Challenge in Bayesian inference for climate reconstruction

Yu-Pei Sung

January 2026

We consider a stationary heat conduction problem on the unit square $\Omega = [0,1]^2$, governed by the Laplace equation

$$\Delta u(x,y) = 0 \in \Omega,$$

with homogeneous Dirichlet boundary conditions on three sides and an unknown, spatially varying Dirichlet boundary condition $u(0,y) = f(y)$ on the left boundary. The goal is to infer the unknown boundary function $f(y)$ from noisy temperature measurements taken at interior locations. Interior measurements are assumed to be related to the solution via

$$u_{\text{meas}} = \Phi(\alpha) + \eta,$$

where $\Phi(\alpha)$ denotes the forward model mapping boundary parameters to interior measurements and η the measurement noise.

The unknown boundary function is represented using a finite Radial Basis Function (RBF) expansion,

$$f(y) = \sum_{k=1}^{K} \alpha_k \exp\left(-\frac{(y-y_k)^2}{2\ell^2}\right),$$

where α_k are unknown coefficients, y_k uniformly distributed RBF centers along the boundary, and ℓ the RBF width. This representation enforces smoothness by construction. Given α, the forward model consists of solving the Laplace equation using a finite difference discretization with the prescribed boundary condition, and evaluating the solution at measurement locations using a Gaussian mollifier to model spatially averaged measurements.

Measurement noise is modeled as additive, independent Gaussian noise, leading to a Gaussian likelihood. This choice is motivated by the aggregation of multiple small and independent sources of experimental uncertainty, for which Gaussian noise is a natural approximation. Furthermore, the provided noise variance σ^2 supports this modeling assumption. A Gaussian prior is placed on the RBF coefficients, implying the belief that the boundary temperature is smooth and of moderate magnitude. The Gaussian prior acts as a regularization mechanism for the ill-posed inverse problem, penalizing unrealistically large boundary amplitudes while remaining weakly informative. The prior variance τ^2 is chosen on the scale of the observed data, allowing the measurements to dominate where informative. Finally, combining prior and likelihood yields the posterior distribution.

The posterior distribution is explored using a random-walk Metropolis–Hastings Markov Chain Monte Carlo (MCMC) sampler. This method is chosen for its simplicity, robustness, and suitability for non-Gaussian posteriors arising from nonlinear forward models. The sampler is initialized at $\alpha = 0$, corresponding to a homogeneous boundary condition, and the proposal step size is tuned to achieve a reasonable acceptance rate. Convergence is assessed using trace plots, running means, and autocorrelation analysis of representative coefficients. While the sampler exhibits autocorrelation due to correlated

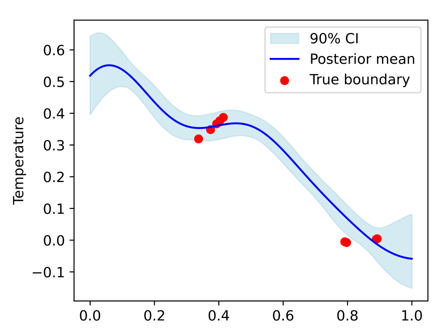

Figure 1: Bayesian reconstruction of the unknown boundary temperature. The solid line shows the posterior mean, the shaded region indicates the 90% credible interval, and red markers denote sparse samples of the true boundary provided for reference.

RBF coefficients and the smooth forward model, the running means after burn-in are stable and sufficient for uncertainty quantification.

Posterior samples are used to reconstruct the boundary function in a Bayesian sense. We report the posterior mean of $f(y)$ together with pointwise credible intervals, illustrating spatially varying uncertainty in the inferred boundary condition. Figure 1 shows the Bayesian reconstruction of the unknown boundary temperature $f(y)$.

The posterior mean captures the overall trend of the true boundary well, despite the limited and indirect information provided by interior measurements. Most reference samples lie within the credible interval, indicating that the inferred uncertainty is consistent with the observed discrepancy between model predictions and data. The width of the credible interval varies along the boundary. Regions with smaller uncertainty correspond to boundary segments that are more strongly informed by the interior measurements, while increased uncertainty near the endpoints reflects reduced identifiability due to weaker sensitivity of interior temperatures to those boundary regions.

ChatGPT

Hi! I want to work on a programming challenge on Bayesian inference, can I ask for your help?

Absolutely — ask away. 😊

If you paste the challenge prompt (or the part you're stuck on) and tell me what language you're using (Python/R/etc.), I can help you:

- translate the problem into a Bayesian model (priors, likelihood, posterior)
- derive the math (conjugacy, marginal likelihood, posterior predictive)
- implement it (MCMC, VI, or exact updates)
- debug code and check results
- explain what the outputs mean

Send the prompt + what you've tried so far, and where you want the most help.

Thank you! The challenge is about an inverse problem in the context of the heat distribution, expressed by the Poisson equation on the domain Ω=[0,1]2. In particular, we assume to have boundary conditions for three boundaries given, as well as selected measurements of the solution, i.e. temperatures inside the domain, and aim at recovering the fourth boundary condition from that information. I have learned about Bayesian inference at university, but there are so many possibilities. how do I know how to choose the distribution of prior or likelihood, or other specifications of the model?

In inverse PDE problems like this, you usually don't "pick priors/likelihoods" from scratch so much as you *encode two things you actually believe*:

1. how the data were generated (measurement + model error) → **likelihood**

2. what kinds of boundary functions are plausible (smoothness, scale, sign, etc.) → **prior**

A really practical way to decide is: **start from physics + measurement process + regularity**, then choose the simplest distributions consistent with those assumptions.

1) Likelihood: start from the measurement model

You have interior temperature measurements y_i at points x_i and a forward map

$$u = \mathcal{F}(g) \quad \text{(solve Poisson with boundary condition } g \text{ on the unknown side)}$$

and predicted measurements

$$\hat{y}_i(g) = u(x_i).$$

Most common default

Additive i.i.d. Gaussian noise:

$$y_i = u(x_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Then

$$p(y \mid g, \sigma) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - u(x_i))^2\right).$$

Why this is often the right first choice:

- temperature sensors often have roughly symmetric noise,
- sums of many small effects → Gaussian is a good approximation,
- it gives a clean quadratic misfit (ties nicely to Tikhonov regularization).

When you'd choose something else

- **Outliers / bad sensors** → Student-t likelihood (robust).
- **Different sensor accuracies** → heteroscedastic Gaussian $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$.
- **You suspect the PDE model is imperfect** → add a *model discrepancy term* (often another GP / correlated error).

A great "Bayesian hygiene" move: treat σ as unknown with a weakly informative prior (half-normal / half-Cauchy) instead of fixing it.

2) Prior: encode what boundary conditions "should look like"

The unknown boundary condition is a *function* on one edge (say $x = 1, y \in [0, 1]$). This is an **infinite-dimensional** unknown. Priors that work well are basically "smooth random functions".

The workhorse choice: Gaussian process / Gaussian random field prior

Discretize the boundary into m nodes and let $g \in \mathbb{R}^m$. Put

$$g \sim \mathcal{N}(\mu,\, \Sigma(\ell, \tau))$$

where Σ comes from a kernel like squared-exponential or Matérn.

Interpretation:

- τ controls amplitude (how big the boundary values can be),
- ℓ controls smoothness / length-scale (how wiggly it can be).

This is not arbitrary: for inverse elliptic problems, a smoothness prior is the Bayesian version of classical regularization.

Equivalent viewpoint (often easier in code): basis expansion + priors on coefficients

Represent

$$g(s) = \sum_{k=1}^{K} a_k \phi_k(s)$$

(Fourier, splines, FEM basis on the boundary), then put priors on a_k, e.g.

- $a_k \sim \mathcal{N}(0, \tau^2 k^{-2p})$ to enforce smoothness,
- or a random-walk prior $g_j - g_{j-1} \sim \mathcal{N}(0, \eta^2)$ (1D boundary → very convenient).

How to pick hyperparameters without guessing

Instead of fixing ℓ, τ, put **hyperpriors** and infer them:

- $\tau \sim$ half-normal(…) or half-Cauchy(…)
- $\ell \sim$ log-normal(…) or Gamma(…) on a reasonable scale (e.g. boundary length is 1, so ℓ in $[0.05, 1]$ is already meaningful)

This is how you avoid "so many possibilities": you choose a *family* (smooth functions) and let the data tune the level of smoothness/scale.

3) The "Bayesian checklist" for choosing a model

When you're unsure, answer these questions:

A. What do you know physically?

- Is the boundary temperature bounded? nonnegative? within a range?
  → consider transforming $g = \exp(h)$ or truncation, or a prior centered in that range.
- Do you expect smooth variation along the boundary?
  → GP / random-walk / spline coefficient priors.

B. What do you know about the measurements?

- Sensor noise level known?
  → fix σ or put a prior near that value.
- Outliers possible?
  → Student-t likelihood.
- Any spatial correlation in errors?
  → correlated noise model.

C. What does the inverse problem "want" numerically?

These problems are often ill-posed: many boundary functions fit the data similarly well.
A smoothness prior is not just "a choice", it's what makes the posterior stable.

4) A strong default model (good starting point)

If you want one concrete baseline that is widely accepted:

- Unknown boundary discretized $g \in \mathbb{R}^m$
- Prior: $g \sim \mathcal{N}(0,\ (\alpha L^\top L + \beta I)^{-1})$
  where L is a 1D second-difference matrix along the boundary (penalizes curvature → smoothness)
- Likelihood: $y \mid g, \sigma \sim \mathcal{N}(A(g),\ \sigma^2 I)$
  where $A(g)$ is "solve PDE then sample at measurement points"
- Hyperpriors: $\sigma \sim$ half-normal, $\alpha \sim$ Gamma (or log-normal)

That's basically Bayesian Tikhonov, but with uncertainty and hyperparameter learning.

5) How you'll know you picked well

Do **posterior predictive checks**:

- sample g from posterior → solve PDE → predict temperatures → compare to observed y
- if residuals look heavy-tailed → likelihood too light (switch to t)
- if posterior g is too wiggly / too flat compared to what you believe → adjust prior family or hyperpriors

If you paste the exact PDE setup (Poisson with what boundary types? Dirichlet/Neumann/mixed?), how the measurements are given, and whether the unknown boundary is Dirichlet or Neumann, I can suggest a *clean* model in equations and a practical inference approach (often: MAP + Laplace / HMC on coefficients / or ensemble Kalman inversion as an approximate Bayesian method).

this is specified in the description of the challenge: It is by intention that the description does not cover all aspects of the task. Submitted material will allow to better understand, how par
ticipants would approach and document a given task. It should also be noted that the quality of the inferred solution is not the central
point of interest. Instead the evaluation rather identifies a working solution, which can be clearly understood in its functionality, as well
as explanations that give a clear guidance on the developed solution strategy. Still, as a point of reference, a few samples of the "true"
boundary condition are provided. does this meaning I don't have to fine tune all the parameters?

Short answer: **yes — that wording very strongly suggests you are *not* expected to finely tune everything** 👍
What they care about is *how you think*, *how you justify choices*, and *whether the solution is coherent and reproducible*, not whether you squeezed out the last bit of accuracy.

Let me unpack what that paragraph is really signaling.

What the challenge description is *implicitly* telling you

1) "It is by intention that the description does not cover all aspects"

This is a **deliberate under-specification** test.

They want to see:

- how you deal with ambiguity,
- whether you can make *reasonable assumptions*,
- whether you document those assumptions clearly.

In Bayesian terms, this maps perfectly to:

> "State your modeling assumptions and encode them probabilistically."

So **choosing a reasonable prior + likelihood and explaining why** already scores highly.

---

2) "Quality of the inferred solution is not the central point"

This is the key sentence.

It means:

- You are **not** being benchmarked like a Kaggle competition.
- You are **not** expected to tune hyperparameters until the posterior mean matches the truth samples.

Instead, they want:

- a working forward–inverse loop,
- uncertainty quantification,
- a transparent inference pipeline.

A *moderately wrong but well-explained* Bayesian model beats a highly tuned black box.

---

3) "Evaluation identifies a working solution, clearly understood in its functionality"

This strongly favors:

- **simple, standard choices**,
- minimal but justified hyperparameters,
- plots and explanations over numerical performance.

For example:

> “We assume i.i.d. Gaussian noise with unknown variance, as is standard for temperature sensors, and infer it jointly with the boundary.”

This is *exactly* the kind of reasoning they are looking for.

---

4) “Samples of the ‘true’ boundary condition are provided”

This is **not** an invitation to overfit.

It’s mainly there to:

- give you a **sanity check**,
- help you argue whether your posterior is *reasonable*,
- validate qualitative features (smoothness, scale, sign).

A very strong use of these samples is **diagnostic**, not tuning:

> “The provided samples indicate a smooth boundary with no sharp discontinuities, which motivates our choice of a smoothness-promoting prior.”

That’s a modeling justification, not parameter hacking.

What this means concretely for your Bayesian choices

✔ You do not need to:

- grid-search length-scales
- optimize kernel hyperparameters aggressively
- match the truth samples numerically
- compare 5 different priors “just in case”

✔ You should:

- pick one *reasonable* prior family,
- put weakly informative hyperpriors if possible,
- explain what each modeling choice represents,
- show that inference runs and produces uncertainty.

A “safe” and well-aligned strategy for this challenge

Here is a strategy that would score very well under these criteria:

Step 1: State assumptions explicitly (this is part of the solution!)

Example:

> We assume that the unknown boundary condition varies smoothly along the boundary and that interior temperature measurements are corrupted by additive Gaussian noise.

This already shows maturity.

Step 2: Use a minimal, standard Bayesian model

- Likelihood: Gaussian with unknown σ
- Prior: smooth Gaussian prior (GP or discrete Laplacian penalty)
- Hyperpriors: weak, broad, not tuned

Step 3: Do one reasonable inference method

- MAP + Laplace approximation
  or
- MCMC on a reduced-dimensional parameterization

No need to compare methods unless you want to.

Step 4: Validate qualitatively

- plot posterior mean + credible bands of the boundary,
- compare with provided true samples,
- discuss where uncertainty is high and why (e.g. far from sensors).

This shows *understanding*, not overfitting.

How to phrase this in your write-up (important!)

This kind of sentence is gold in such challenges:

> Given the exploratory nature of the task, hyperparameters were not extensively tuned. Instead, weakly informative priors were chosen to encode smoothness assumptions while allowing the data to dominate where informative.

That tells evaluators you *know* what you're doing and are making conscious trade-offs.

Bottom line

You are absolutely right to suspect this:

> **You are being evaluated on modeling judgment and clarity, not hyperparameter optimization.**

If you want, next we can:

- draft a **clean "Model & Assumptions" section** you can almost reuse verbatim, or
- choose *one* concrete prior + likelihood and write it mathematically in a challenge-ready form, or
- discuss how to justify ignoring or fixing certain parameters without sounding sloppy.

Just tell me how far along you are.

Thank you! So I also can approximate or sample from the posterior I get. But how do I judge what is the best?

That's a *very* good question — and in the context of **this challenge**, the answer is quite different from "whichever gives the lowest error".

The key idea is:

> **There is no single "best" posterior approximation in absolute terms.**
> **There is only a "best given the purpose, assumptions, and constraints".**

Below is a practical framework you can actually use *and explain*.

---

1) First: what does "best" even mean here?

For Bayesian inference, "best" could mean:

- closest to the true posterior (theoretical accuracy),
- stable and reproducible,
- interpretable and explainable,
- computationally feasible,
- appropriate for the information content of the data.

This challenge prioritizes the middle three, not the first one.

So don't judge methods by:

- how fancy they are,
- whether they asymptotically converge to the exact posterior,
- whether they beat another method by a tiny margin.

Judge them by **fitness for purpose**.

---

2) A simple decision table (you can literally include this reasoning)

Method	**When it's a good choice**	**When it's not**
MAP only	Very ill-posed, need fast solution	No uncertainty quantification
MAP + Laplace	Posterior unimodal, near-Gaussian	Strong nonlinearity, multimodality
MCMC	Gold standard for uncertainty	Expensive, hard to diagnose
Variational inference	Fast, scalable	Underestimates uncertainty
Ensemble Kalman	PDE inverse problems, Gaussian-ish	Strong non-Gaussian posteriors

For Poisson inverse problems with smooth priors, **MAP + Laplace or low-dim MCMC are both very defensible**.

---

3) What evaluators expect you to do

They do **not** expect:

- method comparison studies,
- convergence proofs,
- ESS tables.

They *do* expect:

- that you can explain *why* you chose a method,
- that the approximation behaves sensibly,
- that uncertainty looks reasonable.

So the “best” method is the one you can **defend clearly**.

---

4) How to judge quality *without* knowing the true posterior

A) Posterior predictive checks (this is the strongest criterion)

This is your main diagnostic.

Procedure:

1. Sample $g^{(k)} \sim p(g \mid y)$
2. Solve PDE → predict interior temperatures
3. Compare predicted vs observed measurements

Questions to ask:

- Do predictions cover the observed data?
- Are residuals roughly consistent with the assumed noise?
- Are extreme deviations systematic?

If yes → your posterior is *doing its job*.

You can explicitly say:

> Posterior predictive checks indicate that the inferred boundary conditions produce temperature fields consistent with the observed data.

That’s textbook Bayesian validation.

B) Stability under small perturbations

Try one of these (no need to do all):

- change prior variance slightly,
- change initial guess,
- remove a few measurements.

If the posterior:

- changes slightly → good
- collapses or explodes → model too brittle

You can phrase this as:

> The inferred solution was found to be stable under moderate changes of prior hyperparameters.

That's a strong modeling signal.

C) Qualitative comparison with provided "true" samples

Use them carefully:

- compare smoothness,
- amplitude scale,
- spatial features.

Do *not* optimize against them.

Correct framing:

> The posterior mean captures the dominant features present in the provided reference samples, while uncertainty increases in regions weakly informed by data.

That shows insight.

---

5) Choosing between MAP+Laplace vs MCMC (very concrete guidance)

MAP + Laplace is "best" if:

- posterior looks unimodal,
- boundary prior is Gaussian,
- noise is Gaussian,
- you want clarity and speed.

It gives:

- point estimate,
- covariance,
- uncertainty bands.

This is *perfectly acceptable* and often preferred in inverse problems.

MCMC is "best" if:

- you want to demonstrate full Bayesian inference,

- dimensionality is moderate (after basis truncation),
- you can show trace plots and autocorrelation.

But it's **not automatically superior** in this context.

A great sentence:

> Given the apparent unimodality of the posterior and the Gaussian prior structure, a Laplace approximation around the MAP estimate was deemed sufficient.

---

6) What *not* to worry about

You do **not** need to:

- rank approximations by KL divergence,
- compare marginal likelihoods,
- tune samplers for hours,
- prove convergence.

If it runs, makes sense, and you can explain it → it's good.

---

7) A very strong "evaluation paragraph" you can adapt

> Since the true posterior is unavailable, the quality of the approximation was assessed using posterior predictive checks, stability with respect to modeling choices, and qualitative comparison with provided reference boundary samples. Given the smoothness of the inferred posterior and the consistency of predicted measurements with observed data, the chosen approximation was deemed adequate for the purpose of this study.

That paragraph alone signals maturity.

Bottom line

The best posterior approximation is the one that:

1. matches your assumptions,
2. produces sensible uncertainty,
3. passes posterior predictive checks,
4. can be clearly explained and justified.

If you want, next we can:

- help you **choose one method and justify it explicitly**, or
- design **2–3 diagnostic plots** that communicate posterior quality very clearly, or
- draft a **"Method choice and validation"** subsection tailored to this challenge.

Just tell me what you're leaning toward (MAP, Laplace, MCMC, etc.).