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

Nicholas Lewis
Independent climate scientist

Reading University 21 October 2016

Thank you for inviting me; its a pleasure to be here today.

If anyone finds anything unclear, please interrupt me – I know that for most of you statistics is not a specialism.

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

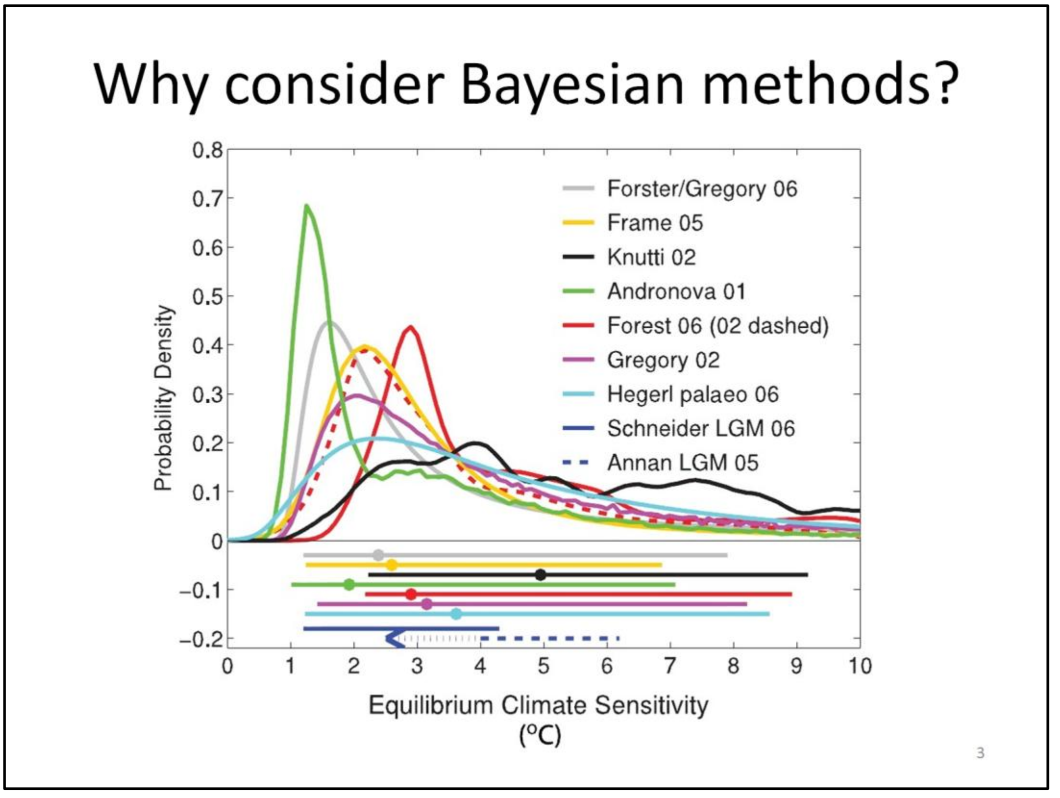
2

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

I focus on Bayesian methods, concluding that Subjective ones are unreliable, but better Objective ones require more effort

I start by contrasting Subjective and Objective Bayesian parameter estimation methods in general terms, and go on to illustrate their use in estimating equilibrium climate sensitivity or ECS.

I then show how correctly to combine evidence, in the simple case where it comes from two independent datasets.



This plot from IPCC AR4 shows estimated probability densities for ECS. These only arise in a Bayesian statistical framework
Sadly, most of the studies used a Subjective Bayesian method – an unsuitable approach for estimating ECS, as I will show.

Probability densities for ECS have practical uses; e.g., as an input when computing optimal carbon pricing.

If uncertainty ranges and median estimates are sufficient, then frequentist methods, deriving CIs, are often easier & safer.

Probability is not a settled field

- Probabilistic inference has a troubled history
- Deep disagreements over fundamental issues
- A few main belief sets and multiple variants
- Bayesian inference disdained for much of 20^{th} C
- Subjective Bayes now strong: suits computers

4

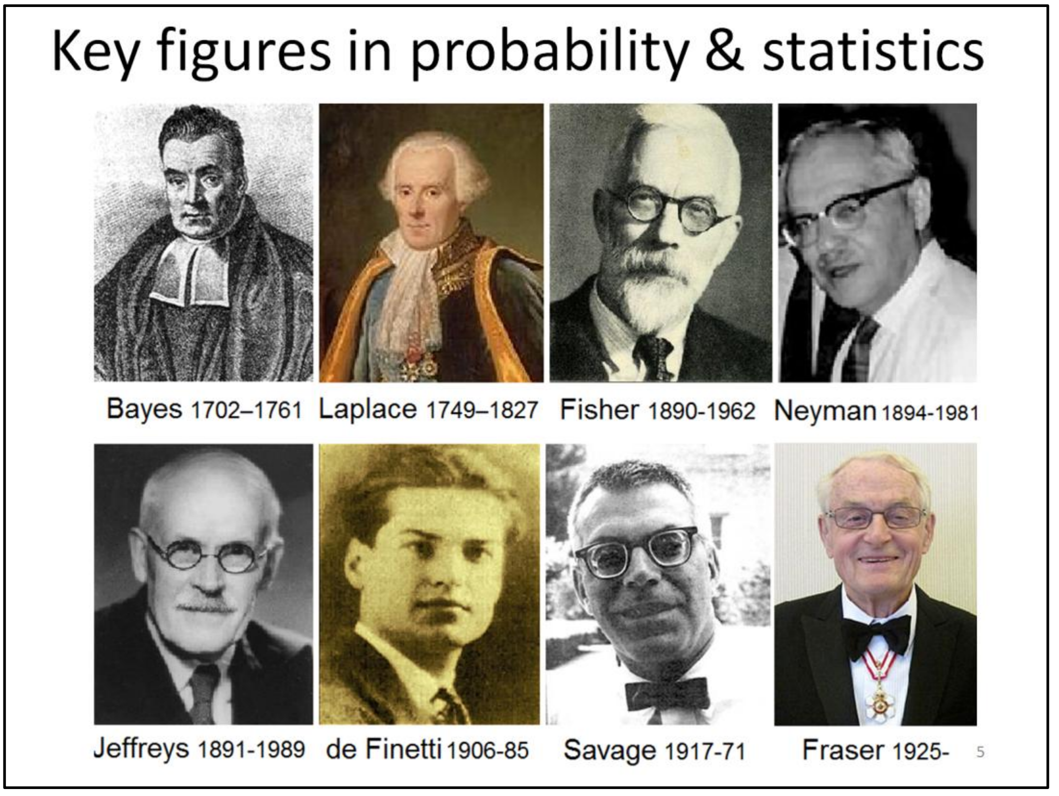
The important statistician Bradley Efron wrote (1998) of the “250-year search for a dependable objective Bayesian theory”

Still disagreements over fundamentals of probability and doubts over applicability of supposedly proven theorems.

Main belief sets are frequentist and Bayesian (subjective dominant, also objective – and others, e.g. empirical B, linear B)

Bayes theorem oscillated between boom and bust; its use is still controversial.

Computing power has made it easy to apply powerful Bayesian methods, so they 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); neither frequentist nor Bayesian

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

Jeffreys – father of objective Bayesianism: developed first noninformative prior for a estimating fixed parameter

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

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

Fraser – Deep insight into probabilistic parameter estimation + how to improve B & likelihood-based 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 $\int p(\theta\,|y)d\theta = 1$
- *a*%–*b*% range: posterior CDF credible interval; CrI

6

Bayes in a nutshell:

Bayes theorem involves multiplication by a prior to convert Likelihood (a probability density for the observed data, expressed as a function of the parameter value) 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. Often a fairly diffuse but bounded PDF is used.

Posterior PDF is normalised so that its ∫, the cumulative distribution function (CDF) reaches 1; CDF gives Cred Intvl 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 <u>random with known prior PDF</u>
- <u>Not</u> satisfied for a fixed but unknown parameter
- Subjective Bayes provides coherent personal beliefs about parameters, **not** scientific validity

7

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. Bayes' original example involved a RV.

If parameter value is fixed, prior is not a known PDF for a KRV and conditional probability does NOT apply.

Coherent: bets using your probabilities offer no arbitrage possibility. Too weak a constraint for science, which requires corrrespondence to observed reality [Gelman]. So, if you are a scientist, be very wary of Subjective Bayes.

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 Bayesian estimation

- Objective Bayes methods retain Bayesian *form*
- Objective Bayes uses 'noninformative' prior (NIP)
- NIP lets info. in data dominate; typically → CrI≈CI
- NIP is a weighting factor: no probability meaning

8

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

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

A noninformative prior doesn't indicate relative probabilities of parameter values: it is a mathematical weighting function.

NIPs are known for various cases, e.g, uniform for a parameter measured with absolute errors, otherwise have to calculate

Information theory based *reference priors*, the 'gold standard' NIPs, are noninformative about parameter values and achieve such objectivity**.**

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

9

The original NIP – Jeffreys prior – was developed to make inference invariant to a change in parameter, e.g x to x^3.

Jeffreys prior is the reference prior for all parameters jointly; it can need modifying for subsets of vector parameters

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

Jacobian-type 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.

If the data uniquely define the parameter value, can get its posterior PDF by sampling from data uncertainty distributions

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

10

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 are strong. But estimation of ECS involves weak data and highly nonlinear data-parameter relations for ECS – and also effective ocean vertical diffusivity K_v , often jointly estimated with ECS.

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

11

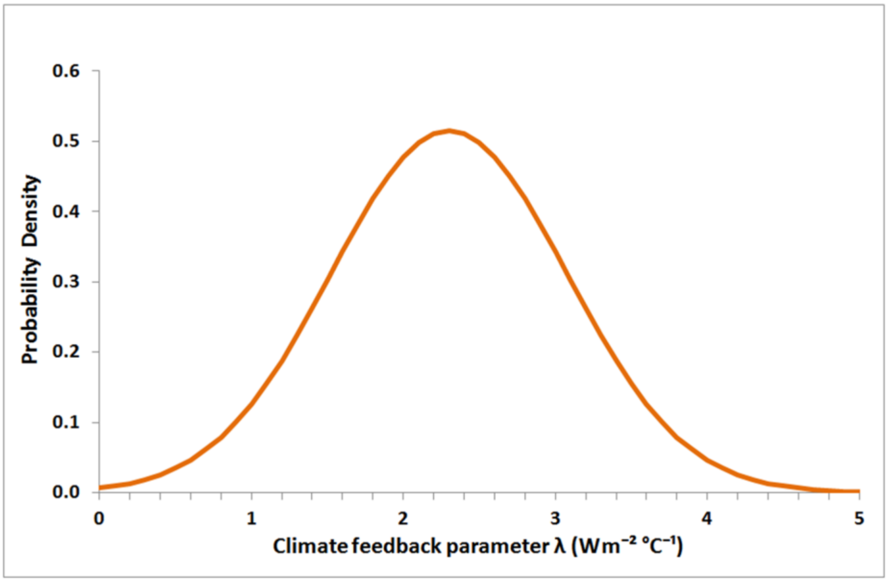
Uniform priors have typically been used, e.g. in AR4, but nonlinear relationships mean they are strongly informative about the parameters and usually result in unrealistic ECS estimation.

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



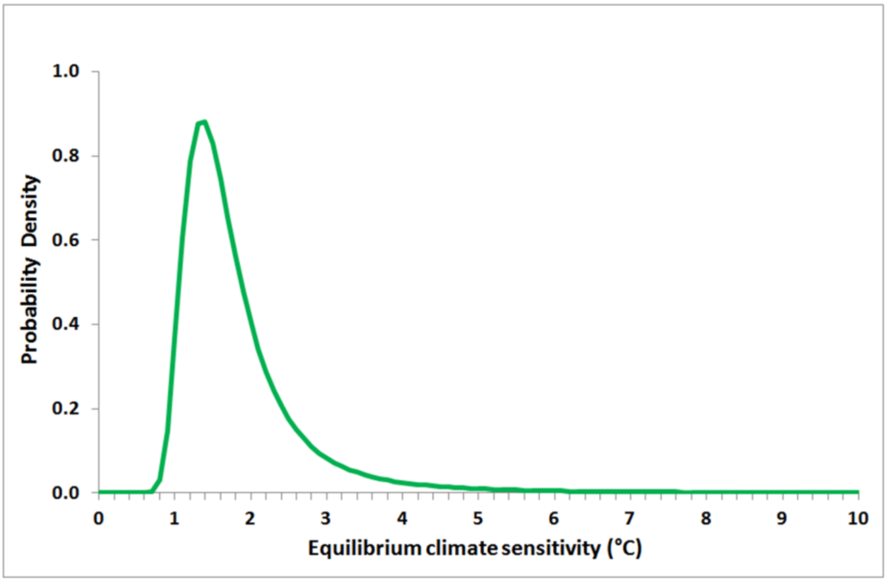
12

I'll use Forster & Gregory 2006's results to demonstrate 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 almost Gaussian uncertainty.

They noted that their regression-based estimate in effect used an almost uniform-in-λ prior, which is noninformative here.

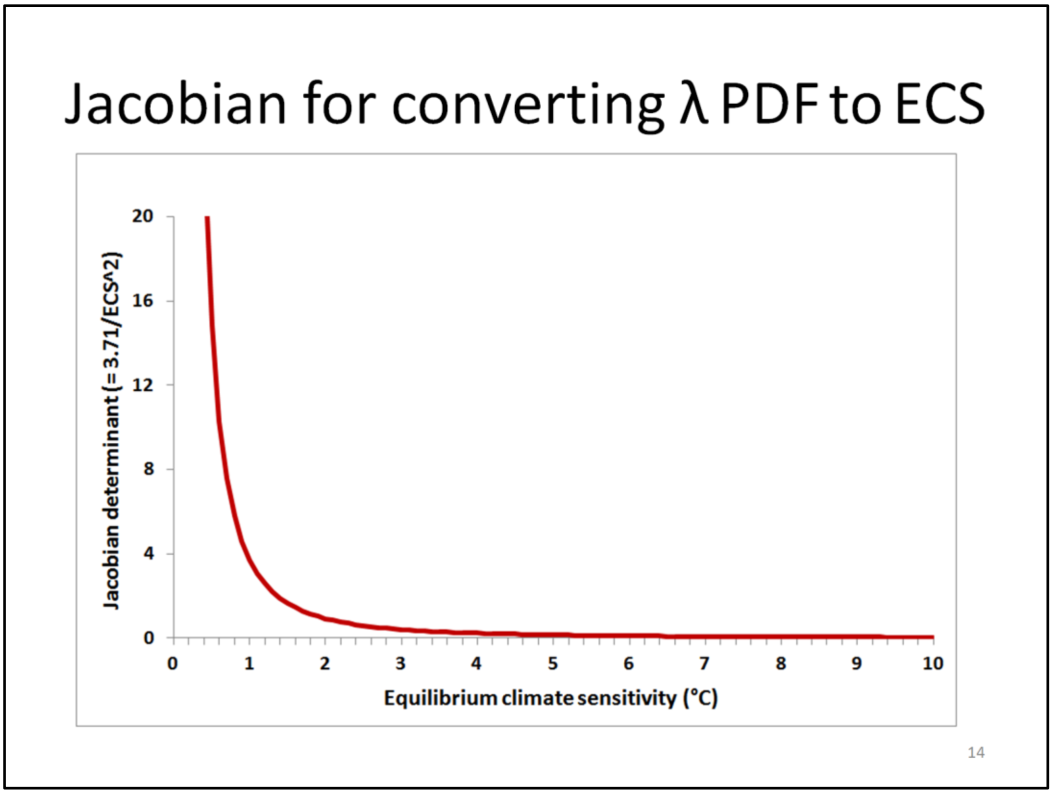
PDF converted (Jacobian) from λ to ECS



13

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, being 3.71 ECS^{-2}.

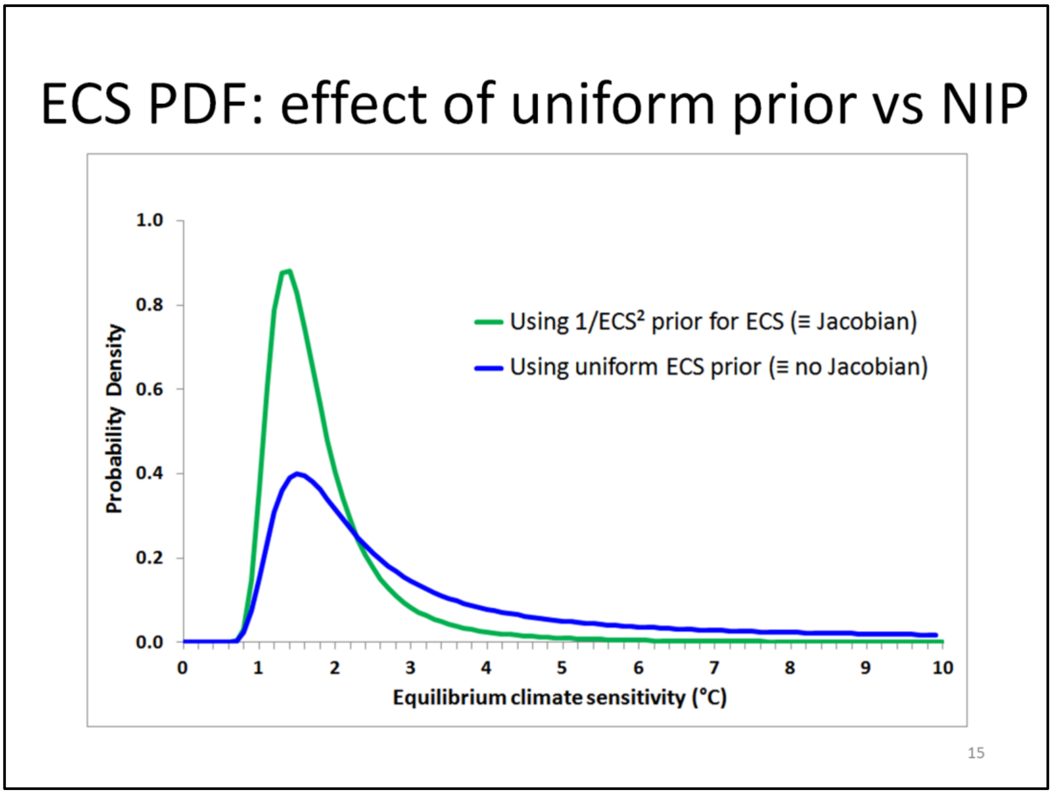
The resulting median ECS estimate is 1.6 K; 95% bound is 3.6 K (per IPCC AR5 uniform-in-λ prior based PDF)



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. Jeffreys prior is always the best NIP when a single parameter is being estimated.



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

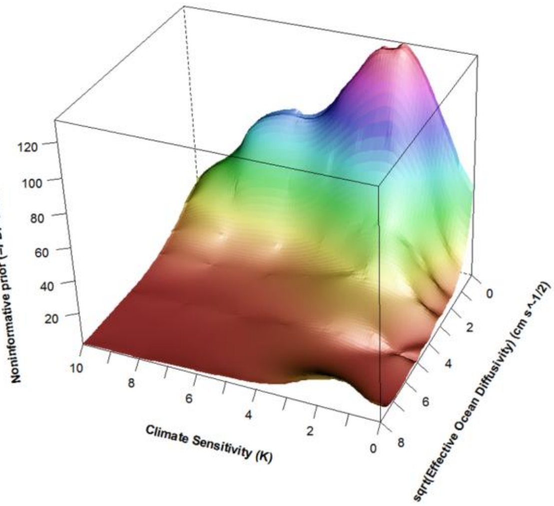
Resulting median estimate was over 60% higher at 2.65 K; 95% bound was nearly quadrupled at 14.2 K.

(Both median and 95% bound 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 a Jacobian factor to convert the PDF.

On a change of variable, a noninformative prior transforms like a normal PDF, so the NIP for ECS is proportional to ECS^{-2}.

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



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

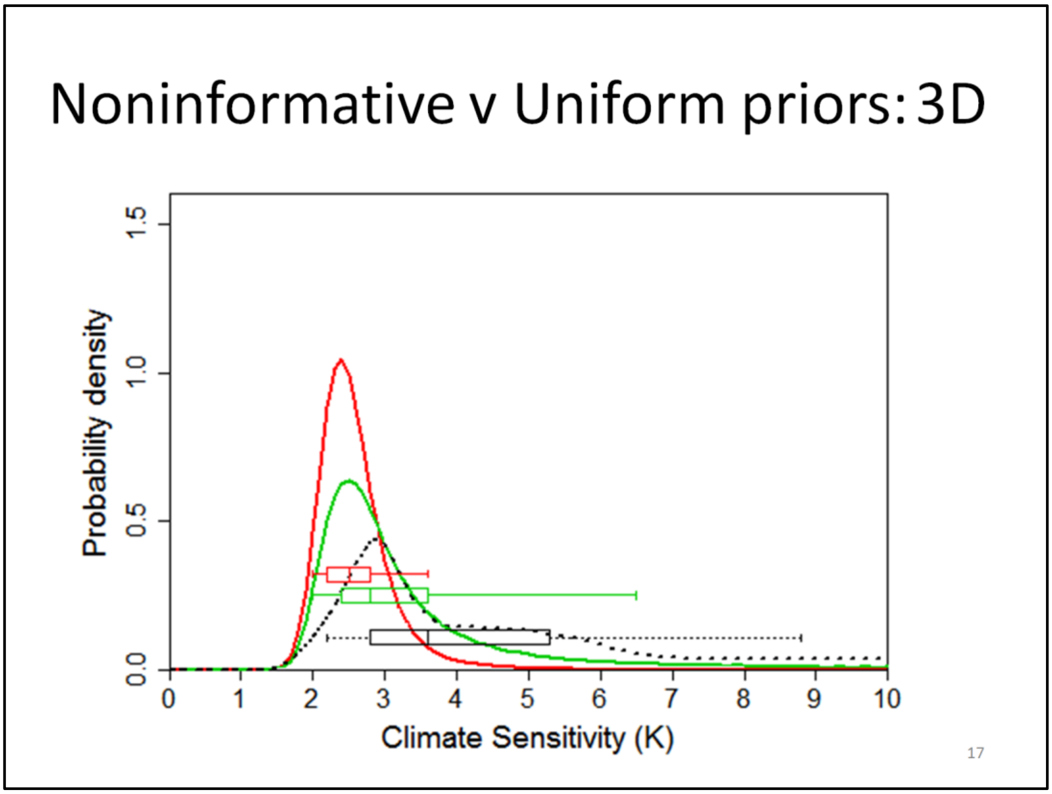
Now an example of a NIP involving 3 estimated parameters. The plotted joint prior for ECS and sqrt(K_v) is averaged over aerosol forcing, with which its shape varies little. Computed by numerical differentiation of data-parameter functions.

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 $\neq$ the product of 3 separable individual priors

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



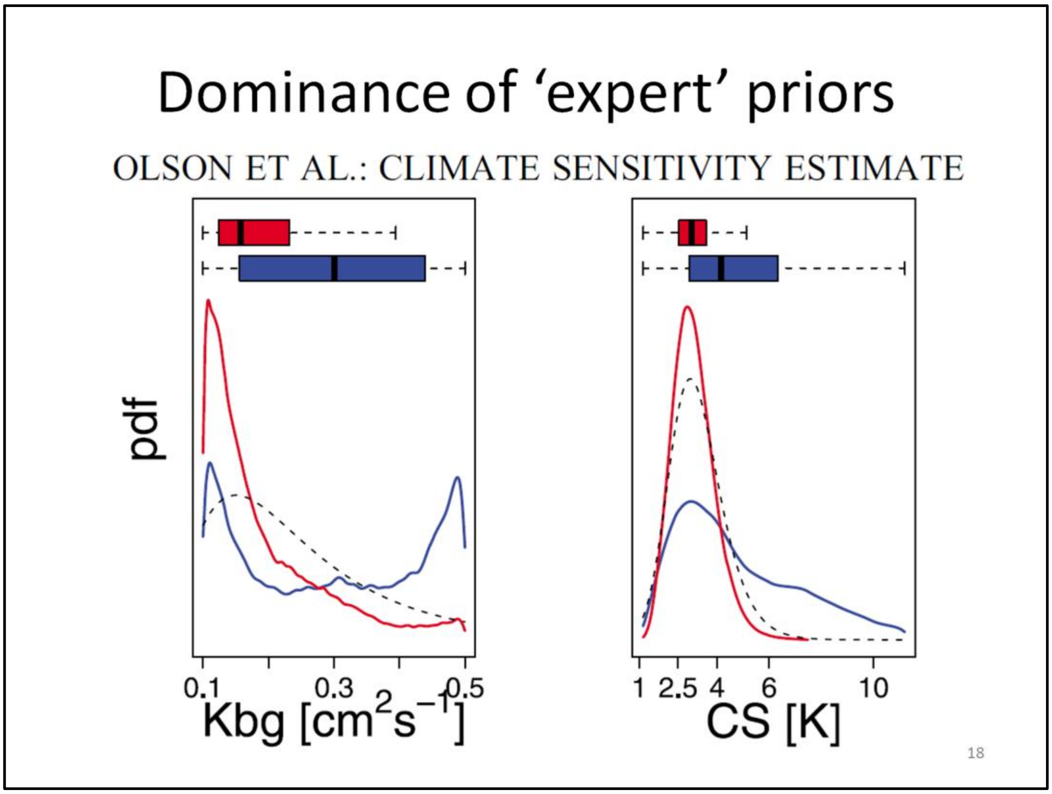
Posterior PDFs from Lewis 2013 using the original (Forest et al 2006) data/diagnostics.

Red= Noninformative 3D joint Jeffreys prior

Green= Uniform prior in ECS, sqrt(Kv) and Faer

Black= as reported in Forest 2006 and IPCC AR4; it also had several other statistical and coding errors

Box plots show medians and 25–75% & 5–95% ranges



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.

Such parameter estimates are scientifically valueless.

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.

19

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

All instrumental data well suited to estimating ECS are non-independent, so use one good representative estimate.

Paleo evidence is (one hopes) independent of instrumental period evidence, but mainly comes from LGM.

Updating is the standard Bayesian way to combine evidence: use existing posterior PDF as the prior, & new data likelihood.

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

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

Bayesian updating is unsatisfactory

- NIPs likely to differ for each source of evidence
- Objective Bayes + updating: order-dependence!
- Π(likelihoods) fixed, but prior for 1st est changes
- Objective Bayes & Bayesian updating incompatible

20

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 not valid for objective estimation of fixed parameters.

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

21

One can avoid unsatisfactory Bayesian updating by instead computing a noninformative prior to use with the combined-evidence data likelihood, in a single application of Bayes theorem. Covered in my papers: not seen it elsewhere.

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, must represent any existing knowledge by a likelihood function and related NIP, not by 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 NIP known
- ECS ~$\propto$ ratio of two normals: ΔT / (ΔF – ΔQ)
- Good approximation to this distribution: RS93
- Method gives CrI=CI for combined ratio-normals

22

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

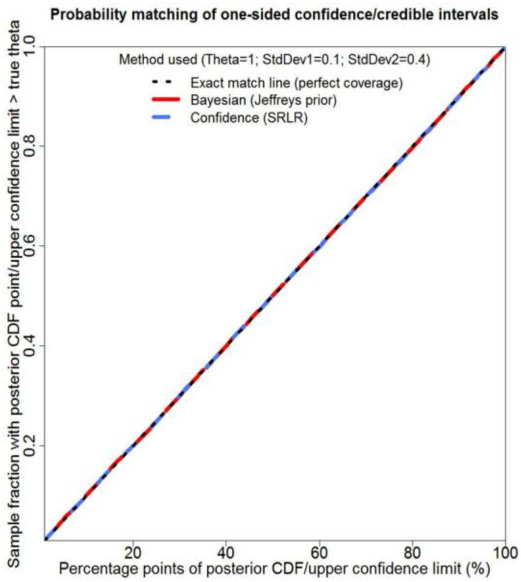
ECS energy budget equation; uncertainty in ΔT and (ΔF – ΔQ) is ~ Normal.

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 [ΔF – ΔQ]) / sqrt(sigma $_{\Delta T}{}^2$ + ECS^2 sigma $_{\Delta F}{}^2$) } ~ N(0,1)

Testing RS93-based probability matching

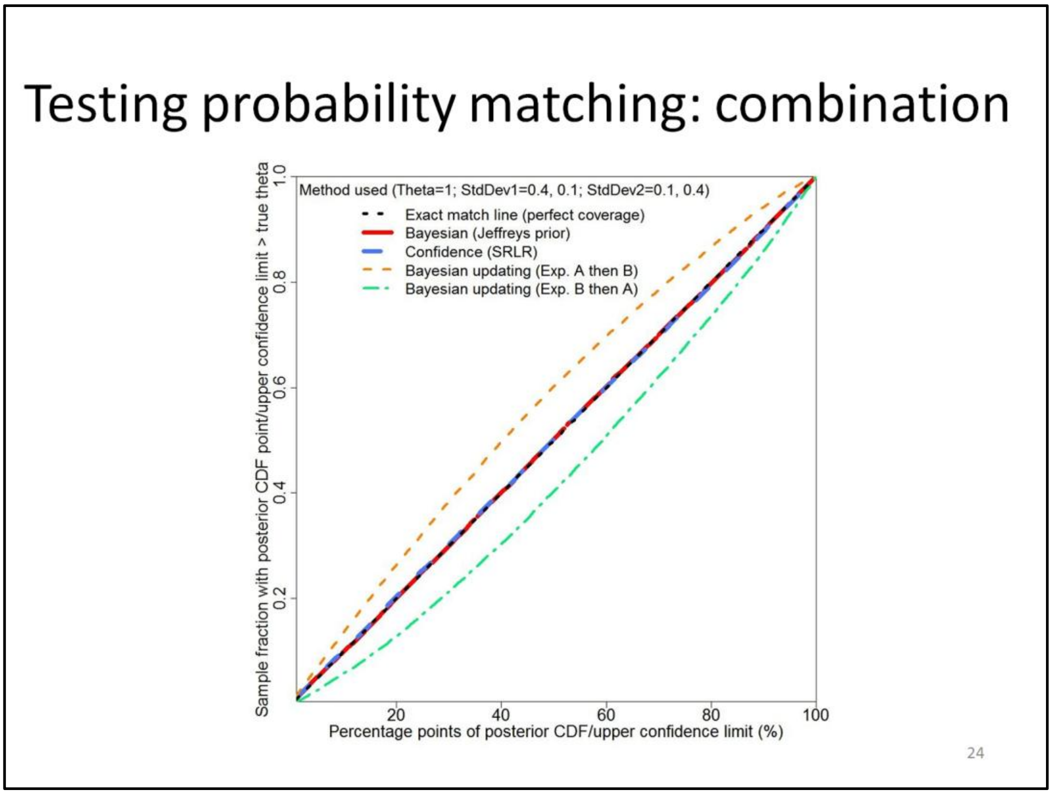


23

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

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.



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).

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.

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

Combining recent & paleo evidence

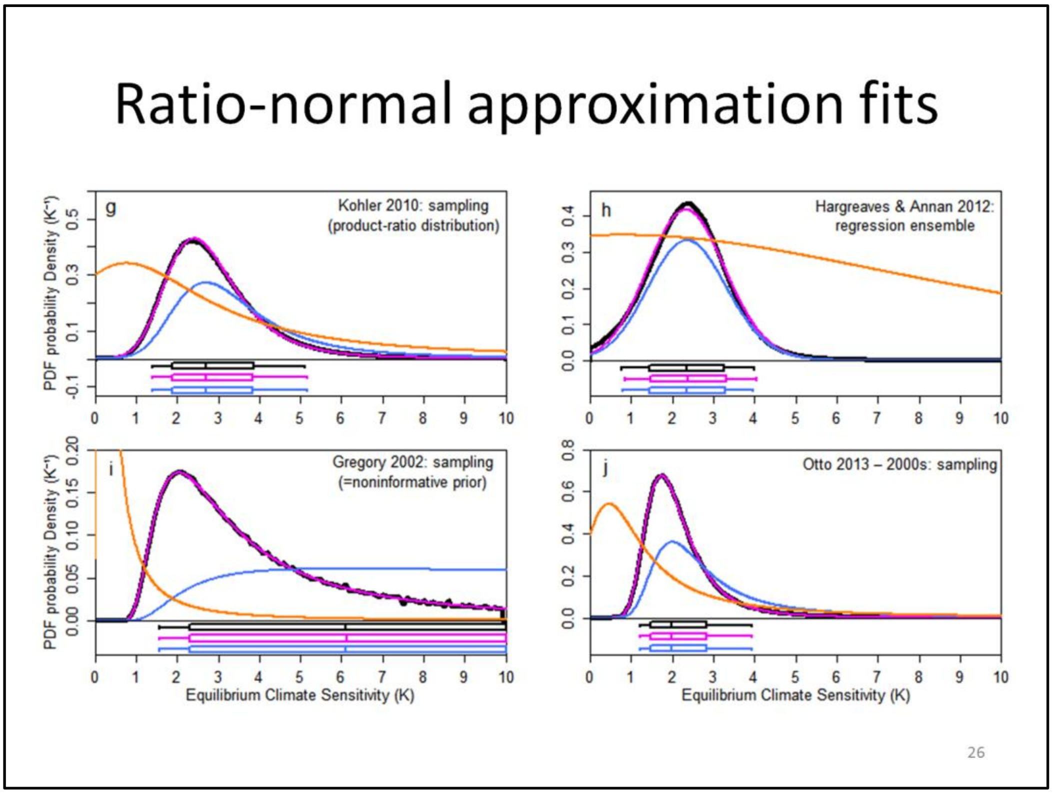
- Represent each ECS estimate by RS93 formula
- RS93 3-parameter formula fits ECS PDFs well
- Median and range suffice to uniquely specify

25

Ratio-normal distribution has 3 free parameters.

So gives a unique fit to median + range.

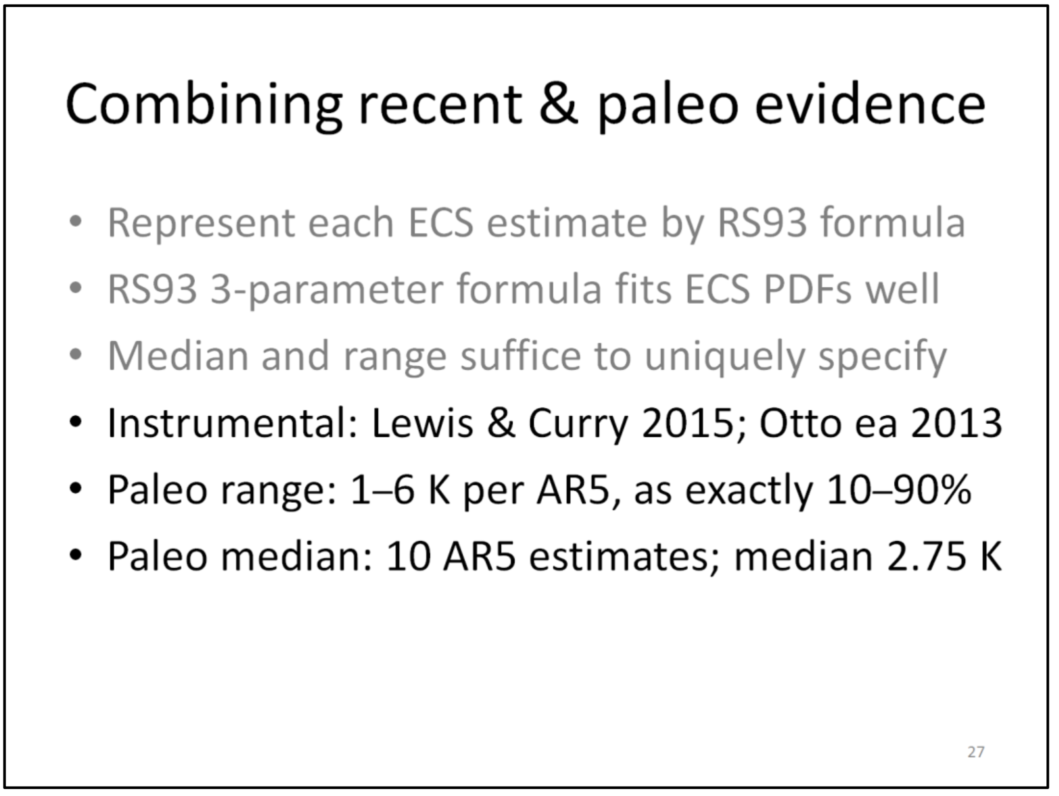
Can fit ECS PDFs well provided they don’t use a highly informative prior.



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

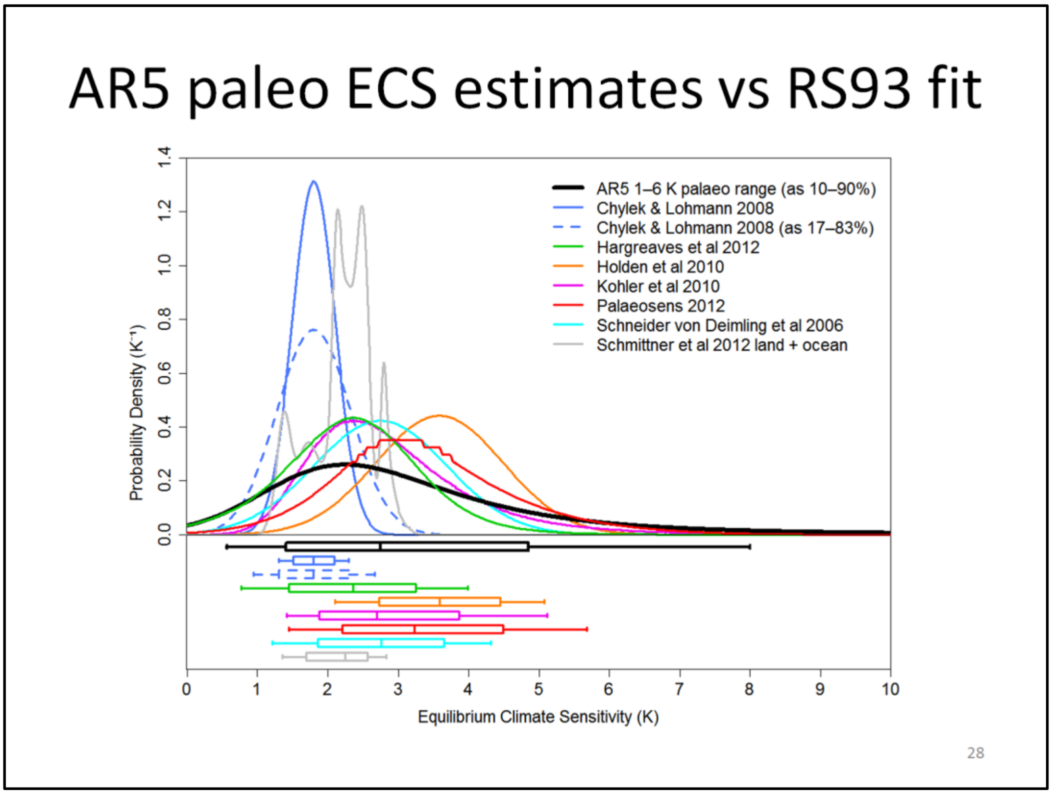
Representative objectively-based estimated PDFs for ECS: top – paleo; bottom – instrumental period.

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

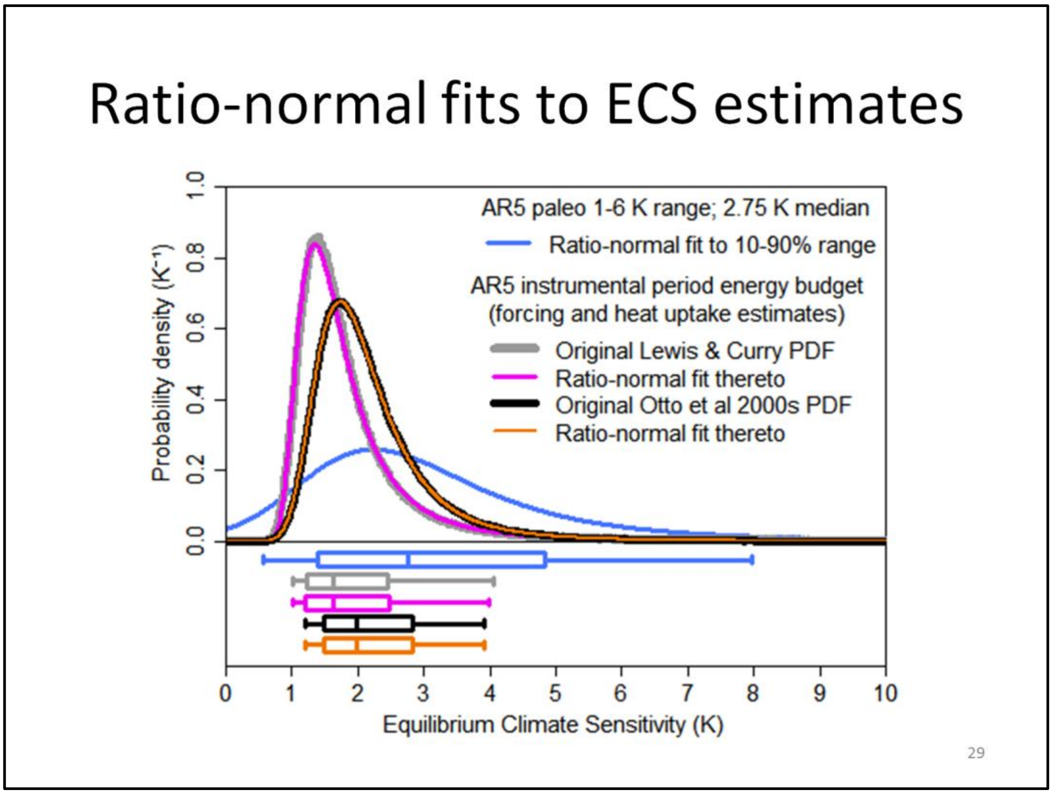


Use two alternative instrumental period energy budget estimates that reflect AR5 forcing and heat uptake data.

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



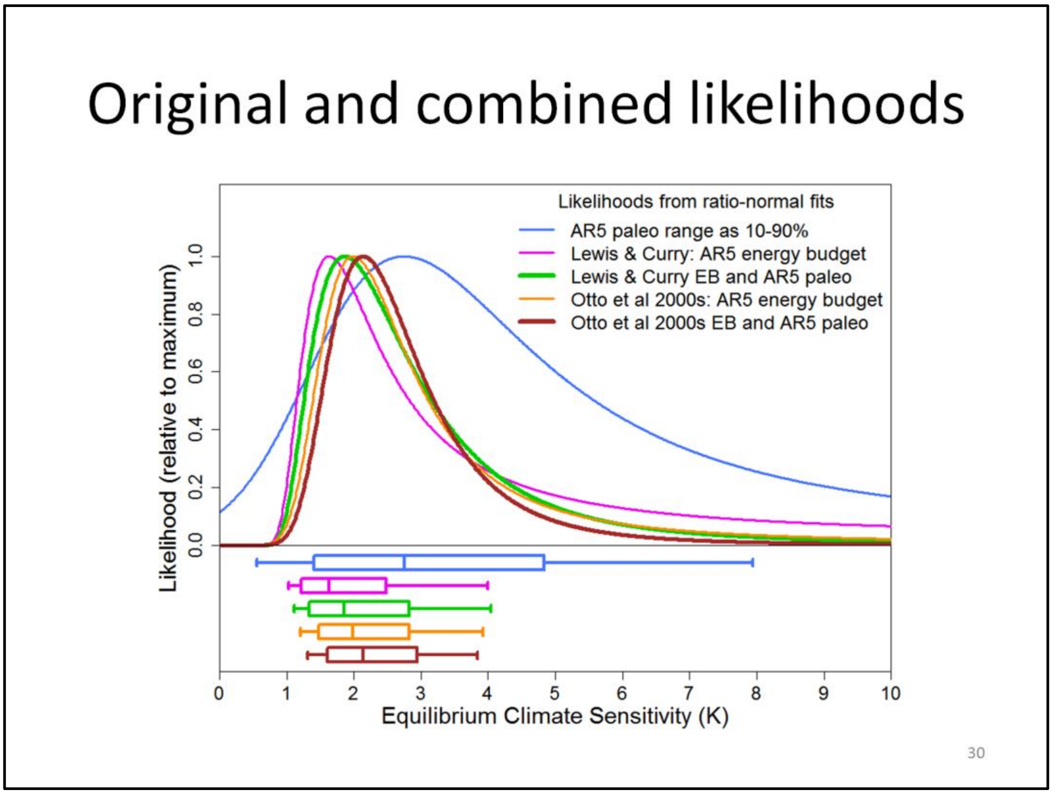
The black fitted RS93 PDF has the same 2.75 K median as the overall median of all the paleo estimates given in IPCC AR5.



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

Blue line is the fitted paleo evidence PDF.

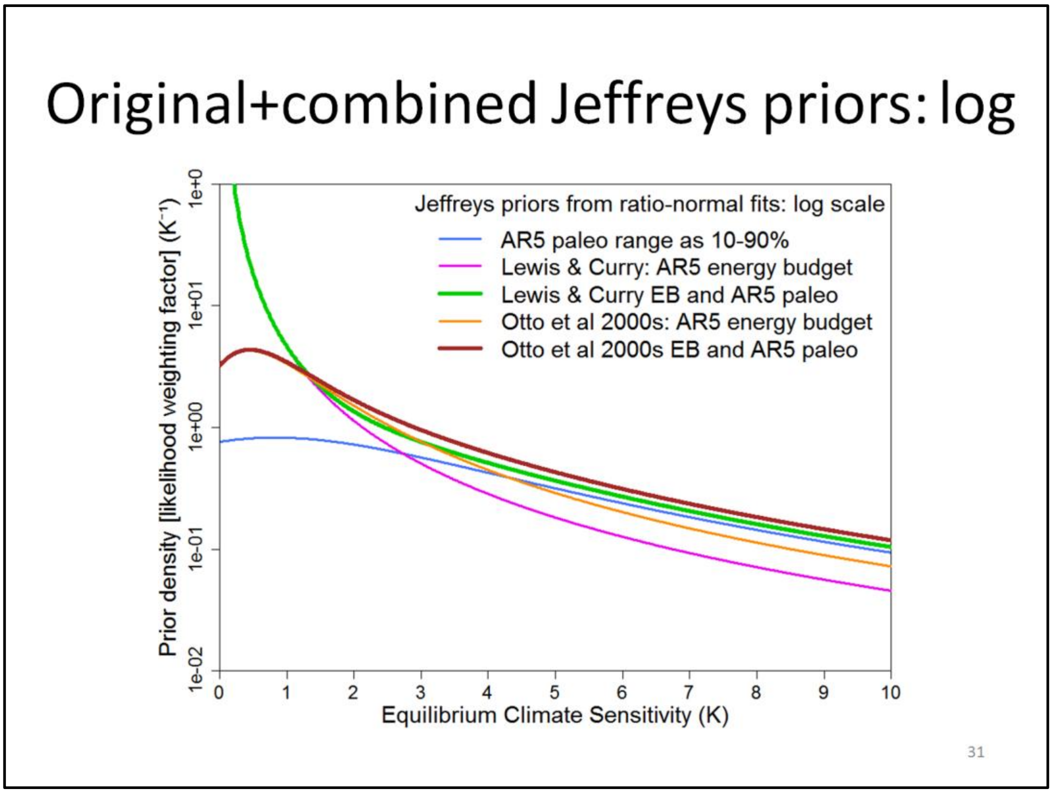
Box plots show medians, 17–83% and 5–95% credible intervals from PDFs.



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 the frequentist likelihood ratio method (SRLR).

For the original ECS estimates, SRLR 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.

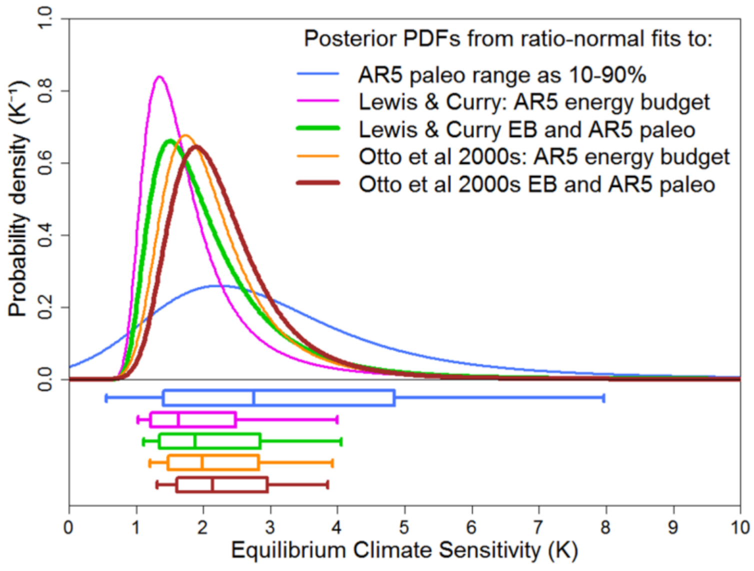
All priors decline – eventually with ECS^{-2} – beyond some ECS value, as data becomes increasingly uninformative about ECS.

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

Original and combined PDFs for ECS



32

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

33

Subjective Bayes not valid for scientific parameter estimation, although strong data will dominate an informative prior.

Either use OB or a frequentist method: signed-root likelihood-ratio good if model is close to normal, maybe transformed.

May be easier to find NIP & use Bayes theorem in data space, then convert the posterior data PDF to the parameter space.

Don't use Bayesian updating, as it is generally not valid for Objective Bayesian parameter estimation.

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

A dimensionally-reducing PDF conversion for a change from data to parameter space can be carried out if there are more data

Thank you for listening

Nicholas Lewis

niclewis.wordpress.com/objective-bayesian-parameter-inference-and-combining-ecs-evidence

References

Lewis N, 2016. Combining independent Bayesian posteriors into a confidence distribution, with application to estimating climate sensitivity. J Stat Planning & Inference (accepted)

Lewis N and P Grunwald. Objectively combining AR5 instrumental period and paleoclimate climate sensitivity evidence. (Submitted, 2016)

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

34

Contact: nhlewis@btinternet.com