

RESEARCH ARTICLE

A generalized Bayesian framework for maximizing information gain and model selection

Prem Jagadeesan^{1,2,3}, Karthik Raman^{3,4*}, Arun K. Tangirala^{1,2,3*}

1 Department of Chemical Engineering, Indian Institute of Technology (IIT) Madras, Chennai, India, **2** Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), IIT Madras, Chennai, India, **3** Centre for Integrative Biology and Systems mEdicine (IBSE), Wadhvani School of Data Science and AI (WSAI), IIT Madras, Chennai, India, **4** Department of Data Science and AI, Wadhvani School of Data Science and AI (WSAI), IIT Madras, Chennai, India

* kraman@iitm.ac.in (KR); arunkt@iitm.ac.in (AKT)



Abstract

Computational modelling of dynamical systems often involves many free parameters estimated from experimental data. The information gained from an experiment plays a crucial role in the goodness of predictions and parameter estimates. Optimal Experiment Design (OED) is typically used to choose an experiment containing maximum information from a set of possible experiments. This work presents a novel Bayesian Optimal Experiment Design Selection principle for generalised parameter distributions. The generalization is achieved by extending the β -information gain to the discrete distributions. The β -information gain is based on what is known as the Bhattacharyya coefficient. We show that maximising the β -information gain is equivalent to maximising the angle between the prior and posterior distributions. We analytically show, with uniform prior, selecting an experiment that maximises β -information gain reduces the posterior's uncertainty. Further, we apply the proposed experiment selection criteria for two realistic experiment designs in systems biology. Firstly, we use the β -information gain to choose the best measurement method for parameter estimation in a Hes1 transcription model. The measurement method selected by the β -information gain results in the minimum mean square error of the parameter estimates. In the second case, we employ the proposed information gained to select an optimal sampling schedule for the HIV 1 2 LTR model. The sampling schedule chosen by the presented method reduces both prediction and parameter uncertainty. Finally, we propose a novel method for model selection using β -information gain and demonstrate the working of the proposed method in the model selection in compartmental models.

OPEN ACCESS

Citation: Jagadeesan P, Raman K, Tangirala AK (2026) A generalized Bayesian framework for maximizing information gain and model selection. PLOS Complex Syst 3(1): e0000082. <https://doi.org/10.1371/journal.pcsy.0000082>

Editor: Réka Albert, Pennsylvania State University, UNITED STATES OF AMERICA

Received: June 30, 2025

Accepted: November 23, 2025

Published: January 14, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcsy.0000082>

Copyright: © 2026 Jagadeesan et al.

This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The link for the code for estimating β -information gain is provided in the J2_supporting information. The code for all the examples and data can be found in this (<https://www.dropbox.com/scl/fo/4c22adh3pd58njo3d8gmh/AAdxarKxJetb0Kr8ZBz81F0?rlkey=4dylyx6q8g7ew63yfwjnnngdq0&st=hkq3aw89&dl=0>).

Funding: This work was supported by the Ministry of Education, Government of India to PJ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The author has declared that no competing interests exist.

Author summary

In this work, we present a generalized Bayesian framework for designing informative experiments and selecting suitable models in biological systems. In simple terms, our method identifies which experiments or measurements are most useful in improving parameter estimates and model predictions. The key idea is based on a new information measure called β -information gain, which uses the Bhattacharyya coefficient to quantify how much knowledge is gained from an experiment. We show that maximizing this gain is equivalent to reducing uncertainty and improving model confidence. Through case studies on the Hes1 transcription model and HIV-1 2-LTR dynamics, we demonstrate how this approach efficiently chooses the best experiments and sampling schedules. Our method also provides a novel and interpretable tool for model selection. Overall, this study provides a practical and computationally simple way to perform optimal experiment design in data-driven modeling in systems biology.

1 Introduction

System identification of biological processes poses numerous challenges at each stage of system identification [1]. Experiment design, model structure selection and parameter estimation are three crucial stages in system identification [2]. Obtaining precise/practically identifiable parameter estimates from noisy and limited data is one of the prevailing challenges in systems biology. Further, the nature of the multi-parameter non-linear models being sloppy hampers the quality of the parameter estimates [3–6]. Anisotropic sensitivity in the parameter space is the prime reason. It is well known that the factors contributing to the sloppiness are both experimental conditions and the nature of the model structure [4,7,8]. Hence, maximising the information content in the experimental data is an ineludible solution to obtain good predictions and parameter estimates.

Fisher Information Matrix (FIM) and covariance matrix of the parameter estimates are the most commonly used Optimal Experiment Design (OED) criteria [9,10]. Following are the three frequently used optimal design criteria to minimise variability in the parameter estimates: (i) *A-optimal*, maximising the trace of the FIM, (ii) *E-optimal*, maximising the minimum eigenvalue, and (iii) *D-optimal*, maximising the determinant of the FIM. Some works based on optimising FIM-based criterion are [11–13]. This work focuses on maximising the information gained in the Bayesian framework. One of the perceived advantages of the Bayesian framework is its ability to incorporate the prior information about a parameter in the form of a p.d.f (probability density function); moreover, posterior parameter distributions obtained from Bayesian inference allows us to use global methods rather than the aforementioned local Fisher information-based methods as the Fisher information for nonlinear predictors is usually numerically computed at the optimal parameter estimates.

Bayesian Optimal Experiment Design has seen several developments in the past two decades. A concise review of BOED is given in [14,15]. Initial attempts were focused on reducing the information entropy in the posterior parameter distribution, as entropy is a measure of uncertainty [16–18]. Juliane *et al.* use the concept of entropy and mutual information to design experiments that maximise the information content in terms of both parameter estimates and predictions [19]. BOED based on reducing the prediction variability was proposed in [20]. A computationally efficient method is proposed in [21]. A decision-theoretic approach with Kullback-Leibler divergence as the design criterion was proposed to design experiments for non-linear systems [22]. A recent FIM-based method to choose experiments to optimise the confidence region of the parameter estimates has been proposed in [23].

The problem of model selection is also linked to the information contained in the data set. The problem of model selection and the problem of optimal experiment design are two sides of the same coin. In the problem of model selection, the experimental data is fixed, and the central idea is to choose the model that absorbs maximum information in the data. In contrast, in the optimal experiment design, it is *vice versa*. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are the most commonly used model selection criterion. They attempt to strike a balance between the goodness of fit and parsimony of models [2]. However, they do not guarantee the goodness of the parameter estimates. To circumvent this issue, an experiment design solely for model selection was proposed in [24]; they used the experiment design criterion for Jensen–Shannon divergence. Daniel *et al.* proposed a Hellinger distance-based experiment design for model selection [25].

In all of the aforementioned information criteria for experiment design and model selection, one of the two following challenges prevails (i) Boundedness, (ii) Interpretability of bounds in terms of parameter estimates. From a system identification perspective, bounded information measures and the interpretability of the bounds have significant utility. Also, assuming joint prior and posterior parameter distributions to be joint Gaussian is a strong assumption in most cases. Hence, in this work, we extend the β -information gain-based BOED proposed in [29] to non-Gaussian prior and posterior distributions. In addition to the advantages proposed above, estimating the Bhattacharyya coefficient is computationally easier than the Kullback-Liebler divergence [26]. We work with the discretised sample prior and posterior distributions. The β -information gain proposed is bounded and has a natural interpretation in terms of the precision of the parameter estimates [29]. The β -information gain is based on what is known as a Bhattacharyya coefficient, and Bhattacharyya distance [27]. The Bhattacharyya distance measures the distance between two probability distributions, while the Bhattacharyya coefficient can be interpreted as the amount of sample overlap between two distributions. Further, we show that maximising the β -information gain is equivalent to maximising the angle between prior and posterior distributions resulting in a reduction of posterior uncertainty. We propose a machine learning-based method to estimate the β -information gain for discrete distributions.

We demonstrate the proposed BOED on two benchmark systems biology problems, (i) select a measurement system for optimal parameter estimation in Hes 1 model, the measurement method selected by the proposed method resulted in the minimum mean square error of the parameter estimates, (ii) select an optimal six-point sampling schedule for parameter estimates in HIV patients under treatment intensification, the parameters obtained from the sampling schedule resulted in minimum prediction and parameter uncertainty. Finally, we propose a novel method for model selection based on the β -information gain and demonstrate the working of the method in two-compartmental pharmacokinetic models.

The rest of the paper is organised as follows: [Sect 2.1](#) contains the preliminary concepts and definitions. [Sect 2.2](#) presents the preliminary results on extending the β -information gain to discrete distributions and its application to experiment design and model selection and illustrates the working of the proposed method with numerical examples. [Sect 3](#) contains the method for estimating the β -information gain and model selection. The paper ends with concluding remarks in [Sect 4](#).

2 Results

2.1 Preliminaries

Experiment design problem in dynamical systems: Consider a general non-linear ODE model with discrete time noisy measurements. The mathematical description of the model is given in Eq (1).

$$\begin{cases} \dot{x}(t) = f(x(t), \theta, u(t)) \\ y^*(t) = g(x(t), u(t)) \\ y(k) = y^*(k) + \xi(k) \end{cases} \quad (1)$$

The structure of the state equation is assumed to be known. The unknown variables are the parameter vector θ , the input $u(t)$, the noise characteristics of $\zeta(k)$, the number of data points N , and the sampling interval T . In some cases, the structure of $y(k)$ and the way the input enters the model are also unknown. The goal of the experiment design is to design one or more experiments to estimate the parameters θ from data based on optimal criteria.

Degrees of freedom for optimal experiment design of dynamical systems. The goal of any optimal experiment design problem is to select the experiment that has maximum information content. Information supplied by an experiment is a function of the quality and quantity of data [18]. Fig 1 shows the dynamical system formulation, factors affecting the quality and quantity of data.

The quality of data is affected by both input and output variables; from input direction, the number of inputs, the nature of input functions and the way the input enters the system affect the data quality. From the output direction, the number of outputs measured, signal-to-noise ratio, nature of the output function, number of data points, duration of the output and

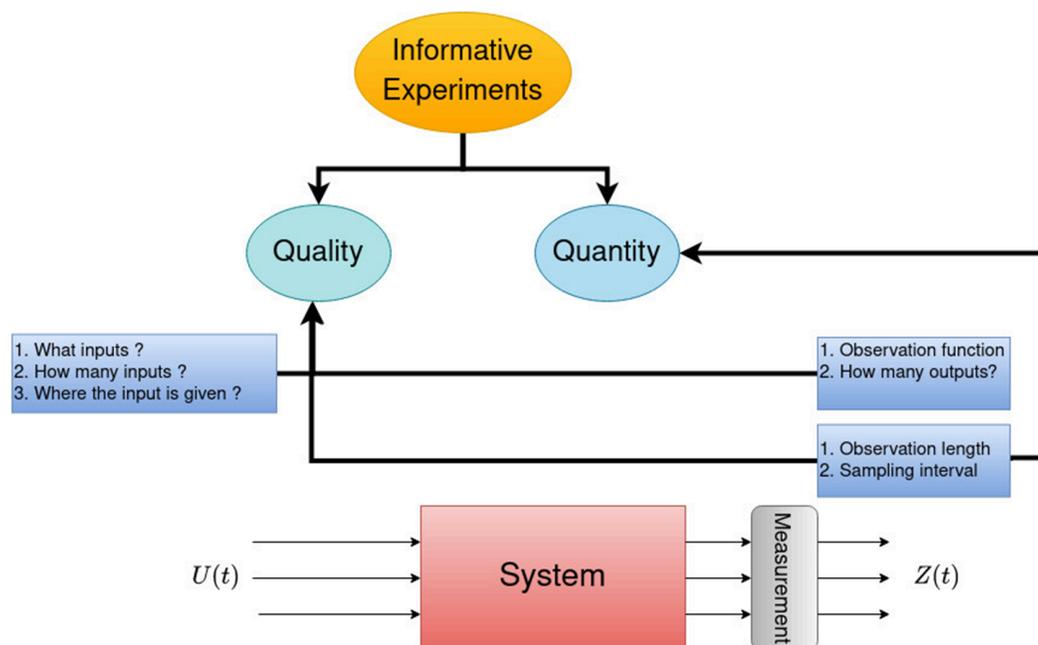


Fig 1. The diagram shows the relationship between system inputs, measurements, and the aspects of quality and quantity. The concept of informative experiments is divided into two major components, quality and quantity. The quality aspect focuses on factors such as the type, number, and location of inputs, the observation function, and the number of outputs. The quantity aspect deals with the observation length and sampling interval, influencing the amount and resolution of data collected.

<https://doi.org/10.1371/journal.pcsy.0000082.g001>

sampling interval affect the data quality. The quantity of data is affected by the number of data points, sampling interval and duration of the measurement.

Bayesian optimal experiment design: In the Bayesian framework, OED is defined as follows: given a set of discrete experimental conditions $\chi = \{z_1 \dots z_n\}$, prior information of the parameters in the form of a probability distribution function and a model structure \mathcal{M} , choose an experiment z_i that maximises the distance between the prior and posterior of the parameters:

$$\max_{z_i} d(p(\theta), p(\theta|z)) \tag{2}$$

where $d(\cdot)$ any distance measure between probability distributions, $p(\theta)$ is the joint prior distribution of parameters, $p(\theta|z)$ is the joint posterior distribution of parameters. This class of Bayesian optimal design problem's primary objective is to reduce the posterior distribution's uncertainty. However, unlike FIM based metric, where the focus is only on reducing the variability in the posterior distribution, obtaining a posterior distribution much different (in terms of assumed statistical/probabilistic distance) from the prior distribution is informative not only from the perspective of reduction of variability but also from moving the Maximum A Posteriori (MAP) estimate close to the true parameters, if the data is informative.

Bhattacharyya coefficient and Bhattacharyya distance: Bhattacharyya distance (B_d) is a measure of similarity between two statistical distributions [27]. The B_c can be used to quantify the relative closeness of two samples. The B_c of two densities $p_1(\theta)$ and $p_2(\theta)$ is given as

$$B_c = \int_{-\infty}^{\infty} \sqrt{p_1(\theta)p_2(\theta)} d\theta \tag{3}$$

where $p_1(\theta)$ and $p_2(\theta)$ have identical outcome space. B_c is bounded between $0 \leq B_c \leq 1$. A distance measure associated with this is the B_d :

$$B_d = -\ln(B_c) \tag{4}$$

The Bhattacharyya distance is bounded between $0 \leq B_d \leq \infty$. It is noteworthy that the B_d is not metric as it does not obey the triangle inequality.

In this work, we work with arbitrary discrete prior/posterior distributions where the closed expression for the distribution is unavailable; hence we turn towards the B_c for discrete distributions.

$$B_c = \sum_n \sqrt{p(\theta)p(\theta|z)} \tag{5}$$

where n is the number of bins in the multivariate histogram. The goodness of the estimate of the discrete Bhattacharyya coefficient depends on the number of bins. The Bhattacharyya coefficient is widely used in feature extraction [28] and optimal signal selection [26].

2.2 β -information gain for generalized distributions

In our previous work [29], we showed that the β -information gain index could be used as a Bayesian Optimal Experiment Design criterion. We demonstrated that β -information gain not only reduces the uncertainty in the posterior but also prediction uncertainty. The primary assumption of the previous work is that the joint prior and posterior distributions are Gaussian. In this work, we extend the β -information gain to non-Gaussian prior and posterior distributions.

In the case of discrete distributions, the information gain is defined as

$$\beta = 1 - \sum \sqrt{p(\theta)p(\theta|z)} \tag{6}$$

$$0 \leq \beta \leq 1$$

Case 1: If $p(\theta) = p(\theta|z)$ then $\beta = 0$. No new information is present in the experimental data apart from the prior.

Case 2: In the case of discrete probability mass functions, if only one of the samples is accepted in the posterior and if that sample has negligible probability in the prior, then $\beta \approx 1$. This indicates that the experiment is highly informative such that it has picked a low probability sample from the and Fig 2 shows the space of PMFs.

Geometric interpretation of the new information gain: Consider an n -dimensional vector space V where n is the cardinality of the outcome space of possible parameter values as shown in Fig 2. Each element v of the vector space is the square root of probabilities. Hence, each vector can be considered as a transformed probability mass function.

$$v = \{[\sqrt{p_1} \quad \sqrt{p_2} \quad \dots \quad \sqrt{p_n}] : \sum_{i=1}^N p_i = 1, 0 \leq p_i \leq 1\} \tag{7}$$

The domain of the vector space is restricted to,

$$\{v \in V : \|v\|_2 = 1, p_i \geq 0\} \tag{8}$$

The basis of the vector space are orthonormal basis

$$B = [e_1 \quad e_2 \quad \dots \quad e_n] \tag{9}$$

The basis vectors are PMF. of type Kronecker-delta function, i.e., the PMF has the form of Kronecker-delta function (Deterministic). Given a vector v , the angle that it makes with a basis vector is inversely proportional to the probability of the corresponding outcome.

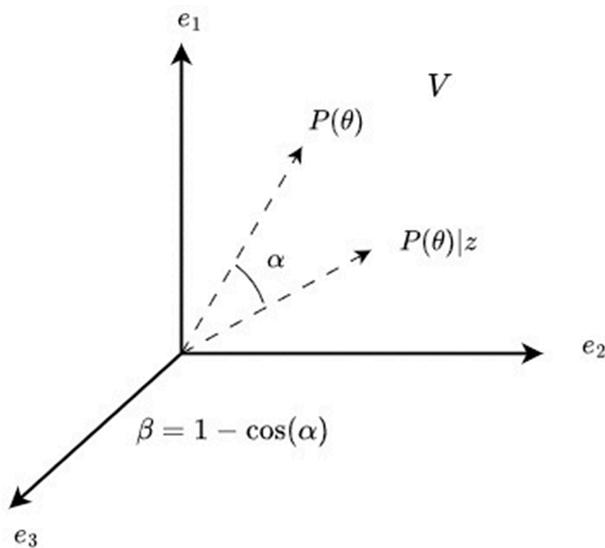


Fig 2. Three-dimensional view of the vector space V . In V , each point is considered as a Probability Mass Function. The Set of all probability mass functions from a probability simplex.

<https://doi.org/10.1371/journal.pcsy.0000082.g002>

In a Bayesian framework, the sampled prior and posterior probability density function can be represented as vectors in \mathcal{V} . The proposed information gain index is

$$\beta = 1 - \rho = 1 - \sum_{i=1}^n \sqrt{p(\theta)p(\theta|z)}, 0 \leq \rho \leq 1 \tag{10}$$

where ρ is the Bhattacharyya coefficient, the Bhattacharyya coefficient can be seen as \cos of the angle between two vectors in \mathcal{V} . We also know that the Bhattacharyya coefficient quantifies the amount of overlap between the distributions; hence, the angle represents the amount of overlap. The lesser the angle between two vectors, the more they are similar.

$$\cos(\alpha) = \sum_{i=1}^N \sqrt{p(\theta)p(\theta|z)}, 0 \leq \alpha \leq \frac{\pi}{2} \tag{11}$$

Case 1: When $\alpha = 0 \implies \rho = 1 \implies \beta = 0$. There is no new information in the data. Prior and posterior are identical.

Case 2: When $\alpha \approx \frac{\pi}{2} \implies \rho \approx 1 \implies \beta \approx 0$. Data has improved prior to the maximum in terms of reducing uncertainty.

Even when the posterior distribution is broad and overlaps significantly with the prior, β retains a well-defined geometric meaning: it quantifies the reduction in overlap between $\sqrt{p(\theta)}$ and $\sqrt{p(\theta|z)}$ in the probability vector space. Maximizing β therefore corresponds to moving these distributions apart, or equivalently, increasing the information gained about the parameter without assuming any delta-like behavior.

Theorem 1. *If the prior distribution of a parameter θ is uniform with arbitrarily large countably finite outcome space and posterior distribution in the form of Kronecker-delta for a given experiment e_k , then in the space \mathcal{V} , the angle between prior and the posterior distributions is $\frac{\pi}{2}$ resulting in maximum β -information gain.*

Proof: The angle between the prior and posterior distribution (Bhattacharyya coefficient B_c) is derived as follows,

$$B_c = \cos \alpha = \left[\sqrt{\frac{1}{n}} \sqrt{\frac{1}{n}} \quad \dots \quad \sqrt{\frac{1}{n}} \right] \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \sqrt{\frac{1}{n}} \tag{12}$$

For arbitrarily large δ ,

$$\lim_{n \rightarrow \delta} \alpha = \cos^{-1}\left(\sqrt{\frac{1}{n}}\right) \implies \alpha \rightarrow \frac{\pi}{2}$$

$$\beta = 1 - B_c \implies 1 - \cos\left(\frac{\pi}{2}\right) \implies 1 \tag{13}$$

This concludes the proof □

Thus maximising the angle is equivalent to reducing the uncertainty in the prior. In all other cases, maximising the angle will drive the posterior vector towards one of the basis. The angle cannot be $\frac{\pi}{2}$ because prior either has to be in an infinite dimensional space or one of the basis as prior, which is not possible in practical scenarios. Table 1 summarizes the extreme cases of Theorem 1.

Remark: The theorem considers an idealized boundary scenario often used in inference theory, where an uninformative (uniform) prior becomes fully informative through data, leading to a delta-like posterior centered on the

Table 1. Entropy and variance of the uniform prior and the Kronecker-delta posterior.

| PMF | Entropy | Variance |
|---------------------------|--|---|
| Prior $(p(\theta))$ | $\lim_{n \rightarrow \infty} H = \log(n) \rightarrow \infty$ | $\lim_{n \rightarrow \infty} \frac{n^2 - 1}{12} \rightarrow \infty$ |
| Posterior $(p(\theta z))$ | 0 | 0 |

<https://doi.org/10.1371/journal.pcsy.0000082.t001>

true parameter. This limiting case illustrates the upper bound of the β -information gain; however, the geometric and optimization-based interpretation presented in this work remains valid for general, non-delta posteriors and non-uniform priors.

2.3 Application in model selection

In the dynamical modelling of biological processes, model selection is considered one of the three classical problems from a system identification point of view [1]. Given a set of possible model structures, prior information on the parameters and a data set, the problem is to find the model that best explains the given data set. Factors influencing model selection are (i) the parameter estimates and (ii) the goodness of the predictions. The ensemble and asymptotic properties of the estimates generally assess the goodness of the estimates. However, minimising the uncertainty in the parameter estimates and prediction for small sample scenarios is considered sufficient. This work uses a new information gain-based approach for model selection in the Bayesian framework.

Problem statement: Given a model set $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$, a data set Z , and the priors of the parameters in the form of a p.d.f, choose an appropriate model structure that best explains the data.

Measure of goodness of fit: Compute the sum-squared error for every θ in the posterior distribution

$$\zeta_i = \sum_{t=1}^N \frac{(z(t) - y(t, \theta_i))^2}{N} \tag{14}$$

Then the measure of goodness of the predictions is computed by

$$\Delta = \hat{E}(\zeta) \tag{15}$$

In this work, we use the sample mean as an estimator of $E(\zeta)$.

2.4 Numerical results

Hes 1 oscillator: We consider a three-state model with mRNA regulatory dynamics of Hes 1 oscillator [19]. The state variable m , p_1 , and p_2 represent the Hes1 mRNA concentration, cytoplasmic concentration and nuclear protein concentration. The parameter P_0 is the amount of Hes1 protein in the nucleus, v is the translation of Hes1 mRNA, h is the hill coefficient, and k_1 is the rate of transport of Hes1 protein. The model equations are given in Eq (2)

$$\mathcal{M} : \begin{cases} \dot{m} = -k_{deg}m + \frac{1}{(1 + p_2/P_0)^h} \\ \dot{p}_1 = -k_{deg}p_1 + v_m - k_1p_1 \\ \dot{p}_2 = -k_{deg}p_2 + k_1p_1 \end{cases} \tag{16}$$

The parameter k_{deg} is experimentally measured as 0.03 min^{-1} . In this case, using Western Blots, it is possible to measure either the mRNA using real-time PCR or total Hes1 protein concentration ($p_1 + p_2$). In this example, we use the

proposed method to investigate which mode of measurement results in minimum parameter uncertainty. The reference parameters of the model are $\theta = [1.4 \ 5.7 \ 0.02 \ 0.09]$. The initial conditions are $x_0 = [5 \ 2.5 \ 2.5]$ and the model is simulated for $t = 0$ to $t = 200$ seconds with sampling interval $T_s = 5$ seconds.

The parameters are estimated using Markov Chain Monte Carlo-based Approximate Bayesian Computation (MCMC-ABC). Figs 3 and 4 show the histogram of the prior and posterior distributions along with the actual parameters. The data has significantly shrunk the uncertainty in the parameters P_0 and h . However, the parameter v_m has been estimated with poor precision. On the other hand, in the combined measurement of p_1 and p_2 , the parameters P_0 and k_1 have been estimated with considerable precision, but the parameters h and v_m are estimated with poor precision. The information gain is maximum for the experiment where mRNA is measured. From Table 2, it can be observed that both bias and reduction in the variance of the parameter estimates. By measuring mRNA alone, three out of four parameters are precisely measured while measuring two proteins, only two parameters are precisely measured, and there is a significant bias in the parameters h and P_0 . In addition, from Table 3, the experiment with maximum information gain has minimum Mean Square Error (MSE). Detailed analysis is supporting information file *S1 File - S4 File*.

This case study shows that the experiment chosen by β -information gain has minimum mean square error. Further, it reduces both parameter uncertainty and bias for the stiff parameter estimates.

HIV 1 2-LTR dynamics: In this example, we design optimal experiments to select optimal sampling times for estimating model parameters in the HIV 2-LTR model. This model is developed in [30] to predict 2-LTR concentration post-treatment intensification with an integrase inhibitor. The clinical trials are primarily costly, and along with it, the burden experienced by the patients must also be considered. Institutional Review Board (IRD) regulates clinical trials on patients. Hence optimal scheduling becomes extremely important in clinical trials. The model equations are given below:

$$\dot{\mathbf{y}}(\mathbf{t}) = -(1 - (1 - \eta_{\Pi} \mathbf{u}_{\Pi})R)\mathbf{a}\mathbf{y} + \mathbf{y}_e \tag{17}$$

$$\dot{\mathbf{c}}(\mathbf{t}) = \phi k_{\Pi}(1 - \eta_{\Pi} \mathbf{u}_{\Pi})R\mathbf{a}\mathbf{y} + k_{\Pi}\eta_{\Pi} \mathbf{u}_{\Pi}R\mathbf{a}\mathbf{y} - \delta \mathbf{c} \tag{18}$$

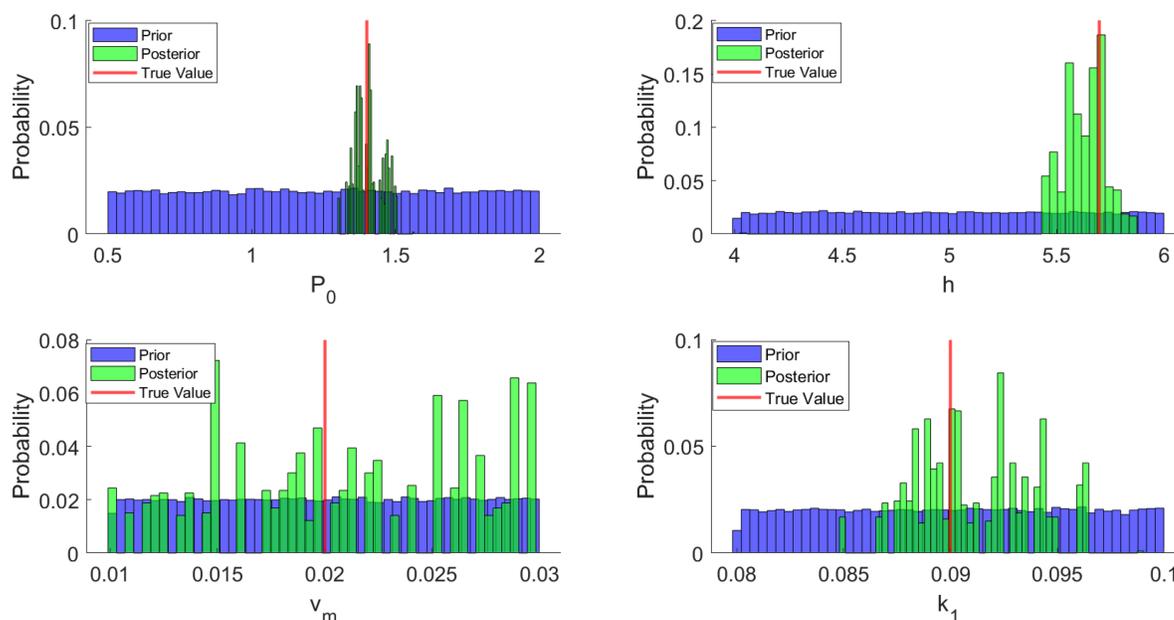


Fig 3. Prior and posterior distribution of all the parameters estimated by measuring both proteins. The variability in parameters P_0 and k_1 has considerably reduced compared to the parameters v_m and h .

<https://doi.org/10.1371/journal.pcsy.0000082.g003>

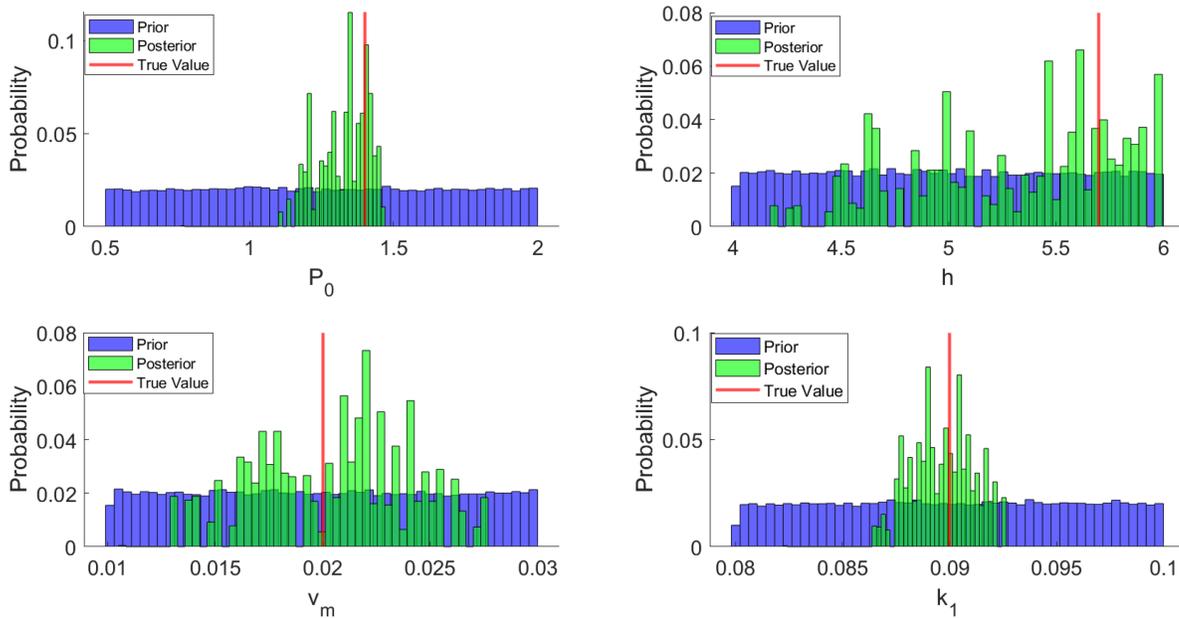


Fig 4. β -information gain for each sampling schedule.

<https://doi.org/10.1371/journal.pcsy.0000082.g004>

Table 2. Summary statistics of sample posterior distributions.

| Parameter | Estimate | | Bias | | Reduction in σ_θ^2 | |
|-----------|----------|-------------|---------|-----------------------|--------------------------------|-------------|
| | m | $p_1 + p_2$ | m | $p_1 + p_2$ | m | $p_1 + p_2$ |
| $y(t)$ | m | $p_1 + p_2$ | m | $p_1 + p_2$ | m | $p_1 + p_2$ |
| P_0 | 1.4066 | 1.3259 | 0.0066 | -0.0741 | 99.7% | 95.88% |
| h | 5.6319 | 5.2985 | -0.0681 | -0.4015 | 96.5% | 32.47% |
| v_m | 0.0211 | 0.0206 | 0.0011 | 5.7×10^{-4} | 0.28% | 62.12% |
| k_1 | 0.0911 | 0.0897 | 0.0011 | -3.1×10^{-4} | 77.3% | 93.57% |

<https://doi.org/10.1371/journal.pcsy.0000082.t002>

Table 3. Information gain β and MSE for both the experiments.

| $y(t)$ | β | MSE |
|-------------|---------|--------|
| m | 0.84 | 0.0185 |
| $p_1 + p_2$ | 0.74 | 0.3993 |

<https://doi.org/10.1371/journal.pcsy.0000082.t003>

where \mathbf{c} is a concentration of 2-LTR circles and \mathbf{y} is a concentration of actively infected CD4+ T-cells, \mathbf{y}_e -exogenous sources of infected cells unaffected by treatment, $\phi k_{\Pi}(1 - \eta_{\Pi} \mathbf{u}_{\Pi})R \mathbf{a} \mathbf{y}$ - rate at which 2-LTR circle forms post intensification, $(1 - (1 - \eta_{\Pi} \mathbf{u}_{\Pi})R) \mathbf{a} \mathbf{y}$ - turnover rate of the infected cells completing a cycle, \mathbf{u}_{Π} is a binary parameter indicating the presence or absence of the treatment and δ being the decay rate of the 2-LTR circles' formation. From [30], it is assumed that the dynamics have attained steady-state prior to treatment intensification; Eqs (20) and (21) are steady-state equations

$$c(t) = c(\infty) + (c(0) - c(\infty))e^{-\delta t} + c(\infty) \frac{\delta \eta_{\Pi} R (e^{-\delta t} - e^{-a(1-(1-\eta_{\Pi})R)t})}{(1-R)(a(1-(1-\eta_{\Pi})R) - \delta)} \quad (19)$$

$$c(0) = \frac{k_{\Pi} \mathbf{y}_e \phi R}{\delta(1-R)} \quad (20)$$

$$c(\infty) = \frac{k_{\Pi} \mathbf{y}_e (\phi + \eta_{\Pi} - \phi \eta_{\Pi} R)}{\delta(1 - (1 - \eta_{\Pi})R)} \quad (21)$$

Experimental conditions: In this work, we replicate the experimental conditions used in [30]. We consider the production of 2-LTR circles in the presence of 2-LTR circle treatment intensification. The replication of the 2-LTR circles in the presence of treatment is quite high, as predicted by the model and has been clinically observed in [30]. The measurement technique used to measure 2-LTR circles was a PCR process which introduces log-normal uncertainty; the measurement equation, along with measurement noise, is given below.

$$z(k) = c(k) + e(k), \quad e(k) \sim \mathcal{LN}(0, \sigma^2) \tag{22}$$

$$\sigma = 10^{-0.21 - 0.24 \log_{10}(42 \times 1.2)} \tag{23}$$

The variance of the noise in Eq (22) is taken from [30], which replicates the measurement error induced by a PCR process. The parameter values used for data generation are given in Table 4.

This work considers four different six-point sampling schedules obtained from four different optimal design criteria [31]. We investigate four sampling schedules informative with respect to β information index. We use MCMC-ABC to estimate parameters. We generate two hundred noise realisations with the same SNR and estimate β -information gain.

From Table 5, we can see that the average β -information gain is maximum for the *D*-optimal sampling schedule. Followed by *E*-optimal, *A*-optimal and *ELK*-optimal. From Table 5, it is evident that all the estimates of β are tightly constrained. The mean square error is also the minimum for the *D*-optimal sampling schedule.

From Table 6, all the sampling schedules are informative with respect to the parameter *R* (probability of infected cells infecting a target cell). *D*-optimal schedule is more informative with respect to the parameter ϕ (Ratio of probability of 2-LTR circle formation) and δ (decay rate of 2-LTR circles). With respect to the parameters η_π and $k_\pi y_e$, *E*-optimal sampling schedule contains more information.

Table 4. Parameter values estimated from the experiment.

| parameter | Value | units |
|-------------|--------|---------------------|
| ϕ | 0.0018 | — |
| δ | 0.46 | day ⁻¹ |
| <i>R</i> | 0.9994 | — |
| η_π | 0.19 | — |
| $k_\pi y_e$ | 0.39 | 2LTR/infected cells |

<https://doi.org/10.1371/journal.pcsy.0000082.t004>

Table 5. Sampling times selected from different optimal criteria.

| Experiment ID | Optimal criterion | 1 | 2 | 3 | 4 | 5 | 6 | $\hat{E}(\beta)$ | MSE |
|---------------|-------------------|---|---|---|----|----|----|------------------|-------|
| Ex1 | <i>A</i> -optimal | 0 | 1 | 4 | 9 | 27 | 28 | 0.859(±0.0006) | 0.65 |
| Ex2 | <i>D</i> -optimal | 0 | 1 | 2 | 4 | 11 | 26 | 0.893(±0.0008) | 0.54 |
| Ex3 | <i>E</i> -optimal | 0 | 2 | 4 | 10 | 26 | 27 | 0.867(±0.004) | 0.69 |
| Ex4 | EKLD | 1 | 2 | 3 | 11 | 27 | 46 | 0.846(±0.006) | 0.745 |

<https://doi.org/10.1371/journal.pcsy.0000082.t005>

Table 6. Reduction in variance for all the parameters.

| Parameter | <i>A</i> -opt | <i>D</i> -opt | <i>E</i> -opt | EKLD |
|-------------|---------------|---------------|---------------|-------|
| ϕ | 38.7% | 58.0% | 39.5% | 25.9% |
| δ | 80.9% | 84.8% | 80.6% | 47.8% |
| <i>R</i> | 99.7% | 98.2% | 99.9% | 99.1% |
| η_π | 47.2% | 14.7% | 63.0% | 11.2% |
| $k_\pi y_e$ | 9.5% | 22.3% | 69.3% | 29% |

<https://doi.org/10.1371/journal.pcsy.0000082.t006>

From Fig 5, it is clear that the samples obtained from D -optimal sampling schedule constraints the predictions most than all other sampling schedules. The $EKLD$ -optimal sampling schedule having the least information gain has poor prediction uncertainty. The β -information gain has chosen the experiment (D -optimal) with the minimum mean square error. Further, the shared parameters ϕ, α and replication rate (R) (crucial for detecting an ongoing infection) has been estimated with very high precision in the chosen experiment.

An illustrative example on model selection: In this example, we demonstrate the working of the proposed method in the leak ambiguity problem in two compartmental multi-input, multi-output models given in [32]. Fig 6 shows the candidate models and the data generating process is (M_1) with nominal parameter vector $\theta^* = [1 \ 1.5 \ 2 \ 1]^T$. The output is measured as $z(t) = x_1(t) + x_2(t) + e(t)$. The noise variance is adjusted such that the signal-to-noise ratio of the measured data is 30. The input (u_1) is given as a unit impulse given at $t = 0$, and the model is simulated for $t = 1$ to $t = 10$ seconds. Parameters are estimated using MCMC-ABC.

From Fig 7A and Fig 7B, the information gain index β and the prediction uncertainty measure $E(\zeta)$ are maximum and minimum, respectively, for the true model (M_1) from which the data is generated. Thus, the proposed method has selected the best model that explains the data in terms of parameter and prediction uncertainty.

Table 7 contains the parameter estimates and standard errors for all the candidate models. Also, from Fig 8, we learn that the parameters k_{01} and k_{21} are sensitive/stiff parameters and k_{02} and k_{12} are sloppy/insensitive parameters. The proposed method has chosen the model with low parameter bias and uncertainty on stiff parameters. It can also be seen that the point estimates of the stiff parameter k_{21} in M_2 have a high bias. Similarly, the point estimates of k_{01} and k_{21} in M_3 have high bias.

Fig 9 contains predictions from sample posterior distributions generated from each candidate model. The predictions from M_2 and M_3 are highly uncertain and have a significant systematic error compared to M_1 . The bias in the estimates of the stiff parameters is the prime reason. Hence, the proposed model selection method selects the best model based on

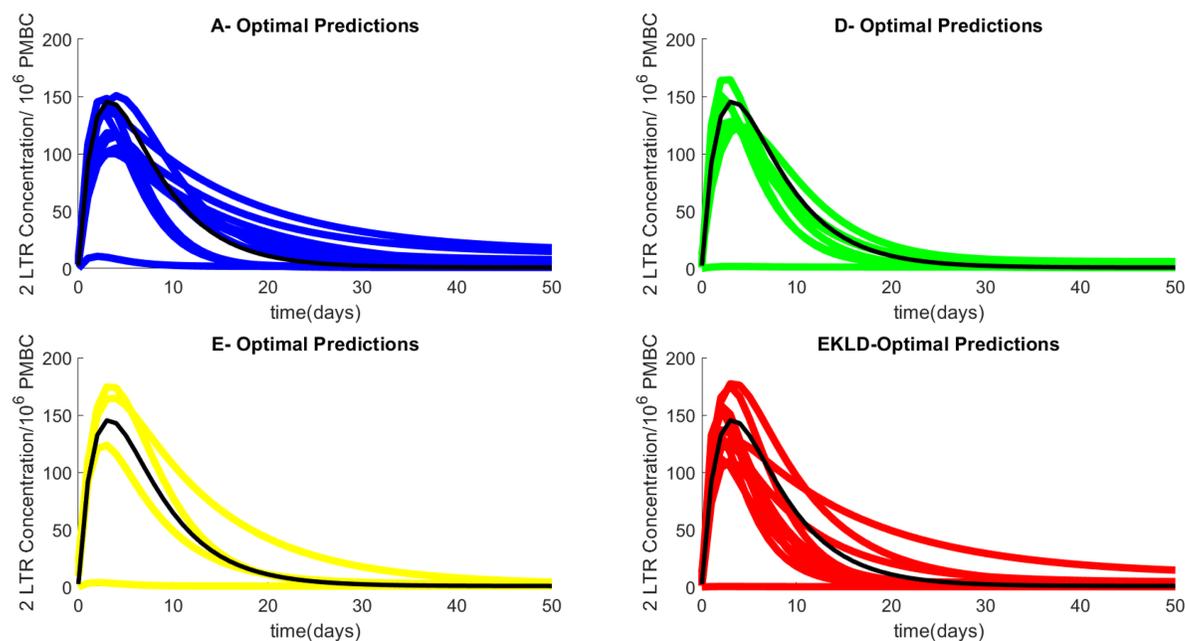
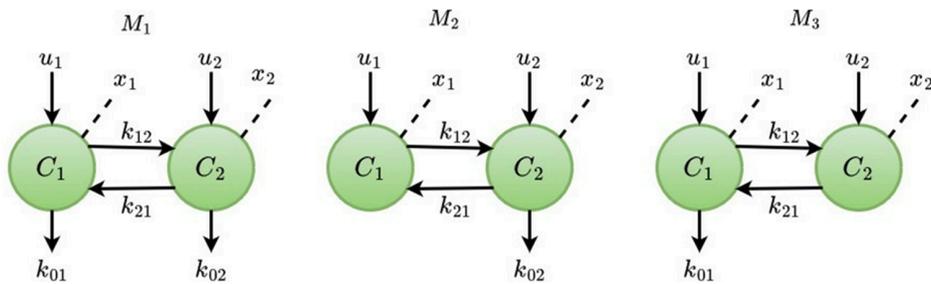


Fig 5. Predictions from the sample posterior parameter distributions for all the sampling schedules. Predictions from D -optimal are well constrained around the predictions from true parameters. Black-Predictions from true parameters, Blue-Predictions from A-optimal, Green-predictions from D -optimal, Yellow-predictions from E-optimal, Red-predictions from $EKLD$ -optimal.

<https://doi.org/10.1371/journal.pcsy.0000082.g005>



$$\begin{aligned} \dot{x}_1 &= -(k_{01} + k_{21})x_1 + k_{12}x_2 + u_1 & \dot{x}_1 &= -k_{21}x_1 + k_{12}x_2 + u_1 & \dot{x}_1 &= -(k_{01} + k_{21})x_1 + k_{12}x_2 + u_1 \\ \dot{x}_2 &= -(k_{02} + k_{12})x_2 + k_{21}x_1 + u_2 & \dot{x}_2 &= -(k_{02} + k_{12})x_2 + k_{21}x_1 + u_2 & \dot{x}_2 &= -k_{12}x_2 + k_{21}x_1 + u_2 \end{aligned}$$

Fig 6. Candidate models. Possible models that can represent the data generating process.

<https://doi.org/10.1371/journal.pcsy.0000082.g006>

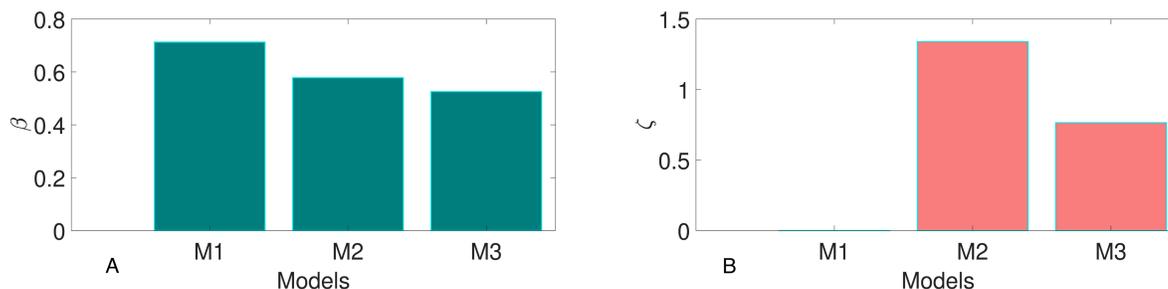


Fig 7. Model M1 has the highest information gain (β) at the same time lowest average error (ζ).

<https://doi.org/10.1371/journal.pcsy.0000082.g007>

Table 7. Parameter estimates of all the candidate models.

| Models | k_{01} | k_{02} | k_{12} | k_{21} |
|--------|----------------------|---------------------|---------------------|---------------------|
| M_1 | 0.50 (± 0.05) | 1.59 (± 0.22) | 2.09 (± 0.22) | 1.49 (± 0.05) |
| M_2 | - | 1.58 (± 0.18) | 2.09 (± 0.19) | 2.65 (± 0.02) |
| M_3 | 0.83 (± 0.051) | - | 2.10 (± 0.19) | 2.48 (± 0.07) |

<https://doi.org/10.1371/journal.pcsy.0000082.t007>

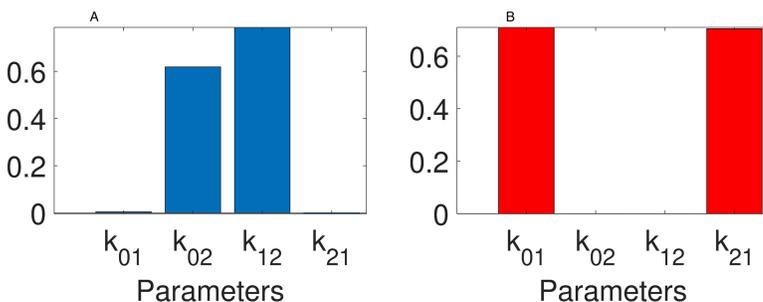


Fig 8. Sloppy and stiff eigenvectors. Parameters k_{02} and k_{12} are stiff, while k_{01} and k_{21} are sloppy.

<https://doi.org/10.1371/journal.pcsy.0000082.g008>

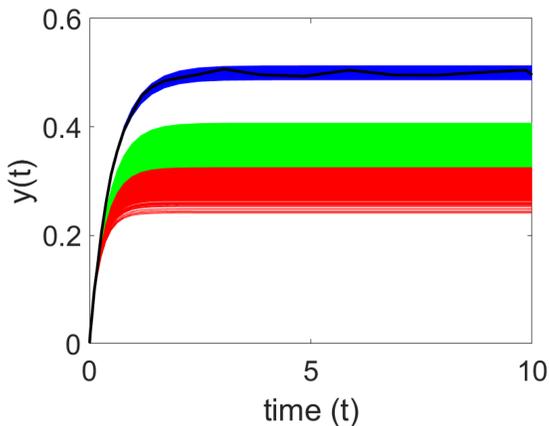


Fig 9. Comparison of prediction uncertainty of candidate models; Blue - predictions from sample posterior of model (M_1). Green - predictions from sample posterior of model (M_2). Red - predictions from sample posterior of model (M_3). Black - data (z).

<https://doi.org/10.1371/journal.pcsy.0000082.g009>

the goodness of predictions and parameter estimates. Maximising β -information gain index for joint distributions minimises the variance and bias of stiff (sensitive) parameters. Consequently, The prediction uncertainty is tightly bounded around the data points.

3 Discussion

The design of optimal experiments is one of the crucial aspects in the system identification of biological systems, primarily due to the difficulty and cost associated with conducting experiments. In this work, we extend our previous work [29] on Bayesian Optimal Experiment Design to the generalised parameter distributions and their application in model selection.

Firstly, using the numerical Bayesian estimation algorithm, the priors and posteriors are not always analytically tractable distributions, except in a few special cases. Hence, we estimate the β -information gain using the discrete version of the Bhattacharyya coefficient. The proposed version of the β -information gain has a beautiful geometric interpretation; maximising the information gain is equivalent to increasing the angle between the prior and posterior PMFs. Further, we show that for a uniform prior (uninformative) and Kronecker delta (deterministic) posterior with large countably finite sample space, the angle between prior and posterior is $\pi/2$. Further, we propose a method to estimate the discrete version of β -information gain.

We demonstrated the working of our method in two realistic experiment selection problems with high relevance in systems biology. Firstly, we used the proposed method to choose the best measurement function (implicitly measurement method) for precise and accurate parameter estimation in Hes 1 transcription network. The β -information gain chose the experiment with minimum mean square error, and all the stiff parameters were estimated with minimum bias and variance. Secondly, we employed the proposed method to select the optimal six-point sampling schedule for the HIV1 2-LTR dynamics model. Given the regulations and costs of clinical trials, this problem is paramount in healthcare. Out of four sampling schedules considered, the β -information gain again chose the six-point sampling schedule, resulting in a minimum mean square error of parameter estimates. Further, the predictions using the parameters from the β -optimal resulted in the least prediction uncertainty.

Finally, we proposed a novel method for model selection in systems biology using the β -information gain. We demonstrated the working of the method in compartmental models. The proposed method has rightly identified the true model. Thus, we believe that the proposed β -information gain has vast potential to be applied in experiment design selection and

model selection. The demonstrated examples show that the use of β -information gain in the experiment design for systems biology unambiguously chooses experiments that result in minimum parameter and prediction uncertainty. Hence, we believe that generalized β -information gain has significant utility in several aspects of system identification, such as OED and model selection.

4 Methodology

In this section, we discuss the procedure for Bayesian Optimal Experiment Design and model selection for generalised priors. Given a model M_i , a prior of the parameters in the form of probability distribution and a set of experimental conditions $\mathcal{X} = z_1, z_2, \dots, z_n$, the aim is to select the experiment condition z_i that maximises the β -information gain.

4.1 Procedure for estimating β -information gain

In order to estimate the Bhattacharyya coefficient for discrete distributions, we need to create bins in the parameter space and estimate the PMF. In this work, we employ the K -means algorithm to create bins in the parameter space. We choose the number of bins for each parameter as \sqrt{N} , where n is the number of samples in the prior parameter distribution. The procedure for estimating the new information gain is given below (Fig 10).

Procedure.

1. Sample N points from joint prior distribution $P(\theta)$
2. Compute the number of bins using

$$n_b = n_p \sqrt{N} \quad (24)$$

where n_p - dimension of the parameter space, n_b - is the number of bins and N - is the number of samples.

3. Create n_b number of bins in the prior parameter space using k -means clustering algorithm.
4. Estimate the PMF of the prior $P(\theta)$ by computing the probability for each bin.
5. Initialise $\theta_i, \theta_i = 1$
6. Propose θ^* from prior distribution $P(\theta)$
7. Simulate a data set z_d^* from the model $\mathcal{M}(\theta^*)$
8. If $d(z_d, z^*) < \epsilon$, go to Step 9, else set $\theta_{i+1} = \theta_i$ and go to Step 10.
9. Set $\theta_{i+1} = \theta^*$ with probability

$$\alpha = \min\left(1, \frac{f(\theta^*)q(\theta_i|\theta^*)}{f(\theta_i)q(\theta^*|\theta_i)}\right) \quad (25)$$

and $\theta_{i+1} = \theta_i$, with probability $1 - \alpha$

10. Accept θ^* into posterior $P(\theta|z_d)$, increment $i = i + 1$ and go to Step 6
11. Estimate the PMF of the posterior distribution $P(\theta|z_d)$
12. Estimate the β -information gain using Eq (6)

We use MCMC-ABC to estimate the posterior distribution of parameters [33]. The plain ABC rejection algorithm will have a low acceptance rate when the prior distribution of parameters is different from the true posterior distribution.

4.2 Procedure for model selection

As a first step, for each model structure $M_i \in M$, we estimate the β -information gain index between the joint prior and posterior of the parameter estimates. From Eq (13), it is evident that maximising the information gain index β minimises

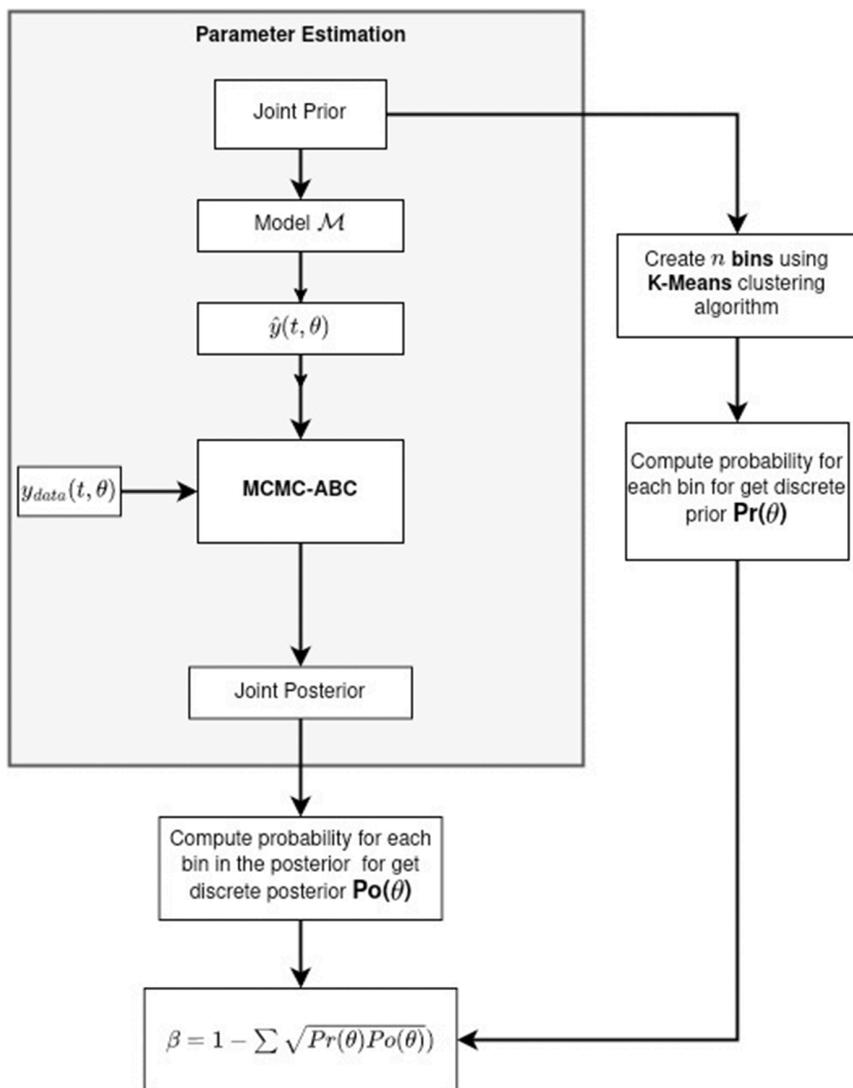


Fig 10. Estimating β - information gain. Algorithm for estimating β -information gain for generalised prior and posterior distributions.

<https://doi.org/10.1371/journal.pcsy.0000082.g010>

the uncertainty in the parameter estimates. This gain in information indicates the practical identifiability of parameter estimates for each model M_i . Further, we compute the goodness of predictions using Eq (15). Table 8 outlines the procedure for the model selection.

Table 8. Procedure for model selection.

| |
|--|
| Step 1: Compute β - information gain. |
| Step 2: Compute ζ_j for all the samples in the posterior parameter distribution Eq (14) |
| Step 3: Compute $E(\zeta)$ using |
| Step 4: Repeat, step 1 to step 3 for each $M_i \in \mathcal{M}$ |
| Step 5: Choose the model that has maximum β and minimum $E(\zeta)$ |

<https://doi.org/10.1371/journal.pcsy.0000082.t008>

$E(\zeta)$ is estimated using sample mean. The model with minimum $E(\zeta)$ has the least prediction uncertainty. Thus, the model that results in the least prediction and parameter uncertainty is the winner. However, when the model set does not contain the true model that generated the data, a single model resulting in low predictions and parameter uncertainty is not guaranteed; in such scenarios, the user can choose a model with a trade-off between prediction and parameter uncertainty.

Supporting information

S1 File. Sample prior and posterior parameter distributions for A-optimal sampling schedule. Blue: sample prior distribution; Green: sample posterior distribution; Red: true parameter.

(PDF)

S2 File. Sample prior and posterior parameter distributions for D-optimal sampling schedule. Blue: sample prior distribution; Green: sample posterior distribution; Red: true parameter.

(PDF)

S3 File. Sample prior and posterior parameter distributions for E-optimal sampling schedule. Blue: sample prior distribution; Green: sample posterior distribution; Red: true parameter.

(PDF)

S4 File. Sample prior and posterior parameter distributions for ELK-optimal sampling schedule. Blue: sample prior distribution; Green: sample posterior distribution; Red: true parameter.

(PDF)

S5 File. Variation of computational cost and estimated β -information gain with the number of samples (N). (a) Computational cost increases approximately linearly with N , scaling close to $\mathcal{O}(N)$, indicating predictable computational growth. (b) Estimated $E(\hat{\beta})$ remains stable within 0.84–0.90 as N increases, showing that the β -information gain estimator is statistically stable and computationally tractable.

(PDF)

S6 File. MATLAB code for β -information gain estimation and numerical examples. Code and data are available at the Dropbox link provided in the manuscript.

(PDF)

Author contributions

Conceptualization: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

Data curation: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

Formal analysis: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

Funding acquisition: Karthik Raman, Arun K. Tangirala.

Investigation: Prem Jagadeesan.

Methodology: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

Project administration: Arun K. Tangirala.

Supervision: Karthik Raman, Arun K. Tangirala.

Validation: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

Visualization: Prem Jagadeesan.

Writing – original draft: Prem Jagadeesan.

Writing – review & editing: Prem Jagadeesan, Karthik Raman, Arun K. Tangirala.

References

- Villaverde AF, Banga JR. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface*. 2013;11(91):20130505. <https://doi.org/10.1098/rsif.2013.0505> PMID: 24307566
- Tangirala AK. Principles of system identification: theory and practice. CRC Press; 2014.
- Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*. 2007;3(10):1871–8. <https://doi.org/10.1371/journal.pcbi.0030189> PMID: 17922568
- Apgar JF, Witmer DK, White FM, Tidor B. Sloppy models, parameter uncertainty, and the role of experimental design. *Mol Biosyst*. 2010;6(10):1890–900. <https://doi.org/10.1039/b918098b> PMID: 20556289
- Waterfall JJ. Universality in multiparameter fitting: sloppy models. Cornell University; 2006.
- Raman DV, Anderson J, Papachristodoulou A. Delineating parameter unidentifiabilities in complex models. *Phys Rev E*. 2017;95(3–1):032314. <https://doi.org/10.1103/PhysRevE.95.032314> PMID: 28415348
- Tönsing C, Timmer J, Kreutz C. Cause and cure of sloppiness in ordinary differential equation models. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2014;90(2):023303. <https://doi.org/10.1103/PhysRevE.90.023303> PMID: 25215847
- Jagadeesan P, Raman K, Tangirala AK. Sloppiness: fundamental study, new formalism and its application in model assessment. Cold Spring Harbor Laboratory; 2022. <https://doi.org/10.1101/2022.04.02.486816>
- DiStefano JJ. Dynamic systems biology modeling and simulation. *Science Progress*. 2019;102(4):378.
- Atkinson AC, Donev AN, Tobias RD. Optimum experimental designs, with SAS. Oxford University Press; 2007.
- Zhang X-Y, Trame MN, Lesko LJ, Schmidt S. Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol*. 2015;4(2):69–79. <https://doi.org/10.1002/psp4.6> PMID: 27548289
- Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, et al. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *IET Syst Biol*. 2007;1(3):190–202. <https://doi.org/10.1049/iet-syb:20060065> PMID: 17591178
- Transtrum MK, Qiu P. Optimal experiment selection for parameter estimation in biological differential equation models. *BMC Bioinformatics*. 2012;13:181. <https://doi.org/10.1186/1471-2105-13-181> PMID: 22838836
- Jiang H, Zhao Y. A review of Bayesian optimal experimental design on different models. *Modern Statistical Methods for Health Research*. Springer; 2021. p. 205–20.
- Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat Sci*. 1995;10(3):273–304.
- Bernardo JM. Expected information as expected utility. *Ann Stat*. 1979;7(3):686–90.
- DeGroot MH. Uncertainty, information, and sequential experiments. *Ann Math Stat*. 1962;33(2):404–19.
- Lindley DV. On a measure of the information provided by an experiment. *Ann Math Stat*. 1956;27(4):986–1005.
- Liepe J, Filippi S, Komorowski M, Stumpf MPH. Maximizing the information content of experiments in systems biology. *PLoS Comput Biol*. 2013;9(1):e1002888. <https://doi.org/10.1371/journal.pcbi.1002888> PMID: 23382663
- Vanlier J, Tiemann CA, Hilbers PAJ, van Riel NAW. A Bayesian approach to targeted experiment design. *Bioinformatics*. 2012;28(8):1136–42. <https://doi.org/10.1093/bioinformatics/bts092> PMID: 22368245
- Paulson J, Martin-Casas M, Mesbah A. Optimal Bayesian experiment design for nonlinear dynamic systems with chance constraints. *J Process Control*. 2019.
- Huan X, Marzouk YM. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J Comput Phys*. 2013;232(1):288–317.
- Gottu Mikkula AR, Paulen R. Optimal experiment design in nonlinear parameter estimation with exact confidence regions. *Journal of Process Control*. 2019;83:187–95. <https://doi.org/10.1016/j.jprocont.2019.01.004>
- Vanlier J, Tiemann CA, Hilbers PAJ, van Riel NAW. Optimal experiment design for model selection in biochemical networks. *BMC Syst Biol*. 2014;8:20. <https://doi.org/10.1186/1752-0509-8-20> PMID: 24555498
- Silk D, Kirk PDW, Barnes CP, Toni T, Stumpf MPH. Model selection in systems biology depends on experimental design. *PLoS Comput Biol*. 2014;10(6):e1003650. <https://doi.org/10.1371/journal.pcbi.1003650> PMID: 24922483
- Kailath T. The divergence and bhattacharyya distance measures in signal selection. *IEEE Trans Commun*. 1967;15(1):52–60. <https://doi.org/10.1109/tcom.1967.1089532>
- Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull Calcutta Math Soc*. 1943;35:99–109.

28. Coleman GB, Andrews HC. Image segmentation by clustering. *Proc IEEE*. 1979;67(5):773–85. <https://doi.org/10.1109/proc.1979.11327>
29. Jagadeesan P, Raman K, Tangirala AK. Bayesian optimal experiment design for sloppy systems. *IFAC-PapersOnLine*. 2022;55(23):121–6. <https://doi.org/10.1016/j.ifacol.2023.01.026>
30. Luo R, Cardozo EF, Piovoso MJ, Wu H, Buzon MJ, Martinez-Picado J, et al. Modelling HIV-1 2-LTR dynamics following raltegravir intensification. *J R Soc Interface*. 2013;10(84):20130186. <https://doi.org/10.1098/rsif.2013.0186> PMID: 23658114
31. Cannon L, Vargas-Garcia CA, Jagarapu A, Piovoso MJ, Zurakowski R. HIV 2-LTR experiment design optimization. *PLoS One*. 2018;13(11):e0206700. <https://doi.org/10.1371/journal.pone.0206700> PMID: 30408070
32. DiStefano JJ. *Dynamic systems biology modeling and simulation*. Academic Press; 2013.
33. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface*. 2009;6(31):187–202. <https://doi.org/10.1098/rsif.2008.0172> PMID: 19205079