Valuing science policy: Dynamic decisionmaking with generalized Bayesian learning

Ivan Rudik – Cornell University
Maxwell Rosenthal – Georgia Tech
Derek Lemoine – University of Arizona and NBER

TWEEDS

Also known as..

Ivan Rudik – Cornell University
Maxwell Rosenthal – Georgia Tech
Derek Lemoine – University of Arizona and NBER

TWEEDS

How to approximate and integrate high dimensional objects quickly and accurately

Ivan Rudik – Cornell University
Maxwell Rosenthal – Georgia Tech
Derek Lemoine – University of Arizona and NBER

TWEEDS

Lots of economics research on mitigation and tech R&D, little research on science policy

Monitoring and modeling are major components of climate science

Lots of economics research on mitigation and tech R&D, little research on science policy

Monitoring and modeling are major components of climate science

Current expenditures: $2-3 billion/year, almost 50% of US climate change expenditures (GAO 2018)

Lots of economics research on mitigation and tech R&D, little research on science policy

Monitoring and modeling are major components of climate science

Current expenditures: $2-3 billion/year, almost 50% of US climate change expenditures (GAO 2018)

What is the marginal benefit of funding climate science?

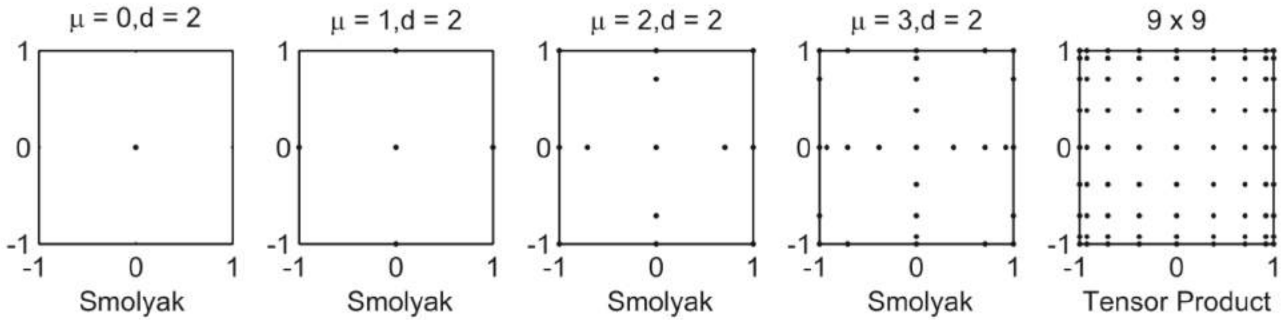
We estimate the value of science policy that helps us learn about climate sensitivity

(sort of) new computational methods we use:

We estimate the value of science policy that helps us learn about climate sensitivity

(sort of) new computational methods we use:

Sparse grids

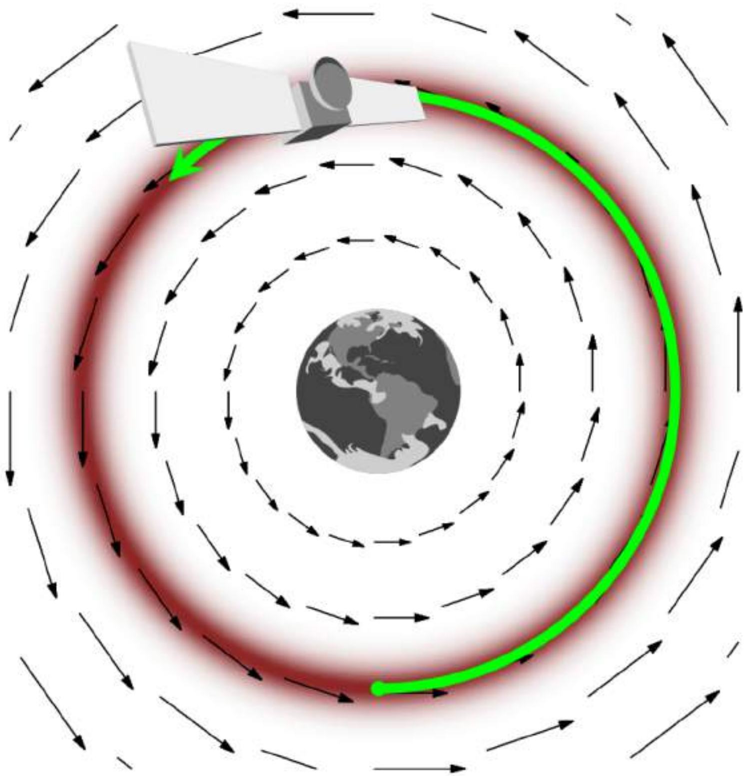


We estimate the value of science policy that helps us learn about climate sensitivity

(sort of) new computational methods we use:

Sparse grids

Hamiltonian Monte Carlo



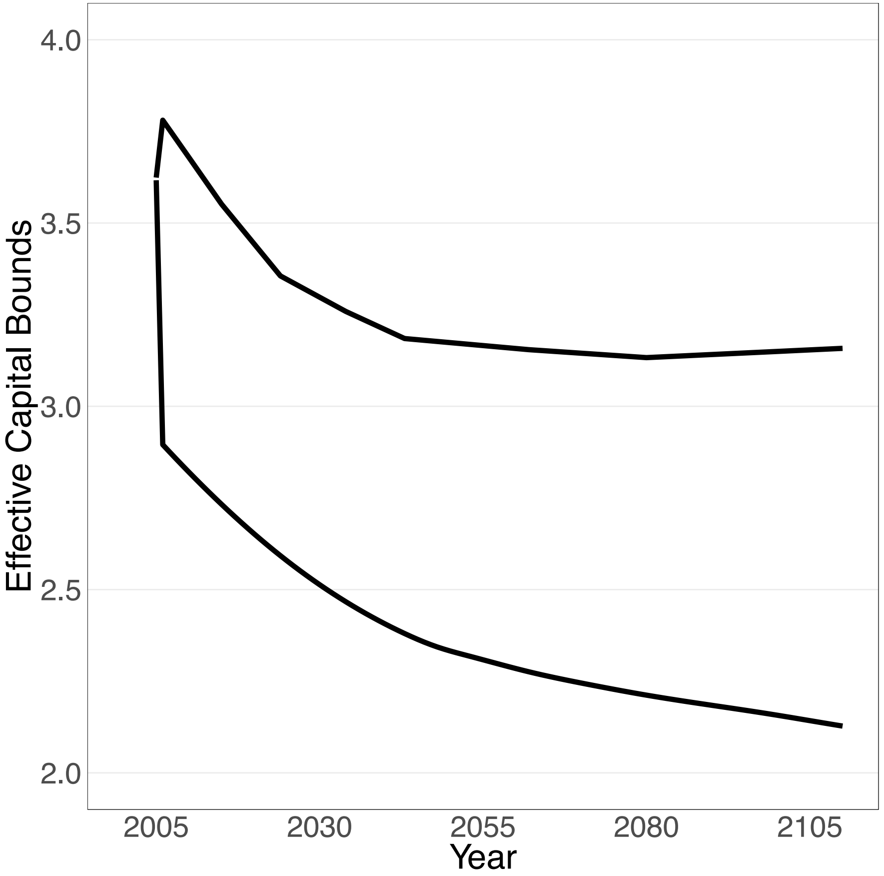
We estimate the value of science policy that helps us learn about climate sensitivity

(sort of) new computational methods we use:

Sparse grids

Hamiltonian Monte Carlo

Adaptive grids / stochastic simulation



Why do we need these methods?

Sparse grids

- 200 states

Hamiltonian Monte Carlo

- Nonconjugate distributions
- High dimensional Bayesian estimation

Adaptive grids / stochastic simulation

- 200 states

Why do we need these methods?

Why do we need these methods?

Approximate time to solve this model once with these methods:

Why do we need these methods?

Approximate time to solve this model once with these methods:

1 million core-hours

Why do we need these methods?

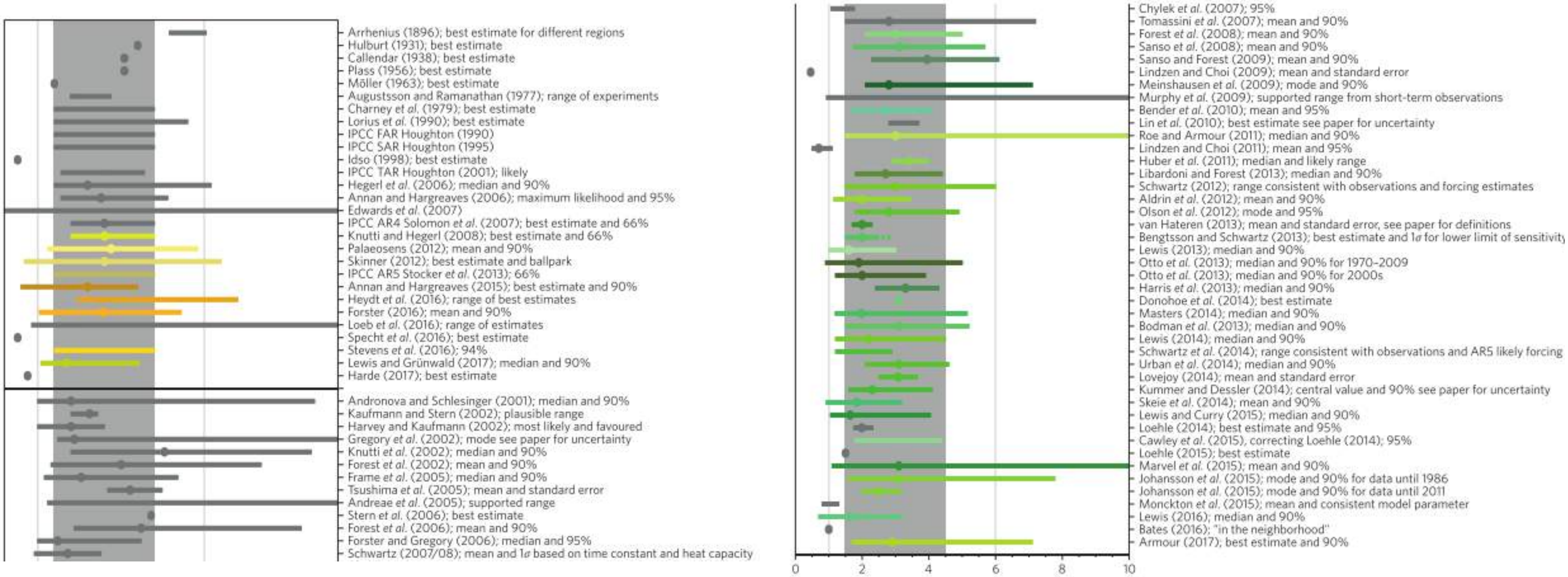
Approximate time to solve this model once with these methods:

1 million core-hours

Optimistic human time equivalent:

1 month on UA Ocelote (~1,400 cores)

Climate sensitivity: equilibrium warming from a doubling of CO_2



"The estimated range of the ECS has not changed much despite massive research efforts." - Knutti et al. (2017)

What is the value of science policy?

1) Science is **extremely** valuable

Accelerating learning is worth up to:

- $100s of billions annually
- $1000s per capita lump sum today
- 1% permanent consumption gain

What is the value of science policy?

1) Science is **extremely** valuable

Accelerating learning is worth up to:

- $100s of billions annually
- $1000s per capita lump sum today
- 1% permanent consumption gain

2) Science should be funded at **large scale**

There are increasing returns to scientific information

IAM + generalized Bayesian estimation

The model is:

- Dynamic stochastic DICE with storage of observed histories
- Bayesian estimation routine for the temperature transition

$$T_{t+10} = T_t + C1\left[F(M_{t+10}, t + 10) - \frac{f}{\mathbf{CS}} T_t - C3(T_t - O_t)\right] + \nu_{t+10}^{T}$$

We want the **CS** distribution (state of climate knowledge)

The equation doesn't admit a closed form Markov updating rule → we need to store observed histories, estimate generic posteriors

How we do it: 5 steps

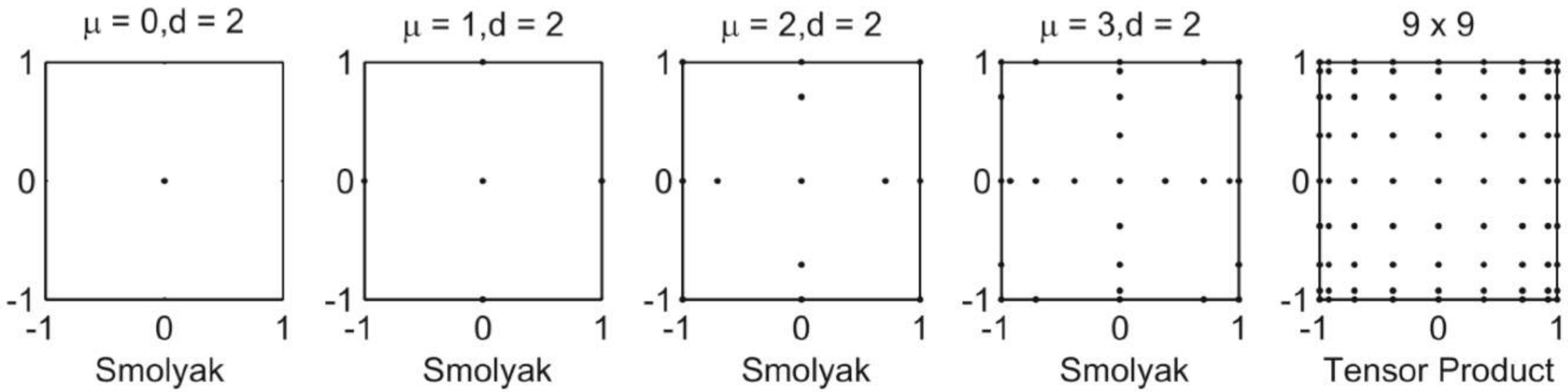
1) Collocation
2) Interpolation
3) Learning
4) Belief computation
5) Dynamic programming

Step 1: Make an approximation grid

Collocation

Taking nonconjugate learning to a climate-economy model: collocation sparsity

We use a sparse collocation grid (Smolyak, 1963)



Intuition: Optimal scheme to minimize approximation errors with a given number of grid points

Delivers: polynomial complexity instead of exponential

Why: we need to store observed histories over centuries

Step 2: Generate data

Interpolation

Issue: model is way too big on a 1 year timestep

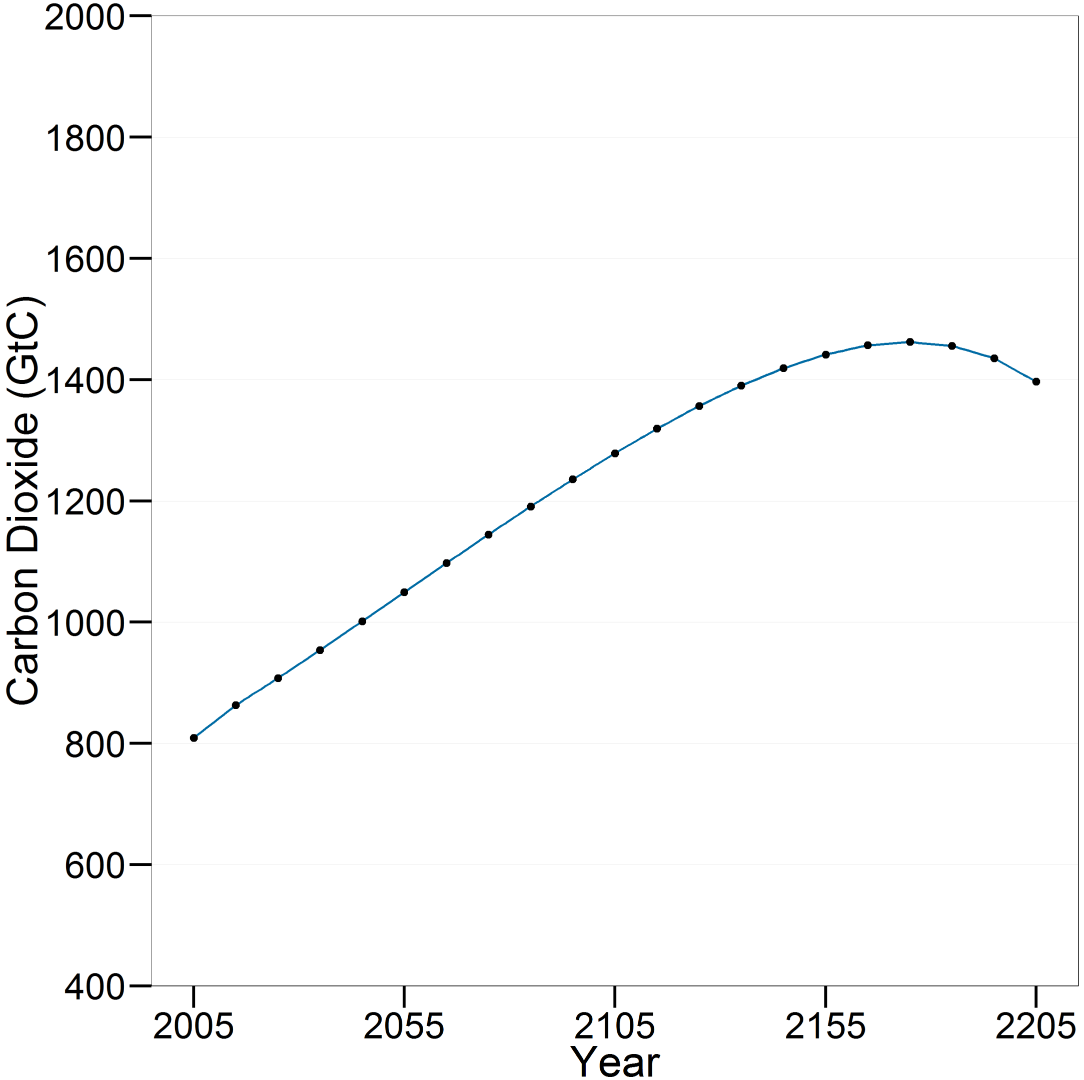
We use a 10 year time step for policymaking because of computational limitations (hundreds vs thousands of states)

This is obviously bad for learning about climate sensitivity

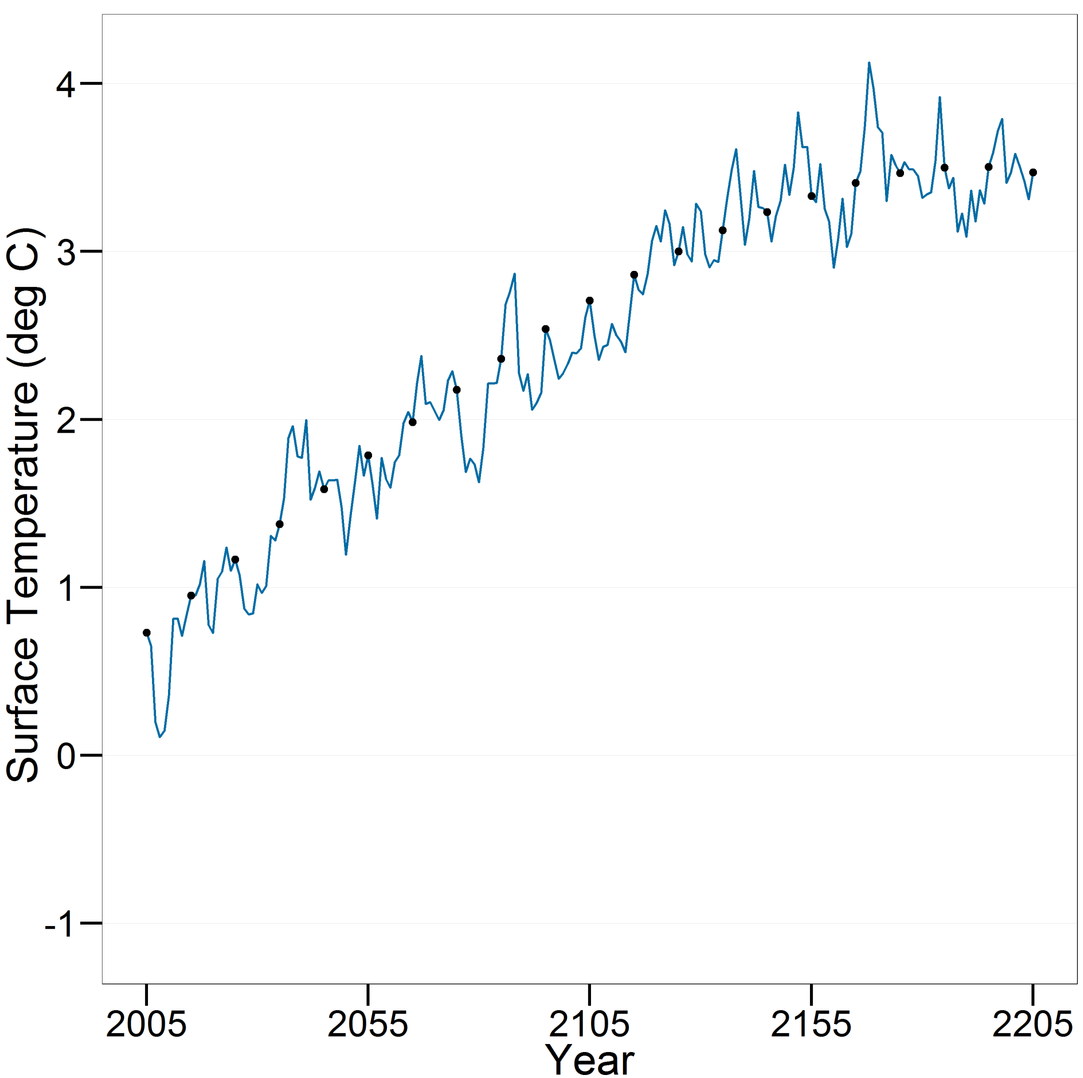
Solution: interpolate between 10 year histories using Brownian bridges + annual climate dynamics

Our interpolated history

Linearly interpolate CO_2



Bridge + annualized transition



Step 3: Bayes' rule

Learning

Step 3: Bayes' rule

Step 4: Hamiltonian Monte Carlo

Beliefs

We need a usable posterior for our model

Bayes gives us a posterior, now we need a way to approximate it to use it in the model

Hamiltonian Monte Carlo: new and efficient method for sampling high dimensional distributions

We need HMC vs MCMC because we will be estimating a ~70 dimensional posterior (CS + data + volatility) to calibrate our model (e.g. Aldrin et al. 2012; Skeie et al. 2014)

Why HMC is the coolest

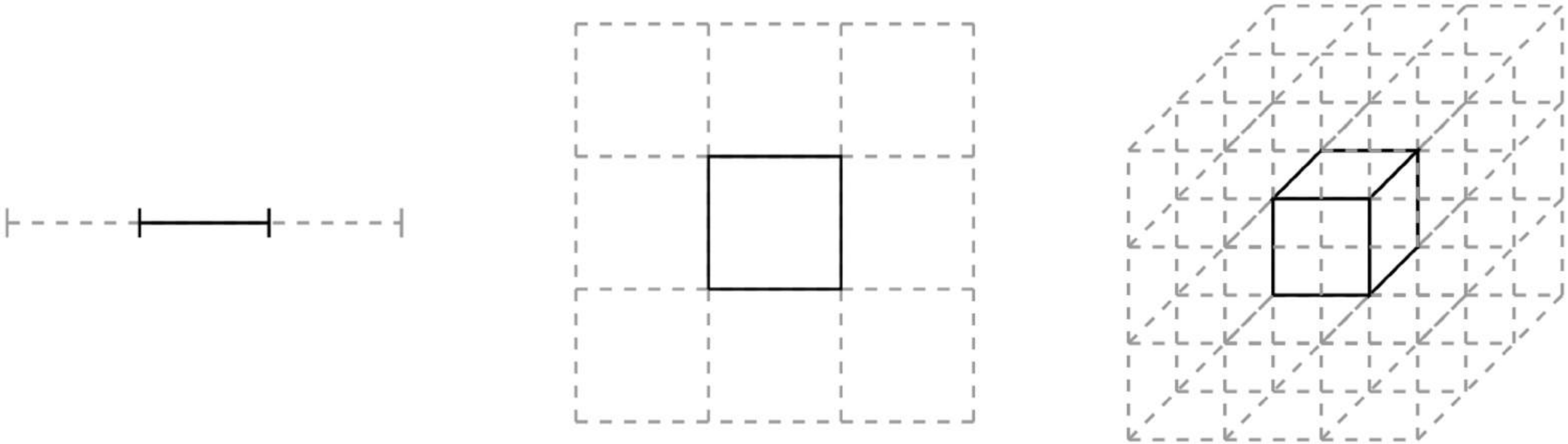
We commonly take expectations by exploring the distribution through random walks (e.g. Metropolis-Hastings)

High dimensional spaces pose problems for random walks

Don’t efficiently traverse the important parts of the distribution

Why? See Betancourt (2017) for details and pictures

Volume scales exponentially in # of dimensions



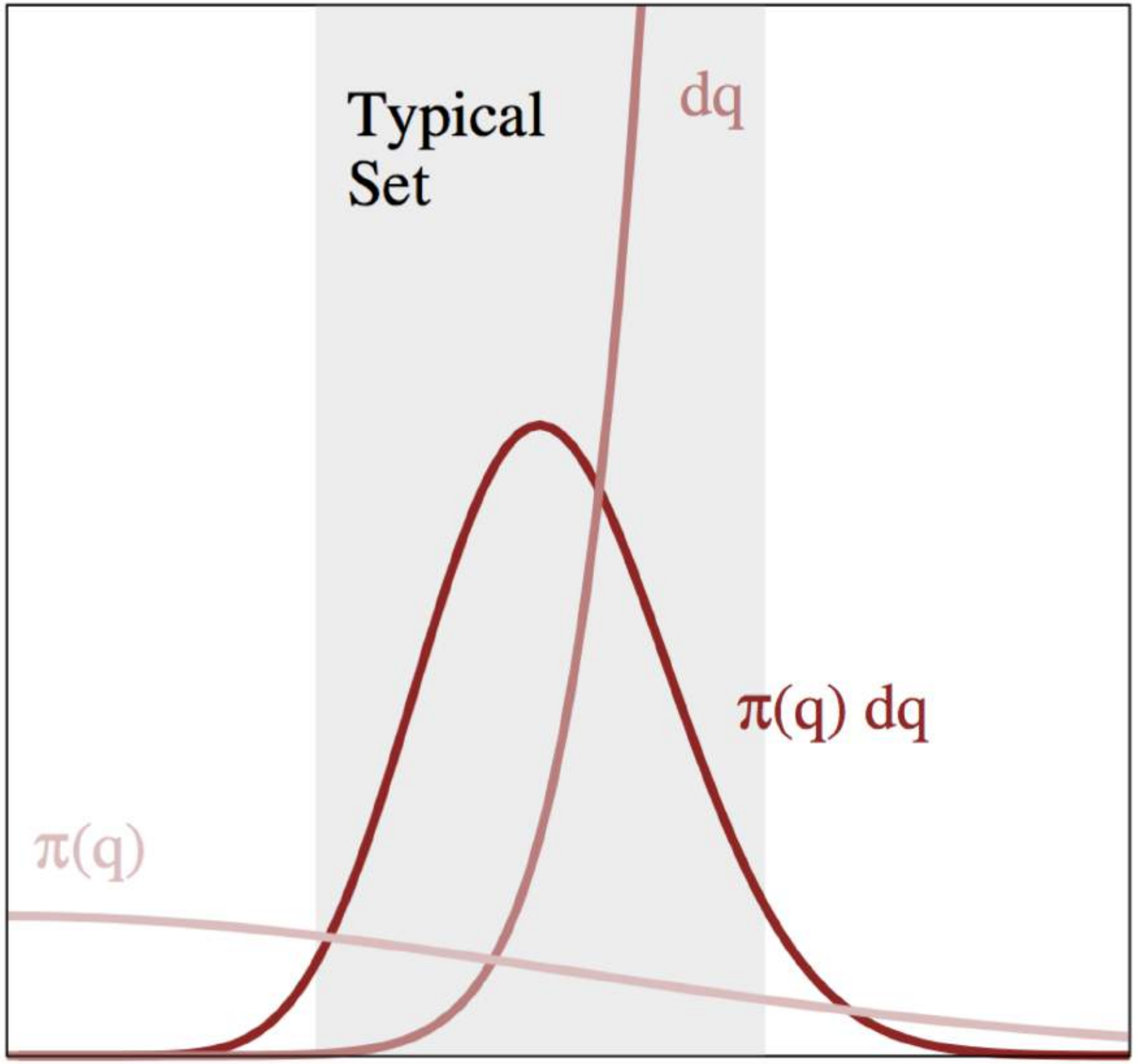
High density regions (e.g. the distributional mode) take up smaller and smaller volumes as the dimensionality increases

Density vs volume tension in high dimensions

A density concentrates around its mode but the vast majority of volume is away from the mode

Contributions to the expectation are determined by the product of density and volume

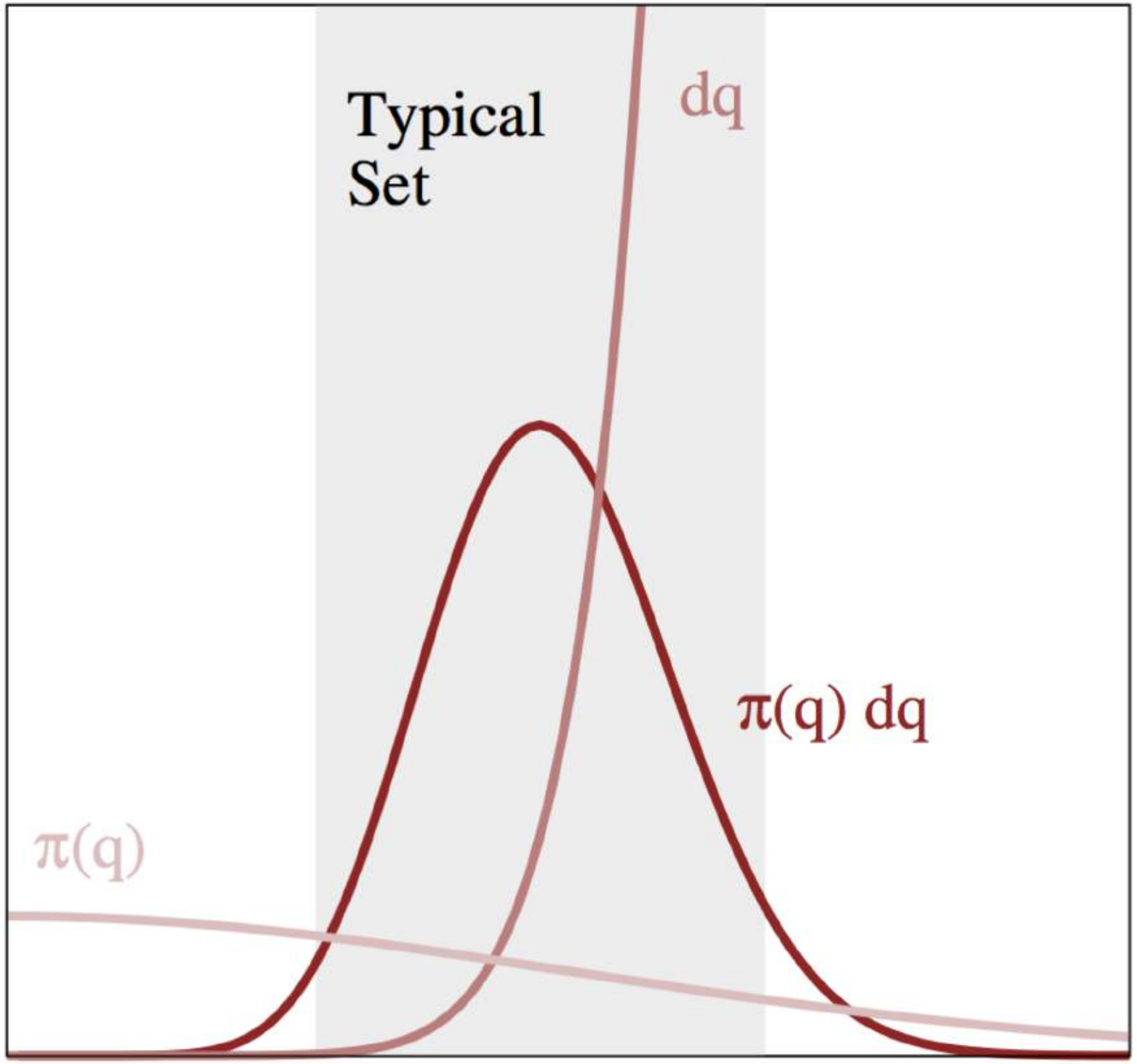
We don't need (or want) to explore the entire distribution



Density vs volume tension in high dimensions

The contributions are centered in an (small) area called the **typical set**

To efficiently take expectations we need to identify and focus on the typical set



It’s just physics

HMC takes standard MCMC approaches,

but informs transitions so they closely follow the typical set

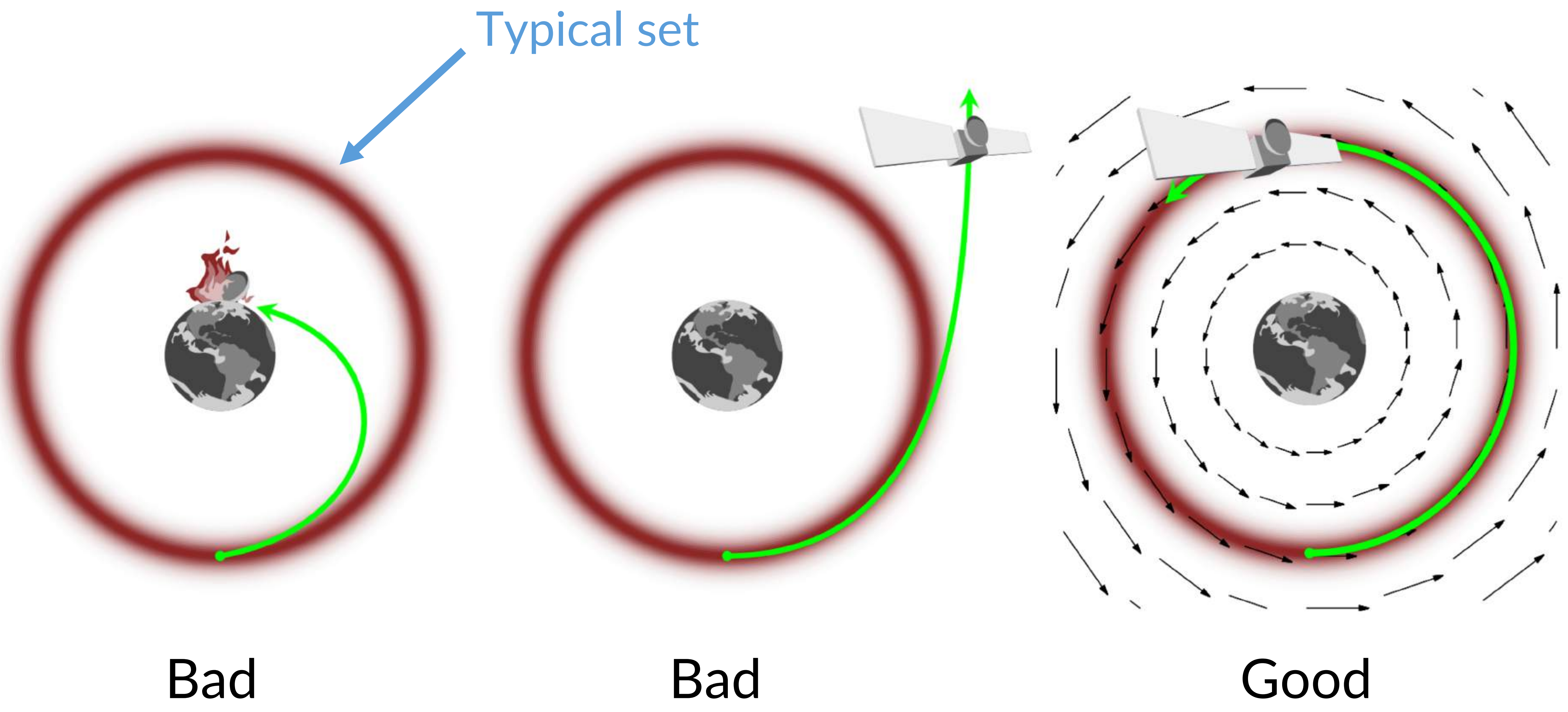
How? Ideas from classical physics

The typical set is actually very similar to stable orbits →

We need enough momentum to offset gravity (gradient’s)

pull toward Earth (the mode)

It's just physics



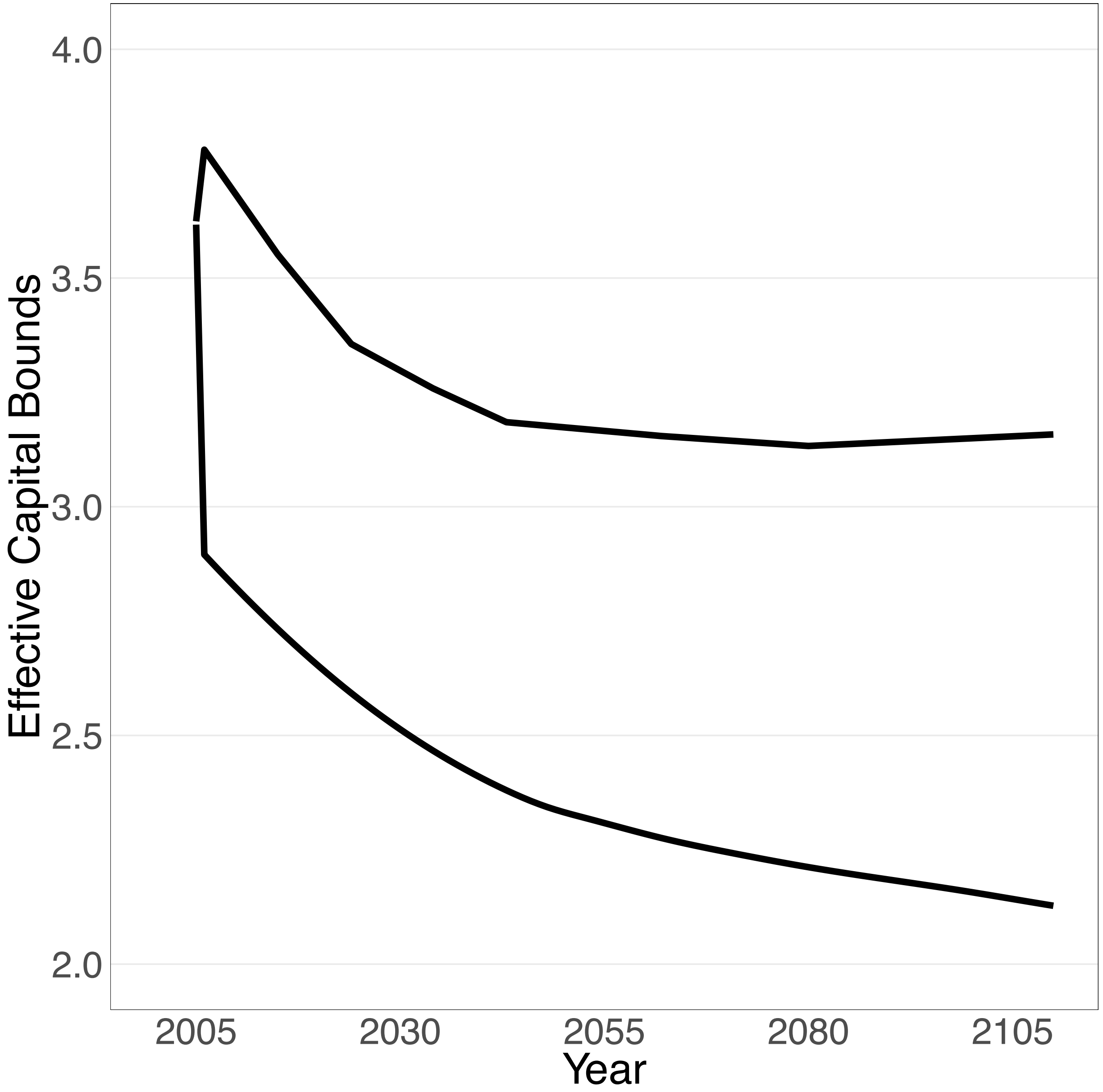
Step 5: Solve the model (twice)

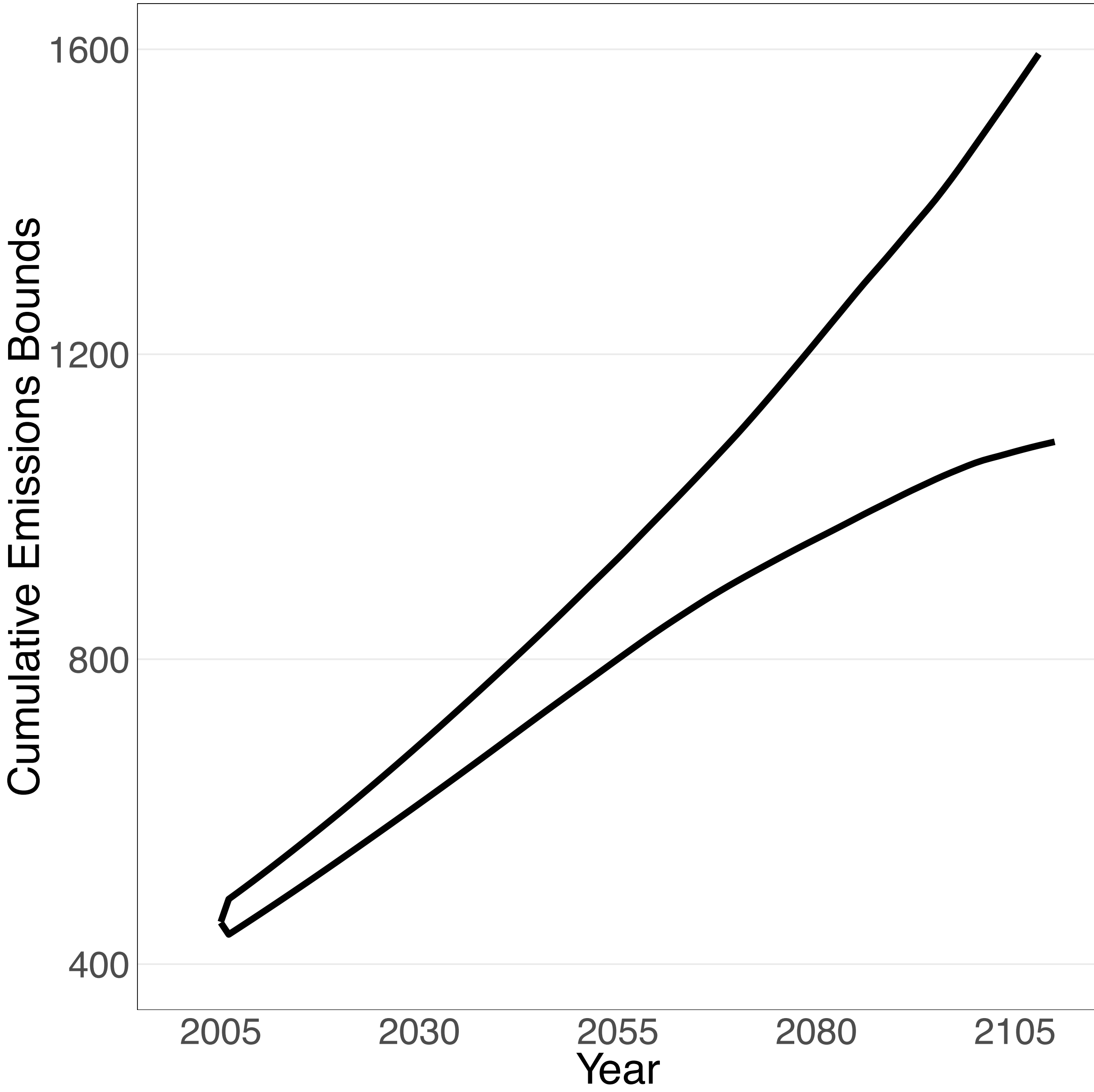
Dynamic Programming

The final step(s)

1) Do standard value function iteration
2) Simulate a bunch of potential state paths
    - The envelope of these paths will be time dependent
3) Generate a time-dependent / adaptive grid based on the sims
4) Repeat everything once

Adaptive grids





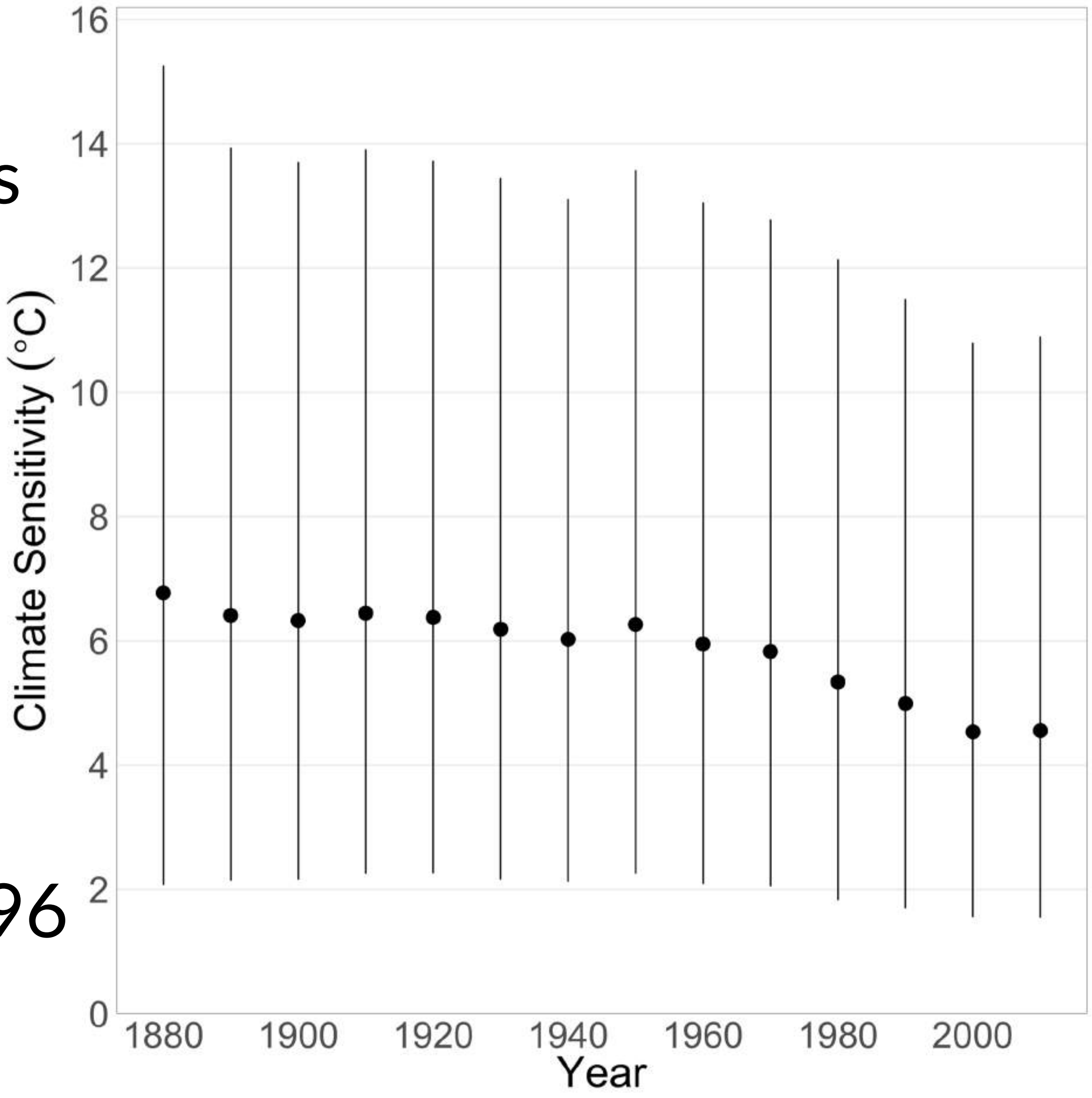
Finally,
the results

Backcast the learning model to validate

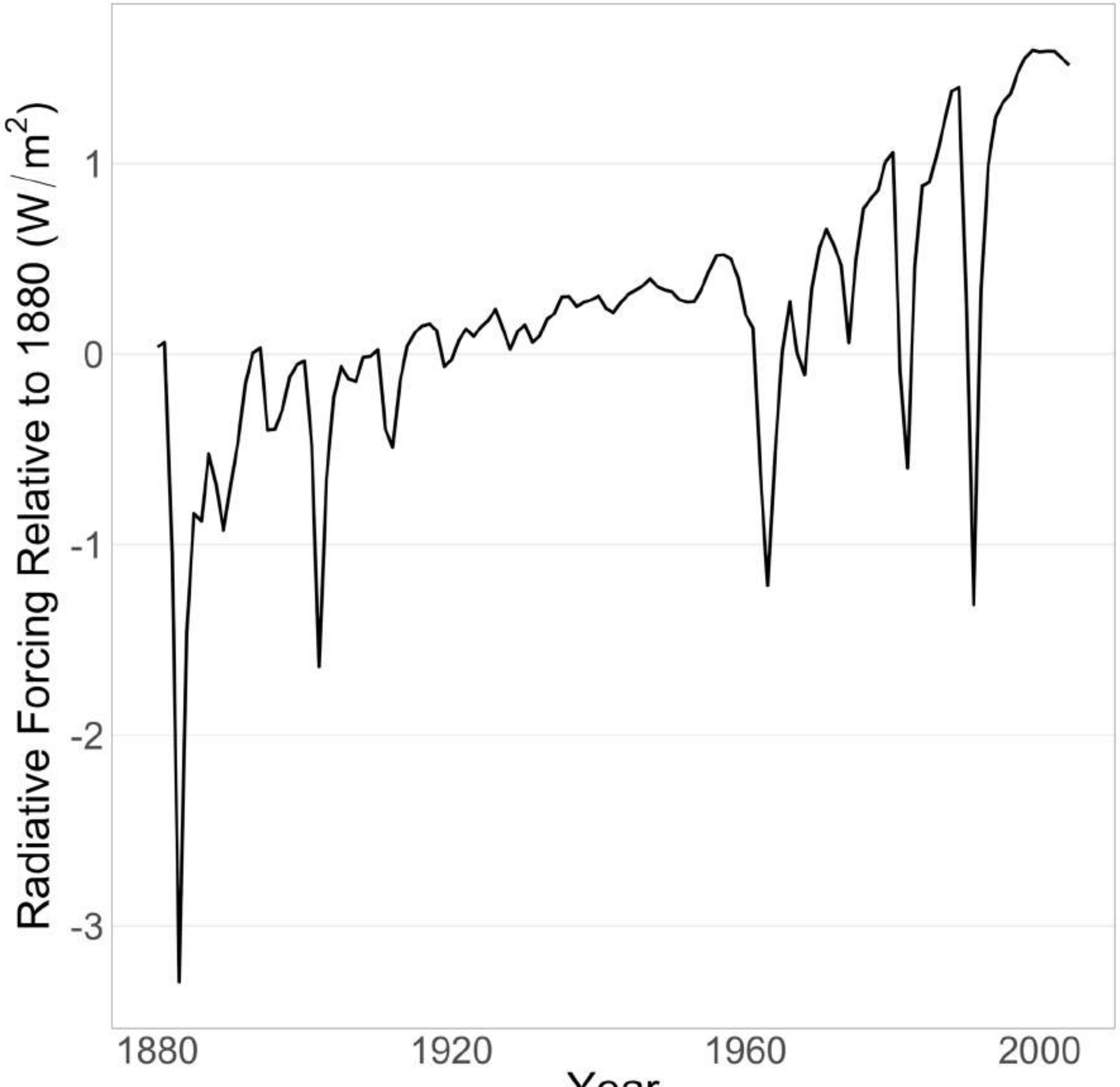
Conditional on the instrumental record, what belief trajectory gets us our current prior?

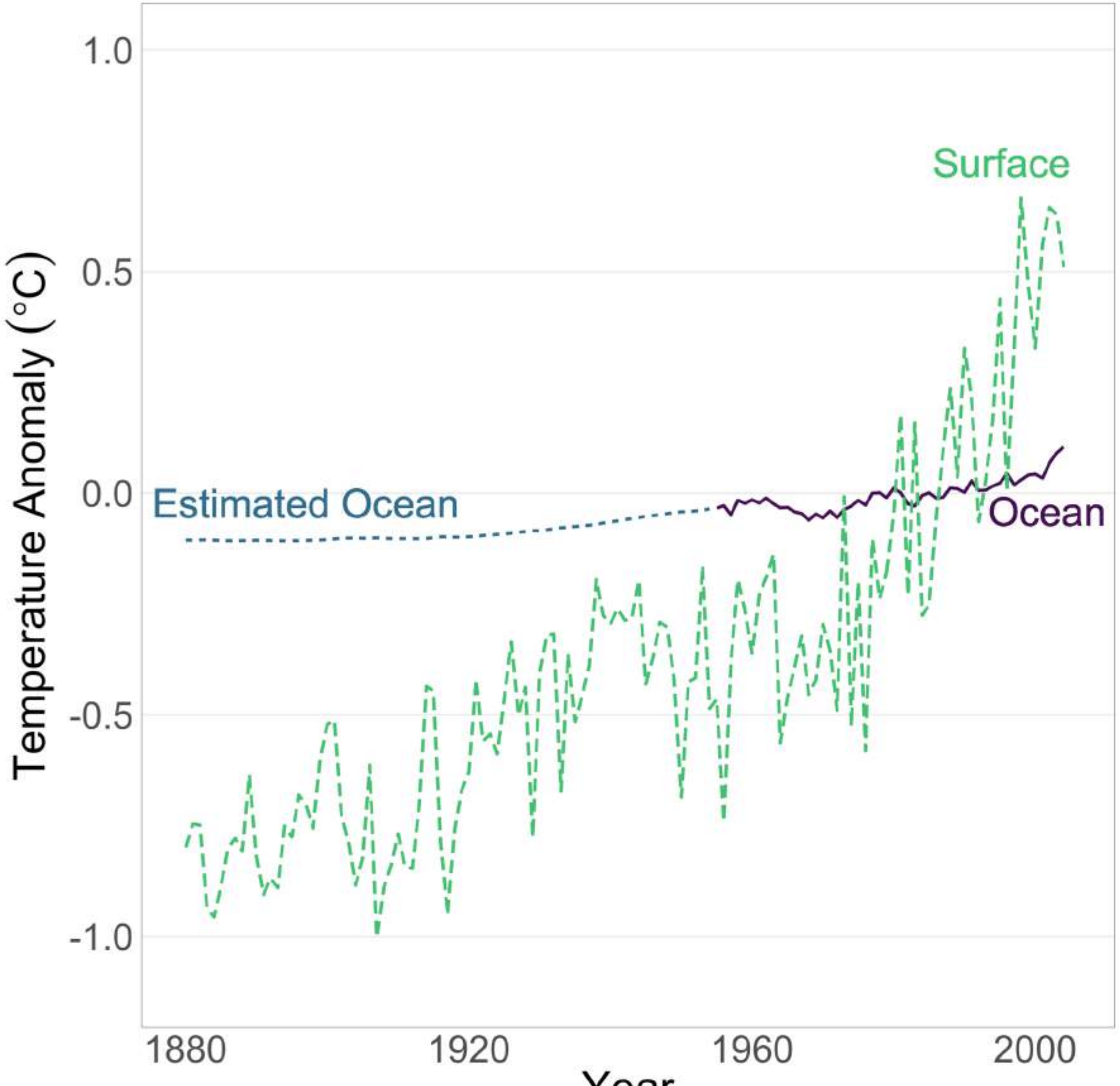
Consistent with historical beliefs over the 140 years

Even going back to the first climate sensitivity estimate in 1896



The data and data estimates





Advancing climate science increases the informativeness of the climate signal

We model improvements in modeling/monitoring as increasing the strength of the climate signal in the data

This is nearly identical to the theoretical information literature

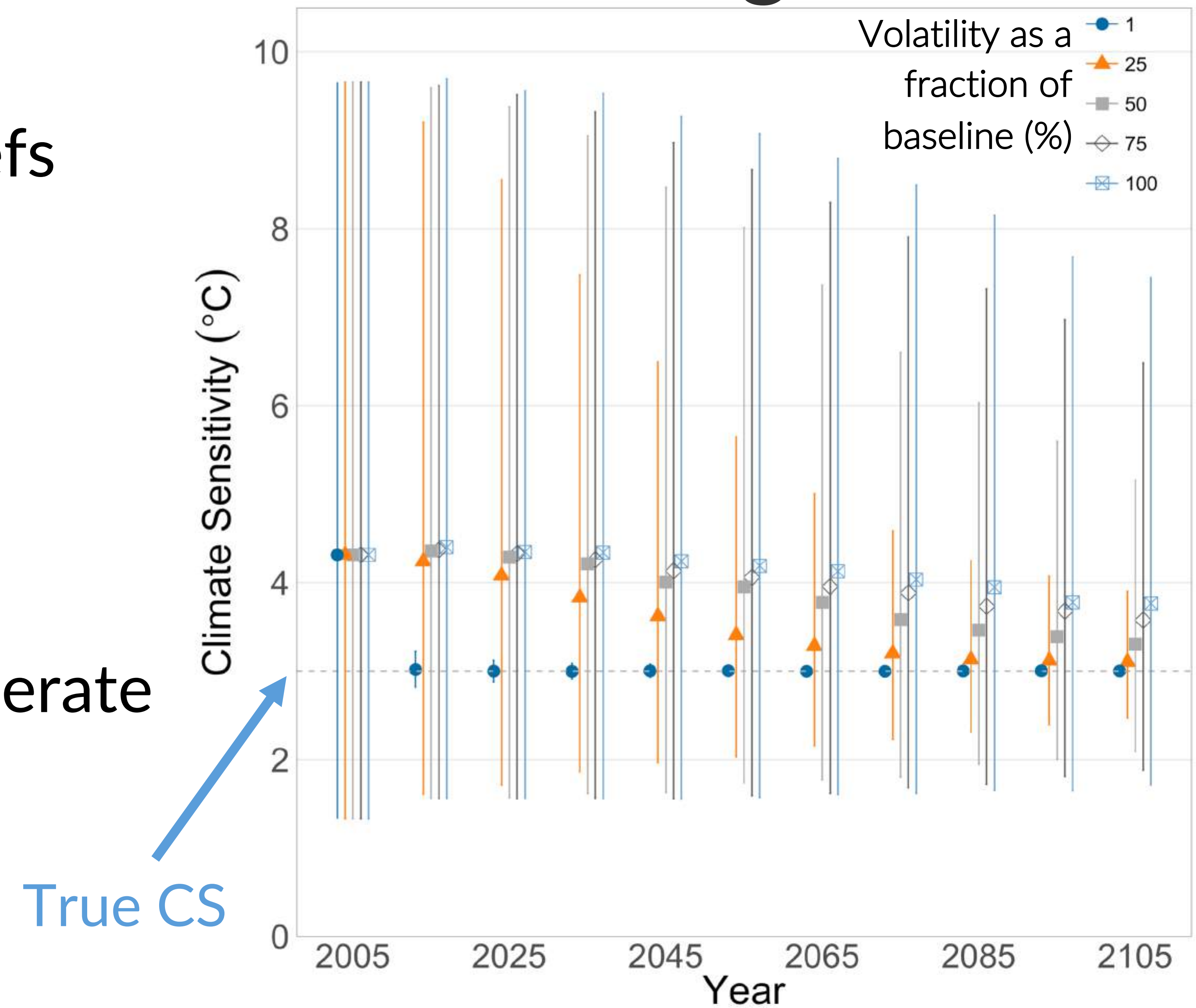
- Radner and Stiglitz (1984); Chade and Schlee (2002); Moscarini and Smith (2002); Keppo et al. (2009)

In practice, we reduce the volatility (unexplained variation) in temperature

Reducing volatility accelerates learning

Mean and 90% CI of beliefs about CS under different temperature volatilities

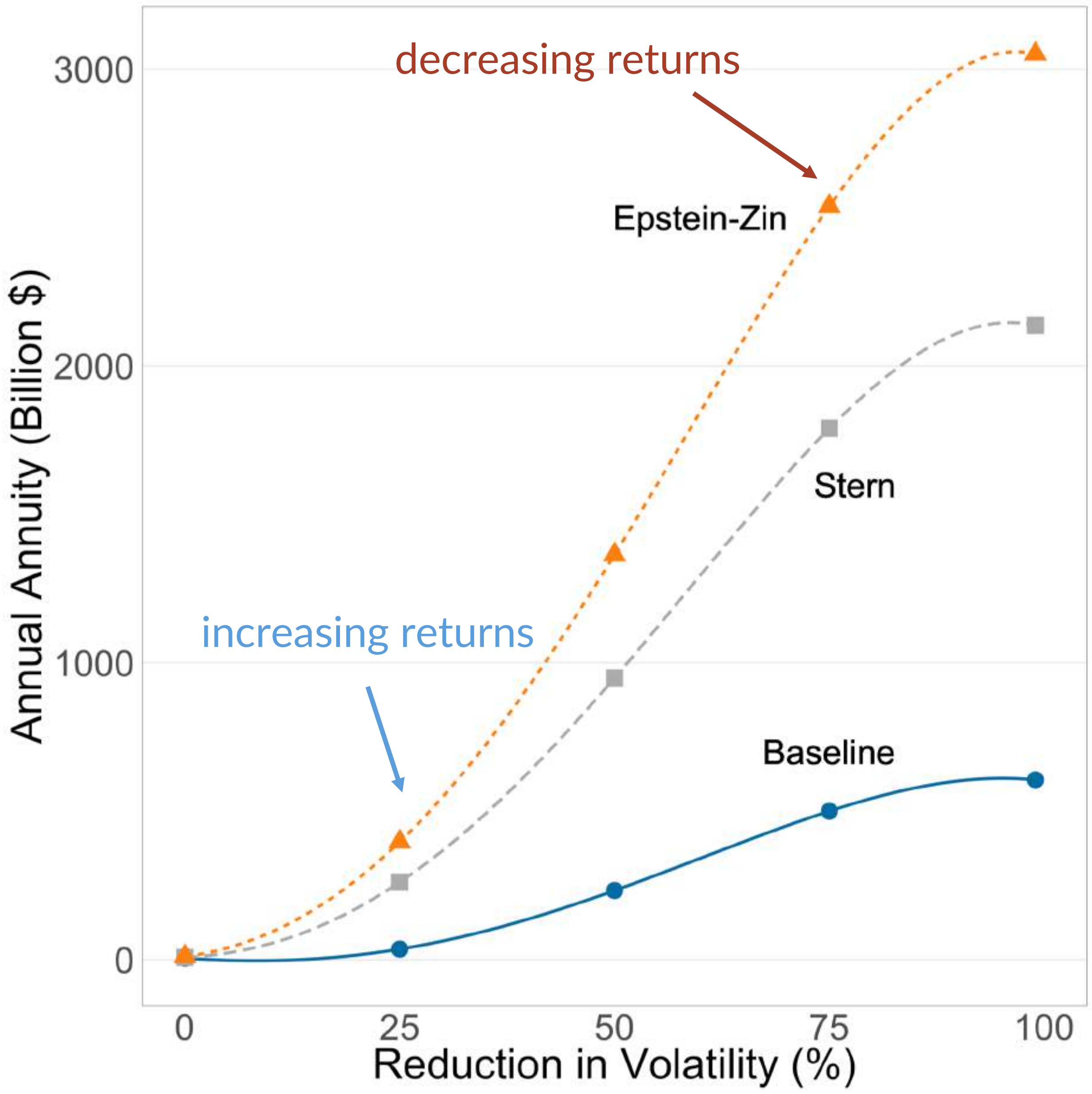
Volatility reductions accelerate learning, allow for better policymaking



Accelerating learning has significant value

Annuity from reducing unexplained variation in temperature

Non-concavities (upward sloping demand) are consistent with theoretical predictions



We estimate the gains from accelerating science

Science policy is highly valuable

Optimal funding is zero or large,
need MC estimates for optimal funding levels

Structural and numerical models often use
similar methods to approximate functions and distributions

These methods make high dimensional problems feasible/faster

Learn high dimensional optimization/approximation basics

Basis functions?

Gradient descent?

Sparse grids?

MCMC/HMC?

Julia?

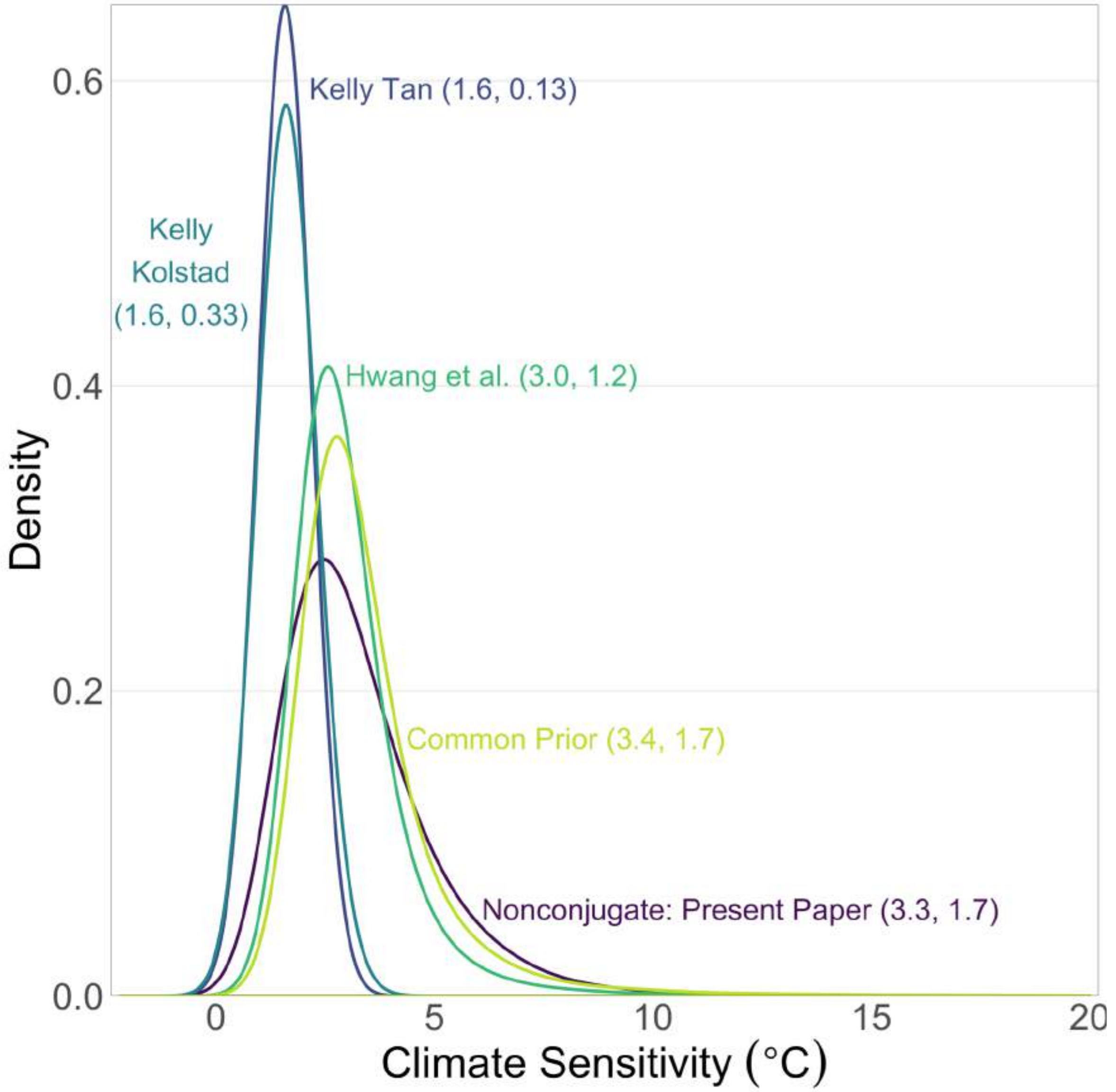
Step 4: Construct the grid and basis matrix

```julia
# Chebyshev polynomial function
function cheb_polys(x, n)
    if n == 0
        return 1                    # T_0(x) = 1
    elseif n == 1
        return x                    # T_1(x) = x
    else
        cheb_recursion(x, n) =
            2x.*cheb_polys.(x, n - 1) .- cheb_polys.(x, n - 2)
        return cheb_recursion(x, n) # T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)
    end
end;
```

https://github.com/AEM7130

Standard conjugate approach says we can now rule out disastrous climate sensitivity

Beliefs since the Charney report under four different IAM learning models



Alternative nonconjugate learning models are biased in *a priori* nonobvious ways