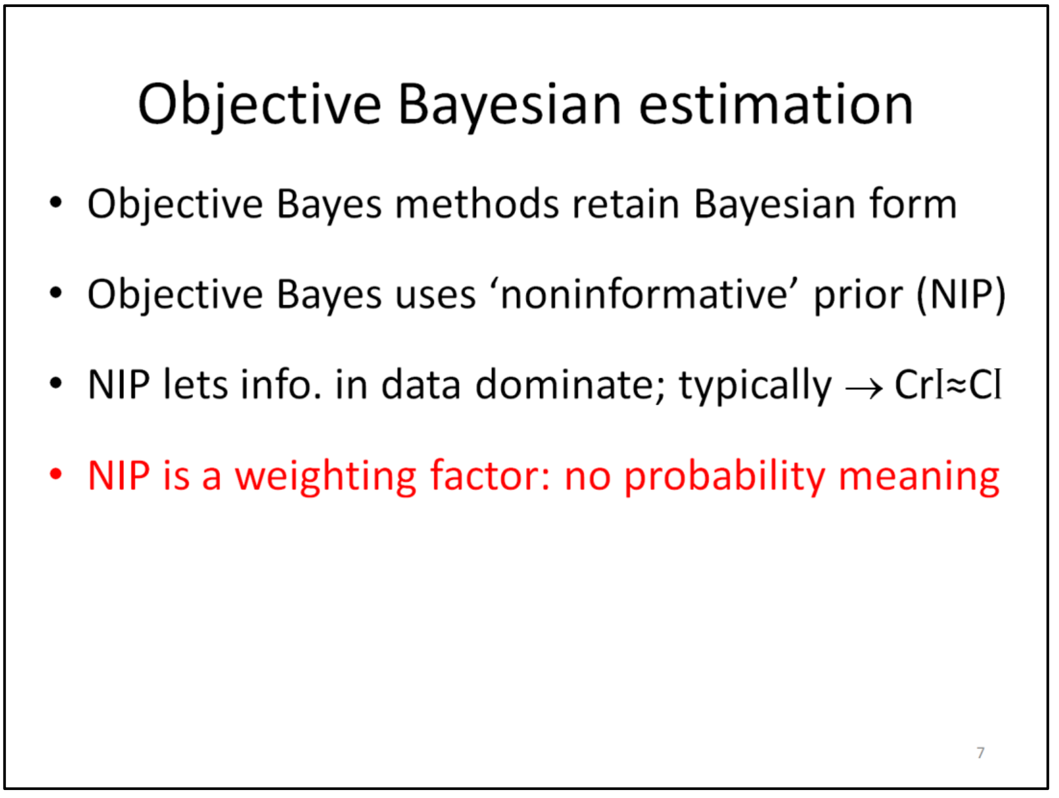
Coherent: bets using your probabilities offer no arbitrage possibility. Too weak a constraint for science, which requires corrrespondence to observed reality [Gelman]. So don't be a Subjectivist.

Fisher: "any statement of probability to be objective must be verifiable as a prediction of frequency"

Gelman: 'correspondence to observed reality [is] a core concern regarding objectivity' and 'the ultimate source of scientific consensus.'

Barnard 1994: a fixed, unknown parameter is not a KRV.

Fraser 2010: using the conditional probability lemma does not produce probabilities from no probabilities; it needs two probability inputs.



'Objective' in the sense that posterior PDFs for parameters are based only on the statistical model and observed data.

Information theory based *reference priors* achieve this, as they are noninformative about parameter values.

NIPs usually lead to CrI ~ confidence interval (CI), so ranges correctly reflect data uncertainties.

Cannot interpret a noninformative prior as indicating relative probabilities of parameter values.

Bernardo 2003: "posterior distributions encapsulate inferential conclusions on the quantities of interest solely based on the assumed model & the observed data", as in frequentist approach. (Bayesian Statistics chapter, in the volume Probability and Statistics of the Encyclopedia of Life Support Systems, Oxford.)

Noninformative priors

- Jeffreys prior: sqrt(|Fisher information matrix|)
- NIP => inference invariant on reparameterisation
- NIP *inter alia* converts probability between data and parameter spaces: includes a Jacobian factor
- NIP is usually simple in data space−often uniform
- Can use Bayes in data space and change variables to parameter space – even if lower dimensional

8

The original NIP – Jeffreys prior – was developed to make inference invariant.

Jeffreys prior is the reference prior for all parameters jointly; itcan need modifying for parameter subsets.

Fisher information: expected data informativeness about parameters at each point in parameter space.

Jacobian factor in NIP means it will be highly nonuniform if data–parameter relationships are strongly nonlinear.

May be easier to use Bayes theorem in data space and convert the posterior data PDF to a parameter PDF.

Fisher information & JP are related to Kullback-Leibler divergence & metric densities.

NIP typically improper; its integral is infinite – but posterior PDF normally fine.

Subjective vs Objective ECS estimation

- Many ECS estimates use Subjective Bayes
- Subjective ~ Objective Bayes if data strong
- ECS data too weak to dominate most priors

9

Studies usually provide an estimated PDF for ECS, implying a Bayesian paradigm: frequentism doesn’t give parameter PDFs

ECS studies rarely use objective Bayes explicitly, but may do so implicitly.

One can get away with Subjective Bayes if data strong. But estimation of ECS involves weak data and highly nonlinear data-parameter relations for ECS and also ocean diffusivity K_v (often jointly estimated).

Subjective vs Objective ECS estimation

- Many ECS estimates use Subjective Bayes
- Subjective ~ Objective Bayes if data strong
- ECS data too weak to dominate most priors
- Subjective Bayesian posterior PDFs for ECS generally don't reflect data error distributions
- Uniform ECS & K_v priors greatly bias estimates
- 'Expert' ECS prior usually dominates over data

10

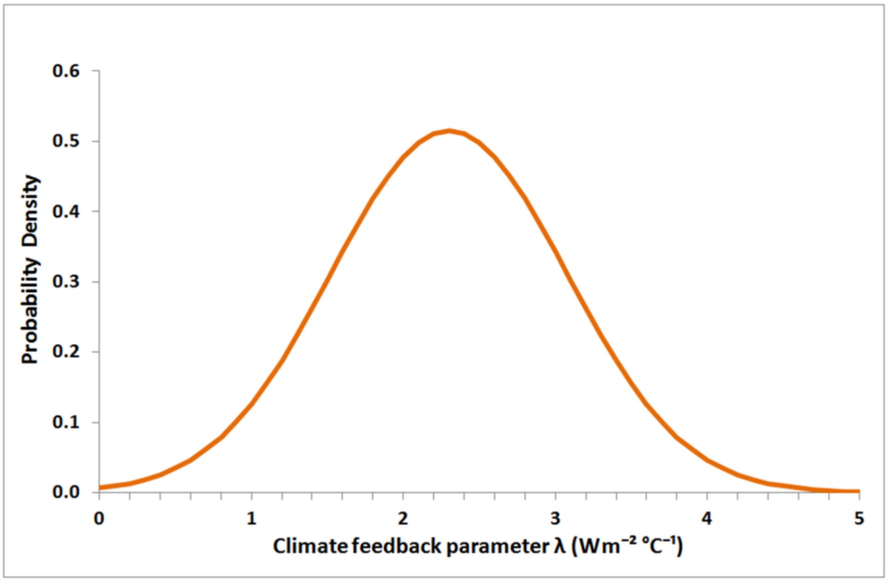
Uniform priors have typically been used, but nonlinear relationships mean they are strongly informative about the parameters and usually result in unrealistic ECS estimation.

Some ECS studies use so-called expert priors, reflecting consensus estimates. However, in science it is standard to report results that reflect the information provided by the experimental data used.

So be an Objectivist and use a noninformative prior!

Bernardo 2009: 'If no prior information is to be assumed, a situation often met in scientific reporting and public decision making.'

Estimating λ(∝1/ECS): Gaussian errors

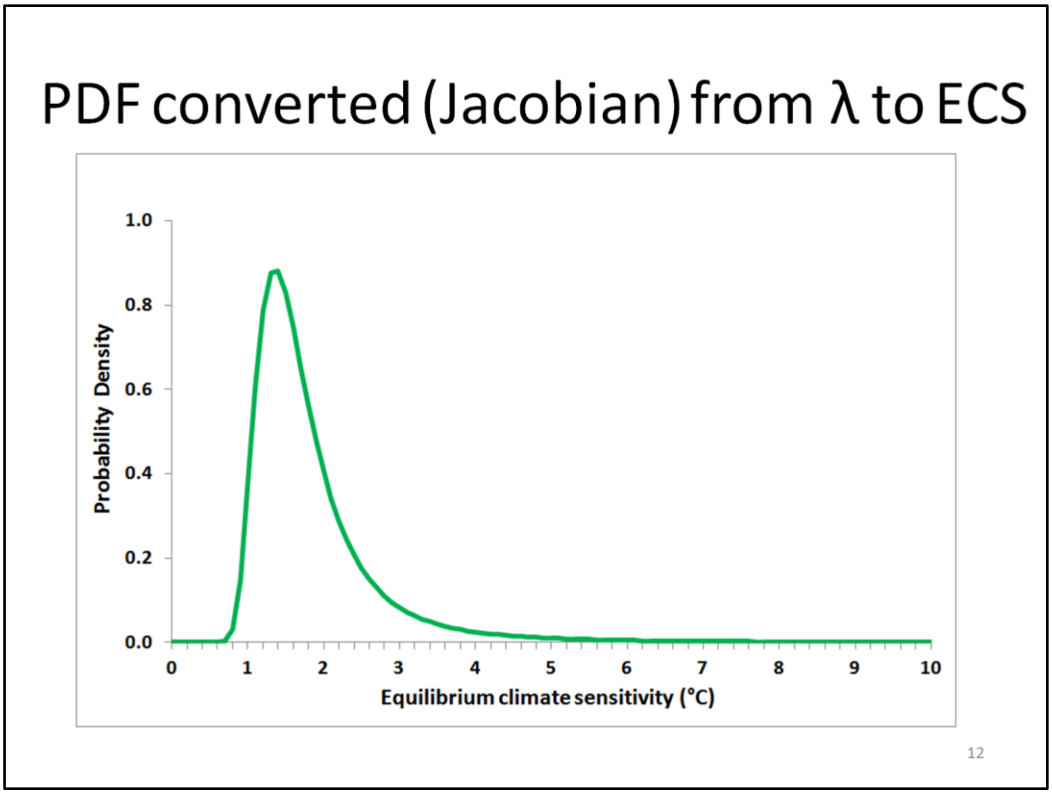


11

I will use Forster & Gregory 2006's results to demonstrate the effect of using a uniform prior for ECS vs a noninformative prior.

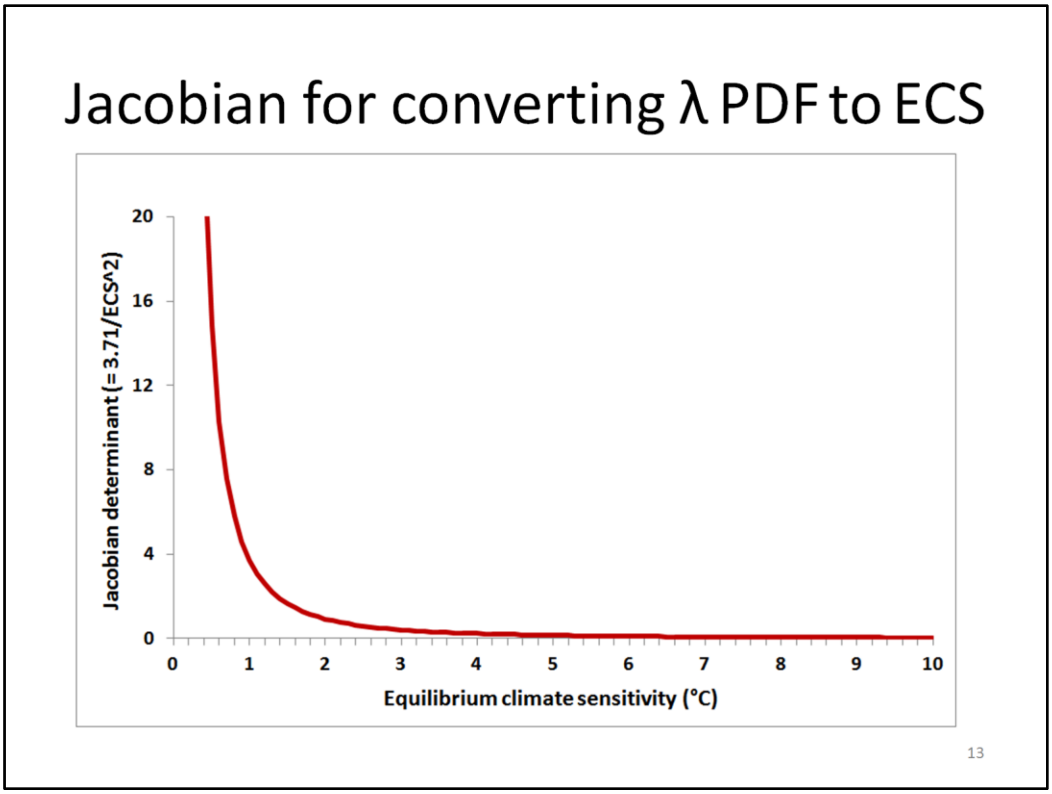
FG06 used an objective method and gave a median estimate for the climate feedback parameter λ of 2.3 $Wm^{-2}K^{-1}$, with Gaussian uncertainty.

They noted that their estimate effectively used an almost uniform-in-λ prior.



Using ECS = 3.71/λ, the FG06 PDF for λ can be converted into a PDF for ECS by changing variable, restating the PDF in terms of ECS and multiplying it by the applicable Jacobian factor.

The resulting median estimate is 1.6 K; 95% bound is 3.6 K (per IPCC AR5 uniform-in-λ prior based PDF)

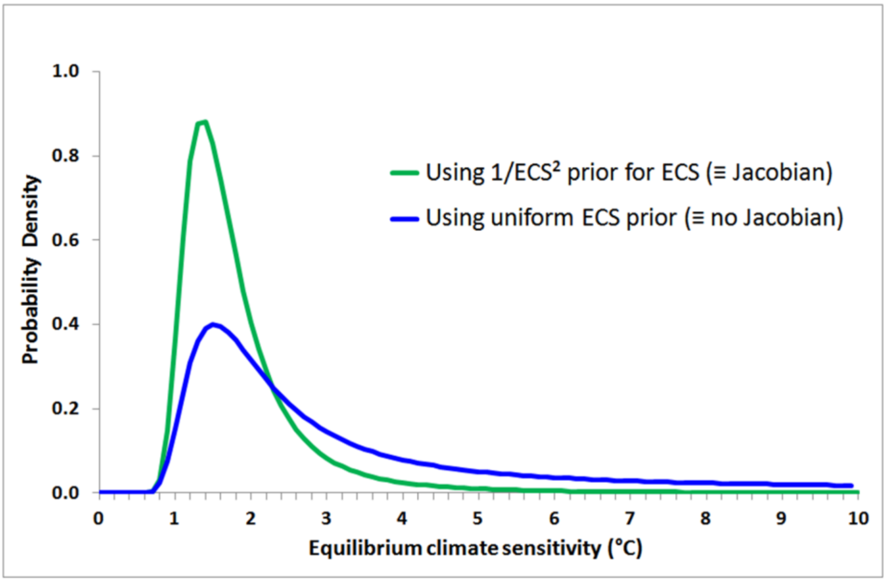


The Jacobian factor is the absolute derivative of λ wrt ECS, being 3.71 ECS^{-2}.

NIPs transform like PDFs so, as the NIP for λ is uniform, the NIP for ECS is proportional to ECS^{-2}, like the Jacobian.

The NIP for each parameter is the Jeffreys prior for estimating that parameter from the experimental data.

ECS PDF: effect of uniform prior vs NIP



14

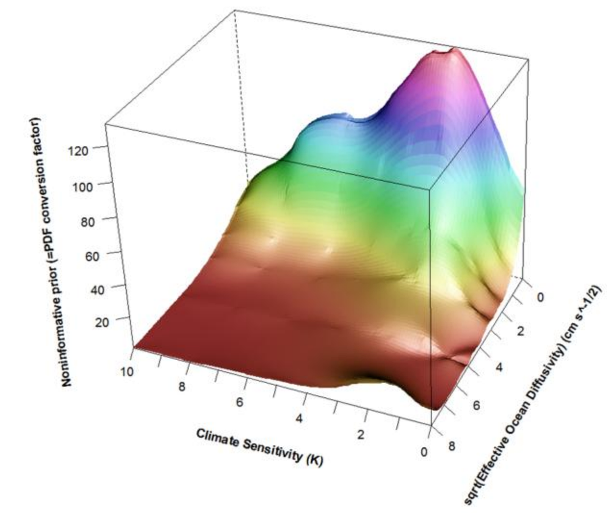
IPCC AR4 restated the FG06 results using a uniform-in-ECS prior over 0 to 18.5 K – giving the blue posterior PDF.

Resulting median estimate was over 60% higher at 2.65 K; 95% bound was nearly quadrupled at 14.2 K.

(Both would be infinite but for the 18.5 K limit on the prior.)

Using a uniform prior for ECS is equivalent here to changing variable without using the Jacobian factor to convert the PDF.

3D: Jeffreys prior for ECS, $\sqrt{K_V}$ & F_{aer}



Lewis (2013) An objective Bayesian improved approach... J Clim

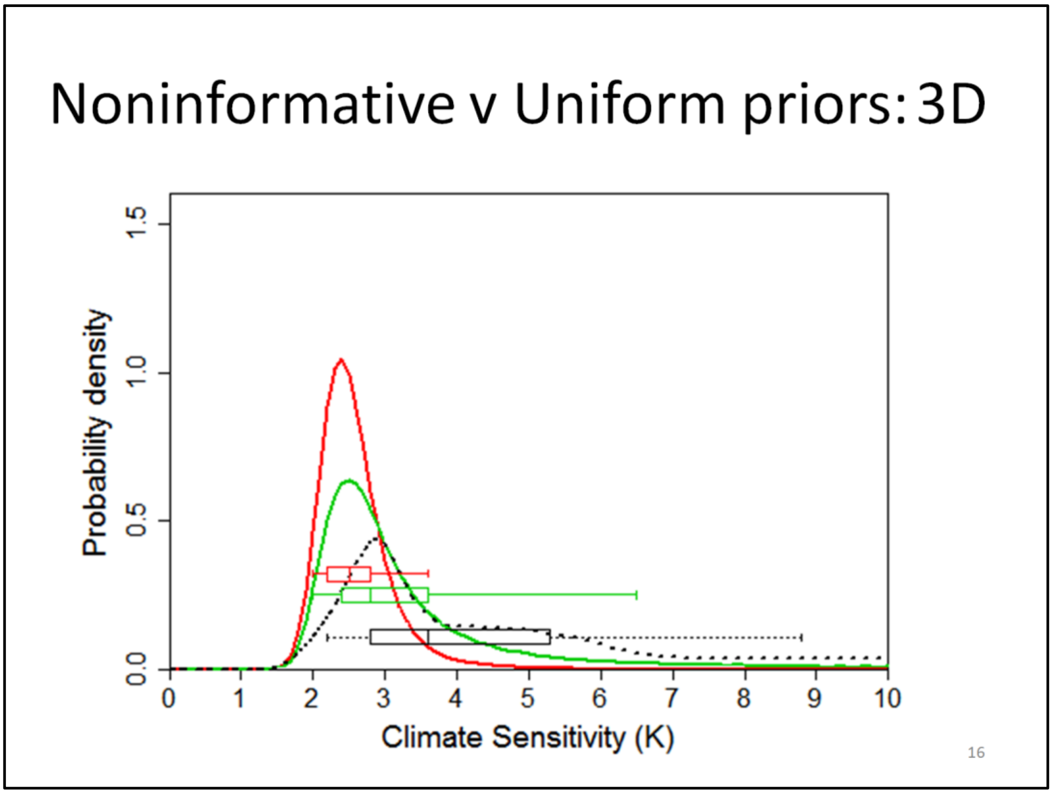
Now an example of a NIP involving 3 estimated parameters. The plotted joint prior is averaged over aerosol forcing, with which its shape varies little.

The parameter origin is in far RH corner. Bumpiness is due to residual climate model internal variability.

Prior ≡ PDF conversion factor for the dimensionally-reducing transformation from 17D whitened data space (in which Jeffreys joint prior is uniform) to 3D parameter space.

Prior ≠ the product of 3 separable individual priors

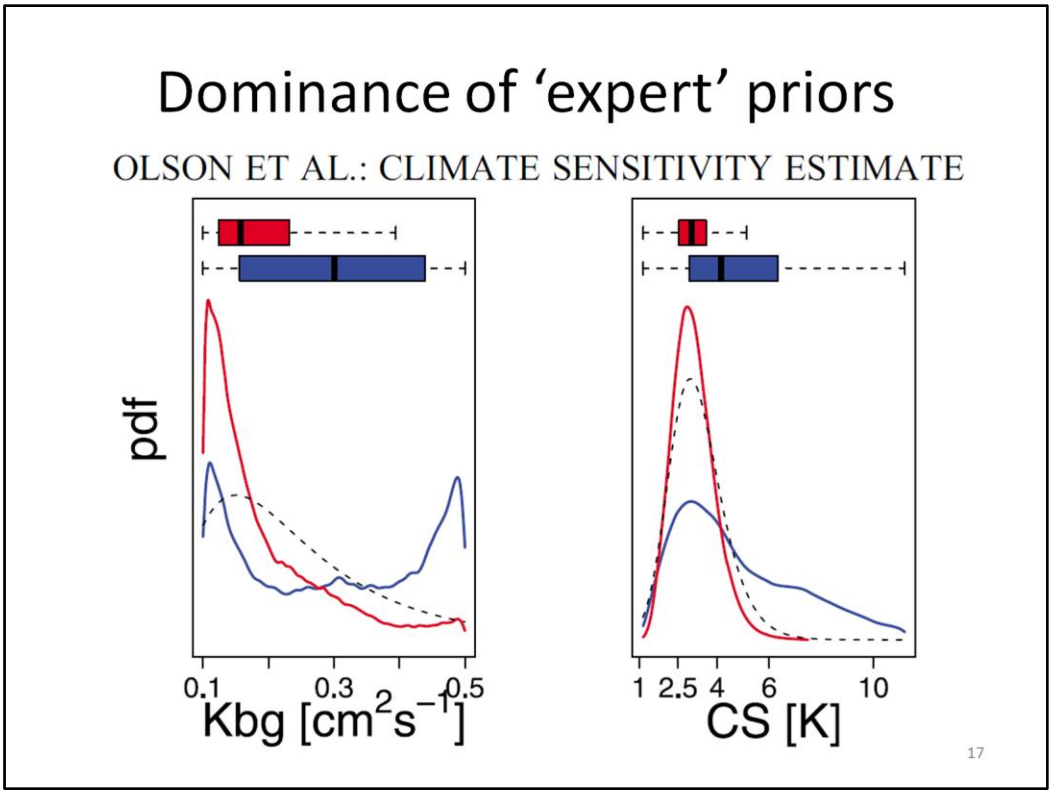
Optimal fingerprint method reduces observations to 17 independent *N*(0,1) whitened variables.



Posterior PDFs from Lewis 2013 using Forest et al 2006 data/diagnostics. Box plots show medians and 25–75% & 5-95% ranges

Red= Noninformative Jeffreys prior; Green= Uniform prior in ECS, sqrt(Kv) and Faer

Black= as reported in Forest 2006; it also had several other statistical and coding errors



Olson et al 2012 Fig.3. PDFs for ocean heat diffusivity and ECS. Posterior PDFs: blue line – with uniform priors, red lines – with 'expert' priors (black dashed lines)

Neither choice provides sensible estimation. With the 'expert prior' the ECS estimate is close to the prior, particularly for ECS. This is not how science should work.

(Kbg is related to rapidity of heat transfer between the surface and deep ocean; similar to $K_{v.}$)

Combining ECS evidence

- Instrumental & paleo evidence ~ independent
- Standard Bayes method: combine by updating
- Final posterior= Π(likelihoods) x prior for 1st est.

18

Several studies have combined instrumental and paleo evidence using subjective Bayesian methods.

Annan & Hargeaves 2006; Hegerl et al 2006; Aldrin et al 2012

The instrumental data most suitable for estimating ECS is non-independent, so use one good representative estimate.

Updating is the standard Bayesian way to combine evidence; use an existing posterior PDF as the prior.

Subjective Bayesian method => unique result, $\because$ prior is the same whatever data is to be analysed.

Bayesian updating is unsatisfactory

- Objective Bayes + updating: order-dependence!
- NIPs likely to differ for each source of evidence
- Objective Bayes & Bayesian updating incompatible

19

Use of a NIP (which varies with the experimental setup) when combining evidence using Bayesian updating will usually give a different final posterior depending on which evidence is analysed first (and thus which NIP is used).

Order dependency can't be right. I conclude Bayesian updating is faulty.

Avoiding Bayesian updating

- Solution: compute NIP for combined evidence
- Bayes on combined prior+combined likelihood
- To combine Jeffreys priors, add in quadrature
- Represent any prior info by data likelihood+NIP

20

One can avoid unsatisfactory Bayesian updating by instead computing a noninformative prior to use with the combined-evidence data likelihood, in a single use of Bayes theorem.

Jeffreys priors for independent evidence are simple to combine. In 1D, just add JPs in quadrature, since they are the square root of Fisher information, which is additive.

Since updating is invalid for Objective Bayes, represent any existing knowledge by a likelihood function and related NIP, not a prior PDF; use a notional datum: Hartigan 1965

Objectively combining ECS evidence

- Bayes on combined prior+combined likelihood
- To combine Jeffreys priors, add in quadrature
- Apply to ECS estimation using parameterised distributions for which Jeffreys-like NIP known
- ECS ~$\propto$ ratio of two normals: ΔT / ($\Delta F - \Delta Q$)
- Good approximation to this distribution: RS93
- Method gives CrI=CI for combined ratio-normals

21

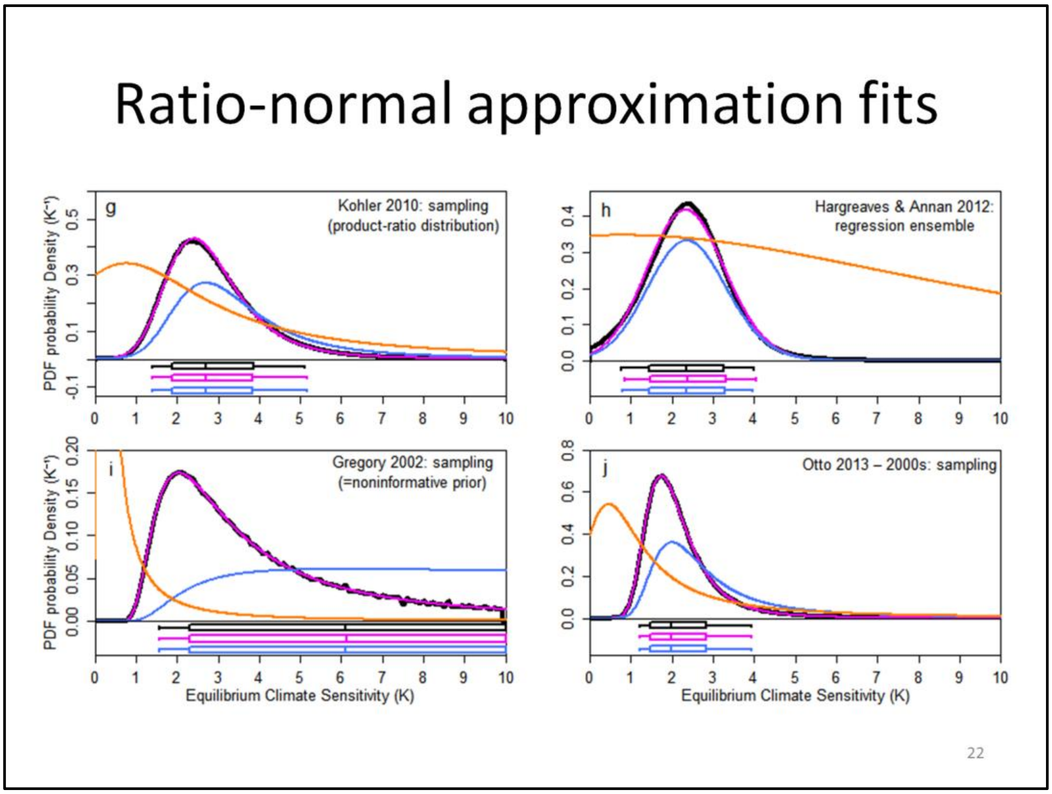
Likelihood & Jeffreys prior for ECS evidence typically not available, so use parameterised distributions.

ECS energy budget equation; uncertainty in ΔT and ($\Delta F - \Delta Q$) is ~ Normal.

Use Raftery & Schweder (RS93) ratio-normal approximation: posterior PDF is uniquely factorizable into constituent likelihood & Jeffreys prior.

Probability matching tested & ~perfect; using Bayesian updating it is poor & depends on order of analysis.

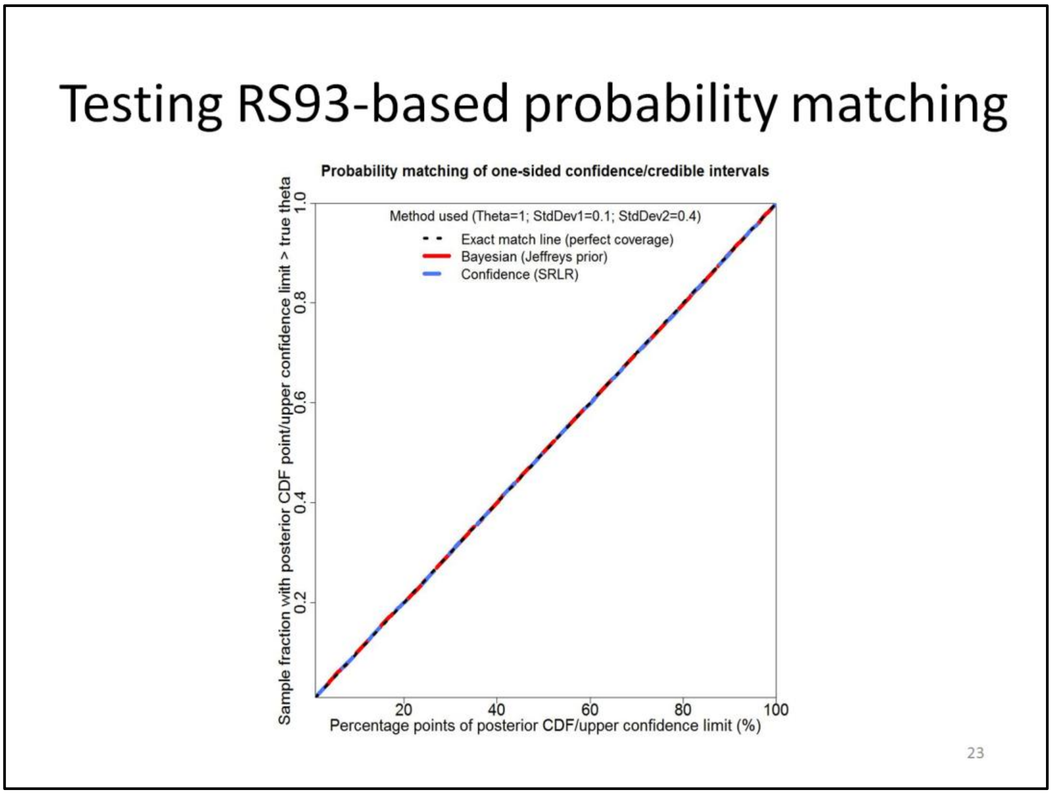
RS93 approximation: Posterior PDF of { (ΔT – ECS [$\Delta F - \Delta Q$]) / sqrt(sigma $_{\Delta T}{}^{2}$ + ECS^2 sigma $_{\Delta F}{}^{2}$) } ~ N(0,1)



The RS93 distribution, with only 3 free parameters, fits ECS PDFs very well (except those using highly informative priors).

Representative objectively-based estimated PDFs for ECS: top paleo; bottom instrumental period.

Black line = original PDF; magenta = RS93 fitted PDF; blue = derived likelihood function and SRLR-based confidence intervals from it; orange = derived Jeffreys prior

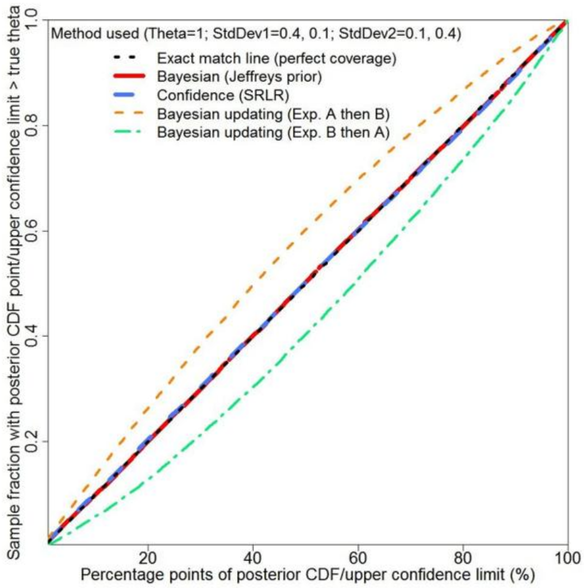


Test by sampling from data error distributions in what % of cases parameter value < 1%, 2%,...99% uncertainty bounds given by RS93-Bayesian and frequentist Signed Root Log-likelihood Ratio (SRLR) methods.

Perfect Bayesian and SRLR probability matching shows: a) RS93 is an excellent approximation to the ratio-normal; b) likelihood function & Jeffreys prior are correctly derived.

15,000 samples drawn randomly from two unit mean normal distributions and their ratio taken: numerator and denominator normals have fractional standard deviations of 0.1 and 0.4. Same matching if the two sds are swapped.

Testing probability matching: combination



24

Probability matching for evidence from two experiments involving ratio-normal estimates with very different numerator and denominator SDs, combined objectively using RS93 distributions and added-in-quadrature Jeffreys prior (red line).

Almost straight and 45° => ~ perfect matching of CrIs to CIs. Based on many random draws from data error distributions.

Probability matching of added-in-quadrature Jeffreys prior method of objectively combining evidence is ~ exact; likelihood ratio confidence inference likewise. Using Bayesian updating it is poor and depends on which dataset is analysed first.

Combining recent & paleo evidence

- Represent each ECS estimate by RS93 formula
- RS93 3-parameter formula fits ECS PDFs well
- Median and range suffice to uniquely specify

25

Ratio-normal distribution has 3 free parameters.
So gives a unique fit to median + range.
And can fit ECS PDFs well provided they don't use a highly informative prior.

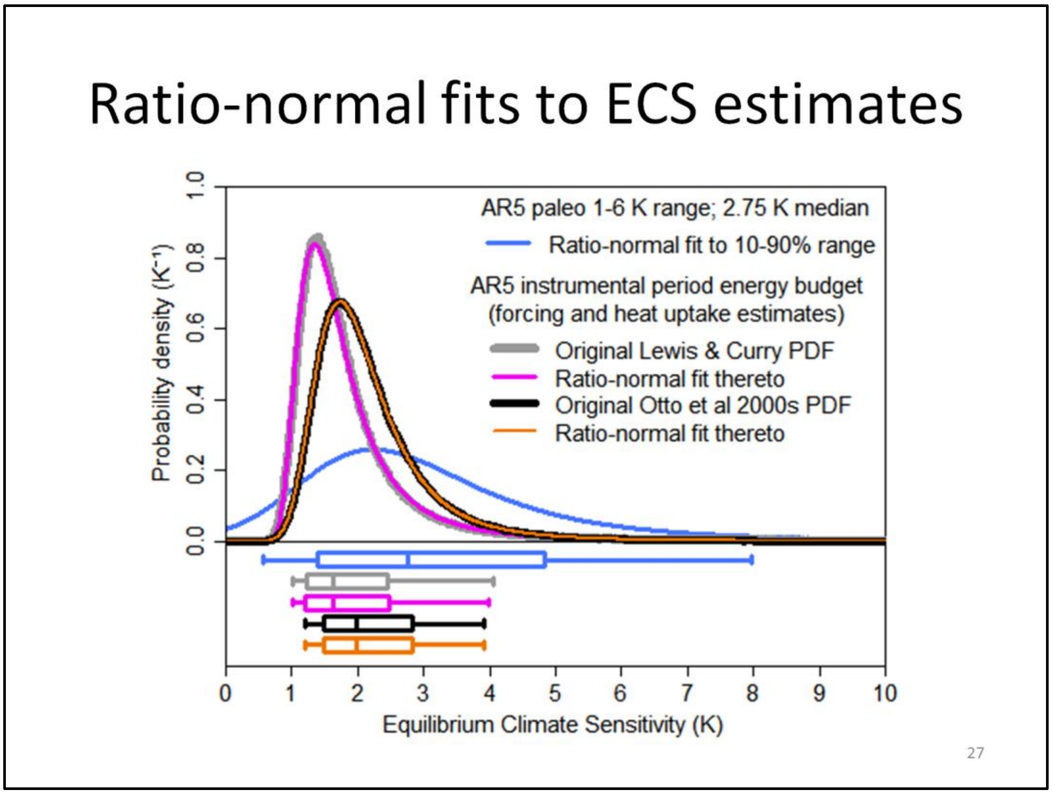
Combining recent & paleo evidence

- Represent each ECS estimate by RS93 formula
- RS93 3-parameter formula fits ECS PDFs well
- Median and range suffice to uniquely specify
- Instrumental: Lewis & Curry 2015; Otto ea 2013
- Paleo range: 1–6 K per AR5, as exactly 10–90%
- Paleo median: 10 AR5 estimates; median 2.75 K

26

Use two instrumental period energy budget estimates that reflect AR5 forcing and heat uptake data.

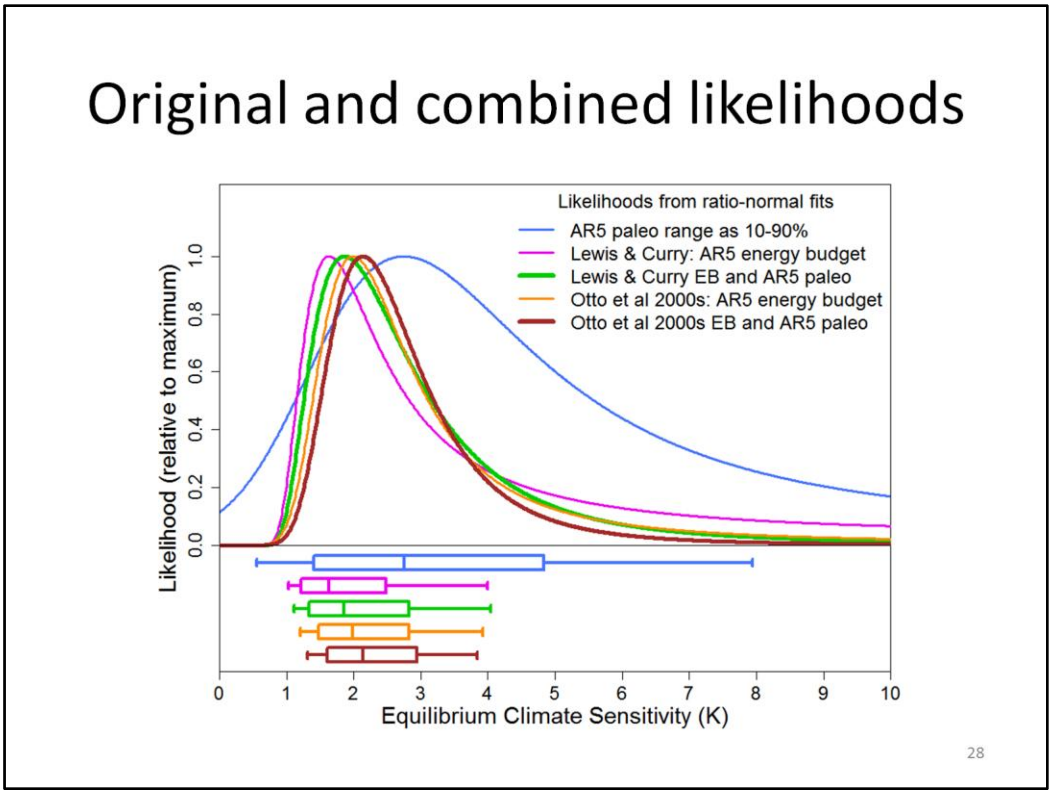
Use range and median for paleo that reflect the assessment made and evidence cited in AR5.



RS93 ratio-normal fit to black Otto et al original PDF is ~ perfect; it is very close for grey Lewis & Curry PDF.

Blue line is the fitted paleo evidence PDF.

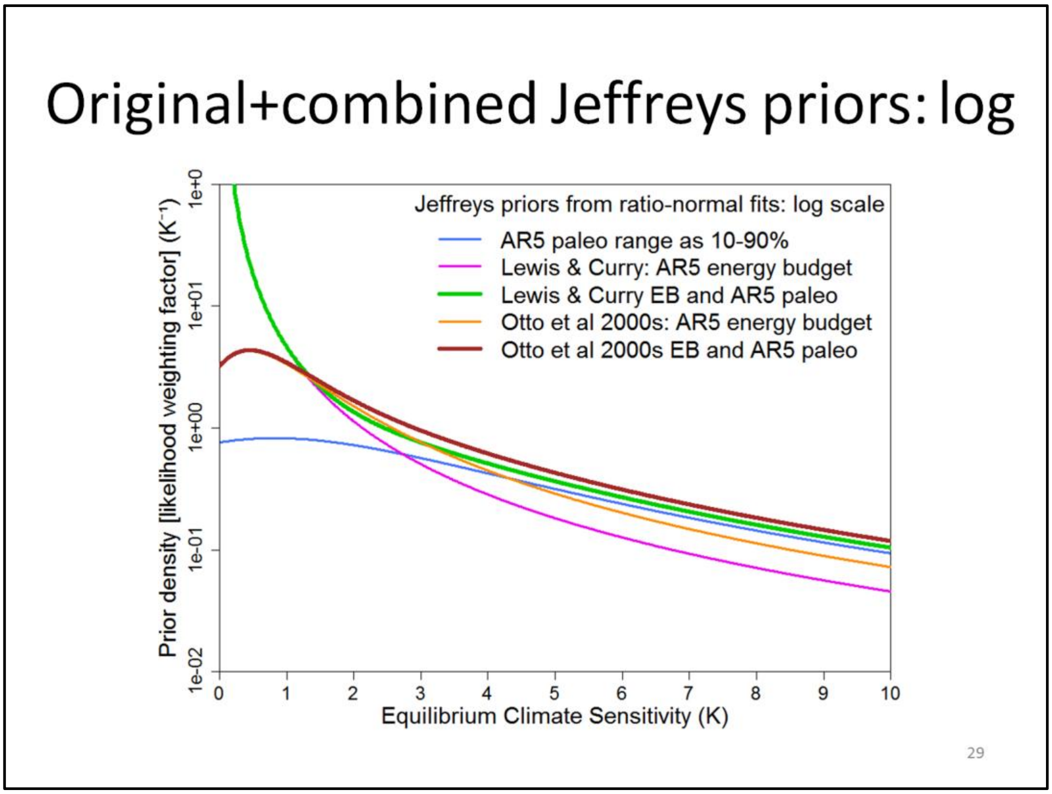
Box plots show medians, 17–83% and 5–95% credible intervals from PDFs.



Adding paleo evidence shifts instrumental likelihoods slightly to right, and reduces them at high ECS.

Box plots show medians, 17–83% & 5–95% CIs from frequentist likelihood ratio method (SRLR).

For the original ECS estimates, medians and ranges are identical to those from Bayesian posterior credible intervals.



Log y-scale used, to make it easier to see differences between the priors at high ECS.

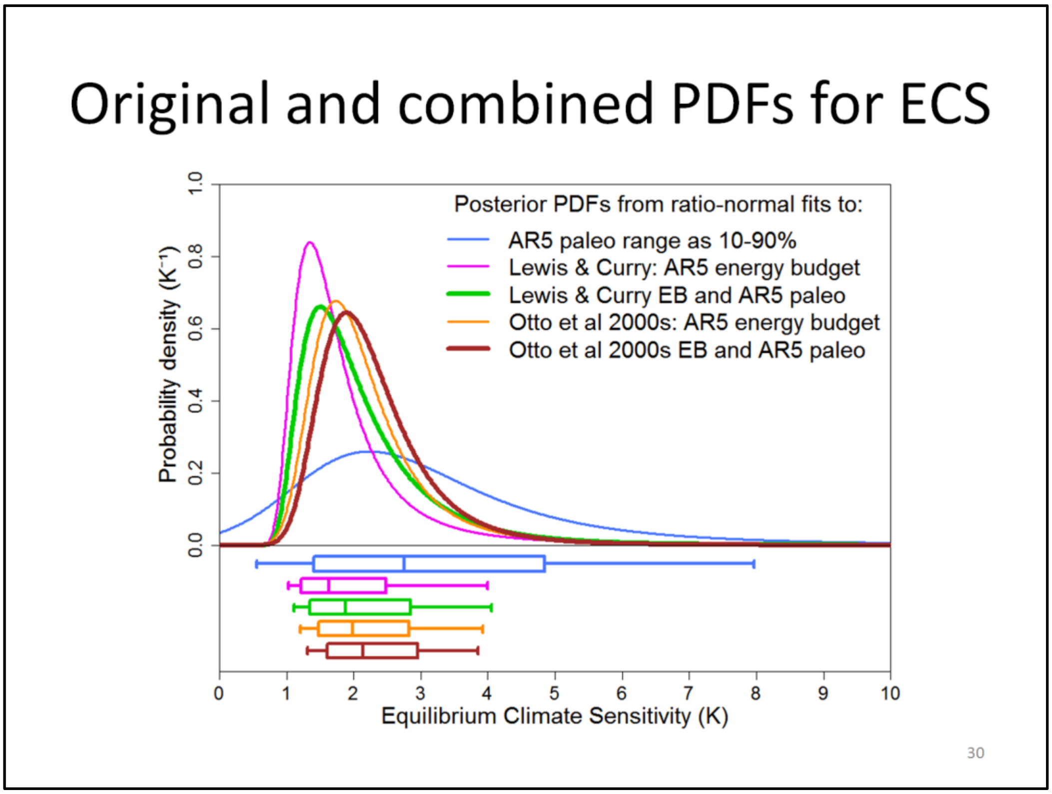
Jeffreys prior for paleo relatively flat over 0–5 K. JP for instrumental estimates peaks at very low ECS.

All priors decline - eventually with ECS^{-2} - beyond some ECS value, as data becomes increasingly uninformative about ECS.

The combined priors (brown and green lines) reflect the total informativeness of the combined data.

At low ECS, the instrumental data (especially Lewis & Curry) is much more informative than the paleo data, which has much higher fractional uncertainty in the numerator normal.

At high ECS, the paleo data is more informative than the instrumental data (especially Lewis & Curry), which has higher fractional uncertainty in the denominator normal; small ($\Delta F - \Delta Q$) range just above zero corresponds to ECS varying from highish to infinity



Adding evidence from the higher median but less constrained paleo ECS estimate slightly increases the instrumental estimate percentile ECS points below 95%; beyond there they sharply decrease.

Bayesian estimation: Conclusions

- Use Objective not Subjective Bayes if data weak
- Need to derive & compute noninformative prior
- Can infer 'true' data values & change variables
- Add Jeffreys priors[2] to combine info: don't update
- Represent prior knowledge by data likelihood+NIP

31

Subjective Bayes not valid for scientific parameter estimation, although strong data will dominate an informative prior. (If finite parameter space, SB normally gives same results as OB $\because$ NIP then always uniform.)

May be easier to find NIP and use Bayes theorem in data space, then convert the posterior data PDF to the parameter space by carrying out a change of variables (dimensionally-reducing if necessary).

Don't use updating, as it is invalid for Objective Bayesian parameter estimation.

So any existing knowledge must be represented by a likelihood function & related NIP, not a prior PDF.

Signed-root likelihood ratio method gives good uncertainty bounds if a close to normal model or transformation thereof is involved.

Thank you for listening

Nicholas Lewis

References

Lewis N, 2016. Combining independent Bayesian posteriors into a confidence distribution, with application to estimating climate sensitivity. JSPI (accepted)

Bernardo J, 2009. Modern Bayesian Inference: Foundations and Objective Methods.

Lewis N, 2013. An objective Bayesian improved approach for applying optimal fingerprint techniques to estimate climate sensitivity. J Clim, 26, 7414-29

Lewis N, 2013. Modification of Bayesian Updating where Continuous Parameters have Differing Relationships with New and Existing Data. *arXiv*:1308.2791

Lewis N, Curry JA, 2015. The implications for climate sensitivity of AR5 forcing and heat uptake estimates. Clim Dyn, *45, 1009–1023* DOI 10.1007/s00382-014-2342-y

Otto A et al, 2013. Energy budget constraints on climate response. Nat Geosci, 6, 415-6

32

Contact: nhlewis@btinternet.com