UNIVERSITY OF CALIFORNIA,
IRVINE

Hierarchical Bayesian Modeling, Model Selection, and Optimal Experimental Design for Hematopoiesis Dynamics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical, Computational and Systems Biology

by

Yushang Lai

Dissertation Committee:
Professor John S. Lowengrub, Co-Chair
Professor Volodymyr M. Minin, Co-Chair
Professor Jack Xin

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACKNOWLEDGMENTS

First of all, I want to express my deepest gratitude to my PhD advisors Dr. John Lowengrub and Dr. Volodymyr Minin, who inspired, guided and supported me throughout this journey. I could never earn this degree without the feedback regulations and feedforward instructions from them. John and Volodymr are exceptional, not only for their breadth of knowledge and creativity, but also for their kindness, philosophies on being a researcher, and thoughts on life. I am more than honored and fortunate to meet John when taking his Numerical Analysis course the summer seven years ago. John is an awesome instructor and advisor with a deep insight in biological-justified mathematical modeling and numerical algorithm of real applied problems. John is the morning star in my life to be a researcher in computational biology. I also could not have imagined a better adviser than Volodymyr, who has a deep insight in developing statistically rigorous solutions to problems that arise in biological sciences. Volodymyr taught me not only amazing dynamic uncertainty Bayesian modeling approaches for hematopoiesis, but also life-long lessons that benefit me for my whole lifetime: that only change one thing at a time and that finishing things is more important and harder than starting new things. Truly, thanks my advisors John and Volodymyr for everything.

I would also like to thank my committee member Dr. Jack Xin for his inspiring suggestions and comments on the minimax optimization project during my final exam. Also, many thanks to Dr. Anglela Fleischman for her support of the experimental datasets across all the projects, which are significant in validating mathematical and statistical models. I appreciate all the support from members of Lowengrub group and Minin group. I learned a lot from the wonderful presentations and fruitful discussions during our weekly group meetings.

This work is supported by NSF grants DMS-1936833, DMS-1714973, DMS-1763272, Simons Foundation grant 594598QN, and NIH grants 1U54CA217378-01A1 and P30CA062203.

VITA

Yushang Lai

EDUCATION

Doctor of Philosophy in
Mathematical, Computational and Systems Biology **2023**
University of California, Irvine *Irvine, California*

Bachelor of Science in Applied Mathematics **2018**
University of California, Irvine *Irvine, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2018–2023**
University of California, Irvine *Irvine, California*

SOFTWARE

h-bayes-hema-model `https://github.com/yushang-lai/h-bayes-hema-model`
Julia, R and Python code for Projects

ABSTRACT OF THE DISSERTATION

Hierarchical Bayesian Modeling, Model Selection, and Optimal Experimental Design for Hematopoiesis Dynamics

By

Yushang Lai

Doctor of Philosophy in Mathematical, Computational and Systems Biology

University of California, Irvine, 2023

Professor John S. Lowengrub, Co-Chair

Professor Volodymyr M. Minin, Co-Chair

Hematopoiesis is the complex mechanism by which hematopoietic stem cells produce a variety of functional blood cells through multiple stages of differentiation. Since the numbers of various blood cell types need to be maintained in homeostasis, with occasional short-lived departures from it, hematopoiesis must have multiple regulatory mechanisms. However, these are still not fully understood. Although many mathematical models of hematopoiesis regulation have been proposed, more work on developing methods for fitting and interpreting experimental data that integrate statistical and mechanistic models is needed.

Here, using a new chemical reaction ordinary differential equation model of negative feedback regulation in hematopoiesis, we develop a scalable, hierarchical Bayesian framework using a latent variables approach that takes cross heterogeneity into account and infers division, differentiation, and feedback regulation parameters of hematopoietic cells. We designed and performed an experiment where mice were injected with the chemotherapy drug 5-FU that reduces the number of stem and progenitor cells by blocking DNA synthesis and repair, to perturb the hematopoietic equilibrium. In order to count the number of cells in the BM, the mouse must be sacrificed. Therefore, each mouse can contribute their cell count data at

a one time point only. To work with partially observed datasets, we use an ODE model to interpolate the noisy means of the experimental cell count data (the missing data is inferred). We evaluate the performance of the new model and inferential framework using synthetic data and find that we are able to distinguish between models that account for biological variation and models that include only technical variation/measurement error. We find that the experimental data are best described by a hierarchical model in which the hematopoiesis model parameters are allowed to vary among mice, suggesting the presence of significant biological variability. Our experimental data and the model show that, after perturbation, hematopoiesis returns to equilibrium via damped oscillations, with a notable overshoot of depleted cell counts that happens shortly after the system is perturbed from equilibrium.

We then explore an alternative way of accounting for data heterogeneity by employing stochastic differential equations instead of letting division and feedback regulation parameters vary across mice. Computational tractability of the likelihood in a Bayesian inference framework is achieved by using the linear noise approximation (LNA) derived from the chemical Langevin equation. This enables us to approximate the joint posterior density for the hematopoietic rate value parameters and missing data. We evaluate the performance of the new Bayesian LNA model framework and compare it to the Bayesian ODE model frameworks we developed previously. We find that the new framework can further improve the out-of-sample prediction, as indicated by leave-one-out cross-validation. We identify limitations of inference for our LNA model when multiple sources of biological and technical variation of the dataset are significant and then develop a procedure for overcoming them.

Finally, we investigate experimental designs that optimize the amount of information gained about the model parameter and missing data. We employ a new adversarial approach that uses a game theory framework for experimental design without the need for calculation of the posterior probability distributions. This enables us to overcome the cost of traditional Bayesian optimal design methodology that requires repeated approximations of the pos-

terior distributions, which are expensive to generate and are prohibitively costly for high dimensional models.

Chapter 1

Introduction

1.1 Hematopoietic System and Cell Lineage

Hematopoiesis is the complex mechanism by which hematopoietic stem cells (HSCs) produce diverse functional myeloid and lymphoid cells through multiple stages of differentiation involving multiple types of intermediate progenitor cells in the bone marrow [52]. Although the hematopoietic system is highly regulated and this regulation ensures the appropriate numbers and proportions of HSCs and mature cells are produced, the details of the regulation are still not well understood. In particular, which hematopoietic cells interact and how they interact with each other is elusive. Understanding cell fate regulation during hematopoiesis can provide insight into new regenerative therapeutic options for blood-related diseases and can help biologists understand the role HSCs play in oncogenesis and aging [31]. In addition, in clinical trials, stem cell transplantation is a well-known treatment for malignant diseases like leukemia since such transplantation can reconstruct the hematopoietic system in a cancer-free manner [30, 64, 34, 58]. Thus, understanding how the hematologic system responds to and recovers from perturbations is important.

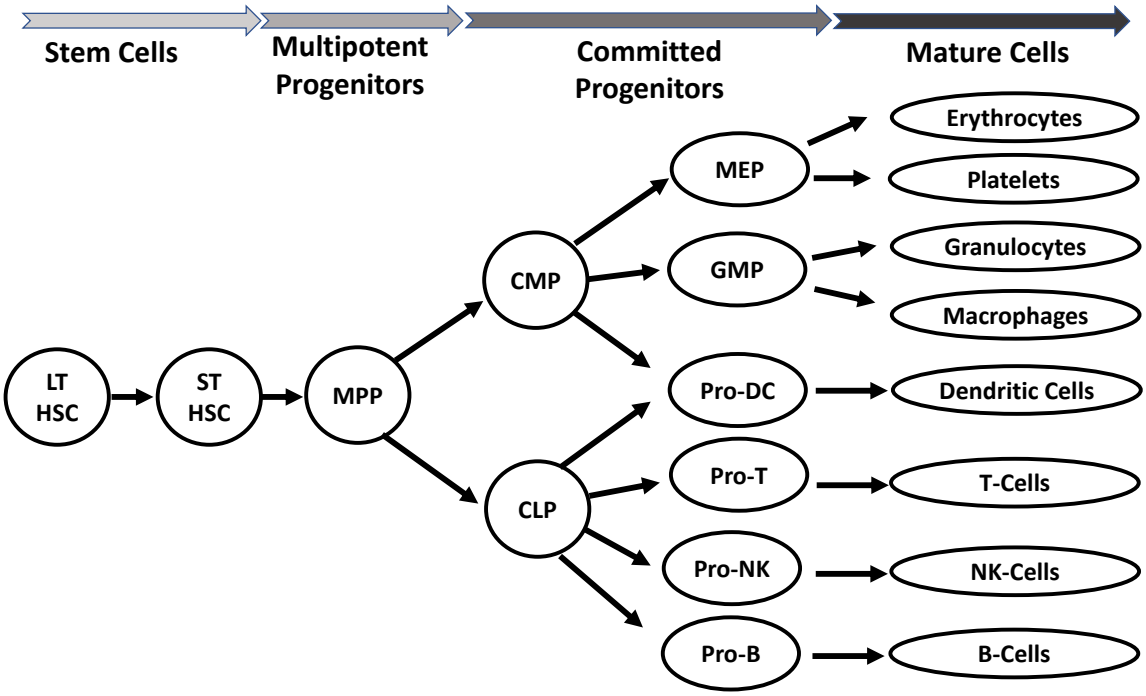


Figure 1.1: Cell lineage of hematopoietic cells described in [52]: from left to right, long-term hematopoietic stem cells (LTHSCs) differentiate into short-term hematopoietic stem cells (STHSCs) that have shorter reconstitution capacity; STHSCs differentiate into multipotent progenitors (MPPs) that can briefly self-renew; MPPs then differentiate into committed progenitor cells: common myeloid progenitor cells (CMPs) and common lymphoid progenitor cells (CMPs); CMPs differentiate into Granulocyte-Macrophage Progenitors (GMPs), Megakaryocte-Erythroid Progenitors (MEPs); cells of CLPs downstream and mature cells are also shown in the figure

The hematopoietic lineage is generated by pluripotent hematopoietic stem cells (HSCs) that are located inside the bone marrow (BM) and have a strong reconstitution capability. HSCs then differentiate into multipotent progenitor cells (MPPs) that may self-renew for a short period of time. MPPs then differentiate into committed progenitor cells and ultimately end with mature myeloid and lymphoid blood cells located in the peripheral blood (PB) [52]. The hematopoietic system produces $10^{10} \sim 10^{11}$ mature blood cells on a daily basis [59]. The hematopoietic system is tightly regulated with secreted biochemical function regulatory cell fate choices and renewal probabilities. Only a few of these biochemical signaling functions are known. However, defining cell fate and renewal regulations is an active area of research.

1.2 Hematopoiesis as a Dynamic System

For about half a century, scientists have been devising mathematical and statistical models that describe the hematopoiesis process both under normal conditions and in diseased states [14, 33, 57, 43, 23]. Traditionally, hematopoiesis has been modeled as a multi-type branching process [46, 70, 78, 20]. A branching process approach represents hematopoiesis as a stochastic compartmental model where cells self-replicate and differentiate [5, 67]. Recently, barcoding (DNA segments are marked to identify cell fate and track cell lineage) based on sequencing technology has become available, and researchers have been able to fit branching process models to data [77, 38, 78]. Lineage tracking strategies allow researchers to inform interpretable probabilistic models of hematopoiesis that provide insight into cell fate decisions. However, in branching process models, cells are independent of each other, and thus feedback regulation cannot be accommodated. Other models for hematopoiesis use ordinary differential equations (ODEs) [30, 47, 9, 15, 27, 45]. These systems can describe the mean-field dynamics (assume that cellular system size is large enough to ignore stochasticity) of the various hematopoietic cell types and typically involve Hill functions that model sigmoidal responses to biological signal [68]. ODE models can explain the dynamic behavior of cell compartments in the hematopoietic lineage. However, the Hill equation is non-linear, and it is not obvious how to extend this approach to a stochastic model. There are also hybrid models, which are combinations of stochastic agent-based and ODE models [64, 34, 62, 17, 37]. Agent-based models are used to describe the population dynamics of small numbers of cells (e.g., stem and progenitor cells), while ODEs are used to describe the dynamics of more mature cell compartments where there are many cells. While some models also account for uncertainty via Bayesian or approximate Bayesian methods for model calibration [78, 28, 22, 50], they typically do not account for feedback regulation that plays an important role in maintaining homeostasis and response to pathological stresses. Thus, although many mathematical models have been developed, fitting differential equation models

to experimental data is still an active area of research, mainly stemming from the fact the data is typically sparse and highly variable due to the inherent stochasticity in hematopoiesis as well as measurement errors.

1.2.1 Motivating Example of Modeling Hematopoiesis

Previously, a nonlinear Hill equation ODE model has been proposed for the dynamics HSCs and MPPs in hematopoiesis dynamics [42] with feedback regulation [16]. This model involves two negative feedback regulations: (i) HSCs negatively regulate the division rate of MPPs [61] and (ii) MPPs negatively feedback on HSC self-renewal probability [69]. This model, schematically represented in Fig (1.2), is described by the following ODEs:

$$\begin{cases} N'_{HSC} = \left(\dfrac{2p_0^*}{1+\gamma_1 \cdot N_{MPP}} - 1\right) \cdot \eta_1 \cdot N_{HSC}, \\ N'_{MPP} = 2(1 - \dfrac{p_0^*}{1+\gamma_1 \cdot N_{MPP}}) \cdot \eta_1 \cdot N_{HSC} + \dfrac{\eta_2}{1+\gamma_2 N_{HSC}} \cdot N_{MPP}, \end{cases} \tag{1.1}$$

where p_0^* is the maximum self-renewal probability of HSCs, γ_1 is the feedback parameter that modulates the effect of MPP cell count on the HSC self-renewal probability, η_1 is HSC total division rate, η_2 is MPPs effective division rate, and the feedback regulation rate from HSCs on η_2 is γ_2.

[42] developed a Bayesian inference ODE model to calibrate the ODE model and learn the model parameters. The cell count datasets are collected by using the bone marrow extraction methods, which are destructive, and consequently, each mouse can contribute their cell count data at a one time point only since mice must be sacrificed during the data collection. The cell counts of each mouse prior to the observation time are thus treated as missing data and consequently need to develop a latent variable approach to interpolate them. Besides, both the sparse control and perturbation experimental dataset in Figure 1.3 (a-b) show a large

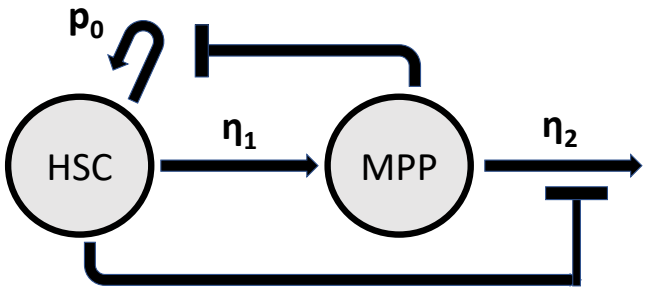
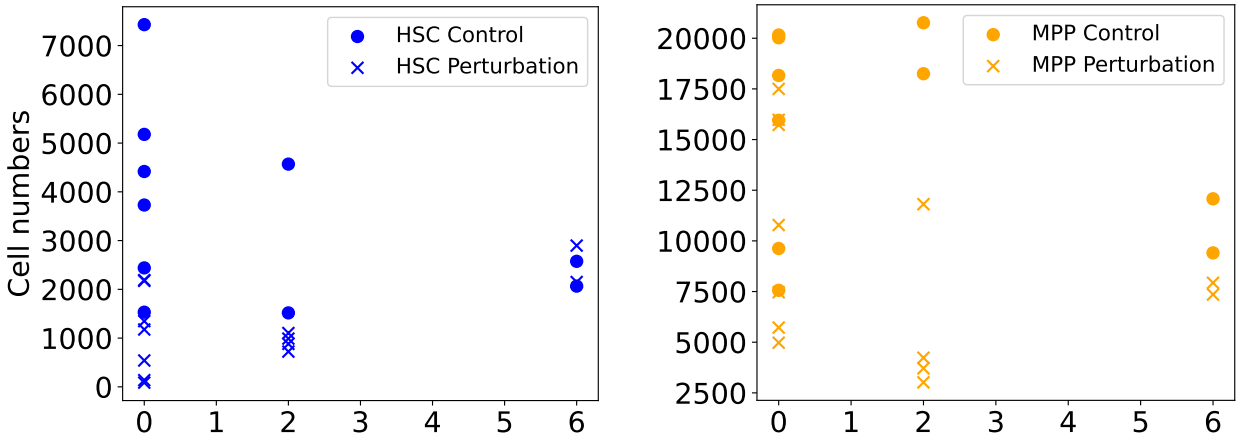


Figure 1.2: Diagram of a two-compartment Hill equation model for hematopoiesis dynamics: MPPs negatively feedback on the HSC self-renewal probability p_0; HSCs negative feedback on the MPPs effective division rate η_2;

data heterogeneity across different cell count magnitudes at observation days 0, 2, 6. These highly induce that there should be more variability in rate value parameters among different mice, and the hematopoiesis dynamic should be a stochastic process when recovering to the hematopoietic equilibrium once the system is perturbed.



(a) HSC Dataset for motivation example

(b) MPP Dataset for motivation example

Figure 1.3: Experiemntal data using by motivating example 2-compartment Hill equation ODE model. Dots refer to the control cell count data, and cross marks refer to the data after perturbation. Y-axis refers to the number of cell counts, and the x-axis refers to the day after the first observation. Both control and perturbation data show large variability.

1.2.2 Motivation of Projects

The variability of the hematopoietic dataset motivates us to build a new hierarchical Bayesian framework to quantify the biological uncertainty in rate value parameters, which could vary from one individual to, another and the state of the system just prior to the treatment. Alternatively, biological variability can be accounted for by replacing the deterministic ODE model with stochastic models. However, it is nontrivial how to extend the Hill equation (HE) ODE model to a stochastic model. In addition, most SDE models do not yield the tractable likelihoods needed for Bayesian inference. Thus, in the third chapter, based on the motivating differential equation model Eqn.(1.1), we develop a new chemical reaction model for hematopoiesis that can be extended to an SDE (see section 2.2.2 for details). We then derive a new hierarchical Bayesian framework that takes model parameter variability across subjects into account. After checking that model selection methods can reliably identify the right model compared to the model in which the parameters are shared across individuals, we find that the experimental data are best described by the hierarchical chemical reaction model. As an alternative approach to explain data heterogeneity, instead of letting hematopoietic transition rate parameters vary across subjects, we replace the deterministic ODE model with a stochastic model with a tractable likelihood. We probe the difference between ODE-based and LNA-based inference by performing a series of simulation studies using a synthetic hematopoietic dataset where initial condition cell counts are perturbed from their homeostatic values.

Since our data collection method, which involves doing bone marrow extraction operation for each mouse, is destructive and consequently expensive, we also want to select the optimal Bayesian experimental design to optimize the information gain from the partially observed data while controlling the experimental costs. Traditional methods for Bayesian experimental design require the repeated calculations of posteriors probability distributions of the parameters, which is computationally prohibitive for high-dimensional models. We employ

an adversarial method that uses a game theory framework for experimental design without calculating posteriors. This reduces the optimal design problem to a nested minimax optimization problem which can be solved by a gradient descent-ascent algorithm. We test this approach with different Bayesian settings and regulation conditions. By evaluating the optimal designs generated by the optimization algorithm using Bayesian utility, we find that the minimax optimization algorithm using the game theory framework reliably returns optimal experimental designs.

Chapter 2

Technical Background

2.1 Bayesian Inference and Markov Chain Monte Carlo

2.1.1 Bayesian Inference

Bayesian inference gained popularity in the physical and biological sciences [76, 29, 44, 63] because of its ability to incorporate various sources of uncertainty and empirical prior knowledge about model parameters. Let $\boldsymbol{\theta}$ denote parameters of interest governing the distribution of observed data $\boldsymbol{Y}$. Then the uncertainty of parameters $\boldsymbol{\theta}$ can be quantified by Bayesian inference through posterior probability distribution:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{\pi(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{Y})}, \tag{2.1}$$

where $\pi(\boldsymbol{Y}|\boldsymbol{\theta})$ is the likelihood function calculated using sampling distributions of $\boldsymbol{Y}$ and $\pi(\boldsymbol{\theta})$ is the given prior distribution of $\boldsymbol{\theta}$ based on previous empirical knowledge gained before seeing $\boldsymbol{Y}$. Since the marginal distribution of data $\boldsymbol{Y}$, $\pi(\boldsymbol{Y}) = \int \pi(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}$, does not

depend on $\boldsymbol{\theta}$, we often work with the unnormalized posterior distribution,

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \pi(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{2.2}$$

In many realistic cases, such as working with a partially observed dataset, the intermediate state unobserved data can also be treated as latent variables $\boldsymbol{X}$ and for each intermediate state. Accordingly, diagramed in Figure 2.1. The sampling distributions can be written as $Pr(\boldsymbol{X_i}|\boldsymbol{X_{i-1}}, \boldsymbol{\theta})$. Then likelihood function is

$$Pr(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta})) = \prod_{j=1}^{N} Pr(\boldsymbol{Y_j}|\boldsymbol{X}, \boldsymbol{\theta})) = \prod_{j=1}^{N} Pr(\boldsymbol{Y_j}|\boldsymbol{X_j}, \boldsymbol{\theta}) \prod_{i=2}^{M_j} Pr(\boldsymbol{X_j^{t_i}}|\boldsymbol{X_j^{t_{i-1}}}, \boldsymbol{\theta}) Pr(\boldsymbol{X_j^{t_1}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.3}$$

where t_i for $i = 1, \cdots, M_j$ denote the unobserved intermediate times, j denotes the individual (mouse), and $\boldsymbol{X}_j^{t_i}$ denotes the corresponding data. Further $\boldsymbol{Y_j}$ denotes the partially observed data for mouse j. The augmented posterior consists of parameter $\boldsymbol{\theta}$ and latent variable $\boldsymbol{X}$ is

$$Pr(\boldsymbol{\theta}, \boldsymbol{X}|\boldsymbol{Y})) \propto Pr(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.4}$$

Suppose we are interested in a function $f(\boldsymbol{\theta})$, the expected posterior distribution can be represented as

$$E_{\boldsymbol{\theta}|\boldsymbol{Y}}(f(\boldsymbol{\theta})) = \int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}. \tag{2.5}$$

This intractable integral can be approximated by Monte Carlo integration. Suppose parameters $\boldsymbol{\theta_j}$ are sampled independently from the posterior or joint posterior distributions. By the strong law of large numbers, Monte Carlo estimates of the mean and variance can,

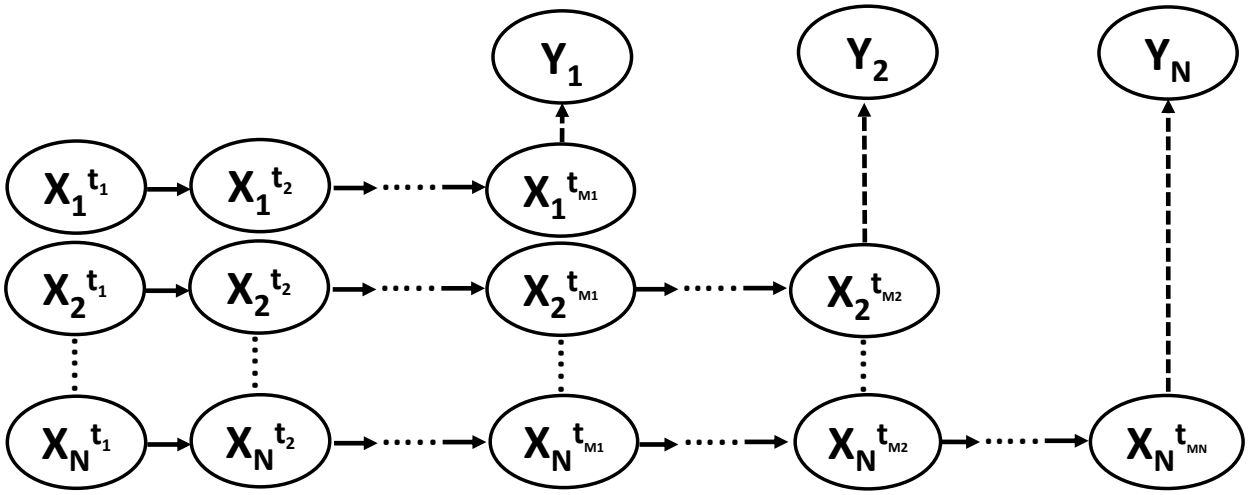


Figure 2.1: Diagram of hidden Markov model for partially observed dataset. $\boldsymbol{Y_n}$ refers to the discrete time partially observed data for mouse n where $\forall I \in 1, \cdots, N$ $\boldsymbol{X_n^{t_i}}$ is the sequence of latent process unobserved intermediate state variables for observed data $\boldsymbol{Y_n}$ for mouse n. t_{Mn} is the observation time for data for mouse n.

respectively, be written as,

$$\boldsymbol{\mu_n} = \sum_1^n f(\boldsymbol{\theta_j}), \text{ where } \lim_{n\to\infty} \boldsymbol{\mu_n} \to E_{\boldsymbol{\theta}|\boldsymbol{Y}}(f(\boldsymbol{\theta})), \tag{2.6}$$

$$\boldsymbol{\delta_n^2} = \frac{1}{n}\sum_1^n (f(\boldsymbol{\theta_j}) - \boldsymbol{\mu_n})^2, \text{ where } \lim_{n\to\infty} \boldsymbol{\delta_n^2} \to \mathrm{Var}_{\boldsymbol{\theta}|\boldsymbol{Y}}(f(\boldsymbol{\theta})), \tag{2.7}$$

and the posterior distribution of the ordinary Monte Carlo estimate can be calculated by the central limit theorem,

$$\frac{\boldsymbol{\mu_n} - E_{\boldsymbol{\theta}|\boldsymbol{Y}}(f(\boldsymbol{\theta}))}{\boldsymbol{\delta_n}/\sqrt{n}} \to Z \sim N(0,1) \tag{2.8}$$

where $\mathrm{Var}(\boldsymbol{\mu_n}) = \boldsymbol{\delta_n}^2/n$.

In general, for complex problems, one cannot derive the closed form solutions of the posterior distributions or draw independent samples from the posterior distribution. Thus, it is necessary to use Markov Chain Monte Carlo (MCMC) numerical approximation to draw a series of dependent samples that will converge to the target distribution [49].

2.1.2 Markov Chain Monte Carlo

In Bayesian inference, MCMC is a framework for numerical integration that approximates the posterior sample distribution and is especially useful for models with large numbers of parameters $\boldsymbol{\theta}$, whose posterior $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ or $\pi(\boldsymbol{\theta}, \boldsymbol{X}|\boldsymbol{Y})$ cannot be sampled directly through ordinary Monte Carlo methods. The core idea of MCMC is to construct an ergodic Markov chain $\{\boldsymbol{\theta_1}, ..., \boldsymbol{\theta_n}\}$ with the stationary distribution equal to the posterior distribution. The selected transition kernel is $K(\boldsymbol{\theta}, \boldsymbol{\theta'})$. To preserve the stationary distribution $K(\boldsymbol{\theta}, \boldsymbol{\theta'})$ gives the transition probability/density from the current state $\boldsymbol{\theta}$ to the next potential state $\boldsymbol{\theta'}$ is required to satisfy the global balance condition:

$$Pr(\boldsymbol{\theta}) = \int K(\boldsymbol{\theta}, \boldsymbol{\theta'}) Pr(\boldsymbol{\theta'}) d\boldsymbol{\theta'}, \tag{2.9}$$

where $Pr(\boldsymbol{\theta})$ is the stationary distribution of state $\boldsymbol{\theta}$. In general, it is difficult to directly verify that the global balance condition is satrisfied. An alternative sufficient condition is called detailed balance,

$$Pr(\boldsymbol{\theta}) K(\boldsymbol{\theta'}, \boldsymbol{\theta}) = Pr(\boldsymbol{\theta'}) K(\boldsymbol{\theta}, \boldsymbol{\theta'}) \text{ for } \forall \boldsymbol{\theta}, \boldsymbol{\theta'}. \tag{2.10}$$

Based on the ergodic theorem, the mean of the integrable function $f(\boldsymbol{\theta})$ will converge to its target Monte Carlo estimate mean Eqn.(2.6). Due to the autocorrelation of MCMC samples, the Monte Carlo variance can be written as,

$$\boldsymbol{\delta_n^2} = \frac{1}{n} \frac{1}{n_{eff}} \sum_{1}^{n} (f(\boldsymbol{\theta_j}) - \boldsymbol{\mu_n})^2, \tag{2.11}$$

where $n_{eff} = n/(1+2\sum_{\ell=1}^{\infty} \hat{p}(\ell))$ is the effective sample size that measures the effectiveness of a sample chain and $\hat{p}(\ell)$ is the estimated autocorrelation at lag ℓ. To work with hierarchical Bayesian models whose parameters vary over different magnitudes, we consider a MCMC

chain sampler that can be free of parameter tuning. In the next sections, we will give a brief overview of the main MCMC algorithm we use here that can adaptively choose the hyperparameters and avoid undesirable random behavior and extra inefficient computation.

Hamiltonian Monte Carlo Sampler

A Hamiltonian Monte Carlo (HMC) sampler [7, 8] is a MCMC algorithm that uses the samples of density function derivatives to generate efficient transitions $K(\boldsymbol{\theta}, \boldsymbol{\theta'})$ spanning the posterior. This approach uses an auxiliary momentum vector variable ρ sampled from the joint posterior distribution:

$$p(\rho, \theta) = p(\rho)p(\theta), \tag{2.12}$$

where ρ is sampled from a multivariate normal distribution that does not depend on θ and $p(\theta)$ is a density for parameter θ (which is typically Bayesian posterior for θ),

$$\rho \sim MvN(0, \tilde{\Sigma}), \tag{2.13}$$

where $\tilde{\Sigma}$, usually estimated from MCMC warm-up steps, acts as a change of basis matrix for the target distribution. Then the Hamiltonian can be defined as,

$$H(\rho, \theta) = -\log\ p(\rho, \theta) = -\log\ p(\rho) - \log p(\theta) \tag{2.14}$$

let $T(\rho|\theta) = -\log p(\rho, \theta)$ denote the kinetic energy and $V(\theta) = \log p(\theta)$ denote the potential energy. Then, the transition derivative from the current state θ can be written as,

$$\frac{d\theta}{dt} = \frac{\partial T}{\partial \rho}, \quad \frac{d\rho}{dt} = -\frac{\partial V}{\partial \theta} \tag{2.15}$$

HMC solves this differential equation system by the leapfrog integrator algorithm, which takes discrete steps in small time intervals ϵ. In this method, a new initial momentum vector

ρ is drawn at every iteration step. A step of the Leapfrog algorithm then updates momentum and the parameters of interest following the rule,

$$\begin{aligned}
\rho(t+\frac{\epsilon}{2}) &= \rho(t) - \frac{\epsilon}{2}\frac{\partial V}{\partial \theta}(\theta(t)), \\
\theta(t+\epsilon) &= \theta(t) + \epsilon \sum \rho(t+\frac{\epsilon}{2}), \\
\rho(t+\epsilon) &= \rho(t+\frac{\epsilon}{2}) - \frac{\epsilon}{2}\frac{\partial V}{\partial \theta}(\theta(t+\epsilon).
\end{aligned} \tag{2.16}$$

The integrator algorithm then applies L number of leapfrog steps and changes the momentum and parameter to (ρ^*, θ^*) in time $L\ \epsilon$. Finally, the Metropolis acceptance step is used to account for numerical errors during the integrator algorithm by accepting or rejecting (ρ^*, θ^*) with probability

$$\min(1, \exp(H(\rho, \theta) - H(\rho^*, \theta^*)))$$

In general, the computational efficiency of HMC is highly affected by the selection of the discretization time size ϵ and the number of leapfrog steps L. If ϵ is too small, then too many small steps are required to be calculated per interval. If ϵ is too large, then the leapfrog numerical integrator will be inaccurate, and the updates will be frequently rejected. If L is too small, then the leapfrog trajectory per iteration will be very short, leading to random walk behavior. If L is too large, then the leapfrog algorithm needs to trace out too long trajectories. Thus, HMC requires an adaptive algorithm to avoid inefficient computations. An undesirable selection of either of these two ϵ or L will lead to a dramatic slow down in MCMC computational performance.

No-U-Turn Sampler

The No-U-Turn sampler (NUTs) can adaptively choose an appropriate number of leapfrog steps L in each MCMC iteration to enable the updates to span the posterior distributions efficiently, and the method can also adapt the discretization time size ϵ through primal-

dual averaging. The motivation for NUTs is to avoid the slow exploration of the state space during which HMC performs random walk behavior and to maximize the expected squared jump distance [25] without any "hand engineering" of tuning parameters. Similarly to HMC, NUTS also first generates an update based on the last iteration and samples a new momentum vector $\boldsymbol{\rho}$ from the multivariate normal distributions. Instead of HMC's Metropolis correction to deal with the numerical integrator error, NUTs first samples an entire trajectory that integrates both forward and backward in time to generate a balanced binary tree for detail balance criteria from the set of numerical trajectories and then samples updates based on appropriately chosen weight functions. In each iteration of the NUTs algorithm, the tree depth is increased by one, which leads to a doubling of the number of leapfrog steps L. Configuring the NUTs sampler places a cap on the maximum tree depth. The NUTs algorithm ends either when the depth of the tree reaches a given maximum depth (default is ten which leads to 2^{10} leapfrogs) or there is a U-turn in Euclidean space on a subtree or the completed tree. Therefore, this algorithm restricts the number of leapfrogs per iteration as follows:

$$L \leq 2^{\textbf{maximum tree depth}} - 1.$$

The NUTs proposal acceptance probability can be written as

$$\frac{1}{|\mathcal{B}_t^{final}|} \sum_{\theta, r \in \mathcal{B}_t^{final}} min \left\{ 1, \frac{p(\theta, r)}{p(\theta^{t-1}, r^{t,0})} \right\},$$

where $\mathcal{B}_t^{final}$ is the set of states explored during the final doubling of the maximum tree depth, $\theta^{t-1}, \rho^{t,0}$ are the initial state and momentum for t^{th} iteration of the Markov chain. NUTs then adaptively choose a reasonable initial parameter value ϵ according to a heuristic algorithm that repeatedly doubles or halves ϵ until the acceptance probability of a Langevin update with current step size ϵ crosses 0.5, making sure that ϵ is small enough for the NUTs algorithm to approximate the Monte Carlo integral accurately without too much computation.

2.2 Bayesian Inference with Differential Equations

Mathematical differential equations modeling is frequently used to understand the time evolution of biological systems at the population level and to explain the complex system dynamic behavior when cell compartment numbers are perturbed from their equilibrium. These differential equations contain parameters that need to be specified. Popular approaches to estimate these parameters included machine learning methods (e.g., neural ODEs, optimization) and probabilistic learning methods (e.g., Bayesian inference). Machine learning methods typically require a lot of training data that are always not available. In such situations, Bayesian inference is more attractive because it can quantify parameter uncertainty without a large training dataset and can incorporate prior knowledge about those parameters. Also, deterministic methods may provide a best fit parameter set, but small errors in data could lead to large differences in the optimal parameter values. Bayesian parameter inference consists of two parts: a forward part that numerically integrates the differential model and simulates it with specific parameters and initial conditions, and a backward part that the parameters of interests change based on the corresponding posterior density function value which is the combination of the likelihood density value and prior density value. In this section, we review Bayesian inference with different types of differential equation models and will use a simple one-dimensional model to gain some insights into strengths and limitations. Generalizations to higher dimensions are similar, but the notation for the one-dimensional model is much simpler.

2.2.1 ODE Parameter Inference

To introduce the concept of Bayesian inference with an ODE model, we consider a simple logistic population growth model. Let the time-varying total population size be $P(t)$ with initial condition $P(0) = P_0$. Then, the population growth rate, $kP(t)$ can represent the ex-

ponential population growth tendency without any limitations, where k is the rate constant. Let the total capacity of the environment be K. A negative feedback term from the total capacity of the population size growth rate can be written as $1 - P/K$. The population changes through time according to the corresponding logistic differential equation model:

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{K}\right), \quad P(t_0) = P_0 \tag{2.17}$$

Let $\boldsymbol{S}$ denote the stoichiometry matrix that provides information about reactants and products in model reactions, where each row corresponds to a reaction compartment, and each column refers to a certain reaction channel. Let $\boldsymbol{h}(t)$ be the reaction propensity vector that gives the probability of a reaction occurring in the time interval (t,t+Δt). The logistic model can also be represented with the following biochemical reactions,

$$\mathcal{R}_1 : P \xrightarrow{k} 2P; \quad \mathcal{R}_2 : P + P \xrightarrow{k/K} \emptyset, \tag{2.18}$$

with $\boldsymbol{S} = [[1, -1]]$ the 1×2 stoichiometry matrix and $\boldsymbol{h} = [kP(t), (k/K)P(t)^2]^t$ the 2×1 reaction propensity vector. The logistic differential equation model can then be written in the chemical reaction form,

$$\frac{dP}{dt} = \boldsymbol{S}\boldsymbol{h}(t) \tag{2.19}$$

By using separation of variables, we can obtain the solution to the logistic differentiation equation:

$$P(t) = \frac{K}{1 + Be^{-kt}}, \quad \text{where } B = \frac{K - P_0}{P_0}. \tag{2.20}$$

We can see that the equilibrium solution when $\frac{dP}{dt} = 0$ of the ODE is

$$\overline{P} = \lim_{t\to\infty} P(t) \to K \text{ stable equilibrium or } \overline{P} = 0. \tag{2.21}$$

In the next subsection, we will see that this logistic differential equation model can be regarded as a deterministic approximation of a discrete stochastic Markov process. Our parameters of interest $\boldsymbol{\theta}$ in this model are k, K, and P_0.

2.2.2 SDE Parameter Inference

Biological processes such as hematopoiesis are stochastic in nature [79, 64, 34], and one of the classic stochastic models for these biological processes is the Markov Jump Process (MJP).

Markov Jump Process

MJP assumes the population dynamics follow a continuous time Markov chain (CTMC) model with d-dimensional discrete state space $\mathcal{S}$. Let i, j be the state in $\mathcal{S}$ and the transition probability that the chain in further state j given current i after time unit $t \geq 0$ be p_{ij}(t). Then the $d \times d$ dimensional transition matrix can be written as $\boldsymbol{p}(t) = \{p_{ij}(t)\}$ where $\boldsymbol{p}(0) = \boldsymbol{I}$ and each row sums up to 1. Let $\Delta t \geq 0$ be a time increment, then the transition matrix satisfies the Chapman-Kolmogorov equation [48]:

$$\boldsymbol{p}(t + \Delta t) = \boldsymbol{p}(t)\boldsymbol{p}(\Delta t), \tag{2.22}$$

and for each pair of states $\{i, j\}$:

$$\boldsymbol{p}_{ij}(t + \Delta t) = \sum_{k\in\mathcal{S}} \boldsymbol{p}_{ik}(t)\boldsymbol{p}_{ik}(\Delta t), \tag{2.23}$$

Let Λ be the transition rate matrix for the MPJ, then the transition rate from state i to j $(j \neq i)$ can be written as

$$\Lambda_{ij} = \lim_{\Delta t \to 0} \frac{\boldsymbol{p_{ij}}(\Delta t)}{\Delta t}. \tag{2.24}$$

If CTMC is stable, then $\Lambda_{ij} < \infty$ for $\forall i, j \in \mathcal{S}$. If CTMC is also conservative, then

$$\Lambda_{ii} = -\sum_{j \in \mathcal{S}, j \neq i} \Lambda_{ij}, \tag{2.25}$$

for $i \in \mathcal{S}$. By using the Kolmogorov backward equation [48] we obtain

$$\frac{d\boldsymbol{p}}{dt} = \Lambda \boldsymbol{p}(t), \text{ where } \boldsymbol{p}(0) = \boldsymbol{I} \tag{2.26}$$

and since the chain we consider is ergodic (it is possible to go from every state to every state), that equation has and converges to a unique solution. Then the unique solution can be written as:

$$\boldsymbol{p}(t) = exp(\Lambda(t)) = \sum_{k=0}^{\infty} \frac{(\Lambda(t)^k)}{k!}. \tag{2.27}$$

Let us consider the Markov jump process of the Logistic differential equation from section 2.2.1, then the transition and related rate can be written as:

$$\Lambda_{P \to P'} = \begin{cases} kP(t), & \text{if } P' = P + 1 \\ k/K \cdot P(t)^2, & \text{if } P' = P - 1 \\ -k/K \cdot P(t)^2 - kP(t), & \text{if } P' = P \\ 0, & \text{otherwise} \end{cases} \tag{2.28}$$

The solution of the MJP can be simulated using the Gillespie algorithm [26]. However, if the population size is very large, then calculating transition probability can be computationally prohibitive [36]. Besides, if the rates are not constant in time, calculating the transition

probability used in the Gillespie algorithm becomes even more challenging. Thus, we want to use an alternative approach for simulations.

Derivation of SDE

Deterministic models such as ordinary differential equations (ODE), which can be viewed as mean-field approximations of the Markov jump process, usually underestimate the uncertainty of the parameters of interest and can consequently lead to inaccurate inferences. A standard computational approach is to generate realizations from the Gillespie method, but there is not a direct connection between the chemical master equation Eqn.(2.19) and the Gillespie sample path of the process. Another method of accelerating stochastic process simulations is to simulate from the diffusion approximation of the true process. The diffusion approximation refers to the continuous state Markov process with continuous time evolution. Let Y be the d-dimensional Ito diffusion process, then the stochastic differential equation can be written as

$$dY(t) = \mu(Y(t))dt + \Sigma(Y(t))dW_t, \tag{2.29}$$

where $\mu : \mathbb{R}^d \to \mathbb{R}^d$ refers to the drift vector function and $\Sigma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ refers to the diffusion matrix function. This diffusion process SDE can be regarded as a d-dimensional Wiener process (Brownian motion) with drift. The SDE can be regarded as the limit of the diffusion process when the time finite interval Δt tends to zero. Then, the increment in a time interval Δt can be written as

$$\Delta Y(t) = \mu(Y(t))\Delta t + \Sigma(Y(t))\Delta W_t, \tag{2.30}$$

where $\Delta W_t \sim N(0, \Delta t)$ is the increment in a multivariate Brownian motion. The SDE solution can be approximated by numerically solving $\Delta Y(t)$ using the Euler-Maruyama algorithm [2].

The Random Time-Change Representation

From the previous section, we can see that reactions can be represented by the chemical master equation and can be simulated by the Gillespie algorithm. However, there is no direct relation between the chemical master equation Eqn.(2.19) and the simulation path X_t of the Gillespie algorithm. Another equation to help mathematically analyze the Markov jump process is called random time-change representation [40], which aims to fill this gap and directly connect the sample path to the master equation. Let $\tilde{N}_i(t)$,i$\in [1, 2, .., d]$ be the count functions for d independent unit Poisson processes

$$\tilde{N}_i(t) \sim Po(h_i(X_t, \theta_i)t) \text{ for all } t \tag{2.31}$$

and $h_i(X_t, \theta_i)$ is the event hazard function (rate law) of reaction i, which measures the event density, and θ_i is the rate parameters for reaction i, then the number of reaction events i happens up to time t can be written as

$$R_{it} = \tilde{N}_i \overbrace{\left(\int_0^t (h_i(X_\tau, \theta_i))d\tau\right)}^{culmulative\ reaction\ propensity} \quad \text{for } \forall i \in [1, 2, .., d]. \tag{2.32}$$

Let $\tilde{N} = (\tilde{N}_1, \tilde{N}_2, ..., \tilde{N}_d)$ and $R_t = (R_{1t}, ..., R_{dt})$ then

$$R_t = \tilde{N}\left(\int_0^t (h_i(X_\tau, \theta))d\tau\right). \tag{2.33}$$

Let X_0 be the initial condition of the sample path, S be the stoichiometry matrix, then the master equation model satisfies the random time-change representation stochastic integral equation

$$\boldsymbol{X}_t - \boldsymbol{X}_0 = \boldsymbol{S}R_t = \boldsymbol{S}\tilde{N}\left(\int_0^t (h_i(X_\tau, \theta))d\tau\right). \tag{2.34}$$

Chemical Langevin Equation

Let's approximate the Poisson distribution above by the Gaussian distribution; then the Poisson process can be approximated by

$$\tilde{N}_i(t) \approx t + W_i(t), \tag{2.35}$$

where $W_i(t)$ refers to the independent Wiener process for reaction type i. The time-change representation then can be written as

$$\boldsymbol{X}_t - \boldsymbol{X}_0 \approx \boldsymbol{S}\left[\int_0^t (h(X_\tau, \theta))d\tau + \boldsymbol{W}\left(\int_0^t (h(X_\tau, \theta))d\tau\right)\right], \tag{2.36}$$

and the stochastic differential equation can be written as

$$d\boldsymbol{X}_t = \boldsymbol{S}\boldsymbol{h}(\boldsymbol{X}_\tau, \boldsymbol{\theta}) + \boldsymbol{S}Diag\{\sqrt{\boldsymbol{h}(\boldsymbol{X}_\tau, \boldsymbol{\theta})}\}d\boldsymbol{W}_t, \tag{2.37}$$

where $d\boldsymbol{W}_t$ is the increment of a d-dimensional Wiener process corresponding to the d-dimensional reaction propensity vector. Diag$\{\sqrt{\boldsymbol{h}_i(\boldsymbol{X}_\tau, \boldsymbol{\theta})}\}$ refers to the diagonal matrix with a leading diagonal defined by the vector $\{\sqrt{\boldsymbol{h}_i(\boldsymbol{X}_\tau, \boldsymbol{\theta})}\}$. Note that this equation reduces to the ODE model described in the previous section by letting $d\boldsymbol{W}_t = \boldsymbol{0}$

Since typical chemical master equations usually have more reaction types, $\dim(\boldsymbol{W}_t(v))$ than reactant compartments, $\dim(\boldsymbol{X}_t(t))$, by multivariate statistics, we can rewrite the CLE as

$$d\boldsymbol{X}_t = \boldsymbol{S}\boldsymbol{h}_i(\boldsymbol{X}_\tau, \boldsymbol{\theta}) + Diag\{\sqrt{\boldsymbol{S}\boldsymbol{h}_i(\boldsymbol{X}_\tau, \boldsymbol{\theta})\boldsymbol{S}^t}\}d\boldsymbol{W}_t, \tag{2.38}$$

where now the dimension of $(\boldsymbol{W}_t)$ is equal to the dimension of $(\boldsymbol{X}_t)$.

Let us consider a stochastic version of the Logistic differential equation from section 2.2.1:

$$dP = kP\left(1 - \frac{P}{K}\right)dt + \sqrt{kP\left(1 + \frac{P}{K}\right)}dW_t. \tag{2.39}$$

The limitation of SDE inference is that the likelihood is usually intractable. This makes it challenging for the NUTS MCMC algorithm to converge since the gradient is changing in every calculation. Some possible methods for dealing with this include using the NUTS algorithm with a low target acceptance rate or using a stochastic gradient Hamilton Monte Carlo (SGHMC) algorithm [12].

2.2.3 LNA Parameter Inference

Motivating the using of LNA

In general, the stochastic simulation trajectory distribution is not Gaussian and the transition density $K(\boldsymbol{\theta}, \boldsymbol{\theta'})$ is unknown. The linear noise approximation (LNA) is derived from the stochastic differential equation (SDE) by performing the first order of Ito-Taylor expansion and taking the first order contribution. With the Gaussian transition density, LNA can decompose stochastic processes into separate mean and variance evolution processes. LNA can also be regarded as a compromise between ODEs and SDEs. In certain situations, LNA inference is more statistically efficient than ODE inference since it takes more stochasticity into account LNA is more computationally efficient than SDE inference where the likelihood is not computationally tractable and the posterior is sensible through the sensitivity analysis.

Derivation of LNA

LNA will be the methodological basis for the hierarchical Bayesian inference for stochastic modeling of hematopoiesis presented in Chapter 4. Here we provide an informal derivation of LNA from CLE Eqn.(2.38) based on [2, 75]. The general idea of LNA is to decompose the sample path $X(t)$ into a deterministic ODE process $\eta(t)$ and a stochastic Gaussian process M(t),

$$X(t) = \eta(t) + M(t). \tag{2.40}$$

Let the Jacobian matrix of the deterministic part $\eta(t)$, be

$$J(\eta(t), \theta) = \frac{\partial Sh(\eta(t), \theta)}{\partial \eta(t)}. \tag{2.41}$$

Consider rescaling compartment counts and transition rates based on concentrations. Let v be the system volume, and

$$\tilde{X}(t) = X(t)/v, \quad \tilde{\theta}_i = \theta_i / v^{1-m_i}, \tag{2.42}$$

where m_i is the reactants number in reaction i. Let $\tilde{\eta} = \eta/v$ be the rescaled deterministic process solution. Then, it can be shown that the stoichiometry matrix and reaction propensity vector forms can also used in the scaled version of the chemical master equation

$$d\tilde{\eta}(t) = Sh(\tilde{\eta}(t), \tilde{\theta}). \tag{2.43}$$

Then, if the rate laws $h(X_\tau, \theta)$ for the reactions are also in mass action form e.g., $(\theta_1 \cdot X_1 \cdot X_2)$, then we will get

$$h(\tilde{X}, \tilde{\theta}) = h(X, \theta)/v, \quad J(\tilde{\eta}(t), \tilde{\theta}) = J(\eta(t), \theta). \tag{2.44}$$

Then, the rescaled version of the CLE Eqn.(2.38) using Eqn.(2.43) and Eqn.(2.44) can be rewritten as

$$d\tilde{X} = Sh(\tilde{X}, \tilde{\theta})dt + \frac{1}{\sqrt{v}}\sqrt{S\ Diag(h(\tilde{X}, \tilde{\theta}))\ S^t}dW_t. \tag{2.45}$$

Let $\tilde{M} = (X(t) - \eta(t))/\sqrt{v}$ be the rescaled residual, then it can be shown that the rescaled version of Eqn.(2.40) is

$$\tilde{X}(t) = \tilde{\eta}(t) + \frac{1}{\sqrt{v}}\tilde{M}. \tag{2.46}$$

Then performing a Taylor expansion on the CLE Eqn.(2.45) around η, we can obtain a linearized SDE:

$$\begin{aligned} d\tilde{X} &= S\ h(\tilde{\eta} + \frac{1}{\sqrt{v}}\tilde{M}, \tilde{\theta})dt + \frac{1}{\sqrt{v}}\sqrt{S\ Diag(h(\tilde{\eta} + \frac{1}{\sqrt{v}}\tilde{M}, \tilde{\theta})\ S^t}dW_t \\ &\approx S\left(h(\tilde{\eta}, \tilde{\theta}) + J(\tilde{\eta}, \tilde{\theta})\frac{1}{\sqrt{v}}\tilde{M} + \mathcal{O}(\frac{1}{v})\right)dt + \frac{1}{\sqrt{v}}\sqrt{S\ Diag(h(\tilde{\eta}, \tilde{\theta}) + \mathcal{O}(\frac{1}{\sqrt{v}}))\ S^t}dW_t \\ &= S\left(h(\tilde{\eta}, \tilde{\theta}) + \frac{1}{\sqrt{v}}J(\tilde{\eta}, \frac{1}{\sqrt{v}}\tilde{\theta})\psi\right)dt + \frac{1}{\sqrt{v}}\sqrt{S\ Diag(h(\tilde{\eta}, \tilde{\theta})\ S^t}dW_t + \mathcal{O}(\frac{1}{v})dt + \mathcal{O}(\frac{1}{\sqrt{v}}))dW_t, \end{aligned} \tag{2.47}$$

Since $\tilde{M} = \sqrt{v}(\tilde{X}(t) - \tilde{\eta}(t))$, and using Eqn.(2.47), the derivative of the rescaled residual can be written as

$$d\tilde{M} = J(\tilde{\eta}, \tilde{\theta})\tilde{M}dt + \sqrt{S\ Diag(h(\tilde{\eta}, \tilde{\theta})\ S^t}dW_t + \mathcal{O}(\frac{1}{v})dt + \mathcal{O}(\frac{1}{\sqrt{v}}))dW_t. \tag{2.48}$$

Multiplying the Eqn.(2.48) by $\sqrt{v}$ and change back to the normal scale using Eqn.(2.46), we get:

$$\begin{aligned}\frac{1}{\sqrt{v}}d(X(t)-\eta(t)) = \frac{1}{\sqrt{v}}\left(J(\eta,\theta)(X(t)-\eta(t))+\mathcal{O}(1)\right)dt \\ + \left(\frac{1}{\sqrt{v}}\sqrt{S\cdot Diag(h(\eta,\theta))\cdot S^t}+\mathcal{O}(1)\right)dW_t.\end{aligned} \tag{2.49}$$

By the local Lipshitz property in time of the Jacobian J(η,θ) and the stochastic part $S\cdot Diag(h(\eta,\theta))\cdot S^t$, we can obtain

$$X_t = \eta_t + M_t + \mathcal{O}(\frac{1}{\sqrt{v}})dt. \tag{2.50}$$

Next we derive the solution for X_t, η_t and M_t. Based Eqn.(2.50), for large system size v, LNA can approximate the original sample path $X(t)$ since

$$X_t^{LNA} = \eta_t + M_t. \tag{2.51}$$

Consider the SDE for the residual process M based on the rescaled residual process $\tilde{M}$ Eqn.(2.48):

$$dM = J(\eta,\theta)Mdt + \sqrt{S\ Diag(h(\eta,\theta)\ S^t}dW_t. \tag{2.52}$$

We can solve dM by further decomposing the equation into mean stochastic and variance stochastic differential equations. To achieve this, first, we need the linear ordinary solution equation system solution and Itô's lemma [1].

Lemma 2.1. *Solution of Linear Ordinary Differential Equation System*
Let $\boldsymbol{F}(t) \in \mathcal{R}^{d\times d}$ *and* $\boldsymbol{X}_t \in \mathcal{R}^d$ *and the linear ODE is defined as*

$$d\boldsymbol{X}_t = \boldsymbol{F}(t)\boldsymbol{X}_t, \tag{2.53}$$

on $t \in (0, \infty)$ *with initial condition* $\boldsymbol{X}(0) = \boldsymbol{X}_0$*, then the solution can be written as*

$$\boldsymbol{X}_t = B(t, 0)\boldsymbol{X}_0, \tag{2.54}$$

where $\boldsymbol{B}(t, 0) \in \mathcal{R}^{d \times d}$ *refers to the solution of the ODE*

$$d\boldsymbol{B}(t, 0) = \boldsymbol{F}(t)\boldsymbol{B}(t, 0)dt, \tag{2.55}$$

with initial condition $\boldsymbol{B}(0, 0) = \boldsymbol{I}$*.*

Lemma 2.2. *Itô's Lemma*

Let the time evolution process be X(t) $\in \mathcal{R}^d$*, the deterministic and stochastic parts of the SDE be f(t,X(t))* $\in \mathcal{R}^d$ *and g(t,X(t))* $\in \mathcal{R}^{d \times r}$*, respectively. Consider a scalar process Y(t)* $= \phi(t, X(t))$ *where the scalar function* $\phi(t, X(t))$ *is twice continuously differentiable in X(t) and continuously differentiable in t, then*

$$\begin{aligned} dY(t) = (\phi_t(t, X(t)) + \phi_x(t, X(t))f(t, X(t)) + \frac{1}{2}tr(\phi_{xx}(t, X(t))g(t, X(t))g^t(t, X(t))))dt \\ + (\phi_x(t, X(t))g(t, X(t)))\, dW_t, \end{aligned} \tag{2.56}$$

where tr denotes the sum of diagonal terms.

Let $\hat{B}$ be the change of basis matrix that one can get from Lemma 2.1 by solving

$$dM_t = J(t)M_t, \tag{2.57}$$

and let the scalar process be $\hat{M} = \hat{B}^{-1}(t, t_0)M(t)$. Then by applying Lemma 2.2 to Eqn.(2.57), we find

$$d\hat{M} = \hat{B}^{-1}(t, t_0)M(t)\sqrt{S\ Diag(h(\eta, \theta)\ S^t}dW_t, \tag{2.58}$$

and the solution of the residual M(t) based on Eqn.(2.58) can be written as

$$M(t) = \overbrace{\hat{B}(s,t_0)M(t_0)}^{\textbf{deterministic part } m(t)} + \overbrace{\hat{B}(t,t_0)\int_{t_0}^{t}\hat{B}^{-1}(t,t_0)\sqrt{S\ Diag(h(\eta,\theta)\ S^t}dW_s}^{\textbf{linear combination of increment Brownian motions part } \Sigma(t)} . \quad (2.59)$$

The latter term can be regarded as a Gaussian noise term with 0 mean since it is the integral of linear combinations of Brownian motions. Then, the mean process $m(t)$ in Eqn.(2.59) can be written as

$$m(t) = \hat{B}(t,t_0)M(t_0), \quad (2.60)$$

and the corresponding ODE for solving Eqn.(2.57) one can get from Lemma 2.1, is

$$dm(t) = J(\eta,\theta)m(t)dt. \quad (2.61)$$

The variance process can be obtained by taking by the square the stochastic part of $M(t)$ in Eqn.(2.59)

$$\Sigma(t) = B(t,t_0)\left(\int_{t_0}^{t} B^{-1}(s,t_0)S\ Diag(h(\eta,\theta)\ S^tB^{-1}(s,t_0)^t ds\right)B(t,t_0)^t. \quad (2.62)$$

Using Eqn.(2.60) and lemma 2.1, it can be shown that

$$dB(t,t_0) = J(\eta,\theta)B(t,t_0)dt, \quad dB(t,t_0)^t = B(t,t_0)^tJ(\eta,\theta)^t dt. \quad (2.63)$$

Based on Eqns.(2.59) and Eqn.(2.60), the ODE for $\Sigma(t)$ can be written as

$$\begin{aligned} d\Sigma(t) &= dB(t,t_0)\cdot\left(\int_{t_0}^{t} B^{-1}(s,t_0)S\ Diag(h(\eta,\theta)\ S^t B^{-1}(s,t_0)^t ds\right)\cdot B(t,t_0)^t \\ &+ B(t,t_0)\cdot d\left(\int_{t_0}^{t} B^{-1}(s,t_0)S\ Diag(h(\eta,\theta)\ S^t B^{-1}(s,t_0)^t ds\right)\cdot B(t,t_0)^t \\ &+ B(t,t_0)\cdot\left(\int_{t_0}^{t} B^{-1}(s,t_0)S\ Diag(h(\eta,\theta)\ S^t B^{-1}(s,t_0)^t ds\right)\cdot dB(t,t_0)^t \\ &= J(\eta,\theta)\Sigma(t) + S\ Diag(h(\eta,\theta)\ S^t + \Sigma(t)J(\eta,\theta)^t\ dt. \end{aligned} \tag{2.64}$$

The stochastic linear noise approximation (LNA) of the original discrete Markov jump process, and CLE $X(t) = \eta(t) + M(t)$, and its discrete-time transition probabilities can be obtained by solving three ordinary differential equations

$$\begin{cases} d\eta(t) = Sh(\eta(t),\theta)dt \\ dm(t) = J(\eta,\theta)m(t)dt \\ d\Sigma(t) = J(\eta,\theta)\Sigma(t) + S\ Diag(h(\eta,\theta))\ S^t + \Sigma(t)J(\eta,\theta)^t\ dt. \end{cases} \tag{2.65}$$

Instead of only inferring the initial conditions of the sample path as in the ODE and SDE inference, the LNA forward simulation process also treats intermediate sampling states as latent parameters to infer. A key parameter in the LNA inference is the time step Δt between intermediate observation states. If Δt is chosen too small, then too many intermediate states need to be inferred which leads to inefficient Bayesian computation. If Δt is chosen too large, then there will be too many cumulative numerical errors for the LNA approximation of the CLE, which leads to inaccurate results. We select the LNA time step Δt by first running the MAP optimization on the Bayesian model and then using those MAP estimated parameters to generate forward LNA solutions with different Δt and compare them with Gillespie and SDE solutions. Here, we select Δt to be 1 day for all the LNA simulations in this thesis.

Putting everything together, the LNA model for the Logistic differential equation can be

written as

$$\begin{cases} d\eta(t) = kP(1 - \frac{P}{K})dt \\ dm(t) = k(1 - 2\frac{P}{K})m(t)dt \\ d\Sigma(t) = k(1 - 2\frac{P}{K})\Sigma(t) + kP(1 + \frac{P}{K}) + \Sigma(t)k(1 - 2\frac{P}{K})^t\, dt. \end{cases} \tag{2.66}$$

Non-Restarting and Restarting Version LNA

Let the observation time point sequence be $\{t_i\}$ with increment Δt. There are two types of LNA initial conditions: the non-restarting version [39], which follows the update rule Algorithm 1 where N is the total number of observation times. Here, D_{obs} is the list of the days when the data is observed, and δ_t is the technical measurement error term, estimated errors arising from measuring of the data. Further, η_{t_0}, m_{t_0} Σ_{t_0} are the initial conditions of the deterministic mean process, the Gaussian mean, and the Gaussian variance processes, respectively. And, $\boldsymbol{X}_0$ is the latent initial condition cell counts, and y is the observed data.

Algorithm 1 Non-restarting version LNA algorithm

Require: $\eta_{t_0} \leftarrow X(0)$, $m_{t_0} \leftarrow 0$, $\Sigma_{t_0} \leftarrow 0_{d\times d}$, X_0, N, D_{obs}, y, Δt, δ_t
  for i in 1:N **do**
    $\eta_{t_i} \leftarrow d\eta(\eta_{t_{i-1}}, \theta, \Delta t)$
    $m_{t_i} \leftarrow dm(X_{t_{i-1}} - \eta_{t_{i-1}}, \theta, \Delta t)$
    $\Sigma_{t_i} \leftarrow d\Sigma(0_{d\times d}, \theta, \Delta t)$
    $X_i | X_{i-1}, \theta \sim N(\eta_{t_i} + m_{t_i}(X_{t_{i-1}}, \eta_{t_{i-1}}), \Sigma_{t_i})$ ▷ Since dm is a homogeneous linear ODE
    if i in D_{obs} **then**
      $y_i \sim N(X_i, \delta_t)$
    end if
  end for

The restarting version of LNA [21] aims to address the numerical errors incurred approximating the transition density of the CLE by a series of ODEs over long time periods which consequently leads to poor approximations to states X_t for large t. In [21], it was shown that by replacing the initial conditions of the deterministic part $\eta_{t_{i-1}}$ by the posterior mean of X_t given data y and the restarting LNA for each time interval, the approximation errors can be avoided for larger Δt. In addition, it was found that the restarting version of LNA is

more accurate for partially observed systems. The restarting version of the LNA algorithm is given in the algorithm 2 below.

Algorithm 2 Restarting version LNA alogrithm

Require: $\eta_{t_0} \leftarrow X(0)$, $m_{t_0} \leftarrow 0$, $\Sigma_{t_0} \leftarrow 0_{d\times d}$, X_0, N, D_{obs}, y, Δt, δ_t
  for i in 1:N **do**
    $\eta_{t_i} \leftarrow d\eta(X_{t_{i-1}}, \theta, \Delta t)$
    $m_{t_i} \leftarrow dm(0_d, \theta, \Delta t)$
    $\Sigma_{t_i} \leftarrow d\Sigma(0_{d\times d}, \theta, \Delta t)$
    $X_i|X_{i-1}, \theta \sim N(\eta_{t_i}(X_{t_{i-1}}), \Sigma_{t_i})$ ▷ Since $m_{t_i} \equiv 0$ and η_t is nonlinear ODE of $X_{t_{i-1}}$
    if i in D_{obs} **then**
      $y_i \sim N(X_i, \delta_t)$
    end if
  end for

By evaluating two algorithms on partially observed datasets by using multiple combinations of parameters, we also found that restarting the version of LNA achieves a better approximation of the Markov jump process and the CLE solution. Thus, we decided to sue the restarting version of the LNA for inference throughout this thesis.

2.2.4 Comparing LNA Inference with ODE and SDE Inference

In general, a Markov jump process that can be simulated by the Gillespie algorithm is the ideal model. We usually consider a fast approach for connecting the chemical master equation and time evolution sample paths by using the SDE to approximate the Markov jump process when the system size is large enough. The SDE, derived by a diffusion approximation, involves a Gaussian approximation to the independent Poisson process (see section 2.2.2). When the system size v is large enough, the SDE solution can be approximated by its deterministic part which refers to the ODE solution (see section 2.2.1). However, SDE inference does not have a tractable likelihood which means that it won't lead to a tractable distribution of $\boldsymbol{X}_t$ based on given the initial conditions $\boldsymbol{X_0}$. ODE inference only describes the mean process of the Markov jump process. LNA, which is derived from Ito-Taylor

Bayesian Inference with Dynamical Systems	
Differential Equation	❋ refers to limitations ✔ refers to advantages
ODE	❋ biological processes are stochastic in nature, therefore ignoring stochasticity in the biological process can lead to inaccurate inference ✔ not computationally expensive, can be applied to high dimensional models
LNA	❋ can be viewed as the first order expansion of Ito-Taylor expansion of SDE which only consider the lowest order fluctuations about the deterministic mean ✔ accurate for any chemical reaction network (CRN) if the master equation contains only first-order reactions (can accurately approximate a linear SDE system) ✔ can approximate Markov jump processes with higher order reactions if the population size is large enough ✔ the restarting version of LNA can avoid the error accumulation during the calculation
SDE	❋ transition density $K(\boldsymbol{\theta}, \boldsymbol{\theta'})$ is generally unknown and thus won't lead to a tractable distribution of $\boldsymbol{X_t}$ given initial condition $\boldsymbol{X_0}$ ❋ likelihood is intractable ✔ account for stochasticity in the time evolution process and can use Approximate Bayesian Computation (ABC) to accelerate the estimation of the posterior distributions

Table 2.1: Limitations and advantages of using different dynamic models in the context of Bayesian inference. Ordinary differential equations (ODEs) only describe the mean field process whereas stochastic differential equations (SDEs) can approximate the Markov jump process when the system size is large enough. The Linear Noise Approximation (LNA) with tractable likelihood can be derived from the SDE through the Ito-Taylor Expansion.

expansion of SDE (see section 2.2.3), can be regarded as a compromise between SDE and ODE inference. LNA is a stochastic process and has tractable likelihood by a decomposing SDE into a deterministic mean process, a Gaussian mean process, and a Gaussian variance process. A comparison of the advantages and limitations of ODE, LNA, and SDE inference can be seen in Table 2.1.

2.3 Bayesian Model Evaluation and Selection

2.3.1 Bayes Factor

To select the most plausible model, we employ Bayesian model selection that relies on the Bayes Factor approach [35]. Let D be the dataset, M_k be the kth candidate model and θ be the parameters. Without loss of generality, the marginal likelihood of the data under the model can be written,

$$p(D|M_k) = \int p(D|\theta, M_k)p(\theta)d\theta \tag{2.67}$$

, which naturally includes a penalty for models that include too many variables since the marginal likelihood integrates through all the parameters. Then by Bayes rule, the posterior probability of the kth model M_k over the observed data can be written as,

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_k p(D|M_k)p(M_k)}, \text{ k=1,2,...} \tag{2.68}$$

In our case we only compare two models at a time, say M_0 and M_1, then the relative marginal likelihood of model M_0 to model M_1 can be written as:

$$\frac{p(M_0|D)}{p(M_1|D)} = \frac{p(D|M_0)}{p(D|M_1)}\frac{p(M_0)}{p(M_1)}. \tag{2.69}$$

Then, the Bayes factor of model M_1, in favor of model M_0, is defined as

$$B_{10} = \frac{p(D|M_1)}{p(D|M_0)} = \frac{p(M_0)}{p(M_1)}\frac{p(M_1|D)}{p(M_0|D)}. \tag{2.70}$$

If the log scale Bayes factor $log_{10}(B_{10})$ is larger than 1/2, then there is substantial evidence that data are generated by model M_1. If $log_{10}(B_{10})$ is larger than 2, then there is decisive evidence that data are generated by model M_1. Details can be found in Table 2.2.

Bayes Factor Interpretation [35]		
$log_{10}(B_{10})$	B_{10}	Evidence against M_0
0 to 1/2	1 to 3.2	Not worth than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

Table 2.2: Bayes factor interpretation table. B_{10}: Bayes factor of model M_1 in favor of model M_0. The table shows the $\log_{B_{10}}$ and B_{10} value ranges when evidence of M_1 against M_0 is not worth than a bare mention, substantial, strong, and decisive respectively.

2.3.2 PSIS-LOO-CV

Pareto Smoothed Importance Sampling (PSIS) Leave-one-out (LOO) cross-validation (CV) is a popular model evaluation metric in both machine learning and probabilistic learning. The predictive density evaluated at data $\boldsymbol{y}_i$ can be approximated by importance sampling, which can adjust Monte Carlo integration to account for the fact that sampling draws could come from the wrong distribution [2, 73]. Let $\boldsymbol{\theta^s}$ be a sampling from the posterior $p(\boldsymbol{\theta}|\boldsymbol{D})$, then the leave-one-out predictive density given the data without i^{th} data can be written as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{D}_{-i}) = \int p(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\theta}) p(\boldsymbol{X}_i, \boldsymbol{\theta}|\boldsymbol{D}_{-i}) d\boldsymbol{\theta} \approx \frac{\sum_{s=1}^{S} p(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\theta^s}) w_i^s}{\sum_{x=1}^{s} w_i^s}, \tag{2.71}$$

where w_i^s are the importance weights defined as

$$w_i^s = \frac{p(\boldsymbol{\theta^s}|\boldsymbol{X}_i, \boldsymbol{D}_{-i})}{p(\boldsymbol{\theta^s}|\boldsymbol{y}_{-i})} \propto \frac{1}{p(\boldsymbol{D_i}|\boldsymbol{\theta^s})}. \tag{2.72}$$

Note that the leave-one-out cross-validation likelihood also naturally includes a penalty for the number of parameters (by integrating over all parameters). However, in many Bayesian estimation situations, the importance ratios may have heavy right tails, which leads to large or infinite variances of the importance weights from the approximated distributions. PSIS is a self-normalized importance sampling estimator by introducing a shape parameter k. This shape parameter aims to smooth the weight distribution and stabilize importance sampling

estimates to ensure finite variances. This consequently results in a more robust and better evaluation of fitted Bayesian models [73]. The expected log point-wise predictive density (elpd) LOO can the calculated by

$$elpd_{LOO} = \frac{1}{n}\sum_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{D}_{-i}, w_i^s(k_i)) \tag{2.73}$$

By using PSIS-LOO-CV, models are selected based on cv_elpd$_{LOO}$, which measures a score related to the out-of-sample prediction accuracy, and naive_elpd$_{LOO}$, which measures a score related to the in-sample prediction accuracy. When comparing two fitted models, we can estimate the difference in their expected predictive accuracy by the difference in cv_elpd$_{LOO}$. Thus, the higher the cv_elpd$_{LOO}$, the more plausible the data is generated by the model. Besides, the models can also be ranked by a set of Akaike-like weights assigned to each model, which can be used in pseudo-Bayesian model averaging based on cv_elpd$_{LOO}$,

$$loo_{weight} = \frac{\exp([\text{cv\_elpd}^1_{LOO}]), \cdots \text{cv\_elpd}^K_{LOO}])}{\sum_{k=1}^{K} \exp(\text{cv\_elpd}^k_{LOO}])} \tag{2.74}$$

where K is the total number of models and cv_elpd$^k_{LOO}$ refer to the cv_elpd$_{LOO}$ of model k.

Chapter 3

Hierarchical Bayesian ODE Modeling and Model Selections for Hematopoiesis Dynamics

3.1 Introduction

To understand the hematopoiesis dynamics feedback regulation mechanism through mathematical and statistical models, we also need to learn the parameters governing these models. One approach to such parameter estimation requries observing how cell counts re-equilibrate after the perturbation from their steady states. Thus, we analyze data from experiments in mice where the hematopoietic system is temporarily perturbed by a single administration of a chemotherapy agent (5-FU). The system re-equilibrates after about 30 days after the 5-FU treatment. A confounding factor is that there are batch effects across the experimental cohorts due to the practiced necessity of using different filters during flow cytometry measurements. Thus, we also want to account for these batch effects during inference. One main

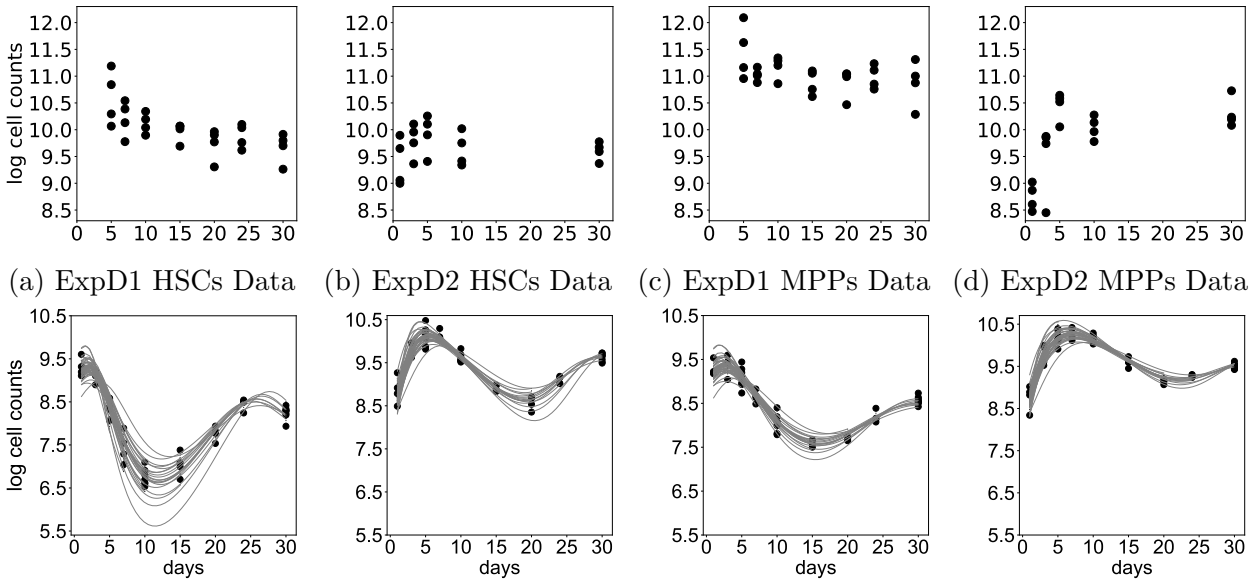
obstacle is that the data for the hematopoiesis process can only be partially observed. More specifically, the hematopoietic cell compartment population counts can not be collected in a longitudinal way because the mice used in the experiment are sacrificed to perform measurements. Each mouse can contribute its cell count data only at one certain time point. Following [42], we use a Bayesian latent variable approach where the unobserved data are treated as latent variables to be inferred and an ODE model is used to interpolate the cell counts until the observations times. Unlike [42], where feedback regulation was modeled using nonlinear Hill functions, here we use a chemical reaction framework that incorporates feedback by assuming hematopoietic cell compartments produce chemical factors that interact with the cell compartments. One advantage of this approach is that our new chemical reaction model can be easily extended to a stochastic model with a tractable likelihood for Bayesian inference using a linear noise approximation [39] – a topic explored in Chapter 4.

We investigate the underlying dynamics and estimate the parameters in the chemical reaction model using a new hierarchical Bayesian inference framework that accounts for heterogeneity in parameter values across individuals and corrects for batch effects through an offset parameter that is also inferred. This enables us to integrate all the experimental data to perform the inference. We validate our approach using synthetic data. Applying this framework to the experimental data and using Bayesian model selection criteria, we find that the experimental data is best described by a hierarchical model, indicating that mouse-to-mouse variations in parameters are a significant source of variability in the data. By accounting for both biological and technical variability (e.g., due to the measurement/model errors), we discover that the system responds to the perturbation via an overshoot followed by damped oscillations, which are characteristic of negative feedback systems and consistent with the equilibrium being a stable spiral.

3.2 Methods

This section provides a comprehensive overview of the experimental data, data generation process, and perturbation methods as well as the proposed new chemical reaction model and the process of simplifying it. We examine the dynamics that arise from interactions between various hematopoietic cells after the perturbation treatment and then use data from perturbation experiments to calibrate and refine our ODE model that accounts for cell-to-cell interactions. We also describe the hierarchical Bayesian methods to infer our models' parameters and perform model selection based on Bayes factors and a cross-validation criteria. Two experiments were conducted on young female mice to perturb the number of hematopoietic cells from their homeostatic values. In both experiments, mice were injected a single time with 5-FU, which is a drug commonly used for cancer chemotherapy. 5-FU acts by inhibiting thymidylate synthase, thereby blocking DNA synthesis and repair and triggering the death of dividing cells. In the first experiment, data is collected 5, 7, 10, 15, 20, and 24 days after 5-FU is administered, and at each time point, the bone marrow (BM) of 4 mice is analyzed. The data is shown in Figure 3.1 (a-d). There are also 4 control mice, whose cell counts are measured before treatment and placed at day 30 since the system returns to homeostasis by about three weeks after 5-FU treatment. Figure 3.1 (a,c) shows the numbers of HSC and multipotent progenitors (MPP) in the bone marrow measured by flow cytometry (see Appendix A.0.1). Although there is a significant amount of variation, the data suggest that the homeostatic values (day 30) for the HSC counts Figure 3.1 (a) are smaller than those recorded at an early time (e.g., day 5), which implies an overshoot consistent with the release of negative feedback regulation [10], likely due to cell death in response to 5-FU treatment. Similar results may also be observed for MPPs in Figure 3.1 (c), but the large variation makes it difficult to draw definitive conclusions. These results motivated a second experiment to be performed that focused on the dynamics at early times 1, 3, 5, and 10 days after 5-FU is administered (see Figure 3.1 (b,d)), referred to as the experimental dataset 2.

As in the first experiment, there were 4 control mice that were placed on day 30. However, in the second experiment, due to the supply problems, a different filter type was used in the flow cytometry measurements (see Appendix A.0.1) for details, which made the absolute cell counts not directly comparable across the two experiments due to this batch effect (note different scales in Figure 3.1 (a,c) versus (b,d)). The results suggest that the HSC and MPP cell numbers first decrease after the 5-FU administration, then increase, possibly overshooting homeostasis, and then settle down to the homeostatic values, although there is still significant variation in the data. As described below, in order to integrate the datasets from both experiments and account for the batch effect, we introduce and infer the value of an offset parameter in the mathematical model that accounts for the different cell count scales in the two experiments.



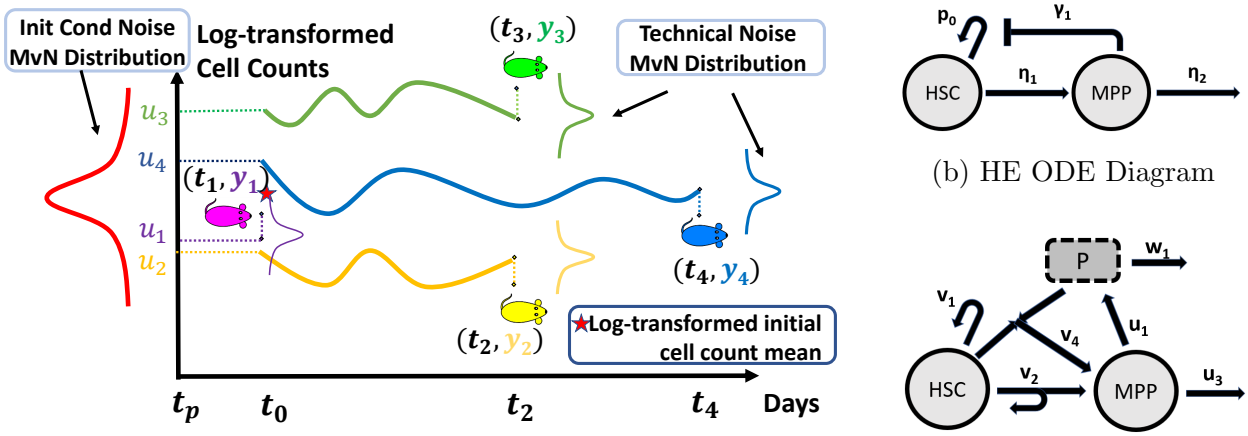
(a) ExpD1 HSCs Data (b) ExpD2 HSCs Data (c) ExpD1 MPPs Data (d) ExpD2 MPPs Data

(e) CRSD HSCs Data (f) CRSD MPPs Data (g) HESD HSCs Data (h) HESD MPPs Data

Figure 3.1: Experimental dataset (top row) and synthetic dataset (bottom row): ExpD: experimental dataset; CRSD: synthetic dataset generated by our chemical reaction model, HESD: synthetic dataset generated by the Hill equation model; points: cell count data; curves: latent ODE trajectories for synthetic data generation process; (a,b) log-scaled cell counts of HSCs in the experimental datasets 1 and 2 (ExpD1 and ExpD2) (c,d) MPPs in the ExpD1 and ExpD2 (e,f) HSCs and MPPs in the CRSD (g,h) HSCs and MPPs in the HESD

3.2.1 Latent Variable Modeling Approach

In order to count the number of cells in the BM, the mouse must be sacrificed. Therefore, each mouse can contribute their count data at a one-time point only. We use log-transformed cell counts of HSCs and MPPs as the data. In our data generation model, prior to the observation time, we assume that each mouse cell counts start at unobserved initial conditions at the time of the perturbation experiment. We then use ODEs to model the expected values of the distribution of the observed cell counts. The variance of this distribution accounts for the experimental measurement noise. Since the experimental cell count data shown in Figure 3.1 (a-d) indicates that the hematopoietic cell trajectories exhibit a high range of mouse-to-mouse variability, we also account for biological variation both in the initial condition (which is not observed) and in the ODE model parameters to account for potential sources of heterogeneity. Figure 3.1 (e-h) show the synthetic data generated by the chemical reaction and Hill equation models using the Bayesian latent variables approach that generates the synthetic data as illustrated in Figure 3.2 (a), where the points indicate the measurements and the curves are the ODE trajectories. Let the day of the perturbation experiment be t_p. The log-transformed initial cell count means on the first observation day t_0 is shown as a log-scale initial condition means (red star) for all the mice. The unobserved cell count initial conditions for mouse i, u_i where $i \in \{1, 2, 3, 4\}$, are sampled from a multivariate normal distribution with mean equal to the red star and a covariance matrix to account for the biological varibility on day t_0. Both the mean and covariance matrix are learned from the data. For each mouse i, y_i denotes the observed values of the cell counts for each mouse when it was sacrificed. We assume each y_i is sampled from the multivariate normal distribution with the ODE solution at that time being the mean and covariance matrice that account for the measurement error. We use two ODE models a nonlinear Hill equation (HE) model (Figure 3.2 (b)) and a new chemical reaction (CR) model (Figure 3.2 (c)); the HE model was previously used in [42].



(a) Data Generation Diagram Illustration

(b) HE ODE Diagram

(c) CR ODE Diagram

Figure 3.2: Data generation process and mechanistic ODE models (a) data generation process which accounts for both biological and technical variations via a latent variable approach: t_p is the 5-FU treatment time; mice are being observed at discrete irregular times e.g. t_0, t_2, t_4; t_0: first observation day; for $i \in$ 1,2,3,4, u_i refers to the latent initial cell count of mouse i at t_0 and y_i refers to the observed cell count of mouse i at the time t_i when it was sacrificed. The curves denote ODE trajectories. (b) two-compartment Hill equation (HE) model for hematopoiesis [42]; where cell count of MPPs negatively feedback on the self-renewal rate of HSCs (c) two-compartment chemical reaction (CR) ODE model for hematopoiesis; chemical factor P is assumed to be produced by MPPs and accelerates HSCs direct differentiation rate.

3.2.2 Mathematical Models

Our data suggest that the most significant change in the hematopoietic system induced by 5-FU is in the stem and early progenitor compartments which is consistent with the experimental studies [46]. A number of experimental studies report that the dynamics of mature hematopoietic cells are independent of hematopoietic stem cells (HSCs) in hematopoiesis for at least half a year after HSCs are pertubed [71, 65, 66, 53, 60]. Thus, we model the dynamics of the HSCs (combining both long-term and short-term) and multipotent progenitor (MPPs) cells, which are the next cells downstream in the lineage. Models that assume feedback on HSCs behavior from other cell types will be considered in future work.

Hill Equation Model

Previously, a nonlinear Hill equation (HE) ODE model has been proposed for the of HSCs and MPPs hematopoiesis dynamics [42] with feedback regulations [16]. The model involves a negative feedback regulation on HSCs self-renewal rate and a feedforward regulation on MPPs decay rate. Here, as shown in Figure 3.2 (b), we simplify the model and only consider negative feedback regulation from MPPs on the HSC self-renewal. The HE model is described by the following ODEs:

$$\begin{cases} N'_{HSC} = \left(\dfrac{2p_0^*}{1+\gamma_1 \cdot N_{MPP}} - 1\right) \cdot \eta_1 \cdot N_{HSC}, \\ N'_{MPP} = 2(1 - \dfrac{p_0^*}{1+\gamma_1 \cdot N_{MPP}}) \cdot \eta_1 \cdot N_{HSC} - \eta_2 \cdot N_{MPP}, \end{cases} \tag{3.1}$$

where p_0^* is the maximum self-renewal probability of HSCs, γ_1 is the feedback gain from MPP cell count on the HSC self-renewal probability, η_1 is HSC division rate, η_2 is the MPPs decay rate.

Chemical Reaction Equation Model

We model the feedback regulation on the HSC self-renewal rate by introducing a chemical factor P that is assumed to be produced by MPPs at a constant rate that interacts with HSCs and induces the HSCs to differentiate to MPPs Figure 3.2 (c). We assume that the MPPs secrete the factor P at a rate u_1 and P decays at a rate w_1. Following [9, 4], we assume that the HSC division rates are constants. We also lump the rates for MPP self-renewal, differentiation, and decay into a parameter u_3, which is also assumed to be constant, we assume the HSC direct differentiation rate to MPP is time-dependent, based on the factor

P. This gives the CR ODE model:

$$\begin{cases} N'_{HSC} = v_1 \cdot N_{HSC} - v_4 \cdot N_P \cdot N_{HSC}, \\ N'_{MPP} = v_2 \cdot N_{HSC} + v_4 \cdot N_P \cdot N_{HSC} - u_3 \cdot N_{MPP}, \\ N'_P = u_1 \cdot N_{MPP} - w_1 \cdot N_P, \end{cases} \tag{3.2}$$

where v_1 is the HSC net proliferation rate which equals the self-renewal rate minus the death rates of HSCs, v_2 is the HSCs asymmetric division rate and v_4 is the HSC direct differentiation parameter. Since the chemical factor typically evolves much faster than the cell self-renewal, differentiation and death process, we further assume that the factor P is in a steady state. As a result, N_p can be regarded as proportional to the number of MPPs. Without loss of generality, let

$$N_p = \frac{u_1}{w_1} N_{MPP}. \tag{3.3}$$

Using Eqn.(3.2.2) and Eqn.(3.2.2) and defining $r_1 = v_4 \cdot u_1/w_1$, then the chemical reaction ODE system can be written as

$$\begin{cases} N'_{HSC} = (v_1 - r_1 \cdot N_{MPP}) \cdot N_{HSC}, \\ N'_{MPP} = (v_2 + r_1 \cdot N_{MPP}) \cdot N_{HSC} - u_3 \cdot N_{MPP}, \end{cases} \tag{3.4}$$

where $r_1 \cdot N_{MPP}$ is the effective direct differentiation rate from HSCs into MPPs. To explore the model behavior numerically, we match the HSC net gain rate, the MPP net gain, and the MPP decay rate between our new chemical reaction ODE model and the previous Hill equation ODE model. We get the following relationship:

$$\begin{cases} v_1 + v_2 = \eta_1, \\ u_3 = \eta_2, \\ v_1 - r_1 \cdot N_{MPP} = (\dfrac{2p_0^*}{1+\gamma_1 \cdot N_{MPP}} - 1) \cdot \eta_1, \\ v_1 + r_1 \cdot N_{MPP} = 2(1 - \dfrac{p_0^*}{1+\gamma_1 \cdot N_{MPP}}) \cdot \eta_1. \end{cases} \tag{3.5}$$

To compare the two models fairly, we need to set the priors for the parameter combinations shown above to be approximated by the same distribution. The plots of four synchronized net priors plots can be seen in Appendix Figure A.1. The 95% confidence intervals for two prior sets (non-hierarchical and hierarchical) can be seen at tables A.1 and table A.2. Let $p_0 = p_0^*/(1+\gamma_1 \cdot N_{MPP}) - 1$. Then, adding the last two equations in the system 3.2.2, we obtain,

$$\begin{cases} 2p_0 - 1 = v1/\eta_1 - r_1/\eta_1 \\ 2(1-p_0) = v1/\eta_1 + r_1/\eta_1 \end{cases} \tag{3.6}$$

from which it follows that

$$p_0 = 1/2 \cdot (1 + v_1/\eta_1 + r_1/\eta_1).$$

Observe that the HSC self-renewal probability $p_0 > 0.5$ as expected for stem cell dynamics.

3.2.3 Scalable Bayesian Hierarchical Framework

To study the regulatory mechanism occurring in an individual mouse, latent variable approaches can be employed to track multiple cell compartment counts over time as described in the data generation process section 3.2.1. We model the experimental measurement error explicitly by including a technical noise parameter. In order to take into account the wide discrepancies between the raw counts of different cell compartments, we decide to transform all the cell count numbers into a logarithmic scale. We assume that the log-transformed HSCs and MPPs counts for all the mice at initial observation day t_0 after the perturbation

experiment follow multivariate normal distributions with means equal to μ_{HSC} and μ_{MPP} respectively and the same standard deviation (std) δ_b. This model variance among the initial cell counts is due to biological differences between the mice. Let $\boldsymbol{\mu}$ be the bivariate vector for the μ_{HSC} and μ_{MPP} and $\boldsymbol{u_i}$ be the log-scale cell count initial conditions for mouse number i. Suppose that the total number of mice is N. Then, we can write the initial conditions for each mouse as

$$\boldsymbol{u_i} \sim MvN(\boldsymbol{\mu}, \delta_b^2 \boldsymbol{I}), \tag{3.7}$$

for all $i \in \{1, 2, \cdots, N\}$

Without loss of generality, we assume that the first m mice are sacrificed and observed during the first observation day of the experiment right after the 5-FU administration. Let $\boldsymbol{y_i}(t)$ be the log-scale observed cell counts data for mouse i at time t and assume the observed data is obtained with technical measurement errors δ_t. Then for the mice that are sacrificed to the day after 5-FU is applied as

$$\boldsymbol{y_i}(t_0) \sim MvN(\boldsymbol{u_i}, \delta_t^2 \boldsymbol{I}), \tag{3.8}$$

for $\forall\ i \in \{1, 2, \cdots, m\}$.

For those mice whose cell counts are measured after t_0, assume that mouse j is sacrificed at day t_j. We assume that the latent trajectories for the cell populations follow an ODE model, denoted as $\boldsymbol{\lambda}(\boldsymbol{u}, \boldsymbol{\theta}, t)$. This allows us to simulate all the cell counts before the mouse is harvested for measurement. Then, the log-scale cell accounts for mouse j at time t_j can be represented as $\boldsymbol{\lambda}(\boldsymbol{u_j}, \boldsymbol{\theta}, t_j)$ and the conditional distribution for these mice can be written as

$$\boldsymbol{y_j}(t_j) \sim MvN(\boldsymbol{\lambda}(\boldsymbol{u_j}, \boldsymbol{\theta}, t_j), \delta_t^2 \boldsymbol{I}), \tag{3.9}$$

for all $j \in \{m+1, m+2, \cdots, N\}$, where $\boldsymbol{\theta} = \{v_1, v_2, u_3, r_1\}$ is the vector of rate parameter values and t_j is the total timespan of the ODE system. Following our data-generating procedure, we also assume that observations are independent of each other, then the likelihood

function is taken to be the product of the multivariate normal densities:

$$p(\boldsymbol{y}|\boldsymbol{\Theta_{NH}}) = \prod_{i=1}^{m} p(\boldsymbol{y_i}|\boldsymbol{u_i}, \delta_t^2 \boldsymbol{I}) \cdot \prod_{j=m+1}^{N} p(\boldsymbol{y_i}|\boldsymbol{\lambda}(\boldsymbol{u_j}, \boldsymbol{\theta}, t_j), \delta_t^2 \boldsymbol{I}), \tag{3.10}$$

where $\boldsymbol{\Theta_{NH}} = (\boldsymbol{\theta}, \boldsymbol{u_{1:N}}, \delta_b, \delta_t)$ refers to the rate parameters for the ODE model, the generated log-scale initial cell counts as well as the biological and technical standard deviation terms, respectively. Then by the Bayes rule, the posterior distribution of $\boldsymbol{\Theta_{NH}}$ conditional on the log-transformed cell count data can be represented as:

$$p(\boldsymbol{\Theta_{NH}}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\Theta_{NH}}) p(\boldsymbol{\Theta_{NH}}), \tag{3.11}$$

where $p(\boldsymbol{\Theta_{NH}}) = p(\boldsymbol{\theta}) \cdot \prod_{i=1}^{N} p(\boldsymbol{u_i}|\boldsymbol{\mu}) \cdot p(\boldsymbol{\mu}) \cdot p(\delta_b) \cdot p(\delta_t)$. We assume that each parameter $\theta^* \in \boldsymbol{\Theta_{NH}}$, θ^* has a Lognormal distribution and consequently $\theta^* \in (0, \infty)$.

3.2.4 Bayesian Cross-Dataset Hierarchical Framework

Traditionally, investigations in hematopoiesis focus on species-level ensemble averages of cell counts and rate parameter values [47, 9]. This approach can smooth out and underestimate the noise from outliers. However, cell populations are generally heterogeneous, and all cells do not respond identically to perturbations. These motivate us to study the case in which the parameters vary among the mice. Thus, we enhance the framework to rationalize mice population variability within the same dataset and extend the framework to integrate data from various experimental datasets. This requires statistical inference for unknown model parameters based on multi-individual data, which accounts for heterogeneity caused by both intrinsic variance (e.g., variations in initial cell counts and rate values) and extrinsic variance (e.g. variation during the wet lab experiments that results in batch effects). Unlike the previous framework, here we capture population-level heterogeneity by letting the rate values

for each mouse differ from the others but are drawn from the same prior distributions:

$$\theta_k^i \sim N(\mu_{\theta_k}, \sigma_{\theta_k}), \tag{3.12}$$

where θ_k represents one of the rate values in $\theta = \{v_1, v_2.u_3, w_1\}$; μ_{θ_k} and σ_{θ_k} are the prior mean and prior standard deviation of the rate values, respectively, which are assumed to be normally distributed:

$$\mu_{\theta_k} \sim N(\mu_{\mu_{\theta_k}}, \sigma_{\mu_{\theta_k}}), \tag{3.13}$$

$$\sigma_{\theta_k} \sim N(\mu_{\sigma_{\theta_k}}, \sigma_{\sigma_{\theta_k}}). \tag{3.14}$$

Here $\mu_{\mu_{\theta_k}}$ and $\sigma_{\mu_{\theta_k}}$ are the prior mean and prior standard deviation of μ_{θ_k} and $\mu_{\sigma_{\theta_k}}$. $\mu_{\sigma_{\theta_k}}$ and $\sigma_{\sigma_{\theta_k}}$ are the prior mean and prior standard deviation of σ_{θ_k}. We also let the biological variance and technical variance be compartment-specific: $\delta_{b_{HSC}}, \delta_{b_{MPP}}$ and $\delta_{t_{HSC}}, \delta_{t_{MPP}}$. Then the log-scale initial cell counts are distributed as:

$$\boldsymbol{u_i} \sim MvN\left(\boldsymbol{\mu}, \begin{bmatrix} \delta_{b_{HSC}}^2 & 0 \\ 0 & \delta_{b_{MPP}}^2 \end{bmatrix}\right), \tag{3.15}$$

for $i \in 1, \cdots, N$ Any offset in cell numbers, due to differences in measurement techniques among different datasets, could be regarded as another technical error. Let the log-scale dataset offset index be $\boldsymbol{\tau}$ between the two experiments due to using different filters: $\boldsymbol{\tau_i} = \boldsymbol{\tau}$:

$$\mathbb{1}_{\text{i is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y_i}(t_0) \sim MvN\left(\boldsymbol{u_i}, \begin{bmatrix} \delta_{t_{HSC}}^2 & 0 \\ 0 & \delta_{t_{MPP}}^2 \end{bmatrix}\right), \tag{3.16}$$

for all i in $\{1, 2, \cdots, m\}$ for mice observed on t_0, while

$$\mathbb{1}_{\text{i is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y}_j(t_j) \sim MvN\left(\lambda(\boldsymbol{u_j}, \boldsymbol{\theta^j}, t_j), \begin{bmatrix} \delta_{t_{HSC}}^2 & 0, \\ 0 & \delta_{t_{MPP}}^2 \end{bmatrix}\right), \tag{3.17}$$

for all j *in* $\{m+1, m+2, \cdots, N\}$ which were sacrificed after t_0. Further, we assume

$$\tau^* \sim N(\mu_{\tau^*}, \delta_{\tau^*}), \tag{3.18}$$

and $\boldsymbol{\tau} = [\tau^*, \tau^*]$ (e.g., we assume the same offset for HSCs and MPPs) where μ_{τ^*} and δ_{τ^*} refer to the prior mean and prior standard deviation of the dataset offset parameter value. Assuming that observations are independent of each other, then the likelihood function can now be written as

$$p(\boldsymbol{y}|\boldsymbol{\Theta_H}) = \prod_{i=1}^{m} p(-\boldsymbol{\tau_i} + \boldsymbol{y_i}|\boldsymbol{u_i}, \delta_t^2 \boldsymbol{I}) \cdot \prod_{j=m+1}^{N} p(-\boldsymbol{\tau_j} + \boldsymbol{y_j}|\boldsymbol{\lambda}(\boldsymbol{u_j}, \boldsymbol{\theta^j}, t_j), \delta_t^2 \boldsymbol{I}), \tag{3.19}$$

where $\boldsymbol{\Theta_H} = (\boldsymbol{\theta_{1:N}}, \boldsymbol{u_{1:N}}, \delta_{b_{HSC}}, \delta_{b_{MPP}}, \delta_{t_{HSC}}, \delta_{t_{MPP}}, \boldsymbol{\tau_{1:N}})$ refers to the rate parameters, initial cell count conditions, standard deviation terms, and the offset index parameter value. By the Bayes rule, the posterior distribution of $\boldsymbol{\Theta_H}$ conditional on the log-transformed real count data can be represented as:

$$p(\boldsymbol{\Theta_H}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\Theta_H}) p(\boldsymbol{\Theta_H}), \tag{3.20}$$

where

$$\begin{aligned} p(\boldsymbol{\Theta_H}) = \prod_{i=1}^{N} \prod_{k=1}^{4} p(\theta_k^i|\mu_{\theta_k}, \sigma_{\theta_k}) \cdot \prod_{i=1}^{N} p(\boldsymbol{u_i}|\boldsymbol{\mu}) \cdot p(\boldsymbol{\mu}) \cdot p(\delta_{b_{HSC}}) \cdot p(\delta_{b_{MPP}}) \cdot p(\delta_{t_{HSC}}) \cdot p(\delta_{t_{MPP}}) \\ \cdot \prod_{k=1}^{4} p(\mu_{\theta_k}|\mu_{\mu_{\theta_k}}, \sigma_{\mu_{\theta_k}}) \cdot \prod_{k=1}^{4} p(\sigma_{\theta_k}|\mu_{\sigma_{\theta_k}}, \sigma_{\sigma_{\theta_k}}) \end{aligned} \tag{3.21}$$

3.2.5 Parameter Initialization of MCMC by Optimization

For the initialization of the MCMC sampling chain, we choose the Maximum a posteriori (MAP) estimate approach, which estimates the mode of posterior distributions [24] obtained with the limited memory Broyden Fletcher Goldfarb Shanno algorithm (LBFGS) optimizer [41].

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{3.22}$$

3.2.6 ODE Parameter Inference Algorithms for Partially Observed Dataset

The algorithms to simulate from the ODE model for the non-hierarchical and hierarchical frameworks are given in algorithms 3 and 4 below. Both algorithms involve forward simulation and backward parameter estimation. $\boldsymbol{\mu_\Theta}$ and $\boldsymbol{\delta_\Theta}$ are the prior means and standard deviations of our parameters of interest. $\boldsymbol{i}_{obs}$ and $\boldsymbol{j}_{obs}$ are the list of observation days and numbers of mice observed at each observation day. $\boldsymbol{D}_{obs}$ is the observed dataset. $Flag$ is the matrix that tells which dataset the mice belong to. $\boldsymbol{\tau}$ is the offset index parameter on log-scale cell counts between two experiment datasets. $\boldsymbol{\theta}$ are the rate value parameters, and $\boldsymbol{u}_{init}$ are the latent initial cell counts for all the mice. $\delta_{init}, \delta_{tech}$ are the cell count initial condition standard deviation and technical standard deviation, respectively. Since our dataset is relatively small, the hierarchical model can be much more efficient by changing the geometric of the priors using non-centered parameterization see Appendix A.0.2 for detail.

Algorithm 3 ODE forward simulation with non-hierarchical Bayesian framework

Require: $\boldsymbol{\mu_\Theta}$, $\boldsymbol{\delta_\Theta}$, $\boldsymbol{i}_{obs}$, $\boldsymbol{j}_{obs}$, Δt, $\boldsymbol{D}_{obs}, Flag$

$N_i = len(\boldsymbol{i}_{obs})$

$\boldsymbol{Z_\Theta} \sim N(\boldsymbol{0}, 1)$

$\boldsymbol{\Theta} = \boldsymbol{\mu_\Theta} + \boldsymbol{\delta_\Theta} \cdot \boldsymbol{Z_\Theta}$

$\boldsymbol{\theta}, \boldsymbol{u}_{init}, \delta_{init}, \delta_{tech}, \boldsymbol{\tau} = \boldsymbol{\Theta}$

for i in 1:N_i **do**

  for j in 1:$\boldsymbol{j}_{obs}[i]$ **do**

    $\boldsymbol{u}[j, i, 1] \sim MvN(\boldsymbol{u}_{init}, \delta_{init} * \boldsymbol{I})$

    if i==1 **then**

      if Flag[j,i]==1 **then**

        $\boldsymbol{D}_{obs}[j, i] \sim MvN(\boldsymbol{u}[j, i], \delta_{tech} * \boldsymbol{I})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{D}_{obs}[j, i] \sim MvN(\boldsymbol{u}[j, i, 1], \delta_{tech} * \boldsymbol{I})$

      end if

    else

      if Flag[j,i]==1 **then**

        $\boldsymbol{D}_{obs}[j, i] \sim MvN(\lambda(\boldsymbol{u}[j, i], i_{lst}[i], \boldsymbol{\theta}), \delta_{tech} * \boldsymbol{I})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{D}_{obs}[j, i] \sim MvN(\lambda(\boldsymbol{u}[j, i], i_{lst}[i], \boldsymbol{\theta}), \delta_{tech} * \boldsymbol{I})$

      end if

    end if

  end for

end for

Algorithm 4 ODE forward simulation with hierarchical Bayesian framework

Require: $\boldsymbol{\mu_\Theta}$, $\boldsymbol{\delta_\Theta}$, $\boldsymbol{i}_{obs}$, $\boldsymbol{j}_{obs}$, $\boldsymbol{D}_{obs}, Flag$

$N_i = len(\boldsymbol{i}_{obs})$

$\boldsymbol{Z_\Theta} \sim N(\boldsymbol{0}, 1)$

$\boldsymbol{\Theta} = \boldsymbol{\mu_\Theta} + \boldsymbol{\delta_\Theta} \cdot \boldsymbol{Z_\Theta}$

$\boldsymbol{\mu_\theta}, \boldsymbol{\sigma_\theta}, \boldsymbol{u}_{init}, \delta_{init}^{HSC}, \delta_{init}^{MPP}, \delta_{tech}^{HSC}, \delta_{tech}^{MPP}, \boldsymbol{\tau} = \boldsymbol{\Theta}$

for i in 1:N_i **do**

  for j in 1:$\boldsymbol{j}_{obs}[i]$ **do**

    $\boldsymbol{u}[j, i, 1] \sim MvN(\boldsymbol{u}_{init}, Diag(\delta_{init}^{HSC}, \delta_{init}^{MPP}))$

    if i==1 **then**

      if Flag[j,i]==1 **then**

        $\boldsymbol{D}_{obs}[j, i] \sim MvN(\boldsymbol{u}[j, i], \delta_{tech} * \boldsymbol{I})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{D}_{obs}[j, i] \sim MvN(\boldsymbol{u}[j, i, 1], Diag(\delta_{tech}^{HSC}, \delta_{tech}^{MPP}))$

      end if

    else

      $\boldsymbol{\theta}_{curr} \sim N(\boldsymbol{\mu_\theta}, \boldsymbol{\sigma_\theta})$

      if Flag[j,i]==1 **then**

        $\boldsymbol{D}_{obs}[j, i] \sim MvN(\lambda(\boldsymbol{u}[j, i], i_{lst}[i], \boldsymbol{\theta}_{curr}), Diag(\delta_{tech}^{HSC}, \delta_{tech}^{MPP})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{D}_{obs}[j, i] \sim MvN(\lambda(\boldsymbol{u}[j, i], i_{lst}[i], \boldsymbol{\theta}_{curr}), Diag(\delta_{tech}^{HSC}, \delta_{tech}^{MPP}))$

      end if

    end if

  end for

end for

3.3 Results

3.3.1 Bayesian Inference and Model Selection for Synthetic Hematopoiesis Dataset

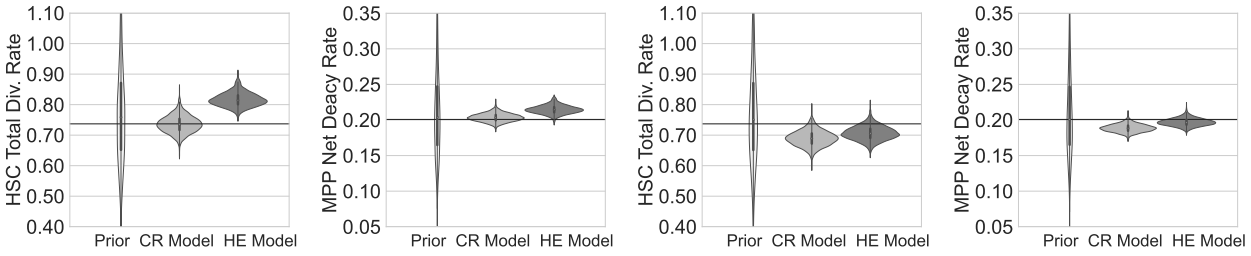
Validation of CR Model and Parameter Inference Framework

In order to test and evaluate our new Chemical Reaction (CR) ODE model to show it can reproduce the dynamic behavior of the non-linear Hill equation (HE) ODE model as well as to demonstrate that the model parameters are identifiable, we first generate synthetic data from the CR model Figures 3.1 (e,f) and fit by both models under non-hierarchical (NH) Bayesian framework (see the prior and posterior plots in appendix Figure A.2). We then generate data from the HE model Figures 3.1 (g,h) and fit the data by both the HE model and the CR model under the NH Bayesian framework (see the prior and posterior plots in appendix Figure A.3). The prior and posterior violin plots for the HSC net division and MPP decay rates are shown in Figure 3.3 (a-d). We observe that both the HE and CR models can recover the corresponding ground truth parameter values (black horizontal lines) used for synthetic data generation. Figure 3.3 (e-h) shows the dynamic uncertainty plots for the CR model synthetic data fit by both CR and HE models. Specifically, the points refer to the synthetic dataset. The shaded areas refer to the 95% credible interval posterior predictive distributions of all the ODE solution trajectories without accounting for the technical errors. The curves refer to the median of the shaded areas. The line plots refer to the 95% credible interval of the posterior predictive plots that account for technical errors. Thus, the technical errors account for differences between the line plot range and the shaded line range on the observation days. As we can see in these posterior predictive distributions, both models can fit the synthetic dataset generated by both models equally well. The dynamic uncertainty plots for the Hill equation model synthetic dataset (HESD)

are shown in Figure 3.3 (e-f) and the chemical reaction synthetic dataset (CRSD) are shown in Figure 3.3 (i-l). Observe that the shaded areas in the dynamic uncertainty plots when fitting the CR Model to the synthetic dataset are generally wider than that of fitting the HE model to the synthetic dataset, and the line plots at the observation times when fitting the CR Model to the synthetic dataset are generally shorter than that of fitting the HE model to the synthetic dataset. This indicates that the CR model is more likely to classify variability in the data as biological variation while the HE model tends to classify the variability as technical measurement error.

Validation of Hierarchical Inference Framework

Next, we will focus on the CR model and validate NH and the hierarchical (H) frameworks as described in section 3.2.3. In Figures 3.4 (a-d), the violin plots for the prior and posterior distributions for the four rate value parameters; the HSC self-renewal rate v_1, the HSC asymmetric division rate v_2, the MPP decay rate u_3, and the HSC effective differentiation rate r_1 using a dataset generated by the NHCR model (NHCRSD). From these figures, we can see that inference using both the hierarchical and non-hierarchical CR models are able to recover the ground truth parameters although the NH CR model is more certain of the parameters. Figure 3.4 (e-h) shows the violin plots with a dataset generated from the HCR model. Inference using both the hierarchical and non-hierarchical models is also able to recover the ground truth values reasonably. The dynamic uncertainty plots for all the synthetic datasets can be seen in Figure 3.4(i-p), we compare the fits using the NHCR and hierarchical chemical reaction (HCR) models. We see that the CR model under both the H/NH frameworks can fit the NHSD and HSD well. The HCR model classifies noise in the data as arising more from biological variance, while the NHCR model classifies the data variability as being due to technical errors. The prior and posterior parameter distribution violin plots for the NHCR model fit NHSD is presented in Appendix Figure A.4. The

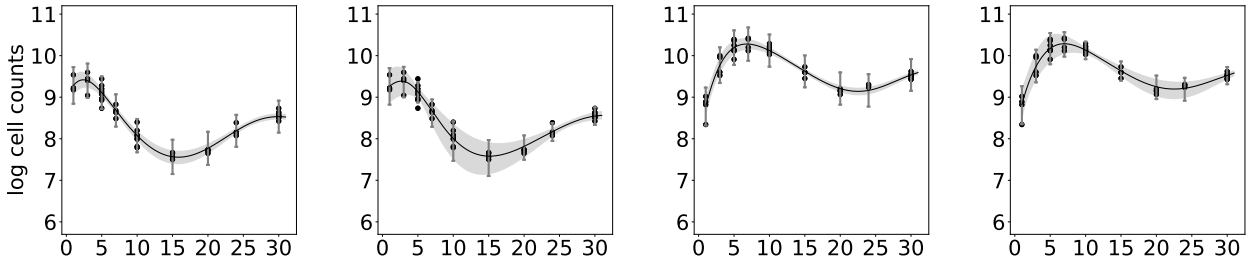


(a) HSC Total Division Rate when fitting models to CRSD

(b) MPP Decay Rate when fitting models to CRSD

(c) HSC Total Division Rate when fitting models to HESD

(d) MPP Decay Rate when fitting models to HESD

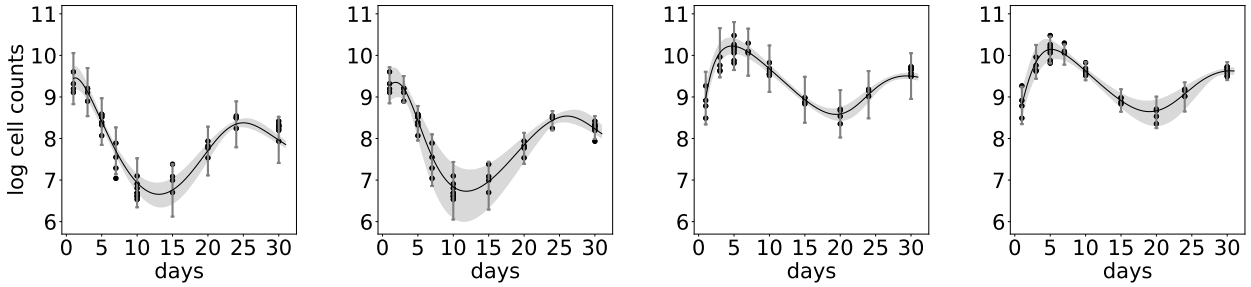


(e) fit HE Model to HESD HSCs Count

(f) fit CR Model to HESD HSCs Count

(g) fit HE Model to HESD MPPs Count

(h) fit CR Model to HESD MPPs Count



(i) fit HE Model to CRSD HSCs Count

(j) fit CR Model to CRSD HSCs Count

(k) fit HE Model to CRSD MPPs Count

(l) fit CR Model to CRSD MPPs Count

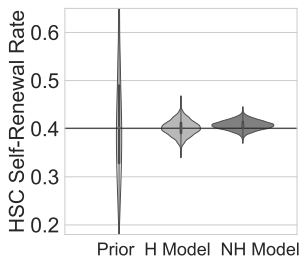
Figure 3.3: Prior and posterior violin plots, dynamic uncertainty plots for fitting both chemical reaction and Hill equation models to the synthetic dataset generated by both models. Fitting both the chemical reaction synthetic dataset (CRSD) and the Hill equation synthetic dataset (HESD) by chemical reaction (CR) and hill equation (HE) models; (a-d) violin plots for the prior and posterior distributions for the net HSC total division rate and the MPP decay rate fitting the CR and HE models to the (a,b) CRSD; (c,d) HESD; (e-l) the dynamic uncertainty plots where the shaded areas refer to the 95% credible interval posterior predictive distributions of all the ODE solution trajectories without accounting for the technical errors. The curves refer to the median of the shaded areas. The line plots refer to the 95% credible interval of the posterior predictive intervals that account for technical errors at the data observation time. (e,f) log scale HESD HSCs data fitted by both ODE models; (g,h) log scale HESE MPP data fitted by both ODE models; (i,j) log scale CRSD HSC data fitted by both ODE models; (k,l) log scale CRSD MPP data fitted by both ODE models.

corresponding plots for the other parameters using various combinations of the datasets can be found in appendix Figures A.5, A.6, A.7 Lastly, as shown in table 3.1, the Bayes factor Eqn.(2.70) decisively selects the correct model for each dataset based on table 3.1 and criteria table 2.2.

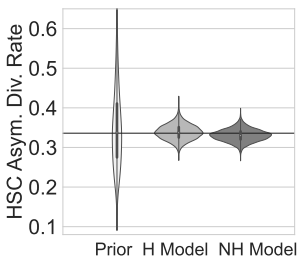
Figure 3.4: Prior and posterior violin plots, dynamic uncertainty plots for fitting non-hierarchical/hierarchical chemical reaction (NHCR/HCR) models to the synthetic data generated by both hierarchical model (H)/non-hierarchical (NH) frameworks (HSD/NHSD). Page(54-55). (a-h) violin plots for priors and inferred posteriors parameter distribution (a-d) using NHSD (e-h) using HSD; prior and posterior distributions violin plots for (a/e) HSCs self-renewal rate; (b/f) HSCs asymmetric division rate; (c/g) MPPs decay rate; (d/h) HSCs effective differentiation rate. (i-p) dynamic uncertainty plots where where the shaded areas refer to the 95% credible interval posterior predictive distributions of all the ODE solution trajectories without accounting for the technical errors. The curves refer to the median of the shaded areas. The line plots refer to the 95% credible interval of the posterior predictive intervals that account for technical errors at the data observation time. (i,j): log scale NHSD HSCs count fitting by the NHCR/HCR model (k,l): log scale NHSD MPPs data fitted by NHCR/HCR models (m,n): log scale HSD HSCs data using the NHCR/HCR models; (o,p): log scale HSD MPPs data fitted by the NHCR/HCR

Validation of the Inference Framework with Bacth Effects

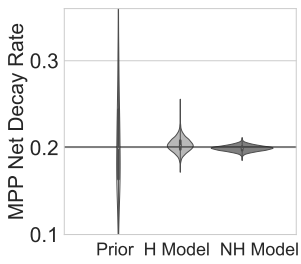
We next incorporate an offset in our synthetic datasets. We denote the offset version of NHSD as ONHSD and the offset version of HSD as OHSD. Figure 3.5 shows the prior and posterior distribution violin plots of the inferred rate parameters using NHSD (a-e) and HSD (f-j). We can see that inference using both HCR/NHCR models recover the ground truth values of the parameters reasonably, except the offset parameter in Figure 3.5 (j), where the NHCR ODE model struggles to locate the ground truth. Figure 3.5 (k-z) shows that the posterior predictive distributions of both the HCR/NHCR ODE models can fit both synthetic datasets well. Again the HCR and NHCR models differ in the classification of the variability in the data. The Bayesian model selection using the Bayes factor Eqn. (2.70) is also able, in the presence of the offset, to decisively select the right mode (Table 3.2). (See



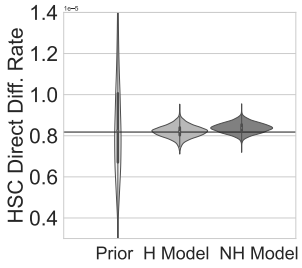
(a) HSC Self-Renewal Rate fit to NHSD



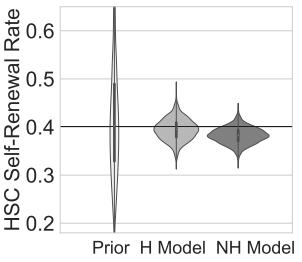
(b) HSC Asym Division Rate fit to NHSD



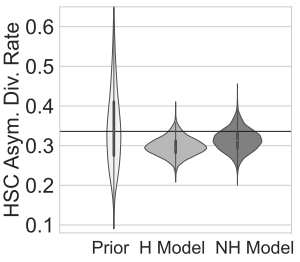
(c) MPP Decay Rate fit to NHSD



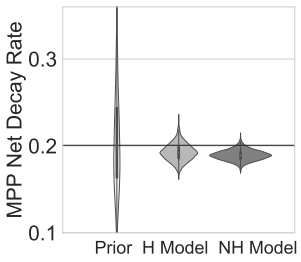
(d) HSC Direct Diff Rate fit to NHSD



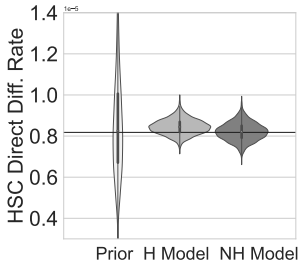
(e) HSC Self-renewal Rate fit to HSD



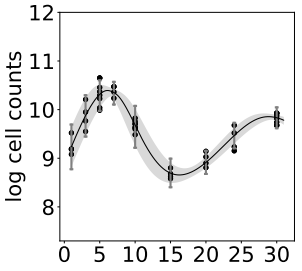
(f) HSC Asym Division Rate fit to HSD



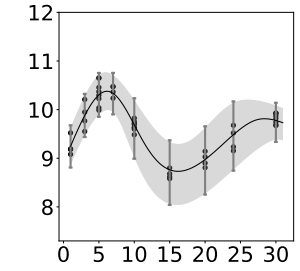
(g) MPP Decay Rate fit to HSD



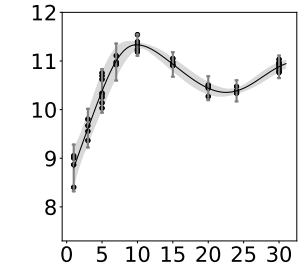
(h) HSC Direct Diff Rate fit to HSD



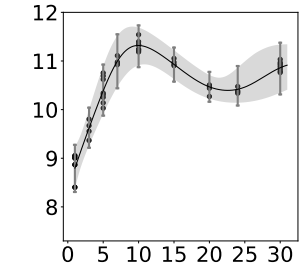
(i) fit NH model to NHSD HSCs Count



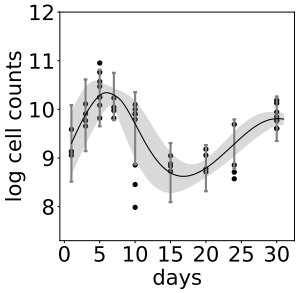
(j) fit H Model to NHSD HSCs Count



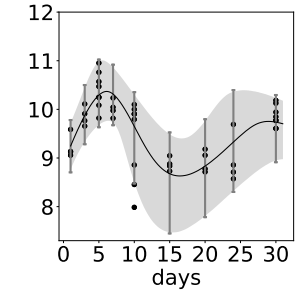
(k) fit NH Model to NHSD MPPs Count



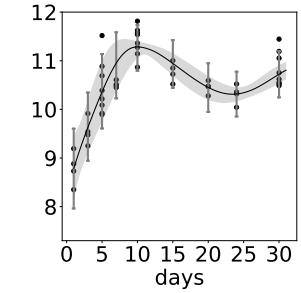
(l) fit H Model to NHSD MPPs Count



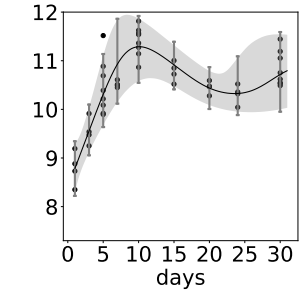
(m) fit NH Model to HSD HSCs Count



(n) fit H Model to HSD HSCs Count

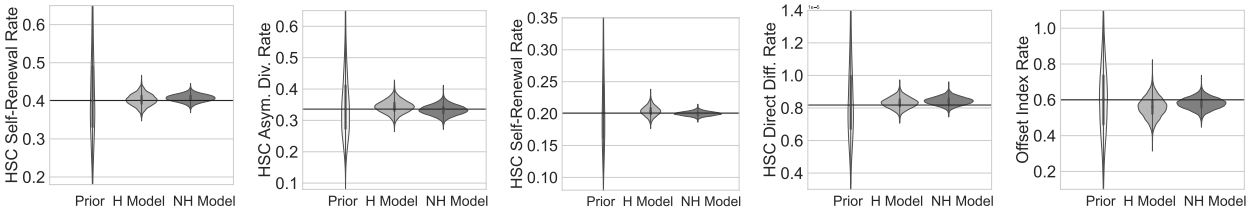


(o) fit NH Model to HSD MPPs Count

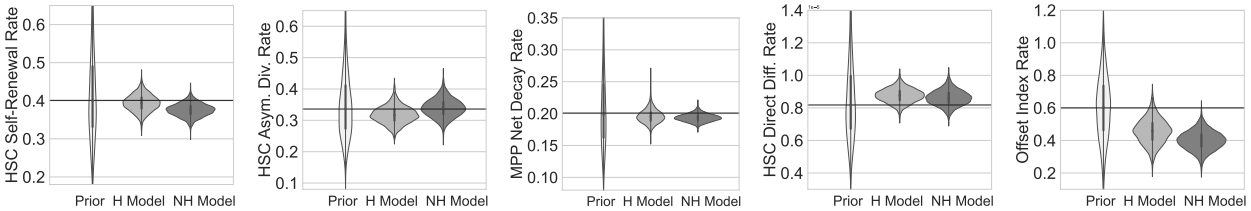


(p) fit H Model to HSD MPPs Count

also Appendix Figures A.8, A.9, A.10, A.11 for prior and posterior distribution plots.)



(a) HSC Self Renewal Rate Fit Models to ONHSD (b) HSC Asym Division Rate Fit Models to ONHSD (c) MPP Decay Rate Fit Models to ONHSD (d) HSC Direct Diff Rate Fit Models to ONHSD (e) Offset Parameter Value Fit Models to ONHSD



(f) HSC Self Renewal Rate Fit Models to OHSD (g) HSC Asym Division Rate Fit Models to OHSD (h) MPP Decay Rate Fit Models to OHSD (i) HSC Effective Diff. Rate Fit Models to OHSD (j) Offset Parameter Value Fit Models to OHSD

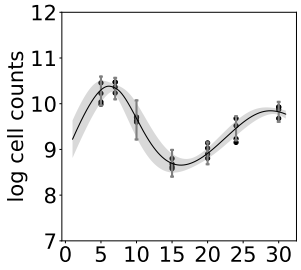
Figure 3.5: Prior and posterior violin plots for fitting both non-hierarchical and hierarchical chemical reaction model to the multiple synthetic datasets with offset parameter generated from both models. Fitting offset version hierarchical chemical reaction (OHCR) and non-hierarchical chemical reaction (ONHCR) model to synthetic dataset generated from both models (OHSD, ONHSD); (a-j): violin plots of prior and inferred posteriors parameter distributions (a-e): using ONHSD (f-j): using OHSD; violin plots of prior and inferred posteriors parameter distributions for (a,f): HSCs self-renewal rate (b,g): HSCs asymmetric division rate (c,h): MPPs decay rate (d,i): HSCs effective differentiation rate (e,j): offset parameter index value between two experimental datasets due to the batch effects

3.3.2 Bayesian Hierarchical Inference and Model Selection for Experimental Data

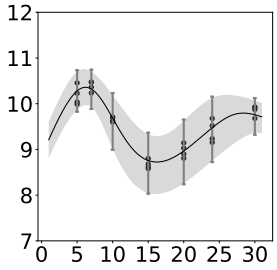
Next, we infer the parameters using the HCR and NHCR models for the experimental datasets. The results are shown in Figure 3.7; in (a-c), we can see that the inference can identify the rate value parameters and the offset parameter with the experimental dataset. Figures 3.7 (f-m) show the posterior predictive distribution plots for the experimental data. We can see that the 95% credible interval posterior plots, without consideration of technical

Figure 3.6: Dynamic uncertainty plots for fitting both non-hierarchical and hierarchical chemical reaction (HCR/NHCR) model to the multiple synthetic datasets with offset parameter generated from both models (HSD/NHSD). Page(57-58). (a-h) dynamic uncertainty plots where the shaded areas refer to the 95% credible interval posterior predictive distributions of all the ODE solution trajectories without accounting for the technical errors. The curves refer to the median of the shaded areas. The line plots refer to the 95% credible interval of the posterior predictive intervals that account for technical errors at the data observation time. (a,e): log scale NHSD1 HSCs, MPPs data fitted by the NHCR model (b,f) log scale NHSD1 HSCs, MPPs data fitted by the HCR model (c,g) log scale NHSD2 HSCs, MPPs data fitted by the NHCR model (d,h) log scale NHSD2 HSCs, MPPs data fitted by the HCR model (i,m): log scale HSD1 HSCs, MPPs data fitted by the NHCR model (j,n) log scale HSD1 HSCs, MPPs data fitted by the HCR model (k,o) log scale HSD2 HSCs, MPPs data fitted by the NHCR model (l,p) log scale HSD2 HSCs, MPPs data fitted by the HCR model

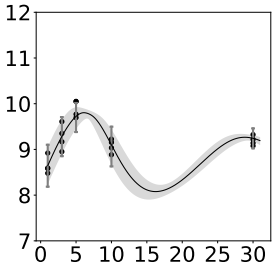
errors in the HCR model, cover more data points than those of the NHCR model. Accordingly, the Bayes factor selects the HCR model decisively (table 3.2). (See Appendix Figures A.12, A.14 in for the prior and posterior distribution)



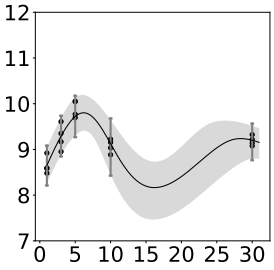
(a) Fit NHCR Model to NHSD1 HSCs Data



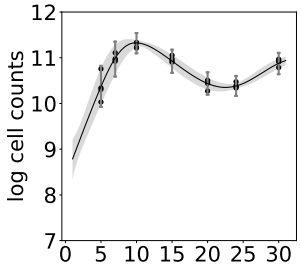
(b) Fit HCR Model to NHSD1 HSCs Data



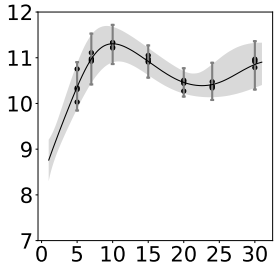
(c) Fit NHCR Model to NHSD2 HSCs Data



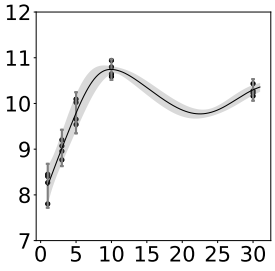
(d) Fit HCR Model to NHSD2 HSCs Data



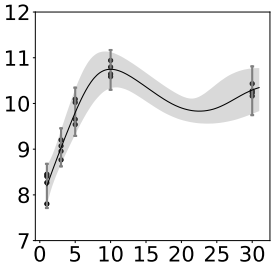
(e) Fit NHCR Model to NHSD1 MPPs Data



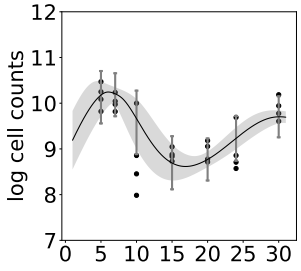
(f) Fit HCR Model to NHSD1 MPPs Data



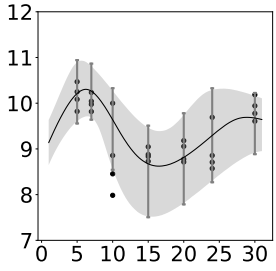
(g) Fit NHCR Model to NHSD2 MPPs Data



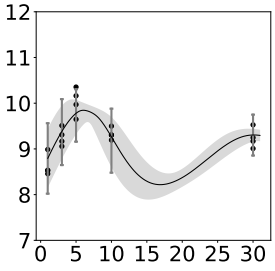
(h) Fit HCR Model to NHSD2 MPPs Data



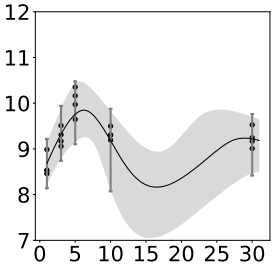
(i) Fit NHCR Model to HSD1 HSCs Data



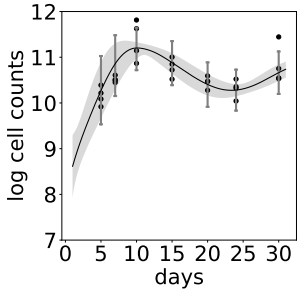
(j) Fit HCR Model to HSD1 HSCs Data



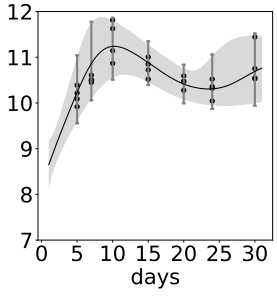
(k) Fit NHCR Model to HSD2 HSCs Data



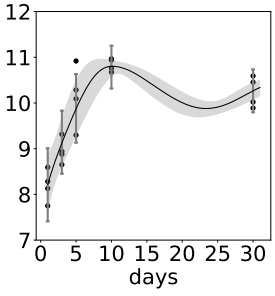
(l) Fit HCR Model to HSD2 HSCs Data



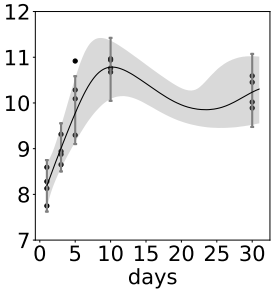
(m) Fit NHCR Model to HSD1 MPPs Data



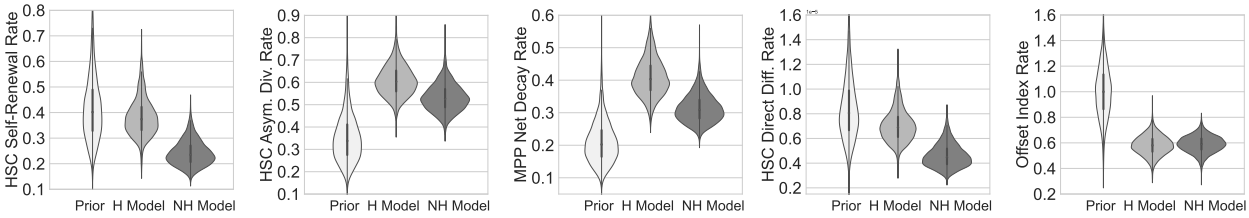
(n) Fit HCR Model to HSD1 MPPs Data



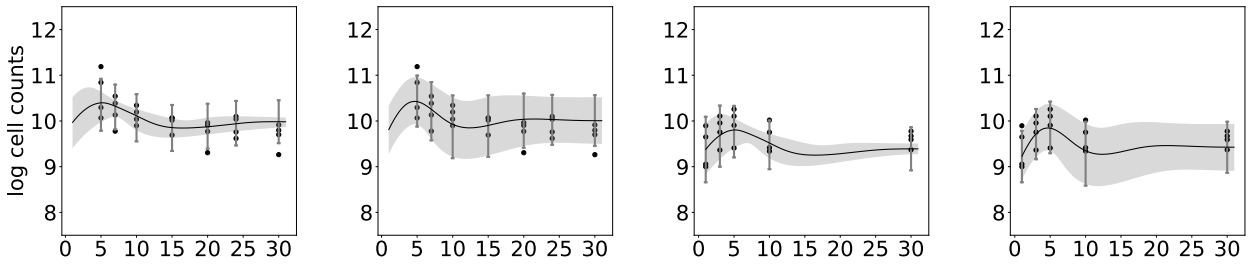
(o) Fit NHCR Model to HSD2 MPPs Data



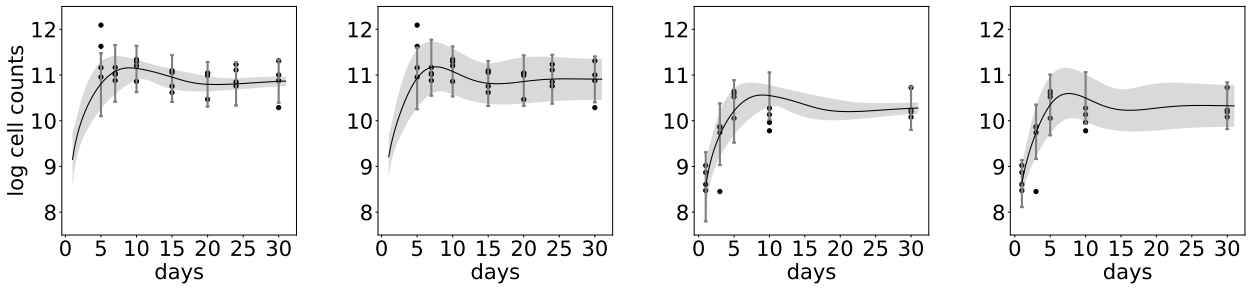
(p) Fit HCR Model to HSD2 MPPs Data



(a) Fit Models to ExpD HSC Self-renewal Rate (b) Fit Models to ExpD HSC Asym. Div. Rate (c) Fit Models to ExpD MPP Decay Rate (d) Fit Models to ExpD HSC Direct Diff. Rate (e) Fit Models to ExpD Offset Parameter Value



(f) Fit NHCR Model to ExpD1 HSCs Data (g) Fit HCR Model to ExpD1 HSCs Data (h) Fit NHCR Model to ExpD2 HSCs Data (i) Fit HCR Model to ExpD2 HSCs Data



(j) Fit NHCR Model to ExpD1 MPPs Data (k) Fit HCR Model to ExpD1 MPPs Data (l) Fit NHCR Model to ExpD2 MPPs Data (m) Fit HCR Model to ExpD2 MPPs Data

Figure 3.7: Fit both hierarchical chemical reaction (HCR) and non-hierarchical chemical reaction (NHCR) model to experimental datasets (ExpD). (a-e) violin plots for priors and inferred posteriors parameter distributions inference (a) HSC self-renewal rate; (b) HSC asymmetric division rate; (c) MPP decay rate; (d) HSC direct differentiation rate; (e) offset parameter between two experimental datasets due to the batch effects; (f-m) dynacmic uncertainty plots for fitting both models to ExpD where the shaded areas refer to the 95% credible interval posterior predictive distributions of all the ODE solution trajectories without accounting for the technical errors. The curves refer to the median of the shaded areas. The line plots refer to the 95% credible interval of the posterior predictive intervals that account for technical errors at the data observation time. (f,j): log scale ExpD1 HSC, MPP data fitted by the NHCR model (g,k) log scale ExpD1 HSC, MPP data fitted by the HCR model (h,l) log scale ExpD12 HSC, MPP data fitted by the NHCR model (i,m) log scale ExpD2 HSC, MPP data fitted by the HCR model

3.3.3 Model Selection for Hematopoiesis Dynamics

We now present the datasets of model selection for the synthetic and experimental datasets. Denote model 0 to be the NHCR model. Let model 1 be the HCR model. Let BF10 defined in Eqn.(2.70) be the Bayes factor of model 1 in favor of model 0 and BF01 be the Bayes factor of model 0 in favor of model 1. In table 3.1, we can see that the Bayes factor can decisively select the right model which synthetic data is generated from based on Bayes factor interpretation table 2.2. Bayes factor decisively selects the hierarchical chemical reaction model for the experimental dataset which indicates that more stochasticity needs to take into account besides the cell count initial condition biological variation and operation technical error. Note that our hierarchical chemical reaction model mimics that real stochastic ODE model, this result agrees with our motivation to replace ODE with a stochastic process model e.g., LNA for the latent variable approach described in section 3.2.1.

Bayes Factor Model Selection Table					
Dataset	NHCR LML	HCR LML	$log_{10}(BF01)$	$log_{10}(BF10)$	Model Selection
NHSD	37.84	18.08	**8.58**	-8.58	NHCR Decisive
HSD	-68.03	-36.72	-13.6	**13.6**	HCR Decisive
ONHSD	30.66	8.1	**9.79**	-9.79	NHCR Decisive
OHSD	-67.56	-39.02	-12.39	**12.39**	HCR Decisive

Table 3.1: Bayes factor model selection table for 2D synthetic hematopoiesis datasets. NHCR: non-hierarchical chemical reaction model; HCR: hierarchical chemical reaction mode; LML: log marginal likelihood; NHSD: synthetic datasets generated by NHCR model; HSD: synthetic datasets generated by HCR model; ONHSD: synthetic datasets generated by NHCR model with offset index between two synthetic dataset due to the batch effects; OHSD: synthetic datasets generated by HCR model with offset index between two synthetic dataset due to the batch effects.

We also fit the non-hierarchical and hierarchical hill equation model to the experimental dataset. The prior posterior violin plots and dynamic uncertainty plots can be found in the Appendix Figures A.13 and Figure A.15. Table 3.2 shows the model selections among four models (HCR, HHE, NHCR, and NHHE) by model selection methods Bayes factor and PSIS-LOO described in section 2.3. Based on 8000 iterations of MCMC for each model,

we can see that the HCR model achieves the highest log model marginal likelihood -71.31 calculated by Eqn.(2.67) and Bayes factor model probability 0.584 calculated by Eqn. (2.68). PSIS-LOO also shows that the HCR model has the highest out-of-sample cross-validation expected log predictive density (cv_elpd) 80.19 and the highest in-sample cross-validation expected log predictive density (naïve_elpd) 86.01 which leads to highest in-sample predict density by Eqn. (2.73). HCR model also achieves the highest PSIS-LOO model weight 0.982 by Eqn.(2.74).

Model Selection Table for Experimental Data				
Metric / Model	HCR	HHE	NHCR	NHHE
Log Marginal Likelihood	**-71.16**	-71.66	-86.95	-93.09
Posterior Model Probability	**0.621**	0.378	8.67e-8	1.86e-10
Log Bayes Factor HCR/Model	**0.00**	0.494	15.78	21.93
PSIS-LOO cv_elpd	**-5.82**	-9.87	-46.81	-47.40
PSIS-LOO naïve_elpd	**80.19**	79.43	-8.98	-6.60
PSIS-LOO loo_weight	**0.982**	0.171	1.5e-18	8.6e-19
Number of Parameters	287	287	105	105

Table 3.2: Bayes factor and PSIS-LOO-CV model selection table for 2D experimental datasets. NHCR: non-hierarchical chemical reaction model; HCR: hierarchical chemical reaction models; NHHE: non-hierarchical hill equation model; HHE: hierarchical hill equation model; LML: log marginal likelihood; NHSD: synthetic datasets generated by NHCR model; HSD: synthetic datasets generated by HCR model; ONHSD: synthetic datasets generated by NHCR model with offset index between two synthetic dataset due to the batch effects; OHSD: synthetic datasets generated by HCR model with offset index between two synthetic dataset due to the batch effects; PSIS-LOO cv_epld: out-of-sample leave-one-out cross-validation expected log predictive density; PSIS-LOO naïve_elpd: in-sample leave-one-out cross-validation expected log predictive density. PSIS-LOO loo_weight: model weight by PSIS-LOO

10 Fold Cross Validation and Leave-Out-A-Day validation

We also performed 10-fold cross-validation (see Appendix Figures A.16-A.19) and leave-out-a-day validation (see Appendix Figure A.20) to test the model and Bayesian framework

robustness. The hierarchical and non-hierarchical posterior predictive summary can be seen in appendix tables (A.3 and A.4). Those validations are measured by median absolute error (MAE), which describes the typical difference between the observed value and their posterior predictive means and the median absolute scaled error (MASE) measures the typical number of standard deviations that the observed values fall from their posterior predictive values. In the 10-fold cross-validation, we can see that both HSC and MPP validation data fall in 1.5 standard deviations from their posterior predictive mean, which means that they are inside their 95% CI posterior prediction intervals for most of the folds. Leaving out the data on day 5 or 7 leads to larger MASE, and consequently, fewer validation data fall in the 95% posterior prediction interval see Appendix Figure A.20. This means that the cell count overshoot produced by feedback regulation plays a significant role in hematopoiesis dynamics.

3.4 Conclusion

In this project, we develop a new chemical reaction (CR) model for the hematopoiesis dynamic, which can be easily extended to a stochastic process model (e.g., SDE, LNA) and a hierarchical Bayesian framework for parameter inference. We validate the CR model and the hierarchical inference framework with/without batch effects by comparing the results of fitted models to synthetic datasets and model selection criteria (Bayes factor and PSIS-LOO-CV). We find that the HCR is the most plausible model that describes experimental datasets. This suggests there is significant biological variability among data. As the dynamic system re-equilibrates, oscillations consisting of stable spirals are observed, which is the characteristic of negative feedback regulation as expected. However, by observing that the HCR fitted posterior distributions of parameters, we find that the HSC self-renewal rate is higher than expected since we lump the rapidly dividing short-term HSCs (STHSCs) and

more slowly dividing long-term HSCs (LTHSCs). For future work, we will invest extended model by separating HSCs into STHSCs and LTHSCs. Besides, we will also incorporate variability through the stochastic process model.

Chapter 4

Stochastic Systems and Parameter Inference for Models of Hematopoiesis

4.1 Introduction

Hematopoiesis is a complex mechanism by which hematopoietic stem cells produce a variety of functional blood cells through multiple stages of differentiation. Since the numbers of various blood cell types need to be maintained in homeostasis, with occasional short-lived departures from it, hematopoiesis must be subject to multiple regulatory mechanisms. However, these regulatory mechanisms are still not well understood. Mathematical models can help us formulate hypotheses about regulatory mechanisms. However, fitting models to data is challenging because the data — typically hematopoietic stem cell, multipotent progenitor, and mature blood cell counts — is noisy. This motivates us to use a latent variable approach to account for the unobserved data. In Chapter 3, using a new chemical reaction (CR) ordinary differential equation model of negative feedback regulation in hematopoiesis, we develop a scalable, hierarchical Bayesian framework that accounts for across-subject hetero-

geneity and infers division and feedback regulation parameters governing hematopoietic cell dynamics. We evaluate the performance of the new model and inferential framework using synthetic data and data from experiments in mice treated with the chemotherapy drug 5FU to perturb the hematopoietic system. The framework distinguishes between models that account for biological variation and those with only technical variation/measurement error. We find that the experimental data are best described by a hierarchical model in which the hematopoiesis model parameters were allowed to vary among mice, suggesting the presence of significant biological variability. As the system re-equilibrates, oscillations are observed, which are characteristic of negative feedback. In this work, we extend the CR model to account for more hematopoietic cell compartment types to explore an alternative way of accounting for data heterogeneity by employing a stochastic differential equation (SDE) model of hematopoiesis instead of letting division and feedback regulation parameters vary across mice. We use a linear noise approximation (LNA) to obtain a tractable likelihood function, which enables us to use Bayesian parameter inference without additional approximations. We test the performance of the stochastic inference framework using synthetic hematopoiesis datasets and find a nontrivial relationship between the variability of the initial condition, measurement error, and variability in the stochastic dynamical system. We find that using the stochastic model to interpolate the latent cell counts can further improve the PSIS-LOO out-of-sample cross-validation expected log predictive density $elpd_{loo}$ Eqn.(2.73) from the non-hierarchical/hierarchical ODE model frameworks we developed in Chapter 3. By fitting the LNA model to the experimental data, we find a limitation in using the stochastic process model to interpolate missing data when the latent initial condition cell count variation at the first observation day is substantial since the different sources of biological variations can compensate for one another. Thus, we propose a solution for this limitation by estimating the initial condition variance and technical measurement errors directly from the experimental dataset. By fixing the initial condition variance and technical measurement error estimated from the experiment dataset, we find the experimental dataset is best described

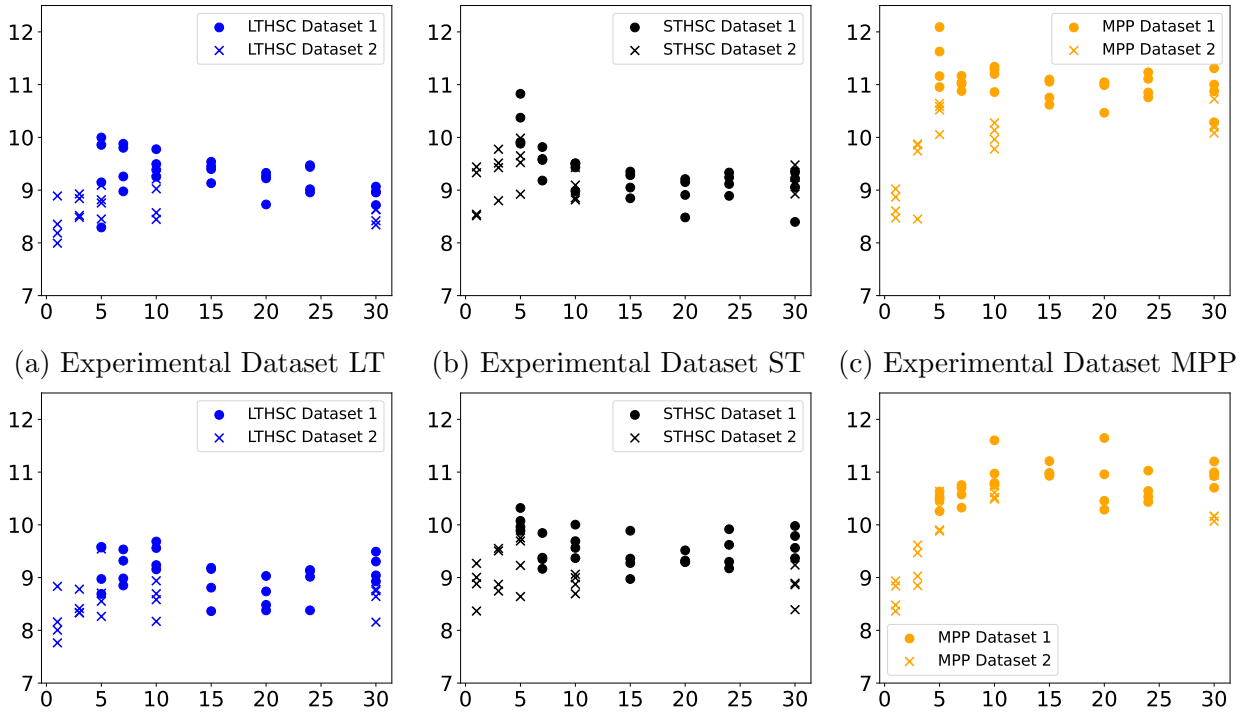
by the stochastic LNA parameter inference model.

4.2 Methods

Hematopoietic stem cells (HSCs) can be divided into long-term HSCs (LTHSCs) and short-term HSCs (STHSCs). LTHSCs are highly self-renewing cells with long-term reconstitution capacity through the entire animal life span and LTHSCs can differentiate into STHSCs, which only have a short-term reconstitution ability. STHSCs can further differentiate into multipotent progenitor cells (MPPs), which are believed to only briefly self-renew or do not have self-renewal ability [13, 52]. For data in Chapter 3, LTHSCs and STHSCs are aggregated. However, since the aggregated HSCs self-renewal rate for the best-fitted HCR model is a bit higher than expected and the larger thae the cell counts for each compartment, the similar NHODE and LNA forward simulation results, in this work, we consider both types of HSCs and extend the two-compartment chemical reaction model used in Chapter 3 to a three-compartment model by separating the HSCs into LTHSCs and STHSCs. The experimental datasets for the three-compartment model can be seen in Figure 4.1 (a-c), where the dots refer to the data from experiment one and cross marks refer to the data from experiment two.

4.2.1 3-Compartment Chemical Reaction ODE Model

Then we consider the following reactions: $\mathcal{R}_1$ LTHSC can self-renewal with rate v_1 refers to the highly self-renewing feature; LTHSC can differentiate into STHSC in two ways: $\mathcal{R}_2$ LTHSC can do asymmetric division, which produces one LTHSC and one STHSC with rate v_2 and $\mathcal{R}_3$ LTHSC can interact with MPP and directly differentiate into STHSC with rate constant r_1. Since STHSC are short-term, we don't directly model its self-renewal rate and



(a) Experimental Dataset LT (b) Experimental Dataset ST (c) Experimental Dataset MPP

(d) Synthetic Dataset LT (e) Synthetic Dataset ST (f) Synthetic Dataset MPP

Figure 4.1: Experimental dataset and synthetic dataset for 3-compartment hematopoiesis dynamics. LT: long-term HSCs; ST: short-term HSCs; dot points are the data from experimental one and cross marks are the data from experimental two (a,b,c) referring to the LT, ST, MPP dataset for the two real experiments; (d,e,f) refer to LT, ST, MPP synthetic dataset which is generated by Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data.

death rate directly. Consider STHSC also differentiates into MPP in two ways: $\mathcal{R}_4$ STHSC can do asymmetric division, which produces one STHSC and one MPP with the rate v_3 and $\mathcal{R}_5$ STHSC can interact with MPP and directly differentiate into MPP with rate constant q_1. $\mathcal{R}_6$ MPP lumped decay with rate u_3. Then those reactions can be written as

$$\mathcal{R}_1 : LTHSC \xrightarrow{v_1} 2LTHSC; \quad \mathcal{R}_2 : LTHSC \xrightarrow{v_2} LTHSC + STHSC;$$

$$\mathcal{R}_3 : LTHSC + MPP \xrightarrow{r_1} STHSC + MPP; \quad \mathcal{R}_4 : STHSC \xrightarrow{v_3} STHSC + MPP;$$

$$\mathcal{R}_5 : STHSC + MPP \xrightarrow{q_1} 2MPP; \quad \mathcal{R}_6 : MPP \xrightarrow{u_3} \phi. \tag{4.1}$$

The 3-compartment ODE system for these reactions can be written as

$$\begin{cases} N'_{LTHSC} = v1 \cdot N_{LTHSC} - r1 \cdot N_{LTHSC} \cdot N_{STHSC} \\ N'_{STHSC} = v2 \cdot N_{LTHSC} + r1 \cdot N_{LTHSC} \cdot N_{STHSC} - q1 \cdot N_{STHSC} \cdot N_{MPP} \\ N'_{MPP} = v3 \cdot N_{STHSC} + q1 \cdot N_{STHSC} \cdot N_{MPP} - u3 \cdot N_{MPP}. \end{cases} \tag{4.2}$$

The diagram illustration for those reactions can be seen in Figure 4.2 where the MPPs cell count negatively feedback on both LTHSCs and STHSCs' self-renewal rate.

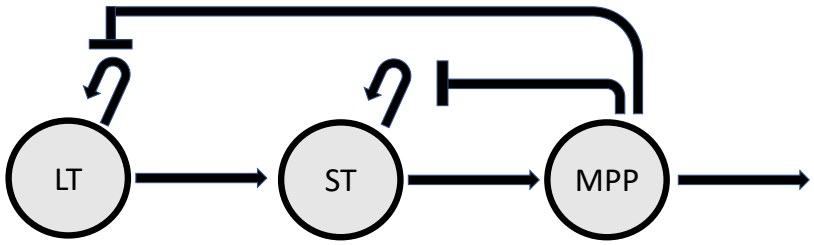


Figure 4.2: Diagram of 3D chemical reaction model for hematopoiesis dynamics. LT, ST, MPP refer to long-term HSC, short-term HSC and MPP respectively. The system involves two feedback regulations: MPPs cell count negatively feedback regulation on both LTHSCs and STHSCs' self-renewal rate. Since STHSC are short-term, we don't directly model its self-renewal rate and death rate directly. The corresponding chemical reaction equations for this diagram can be seen at section .(4.1)

4.2.2 Derivation of LNA for 3D Chemical Reaction Model

Following the LNA derivation in section 2.2.3, we first write the ODE system Eqn.(4.2) into the multiplication form of stochemistry matrix, which rows refer to the different cell

compartments and columns refer to different reaction channels,

$$S = \begin{array}{c} \begin{array}{cccccc} \mathcal{R}_1 & \mathcal{R}_2 & \mathcal{R}_3 & \mathcal{R}_4 & \mathcal{R}_5 & \mathcal{R}_6 \end{array} \\ \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{pmatrix} \end{array} \begin{array}{c} LTHSC \\ STHSC \\ MPP, \end{array} \tag{4.3}$$

and the reaction propensity vector which consist of all the transition rate of Markov jump process for the reactions is given by

$$h(\mathbf{X}(t), \theta) = \begin{pmatrix} v_1 \cdot LTHSC(t) \\ v_2 \cdot LTHSC(t) \\ r_1 \cdot LTHSC(t) \cdot MPP(t) \\ v_3 \cdot STHSC(t) \\ q_1 \cdot STHSC(t) \cdot MPP(t) \\ u_3 \cdot MPP(t) \end{pmatrix} \begin{array}{c} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \mathcal{R}_3 \\ \mathcal{R}_4 \\ \mathcal{R}_5 \\ \mathcal{R}_6. \end{array} \tag{4.4}$$

Let $H(\mathbf{X}(t), \theta = Diag\{h(\mathbf{X}(t), \theta\}$ be the reaction propensity matrix. Based on the derived CLE Eqn.(2.40) by applying the normal approximation to the independent homogenous Possion process, the SDE of this chemical reaction model takes the form

$$d\boldsymbol{X}(t) = \boldsymbol{S}\boldsymbol{h_i}(X(t), \boldsymbol{\theta})dt + Diag\{\sqrt{\boldsymbol{S}\boldsymbol{H}(\boldsymbol{X}(t), \boldsymbol{\theta})\boldsymbol{S}^t}\}dW_t. \tag{4.5}$$

Log Transform of the Differential Equations

Solving differential equations with initial conditions perturbed from their equilibrium can lead the time evolution sample paths to vary in different magnitudes and consequently, it is more reasonable to infer multiplicative noise terms instead of inferring additive noise

terms for Bayesian computation. We can transform our differential equations all into the log scale for better computation performance. Another reason is that for some situations e.g. bone marrow transplantation operation dataset where the initial conditions of compartments can be very small, when solving LNA forward simulations, the compartment number can go negative which may break down the LNA inference process because it can make variance process $\Sigma(t)$ non-positive definite and cannot do Cholesky decomposition. Let $\hat{X}(t) = \phi(X(t), t) = log(X(t)+1) \rightarrow X(t) = exp(\hat{X(t)}) - 1$. By Lemma 2.2, we can get that

$$\frac{\partial \phi(X(t),t)}{\partial X} = \frac{1}{X}, \quad \frac{\partial^2 \phi(X(t),t)}{\partial X^2} = \frac{-1}{X^2}, \quad \frac{\partial \phi(X(t),t)}{\partial t} = 0 \tag{4.6}$$

and the log-transformed SDE becomes

$$\begin{aligned} d\hat{X}(t) &= Diag(exp(-\hat{X}(t)) - 0.5\ exp(-2\hat{X}(t)))f(exp(\hat{X}(t)) - 1, t) \\ &\quad + Diag(exp(-\hat{X}(t)))g(exp(\hat{X}(t)) - 1, t) \\ &= f(\hat{X}, t)dt + g(\hat{X}, t)dW_t \end{aligned} \tag{4.7}$$

For the reasons described above, we log-transform the cell compartment counts into

$$\hat{\boldsymbol{X}}(t) = log(\boldsymbol{X}(t) + 1), \tag{4.8}$$

and $\boldsymbol{X}(t) = exp(\hat{\mathbf{X}}(t)) - 1$. By lemma 2.2, we can log-transform SED Eqn.(4.5) by using changing of variable $\boldsymbol{X(t)}$ Eqn.(4.8) to get

$$\begin{aligned} d\hat{\boldsymbol{X}}(t) = &\overbrace{Diag\left\{exp(-\hat{\boldsymbol{X}}(t)) - 0.5\ exp(-2\hat{\boldsymbol{X}}(t))\right\} \cdot \boldsymbol{Sh}(exp(\hat{\boldsymbol{X}}(t)) - 1, \boldsymbol{\theta})}^{\textit{log-transformed deterministic solution process } \boldsymbol{\eta^*}(\hat{\boldsymbol{X}}(t))} dt \\ &+ \overbrace{Diag\left\{exp(-\hat{\boldsymbol{X}}(t))\right\} \cdot Diag\left\{\sqrt{\boldsymbol{SH}(exp(\hat{\boldsymbol{X}}(t)) - 1, \boldsymbol{\theta})\boldsymbol{S}^t}\right\}}^{\textit{log-transformed Gaussian process } (\boldsymbol{\Phi^*}(\hat{\boldsymbol{X}}(t)))^{1/2}} dW_t. \end{aligned} \tag{4.9}$$

The Jacobian matrix for this equation can be written as $J(\hat{\boldsymbol{X}}(t)) = \frac{\partial \boldsymbol{\eta}(\hat{\boldsymbol{X}}(t))}{\partial \hat{\boldsymbol{X}}}$. From the LNA formula Eqn.(2.65) derived in section 2.2.3, we obtain the following LNA system for this log-transformed problem:

$$\begin{cases} d\boldsymbol{\eta}(t) = \boldsymbol{\eta}^*(\hat{\boldsymbol{X}}(t)) \\ d\boldsymbol{m}(t) = J(\hat{\boldsymbol{X}}(t))\boldsymbol{m}(t)dt \\ d\boldsymbol{\Sigma}(t) = \Big(J(\hat{\boldsymbol{X}}(t))\boldsymbol{\Sigma}(t) + (\boldsymbol{\Phi}^*(\hat{\boldsymbol{X}}(t)))^{1/2}((\boldsymbol{\Phi}^*(\hat{\boldsymbol{X}}(t)))^{1/2})^t + \boldsymbol{\Sigma}(\hat{\boldsymbol{X}}(t))J(\hat{\boldsymbol{X}}(t))\Big)\, dt. \end{cases} \tag{4.10}$$

4.2.3 LNA Framework for Partially Observed Datasets

As the Figure 4.3 shows for LNA diagram illustration, denote the time at which 5-FU is administered as t_p; In order to take into account the wide discrepancies between the raw counts of different cell compartments, we decide to transform all the cell count numbers into a logarithmic scale. We assume that the log-transformed LTHSCs, STHSCs, and MPPs counts for all the mice at initial observation day t_0 after the perturbation experiment follow multivariate normal distributions with means equal to μ_{LT}, μ_{ST} and μ_{MPP} respectively and the same standard deviation δ_{init}. This model variance among the initial cell counts is due to biological differences between the mice. Let $\boldsymbol{\mu}$ (red star in Figure 4.3) be the vector for the μ_{LT}, μ_{ST} and μ_{MPP} and $\boldsymbol{x_0^i}$ be the log-scale cell count initial conditions for mouse number i. Suppose that the total number of mice is N. Then, we can write the initial conditions for each mouse as

$$\boldsymbol{x_i^0} \sim MvN(\boldsymbol{\mu}, \delta_{init}^2 \boldsymbol{I}), \tag{4.11}$$

for all $i \in \{1, 2, \cdots, N\}$. Without loss of generality, we assume that the first m mice are sacrificed and observed during the first observation day t_0 of the experiment right after the 5-FU administration. Let $\boldsymbol{y_i}(t)$ be the log-scale observed cell counts data for mouse i at time

t and assume the observed data is obtained with technical measurement errors δ_{tech}. Then for the mice that are sacrificed to the day t_0 after 5-FU is applied as

$$\boldsymbol{y_i}(t_0) \sim MvN(\boldsymbol{x_i^0}, \delta_{tech}^2 \boldsymbol{I}), \tag{4.12}$$

for $\forall\ i \in \{1, 2, \cdots, m\}$.

We choose the time step Δt for the LNA algorithm by first running the MAP optimization on fitting Bayesian model to the experimental datasets and then using those MAP point estimation parameters to generate LNA forward simulations with different Δt and compare them with Gillespie and SDE solutions. Based on forward simulation results, we select Δt to be a 1-day interval which balances the computational efficiency and accuracy of LNA approximation to the SDE solution. Let variables $\boldsymbol{x}_j^i$ refer to the latent LNA intermediate step j cell counts for mouse i. For each mouse i, the intermediate state cell counts at LNA step j can be calculated by

$$\boldsymbol{x}_i^j \sim N(\boldsymbol{\eta}_j + \boldsymbol{m}_j, \boldsymbol{\Sigma_j}), \tag{4.13}$$

where the deterministic mean process solution $\boldsymbol{\eta}_j$, Gaussian mean process solution $\boldsymbol{m}_j$, Gaussian variance process solution $\boldsymbol{\Sigma_j}$ are calculated by using ODE system Eqn.(4.10) with initial condition $\boldsymbol{\eta}_{j-1}, \boldsymbol{m}_{j-1}, \boldsymbol{\Sigma_{j-1}}$ and rate value parameters $\boldsymbol{\theta} = \{v_1, v_2, v_3, r_1, q_1, u_3\}$ which are sampled from the prior distributions in time interval Δt. For those mice whose cell counts are measured after t_0, assume that mouse j is sacrificed at day t_{LNA_j}. We assume that the latent trajectories for the cell populations follow an LNA model. This allows us to simulate all the cell counts before the mouse is harvested for measurement. Let the log-scale dataset offset index be $\boldsymbol{\tau}$ between the two experiments due to using different filters. Then, the log-scale observed cell accounts for mouse j at time t_{LNA_j} can be represented as

$$\mathbb{1}_{\text{is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y_j}(t_{LNA_j}) \sim MvN(\boldsymbol{x^j_{LNA_j}}, \delta_{tech}^2 \boldsymbol{I}), \tag{4.14}$$

for all $j \in \{m+1, m+2, \cdots, N\}$. Following our data-generating procedure, we also assume that observations are independent of each other, then the posterior density can be represented as Let $\boldsymbol{\Theta} = \{\boldsymbol{\theta}, \boldsymbol{x}_0, \delta_{init}, \delta_{tech}, \boldsymbol{\tau}\}$ be the prior parameters which are sampled by the non-centered parameterization described in Appendix section A.0.2 which the non-centered prior mean $\boldsymbol{\mu_\Theta}$, standard deviation $\boldsymbol{\delta_\Theta}$ and center $\boldsymbol{Z_\Theta}$. Then posterior density value can be written as

$$Pr(\boldsymbol{\Theta}, \boldsymbol{X} | \mathbb{1}_{\text{is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y}) \propto Pr(\mathbb{1}_{\text{is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y} | \boldsymbol{\Theta}, \boldsymbol{X}) Pr(\boldsymbol{X} | \boldsymbol{\Theta}) Pr(\boldsymbol{\Theta}) \quad (4.15)$$

Then the posterior density function can be written as

$$\begin{aligned} &Pr(\mathbb{1}_{\text{ is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y} | \boldsymbol{\Theta}, \boldsymbol{X}) Pr(\boldsymbol{X} | \boldsymbol{\Theta}) Pr(\boldsymbol{\Theta}) = \\ &\qquad \prod_{j=1}^{N} Pr(\mathbb{1}_{\text{j is in experiment 2}} \cdot \boldsymbol{\tau} + \boldsymbol{y_j} | \boldsymbol{X_j}^{\boldsymbol{t_{LNA_j}}}, \delta_{tech} \cdot \boldsymbol{I}) \cdot \\ &\qquad\quad \prod_{i=1}^{LNA_j} Pr(\boldsymbol{X_j^{t_i}} | DoLNA(\boldsymbol{X_j^{t_{i-1}}}, \boldsymbol{\theta})) \cdot Pr(\boldsymbol{X_j^{t_0}} | \boldsymbol{u_{init}^{mean}}, \delta_{init} \cdot \boldsymbol{I}) \\ &\qquad\qquad Pr(\boldsymbol{u_{init}^{mean}}, \boldsymbol{\theta}, \delta_{init}, \delta_{tech} | \boldsymbol{\mu_\Theta}, \boldsymbol{\delta_\Theta}, \boldsymbol{Z_\Theta}) \cdot \pi(\boldsymbol{\mu_\Theta}) \pi(\boldsymbol{\delta_\Theta}) \pi(\boldsymbol{Z_\Theta}) \end{aligned} \quad (4.16)$$

Let $\boldsymbol{i}_{obs}$ and $\boldsymbol{j}_{obs}$ be the list of observation days and numbers of mice observed at each observation day. $Flag$ is the matrix that tells which dataset the mice belong to. $\boldsymbol{\theta}$ are the rate value parameters. The LNA inference algorithm includes forward simulation and backward parameter estimation and is given in Algorithm 5 Restart version of LNA Inference.

Algorithm 5 Restarting version LNA forward simulation

Require: $\boldsymbol{\mu_\Theta}$, $\boldsymbol{\delta_\Theta}$, $\boldsymbol{i}_{obs}$, $\boldsymbol{j}_{obs}$, Δt, $\boldsymbol{y}, Flag$

$N_i = len(\boldsymbol{i}_{obs})$

$\boldsymbol{Z_\Theta} \sim N(\boldsymbol{0}, 1)$

$\boldsymbol{\Theta} = \boldsymbol{\mu_\Theta} + \boldsymbol{\delta_\Theta} \cdot \boldsymbol{Z_\Theta}$

$\boldsymbol{\theta}, \boldsymbol{u}_{init}, \delta_{init}, \delta_{tech}, \boldsymbol{\tau} = \boldsymbol{\Theta}$

for i in 1:N_i **do**

  for j in 1:$\boldsymbol{j}_{obs}[i]$ **do**

    $\boldsymbol{u}[j, i, 1] \sim MvN(\boldsymbol{u}_{init}, \delta_{Init} * \boldsymbol{I})$

    if i==1 **then**

      if Flag[j,i]==1 **then**

        $\boldsymbol{y}[j, i] \sim MvN(\boldsymbol{u}[j, i, 1], \delta_{tech} * \boldsymbol{I})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{y}[j, i] \sim MvN(\boldsymbol{u}[j, i, 1], \delta_{Init} * \boldsymbol{I})$

      end if

    else

      $N_{LNA} = \boldsymbol{i}_{obs}/\Delta t + 1$

      for i_{LNA} in 1:N_{LNA} **do**

        $\boldsymbol{\eta_{curr}} \leftarrow d\boldsymbol{\eta}(\boldsymbol{u}[j, i, i_{LNA} - 1], \boldsymbol{\theta}, \Delta t)$

        $\boldsymbol{\Sigma_{curr}} \leftarrow d\boldsymbol{\Sigma}(\boldsymbol{0}, \boldsymbol{\theta}, \Delta t)$

        $\boldsymbol{u}[j, i, i_{LNA}] \sim N(\boldsymbol{\eta_{curr}}, \boldsymbol{\Sigma_{curr}})$

      end for

      if Flag[j,i]==1 **then**

        $\boldsymbol{y}[j, i] \sim MvN(\boldsymbol{u}[j, i, N_{LNA}], \delta_{tech} * \boldsymbol{I})$

      else

        $-\boldsymbol{\tau} + \boldsymbol{y}[j, i] \sim MvN(\boldsymbol{u}[j, i, N_{LNA}], \delta_{tech} * \boldsymbol{I})$

      end if

    end if

  end for

end for

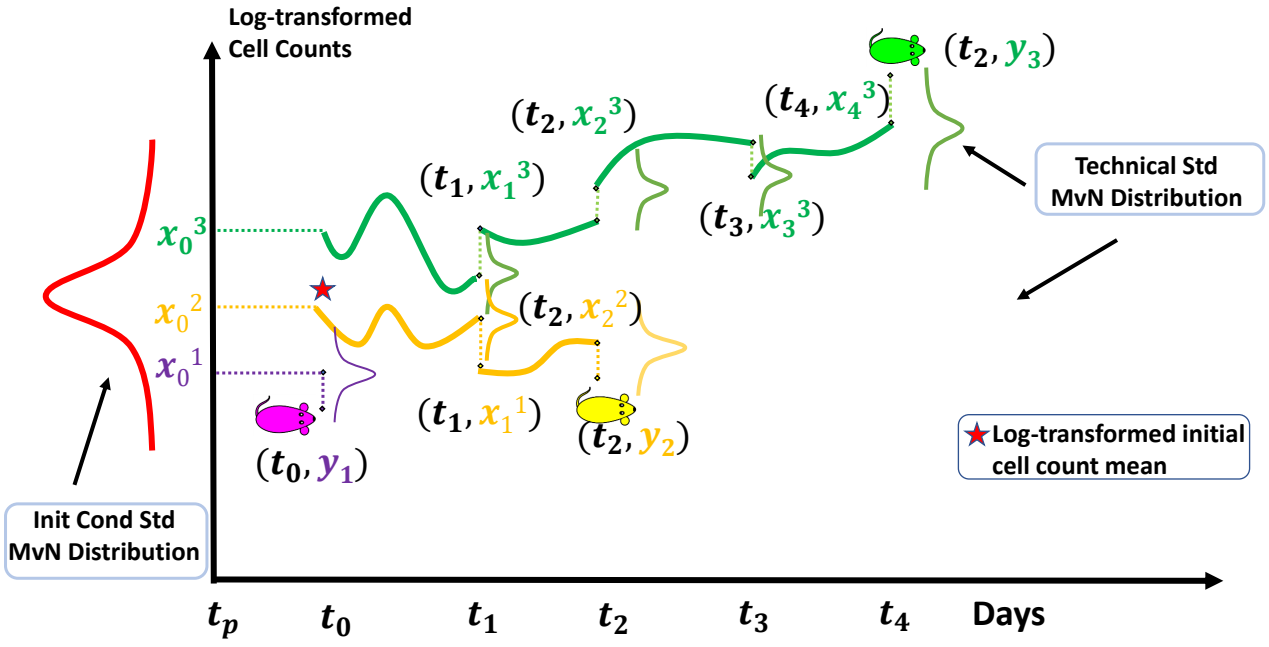


Figure 4.3: Data generation process diagram illustration of LNA model. Data generation process which accounts for both biological variation and technical variation via a latent variable approach; latent cell count trajectories are assumed to be LNA trajectories; mice are being observed at discrete irregular times, e.g., t_0, t_2, t_4; x_0^i refer to the initial cell count of mouse i at t_0 and the red star is the inferred log-scale initial condition cell count mean. y_i refers to the observed cell count of mouse i at the time it is sacrificed. x_j^i refers to the latent LNA intermediate step j cell count for mouse i.

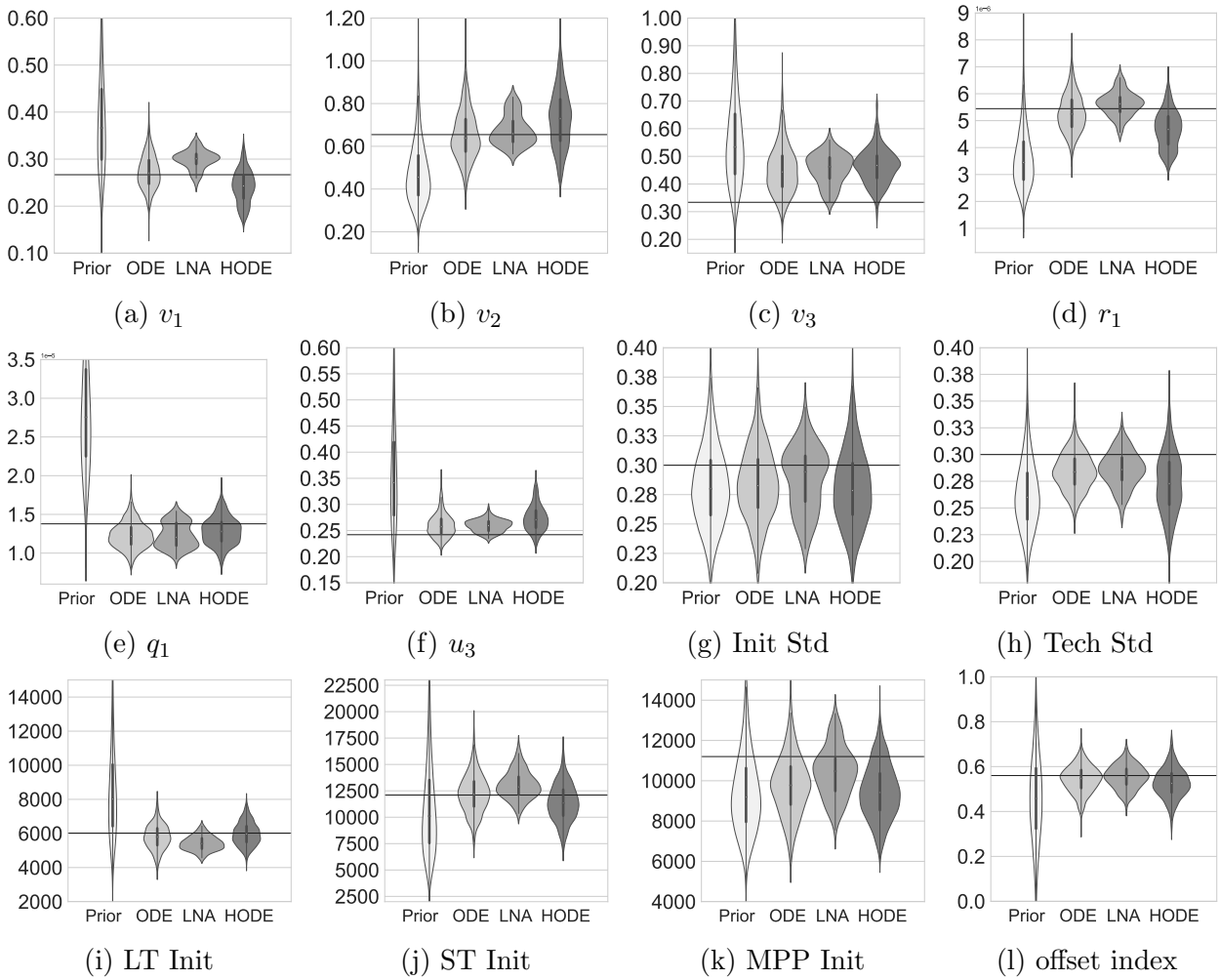
4.3 Results

4.3.1 Simulation Study on Synthetic Hematopoietic Dataset

To check the parameter identifiability for the new 3D LNA CR model described in section 4.2.2, we first fit the model to a synthetic dataset shown in Figure 4.1 (d-f) generated by the LNA algorithm which mimics the experimental dataset. The parameter values used for data generation are based on MAP point estimation optimization described in section 3.2.5. The synthetic dataset has the same observation design as the experimental datasets and has similar cell count trajectory magnitudes and time evolution oscillation dynamics. By fitting

non-hierarchical, hierarchical ODE models (NHODE, HODE models), described in section 3.2.4 and non-hierarchical LNA model, respectively, to this synthetic dataset using Bayesian inference, we find that the posterior distributions of parameter posteriors from all the three models are identifiable and the ground truth parameter values, which used for generating the synthetic dataset, are correctly recovered which can be seen in Figure 4.4 (a-l) Compared to NHODE and HODE fitted posterior posterior violin plots in Figure 4.4 (a-l), LNA posterior violin plots for parameters are generally narrower, which means that the LNA model inference is more certain about those parameter values. A comparison of dynamic uncertainty plots for the three models on synthetic dataset 1 based on 3000 iterations of MCMC using the NUTs algorithm is shown in Figure 4.5 and Appendix Figure B.1 for synthetic datasets 1 and 2 respectively. Specifically, the dotted point marks refer to the synthetic data; the shaded areas indicate the 95% credible intervals of the differential equation solutions at each time point; the trajectories are the median solutions of the differential equations; the line plots are the posterior predictive intervals at the data observation time. Take Figure 4.5, which describes the dynamic uncertainty plots for synthetic dataset 1 for example, the first and second columns show the ODE and LNA dynamic uncertainty plots respectively. We can see that the LNA model can generate wider 95% credible intervals of posterior predictive distribution based on differential equation solutions than that of ODE, even the parameter posteriors of LNA are generally narrower. The third column describes dynamic uncertainty plots for the HODE model. We can see that the shaded areas of the HODE model are much wider than that of ODE and LNA, while the line plots for the HODE model have similar lengths as that of ODE and LNA. This agrees with the posterior violin plots in Figure 4.4 (h) that the technical standard deviation posterior mean is much smaller than that of ODE and LNA. Besides, from Figure 4.4, we can see that the shaded areas of the HODE model near the steady states (day 30) are still wide which is not as expected while that of the LNA model is wide at early observation days but relatively narrow near the steady state as expected since it indicates the large biological variability among mice near the steady states.

By using the Bayesian model selection method PSIS-LOO-CV described in section 2.3.2, we found that the LNA model has the highest out-of-sample $elpd_{LOO}$= -54.17 among the three models followed by the HODE model.



(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) Init Std (h) Tech Std

(i) LT Init (j) ST Init (k) MPP Init (l) offset index

Figure 4.4: Prior and posterior violin plots for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the synthetic dataset which is generated by Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data. (a-l) prior and fitted posterior violinplots for (a) v_1: LTHSCs self-renewal rate; (b) v_2: LTHSCs asymmetric division rate; (c) v_3 STHSCs asymmetric division rate; (d) r_1: LTHSCs effective differentiation rate into STHSCs; (e) q_1 STHSCs effective differentiation rate into MPPs; (f) u_3 MPP lumped decay rate.(g) Init Std: initial condition cell count standard deviation; (h) Tech Std: technical measurement error standard deviation; (i) LT init: LTHSCs initial cell count mean; (j) ST init: STHSCs initial cell count mean; (k) MPP init: MPPs initial cell count means; (l) offset index: offset parameter value between the two synthetic datasets due to the batch effects

4.3.2 Inference Results on Experimental Hematopoietic Dataset

We next fit the NHODE, HODE, and LNA models to the experimental datasets. As we can see in the Bayesian model selection table 4.2, by using the metric PSIS-LOO $elpd_{LOO}$ described in section 2.3.2, we find that the LNA model achieves the highest out-of-sample leave-one-out expected log point-wise predictive density $elpd_{LOO}$= -66.20 among the three models followed by the HODE model. From the prior and fitted posterior violin plots in Figure 4.6 of the NHODE, LNA, and HODE models, we can see that the posterior violin plots for the NHODE and that of the LNA model are very similar. The HODE model has generally higher rate value posterior means and smaller technical standard deviation posterior means than those of the LNA and NHODE models. We next ask why the LNA and NHODE models yield similar parameter inference results. This motivates us to propose a hypothesis:

Hypothesis 4.1. *For Bayesian inference with differential equations using partially observed datasets, the stochastic process model classifies most of the data heterogeneity into the initial condition variation rather than in the stochasticity of solving time evolution stochastic*

PSIS-LOO Model Selection Table for Synthetic Dataset			
Metric / Model	LNA	NHODE	HODE
PSIS-LOO cv_elpd	**-54.17**	-60.08	-55.64
PSIS-LOO naïve_elpd	-36.60	-32.68	-9.52
PSIS-LOO loo_weight	**0.8112**	0.002	0.1865
Number of Parameters	1932	156	287

Table 4.1: Model selection table for hematopoiesis synthetic dataset. PSIS-LOO cv_epld: out-of-sample leave-one-out cross-validation expected log predictive density; PSIS-LOO naïve_lpd: in-sample leave-one-out expected log predictive density; PSIS-LOO loo_weight: model weight by PSIS-LOO; The most important metric we want to maximize is cv_epld. We can see that the chemical reaction (CR) LNA model achieves the highest value of cv_epld using synthetic datasets.

differential equations when both initial condition biological variation and technical variation are substantial. The prior-posterior violin plots and 2σ Bayesian credible intervals (BCIs) of the stochastic process model and NHODE model inference are distinct when the ground truth initial condition cell count variation is relatively small. The inference results become similar as the initial condition variation gets larger.

4.3.3 Parameter Inference using a Synthetic Bone Marrow Transplants Dataset

To test the hypothesis 4.1, we perform a series of simulation experiments. We take the initial condition, a system with much smaller cell counts, to increase the biological variation in solving the time evolution stochastic process LNA model. Let the log-scale initial means of the cell compartments $(\mu_{LT}, \mu_{ST}, \mu_{MPP})$ = (log(12), log(18), log(67)) respectively. This mimics a bone marrow transplantation where cells are introduced to bone marrow in which the cells are removed. Based on the similar inference results of the NHODE model and the LNA model, we want to explore the magnitude levels of initial condition variance and technical variance that can make LNA and NHODE inference results distinct from each other (especially the cases when all the ground truth parameter values are correctly identified by using LNA model while they are not fully correctly recovered by using ODE model). Firstly, instead of partially observed datasets, we want to validate that all the ground truth parameter values used for time-series synthetic data generation can be correctly recovered by the LNA model.

Simulation Study 4.1. *Generate a time series synthetic transplant dataset from the LNA model and show that all the parameters, especially the initial condition standard deviation term are identifiable and the ground truth values used for synthetic data generation are correctly recovered.*

We generate synthetic datasets using the LNA model Eqn.(4.10) with 4 fully observed mice and different initial condition standard deviation levels δ_{init}=[0.05, 0.1, 0.2, 0.3]. Then, we assume that mice are not sacrificed during the bone marrow extraction operation and are periodically observed on days [1, 3, 5, 7, 10, 15, 20, 24, 30]. As we can see in Figure 4.8, both the ground truth parameter values of initial conditional noise δ_{init} (a-d) and technical noise δ_{tech} (e-h) term posteriors are correctly recovered their ground truth values. From Figure 4.8 (i-l), we can see that the LTHSC posterior predictive intervals and 2σ BCIs describe the LTHSC synthetic data well and the 2σ BCIs grow wider as the ground truth initial conditional standard deviation increases. The full prior and posterior violin plots and dynamic uncertainty plot results can be found in the Appendix Figures (B.3-B.6).

Secondly, we want to validate that the LNA model tends to classify biological variability in data into initial condition variance of cell counts instead in the stochastic data generation process by LNA algorithm when initial condition variance can be substantial.

Simulation Study 4.2. *Generate partially observed synthetic transplant datasets from the LNA model, which only uses two observation days at the early times. Fixing all the parameters at their ground truth values except initial condition cell counts and standard deviation, we show that the total standard deviation term (initial condition standard deviation term + technical standard deviation term) is identifiable and the ground truth value can be correctly recovered when the observation number of the later day is small but the total standard deviation will be overestimated when the observation number of the later day getting larger.*

In this study, we generate partially observed synthetic data only on days 1 and 3 using the LNA model with synthetic experimental designs in the form [# of mice being observed at day 1, # of mice being observed at day 3] and the observation numbers are [48,1], [48,3], [48,5], [48,10], [48,20] respectively. During the Bayesian inference, all the rate value parameters are fixed to their ground truth values, and we only infer standard deviation terms and latent initial condition cell counts. From Figure 4.9 (a-e), we can see that as the mice observation

number on day 3 increases, both NHODE and LNA models tend to overestimate the total standard deviation level to compensate for the data variability on day 3, which stems from the stochastic nature of solving LNA differential equations.

Thirdly, we want to explore the relationship between LNA/NHODE inference results and the magnitudes of initial condition variance and technical measurement errors. Thus, we propose two types of simulation experiments for this exploration. One is to fix the initial condition cell count variance at the ground truth value and infer all other parameters.

Simulation Study 4.3. *Generate partially observed synthetic transplant datasets from the LNA model with different initial condition standard deviation and technical standard deviation levels. Fix the initial condition standard deviation level at the ground truth value during the Bayesian inference and compare all other parameter value posterior distributions and dynamic uncertainty plots between fitted LNA and NHODE models.*

In this study, we generate partially observed transplantation synthetic datasets with different log-scale initial conditional standard deviation level δ_{init}=[0.0001, 0.001, 0.01, 0.02, 0.05 ,0.1 ,0.2 ,0.3]. We fix the initial condition standard deviation term to its ground truth value and infer all other parameters. As we can see in Figure 4.10 (a-d), when δ_{init} is small, then both ODE and LNA tend to overestimate the technical standard deviation δ_{tech} level to account for the data variability from the LNA data generation process. As the fixed ground truth value of δ_{init} increases, the posterior means of δ_{tech} for both NHODE and LNA move closer to their ground truth values, and the ground truth values can be correctly recovered when δ_{init} is large enough e.g., δ_{init}=0.2 or 0.3. The NHODE and LNA dynamic uncertainty plots for the LTHSCs can be seen in the (e-h), and (i-l), respectively. We can see that the 2σ BCIs of the NHODE model get closer to that of the LNA model when δ_{init} increases. In general, the NHODE and LNA inference results become similar as δ_{init} increases. The full posterior violin plot and dynamic uncertainty plot can be seen in the Appendix Figures (B.7-B.10).

In Appendix Figure B.11, we explore the technical error term δ_{tech} using the NHODE and the LNA models when the ground truth values of δ_{init}^{real} and δ_{tech}^{real} are relatively large. We find that the larger the δ_{init}^{real} and δ_{tech}^{real}, the better δ_{tech} can be correctly identified. The δ_{tech} posterior mean of LNA is generally closer to the δ_{tech}^{real} than that of the NHODE inference. In Figure Appendix B.12, we explore the technical error term δ_{tech} using the NHODE and LNA models when the ground truth values of δ_{init}^{real} and δ_{tech}^{real} are relatively small. The results are similar to the case when δ_{init}^{real} is relatively large. Then we further decrease the δ_{init}. Figure 4.11 shows the case when δ_{init} = 1e-4, δ_{tech}= 0.1. In Figure 4.11, (a-j), we can observe that LNA inference has narrower posteriors than that of the NHODE model inference and (g) LNA inference can correctly identify the technical standard deviation value while ODE inference using the NHODE model overestimates it.

Fourthly, we do the simulation experiment in another way using the synthetic data generated in simulation study 4.3 by fixing the technical variance at the ground truth value and inferring all other parameters.

Simulation Study 4.4. *Use the same synthetic data generated from simulation study 4.3. Fix the technical standard deviation term at its ground truth value during the Bayesian inference and compare the parameter value posterior distributions and dynamic uncertainty plots between fitted LNA and ODE models.*

In this study, we use the same partially observed synthetic dataset generated in simulation study 4.3. Then, we fix the technical standard deviation term δ_{tech} to its ground truth value and infer all the other parameters. As we can see in Figure 4.12 (a-d), both the LNA and the NHODE inference overestimate the initial condition noise level δ_{init}. The posterior distribution mean of δ_{init} moves closer to the ground truth values as δ_{init} increases. Figures 4.12 (e-h), (i-l) show the NHODE and LNA dynamic uncertainty plots for the LTHSC dataset, respectively. We observe that NHODE 2σ BCIs are closer to that of LNA when δ_{init} increases. The full posterior violin plot and posterior predictive plot results can be seen

in Appendix Figures (B.13-B.16). In Appendix Figure B.17, we explore the initial condition standard deviation term δ_{init} using ODE and LNA inference when the ground truth values of δ_{init}^{real} and δ_{tech}^{real} are in relatively large scales which are respectively selected in [0.05, 0.1, 0.2, 0.3] and [0.1, 0.2, 0.3]. We find that the δ_{init} posterior means of LNA are generally closer to δ_{init}^{real} than those from the NHODE inference. When δ_{tech}^{real} is large enough and δ_{init}^{real} is small enough, both ODE and LNA tend to correctly identify δ_{init}^{real}. The best δ_{init} posterior for both the NHODE and LNA model inferences is achieved in Figure B.17 (i) when δ_{init}^{real} has a relatively small value (e.g., 0.05) and δ_{tech}^{real} has relatively large value (e.g., 0.3). In figure Appendix B.18, we explore the initial condition standard deviation term δ_{init} using ODE and LNA inference when the ground truth values of δ_{init}^{real} and δ_{tech}^{real} are in relatively small scales which are respectively selected in [0.001, 0.005, 0.01, 0.02] and [0.001, 0.01, 0.1, 0.3].

We observe that when δ_{tech}^{real} is 0.001 and 0.01, the LNA inference tends to classify all the data heterogeneity into initial condition variance. When δ_{tech}^{real} increases to 0.1 and 0.3, then the LNA inference is more likely to correctly identify δ_{init} while the NHODE inference overestimates it. We show the prior posterior violin plots and dynamic uncertainty plots for selected case (m) in figure 4.13 when $\delta_{init}^{real} = 0.02$ and $\delta_{tech}^{real} = 0.3$. We can see that most of the LNA inference posteriors can correctly identify the ground truth values for the parameters of interest, while some of the posteriors of the NHODE model inference do not recover the correct value (b,e,f,g). The dynamic uncertainty plots for LTHSCs for fitted NHODE and LNA models are shown in (k,l) respectively. We can observe the LNA 2σ BCIs are much wider than that of ODE while their posterior interval widths are similar.

Fifthly, we want to validate a potential solution for the issue described in hypothesis 4.1 that LNA classifies data heterogeneity into initial condition variation instead of in the stochastic time evolution process when initial condition variation can be substantial.

Simulation Study 4.5. *Generate partially observed synthetic transplant dataset from the LNA model with different initial condition standard deviation levels. Fix both the initial*

condition and technical standard deviation terms at ground truth values during the Bayesian inference and compare the parameter value posterior distributions and dynamic uncertainty plots between fitted LNA and NHODE models. This is motivated by the fact that the technical standard deviation term can be directly estimated if there is enough vehicle data, while the total standard deviation term can be directly estimated if there is enough data observation at day 1. We can then estimate the initial condition standard deviation based on estimated technical and total standard deviation terms.

Based on the fact that technical noise and initial condition noise can be directly estimated when there are enough vehicles and day 1 observation, we perform the Bayesian inference with initial condition noise and technical noise both fixed to the ground truth values during the Bayesian LNA inference with partially observed synthetic datasets. As we can see from the prior posterior violin plots in Figure 4.14 (a-i), all the parameters of interest are identifiable and the ground truth values are recovered correctly.

Those simulation results support our hypothesis 4.1 about the relationship between the level of initial condition standard deviation δ_{init} and technical standard deviation δ_{tech}, and differences between inference results using the LNA and the NHODE models. We find when the biological variation in the initial condition is relatively small and technical standard deviation is relatively large, then the ground truth values can be correctly recovered by the LNA model while that cannot be fully correctly recovered by ODE inference. However, when the initial condition standard deviation is substantial, the LNA model inference and ODE inference results are similar.

4.3.4 Inference Results on Experimental Hematopoietic Dataset with Fixed Initial Condition and Technical Variance

Motived by simulation study 4.5, since the cell count measurement of vehicles (which is also treated as data at day 30) is only affected by technical variation source, we can estimate from eight vehicles that $\delta_{tech} \approx 0.35$ in for log-scale cell counts. On the first observation day, the cell count measurement is only affected by the initial condition cell count and technical variation without any biological variation. We can estimate the sum standard deviation $\sqrt{\delta_{init}^2 + \delta_{tech}^2}$. The initial condition standard deviation can then be estimated by those two terms, and we find that $\delta_{init} \approx 0.3$. We then perform the Bayesian inference for the experimental dataset with fixed initial condition standard deviation and technical measurement standard deviation. The prior posterior violin plot and dynamic uncertainty plot results in Figure 4.15: (a-j) show that LNA inference has narrower posteriors than those of the NHODE and HODE models, which indicates that LNA is more certain about those parameters. (j,k) show the ODE and LNA dynamic uncertainty plots respectively. We can see that the LNA model can generate wider 95% credible intervals of posterior predictive distribution based on differential equation solutions than that of ODE, even the parameter posteriors of LNA are generally narrower. (l) describes dynamic uncertainty plots for the

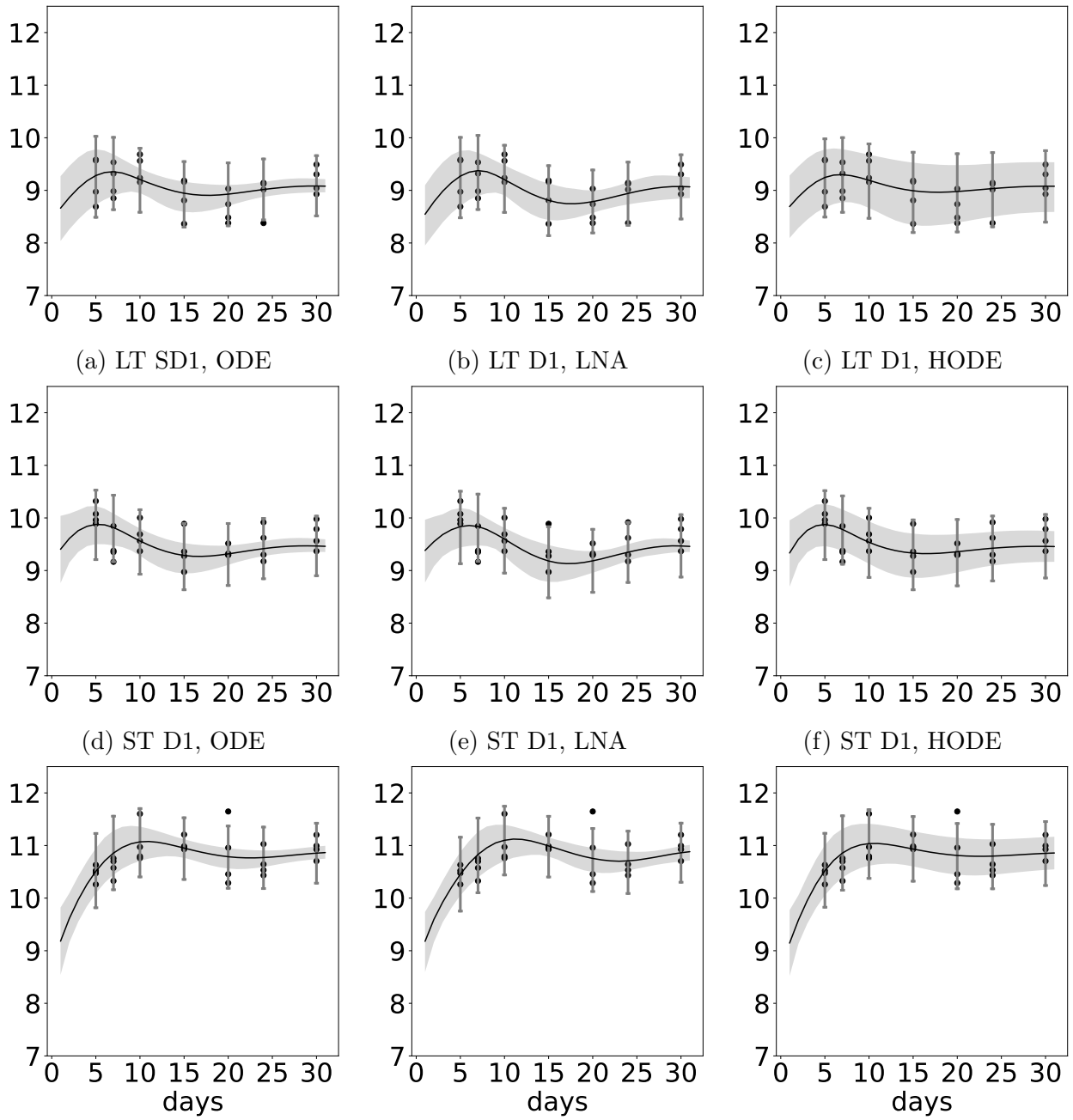
PSIS-LOO Model Selection Table for Experimental Dataset			
Metric / Model	LNA	NHODE	HODE
PSIS-LOO cv_elpd	**-66.20**	-68.22	-66.63
PSIS-LOO naïve_elpd	-40.90	-42.22	-33.88
PSIS-LOO loo_weight	**0.5608**	0.0743	0.3648
Number of Parameters	1932	156	482

Table 4.2: Model selection table for hematopoiesis experimental dataset. PSIS-LOO cv_epld: out-of-sample leave-one-out cross-validation expected log predictive density; PSIS-LOO naïve_lpd: in-sample leave-one-out expected log predictive density; PSIS-LOO loo_weight: model weight by PSIS-LOO; The most important metric we want to maximize is cv_epld. We can see that the chemical reaction (CR) LNA model achieves the highest value of cv_epld using the experimental dataset.

HODE model. We can see that the shaded areas of the HODE model are much wider than that of ODE and LNA. Besides, from Figure 4.15, we can see that the shaded areas of the HODE model near the steady states (day 30) are still wide, which is not as expected, while that of the LNA model is wide at early observation days but relatively narrow near the steady state as expected since it indicates the large biological variability among mice near the steady states. By using the Bayesian model selection method PSIS-LOO-CV described in section 2.3.2, we found that the LNA model has the highest out-of-sample $elpd_{LOO}$= -62.84 among the three models followed by the HODE model in table 4.3.

Model Selection for Experimental Dataset with std terms Fixed			
Metric / Model	LNA	NHODE	HODE
PSIS-LOO cv_elpd	**-62.84**	-67.83	-66.18
PSIS-LOO naïve_elpd	-46.47	-44.76	-33.86
PSIS-LOO loo_weight	**0.9594**	0.006	0.034
Number of Parameters	1932	156	482

Table 4.3: Model selection table for hematopoiesis 3D real dataset when both initial condition and technical measurement standard deviation terms fixed. PSIS-LOO cv_epld: out-of-sample leave-one-out cross-validation expected log predictive density; PSIS-LOO naïve_lpd: in-sample leave-one-out expected log predictive density; PSIS-LOO loo_weight: model weight by PSIS-LOO; The most important metric we want to maximize is cv_epld. We can see that the chemical reaction (CR) LNA model achieves the highest value of cv_epld using the experimental dataset.

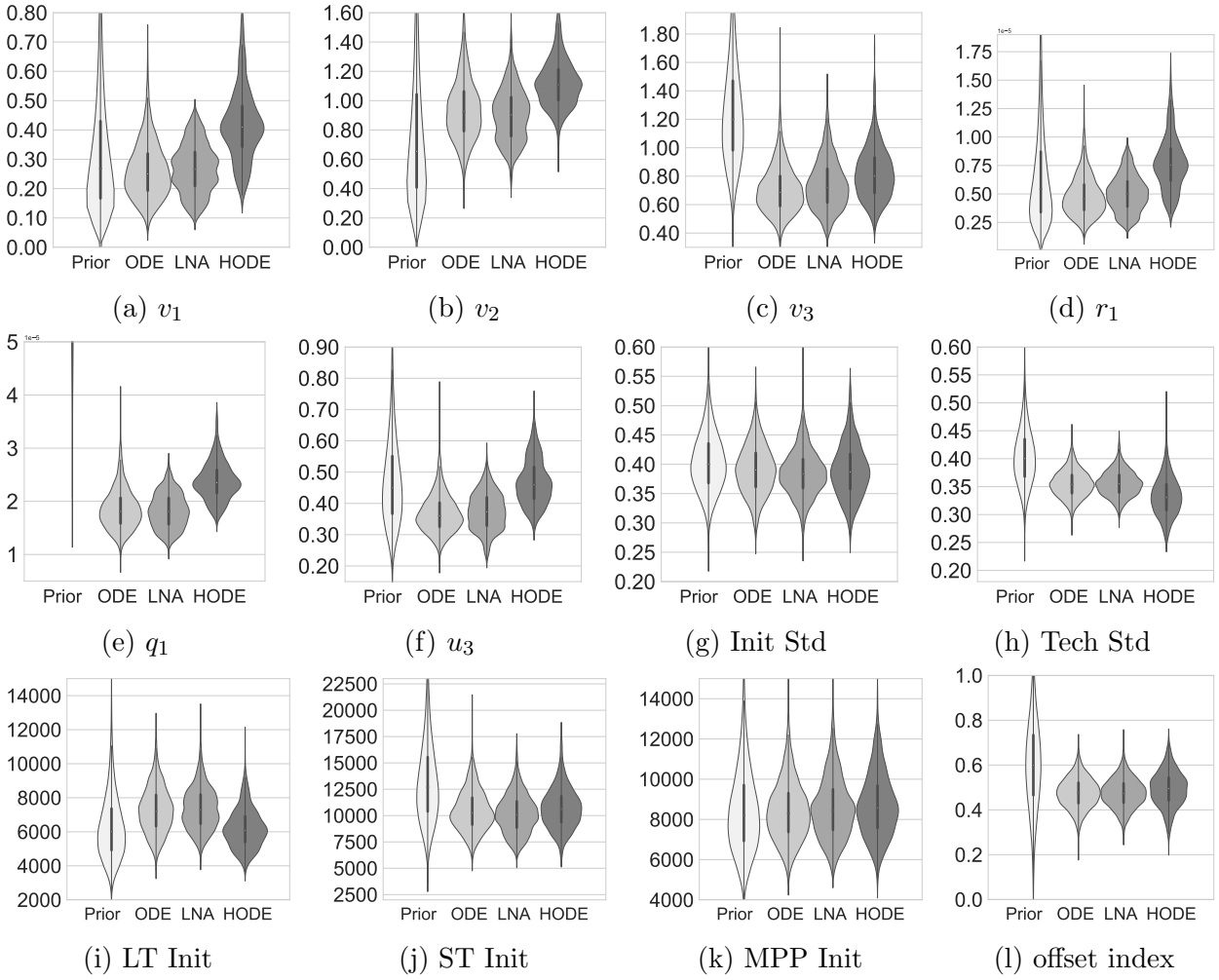


(a) LT SD1, ODE (b) LT D1, LNA (c) LT D1, HODE

(d) ST D1, ODE (e) ST D1, LNA (f) ST D1, HODE

(g) MP D1, ODE (h) MPP D1, LNA (i) MPP D1, HODE

Figure 4.5: Dynamic uncertainty plots for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the synthetic dataset which is generated by Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data (ExpD). LT: LTHSCs; ST: STHSCs; In dynamic uncertainty plots (a-i), the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-c) plots for LT data in SD1; (d-f) plots for ST data in SD1; (g-i) plots for MPP data in SD1

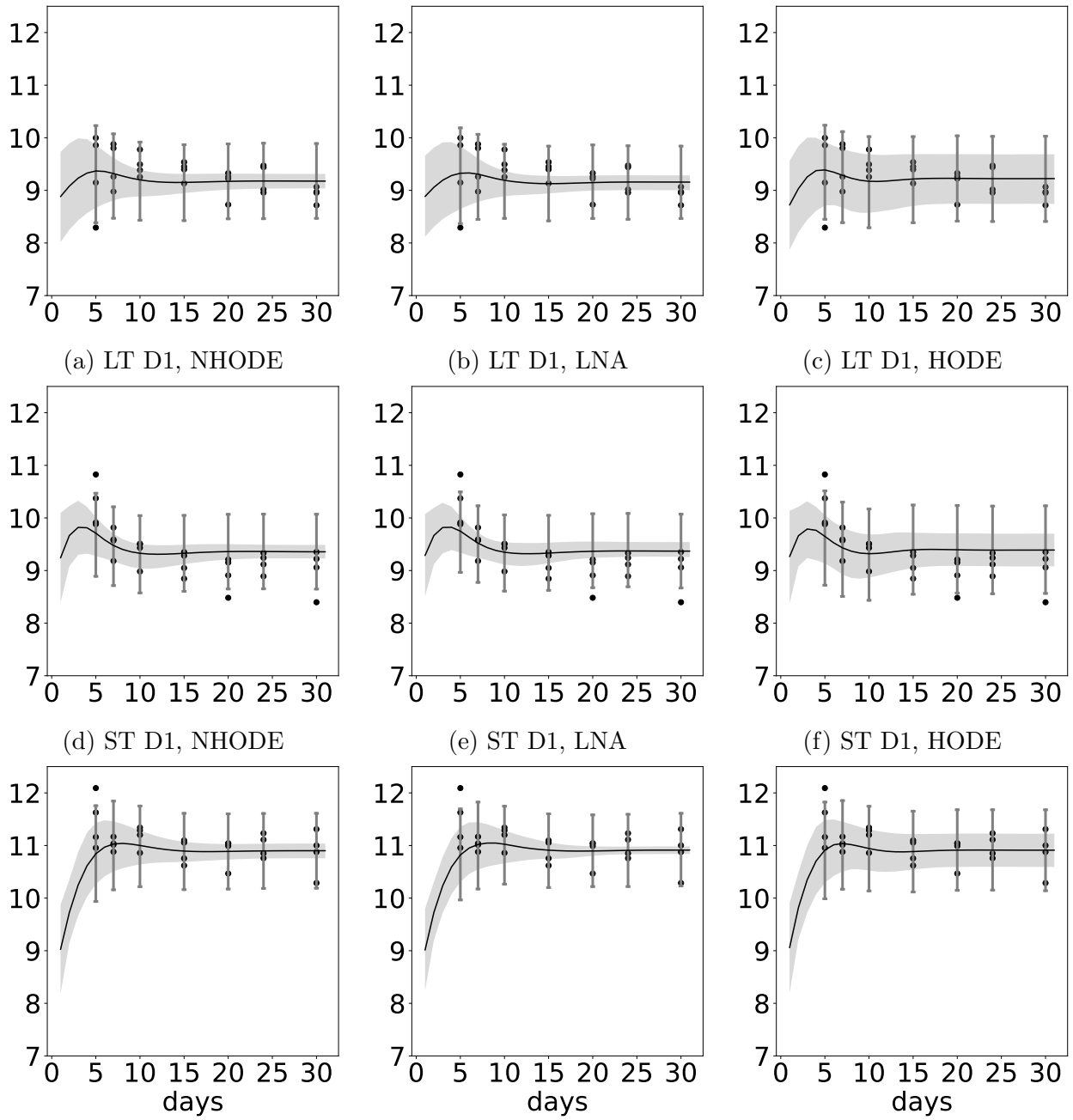


(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) Init Std (h) Tech Std

(i) LT Init (j) ST Init (k) MPP Init (l) offset index

Figure 4.6: Prior and posterior violin plots for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the experimental dataset. (a-l) prior and fitted posterior violinplots for (a) v_1: LTHSCs self-renewal rate; (b) v_2: LTHSCs asymmetric division rate; (c) v_3 STHSCs asymmetric division rate; (d) r_1: LTHSCs effective differentiation rate into STHSCs; (e) q_1 STHSCs effective differentiation rate into MPPs; (f) u_3 MPP lumped decay rate.(g) Init Std: initial condition cell count standard deviation; (h) Tech Std: technical measurement error standard deviation; (i) LT init: LTHSCs initial cell count mean; (j) ST init: STHSCs initial cell count mean; (k) MPP init: MPPs initial cell count means; (l) offset index: offset parameter value between the two experiment datasets due to the batch effects

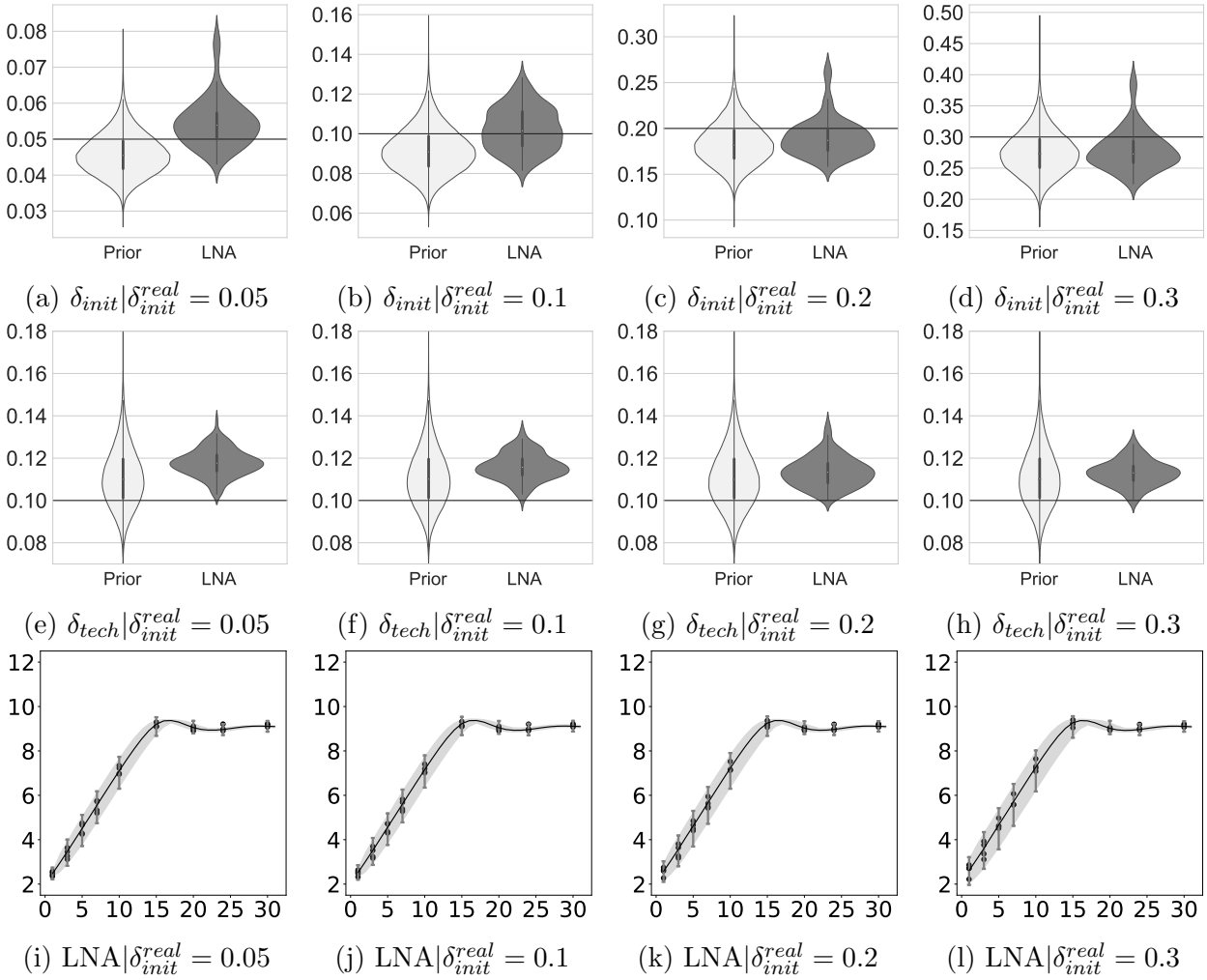


(a) LT D1, NHODE (b) LT D1, LNA (c) LT D1, HODE

(d) ST D1, NHODE (e) ST D1, LNA (f) ST D1, HODE

(g) MP D1, NHODE (h) MPP D1, LNA (i) MPP D1, HODE

Figure 4.7: Dynamic uncertainty plots for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the experimental data (ExpD). LT: LTHSCs; ST: STHSCs; In dynamic uncertainty plots (a-i), the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-c) plots for LT data in ExpD1; (d-f) plots for ST data in ExpD1; (g-i) plots for MPP data in synthetic ExpD1

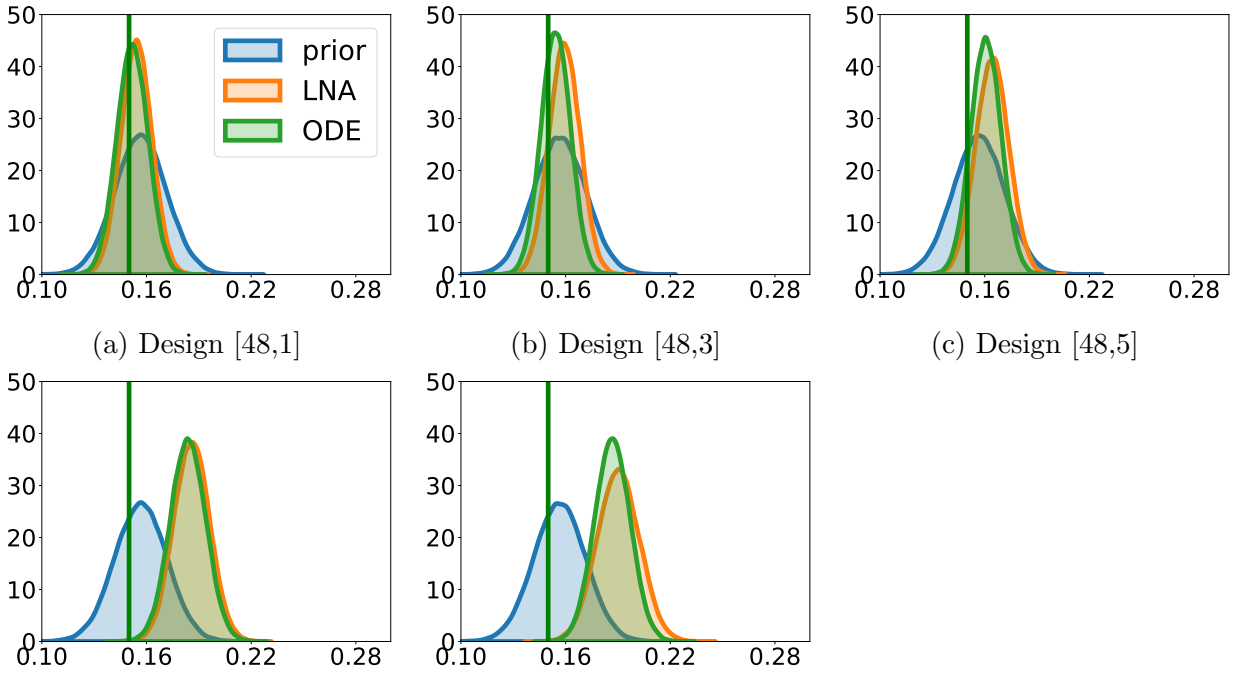


(a) $\delta_{init}|\delta_{init}^{real} = 0.05$ (b) $\delta_{init}|\delta_{init}^{real} = 0.1$ (c) $\delta_{init}|\delta_{init}^{real} = 0.2$ (d) $\delta_{init}|\delta_{init}^{real} = 0.3$

(e) $\delta_{tech}|\delta_{init}^{real} = 0.05$ (f) $\delta_{tech}|\delta_{init}^{real} = 0.1$ (g) $\delta_{tech}|\delta_{init}^{real} = 0.2$ (h) $\delta_{tech}|\delta_{init}^{real} = 0.3$

(i) LNA$|\delta_{init}^{real} = 0.05$ (j) LNA$|\delta_{init}^{real} = 0.1$ (k) LNA$|\delta_{init}^{real} = 0.2$ (l) LNA$|\delta_{init}^{real} = 0.3$

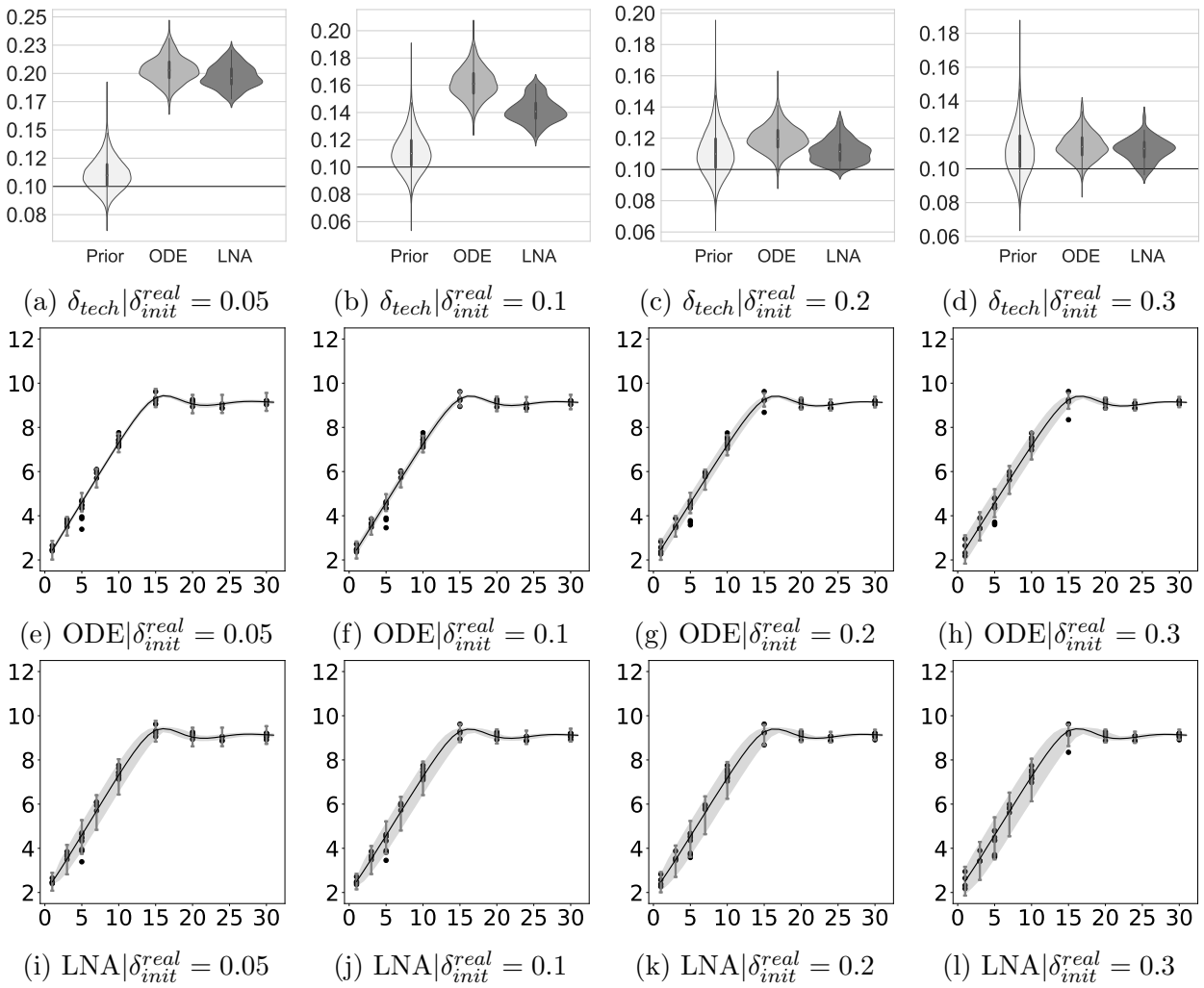
Figure 4.8: Prior and posterior violin plots, dynamic uncertainty plots for simulation study 4.1 using time-series synthetic datasets. δ_{init}: initial condition standard deviation; δ_{tech}: technical measurement standard deviation; δ_{init}^{real}: ground truth values of δ_{init}. (a-d) prior and posterior violin plots for δ_{init} for synthetic dataset generated by different values of δ_{init}^{real} (e-h) prior and posterior violin plots for δ_{tech} for synthetic dataset generated by different values of δ_{init}^{real} (i-l) dynamic uncertainty plots for LTHSCs data fitted by LNA model where the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day



(a) Design [48,1] (b) Design [48,3] (c) Design [48,5]

(d) Design [48,10] (e) Design [48,20]

Figure 4.9: Prior and posterior distributions of the initial condition and technical measurement standard deviation sum for simulation study 4.2 using partially observed synthetic dataset. Blue areas: prior distributions of sum standard deviation; orange areas: fitted LNA posterior distributions of total standard deviation; green areas: fitted ODE posterior distributions of total standard deviation; green line: ground truth values of total standard deviation. The dataset only includes day 1 and day 3 observation after the perturbation exam; the synthetic mice data used on day 1 and 3 are respectively (a) [48,1], (b) [48,3], (c) [48,5], (d) [48,10], (e) [48,20]

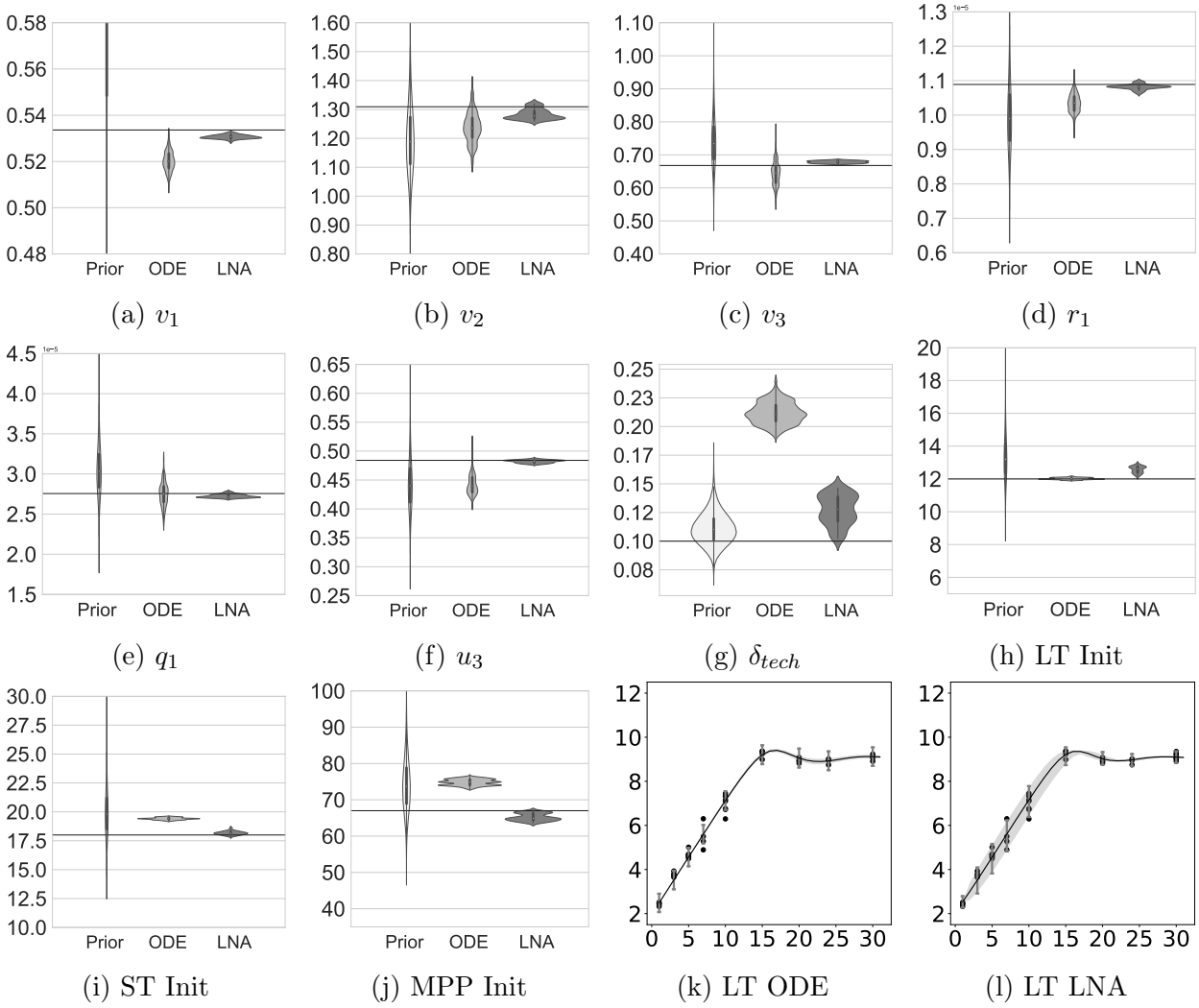


(a) $\delta_{tech}|\delta_{init}^{real} = 0.05$ (b) $\delta_{tech}|\delta_{init}^{real} = 0.1$ (c) $\delta_{tech}|\delta_{init}^{real} = 0.2$ (d) $\delta_{tech}|\delta_{init}^{real} = 0.3$

(e) ODE$|\delta_{init}^{real} = 0.05$ (f) ODE$|\delta_{init}^{real} = 0.1$ (g) ODE$|\delta_{init}^{real} = 0.2$ (h) ODE$|\delta_{init}^{real} = 0.3$

(i) LNA$|\delta_{init}^{real} = 0.05$ (j) LNA$|\delta_{init}^{real} = 0.1$ (k) LNA$|\delta_{init}^{real} = 0.2$ (l) LNA$|\delta_{init}^{real} = 0.3$

Figure 4.10: Prior and posterior violin plots, dynamic uncertainty plots for simulation study 4.3 where initial condition standard deviation are fixed at ground truth value and infer all other parameters during the Bayesian inference. δ_{init}: initial condition standard deviation; δ_{tech}: technical measurement standard deviation; δ_{init}^{real}: ground truth values of δ_{init}. (a-d) prior and posterior violin plots for technical measurement standard deviation δ_{tech}. Dynamic uncertainty plots (e-h) for LTHSCs data fitted by the ODE model. (i-l) for LTHSCs data fitted by the LNA model. The grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day

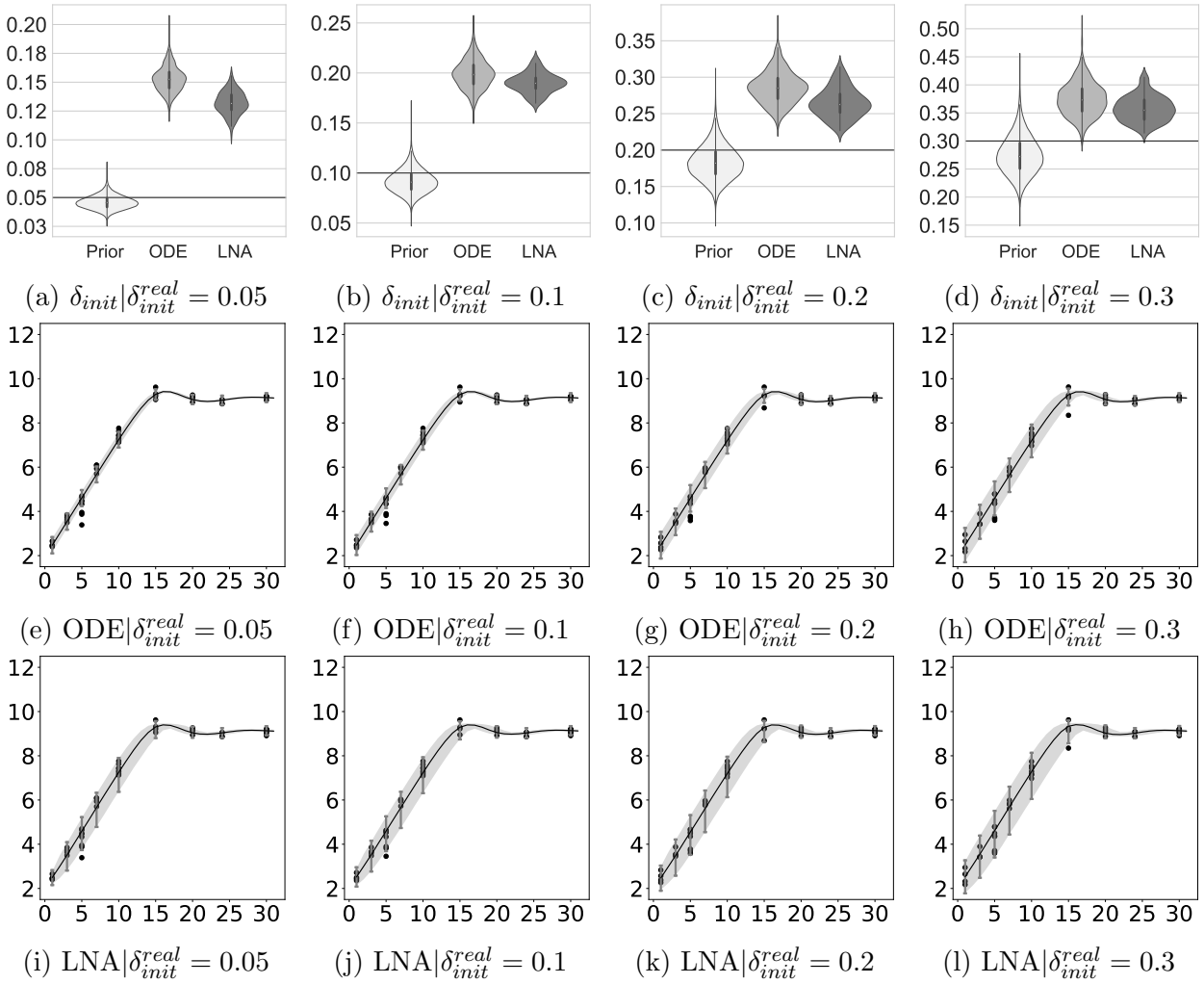


(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) δ_{tech} (h) LT Init

(i) ST Init (j) MPP Init (k) LT ODE (l) LT LNA

Figure 4.11: Prior and posterior violin plots, dynamic uncertainty plots for the selected case when δ_{init}^{real} = 1e-4, δ_{tech}^{real}=0.1 and the ground truth parameter values can be correctly recovered by LNA model while cannot be fully correctly recovered by ODE model for simulation study 4.3 fix initial condition standard deviation at ground truth value and infer all other parameters. (a-j) prior and fitted posterior violin plots for (a) v_1: LTHSCs self-renewal rate; (b) v_2: LTHSCs asymmetric division rate; (c) v_3 STHSCs asymmetric division rate; (d) r_1: LTHSCs effective differentiation rate into STHSCs; (e) q_1 STHSCs effective differentiation rate into MPPs; (f) u_3 MPP lumped decay rate.(g) Tech Std: technical measurement error standard deviation; (h) LT init: LTHSCs initial cell count mean; (i) ST init: STHSCs initial cell count mean; (j) MPP init: MPPs initial cell count means; (k,l) dynamic uncertainty plots for LTHSCs data fitted by ODE and LNA. The grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day.

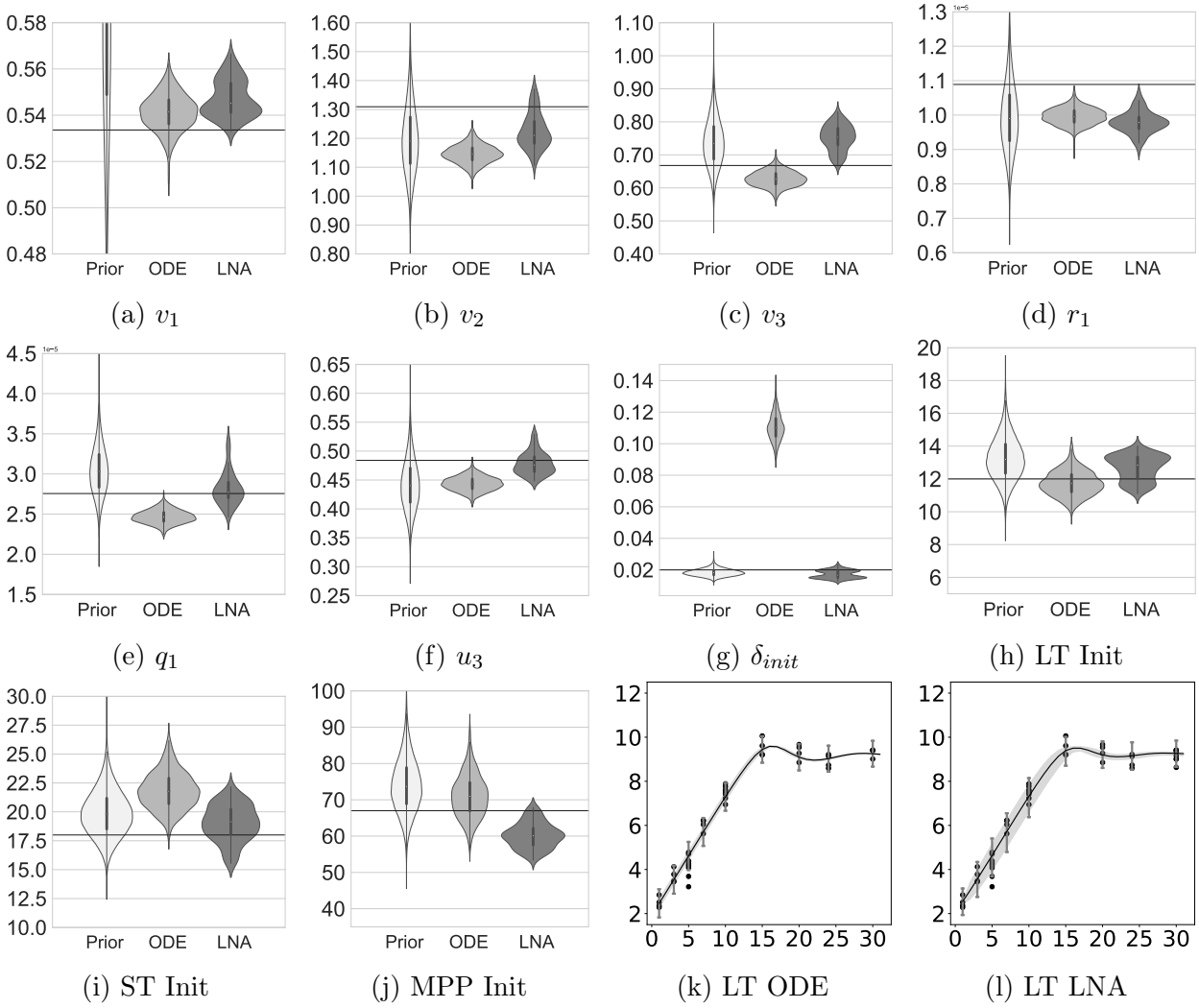


(a) $\delta_{init}|\delta_{init}^{real} = 0.05$ (b) $\delta_{init}|\delta_{init}^{real} = 0.1$ (c) $\delta_{init}|\delta_{init}^{real} = 0.2$ (d) $\delta_{init}|\delta_{init}^{real} = 0.3$

(e) ODE$|\delta_{init}^{real} = 0.05$ (f) ODE$|\delta_{init}^{real} = 0.1$ (g) ODE$|\delta_{init}^{real} = 0.2$ (h) ODE$|\delta_{init}^{real} = 0.3$

(i) LNA$|\delta_{init}^{real} = 0.05$ (j) LNA$|\delta_{init}^{real} = 0.1$ (k) LNA$|\delta_{init}^{real} = 0.2$ (l) LNA$|\delta_{init}^{real} = 0.3$

Figure 4.12: Prior and posterior violin plots, dynamic uncertainty plots for simulation study 4.4 where technical measurement standard deviation is fixed at ground truth value and infer all other parameters during the Bayesian inference. δ_{init}: initial condition standard deviation; δ_{tech}: technical measurement standard deviation; δ_{init}^{real}: ground truth values of δ_{init}. (a-d) prior and posterior violin plots for technical measurement standard deviation δ_{tech}. Dynamic uncertainty plots (e-h) for LTHSCs data fitted by the ODE model. (i-l) for LTHSCs data fitted by the LNA model. The grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day

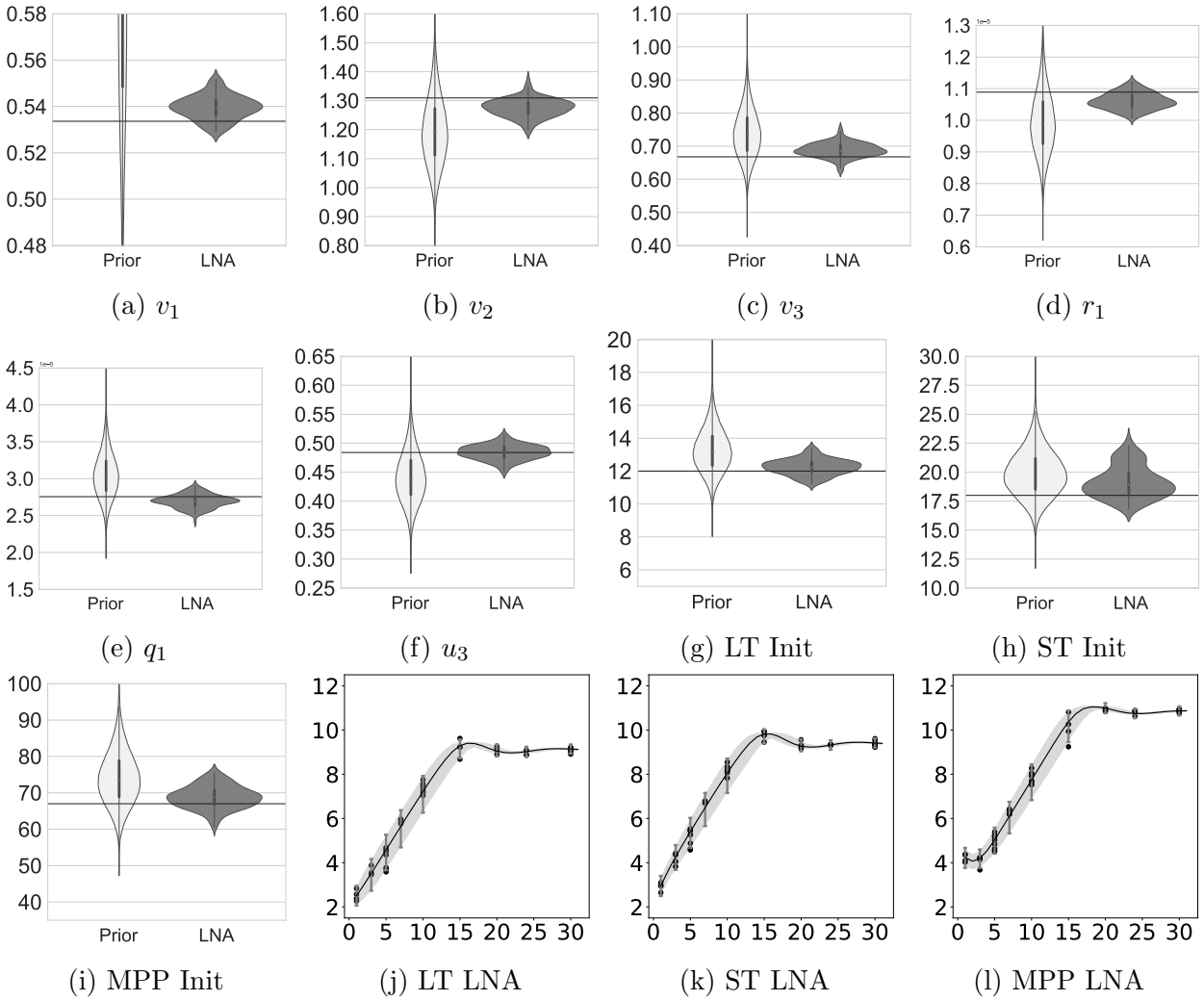


(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) δ_{init} (h) LT Init

(i) ST Init (j) MPP Init (k) LT ODE (l) LT LNA

Figure 4.13: Prior and posterior violin plots, dynamic uncertainty plots for the selected case when $\delta_{init}^{real} = 0.02$, δ_{tech}^{real}=0.3 and the ground truth parameter values can be correctly recovered by LNA model while cannot be fully correctly recovered by ODE model for simulation study 4.4 fix technical measurement standard deviation at ground truth value and infer all other parameters. (a-j) prior and fitted posterior violin plots for (a) v_1: LTHSCs self-renewal rate; (b) v_2: LTHSCs asymmetric division rate; (c) v_3 STHSCs asymmetric division rate; (d) r_1: LTHSCs effective differentiation rate into STHSCs; (e) q_1 STHSCs effective differentiation rate into MPPs; (f) u_3 MPP lumped decay rate.(g) Tech Std: technical measurement error standard deviation; (h) LT init: LTHSCs initial cell count mean; (i) ST init: STHSCs initial cell count mean; (j) MPP init: MPPs initial cell count means; (k,l) dynamic uncertainty plots for LTHSCs data fitted by ODE and LNA. The grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day.

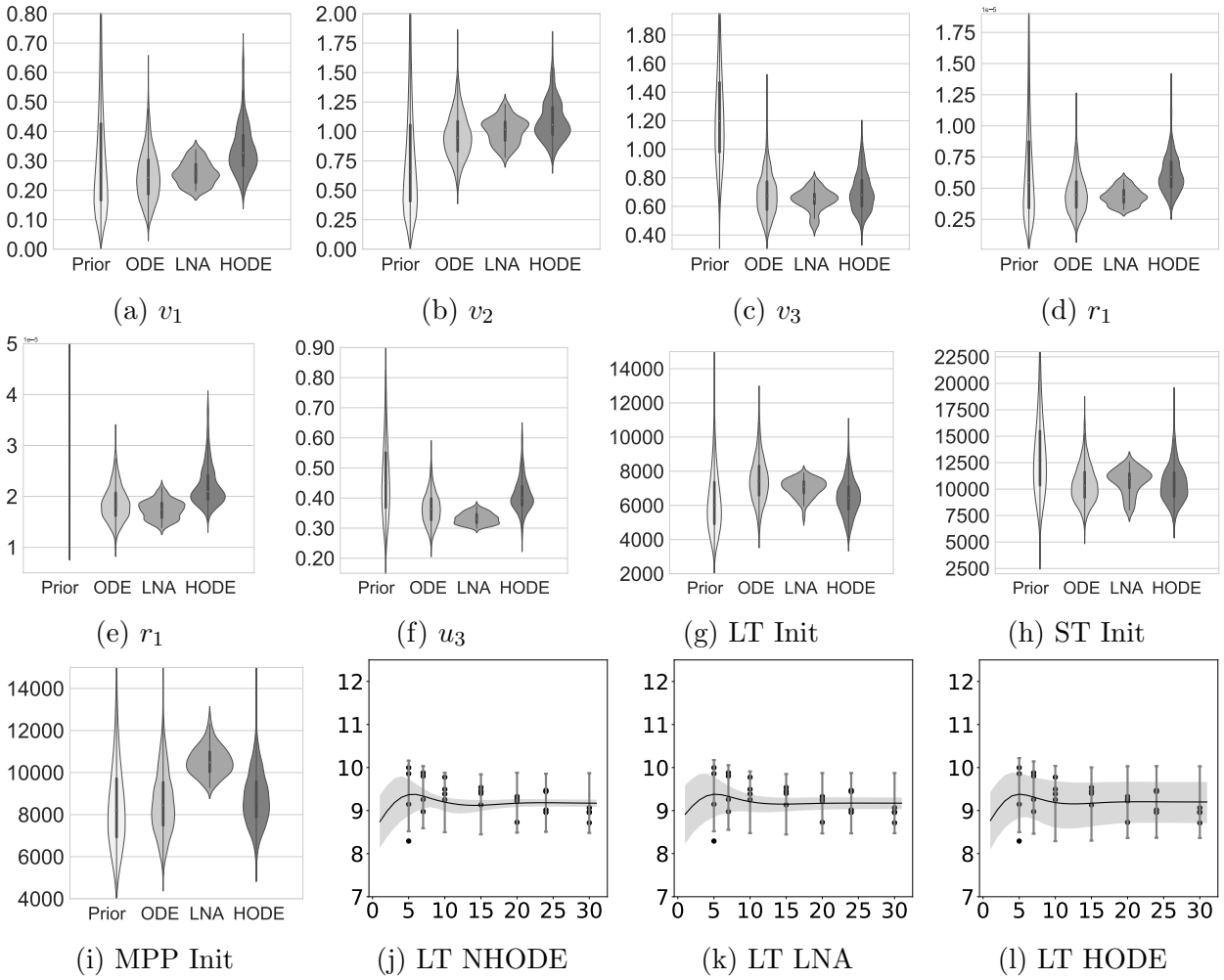


(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) LT Init (h) ST Init

(i) MPP Init (j) LT LNA (k) ST LNA (l) MPP LNA

Figure 4.14: Prior and posterior violin plots, dynamic uncertainty plots when both initial condition standard deviation and technical measurement standard deviation terms are fixed at ground truth values for simulation study 4.5 during the Bayesian inference. (a-i) Violin plots for priors and inferred posteriors parameter distributions. Init condition standard deviation and technical measurement standard deviation are both fixed to ground truth value during the inference time. v_1: LTHSCs self-renewal rate; v_2: asymmetric division rate; r_1: LTHSCs directly differentiate into STHSCs rate. q_1: STHSCs directly differentiate into MPP with rate constant q_1. u_3: MPP lumped decay rate. (a-l) dynamic uncertainty plots for LTHSC, STHSC, and MPP. the grey shaded areas refers to the 95% credible intervals of time evolution differential equation solutions; black dots refer to partially observed data, each dot related to a certain mouse; black trajectory refers to median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day;



(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) r_1 (f) u_3 (g) LT Init (h) ST Init

(i) MPP Init (j) LT NHODE (k) LT LNA (l) LT HODE

Figure 4.15: Prior and posterior violin plots, dynamic uncertainty plots for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the experimental datasets with initial condition standard deviation and technical measurement standard deviation fixed at data estimated level. (a-j) prior and fitted posterior violin plots for (a) v_1: LTHSCs self-renewal rate; (b) v_2: LTHSCs asymmetric division rate; (c) v_3 STHSCs asymmetric division rate; (d) r_1: LTHSCs effective differentiation rate into STHSCs; (e) q_1 STHSCs effective differentiation rate into MPPs; (f) u_3 MPP lumped decay rate.(g) LT init: LTHSCs initial cell count mean; (h) ST init: STHSCs initial cell count mean; (i) MPP init: MPPs initial cell count means; (j,k,l) dynamic uncertainty plots for LTHSCs data fitted by NHODE, HODE and LNA models. The grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the observation day.

4.4 Conclusion

In this project, we consider an alternative approach to account for data heterogeneity by using a stochastic process model under a non-hierarchical Bayesian framework to interpolate the latent cell count time evolution data generation process instead of letting the rate parameter values vary among mice. The computational tractability of Bayesian inference with a stochastic process model is achieved by using LNA which can be regarded as a compromise between the deterministic ODE model and intractable SDE model. Since the HSCs self-renewal rate for the best-fitted 2-compartment model to the experimental dataset is a bit higher than expected and also the larger the counts for each cell compartment, the similar the LNA forward simulation results to the ODE forward simulation results. We extended our model to 3D by decomposing HSCs into LTHSCs and STHSCs then derived the corresponding LNA formula for this model in the log scale. After evaluating models on a synthetic dataset is generated by the Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data. using PSIS-LOO-CV, we found that the Bayesian LNA model can further improve the out-of-sample expected log predictive density than the hierarchical model we previously developed. We found that Bayesian LNA inference can generally lead to narrower posterior distributions which means that it can be more certain to correctly identify the ground truth parameter values. By fitting 3-compartment NHODE, LNA and HODE models to the experimental dataset, we find a limitation that fitted posterior distributions for parameters of interest and dynamic uncertainty posterior predictive plots between Bayesian ODE inference and Bayesian LNA inference are similar since the inferred initial condition variance is substantial. We then propose a hypothesis that the Bayesian inference results of LNA and ODE models are distinct when ground truth initial condition cell count variation is relatively small and the inference results become similar as the initial condition variation gets larger. To validate this hypothesis, we experiment on a series of simulation experiments includes: using time-series synthetic datasets instead

of partially observed synthetic data to see whether all the parameters can be correctly recovered; only using early observation data to show how the LNA model misclassifies the biological variability into initial condition standard deviation instead of to the stochastic nature of solving stochastic differential equations; generate synthetic dataset using different magnitudes of initial condition standard deviation and technical measure standard deviation. During the Bayesian inference, fixing initial condition standard deviation at its ground truth and inferring all other parameters or fixing technical measure error at its ground truth and inferring all other parameters. We find that posterior violin plots and 2σ Bayesian credible intervals (BCIs) of the stochastic process model and ODE model inference is distinct when the real initial condition cell count variation is relatively small while the inference results become similar as initial condition variation gets larger, and finally propose a solution to this limitation by estimating the initial condition standard deviation and technical measurement standard deviation directly from the experimental dataset and fix both standard deviations during the Bayesian inference. We observe similar results to that of the fitting models to the synthetic dataset generated by the Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data: Bayesian LNA inference can generally lead to narrower posterior distributions than that of NHODE and HODE which means that LNA can be more certain to correctly identify the parameter values. The experimental datasets are best described by the LNA model since it achieves the highest value of out-of-sample expected log predictive density.

Chapter 5

Bayesian Optimal Experimental Design for Hematopoesis using Game Theory

5.1 Introduction

Bayesian inference is attractive in situations where we do not have much data but have some prior knowledge about model parameters that can uncover the underlying mechanism of biological processes. Selecting a good experimental design is very important to control the costs of experiments while maximizing information gain from the data. Therefore, optimal experimental design has a long history in pharmacokinetics, ecology, and epidemiology [19, 3, 81, 54, 80]. Take the hematopoiesis experiment in the previous chapter, for example. Mice are sacrificed during the bone marrow extraction operation in order to measure hematopoietic cell counts. Thus, the resulting cell count data obtained from the experiment can be only partially observed for each individual instead of observed as time series. Thus, working with

partially observed data from mice is challenging, so it is crucial to carefully choose a data collection schedule after the perturbation experiment: on what days to sacrifice mice and how many to sacrifice on each day. A classical Bayesian approach to finding an optimal experimental design is seeking to maximize the Kullback-Leibler (KL) divergence between prior and posterior distributions by considering a large number of potential designs [42, 11, 32, 18]. However, this optimization procedure is very expensive when dealing with high-dimensional complex model systems. To overcome this difficulty, a game theory approach was recently developed [55]. We use this new method to find the optimal experimental design for our hematopoiesis perturbation experiment and avoid repeated approximations of the posterior distributions under all possible experimental designs.

5.1.1 Motivating Examples

In a previous work on choosing an optimal experimental design for a hematopoiesis perturbation experiment, the authors selected the optimal design with a Bayesian utility score based on information gain using KL divergence and demonstrated that proper designs can lead to better estimation of parameters of interest with only a few observations [42]. Let a dataset be $\boldsymbol{D}_{obs}$, model parameters of interest be $\boldsymbol{\Theta}$, and a design be $\boldsymbol{\tau}$. Then the expected utility score for design $\boldsymbol{\tau}$ can be written as

$$U(\boldsymbol{\tau}) = E_{\boldsymbol{\Theta},\boldsymbol{D}_{obs}}[u(\boldsymbol{D}_{obs},\boldsymbol{\tau})] = \int_{\boldsymbol{D}_{obs}} \int_{\boldsymbol{\Theta}} u(\boldsymbol{D}_{obs},\boldsymbol{\tau}) p(\boldsymbol{y}_{obs}|\boldsymbol{\Theta},\boldsymbol{\tau}) \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta} d\boldsymbol{y}_{obs}, \tag{5.1}$$

where the KL divergence score function is

$$u(\boldsymbol{D}_{obs},\boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} log\left(\frac{p(\boldsymbol{\Theta}|\boldsymbol{D}_{\boldsymbol{obs}},\boldsymbol{\tau})}{p(\boldsymbol{\Theta})}\right) p(\boldsymbol{\Theta}|\boldsymbol{D}_{obs},\boldsymbol{\tau}) d\boldsymbol{\Theta}. \tag{5.2}$$

The diagram illustration for the optimal experimental design using the utility in Eqn.(5.1)

can be seen in Figure 5.1. Start from selecting various experimental designs e.g., $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3$ that specify different sets of observation days and different numbers of mice to observe on those days. Then, we generate multiple synthetic datasets for each experimental design based on the prior belief of the model parameters under the Bayesian framework, e.g., synthetic dataset j of design k is denoted as D_j^k. Then, we perform the posterior calculation for each synthetic dataset, e.g., we calculate posterior density $p(\boldsymbol{\Theta}|\boldsymbol{D_j^k})$ of synthetic dataset $\boldsymbol{D}_j^k$. Finally, we calculate and compare the expected utilities using Eqn.(5.1) to choose one optimal design among all the potential designs. The best design is selected as the one with the highest expected utility.

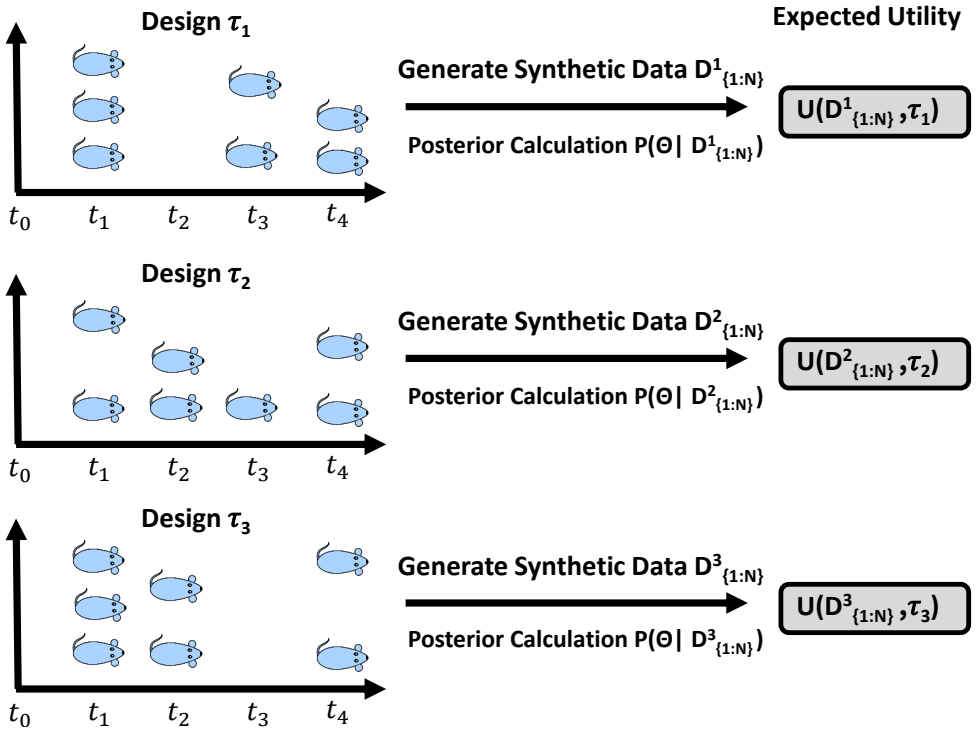


Figure 5.1: Diagram illustration for the optimal experimental design motivating examples. $\boldsymbol{\tau_j}$ refers to the experimental design j. $\boldsymbol{D}_j^k$: refers to synthetic dataset j of design k; $p(\boldsymbol{\Theta}|\boldsymbol{D_j^k})$ refers to posterior density of synthetic dataset $\boldsymbol{D}_j^k$. $U(\boldsymbol{D}_j^k, \boldsymbol{\tau}_j)$ refers to the expected utility for synthetic dataset $\boldsymbol{D}_j^k$ from experimental design τ_j

5.2 Methods

Here we follow [55, 74] to provide a derivation for a score function based on a Fisher information distance and review a framework for the game theory optimization approach.

5.2.1 Fisher Information Matrix (FIM)

The Fisher information matrix (FIM) for parameters of interest $\boldsymbol{\Theta}$ can be regarded as an approximation to the posterior precision of $\boldsymbol{\Theta}$ [72] or the expected value of the observed Fisher information (the negative of the hessian matrix of the log-likelihood function with respect to the parameter vector). Let $\boldsymbol{\tau}$ be the experimental design and $p(\boldsymbol{D}|\boldsymbol{\Theta}, \boldsymbol{\tau})$ be the prior predictive density function. The FIM can be written as

$$\mathcal{I}_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}, \boldsymbol{\tau}) = E_{\boldsymbol{D}\sim p(\boldsymbol{D}|\boldsymbol{\Theta},\boldsymbol{\tau})}(u(\boldsymbol{D}, \boldsymbol{\Theta}, \boldsymbol{\tau})^t u(\boldsymbol{D}, \boldsymbol{\Theta}, \boldsymbol{\tau})), \tag{5.3}$$

where $u(\boldsymbol{D}, \boldsymbol{\Theta}, \boldsymbol{\tau})$ refers to the score function (the gradient of the log-likelihood function with respect to the parameter vector). Let the scoring rule (which provides a summary measure for the evaluation of probabilistic prediction) be $\mathcal{S}(q(\boldsymbol{\Theta}), \boldsymbol{\Theta})$ where q($\boldsymbol{\Theta}$) is the density of $\boldsymbol{\Theta}$. Then the expected utility of design $\boldsymbol{\tau}$ for the Bayesian approach can be chosen as

$$\mathcal{U}_{entropy\ diff} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta}|\boldsymbol{D},\boldsymbol{\tau})}\mathcal{S}(\pi(\boldsymbol{\Theta}|\boldsymbol{D}, \boldsymbol{\tau}), \boldsymbol{\Theta}) - E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}\mathcal{S}(\pi(\boldsymbol{\Theta}), \boldsymbol{\Theta}), \tag{5.4}$$

$$\mathcal{U}_{divergence} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}[\mathcal{S}(\pi(\boldsymbol{\Theta}|\boldsymbol{D}, \boldsymbol{\tau}), \boldsymbol{\Theta}) - \mathcal{S}(\pi(\boldsymbol{\Theta}), \boldsymbol{\Theta})], \tag{5.5}$$

where $\mathcal{U}_{entropy\ diff}$ measures the entropy difference of the posterior and prior density distributions and $\mathcal{U}_{divergence}$ measures the divergence from prior density to posterior density.

5.2.2 Hyvärinen Score Rule

Hyvärinen score rule is very popular for unnormalized densities since the score function $\mathcal{S}(q(\boldsymbol{\Theta}), \boldsymbol{\Theta})$ only depends on derivatives of log density log(q($\boldsymbol{\Theta}$)). The score function, entropy difference and density divergence utilities under Hyvärinen score rule can be written as

$$\mathcal{S}(q(\boldsymbol{\Theta}), \boldsymbol{\Theta})) = 2\Delta log(q(\boldsymbol{\Theta}), \boldsymbol{\Theta})) + ||\nabla log(q(\boldsymbol{\Theta}), \boldsymbol{\Theta})||^2, \tag{5.6}$$

$$\mathcal{U}_{entropy\ diff} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta}|\boldsymbol{D},\boldsymbol{\tau})}||\nabla log\pi(\boldsymbol{\Theta}|\boldsymbol{D},\boldsymbol{\tau})||^2 - E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}||\nabla log\pi(\boldsymbol{\Theta})||^2, \tag{5.7}$$

$$\mathcal{U}_{divergence} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}||\nabla log\pi(\boldsymbol{\Theta}|\boldsymbol{D},\boldsymbol{\tau}) - ||\nabla log\pi(\boldsymbol{\Theta})||^2, \tag{5.8}$$

where $||\cdot||$ refers to the $\mathcal{L}_2$ norm. Let

$$\mathcal{U}_{trace} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}\left[\mathrm{tr}(\mathcal{I}_{\boldsymbol{\Theta}}(\boldsymbol{\Theta},\boldsymbol{\tau}))\right] = \mathrm{tr}\left(E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}\left[\mathcal{I}_{\boldsymbol{\Theta}}(\boldsymbol{\Theta},\boldsymbol{\tau})\right]\right). \tag{5.9}$$

Let

$$\bar{\mathcal{I}} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}\mathcal{I}(\boldsymbol{\Theta},\boldsymbol{\tau}). \tag{5.10}$$

Since the $\mathcal{U}_{trace}$ and $\mathcal{U}_{entropy\ diff}$ and $\mathcal{U}_{divergence}$ are proved to have the same expectation value with only additive constant difference [74], they are equivalent when used to choose an optimal experimental design. An intuitive explanation for this equivalence is that maximizing $\mathcal{U}_{trace}$ is approximately equivalent to maximizing the sum of the eigenvalues $\{\lambda_1, \cdots \lambda_n\}$ of matrix $\bar{\mathcal{I}}$, which is often equivalent to maximizing one of the eigenvalue λ_i of matrix $\bar{\mathcal{I}}$ while other eigenvalues $\lambda_j$are small for $\forall j \in (1, \cdots N), j \neq i$ [74, 51],

$$\max_{\boldsymbol{\Theta},\boldsymbol{\tau}} \mathcal{U}_{trace} \Longleftrightarrow \max_{n\in\{1,\cdots N\}} \sum \lambda_n \Longleftrightarrow \max \lambda_i \text{ while } \lambda_j \text{ are small for } \forall j \in (1, \cdots N), j \neq i. \tag{5.11}$$

Then the Fisher information gain (FIG) is defined as

$$\mathcal{J}_{FIG} = E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})} tr(\mathcal{I}(\boldsymbol{\Theta},\boldsymbol{\tau})) = tr(E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})}\mathcal{I}(\boldsymbol{\Theta},\boldsymbol{\tau})). \tag{5.12}$$

It can be also shown that maximizing $\mathcal{U}_{trace}$ is equvilent to maximizing $\mathcal{J}_{FIG}$ [56]. Then under the Hyvärinen Score Rule, $\mathcal{J}_{FIG}$ can be used as a utility based on the trace of the Fisher information.

5.2.3 Game Theory Framework

The game involves two players — an experimenter and a critic and five decision actions: selecting experimental design $\boldsymbol{\tau}$, selecting an invertible matrix $\boldsymbol{A}$, draring parameters of interest $\boldsymbol{\Theta}$ from prior distributions, generating observed data y and calculating critic reward function $\mathcal{K}$. Note that the reward function for the experimenter is $-\mathcal{K}$. Since the optimal design problem involves players performing actions sequentially, e.g., changing designs at each iteration, it is appropriate to consider the subgame perfect equilibria (SPE) condition: players take action to maximize their expected reward based on the assumption that later actions will also do so.

At each iteration during the game given the current design $\boldsymbol{\tau}$, first, parameters of interest $\boldsymbol{\Theta}$, e.g., rate parameter values and initial condition values, are drawn from the prior distributions. Then the critic selects an invertible change of base matrix $\boldsymbol{A}$, creating new parameter variables

$$\boldsymbol{\Phi} = \boldsymbol{A}^{-1}\boldsymbol{\Theta} \tag{5.13}$$

, that aims to avoid poor designs in practice (e.g., requiring all observations to occur at a single time point when expected utility is optimized given accurate inference for one linear combination of parameters but not for others [55]). By selecting a linear transformation of the parameter Θ to less favorable ones Φ, the framework can then account for optimization under different sets of parameters. Note that posterior densities and the score function are

unchanged by this base change operation if $|det(\boldsymbol{A})| = 1$ since

$$\pi_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}|y, \boldsymbol{\tau}) = \pi_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|y, \boldsymbol{\tau})|det(\boldsymbol{A})| = \pi_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|y, \boldsymbol{\tau}), \text{ and} \tag{5.14}$$

$$log(\pi_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}|y, \boldsymbol{\tau})) = log(\pi_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|y, \boldsymbol{\tau})) + log(|det(\boldsymbol{A})|) = log(\pi_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|y, \boldsymbol{\tau})). \tag{5.15}$$

Next, based on the Eqns.(5.3, 5.12, 5.13), let

$$\mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A}) = -E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})} tr[(Au(\boldsymbol{\Theta}, \boldsymbol{\tau}))^t Au(\boldsymbol{\Theta}, \boldsymbol{\tau})] = -E_{\boldsymbol{\Theta}\sim\pi(\boldsymbol{\Theta})} tr[A^t \mathcal{I}(\boldsymbol{\Theta}, \boldsymbol{\tau}) A] \tag{5.16}$$

be the reward of critics, and then the experimenter will select a new design that tries to maximize his reward -$\mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A})$ or equivalently minimize the critic's reward $\mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A})$. Note that the selection decisions made by the experimenter and the critic are based on the backward auto differentiation of their reward from the last iteration. Then satisfying the SPE criteria is equivalent to solving the nested minimax optimization problem

$$\min_{\boldsymbol{\tau}} \max_{\boldsymbol{A}} \mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A}). \tag{5.17}$$

5.2.4 ODE Model

Here we use the previously developed two-compartment ODE model for hematopoiesis dynamics in Section 3.2.2:

$$\begin{cases} N'_{HSC} = (v_1 - r_1 \cdot N_{MPP}) \cdot N_{HSC}, \\ N'_{MPP} = (v_2 + r_1 \cdot N_{MPP}) \cdot N_{HSC} - u_3 \cdot N_{MPP}, \end{cases} \tag{5.18}$$

where v_1 is the HSCs net proliferation rate; r_1 is the feedback regulation parameter from MPPs on HSC net proliferation rate; v_2 is the HSCs asymmetric division rate; u_3 is the MPPs lumped decay rate. The initial condition mean of HSCs and MPPs μ_{HSC} and μ_{MPP}

Algorithm 6 Nested minimax optimization gradient decent ascent (GDA) algorithm

Require: $\boldsymbol{\tau}_{init}, \boldsymbol{A}_{init}, lr_{\boldsymbol{A}}, lr_{\boldsymbol{\tau}}, Flag_{reg}, N, \boldsymbol{\mu_\Theta}, \boldsymbol{\sigma_\Theta}, \lambda$

$\boldsymbol{\tau}_{curr} = \boldsymbol{\tau}_{init}$

$\boldsymbol{A}_{curr} = \boldsymbol{A}_{init}$

for i in 1:N **do**

$\boldsymbol{\Theta} \sim LN(\boldsymbol{\mu_\Theta}, \boldsymbol{\sigma_\Theta})$

$\boldsymbol{\theta}, \boldsymbol{\mu_{init}}, \sigma_{init}, \sigma_{tech} = \boldsymbol{\Theta}$

$\boldsymbol{u_{sol}} \leftarrow \lambda(\boldsymbol{\theta}, \boldsymbol{\mu_{init}})[\boldsymbol{\tau}_{curr}]$

$\mathcal{J} \leftarrow \partial \boldsymbol{u}_{sol} / \partial \boldsymbol{\theta}$

$\mathcal{I}_{\boldsymbol{\theta}}(\tau_{curr}) = \sigma^2_{tech} \mathcal{J}^t \mathcal{J}$

if $Flag_{reg} == 0$ **then**

$\mathcal{K}(\boldsymbol{\tau}_{curr}, \boldsymbol{A}_{curr}) = -tr[A^t \mathcal{I}(\boldsymbol{\Theta}, \boldsymbol{\tau}) A]$

else

$magnitude = log_{10}(|tr[A^t \mathcal{I}(\boldsymbol{\Theta}, \boldsymbol{\tau}) A]|) - 1$

$\mathcal{K}(\boldsymbol{\tau}_{curr}, \boldsymbol{A}_{curr}) = -tr[A^t \mathcal{I}(\boldsymbol{\Theta}, \boldsymbol{\tau}) A] + 10^{magnitude} * reg(\boldsymbol{\tau}_{curr})$

end if

$\nabla \mathcal{K}(\boldsymbol{\tau}_{curr}, \boldsymbol{A}_{curr}) \leftarrow$ auto differentiation

$\boldsymbol{A}_{curr} -= lr_{\boldsymbol{A}} \cdot \nabla_{\boldsymbol{A}_{curr}} \mathcal{K}(\boldsymbol{\tau}_{curr}, \boldsymbol{A}_{curr})$

$\boldsymbol{\tau}_{curr} += lr_{\boldsymbol{\tau}} \cdot \nabla_{\boldsymbol{\tau}_{curr}} \mathcal{K}(\boldsymbol{\tau}_{curr}, \boldsymbol{A}_{curr})$

end for

are chosen as 30, 67.

5.2.5 Bayesian Framework

For all the Bayesian frameworks, we set the priors of rate value parameters and initial condition cell counts as follows

$$v_1 \sim LN(0.802, 0.1), \quad v_2 \sim LN(0.672, 0.1), \quad u_3 \sim LN(0.4016, 0.1)$$

$$r_3 \sim LN(1.63e-6, 0.1), \quad \mu_{HSC} \sim LN(30, 0.1), \quad \mu_{MPP} \sim LN(67, 0.1) \tag{5.19}$$

Beside, let $\lambda([\mu_{HSC}, \mu_{MPP}], \boldsymbol{\tau}[i])$ be the ODE solution of Eqn.(5.18) with initial cell counts $[\mu_{HSC}, \mu_{MPP}]$ and at observation time $\boldsymbol{\tau}[i]$. We consider three different Bayesian settings as below.

Setting 1

We consider the non-hierarchical Bayesian framework where all the mice start from the same initial condition cell counts, and the technical standard deviation is fixed at 0.1 during the observation time, then the rest of the framework can be written as

$$\delta_{init} = 0, \quad \delta_{tech} = 0.1,$$

and the initial condition cell counts for mouse i can be written as

$$\boldsymbol{u}_i = [\mu_{HSC}, \mu_{MPP}].$$

Then the i^{th} observation data follows the distribution

$$\boldsymbol{y}_i \sim MvN\left(\lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i])), \begin{bmatrix} \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i]))^2_{HSC} & 0 \\ 0 & \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[I]))^2_{MPP} \end{bmatrix} \cdot 0.01 \cdot \boldsymbol{I}\right).$$

Setting 2

We consider the hierarchical Bayesian inference framework where the initial condition cell counts of all mice are sampled from the same prior distribution with fixed initial condition standard deviation and technical standard deviation. Then the rest of the framework can be written as

$$\delta_{init} = 0.1, \quad \delta_{tech} = 0.1,$$

and the initial condition cell counts for mouse i can be written as

$$\boldsymbol{u}_i \sim MvN\left([\mu_{HSC}, \mu_{MPP}], \begin{bmatrix} \mu^2_{HSC} & 0 \\ 0 & \mu^2_{MPP} \end{bmatrix} \cdot 0.01 \cdot \boldsymbol{I}\right).$$

The i^{th} observation data follows the distribution

$$\boldsymbol{y}_i \sim MvN\left(\lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i])), \begin{bmatrix} \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i]))^2_{HSC} & 0 \\ 0 & \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[I]))^2_{MPP} \end{bmatrix} \cdot 0.01 \cdot \boldsymbol{I}\right).$$

Setting 3

We consider the hierarchical Bayesian framework where the initial condition cell counts of all mice are sampled from the same prior distribution. The initial condition standard deviation and technical standard deviation are also being inferred, then the rest of the framework can be written as

$$\delta_{init} \sim LN(0.1, 0.1), \quad \delta_{tech} \sim LN(0.1, 0.1)$$

and the initial condition cell counts for mouse i can be written as

$$\boldsymbol{u}_i \sim MvN\left([\mu_{HSC}, \mu_{MPP}], \begin{bmatrix} \mu^2_{HSC} & 0 \\ 0 & \mu^2_{MPP} \end{bmatrix} \cdot \delta^2_{init} \cdot \boldsymbol{I}\right).$$

The i^{th} observation data follows the distribution

$$\boldsymbol{y}_i \sim MvN\left(\lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i])), \begin{bmatrix} \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[i]))^2_{HSC} & 0 \\ 0 & \lambda(\boldsymbol{u}_i, \boldsymbol{\tau}[I]))^2_{MPP} \end{bmatrix} \cdot \delta^2_{tech} \cdot \boldsymbol{I}\right).$$

5.2.6 Regulation Term

Without any regulation constraint on the optimization reward function, the optimal designs obtained from the framework can only include a few clusters of repeated observation days. This happens when the Bayesian estimation is too certain of the oscillation dynamics based on the given priors and differential equation model. However, in the real data analysis

scenarios, without any model misspecification, we also want to observe more days with some gap between them for model criticism. The reason for this is that the experimental design days will converge to a few (e.g., three) clusters of repeated observation without any regulation [56]. An intuitive explanation for this is that the early observation can learn the initial condition values, the intermediate observations learn the rate parameters, the late observation learns the steady state values, and consequently, all the parameters can be identified through three clusters of repeated observations. Thus, we also propose a regulation term that enforces at most three data that can be observed per day and a 2-day gap between observations to keep a couple of data observed at other times for model criticism.

Algorithm 7 Regulation term for nested minimax optimization algrorithm

Require: $\boldsymbol{\tau}$
  Flag1, Flag2, Reg = 0,0,0
  $\boldsymbol{\tau}_{sorted} = sort(\boldsymbol{\tau})$
  gaps = $\boldsymbol{\tau}_{sorted}[2:N] - \boldsymbol{\tau}_{sorted}[1:N-1]$
  for i in 3:N-1 **do**
    Flag1 = Flag2
    if $gaps[i-2] \leq 0.2$ and $gaps[i-1] \leq 0.2$ **then**
      Flag2 = 1
      if Flag1 == 0 **then**
        Reg += Relu(2.0-gaps[i])
      else
        Reg += Relu(2.0-gaps[i-1])
      end if
    end if
  end for

5.2.7 Evaluation Metric

For design evaluation, we consider the Bayesian utility using KL divergence described in section 5.1.1 that optimize the information gain by maximizing the distance between prior and posterior distributions.

5.3 Results

We employ the adversarial optimization approach with the three Bayesian settings described in Section 5.2.5 and with or without the regulation term developed in Section 5.2.6. Each of the six optimization cases is run for 600,000 iterations. The initial condition design includes 24 mice that are evenly spaced between day 0 and day 35.

5.3.1 Optimization Trace Plots and Final Designs

Optimization results without the regulation term for each Bayesian setting can be seen in Figures 5.2, 5.4, and 5.6, respectively. These optimal designs are similar. Subfigure (a) refers to the observation trace plot of the designs that the experimenter chosen changed in terms of GDA iterations. We can see that designs converge to three clusters of repeated observation days that are at an initial observation time, an intermediate time, and a late time when the system is near the steady state. Subfigure (b) refers to the trace plots of matrix $\boldsymbol{A}$ element values which are selected by the critic in log-scale in terms of the GDA iterations. We can see that the elements in matrix A converge and mix well in the later iterations. Subfigure (c) refers to the reward function value $\mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A})$, which the critic wants to maximize and the experimenter wants to minimize. We can see that object $\mathcal{K}$ value converges and mixes well in the later GDA iterations. Subfigure (d) refers to the experimental design at the last GDA optimization iteration, which we use to calculate the Bayesian utility for design evaluation.

The optimization results with the regulation terms for each Bayesian setting can be seen in Figures 5.3, 5.5, and 5.7, respectively. They have similar optimization designs. Subfigure (a) refers to the observation trace plot of the designs that the experimenter chosen changed in terms of GDA iterations. Different from the cases when there is no regulation term, the designs selected by the experimenter are enforced to have some gaps when the number of a

cluster exceeds the maximum capacity. This leads the designs to converge into nine clusters of repeated observation days with gaps between them. Subfigure (b) refers to the trace plots of matrix $\boldsymbol{A}$ element values which are selected by the critic in log-scale in terms of the iterations. We can see that the elements in matrix A converge and mix well in the later iterations. Subfigure (c) refers to the reward function value $\mathcal{K}$. Different from the trace plot for the cases without regulation term, we can see that in the late iterations, there are some overshoots of the reward score value $\mathcal{K}(\boldsymbol{\tau}, \boldsymbol{A})$. This is caused by the regulation term when there are more than three data that want to converge into the same cluster or two clusters are too close to each other. Then in the next few iterations after the overshoot, the experimenter modifies his designs by enforcing extra data away from the cluster with maximum capacity, and consequently, the reward value return to the steady state value magnitude. Subfigure (d) refers to the experimental design at the last GDA optimization iteration, which we use to calculate the Bayesian utility for design evaluation. We can see that observation data is being assigned to 9 different clusters.

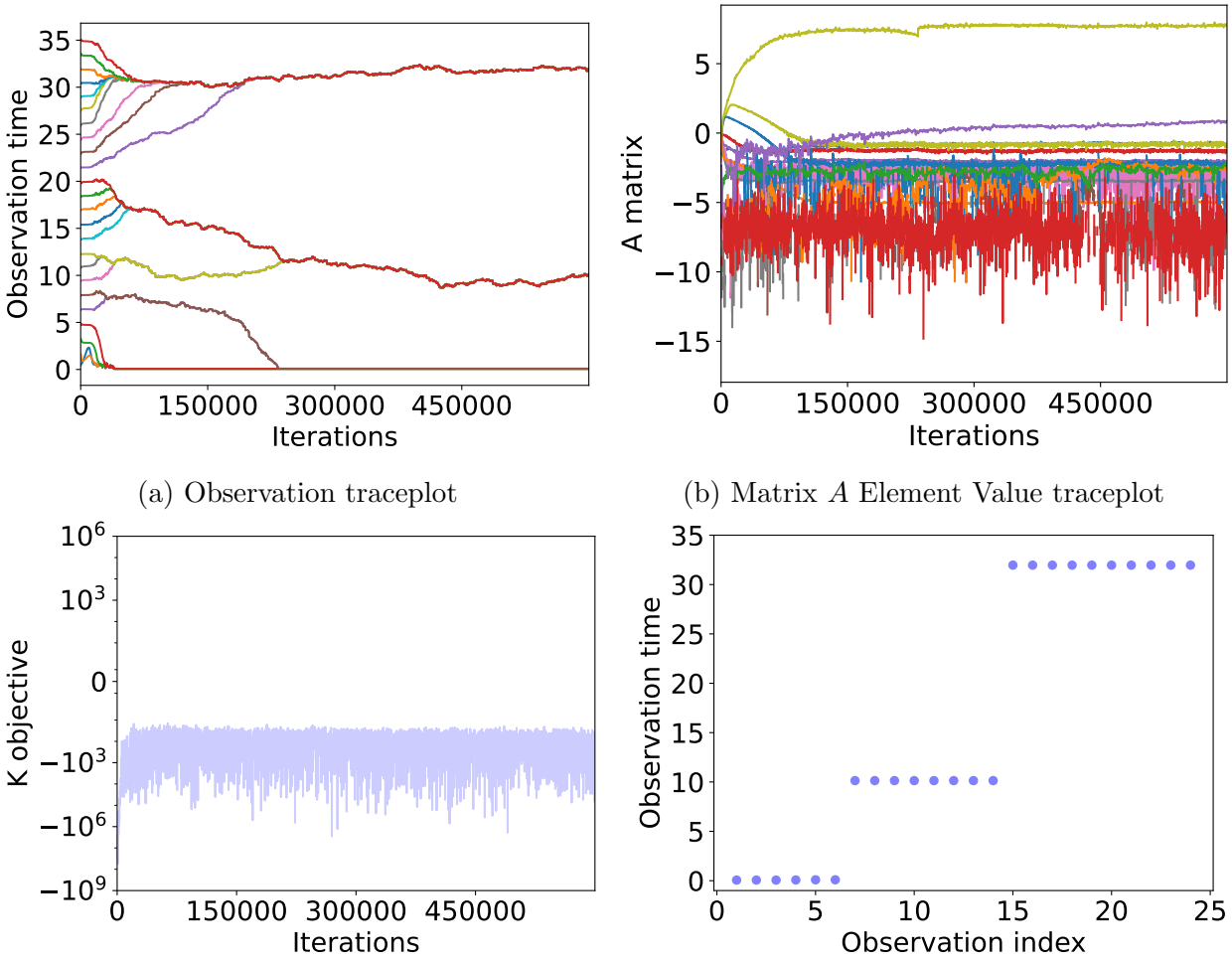
The final experimental designs for all the cases can be seen in Figure 5.8. From the non-regulation (non-gap) cases, we can see that the intermediate and late observation clusters move early as the initial condition standard deviation and technical standard deviation are both fixed to 0.1 in log-scale or further both being inferred compared to the case when there is no initial condition standard deviation and the technical standard deviation is fixed to 0.1 in the log-scale. From the regulation (gap) cases, we can see that the new observation clusters are separated from the corresponding clusters in the non-regulation cases.

5.3.2 Design Evaluation with Different Beyasian Settings and Regulation Condition

We evaluate the final designs for each Bayesian setting and regulation condition by using the Bayesian utility described in Section 5.1.1. The mean Bayesian utility are estimated by 20 synthetic datasets for each design and 3000 iterations of the MCMC algorithm for parameter posterior approximation using each synthetic dataset. From table 5.1, we can see that for each setting, the expected Bayesian utility of the case without regulation are all larger than that of the corresponding cases when regulation is on. This indicates that the minimax optimization algorithm reliably returns the optimal designs for each Bayesian setting.

Design Comparison Table		
Metric / Design	Non-Reg	Reg
Setting 1		
Mean Bayesian utility	**336.7**	333.4
Setting 2		
Mean Bayesian utility	**367.5**	349.5
Setting 3		
Mean Bayesian utility	**272.2**	252.3

Table 5.1: Experimental design evaluation table using Bayesian utility. The mean Bayesian utilities are estimated by Bayesian inference on synthetic datasets generated based on the final GDA iteration design in different Bayesian settings and regulation conditions.



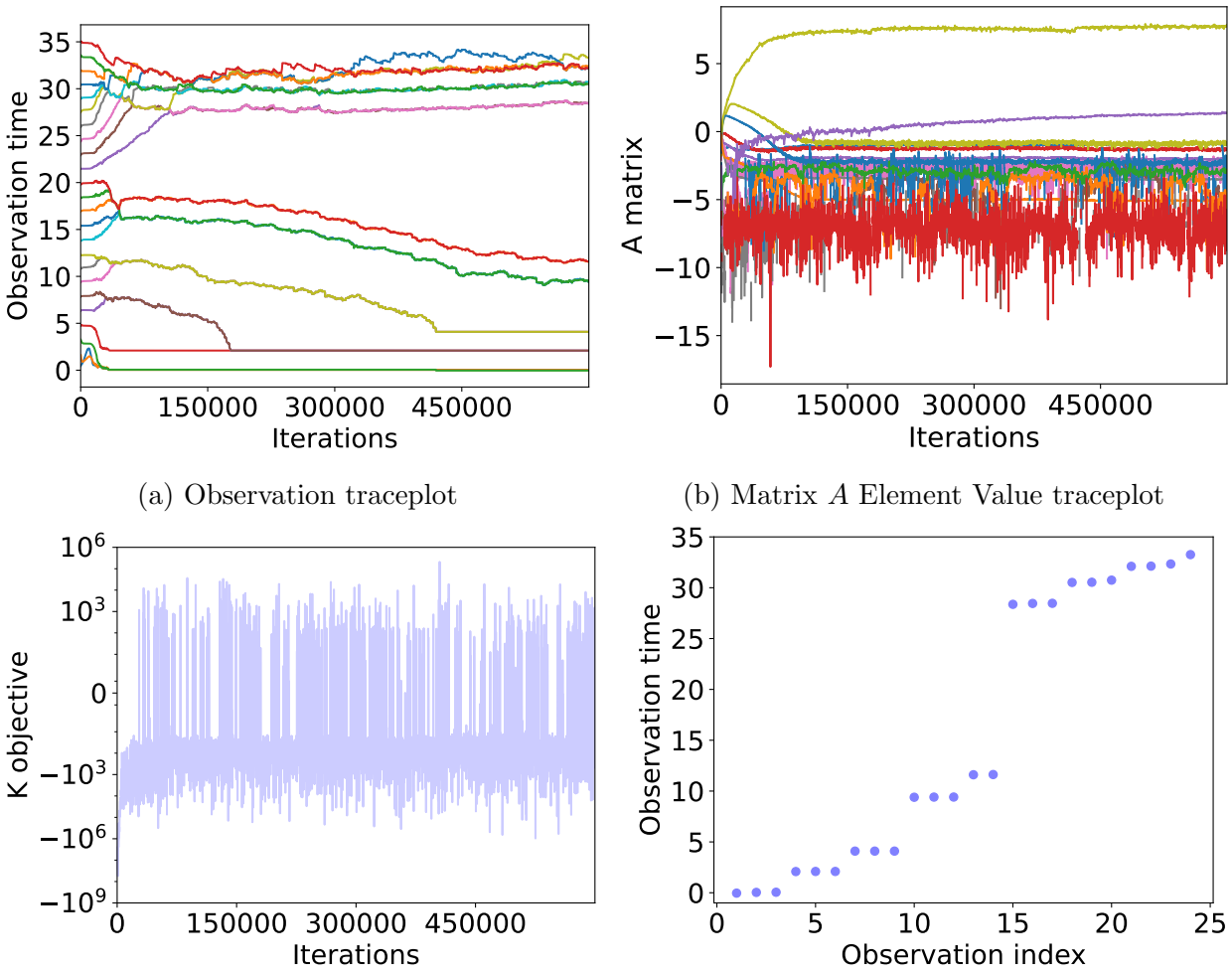
(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.2: Optimal experimental design using two-compartment ODE model under the Bayesian setting that initial condition standard deviation is 0, technical standard deviation term is fixed at 0.1 in log-scale and no regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.



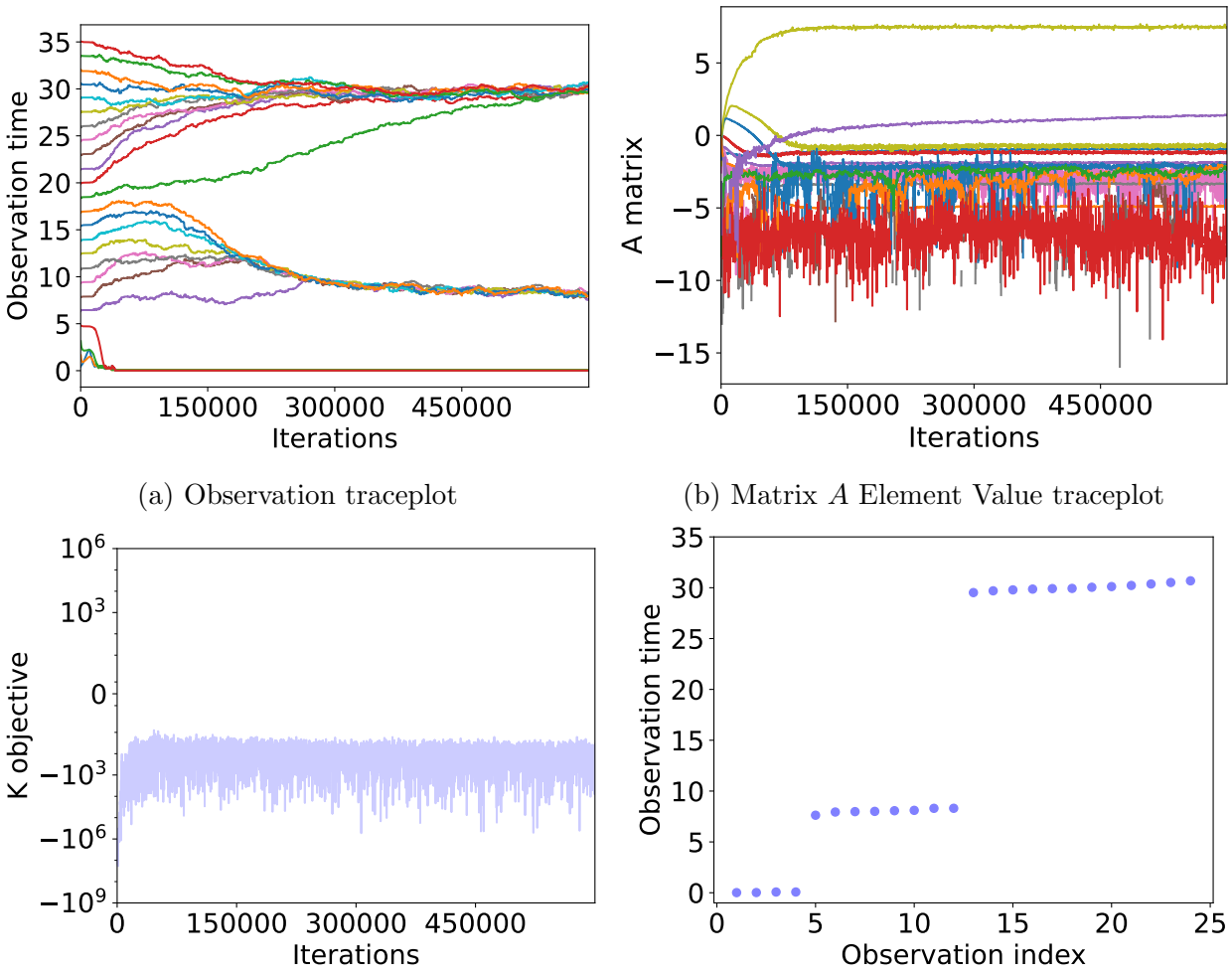
(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.3: Optimal experimental design using two-compartment ODE model under the Bayesian setting that initial condition standard deviation is 0, technical standard deviation term is fixed at 0.1 in log-scale and with the regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.



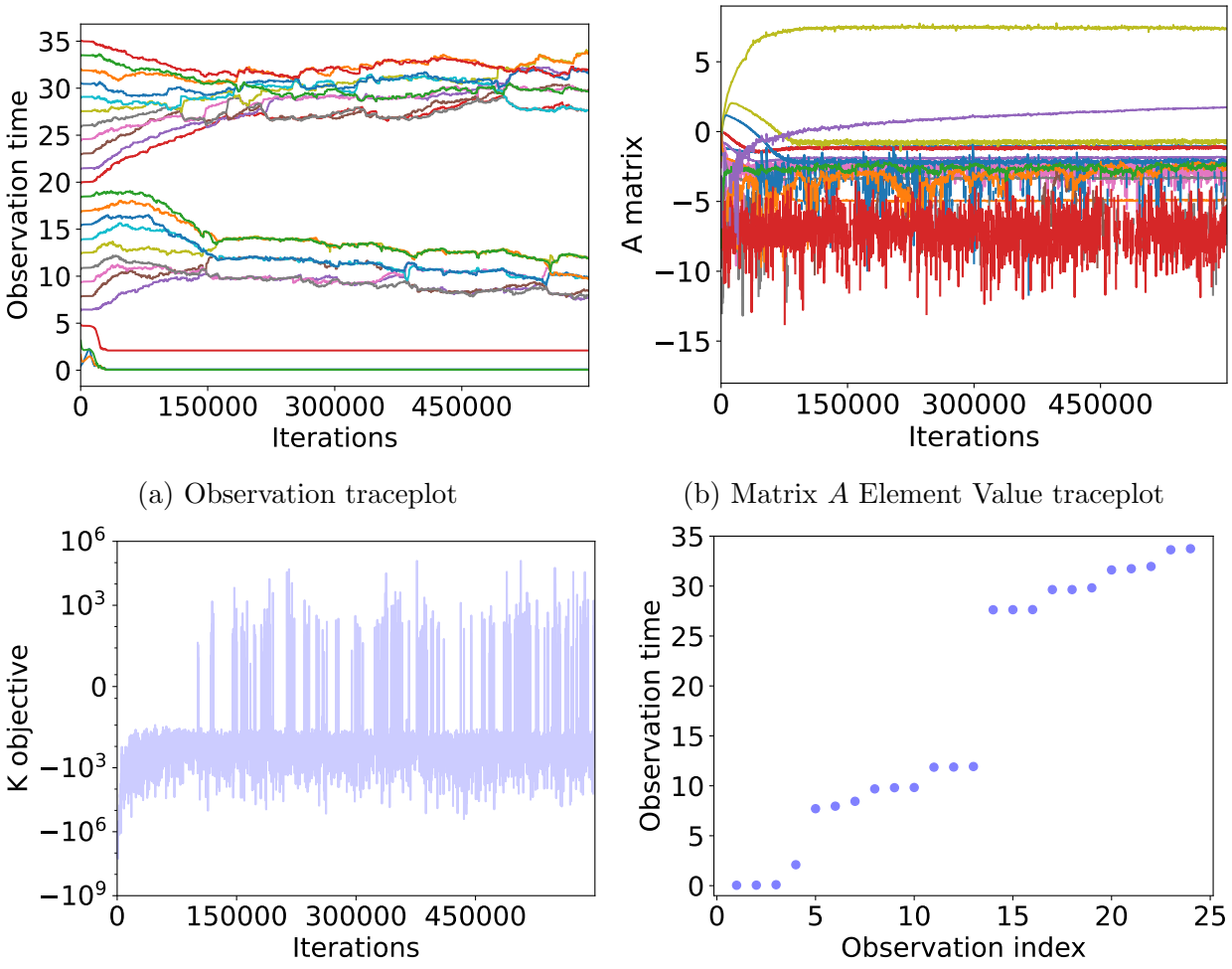
(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.4: Optimal experimental design using two-compartment ODE model under the Bayesian setting that both initial condition standard deviation and technical standard deviation term are fixed at 0.1 in log-scale and no regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.



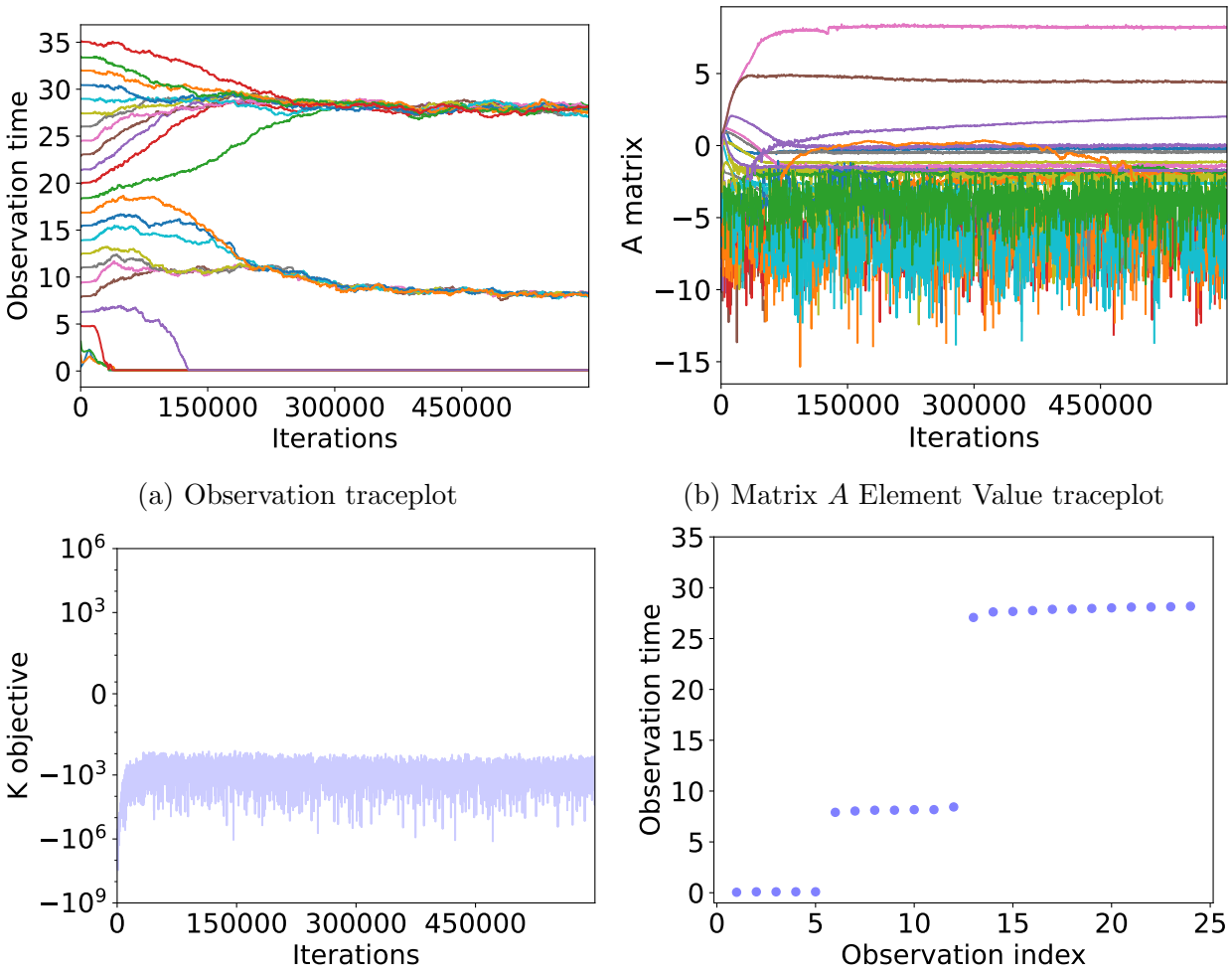
(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.5: Optimal experimental design using two-compartment ODE model under the Bayesian setting that both initial condition standard deviation and technical standard deviation term are fixed at 0.1 in log-scale and with regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.



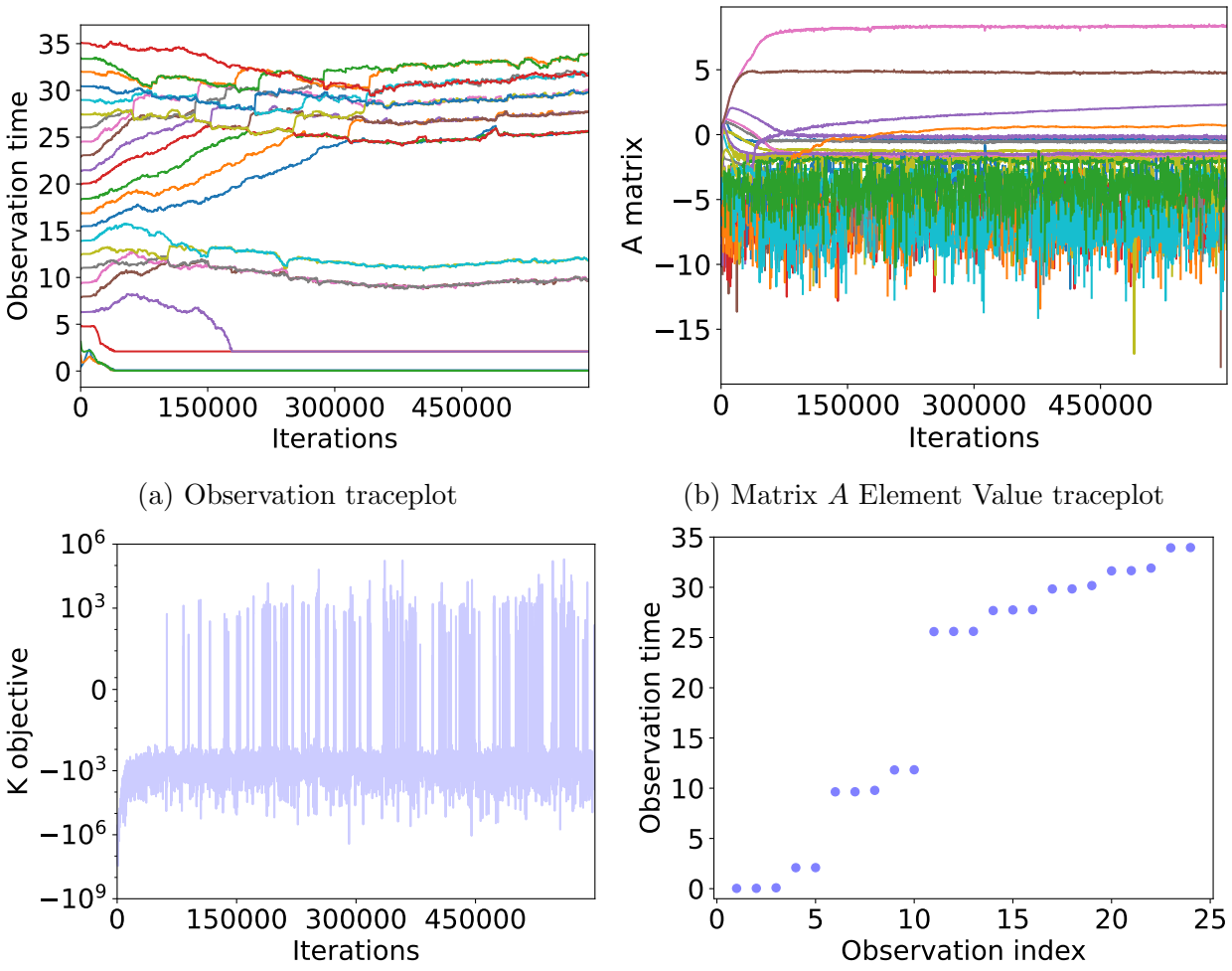
(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.6: Optimal experimental design using two-compartment ODE model under the Bayesian setting that both initial condition standard deviation and technical standard deviation term are both being inferred and no regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.



(a) Observation traceplot

(b) Matrix A Element Value traceplot

(c) $\mathcal{K}$ object value GDA traceplot

(d) Final design

Figure 5.7: Optimal experimental design using two-compartment ODE model under the Bayesian setting that both initial condition standard deviation and technical standard deviation term are both being inferred and with regulation on the number of mice per observation day and the gap between each observation cluster. GDA: gradient decent ascent algorithm. Subfigure (a) refers to the traceplot, which describes the trajectories for each observation time in terms of the GDA iterations. Subfigure (b) refers to the traceplots of the log-scale element values in matrix A which are chosen by the critic to minimize the experimenter's expected rewards. Subfigure (c) traceplot of $\mathcal{K}$ objective value which refers to the criticizer's reward in terms of GDA iterations. Subfigure (d) refers to the experimental design at the final GDA iteration.

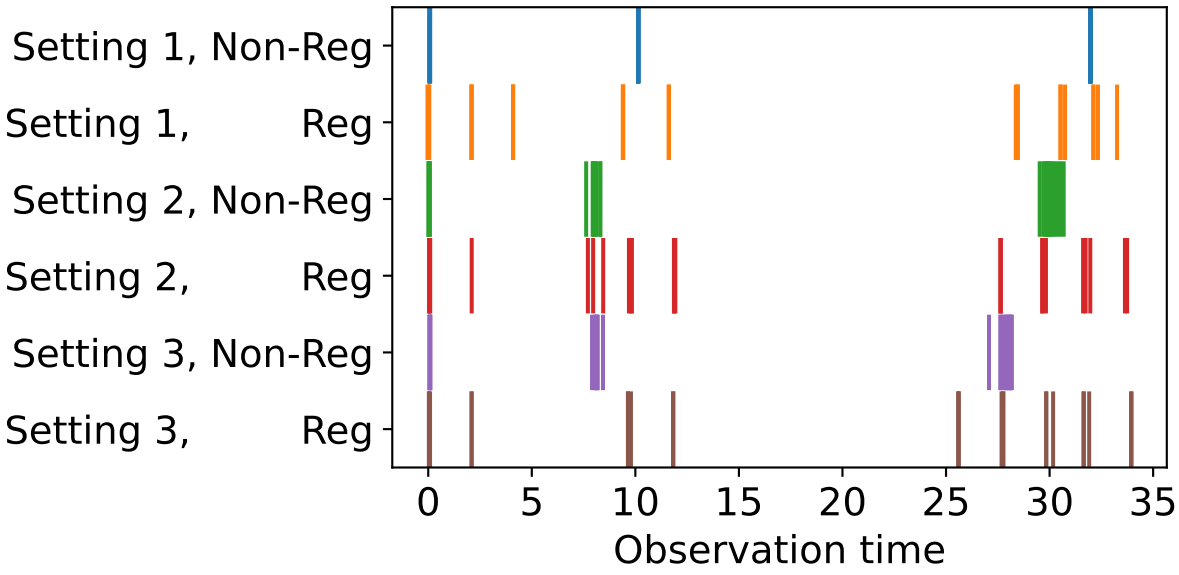


Figure 5.8: Observation time comparison of optimal experimental designs under different Bayesian settings and regulation condtions.

5.4 Conclusion

Since our data collection method, which involves doing bone marrow extraction operation for each mouse, is destructive and consequently expensive, we want to investigate experimental designs that optimize the amount of information gained about the model parameters. We first walk through a traditional method for Bayesian experimental design, which requires the repeated approximations of posteriors probability distributions of the parameters for calculating Bayesian utility. However, this is computationally prohibitive for high-dimensional models. Thus, we employ a new adversarial approach that uses a game theory framework for experimental design without calculating posteriors. We show that this new approach reduces the optimal design problem to a nested minimax optimization problem that can be solved by a gradient descent-ascent algorithm. Next, we propose a regulation term for model criticism and then test the optimization approach under different Bayesian settings from our previously developed Bayesian framework and regulation conditions. By evaluating the optimal designs generated by the GDA algorithm using Bayesian utility, we find that the minimax optimization algorithm using the game theory framework reliably returns optimal experimental designs. For future work, we will experiment with this minimax optimization approach with the ODE model which includes more cell types (e.g., decomposing HSCs into long-term HSCs and short-HSCs or including more downstream cells of MPPs) as well as using the stochastic LNA model to interpolate the missing data instead of using the deterministic ODE model.

Chapter 6

Conclusion & Future Directions

We have proposed a new chemical reaction (CR) ODE model and showed that it can fit data generated by a traditional Hill equation (HE) ODE model in the context of hematopoiesis. We then proposed a hierarchical Bayesian framework, which accounts for technical variation as well as biological variation across individuals through rate parameters. By integrating the chemical reaction ODE model and the hierarchical Bayesian framework, we mimicked the stochastic biological process of hematopoiesis cell dynamics and performed parameter inference for the cell division, differentiation, and decay rates as well as the feedback gains for the various cell compartments. We used real experimental datasets to establish synthetic datasets for a chemical reaction model with both Hierarchical and Non-Hierarchical Bayesian frameworks. We employed a latent variable approach for the data generation process by assuming that the missing data of cell counts follow the ODE trajectories.

By comparing the inference results for CR ODE and HE ODE models on synthetic data, we found that both models can sufficiently recover the ground truth parameter values. However, the CR ODE model is more likely to classify data noise as biological variation while the HE ODE model tends to classify it as technical variation. We found that the Bayes factor model

selection method can decisively select the model that produced the synthetic data. Finally, we tested the chemical reaction ODE model with both frameworks on real experimental datasets and found that the Bayes factor selected the hierarchical framework decisively. This suggests that biological heterogeneity and variability is an important source of noise in the data, and it implies that such variability needs to be considered when perturbing the system to achieve desired results.

We then considered an alternative approach to accounting for data heterogeneity by using a stochastic process model in a non-hierarchical Bayesian framework for the latent cell count time evolution data generation process instead of letting the rate parameter values vary among mice. The computational tractability of Bayesian inference with a stochastic process model is achieved using LNA, which can be regarded as a compromise between the deterministic ODE model and the intractable SDE model. We extended our model to 3D by separating HSCs into LTHSC and STHSC, then derived the corresponding LNA formula for this model in log scale. After evaluating models on a synthetic dataset that mimics our real experimental dataset using PSIS-LOO-CV, we found that the Bayesian LNA model can further improve the out-of-sample prediction accuracy compared to the hierarchical model we previously developed. The real experimental dataset also reliably selects the Bayesian LNA as the appropriate model. We found that Bayesian inference of the LNA model generally leads to narrower posterior distributions, which means that the LNA model is more certain of the parameter values.

To understand the difference between the Bayesian stochastic process model and the Bayesian ODE model, we proposed a hypothesis that: for hierarchical Bayesian inference with differential equations using partially observed datasets, the stochastic process model classifies most of the data heterogeneity into the initial condition variation rather than in the stochasticity of solving time evolution stochastic differential equations if both initial condition biological variation and technical variations are substantial. We did a series of simulation experiments

which shows that the posterior violin plots and 2σ Bayesian credible interval (BCI) of the stochastic process model and ODE model inference are distinct when the ground truth initial condition cell count variation is relatively small while the inference results become similar as initial condition variation gets larger. The results support our hypothesis and point out a limitation of Bayesian inference using a stochastic process model with the partially observed dataset. We then developed a procedure to overcome this limitation by directly measuring the initial condition and technical measurement variances from the dataset and fixing those variances during the Bayesian inference.

Since our data collection method is very expensive, we also want to maximize the information gain from the data while controlling the experimental cost. Thus, we experimented with exploring the best potential experimental design. A fundamental practice is to set different experimental designs and generate synthetic data accordingly. By sampling the posterior for each experimental design and calculating the information gain between priors and those posteriors, one can select the design with the largest information gain as the optimal design. However, this approach is very computationally expensive because of the many MCMC samples that need to be generated. To avoid this computational cost, we employed an adversarial approach to select the optimal experimental design that can be used. Game Theory methods are used to circumvent the costly posterior sampling processes used in Bayesian utility selection methods. Instead of repeatedly calculating the Bayesian utility, we reduced the experiment selection process to an optimization minimax problem for computation efficiency. We experimented with this approach with three Bayesian settings using the two-compartment chemical reaction ODE model and a regulation term. The final designs converge to a few clusters of repeated observation days. By evaluating those designs with Bayesian utility, we found that the optimization results reliably return the optimal experimental design based on information gain from the posterior to prior distributions.

As further work, we will extend our new explanatory and data-driven chemical reaction ODE

model to account for more downstream cells of MPPs e.g. Granulocytes (GRAs) with our new scalable Bayesian LNA Framework. By visualizing the second experimental dataset which focuses on the early 10 days after the perturbation, we find that the GRAs are also perturbed from their steady state value Thus, we will also incorporate GRAs into our model. By using the new stochastic process model with more cell compartments (e.g. GRAs), our current two experimental datasets may not be enough to reliably infer the model parameters. For this situation, we will try to explore the best potential experimental designs using the game theory framework with higher dimensional models which include more downstream cells of MPPs and the Bayesian LNA framework.

Bibliography

[1] *Stochastic differential equations: An introduction with applications*. Springer, 2013.

[2] *Stochastic modelling for Systems Biology*. CRC Press, Taylor et Francis Group, may 2020.

[3] L. Aarons and K. Ogungbenro. Optimal design of pharmacokinetic studies. *Nordic Pharmacological Society*, 106, 250–255, 2010.

[4] M. Barile, K. Busch, A.-K. Fanti, A. Greco, X. Wang, H. Oguro, Q. Zhang, S. J. Morrison, H.-R. Rodewald, and T. Höfer. Hematopoietic stem cells self-renew symmetrically or gradually proceed to differentiation. *bioRxiv*, 2020.

[5] A. Becker, E. McCulloch, and T. J.E. Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. *Nature*, (197):452–454, 1963.

[6] J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. E. Heckerman, A. F. M. Smith, and M. West. Non-centered parameterisations for hierarchical models and data augmentation. 2003.

[7] M. J. Betancourt and M. Girolami. Hamiltonian monte carlo for hierarchical models, 2013.

[8] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, may 2011.

[9] K. Busch, K. Klapproth, M. Barile, M. Flossdorf, T. Holland-Letz, S. M. Schlenner, M. Reth, T. Höfer, and H.-R. Rodewald. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Stem Cells Dev.*, 518, 542–546, 2015.

[10] G. Buzi, A. D. Lander, and M. Khammash. Cell lineage branching as a strategy for proliferative control. *BMC Biology*, 13, 2015.

[11] K. Chaloner and I. Verdinelli. Bayesian Experimental Design: A Review. *Statistical Science*, 10(3):273 – 304, 1995.

[12] T. Chen, E. B. Fox, and C. Guestrin. Stochastic gradient hamiltonian monte carlo, 2014.

[13] H. Cheng, Z. Zheng, and T. Cheng. New paradigms on hematopoietic stem cell differentiation. *Protein Cell .*, 11(1):34-44., 2020.

[14] S. J. Corey, M. Kimmel, and J. N. Leonard. *A Systems Biology Approach to Blood, volume 844 of Advances in Experimental Medicine and Biology*. Springer, 2014.

[15] M. Craig, A. R. Humphries, and M. C. Mackey. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Bull Math Biol.*, Dec;78(12):2304-2357, 2016.

[16] A. M. Czochra, T. Stiehl, A. D. Ho, W. Jarger, and W. Wagner. Modeling of asymmetric cell division in hematopoietic stem cells-regulation of self-renewal is essential for efficient repopulation. *J. R. Soc.*, Stem Cells Dev. 2009 Apr;18(3):377-85.

[17] D. Dingli, A. Traulsen, and J. M. Pacheco. Stochastic dynamics of hematopoietic tumor stem cells. *Cell Cycle.*, Feb 15;6(4):461-6., 2007.

[18] C. C. Drovandi, J. M. McGree, and A. N. Pettitt. Sequential monte carlo for bayesian sequentially designed experiments for discrete data. *Computational Statistics & Data Analysis*, 57(1):320–335, 2013.

[19] M. K. Erdal, K. W. Plaxco, J. Gerson, T. E. Kippin, and J. P. Hespanha. Optimal experiment design with applications to pharmacokinetic modeling. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3072–3079, 2021.

[20] W. Fang, C. M. Bell, A. Sapirstein, S. Asami, K. Leeper, D. J. Zack, H. Ji, and R. Kalhor. Quantitative fate mapping: Reconstructing progenitor field dynamics via retrospective lineage barcoding. *bioRxiv*, 2022.

[21] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70(2):457–466, jan 2014.

[22] Y. Fong, P. Guttorp, and J. Abkowitz. Bayesian inference and model choice in a hidden stochastic two-compartment model of hematopoietic stem cell fate decisions. *Ann Appl Stat*, Dec;3(4):1696-1709, 2009.

[23] C. Fornari, L. O. O'Connor, J. W. T. Yates, S. Y. A. Cheung, D. I. Jodrell, J. T. Mettetal, and T. A. Collins. Understanding hematological toxicities using mathematical modeling. *Clin Pharmacol Ther.*, Oct;104(4):644-654, 2018.

[24] J.-L. Gauvain and C.-H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298, 1994.

[25] A. Gelman, W. R. Gilks, and G. O. Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110 – 120, 1997.

[26] D. T. Gillespies. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, (81(25)):2340–2361, 1977.

[27] I. Glauche, M. Kuhn, C. Baldow, P. Schulze, T. Rothe, H. Liebscher, A. Roy, X. Wang, and I. Roeder. Quantitative prediction of long-term molecular response in tki-treated cml – lessons from an imatinib versus dasatinib comparison. *Sci Rep*, 8,12330, 2018.

[28] D. Golinelli, P. Guttorp, and J. A. Abkowitz. Bayesian inference in a hidden stochastic two-compartment model for feline hematopoiesis. *Math Med Biol.*, Sep;23(3):153-72., 2006.

[29] J. Gomez-Ramirez and R. Sanz. On the limitations of standard statistical modeling in biological systems: a full bayesian approach for biology. *Prog Biophys Mol Biol.*, 113(1):80–91, Sep 2013.

[30] H. Greim, D. A. Kaden, R. A. Larson, C. M. Palermo, J. M. Rice, D. Ross, and R. Snyder. The bone marrow niche, stem cells, and leukemia: impact of drugs, chemicals, and the environment. *Annals of the New York Academy of Sciences*, 1310(1):7–31, 2014.

[31] G. d. Haan and S. S. Lazare. Aging of hematopoietic stem cells. *Blood*, Blood (2018) 131 (5): 479–487., 2018.

[32] X. Huan and Y. Marzouk. Sequential bayesian optimal experimental design via approximate dynamic programming. 04 2016.

[33] T. Höfer, M. Barile, and M. Flossdorf. Stem-cell dynamics and lineage topology from in vivo fate mapping in the hematopoietic system. *Annals of the New York Academy of Sciences*, Curr Opin Biotechnol. 2016 Jun;39:150-156.

[34] F. Jäkel, O. Worm, S. Lange, and R. Mertelsmann. A stochastic model of myeloid cell lineages in hematopoiesis and pathway mutations in acute myeloid leukemia. *PLoS ONE*, 13(10): e0204393, 2018.

[35] R. E. Kass and A. E. Raftery. Bayes factor. *Journal of American Statistical Association*, Vol. 90, No. 430(Jun., 1995),773-795, 1995.

[36] M. Keeling and J. Ross. On methods for studying stochastic disease dynamics. *Journal of The Royal Society Interface*, 5(19):171–181, 2008.

[37] M. Kimmel. Stochasticity and determinism in models of hematopoiesis. *Adv Exp Med Biol.*, 844:119-52, 2014.

[38] S. Koelle, D. Espinoza, C. Wu, X. Jason, R. Lu, B. Li, R. Donahue, and C. Dunbar. Quantitative stability of hematopoietic stem and progenitor cell clonal output in transplanted rhesus macaques. *Blood*, 129, 01 2017.

[39] M. Komorowski, B. Finkenstädt, C. V. Harper, and D. A. Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10:1471–2105, Oct. 2009.

[40] T. G. Kurtz. Representations of markov processes as multiparameter time changes. *The Annals of Probability*, 8(4):682–715, 1980.

[41] D. C. Liu and j. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45, 503-528, 1989.

[42] L. M. Lomeli, A. Iniguez, P. Tata, N. Jena, Z.-Y. Liu, R. Van Etten, A. D. Lander, B. Shahbaba, J. S. Lowengrub, and V. N. Minin. Optimal experimental design for mathematical models of haematopoiesis. *Journal of The Royal Society Interface*, 18(174):20200729, 2021.

[43] A. L. MacLean, C. L. Celso, and M. P H Stumpf. Concise review: Stem cell population biology: Insights from hematopoiesis. *Stem Cells.*, Jan;35(1):80-88, 2017.

[44] A. L. MacLean, H. A. Harrington, M. P. H. Stumpf, and H. M. Byrne. Mathematical and statistical techniques for systems medicine: The wnt signaling pathway as a case study. *Methods in molecular biology (Clifton, N.J.)*, 1386, 2016.

[45] B. Mahadik, B. Hannon, and B. A. C. Harley. A computational model of feedback-mediated hematopoietic stem cell differentiation in vitro. *PLOS ONE*, 14(3):1–21, 030, 2019.

[46] E. Manesso, J. Teles, D. Bryder, and C. Peterson. Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation. *Journal of the Royal Society Interface*, 10(80):20120817, 2013.

[47] A. Marciniak-Czochra, T. Stiehl, A. D. Ho, W. Jäger, and W. Wagner. Modeling of asymmetric cell division in hematopoietic stem cells–regulation of self-renewal is essential for efficient repopulation. *Stem Cells Dev.*, Apr;18(3):377-85., 2009.

[48] J. L. McCauley. Fokker-planck and chapman-kolmogorov equations for ito processes with finite memory. In *Evolution from Cellular to Social Scales*, pages 99–110. Springer Netherlands.

[49] R. M. Neal. Probabilistic inference using markov chain monte carlo methods. 2011.

[50] J. T. Oden, A. Hawkins, and S. Prudhomme. Toward predictive models of tumor growth: A general diffuse-interface continuum theory of mixtures for tumor growth and a bayesian approach to model validation and uncertainty quantification. *Mathematical Models and Methods in Applied Sciences*, Vol. 20, No. 03, pp. 477-517, 2010.

[51] A. M. Overstall. Properties of using fisher information gain for bayesian design of experiments, 2021.

[52] E. Passegue, C. H. Jamieson, l. E. Ailles, and I. L. Weissman. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proceedings of the National Academy of Sciences of the United States of America*, 100:11842–9, 2003.

[53] W. Pei, T. B. Feyerabend, J. Rösslerand, X. Wang, D. Postrach, K. Busch, I. Rode, K. Klapproth, N. Dietlein, C. Quedenau, W. Chen, S. Sauer, S. Wolf, T. Höfer, and H.-R. Rodewald. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, 548, pages456–460, 2017.

[54] C. Pooley, G. Marion, S. Bishop, and A. Doeschl-Wilson. Optimal experimental designs for estimating genetic and non-genetic effects underlying infectious disease transmission. *bioRxiv*, 2022.

[55] D. Prangle, S. Harbisher, and C. S. Gillespie. Bayesian experimental design without posterior calculations: an adversarial approach, 2021.

[56] D. Prangle, S. Harbisher, and C. S. Gillespie. Bayesian Experimental Design Without Posterior Calculations: An Adversarial Approach. *Bayesian Analysis*, pages 1 – 31, 2022.

[57] L. Pujo-Menjouet. Blood cell dynamics: Half of a century of modelling. *Math. Model. Nat. Phenom.*,, Volume 11, Number 1, 2016.

[58] B. Reinhardt, O. Habib, K. L. Shaw, E. Garabedian, D. A. Carbonaro-Sarracino, D. Terrazas, B. C. Fernandez, S. D. Oliveira, T. B. Moore, A. K. Ikeda, B. C. Engel, G. M. Podsakoff, R. P. Hollis, A. Fernandes, C. Jackson, S. Shupien, S. Mishra, A. Davila, J. Mottahedeh, A. Vitomirov, W. Meng, A. M. Rosenfeld, A. M. Roche, P. Hokama, S. Reddy, J. Everett, X. Wang, E. T. L. Prak, K. Cornetta, M. S. Hershfield, R. Sokolic, S. S. D. Ravin, H. L. Malech, F. D. Bushman, F. Candotti, and D. B. Kohn. Long-term outcomes after gene therapy for adenosine deaminase severe combined immune deficiency. *Blood*, 2021 Oct 14;138(15):1304-1316, 2021.

[59] M. A. Rieger and T. Schroeder. Hematopoiesis. *Cold Spring Harbor perspectives in biology*, 4,12, Dex 2012.

[60] A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein, and F. Camargo. Clonal analysis of lineage fate in native haematopoiesis. *Nature*, (Jan 11;553(7687)):212–216, 2018.

[61] A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein, and F. D. Camargo. Clonal analysis of lineage fate in native haematopoiesis. *Nature*, 553, pages212–216, 2018.

[62] I. Roeder and M. Loeffler. A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Experimental hematology*, 30 8:853–61, 2002.

[63] J. M. Rosenberg, M. Kubsch, E.-J. Wagenmakers, and M. Dogucu. Making sense of uncertainty in the science classroom. *Science & Education volume*, 31:1239–1262, June 2022.

[64] A. I. Rozhok, J. L. Salstrom, and J. DeGregori. Stochastic modeling reveals an evolutionary mechanism underlying elevated rates of childhood leukemia. *National Academy of Sciences*, 113 (4) 1050-1055;, 2016.

[65] K. B. Schoedel, M. N. F. Morcos, T. Zerjatke, I. Roeder, T. Grinenko, D. Voehringer, J. R. Göthert, C. Waskow, A. Roers, and A. Gerbaulet. The bulk of the hematopoietic

stem cell population is dispensable for murine steady-state and stress hematopoiesis. *Blood.*, Nov 10;128(19):2285-2296., 2016.

[66] B. N. Sheikh, Y. Yang, J. Schreuder, S. K. Nilsson, R. Bilardi, S. Carotta, H. M. McRae, D. Metcalf, A. K. Voss, and T. Thomas. Moz (kat6a) is essential for the maintenance of classically defined adult hematopoietic stem cells. *Blood.*, Nov 10;128(19):2307-2318., 2016.

[67] L. Siminovitch, M. E.A., and J. Till. The Distribution Of Colony-Forming Cells Among Spleen Colonies. *J Cell Comp Physiol .*, (62):327–36, 1963 Dec.

[68] P. R. Somvanshi and K. V. Venkatesh. *Hill Equation*, pages 892–895. Springer New York, New York, NY, 2013.

[69] R. J. Staversky, D. K. Byun, M. A. Georger, B. J. Zaffuto, A. Goodman, M. W. Becker, L. M. Clavi, and b. Frisch. The chemokine ccl3 regulates myeloid differentiation and hematopoietic stem cell numbers. *Sci Rep*, 8(1):14691, 2018.

[70] T. C. Stutz, J. S. Sinsheimer, M. Sehl, and J. Xu. Computational tools for assessing gene therapy under branching process models of mutation. *Bulletin of Mathematical Biology*, 84, Article number: 15, 2021.

[71] J. Sun, A. Ramos, B. Chapman, J. B. Johnnidis, L. Le, Y.-J. Ho, A. Klein, O. Hofmann, and F. D. Camargo. Clonal dynamics of native haematopoiesis. *Nature volume*, 514, pages322–327, 2014.

[72] A. W. v. d. Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[73] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry. Pareto smoothed importance sampling, 2022.

[74] S. G. Walker. Bayesian information in an experiment and the fisher information distance. *Statistics and Probability Letters*, 112:5–9, 2016.

[75] E. W. J. Wallace. A simplified derivation of the linear noise approximation, 2010.

[76] D. J. Wikinso. Bayesian methods in bioinformatics and computational systems biology. *Briefings in bioinformatics*, 8,2, 2007.

[77] C. Wu, B. L. Li, R. Lu, S. J. Koelle, Y. yang, A. Jares, A. E. Krouse, M. Metzger, F. Liang, K. Loré, C. O. Wu, R. E. Donahue, I. S. Y. Chen, I. Weissman, and C. E. Dunbar. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell .*, (3;14(4)):486–499, 2014.

[78] J. Xu, S. Koelle, P. Guttorp, C. Wu, C. E. Dunbar, J. L. Abkowitz, and V. N. Minin. Statistical inference in partially observed branching processes with application to cell lineage tracking of in vivo hematopoiesis. *Ann. Appl. Stat.*, 13(4): 2091-2119, 2019.

[79] J. Xu, Y. Wang, P. Guttorp, and J. L. Abkowitz. Visualizing hematopoiesis as a stochastic process. *Blood Adv*, Oct 23; 2(20) 2637–2645, 2018.

[80] M. Zemplenyi and J. W. Miller. Bayesian Optimal Experimental Design for Inferring Causal Structure. *Bayesian Analysis*, pages 1 – 28, 2022.

[81] J. F. Zhang, N. E. Papanikolaou, T. Kypraios, and C. C. Drovandi. Optimal experimental design for predator–prey functional response experiments. *Journal of The Royal Society Interface*, 15(144):20180186, 2018.

Appendix A

Chapter 3 Supplementary

A.0.1 Experimental Methods

Mice: Female C57B/6 mice (Jackson Laboratory), aged 8-10 weeks were used for experiments. All experiments were approved by the Institutional Animal Use and Care Committee of University of California, Irvine

Exposure: mice received an intraperitoneal injection of 5-Fluorouracil (5-FU) at 50mg/kg. Control mice received an intraperitoneal injection of 100 μl saline.

Flow cytometry analysis of cell populations: Bone marrow (BM) from tibia and femur were isolated by flushing bones, followed by lysing of red blood cells with Ammonium Chloride Potassium (ACK) buffer. Cells were then resuspended in Phosphate Buffered Saline (PBS) + 2% Fetal Bovine Serum (FBS) with antibody cocktail and incubated at 4°C for 30 minutes prior to washing. Cells were then analyzed on a NovoCyte (ACEA).

Antibodies: BV650 Sca-1 (clone D7, BioLegend), PerCP Cy5.5 CD48 (clone HM48-1, Bi-

oLegend), PE CD150 (clone TC15-12FL2.2 BioLegend), c-kit APC-Cy7 (clone 2B8), FITC CD34, APC lineage cocktail (Ter119, Mac1, Gr1, B220, CD3, CD4 and CD8)

A.0.2 Synchronization for Same Priors

Two Compartment Chemical Reaction ODE Model

$$\begin{cases} N'_{HSC} = v_1 \cdot N_{HSC} - r_1 \cdot N_{HSC} \cdot N_{MPP} \\ N'_{MPP} = v_2 \cdot N_{HSC} + r_1 \cdot N_{HSC} \cdot N_{MPP} - u_3 \cdot N_{MPP} \end{cases}$$

where v_1 is the HSC net proliferation rate which equals to the self-renewal rate minus the death rates of HSCs; v_2 is the HSCs asymmetric division rate; r_1 is the HSC direct differentiation parameter; u_3 is the MPPs decay parameters.

Two Compartment Hill Equation Model

$$\begin{cases} N'_{HSC} = (\frac{2p_0}{1+\gamma_1 \cdot N_{MPP}} - 1) \cdot \eta_1 \cdot N_{HSC} \\ N'_{MPP} = 2(1 - \frac{p_0}{1+\gamma_1 \cdot N_{MPP}}) \cdot \eta_1 \cdot N_{HSC} - \eta_2 \cdot N_{MPP} \end{cases}$$

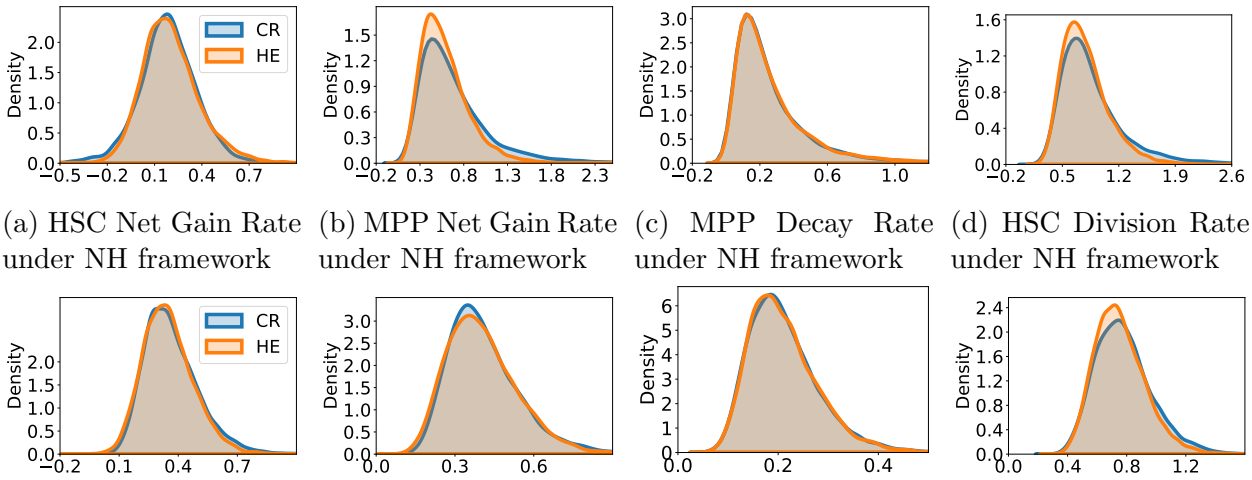
where p_0 is the HSCs maximum self-renewal probability; γ_1 is the feedback parameter from MPPs cell counts to p_0; η_1 is the HSCs total division rate; η_2 is the MPPs decay rate.

Model Comparison

To compare both chemical reaction and hill equation model equally, we group each model into net gain and net loss rate for comparison and to set similar priors.

$$\begin{cases} N'_{HSC} = (v_1 - r_1 \cdot N_{MPP}) \cdot N_{HSC} \\ N'_{MPP} = (v_2 + r_1 \cdot N_{MPP}) \cdot N_{HSC} - u_3 \cdot N_{MPP} \end{cases}$$

$$\begin{cases} N'_{HSC} = (\dfrac{2p_0}{1+\gamma_1 \cdot N_{MPP}} - 1) \cdot \eta_1 \cdot N_{HSC} \\ N'_{MPP} = 2(1 - \dfrac{p_0}{1+\gamma_1 \cdot N_{MPP}}) \cdot \eta_1 \cdot N_{HSC} - \eta_2 \cdot N_{MPP}.. \end{cases}$$



(a) HSC Net Gain Rate under NH framework (b) MPP Net Gain Rate under NH framework (c) MPP Decay Rate under NH framework (d) HSC Division Rate under NH framework

(e) HSC Net Gain Rate under H framewor (f) MPP Net Gain Rate under H framework (g) MPP Decay Rate under H framework (h) HSC Division Rate under H framewor

Figure A.1: Prior synchronization for chemical reaction and Hill equation models under non-hierarchical (NH) and hierarchical (H) Bayesian framework. Blue areas: prior distributions for the chemical reaction model; orange areas: prior distributions for the Hill equation model. (a-d) priors synchronization for non-hierarchical chemical reaction and hill equation models for synthetic and experimental datasets; (e-h) priors synchronization for hierarchical chemical reaction and Hill equation models for synthetic and experimental dataset

Non-Centered Parameterization

In our partially observed data case see experimental materials section 3.2, let μ_u be the initial mean condition and μ_θ be the one-dimensional mean rate value parameter, then for each individual k, we assume that their individual specific parameters follow centered normal probability density functions

$$u_k \sim N(\mu_{u_k}, \delta_{u_k}) \quad \theta_k \sim N(\mu_{\theta_k}, \delta_{\theta_k}).$$

These centered densities can be regarded as capture the absolute behaviors which is independent of latent parameter means μ_u and μ_θ. The posterior density function can be written as

$$Pr(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\mu}_u, \boldsymbol{\mu}_\theta, \boldsymbol{\delta}_u, \boldsymbol{\delta}_\theta | \boldsymbol{y}) \propto \overbrace{\prod_{k=1}^{M} p(y_k | \boldsymbol{u}_k, \boldsymbol{\theta_k})}^{\textit{Likelihood Function}}$$

$$\overbrace{\prod_{k=1}^{M} p(\boldsymbol{u}_k | \boldsymbol{\mu}_{u_k}, \boldsymbol{\delta}_{u_k}) p(\boldsymbol{\theta}_{u_k} | \boldsymbol{\mu}_{\theta_k}, \boldsymbol{\delta}_{\theta_k}) \pi(\boldsymbol{\mu}_{u_k}) \pi(\boldsymbol{\delta}_{u_k}) \pi(\boldsymbol{\mu}_{\theta_k}) \pi(\boldsymbol{\delta}_{\theta_k})}^{\textit{Prior Density Function}}.$$

When there is not much data, the hierarchical model can be much more efficient by changing the geometric of the priors: since the normal distribution family is closed under translations and scaling, one can also consider non-centered parameterization, which generates normal individual parameters from the standard normal distribution [6] and then transform the parameter density distributions by scaling and shifting with latent parameter means $\boldsymbol{\mu}_{u_k}$ and $\boldsymbol{\mu}_{\theta_k}$,

$$u_k^{ncp} \sim N(0,1) \quad u_k = \mu_u + \delta_u u_k^{ncp},$$

$$\theta_k^{ncp} \sim N(0,1) \quad \theta_k = \mu_\theta + \delta_\theta \theta_k^{ncp},$$

Those non-centered variables can be regarded as the deviations of the individual behaviors relative to the latent parameter means. Non-centered parameterization improves MCMC exploration with weakly informed datasets by removing explicit hierarchical correlations to change the geometry of density distributions of hierarchical models and thus drastically increase the performance of Euclidean Hamiltonian Monte Carlo in the HMC algorithm.

The posterior density function under the non-center parameterization can now be written as

$$Pr(\boldsymbol{\theta}^{ncp}, \boldsymbol{\mu}^{ncp}, \boldsymbol{\mu_u}, \boldsymbol{\mu_\theta}, \boldsymbol{\delta_u}, \boldsymbol{\delta_\theta}|\boldsymbol{y}) \propto \overbrace{\prod_{k=1}^{M} p(y_k|\boldsymbol{\mu_{u_k}} + \boldsymbol{\delta_{u_k}}\boldsymbol{u}_k^{ncp}, \boldsymbol{\mu_{\theta_k}} + \boldsymbol{\delta}_{\theta_k}\boldsymbol{\theta}_k^{ncp})}^{Likelihood\ Function}$$
$$\overbrace{\prod_{k=1}^{M} p(\boldsymbol{u}_k^{ncp}|0,1)p(\boldsymbol{\theta}_k^{ncp}|0,1)\pi(\boldsymbol{\mu}_u)\pi(\boldsymbol{\delta}_{\boldsymbol{u}_k})\pi(\boldsymbol{\mu}_{\boldsymbol{\theta}_k})\pi(\boldsymbol{\delta}_{\boldsymbol{\theta}_k})}^{Prior\ Density\ Function}$$

Let the total number of data be M, then in general, centered parameterization achieves better performance with respect to MCMC mixing and enjoys larger effective sample sizes when one has a large and informative dataset with limited data heterogeneity (large M relative to δ_u and δ_θ). Non-center parameterization gets better performance when one has a weak informative dataset and large variability among data (small M relative to δ_u and δ_θ).

Parameter	Symbol	CI 2.5%	CI 50%	CI 97.5%
Chemical Reaction Model				
HSC Self-Renewal Rate	v_1	0.2114	0.401	0.7605
HSC Asym. Div. Rate	v_2	0.0639	0.336	1.7671
MPP Lumped Decay Rate	u_3	0.0405	0.2008	0.9946
MPP Feedback on HSC (e-4)	r_1	0.1466	0.35	0.8354
Hill Equation Model				
HSC Self-Renewal Prob.	p_0	0.5965	0.8194	0.9645
HSC Total Div. Rate	η_1	0.3623	0.737	1.4994
MPP Lumped Decay Rate	η_2	0.0405	0.2008	0.9946
MPP Feedback on HSC (e-5)	γ_1	2.0764	4.54	9.9121
Shared Parameters				
Biological std	δ_b	0.156	0.2	0.2564
Technical std	δ_t	0.078	0.1	0.1282
HSC Initial	μ_{HSC}	6309	10300	16814
MPP Initial	μ_{MPP}	3712	6000	9696

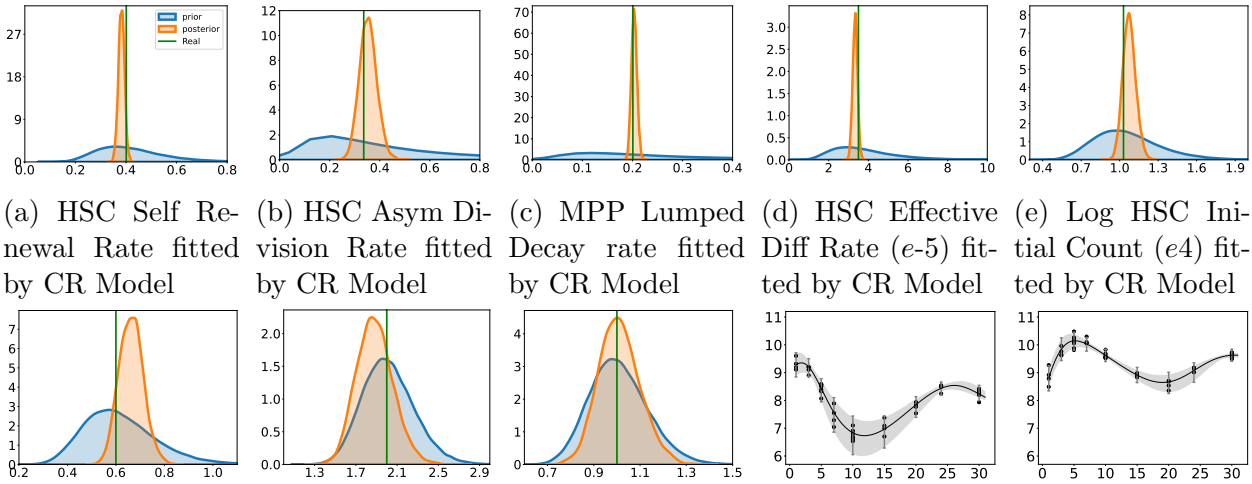
Table A.1: Priors parameter value ranges for non-hierarchical chemical reaction and Hill equation models

Parameter	Symbol	CI 2.5%	CI 50%	CI 97.5%
Chemical Reaction Model				
HSC Self-Renewal Rate	v_1	0.2201	0.401	0.7307
HSC Asym. Div. Rate	v_2	0.1635	0.336	0.6903
MPP Lumped Decay Rate	u_3	0.1102	0.2008	0.3659
MPP Feedback on HSC (e-4)	r_1	0.0449	0.0818	0.149
Hill Equation Model				
HSC Self-Renewal Prob.	p_0	0.6947	0.8171	0.9125
HSC Total Div. Rate	η_1	0.494	0.737	1.0995
MPP Lumped Decay Rate	η_2	0.1102	0.2008	0.3659
MPP Feedback on HSC (e-5)	γ_1	0.6771	1.6702	4.12
Shared Parameters				
Biological std	δ_b	0.156	0.2	0.2564
Technical std	δ_t	0.078	0.1	0.1282
HSC Initial	μ_{HSC}	5652	10300	18767
MPP Initial	μ_{MPP}	3293	6000	10932

Table A.2: Priors parameter value ranges for non-hierarchical chemical reaction and Hill equation models

A.0.3 Comparison between Hill Equation and Chemical Reaction Model

By generating synthetic datasets similar to the experimental dataset from both chemical reaction and Hill equation models and fitting by both models with similar prior distributions. We find the ground truth parameter can be correctly identified b both models in reasonable posterior distributions.



(a) HSC Self Renewal Rate fitted by CR Model

(b) HSC Asym Division Rate fitted by CR Model

(c) MPP Lumped Decay rate fitted by CR Model

(d) HSC Effective Diff Rate (e-5) fitted by CR Model

(e) Log HSC Initial Count (e4) fitted by CR Model

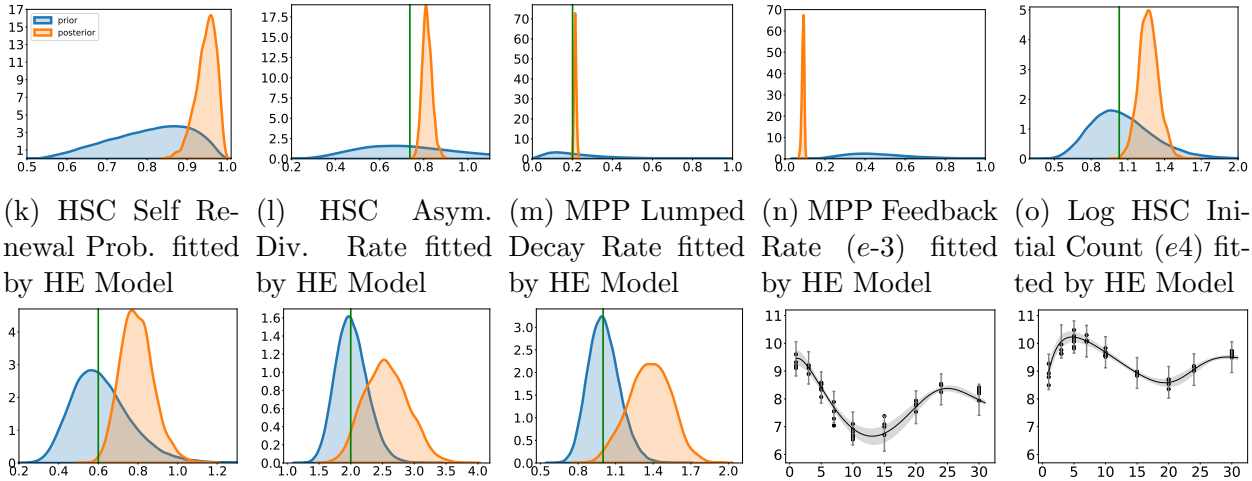
(f) Log MPP Initial Count (e4) fitted by CR Model

(g) Log Biological std (e-1) fitted by CR Model

(h) Log Technical std (e-1) fitted by CR Model

(i) DUP for CRSD HSC fitted by CR Model

(j) DUP for CRSD MPP fitted by CR Model



(k) HSC Self Renewal Prob. fitted by HE Model

(l) HSC Asym. Div. Rate fitted by HE Model

(m) MPP Lumped Decay Rate fitted by HE Model

(n) MPP Feedback Rate (e-3) fitted by HE Model

(o) Log HSC Initial Count (e4) fitted by HE Model

(p) Log MPP Initial Count (e4) fitted by HE Model

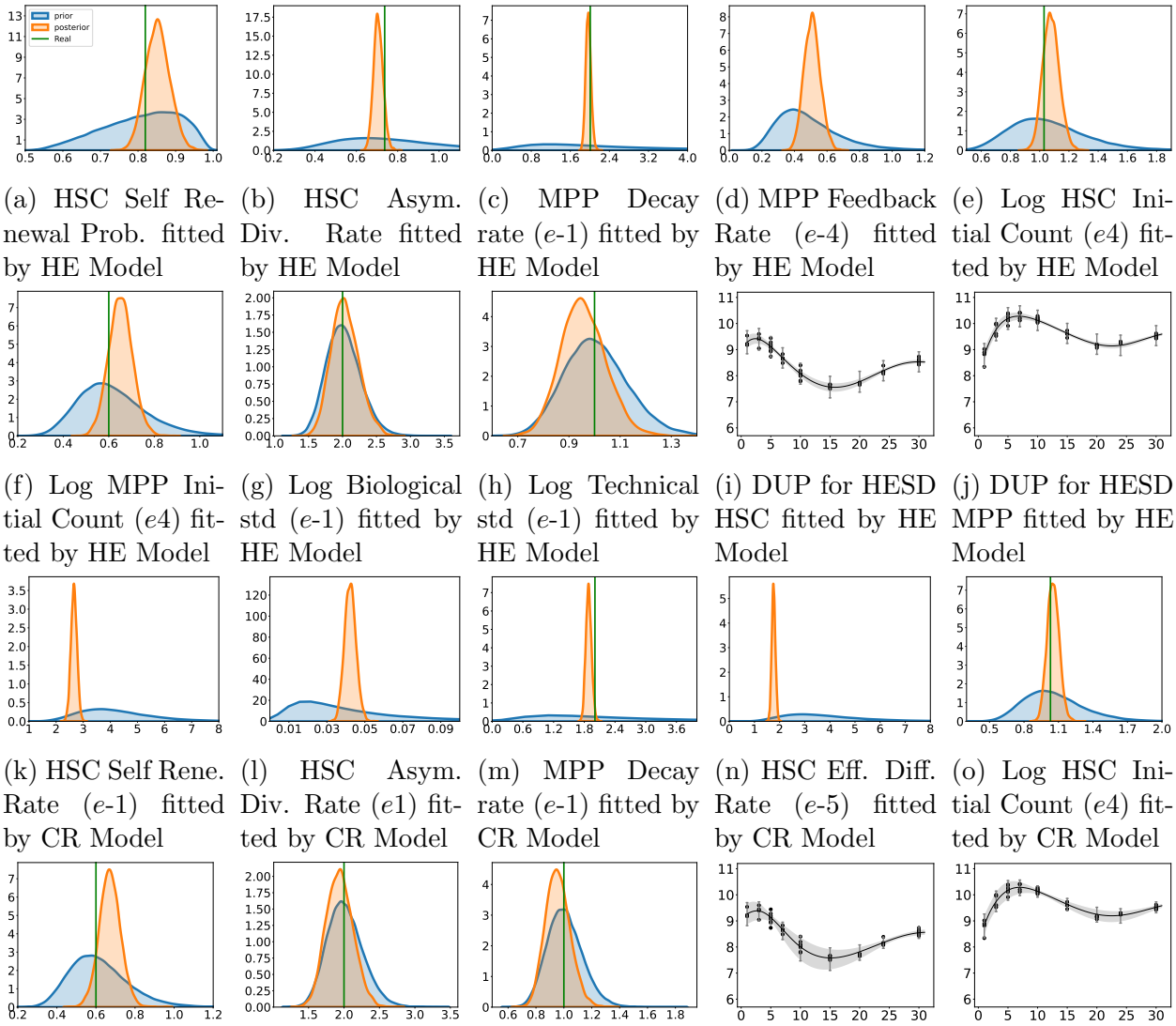
(q) Log Biological std (e-1) fitted by HE Model

(r) Log Technical std (e-1) fitted by HE Model

(s) DUP for CRSD HSC fitted by HE Model

(t) DUP for CRSD MPP fitted by HE Model

Figure A.2: Fit both chemical reaction (CR) and Hill equation (HE) model to the synthetic dataset generated by CR model (CRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-h) prior posterior plots for fitting the CR model to CRSD. (i-j) DUP for CRSD fitted by CR model. (k-r) prior posterior plots for fitting the HE model to CRSD. (s-t) DUP for CRSD fitted by HE model



(a) HSC Self Renewal Prob. fitted by HE Model

(b) HSC Asym. Div. Rate fitted by HE Model

(c) MPP Decay rate (e-1) fitted by HE Model

(d) MPP Feedback Rate (e-4) fitted by HE Model

(e) Log HSC Initial Count (e4) fitted by HE Model

(f) Log MPP Initial Count (e4) fitted by HE Model

(g) Log Biological std (e-1) fitted by HE Model

(h) Log Technical std (e-1) fitted by HE Model

(i) DUP for HESD HSC fitted by HE Model

(j) DUP for HESD MPP fitted by HE Model

(k) HSC Self Rene. Rate (e-1) fitted by CR Model

(l) HSC Asym. Div. Rate (e1) fitted by CR Model

(m) MPP Decay rate (e-1) fitted by CR Model

(n) HSC Eff. Diff. Rate (e-5) fitted by CR Model

(o) Log HSC Initial Count (e4) fitted by CR Model

(p) Log MPP Initial Count (e4) fitted by CR Model

(q) Log Biological std (e-1) fitted by CR Model

(r) Log Technical std (e-1) fitted by CR Model

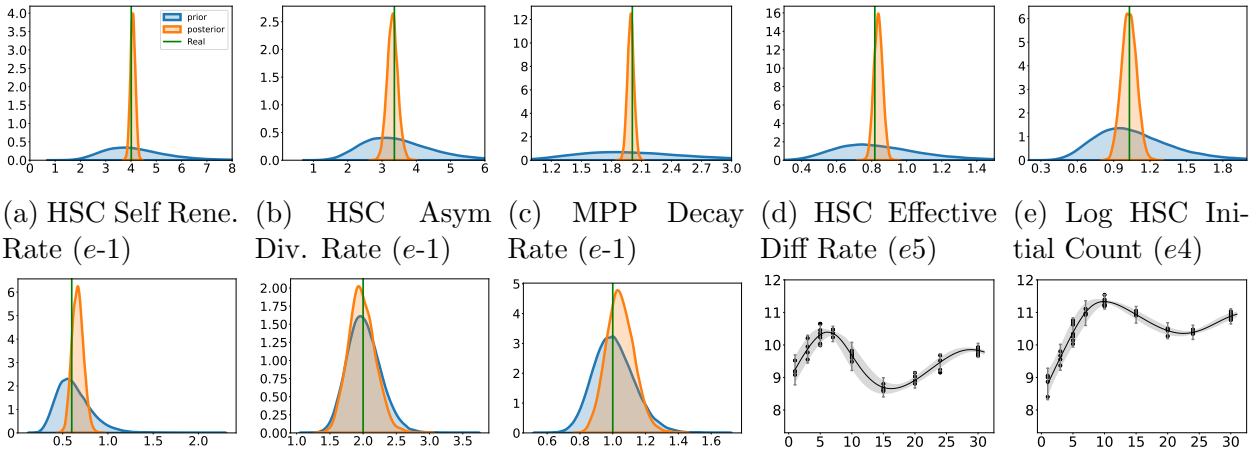
(s) DUP for HESD HSC fitted by CR Model

(t) DUP for HESD MPP fitted by CR Model

Figure A.3: Fit both chemical reaction (CR) and Hill equation (HE) model to the synthetic dataset generated by HE model (HESD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-h) prior posterior plots for fitting the HE model to HESD. (i-j) DUP for HESD fitted by HE model. (k-r) prior posterior plots for fitting the CR model to HESD. (s-t) DUP for HESD fitted by CR model

A.0.4 Comparison between Non-hierarchical Chemical Reaction ODE Model and Hierarchical Chemical Reaction ODE Model

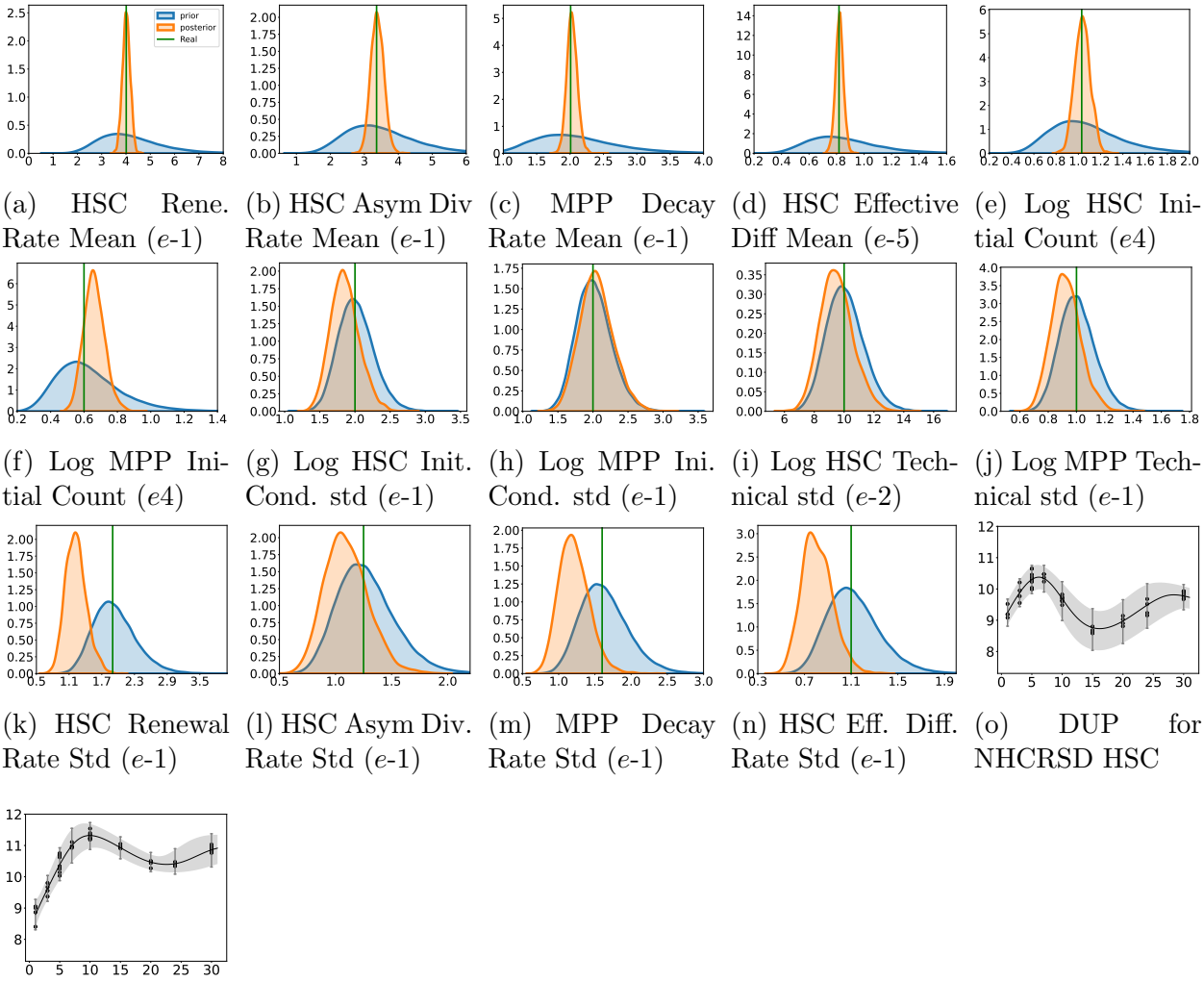
By generating synthetic datasets which mimic the real experimental dataset from both the non-hierarchical chemical reaction ODE Model and hierarchical chemical reaction ODE Model and fitting the synthetic data by both models with similar prior distributions. We found that both models can identify the ground truth parameter in reasonable posterior distributions using their synthetic data. Then, we get an insight into uncertainty classification by fitting a non-hierarchical chemical reaction ODE model with a hierarchical synthetic dataset and fitting a hierarchical model with a non-hierarchical synthetic dataset.



(a) HSC Self Rene. Rate (e-1) (b) HSC Asym Div. Rate (e-1) (c) MPP Decay Rate (e-1) (d) HSC Effective Diff Rate (e5) (e) Log HSC Initial Count (e4)

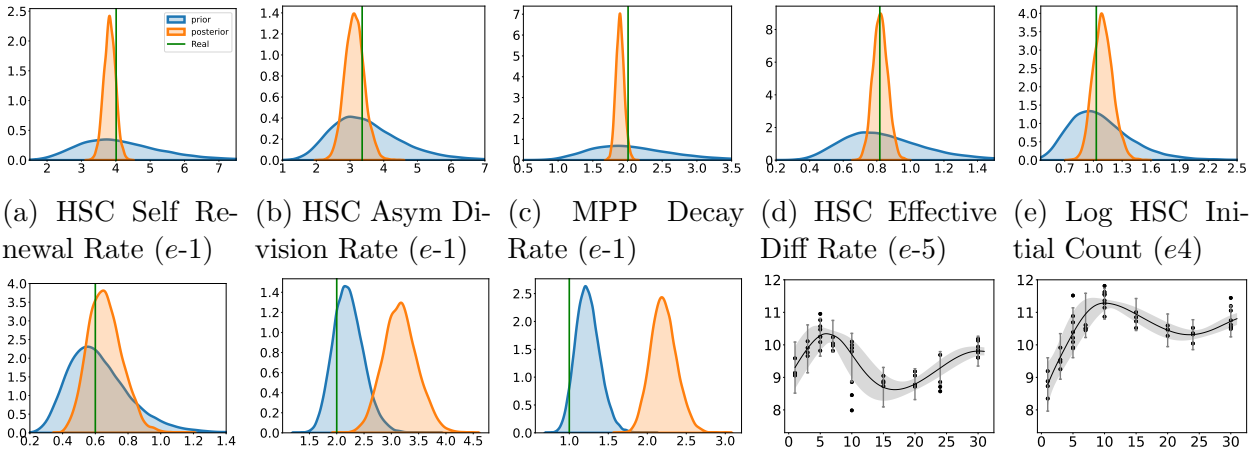
(f) Log MPP Initial Count (e4) (g) Log Init. Cond. std (e-1) (h) Log Technical std (e-1) (i) DUP for NHCRSD HSC (j) DUP for NHCRSD MPP

Figure A.4: Fit non-hierarchical chemical reaction (NHCR) model to the synthetic data generated by NHCR model (NHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-h) prior posterior plots for fitting the NHCR model to NHCRSD. (i-j) DUP for NHCRSD fitted by NHCR model.



(a) HSC Rene. Rate Mean (e-1)
(b) HSC Asym Div Rate Mean (e-1)
(c) MPP Decay Rate Mean (e-1)
(d) HSC Effective Diff Mean (e-5)
(e) Log HSC Initial Count (e4)
(f) Log MPP Initial Count (e4)
(g) Log HSC Init. Cond. std (e-1)
(h) Log MPP Ini. Cond. std (e-1)
(i) Log HSC Technical std (e-2)
(j) Log MPP Technical std (e-1)
(k) HSC Renewal Rate Std (e-1)
(l) HSC Asym Div. Rate Std (e-1)
(m) MPP Decay Rate Std (e-1)
(n) HSC Eff. Diff. Rate Std (e-1)
(o) DUP for NHCRSD HSC
(p) DUP for NHCRSD MPP

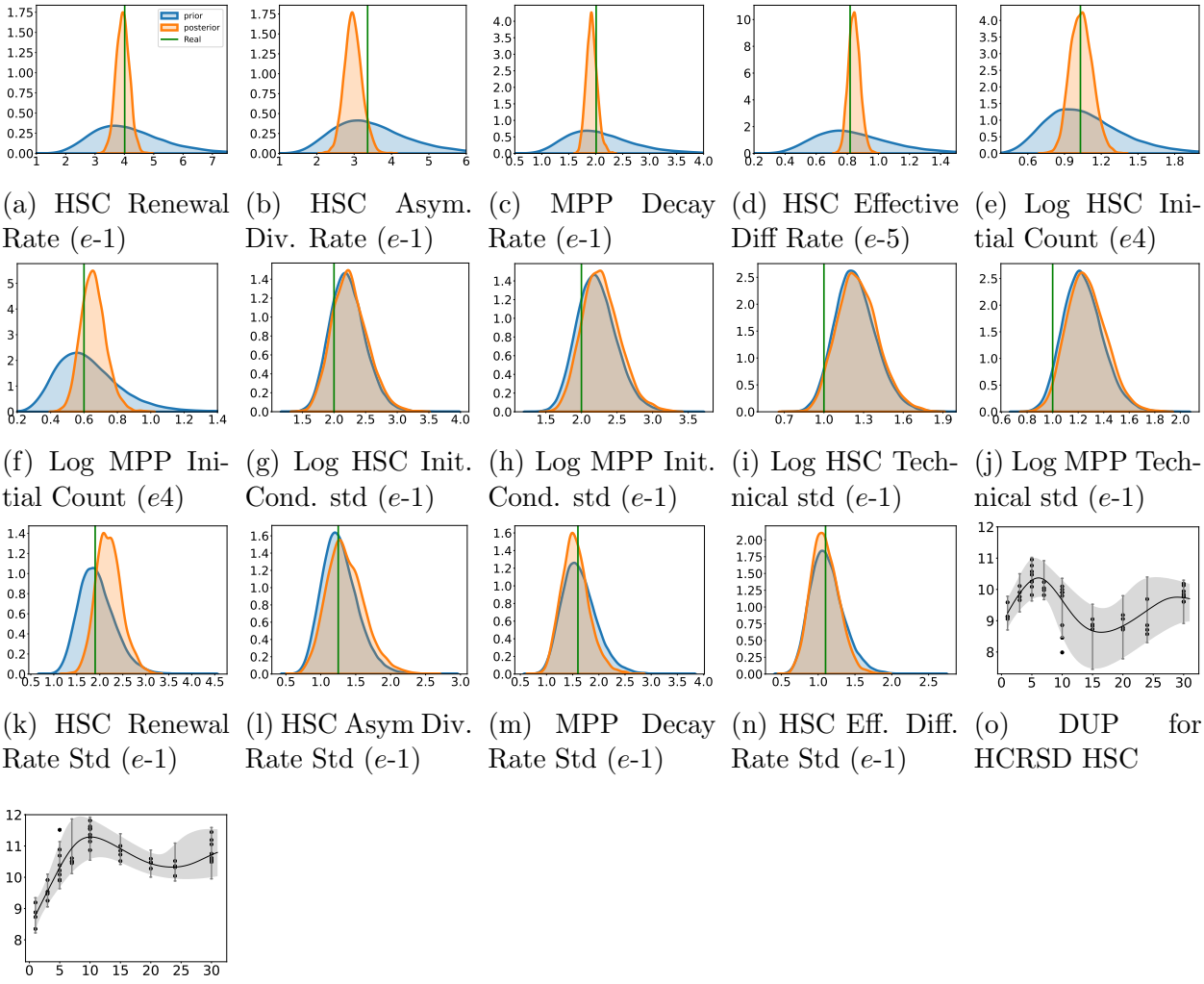
Figure A.5: Fit hierarchical chemical reaction (HCR) model to the synthetic data generated by non-hierarchical chemical reaction (NHCR) model (NHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-n) prior posterior plots for fitting the HCR model to NHCRSD. (o-p) DUP for NHCRSD fitted by HCR model.



(a) HSC Self Renewal Rate (e-1) (b) HSC Asym Division Rate (e-1) (c) MPP Decay Rate (e-1) (d) HSC Effective Diff Rate (e-5) (e) Log HSC Initial Count (e4)

(f) Log MPP Initial Count (e4) (g) Log Init. Cond. std (e-1) (h) Log Technical std (e-1) (i) DUP for HCRSD HSC (j) DUP for HCRSD MPP

Figure A.6: Fit non-hierarchical chemical reaction (NHCR) model to the synthetic data generated by hierarchical chemical reaction (HCR) model (HCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-h) prior posterior plots for fitting the NHCR model to HCRSD. (i-j) DUP for NHCRSD fitted by NHCR model.



(a) HSC Renewal Rate (e-1)

(b) HSC Asym. Div. Rate (e-1)

(c) MPP Decay Rate (e-1)

(d) HSC Effective Diff Rate (e-5)

(e) Log HSC Initial Count (e4)

(f) Log MPP Initial Count (e4)

(g) Log HSC Init. Cond. std (e-1)

(h) Log MPP Init. Cond. std (e-1)

(i) Log HSC Technical std (e-1)

(j) Log MPP Technical std (e-1)

(k) HSC Renewal Rate Std (e-1)

(l) HSC Asym Div. Rate Std (e-1)

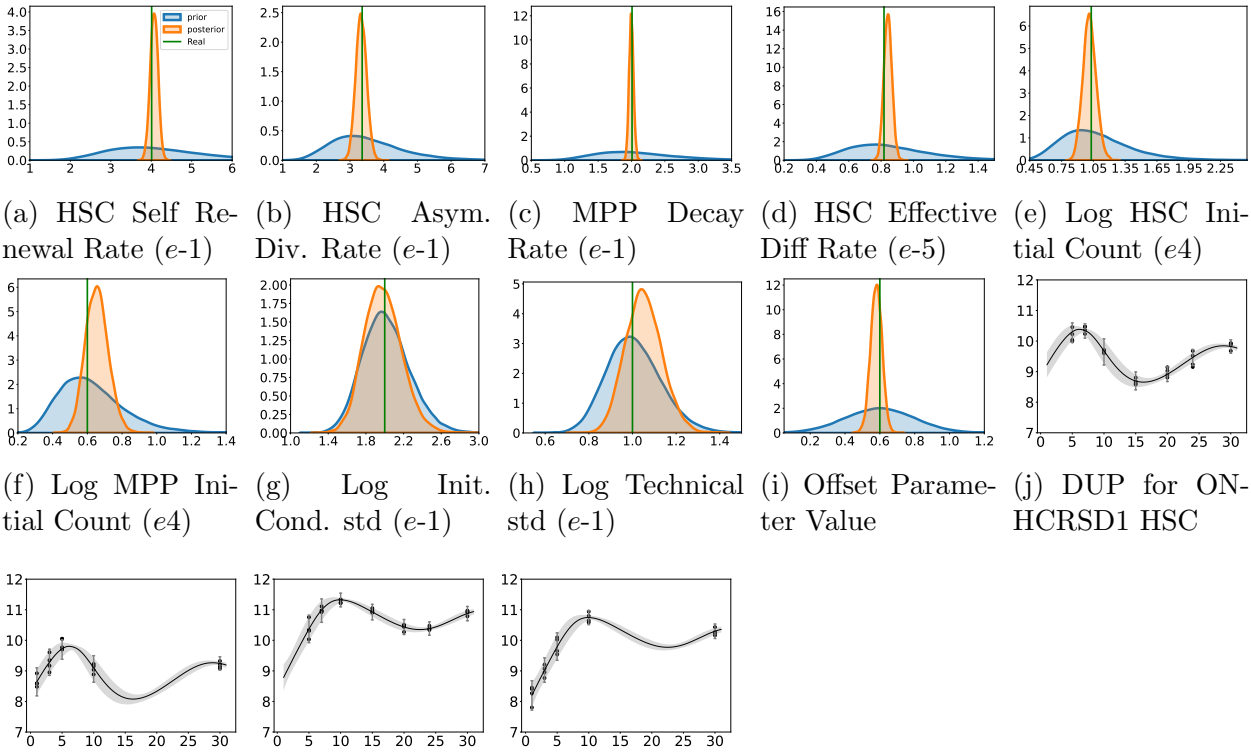
(m) MPP Decay Rate Std (e-1)

(n) HSC Eff. Diff. Rate Std (e-1)

(o) DUP for HCRSD HSC

(p) DUP for HCRSD HSC

Figure A.7: Fit hierarchical chemical reaction (HCR) model to the synthetic data generated HCR model (HCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-n) prior posterior plots for fitting the HCR model to HCRSD. (o-p) DUP for HCRSD fitted by HCR model.

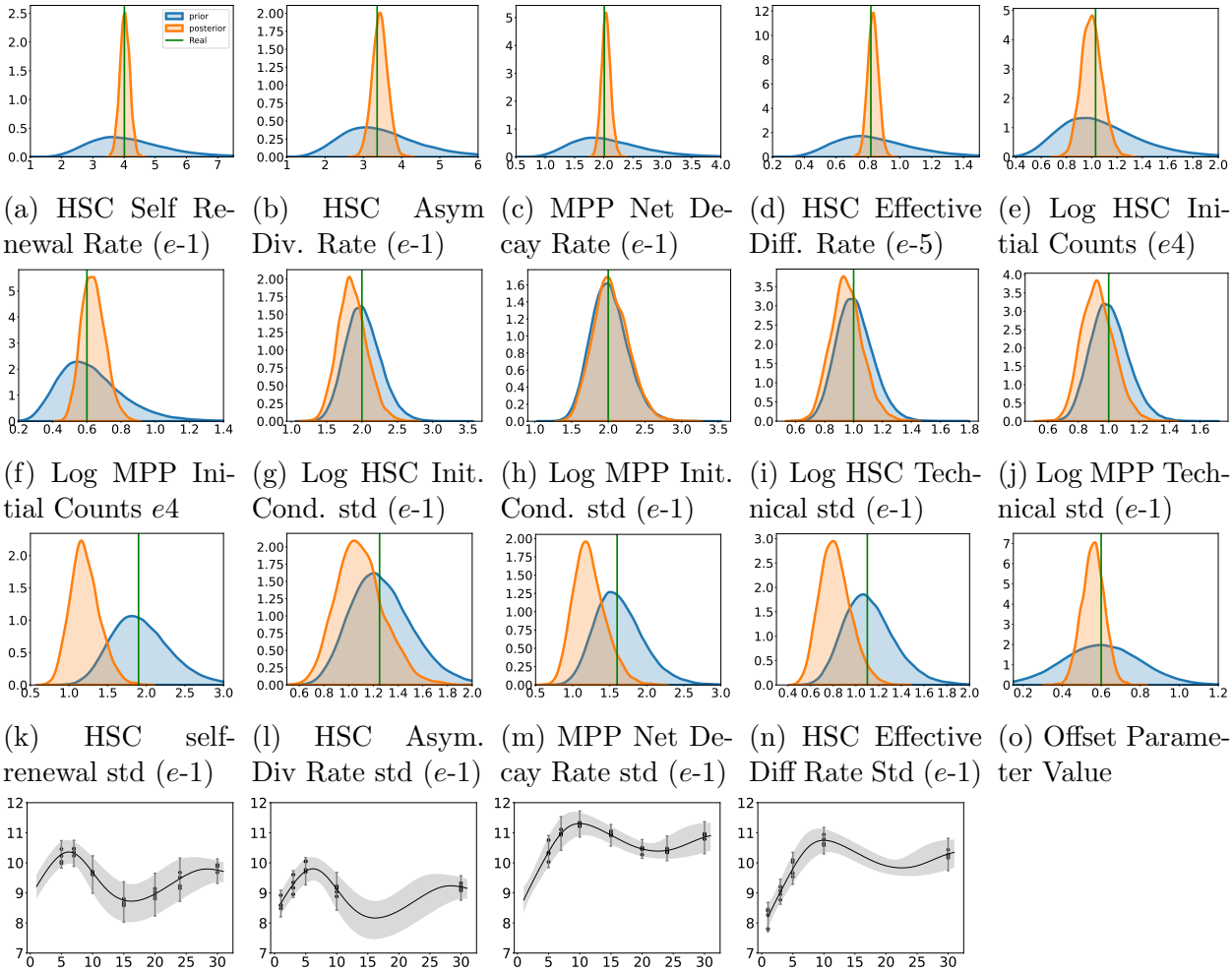


(a) HSC Self Renewal Rate (e-1) (b) HSC Asym. Div. Rate (e-1) (c) MPP Decay Rate (e-1) (d) HSC Effective Diff Rate (e-5) (e) Log HSC Initial Count (e4)

(f) Log MPP Initial Count (e4) (g) Log Init. Cond. std (e-1) (h) Log Technical std (e-1) (i) Offset Parameter Value (j) DUP for ONHCRSD1 HSC

(k) DUP for ONHCRSD1 MPP (l) DUP for ONHCRSD2 HSC (m) DUP for ONHCRSD2 MPP

Figure A.8: Fit the non-hierarchical chemical reaction (NHCR) model to the two synthetic datasets generated by the non-hierarchical chemical reaction (NHCR) model with an offset parameter between the log-scale cell counts of two synthetic datasets (ONHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the NHCR model to ONHCRSD. (j-m) DUP for ONHCRSD fitted by ONHCR model.



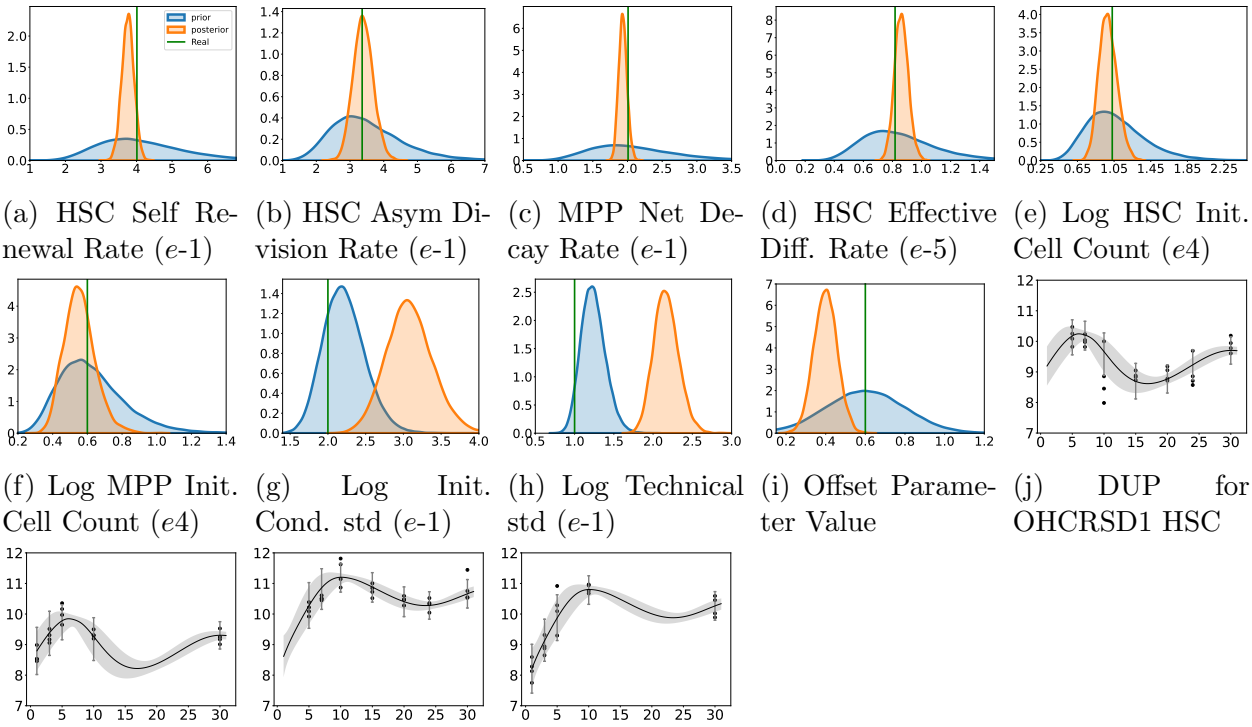
(a) HSC Self Renewal Rate (e-1) (b) HSC Asym Div. Rate (e-1) (c) MPP Net Decay Rate (e-1) (d) HSC Effective Diff. Rate (e-5) (e) Log HSC Initial Counts (e4)

(f) Log MPP Initial Counts e4 (g) Log HSC Init. Cond. std (e-1) (h) Log MPP Init. Cond. std (e-1) (i) Log HSC Technical std (e-1) (j) Log MPP Technical std (e-1)

(k) HSC self-renewal std (e-1) (l) HSC Asym. Div Rate std (e-1) (m) MPP Net Decay Rate std (e-1) (n) HSC Effective Diff Rate Std (e-1) (o) Offset Parameter Value

(p) DUP for ONHCRSD1 HSC (q) DUP for ONHCRSD1 MPP (r) DUP for ONHCRSD2 HSC (s) DUP for ONHCRSD2 MPP

Figure A.9: Fit the hierarchical chemical reaction (HCR) model to the two synthetic datasets generated by the non-hierarchical chemical reaction (NHCR) model with an offset parameter between the log-scale cell counts of two synthetic datasets (ONHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the HCR model to ONHCRSD. (j-m) DUP for OHCRSD fitted by HCR model.

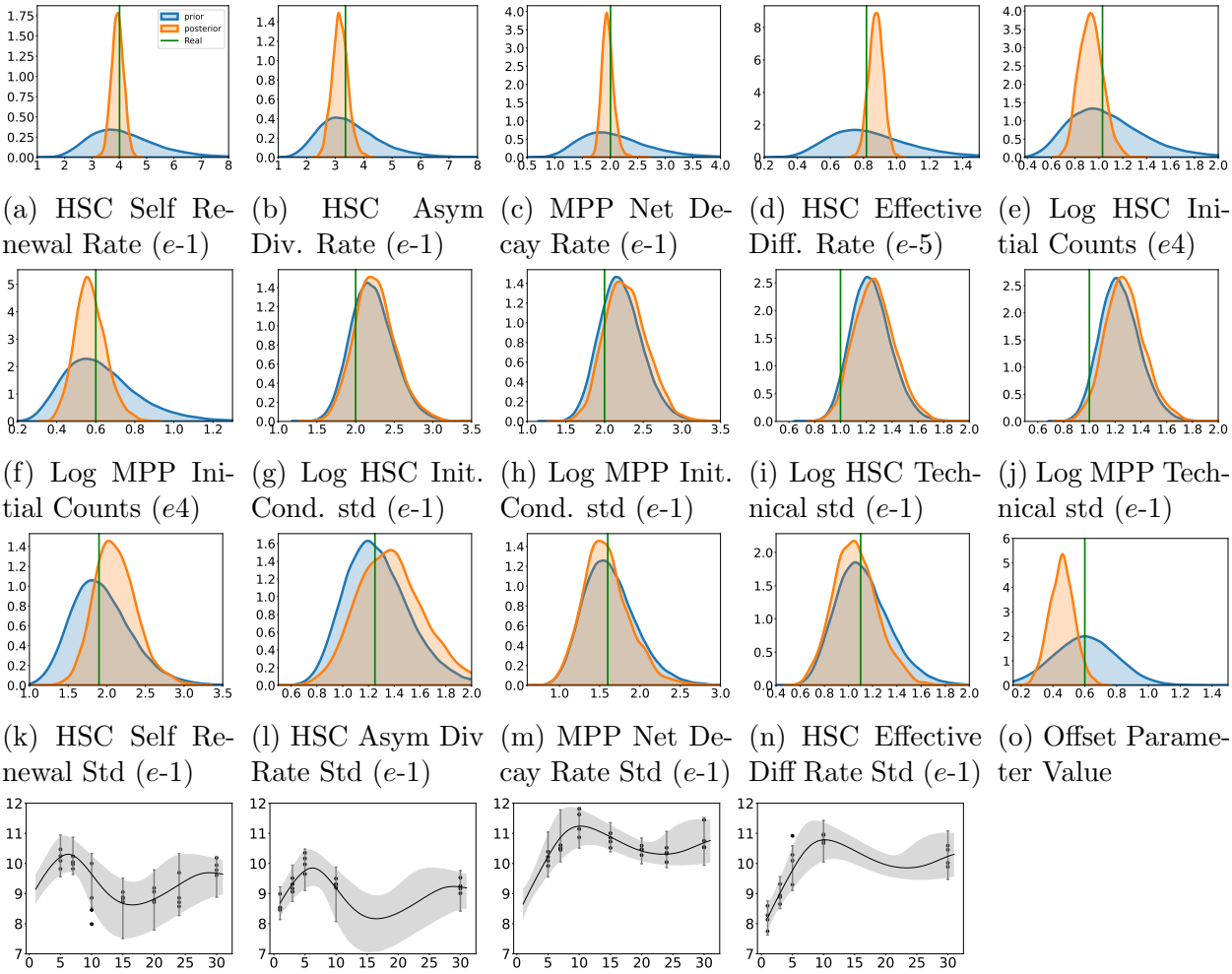


(a) HSC Self Renewal Rate (e-1) (b) HSC Asym Division Rate (e-1) (c) MPP Net Decay Rate (e-1) (d) HSC Effective Diff. Rate (e-5) (e) Log HSC Init. Cell Count (e4)

(f) Log MPP Init. Cell Count (e4) (g) Log Init. Cond. std (e-1) (h) Log Technical std (e-1) (i) Offset Parameter Value (j) DUP for OHCRSD1 HSC

(k) DUP for OHCRSD1 MPP (l) DUP for OHCRSD2 HSC (m) DUP for OHCRSD2 MPP

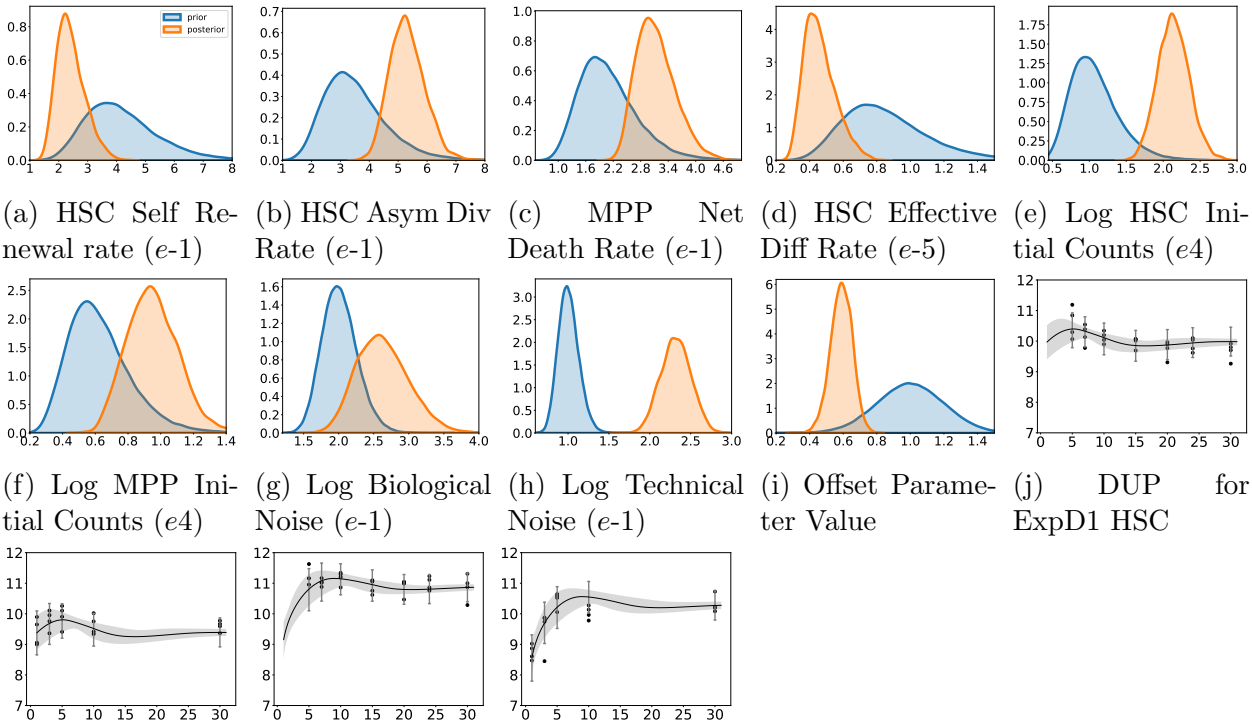
Figure A.10: Fit the non-hierarchical chemical reaction (NHCR) model to the two synthetic datasets generated by the hierarchical chemical reaction (HCR) model with an offset parameter between the log-scale cell counts of two synthetic datasets (OHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the NHCR model to OHCRSD. (j-m) DUP for ONHCRSD fitted by NHCR model.



(a) HSC Self Renewal Rate (e-1)
(b) HSC Asym Div. Rate (e-1)
(c) MPP Net Decay Rate (e-1)
(d) HSC Effective Diff. Rate (e-5)
(e) Log HSC Initial Counts (e4)
(f) Log MPP Initial Counts (e4)
(g) Log HSC Init. Cond. std (e-1)
(h) Log MPP Init. Cond. std (e-1)
(i) Log HSC Technical std (e-1)
(j) Log MPP Technical std (e-1)
(k) HSC Self Renewal Std (e-1)
(l) HSC Asym Div Rate Std (e-1)
(m) MPP Net Decay Rate Std (e-1)
(n) HSC Effective Diff Rate Std (e-1)
(o) Offset Parameter Value
(p) DUP for HCRSD1 HSC
(q) DUP for HCRSD1 MPP
(r) DUP for HCRSD2 HSC
(s) DUP for HCRSD2 MPP

Figure A.11: Fit the hierarchical chemical reaction (HCR) model to the two synthetic datasets generated by the hierarchical chemical reaction (HCR) model with an offset parameter between the log-scale cell counts of two synthetic datasets (OHCRSD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the HCR model to OHCRSD. (j-m) DUP for OHCRSD fitted by HCR model.

A.0.5 Fit Both Non-hierarchical and Hierarchical Chemical Reaction Models to Experimental Dataset

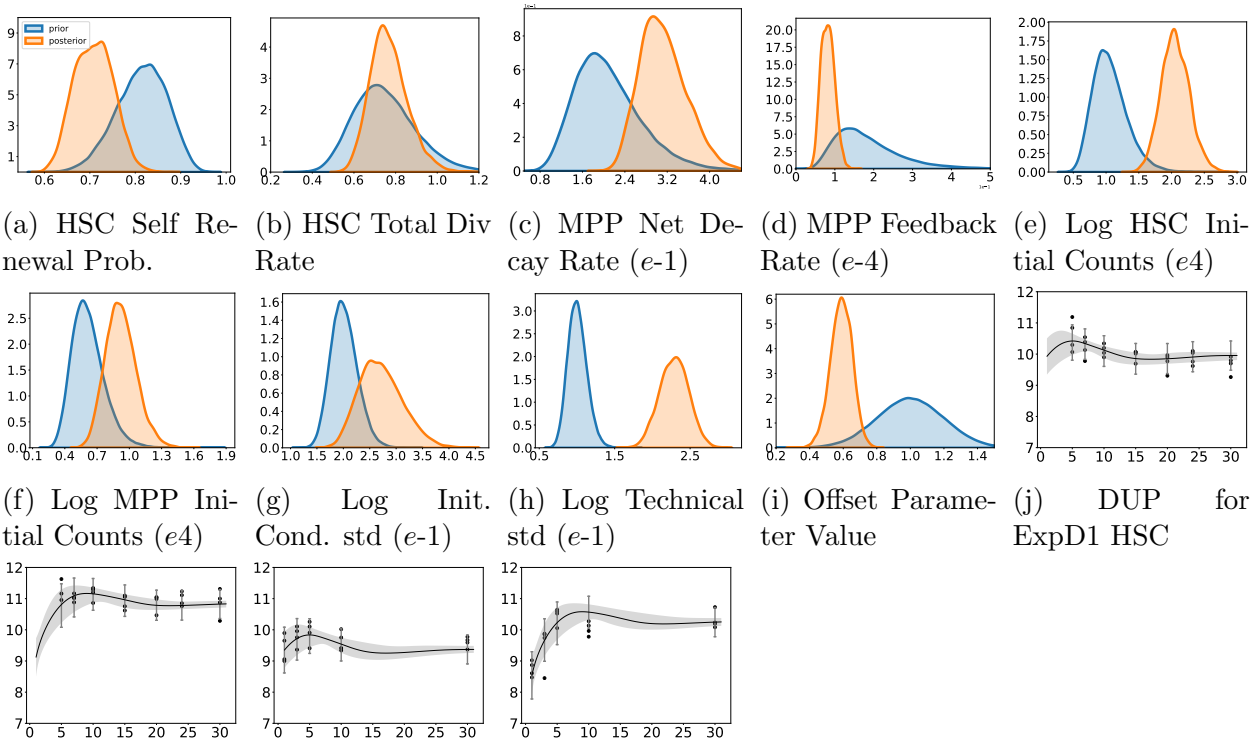


(a) HSC Self Renewal rate (e-1) (b) HSC Asym Div Rate (e-1) (c) MPP Net Death Rate (e-1) (d) HSC Effective Diff Rate (e-5) (e) Log HSC Initial Counts (e4)

(f) Log MPP Initial Counts (e4) (g) Log Biological Noise (e-1) (h) Log Technical Noise (e-1) (i) Offset Parameter Value (j) DUP for ExpD1 HSC

(k) DUP for ExpD1 MPP (l) DUP for ExpD2 HSC (m) DUP for ExpD2 MPP

Figure A.12: Fit non-hierarchical chemical reaction model (NHCR) to experimental datasets (ExpD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the NHCR model to the experimental datasets. (j-m) DUP for ExpD fitted by NHCR model.

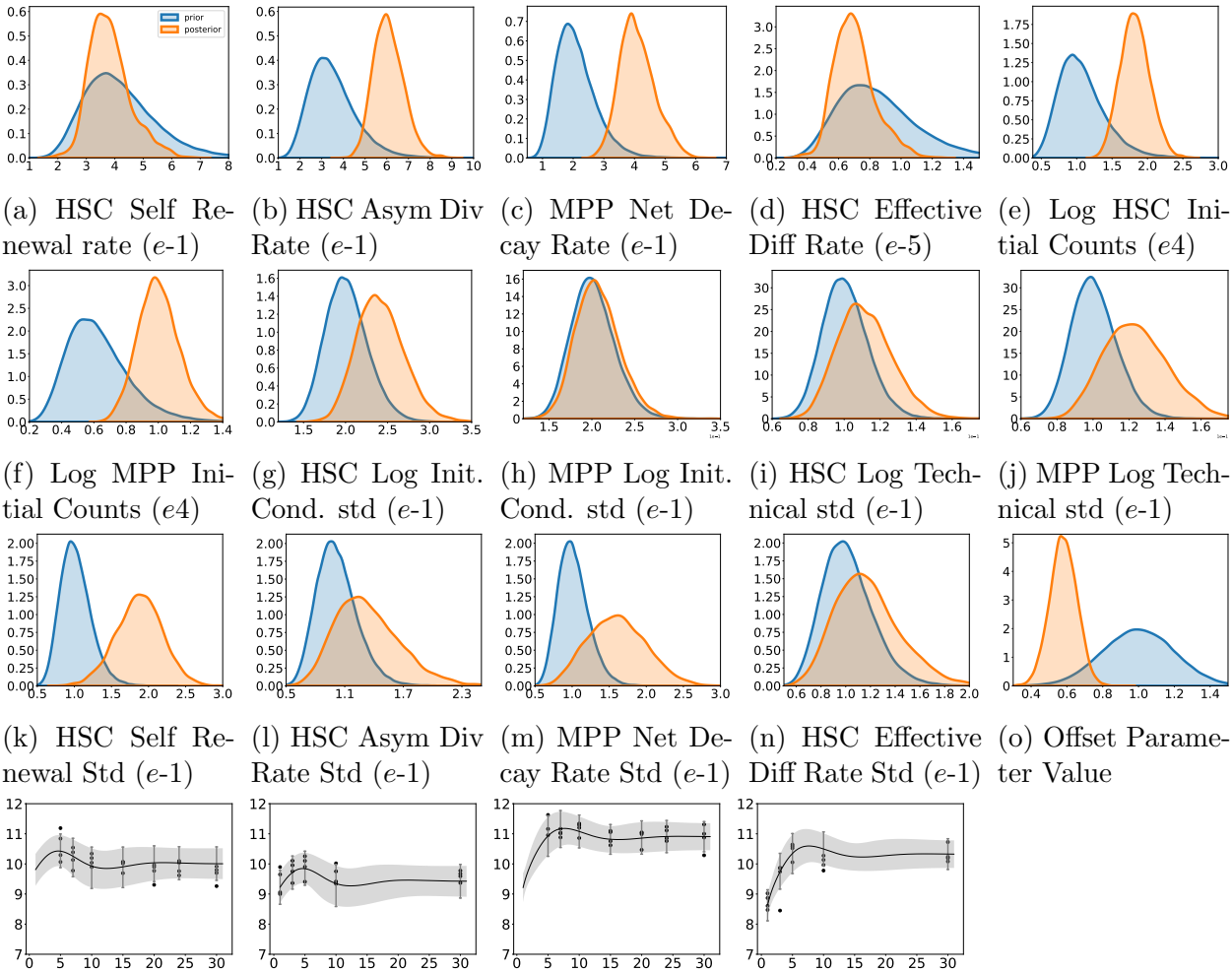


(a) HSC Self Renewal Prob. (b) HSC Total Div Rate (c) MPP Net Decay Rate (e-1) (d) MPP Feedback Rate (e-4) (e) Log HSC Initial Counts (e4)

(f) Log MPP Initial Counts (e4) (g) Log Init. Cond. std (e-1) (h) Log Technical std (e-1) (i) Offset Parameter Value (j) DUP for ExpD1 HSC

(k) DUP for ExpD1 MPP (l) DUP for ExpD2 HSC (m) DUP for ExpD2 MPP

Figure A.13: Fit non-hierarchical Hill equation (NHHE) model to experimental Datasets (ExpD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the HECR model to the ExpD. (j-m) DUP for ExpD fitted by NHHE model.



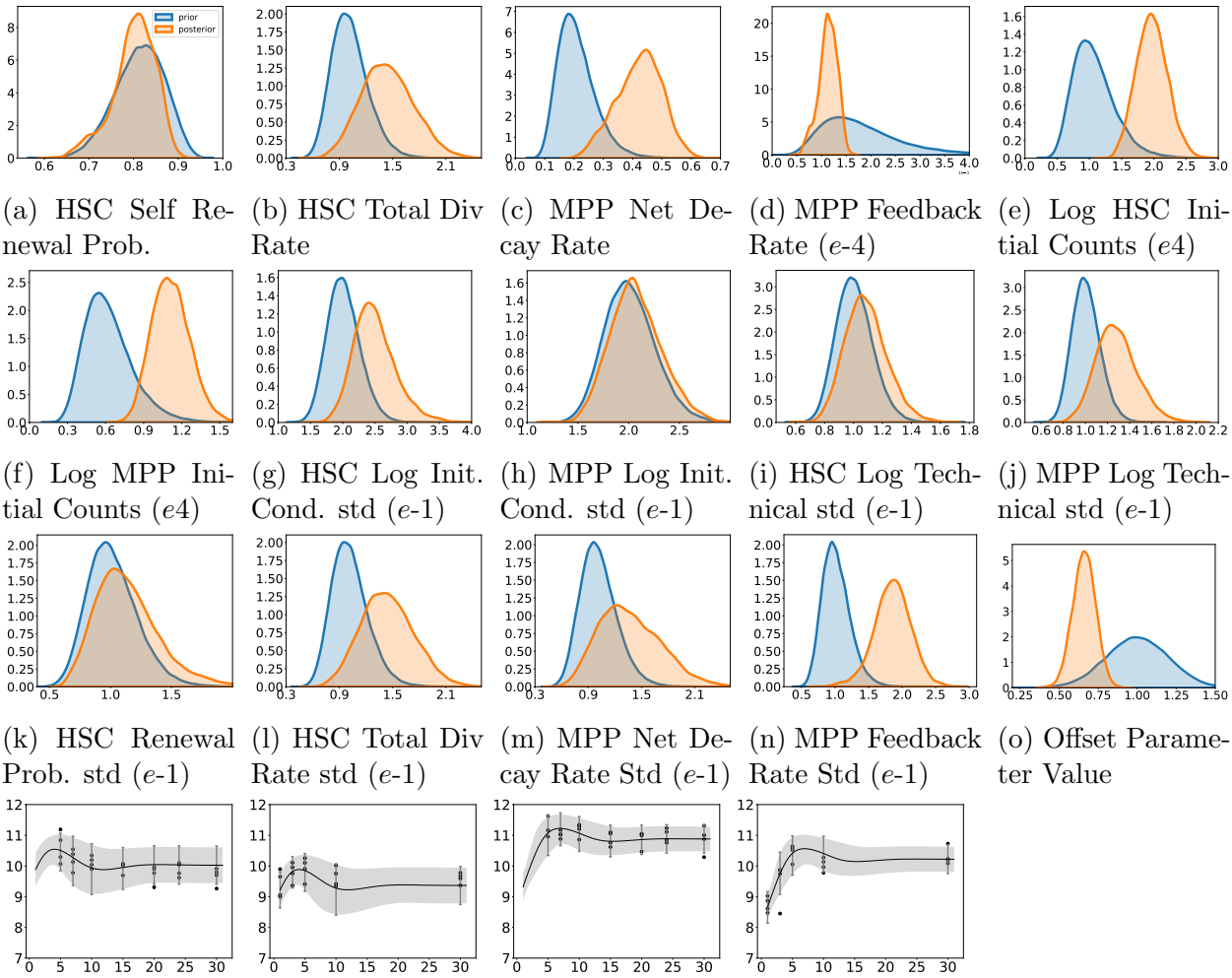
(a) HSC Self Renewal rate (e-1) (b) HSC Asym Div Rate (e-1) (c) MPP Net Decay Rate (e-1) (d) HSC Effective Diff Rate (e-5) (e) Log HSC Initial Counts (e4)

(f) Log MPP Initial Counts (e4) (g) HSC Log Init. Cond. std (e-1) (h) MPP Log Init. Cond. std (e-1) (i) HSC Log Technical std (e-1) (j) MPP Log Technical std (e-1)

(k) HSC Self Renewal Std (e-1) (l) HSC Asym Div Rate Std (e-1) (m) MPP Net Decay Rate Std (e-1) (n) HSC Effective Diff Rate Std (e-1) (o) Offset Parameter Value

(p) DUP for ExpD1 HSC (q) DUP for ExpD1 MPP (r) DUP for ExpD2 HSC (s) DUP for ExpD2 MPP

Figure A.14: Fit the hierarchical chemical reaction (HCR) model to the experimental datasets (ExpD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the HCR model to ExpD. (j-m) DUP for ExpD fitted by HCR model.



(a) HSC Self Renewal Prob. (b) HSC Total Div Rate (c) MPP Net Decay Rate (d) MPP Feedback Rate (e-4) (e) Log HSC Initial Counts (e4)

(f) Log MPP Initial Counts (e4) (g) HSC Log Init. Cond. std (e-1) (h) MPP Log Init. Cond. std (e-1) (i) HSC Log Technical std (e-1) (j) MPP Log Technical std (e-1)

(k) HSC Renewal Prob. std (e-1) (l) HSC Total Div Rate std (e-1) (m) MPP Net Decay Rate Std (e-1) (n) MPP Feedback Rate Std (e-1) (o) Offset Parameter Value

(p) DUP for ExpD1 HSC (q) DUP for ExpD1 MPP (r) DUP for ExpD2 HSC (s) DUP for ExpD2 MPP

Figure A.15: Fit the hierarchical Hill equation (HHE) model to the experimental datasets (ExpD). Blue areas: prior distributions. Orange areas: posterior distributions. Green verticle line: ground truth parameter values. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-i) prior posterior plots for fitting the HHE model to ExpD. (j-m) DUP for ExpD fitted by HHE model.

A.0.6 10-fold Cross Validation Report for Both Non-hierarchical and Hierarchical Models

The median absolute error (MAE) measures the typical difference between the observed Y_i and their posterior predictive means $Y_i^{'}$. The scaled median absolute error measures the typical standard deviations that the observed Y_i fall from their posterior predictive means $Y_i^{'}$. The within50 statistic measures the proportion of observed values that fall within their 50% posterior prediction interval. The within 95% statistic is similar but for 95% posterior prediction intervals.

$$MAE = median(|Y_1 - Y_1^{'}|, \cdots, |Y_N - Y_N^{'}|),$$

where N is the total number of observation mice. The scalded version of MAE, MASE then can be written as

$$MASE = median\left(\frac{|Y_1 - Y_1^{'}|}{\delta_1}, \cdots, \frac{|Y_N - Y_N^{'}|}{\delta_N}\right)$$

where $\delta_1, \cdots, \delta_N$ are the stand deviation of posterior predictive intervals at observation times for mice $1, \cdots, N$

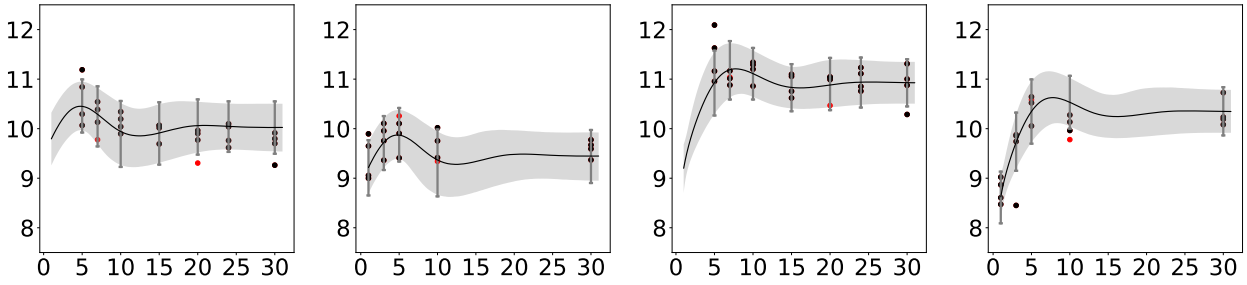
Hierarchical Model Posterior Predictive Summaries				
Fold	MAE	MASE	within 50	within 95
1 HSC	0.443	1.517	25%	75%
2 HSC	0.576	1.740	0%	50%
3 HSC	0.393	1.30	25%	75%
4 HSC	0.248	0.834	50%	100%
5 HSC	0.236	0.792	25%	100%
6 HSC	0.298	0.951	50%	100%
7 HSC	0.223	0.625	50%	100%
8 HSC	0.062	0.204	100%	100%
9 HSC	0.351	1.179	25%	100%
10 HSC	0.219	0.706	50%	100%
average HSC	0.304	0.983	40%	92.5%
1 MPP	0.329	1.14	25%	75%
2 MPP	0.245	0.843	50%	100%
3 MPP	0.254	0.891	25%	75%
4 MPP	0.214	0.658	50%	100%
5 MPP	0.164	0.598	50%	100%
6 MPP	0.258	0.949	0%	75%
7 MPP	0.164	0.576	75%	100%
8 MPP	0.275	0.990	25%	75%
9 MPP	0.325	1.10	25%	75%
10 MPP	0.123	0.483	50%	100%
average MPP	0.295	0.814	37.5%	87.5%
delete 5 HSC	0.387	1.512	25%	50%
delete 5 MPP	0.881	2.812	0%	25%
delete 7 HSC	0.387	1.282	25%	75%
delete 7 MPP	0.385	1.071	50%	75%

Table A.3: 10-Fold cross-validation with Hierarchical Model Posterior Predictive Summaries. The median absolute error (MAE) measures the typical difference between the observed Y_i and their posterior predictive means $Y_i^{'}$. The scaled median absolute error measures the typical number of standard deviations that the observed Y_i fall from their posterior predictive means $Y_i^{'}$. The within50 statistic measures the proportion of observed values that fall within their 50% posterior prediction interval. The within95 statistic is similar but for 95% posterior prediction intervals.

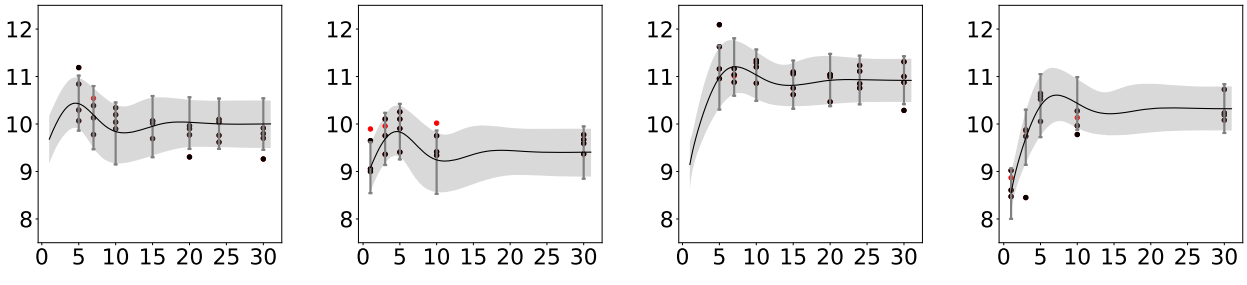
Non-Hierarchical Model Posterior Predictive Summaries				
Fold	MAE	MASE scaled	within 50	within 95
1 HSC	0.513	2.053	0%	50%
2 HSC	0.450	1.314	0%	75%
3 HSC	0.344	1.376	25%	75%
4 HSC	0.292	1.008	0%	100%
5 HSC	0.229	0.722	50%	100%
6 HSC	0.262	1.014	25%	100%
7 HSC	0.100	0.402	75%	100%
8 HSC	0.162	0.653	50%	100%
9 HSC	0.367	1.324	25%	75%
10 HSC	0.241	0.839	50%	100%
average HSC	0.243	1.070	30%	87.5%
1 MPP	0.374	1.358	25%	75%
2 MPP	0.287	0.809	50%	100%
3 MPP	0.284	1.012	25%	75%
4 MPP	0.189	0.587	75%	100%
5 MPP	0.155	0.521	75%	100%
6 MPP	0.299	1.199	25%	75%
7 MPP	0.175	0.629	50%	100%
8 MPP	0.366	0.265	25%	50%
9 MPP	0.311	0/976	25%	75%
10 MPP	0.179	0.668	50%	100%
average MPP	0.2277	0.822	37.5%	87.5%
delete 5 HSC	0.387	1.340	50%	50%
delete 5 MPP	0.894	2.697	0%	25%
delete 7 HSC	0.387	1.321	25%	75%
delete 7 MPP	0.602	1.667	25%	50%

Table A.4: 10-Fold cross validation with Hierarchical Model Posterior Predictive Summaries the median absolute error (MAE) measures the typical difference between the observed Y_i and their posterior predictive means $Y_i^{'}$. The scaled median absolute error measures the typical number of standard deviations that the observed Y_i fall from their posterior predictive means $Y_i^{'}$. The within50 statistic measures the proportion of observed values that fall within their 50% posterior prediction interval. The within95 statistic is similar, but for 95% posterior prediction intervals.

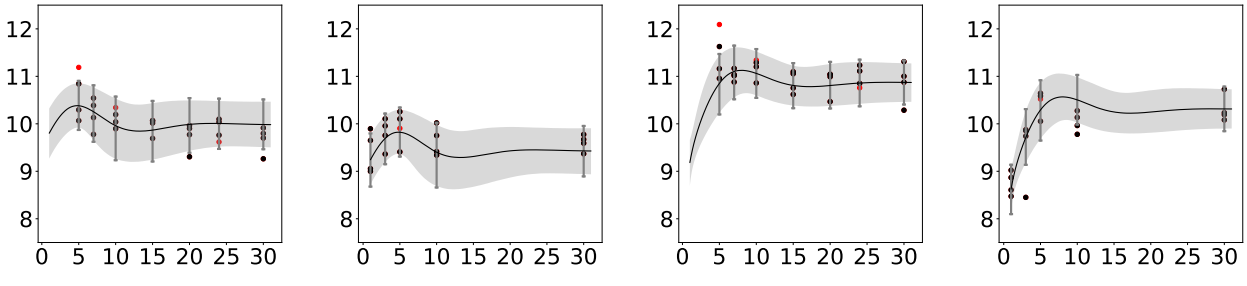
Figure A.16: Dynamic uncertainty plots for hierarchical chemical reaction (HCR) model 10-fold cross-validation (CV) using the experimental dataset (ExpD), page(156-157). CVi: the ith fold training and validation dataset. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; red dot marks refer to the validation dataset; black dot marks refer to the partially observed training dataset; each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-t) DUP for ExpD fitted by HCR model using 10-fold cross-validation.



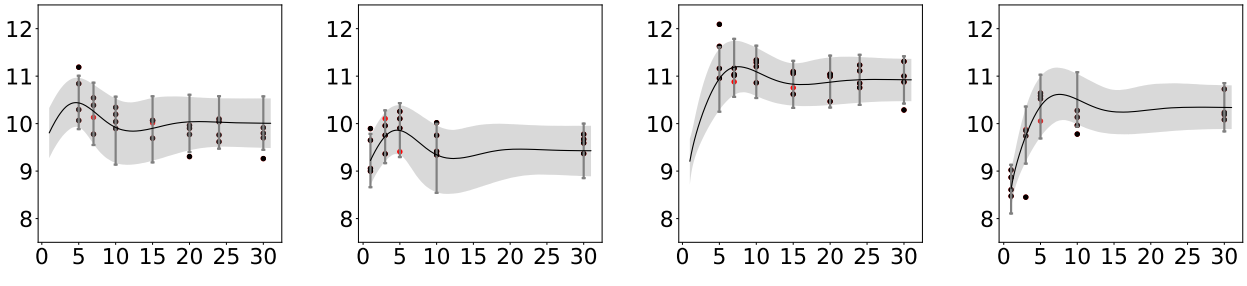
(a) CV1 ExpD1 HSC (b) CV1 ExpD2 HSC (c) CV1 ExpD1 MPP (d) CV1 ExpD2 MPP



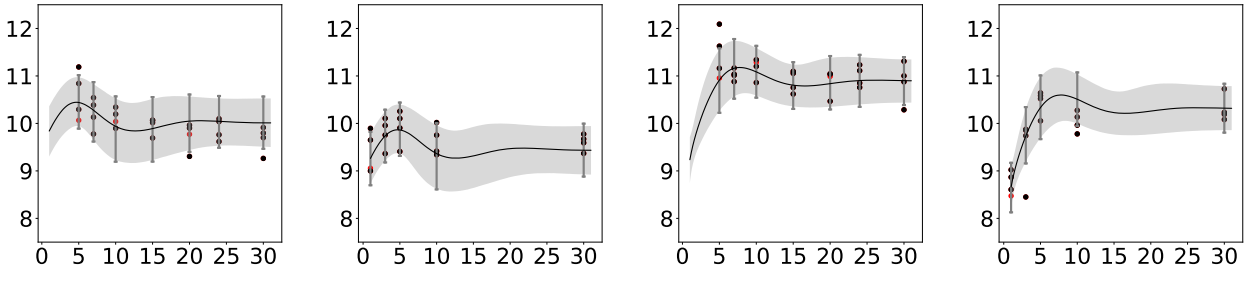
(e) CV2 ExpD1 HSC (f) CV2 ExpD2 HSC (g) CV2 ExpD1 MPP (h) CV2 ExpD2 MPP



(i) CV3 ExpD1 HSC (j) CV3 ExpD2 HSC (k) CV3 ExpD1 MPP (l) CV3 ExpD2 MPP

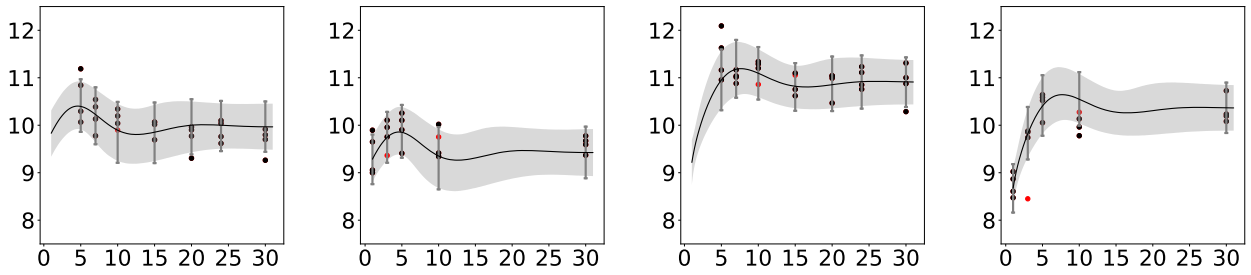


(m) CV4 ExpD1 HSC (n) CV4 ExpD2 HSC (o) CV4 ExpD1 MPP (p) CV4 ExpD2 MPP

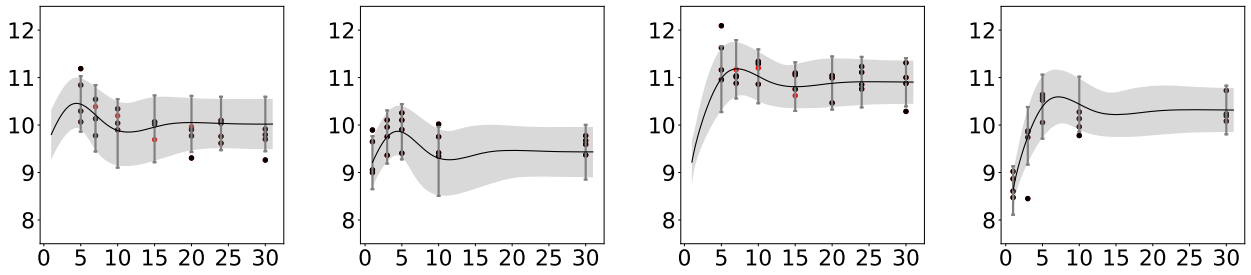


(q) CV5 ExpD1 HSC (r) CV5 ExpD2 HSC (s) CV5 ExpD1 MPP (t) CV5 ExpD2 MPP

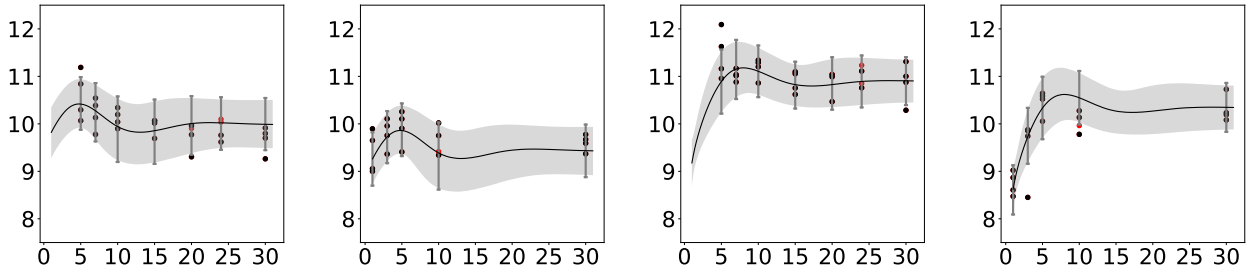
Figure A.17: Dynamic uncertainty plots for hierarchical chemical reaction (HCR) model 10-fold cross-validation (CV) using the experimental dataset (ExpD). Page (158-159) CVi: the ith fold training and validation dataset. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; red dot marks refer to the validation dataset; black dot marks refer to the partially observed training dataset; each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-t) DUP for ExpD fitted by HCR model using 10-fold cross-validation.



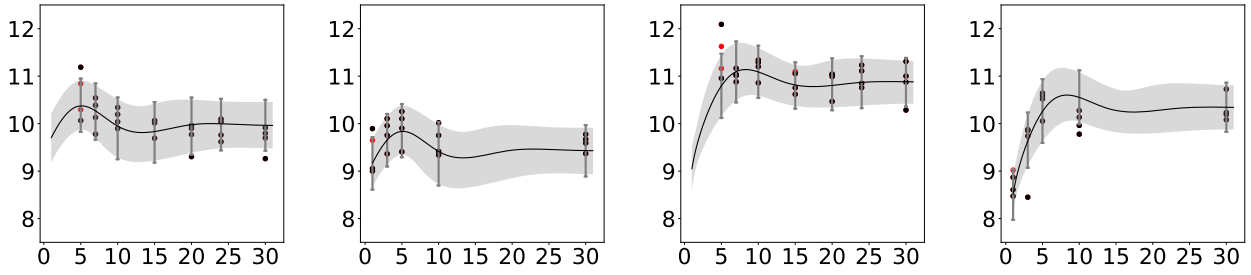
(a) CV6 ExpD1 HSC (b) CV6 ExpD2 HSC (c) CV6 ExpD1 MPP (d) CV6 ExpD2 MPP



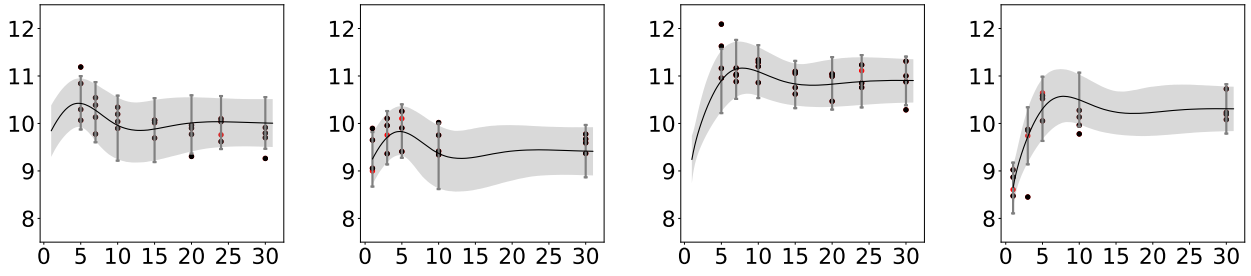
(e) CV7 ExpD1 HSC (f) CV7 ExpD2 HSC (g) CV7 ExpD1 MPP (h) CV7 ExpD2 MPP



(i) CV8 ExpD1 HSC (j) CV8 ExpD2 HSC (k) CV8 ExpD1 MPP (l) CV8 ExpD2 MPP

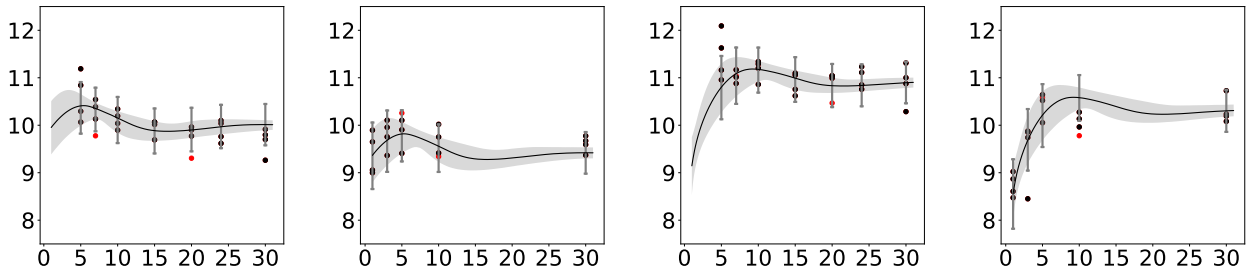


(m) CV9 ExpD1 HSC (n) CV9 ExpD2 HSC (o) CV9 ExpD1 MPP (p) CV9 ExpD2 MPP

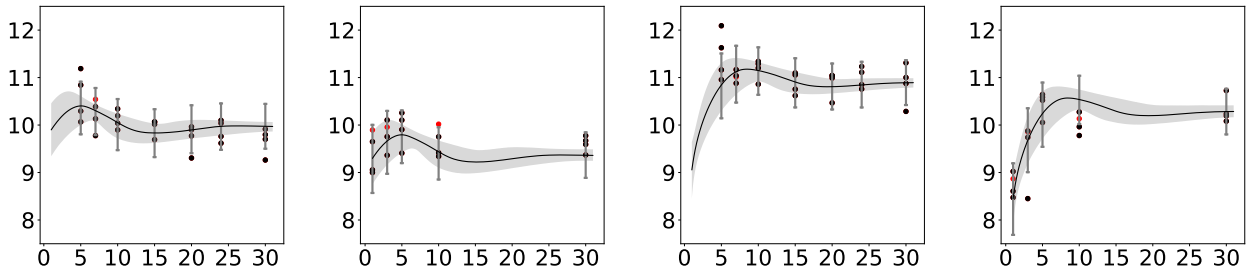


(q) CV10 ExpD1 HSC (r) CV10 ExpD2 HSC (s) CV10 ExpD1 MPP (t) CV10 ExpD2 MPP

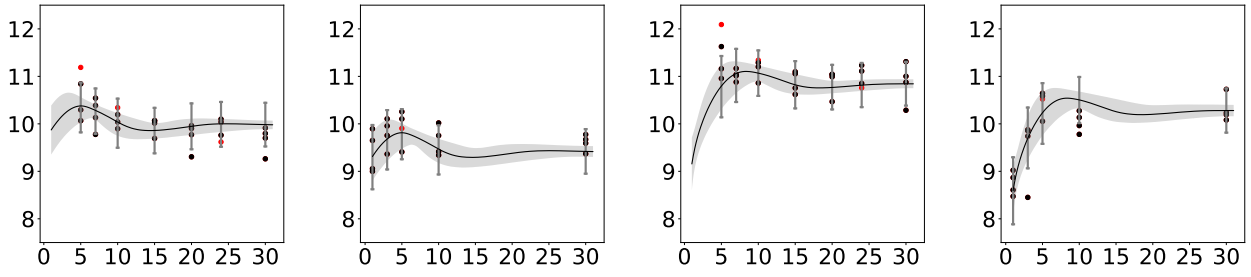
Figure A.18: Dynamic uncertainty plots for non-hierarchical chemical reaction (NHCR) model 10-fold cross-validation (CV) using the experimental dataset (ExpD). Page (160-161) CVi: the ith fold training and validation dataset. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; red dot marks refer to the validation dataset; black dot marks refer to the partially observed training dataset; each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-t) DUP for ExpD fitted by NHCR model using 10-fold cross-validation.



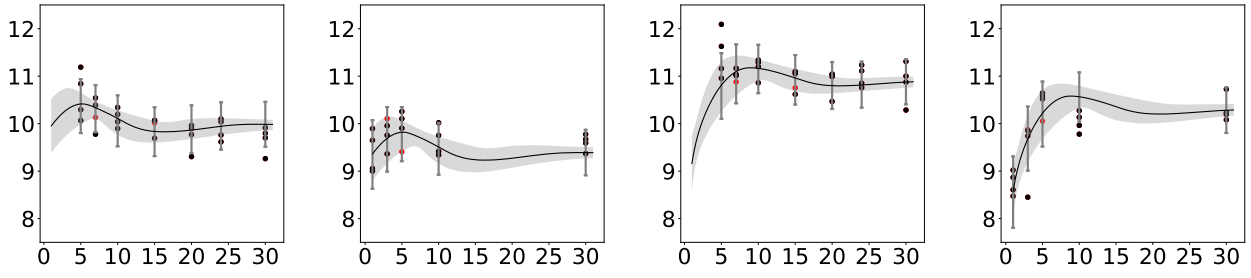
(a) CV1 ExpD1 HSC (b) CV1 ExpD2 HSC (c) CV1 ExpD1 MPP (d) CV1 ExpD2 MPP



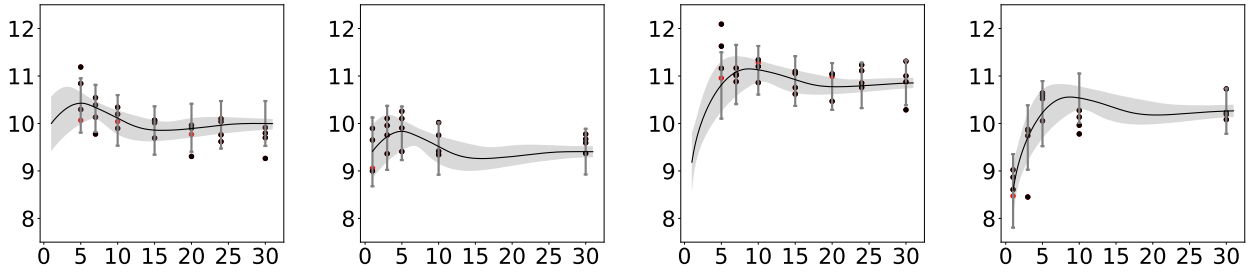
(e) CV2 ExpD1 HSC (f) CV2 ExpD2 HSC (g) CV2 ExpD1 MPP (h) CV2 ExpD2 MPP



(i) CV3 ExpD1 HSC (j) CV3 ExpD2 HSC (k) CV3 ExpD1 MPP (l) CV3 ExpD2 MPP

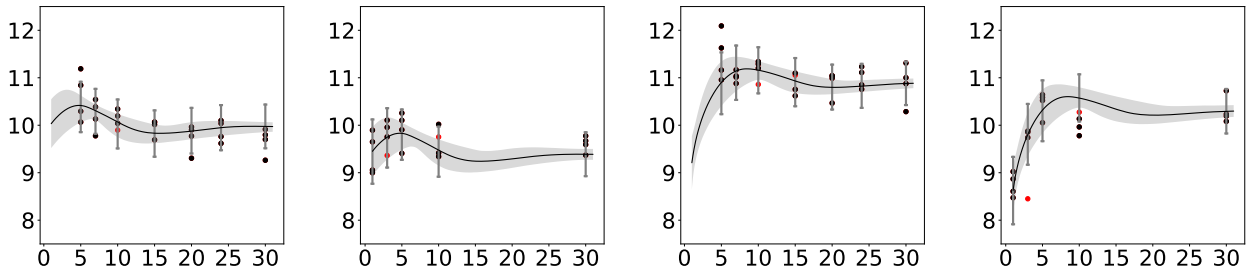


(m) CV4 ExpD1 HSC (n) CV4 ExpD2 HSC (o) CV4 ExpD1 MPP (p) CV4 ExpD2 MPP

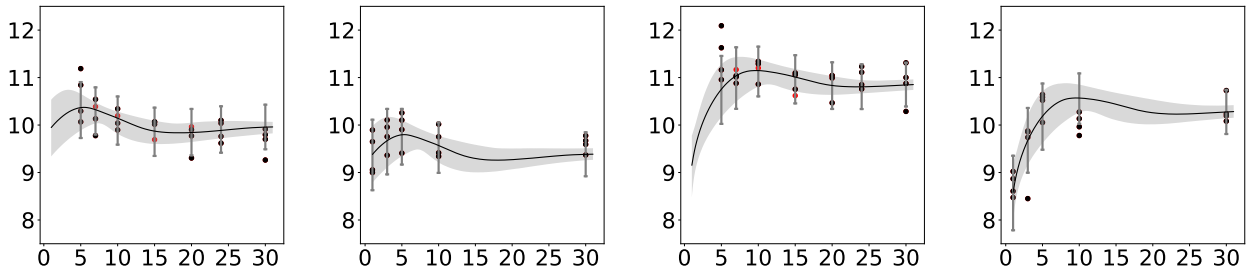


(q) CV5 ExpD1 HSC (r) CV5 ExpD2 HSC (s) CV5 ExpD1 MPP (t) CV5 ExpD2 MPP

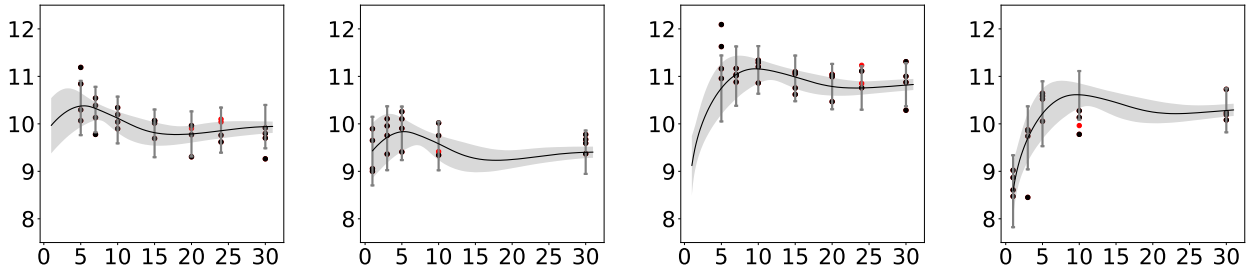
Figure A.19: Dynamic uncertainty plots for non-hierarchical chemical reaction (NHCR) model 10-fold cross-validation (CV) using the experimental dataset (ExpD). Page(162-163) CVi: the ith fold training and validation dataset. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; red dot marks refer to the validation dataset; black dot marks refer to the partially observed training dataset; each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-t) DUP for ExpD fitted by NHCR model using 10-fold cross-validation.



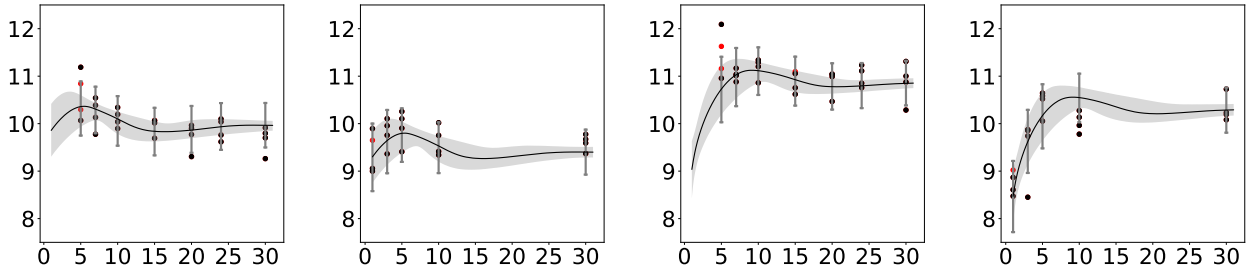
(a) CV6 ExpD1 HSC (b) CV6 ExpD2 HSC (c) CV6 ExpD1 MPP (d) CV6 ExpD2 MPP



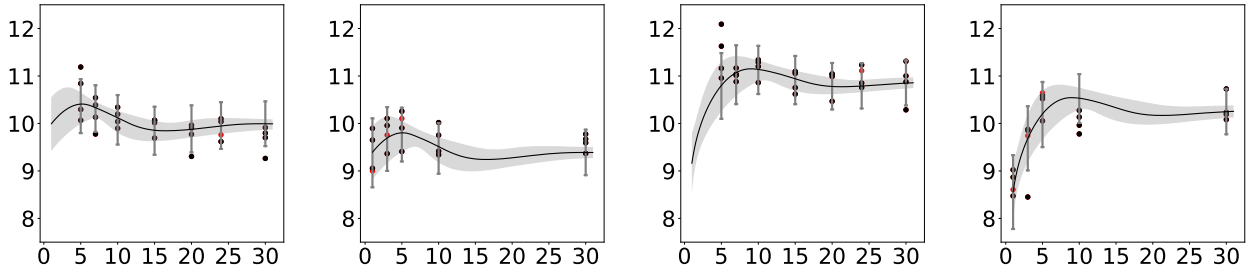
(e) CV7 ExpD1 HSC (f) CV7 ExpD2 HSC (g) CV7 ExpD1 MPP (h) CV7 ExpD2 MPP



(i) CV8 ExpD1 HSC (j) CV8 ExpD2 HSC (k) CV8 ExpD1 MPP (l) CV8 ExpD2 MPP

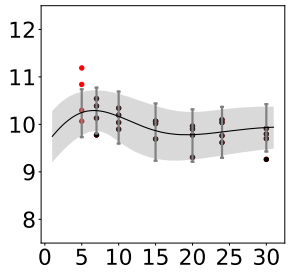


(m) CV9 ExpD1 HSC (n) CV9 ExpD2 HSC (o) CV9 ExpD1 MPP (p) CV9 ExpD2 MPP

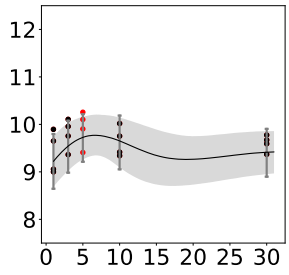


(q) CV10 ExpD1 HSC (r) CV10 ExpD2 HSC (s) CV10 ExpD1 MPP (t) CV10 ExpD2 MPP

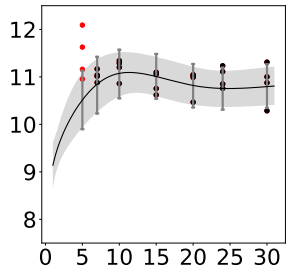
Figure A.20: Dynamic uncertainty plots for non-hierarchical/hierarchical chemical reaction (NHCR/HCR) model leave-out one day cross-validation (CV) using the experimental dataset (ExpD). Page (164-165). Di ExpD: training dataset is ExpD without data at day$i$, and the validation dataset is ExpD data at day$i$. Dynamic uncertainty plots (DUP): the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; red dot marks refer to the validation dataset; black dot marks refer to the partially observed training dataset; each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-h) DUP for ExpD fitted by the NHCR model. (i-p) fitted by the HCR model using leave-out a day cross-validation.



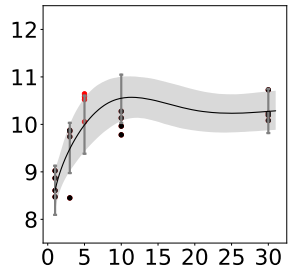
(a) Fit HCR Model to D5 ExpD1 HSC



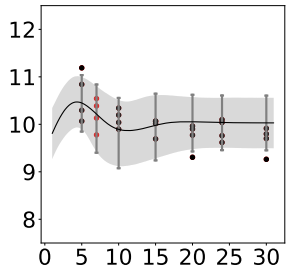
(b) Fit HCR Model to D5 ExpD2 HSC



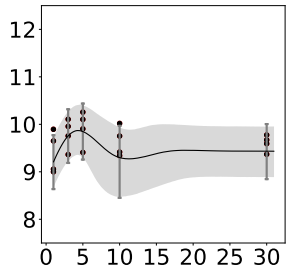
(c) Fit HCR Model to D5 ExpD1 MPP



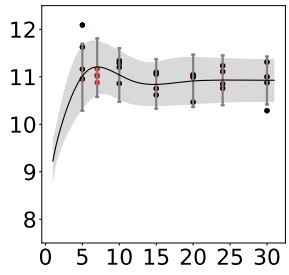
(d) Fit HCR Model to D5 ExpD2 MPP



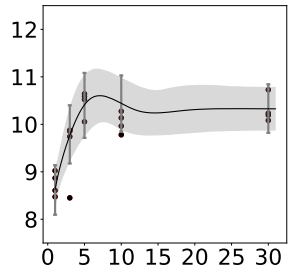
(e) Fit HCR Model to D7 ExpD1 HSC



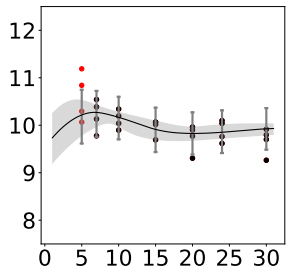
(f) Fit HCR Model to D7 ExpD2 HSC



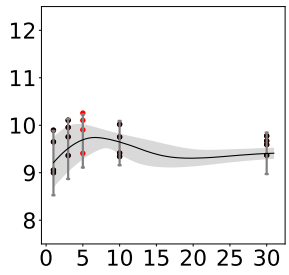
(g) Fit HCR Model to D7 ExpD1 MPP



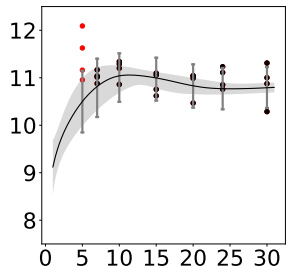
(h) Fit HCR Model to D7 ExpD2 MPP



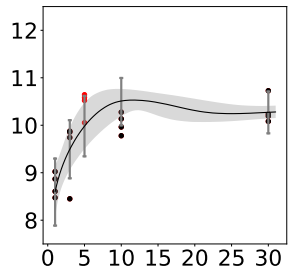
(i) Fit NHCR Model to D5 ExpD1 HSC



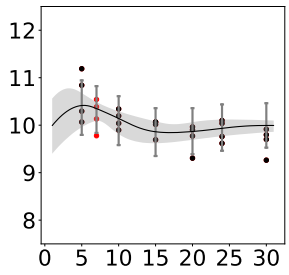
(j) Fit NHCR Model to D5 ExpD2 HSC



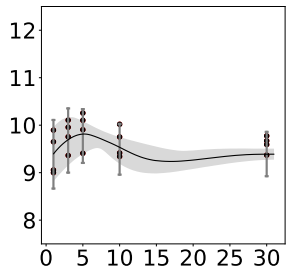
(k) Fit NHCR Model to D5 ExpD1 MPP



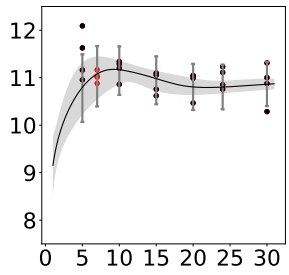
(l) Fit NHCR Model to D5 ExpD2 MPP



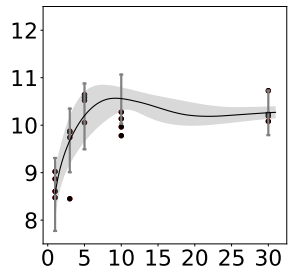
(m) Fit NHCR Model to D7 ExpD1 HSC



(n) Fit NHCR Model to D7 ExpD2 HSC



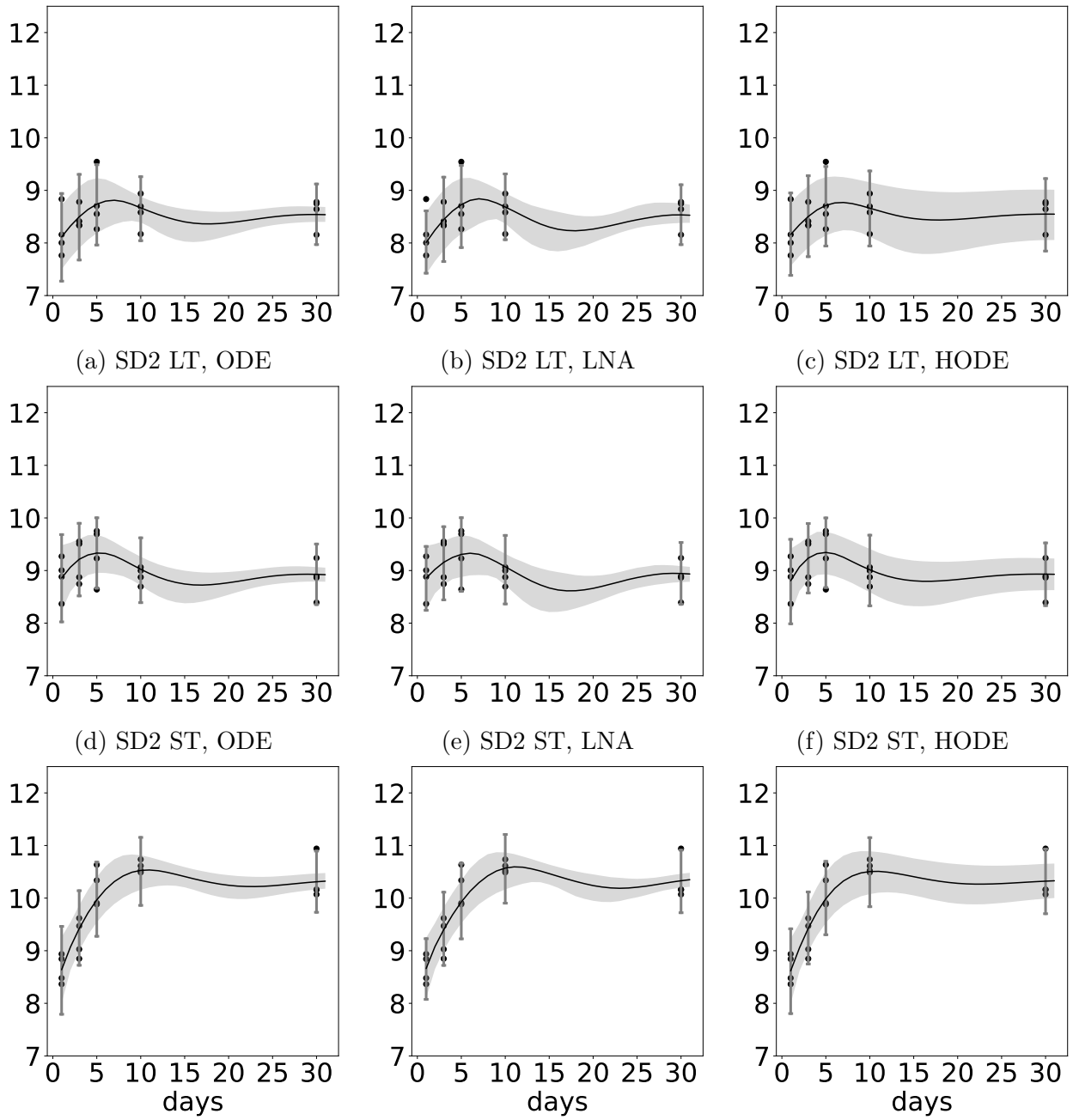
(o) Fit NHCR Model to D7 ExpD1 MPP



(p) Fit NHCR Model to D7 ExpD2 MPP

Appendix B

Chapter 4 Supplementary

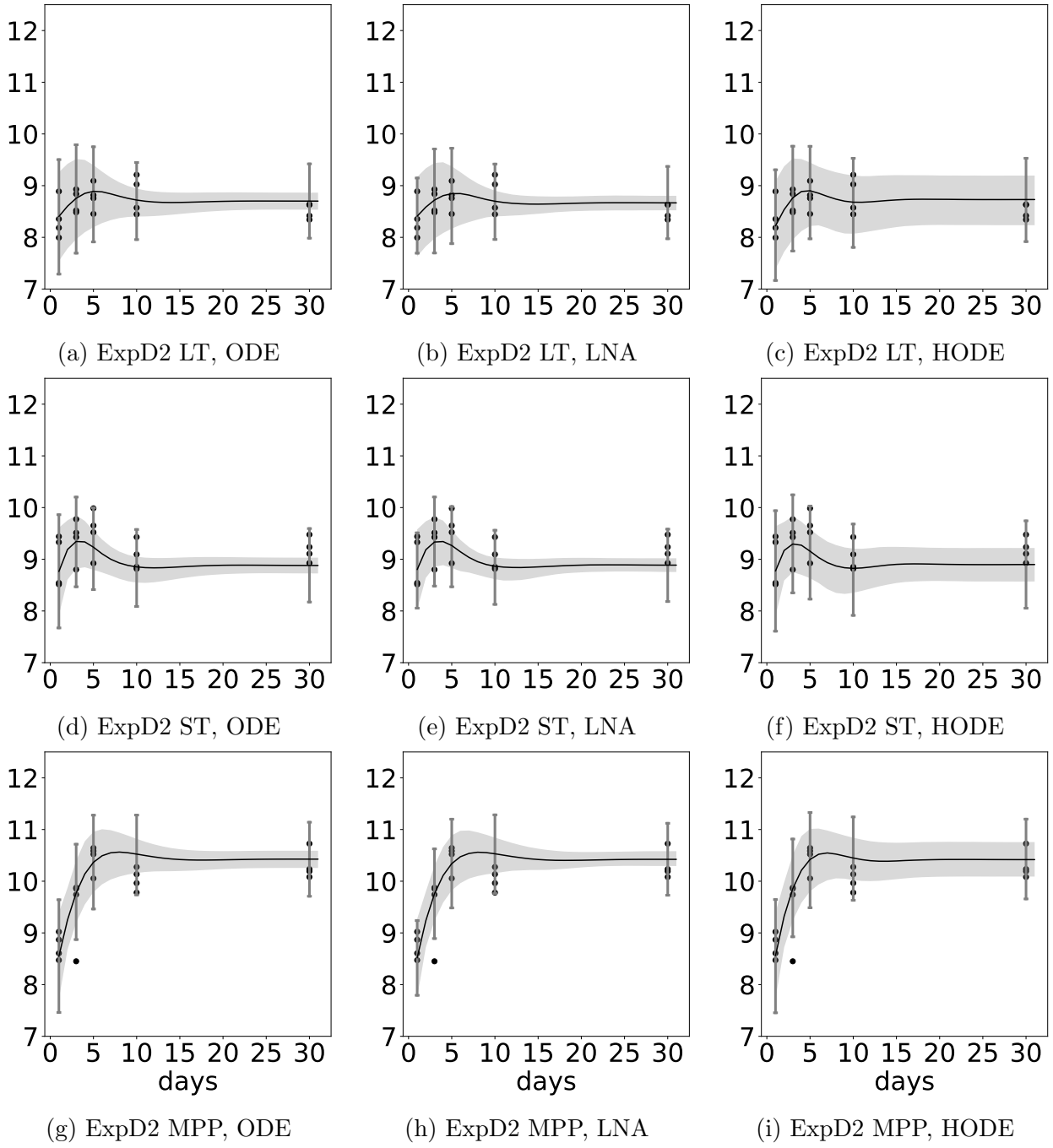


(a) SD2 LT, ODE (b) SD2 LT, LNA (c) SD2 LT, HODE

(d) SD2 ST, ODE (e) SD2 ST, LNA (f) SD2 ST, HODE

(g) MPP D2, ODE (h) MPP D2, LNA (i) MPP D2, HODE

Figure B.1: Dynamic uncertainty plots (DUP) for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the synthetic dataset (SD) which is generated by Gillespie algorithm using the parameter values as MAP optimization point estimation on the experimental data (ExpD). LT: LTHSCs; ST: STHSCs; In DUP (a-i), the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day; (a-c) DUP for LT data in SD2.; (d-f) DUP for ST data in SD2; (g-i) DUP for MPP data in SD2.

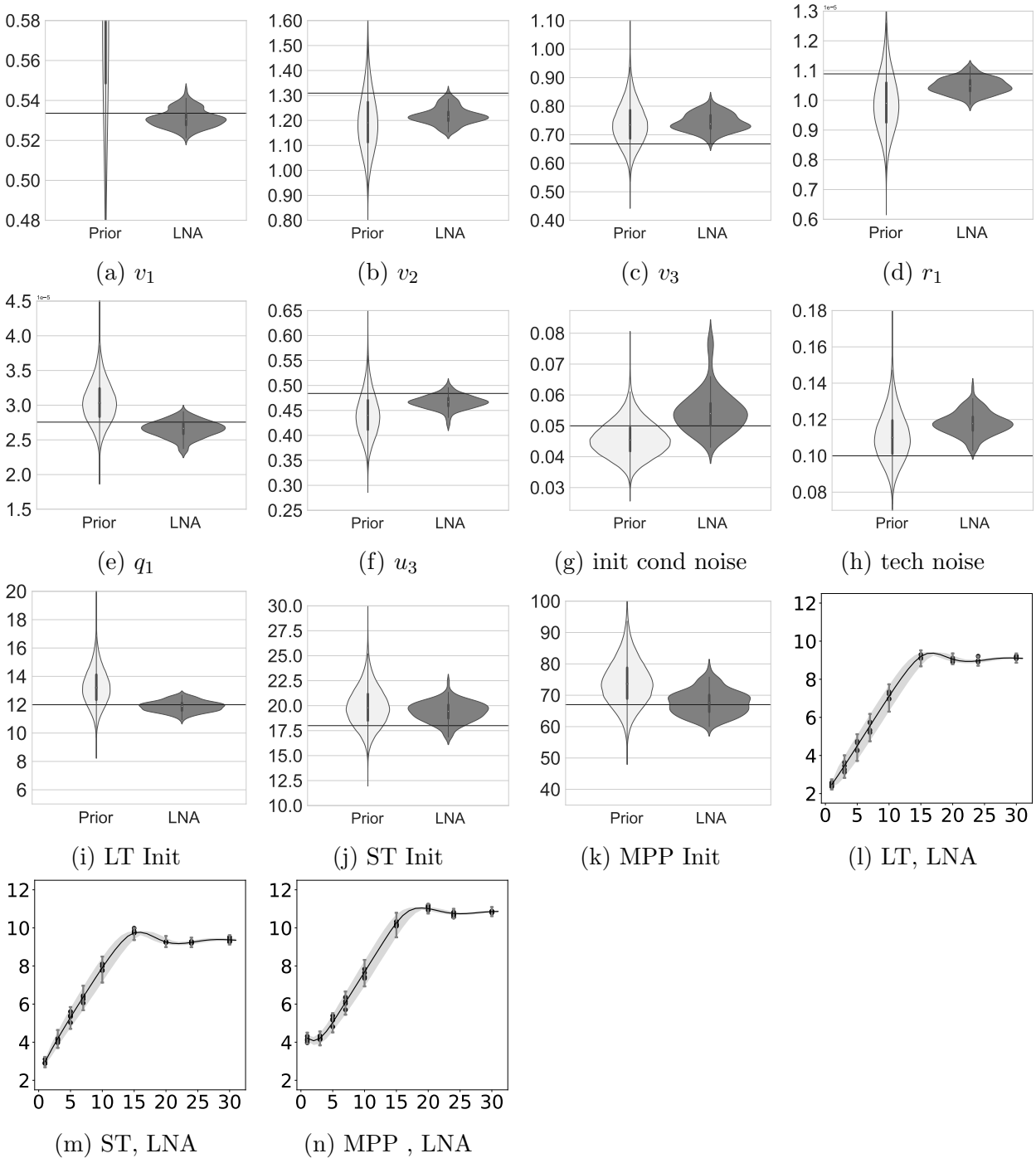


(a) ExpD2 LT, ODE (b) ExpD2 LT, LNA (c) ExpD2 LT, HODE

(d) ExpD2 ST, ODE (e) ExpD2 ST, LNA (f) ExpD2 ST, HODE

(g) ExpD2 MPP, ODE (h) ExpD2 MPP, LNA (i) ExpD2 MPP, HODE

Figure B.2: Dynamic uncertainty plots (DUP) for fitting non-hierarchical/hierarchical ODE (ODE/HODE) and LNA models to the experimental data (ExpD). LT: LTHSCs; ST: STHSCs; In DUP (a-i), the grey shaded areas refer to the 95% credible intervals of time evolution differential equation solutions without technical measurement error; black dot marks refer to partially observed data, each dot related to a certain mouse; black trajectory refers to the median of time evolution differential equation solutions; Line plots refer to the 95% posterior predictive intervals at the experimental observation day. (a-c) DUP for LT data in ExpD1; (d-f) DUP for ST data in ExpD2; (g-i) DUP for MPP data in synthetic ExpD2.



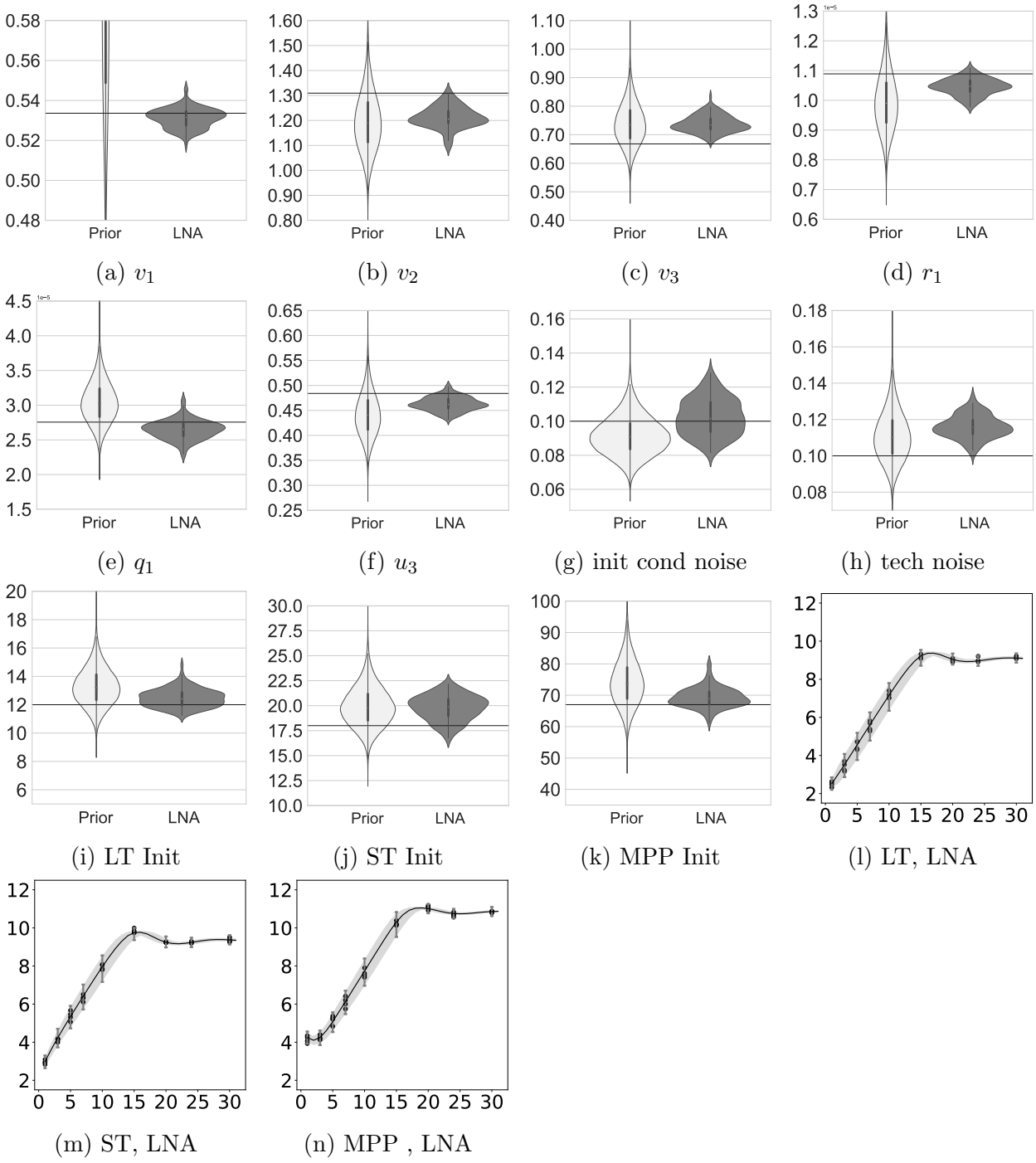
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) tech noise

(i) LT Init (j) ST Init (k) MPP Init (l) LT, LNA

(m) ST, LNA (n) MPP , LNA

Figure B.3: Fit LNA model to time series synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.05$, $\delta_{tech}^{real} = 0.1$. Subfigures (a-k) are prior posterior violin plots for parameters. Subfigures (l-n) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



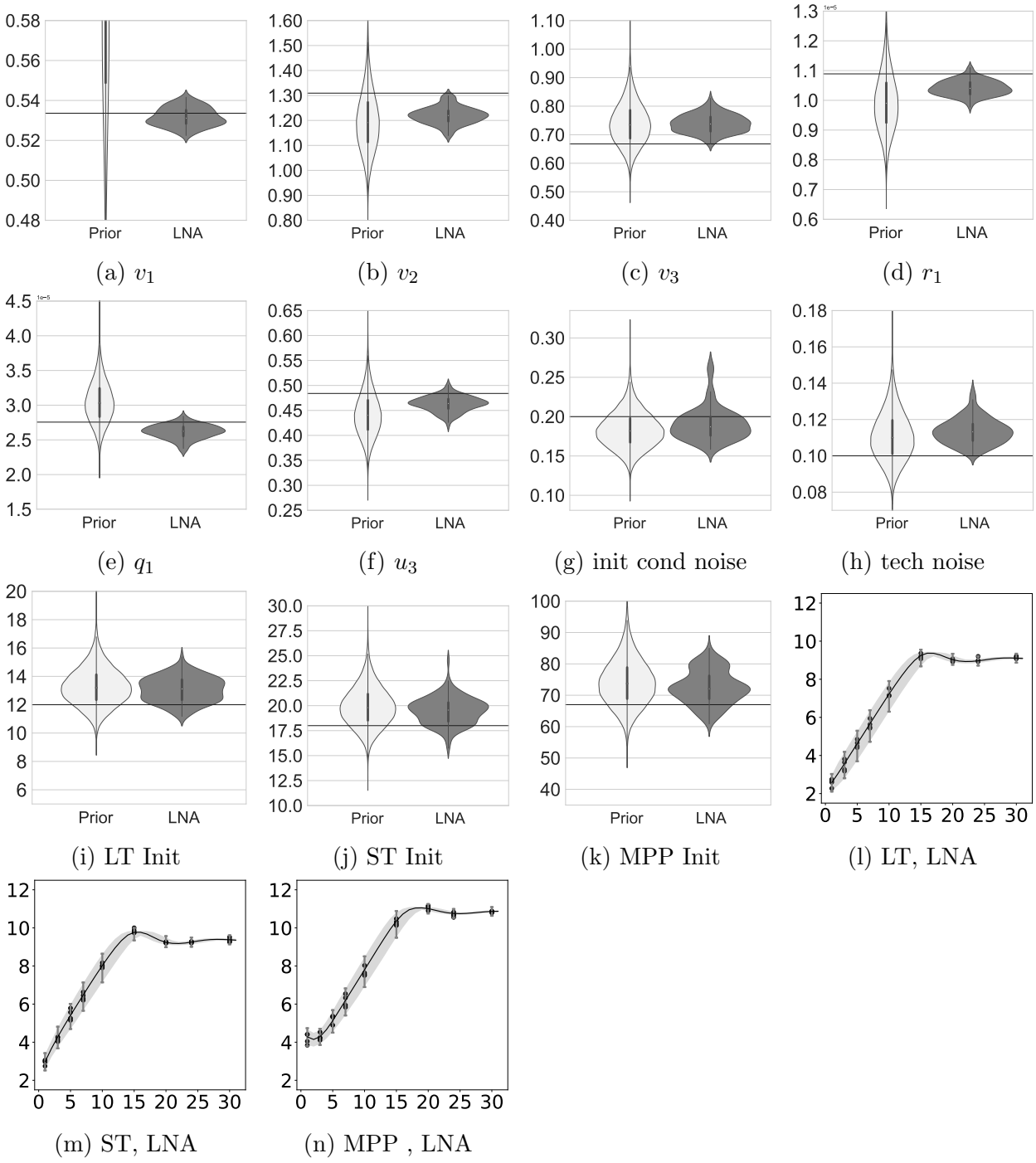
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) tech noise

(i) LT Init (j) ST Init (k) MPP Init (l) LT, LNA

(m) ST, LNA (n) MPP , LNA

Figure B.4: Fit LNA model to time series synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.1$, $\delta_{tech}^{real} = 0.1$. Subfigures (a-k) are prior posterior violin plots for parameters. Subfigures (l-n) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



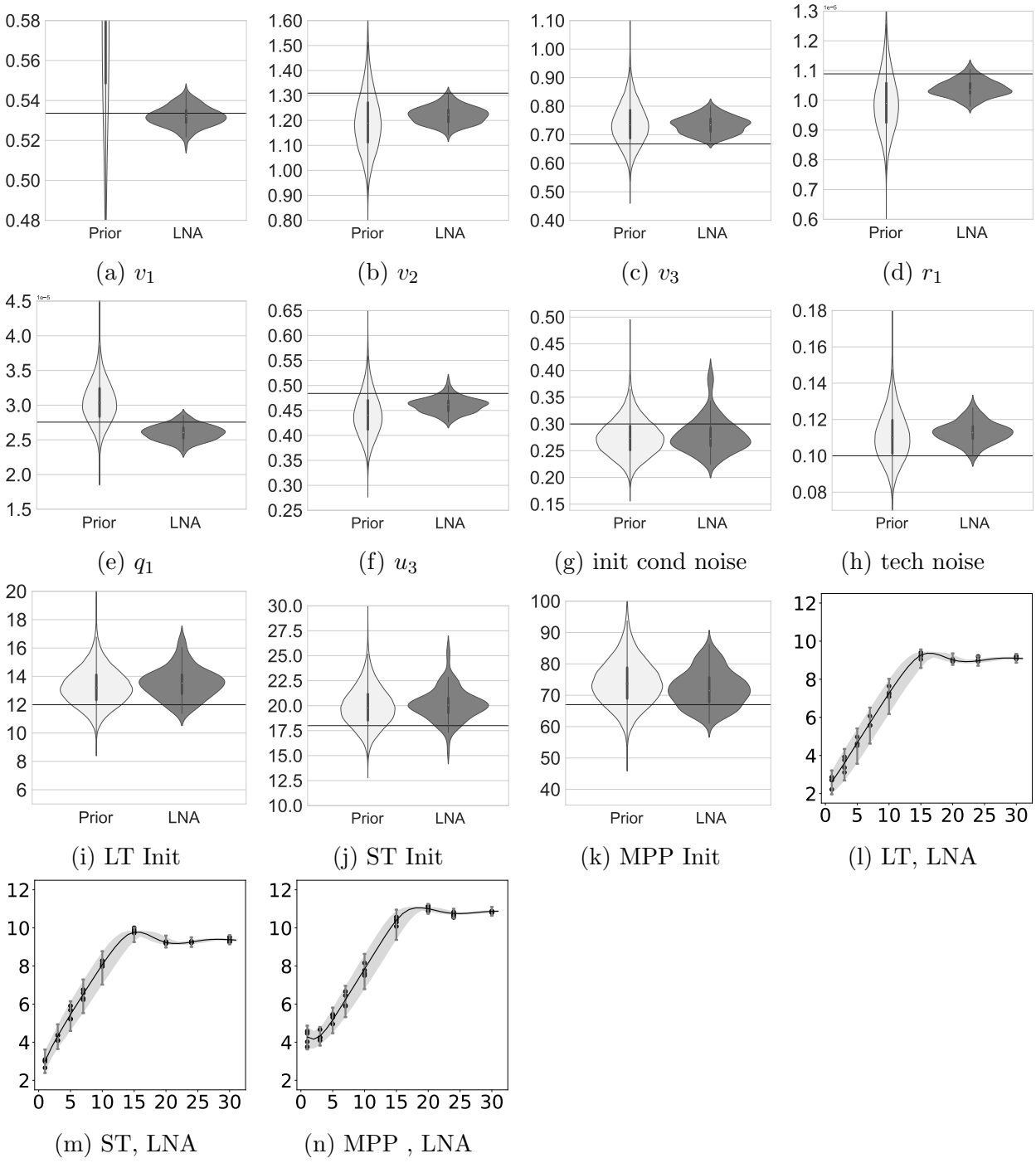
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) tech noise

(i) LT Init (j) ST Init (k) MPP Init (l) LT, LNA

(m) ST, LNA (n) MPP , LNA

Figure B.5: Fit LNA model to time series synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.2$, $\delta_{tech}^{real} = 0.1$. Subfigures (a-k) are prior posterior violin plots for parameters. Subfigures (l-n) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



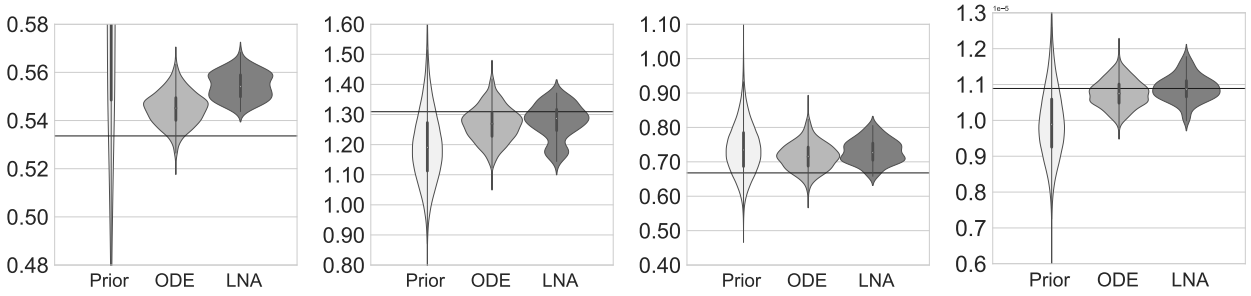
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) tech noise

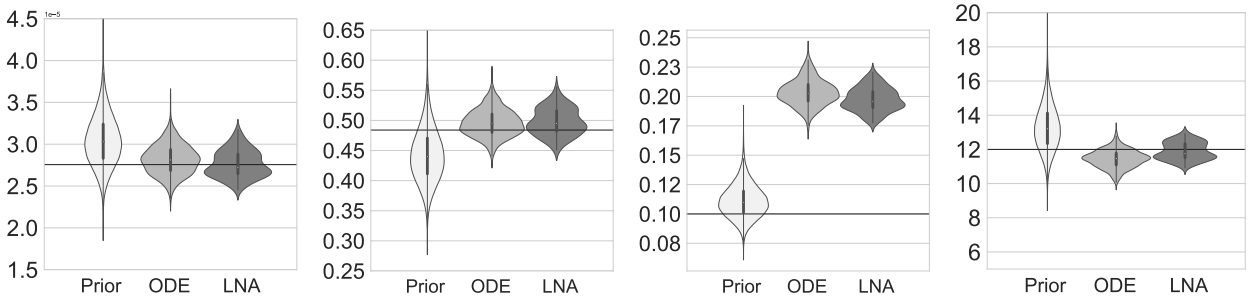
(i) LT Init (j) ST Init (k) MPP Init (l) LT, LNA

(m) ST, LNA (n) MPP , LNA

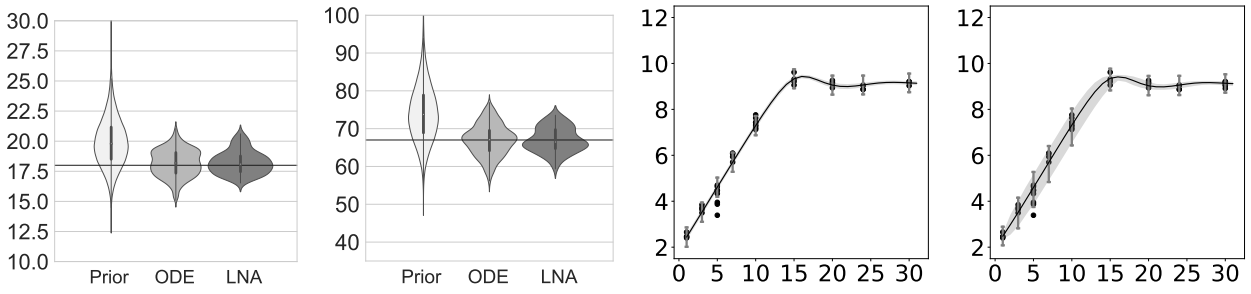
Figure B.6: Fit LNA model to time series synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.3$, $\delta_{tech}^{real} = 0.1$. Subfigures (a-k) are prior posterior violin plots for parameters. Subfigures (l-n) are dynamic uncertainty plots for LT, ST, MPP data fitted by LNA model.



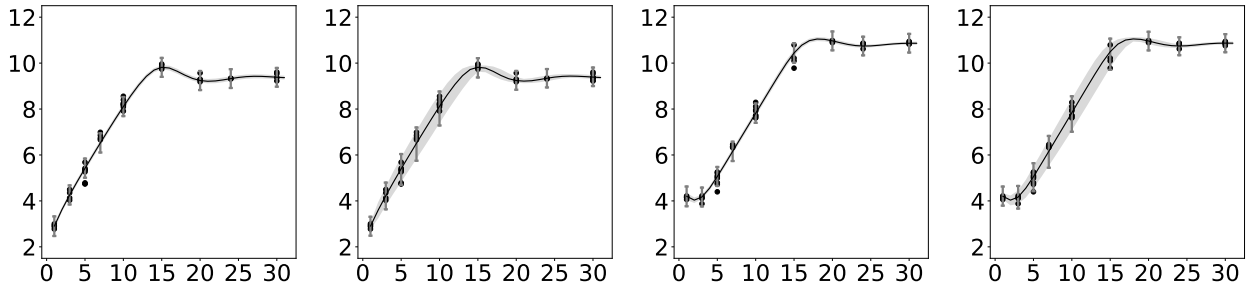
(a) v_1 (b) v_2 (c) v_3 (d) r_1



(e) q_1 (f) u_3 (g) tech noise (h) LT Init

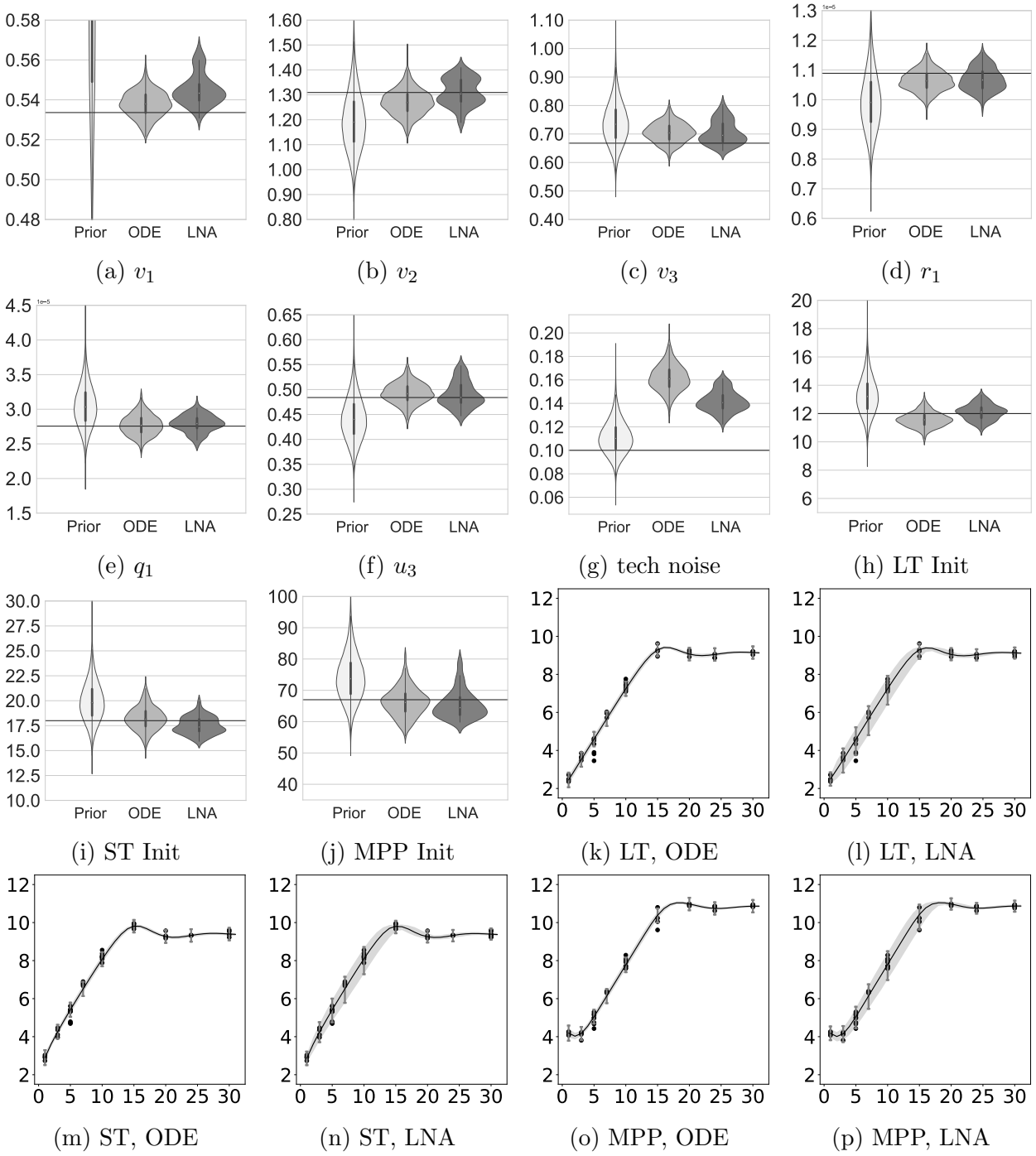


(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA



(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.7: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.05$, $\delta_{tech}^{real} = 0.1$. Fix the initial condition standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



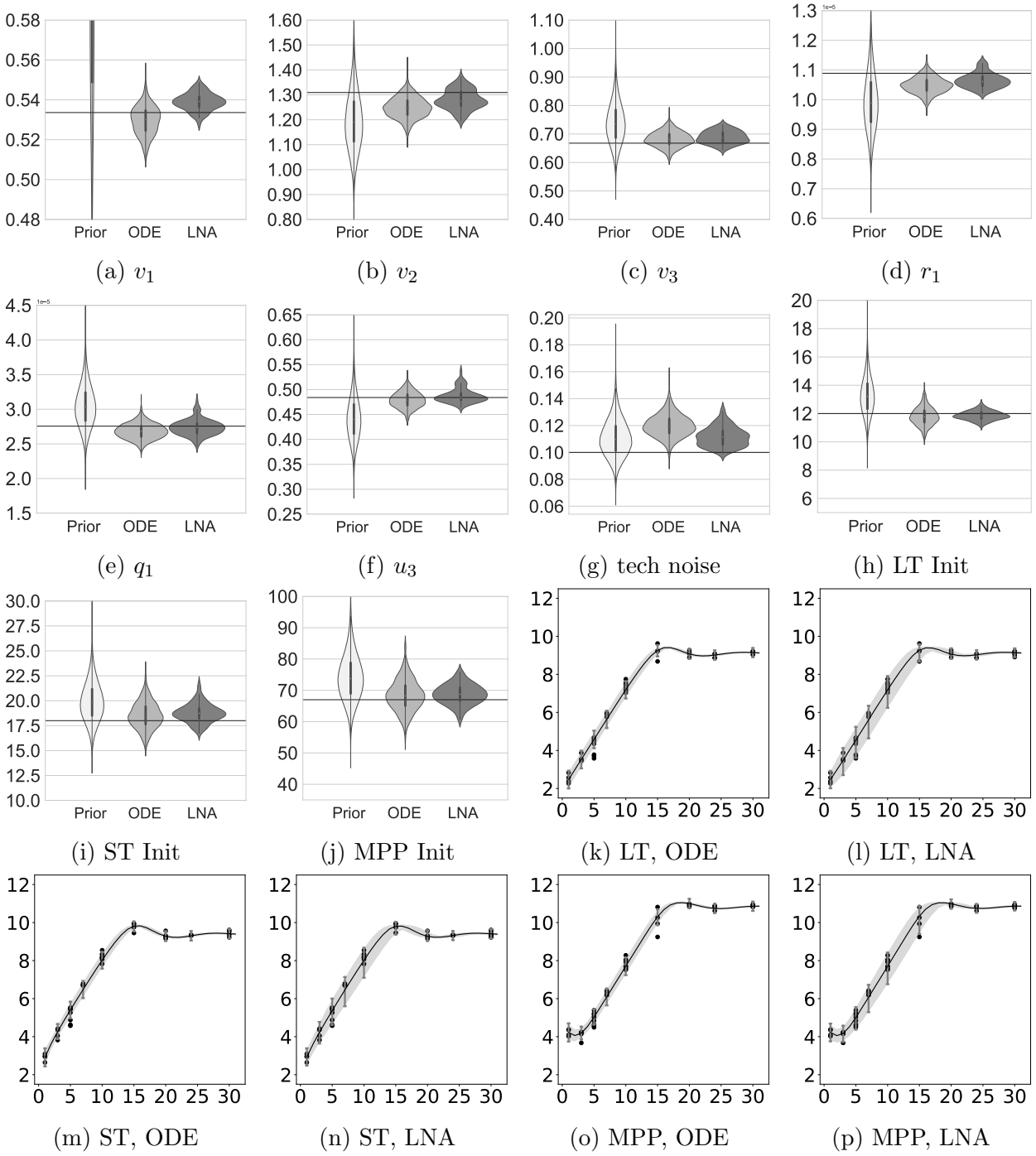
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) tech noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.8: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.1$, $\delta_{tech}^{real} = 0.1$. Fix the initial condition standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



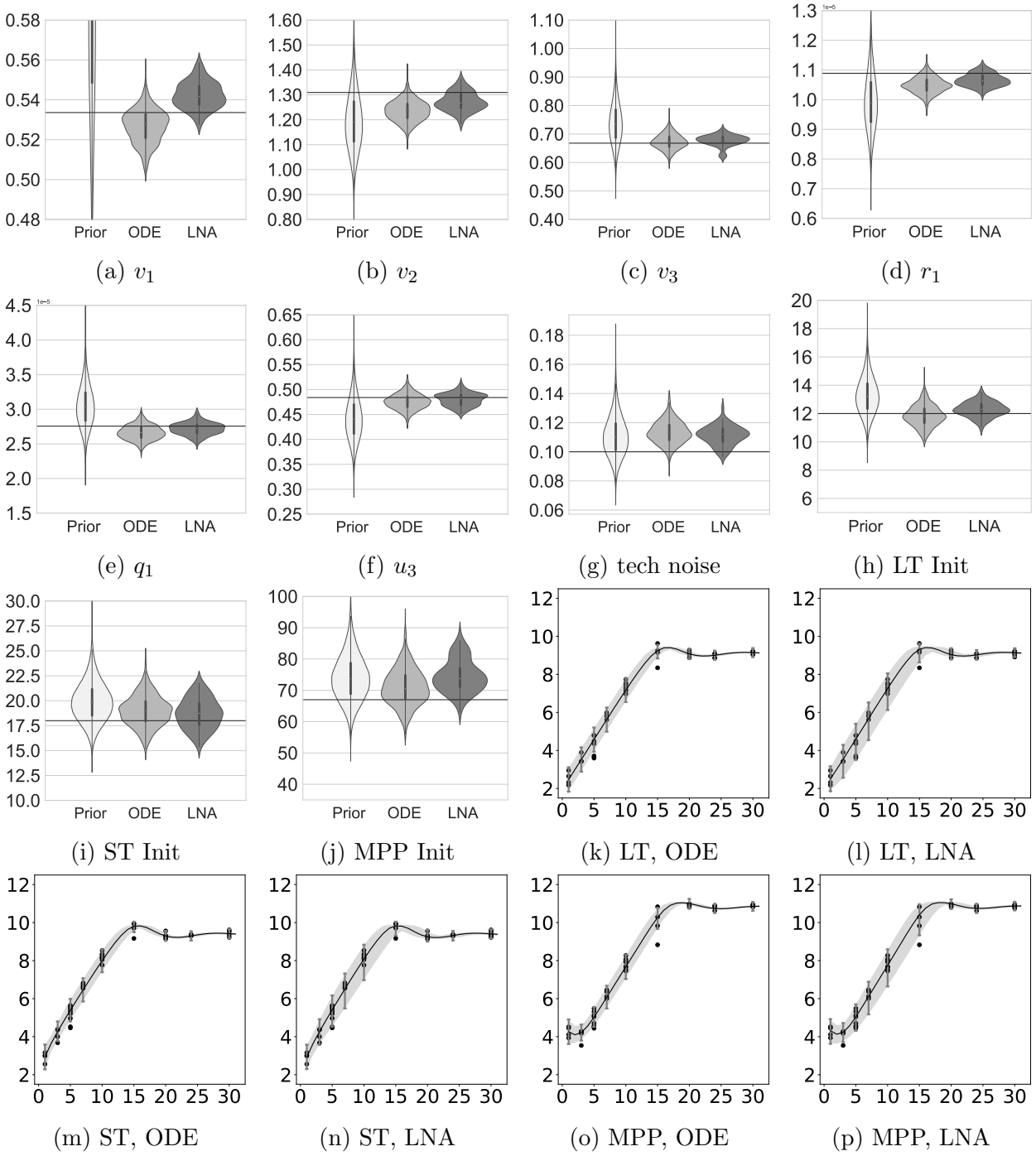
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) tech noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.9: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.2$, $\delta_{tech}^{real} = 0.1$. Fix the initial condition standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



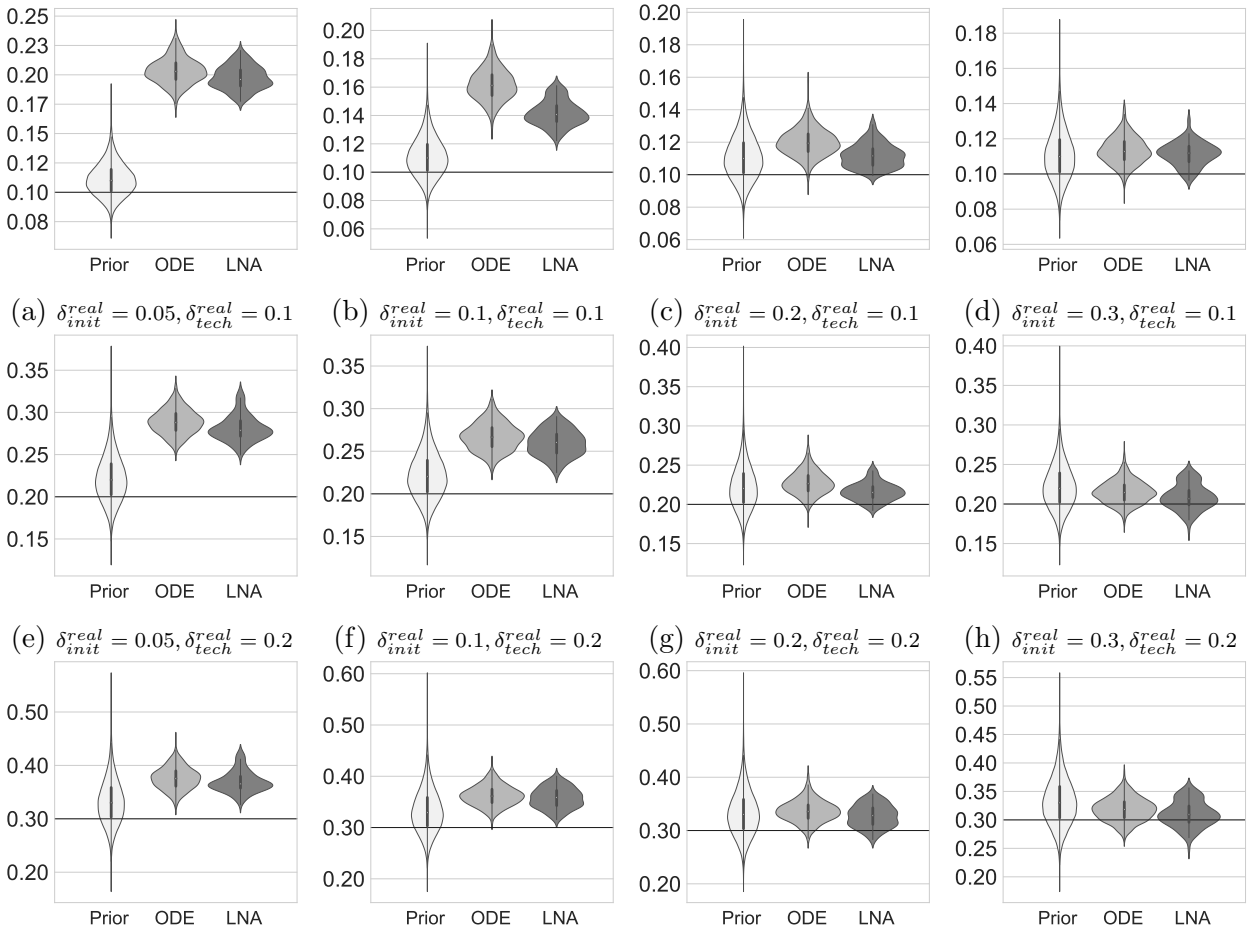
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) tech noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.10: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.3$, $\delta_{tech}^{real} = 0.1$. Fix the initial condition standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



(a) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.1$ (b) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.1$ (c) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.1$ (d) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.1$

(e) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.2$ (f) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.2$ (g) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.2$ (h) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.2$

(i) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.3$ (j) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.3$ (k) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.3$ (l) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.3$

Figure B.11: Prior and posterior violin plots for technical measurement standard deviation term δ_{tech} for simulation study 4.3 that fix initial condition standard deviation δ_{init} at ground truth value and infer all other parameters when the ground truth for standard deviation terms δ_{init}^{real} and δ_{tech}^{real} are in relatively large scales.

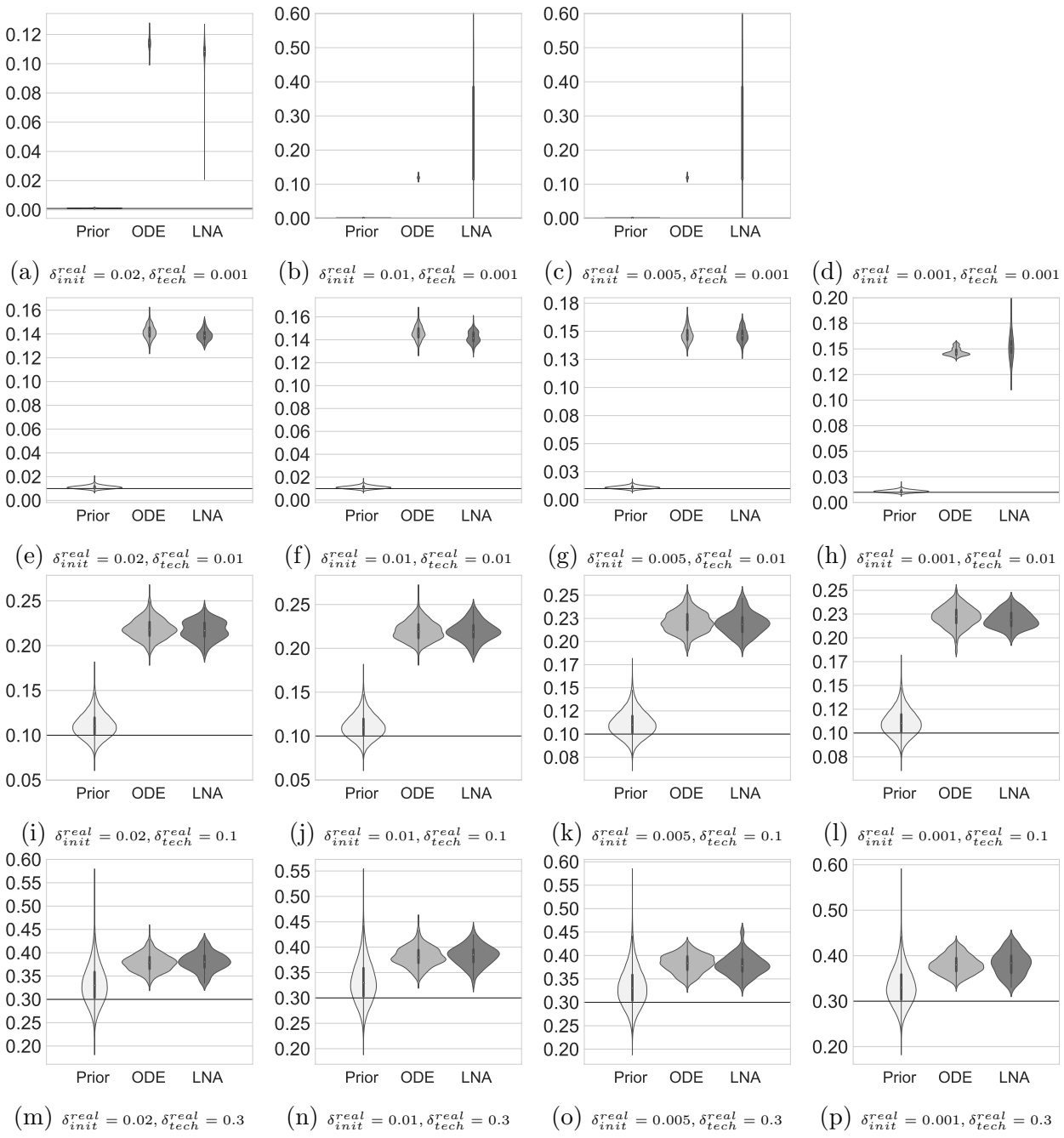
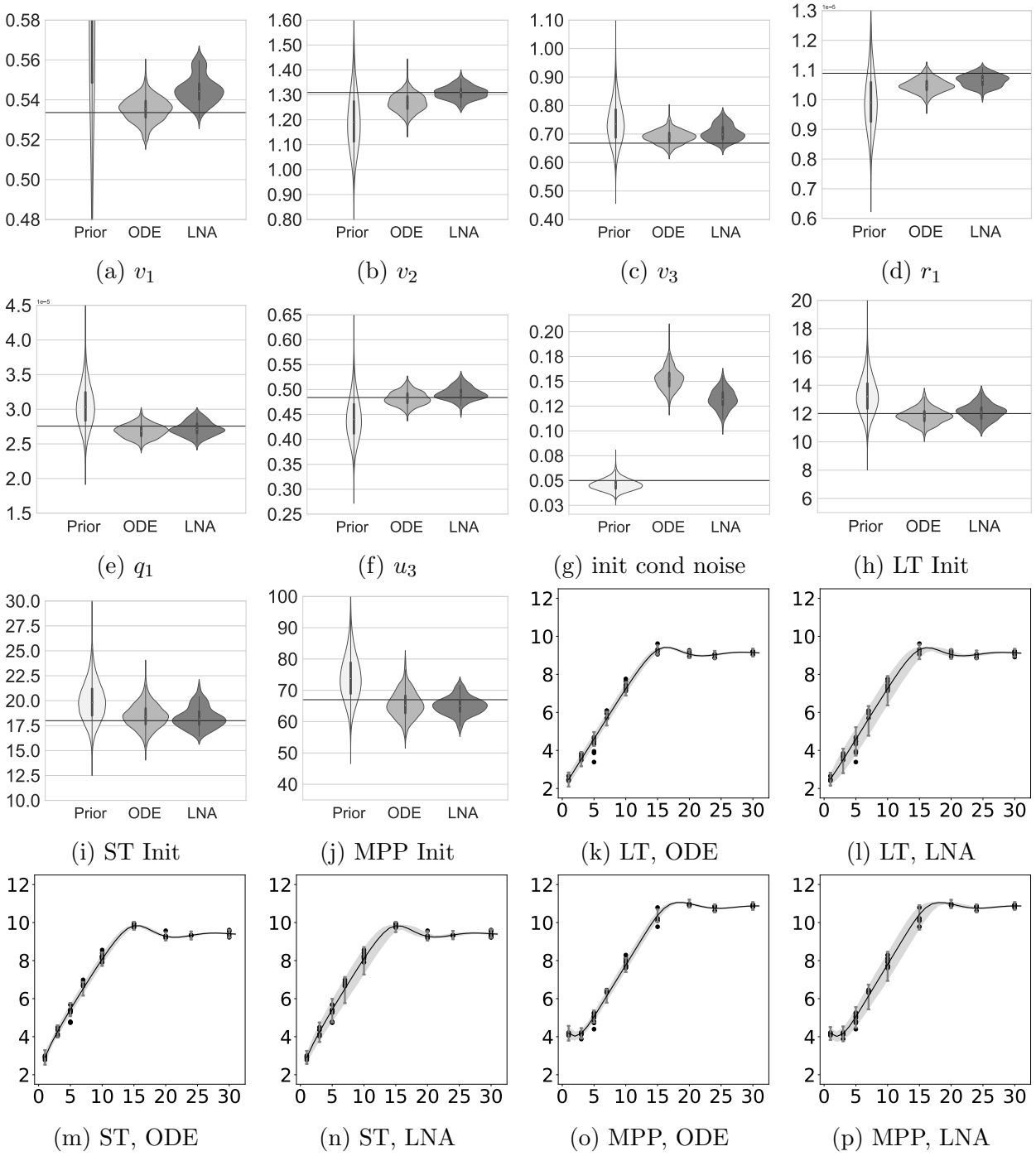


Figure B.12: Prior and posterior violin plots for technical measurement standard deviation term δ_{tech} for simulation study 4.3 that fix initial condition standard deviation δ_{init} at ground truth value and infer all other parameters when the ground truth for standard deviation terms δ_{init}^{real} and δ_{tech}^{real} are in relatively small scales.



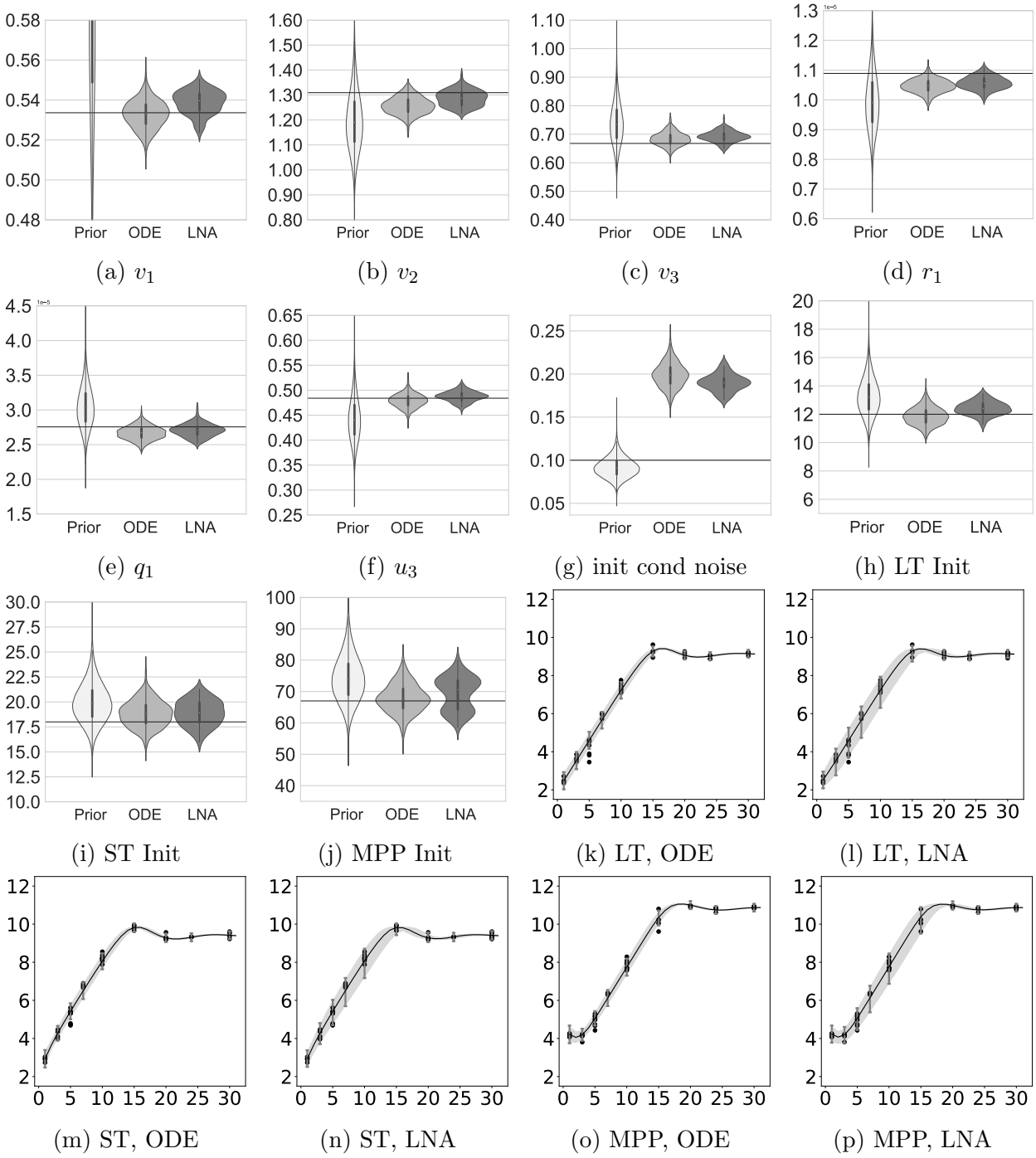
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.13: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.05$, $\delta_{tech}^{real} = 0.1$. Fix the technical measurement standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



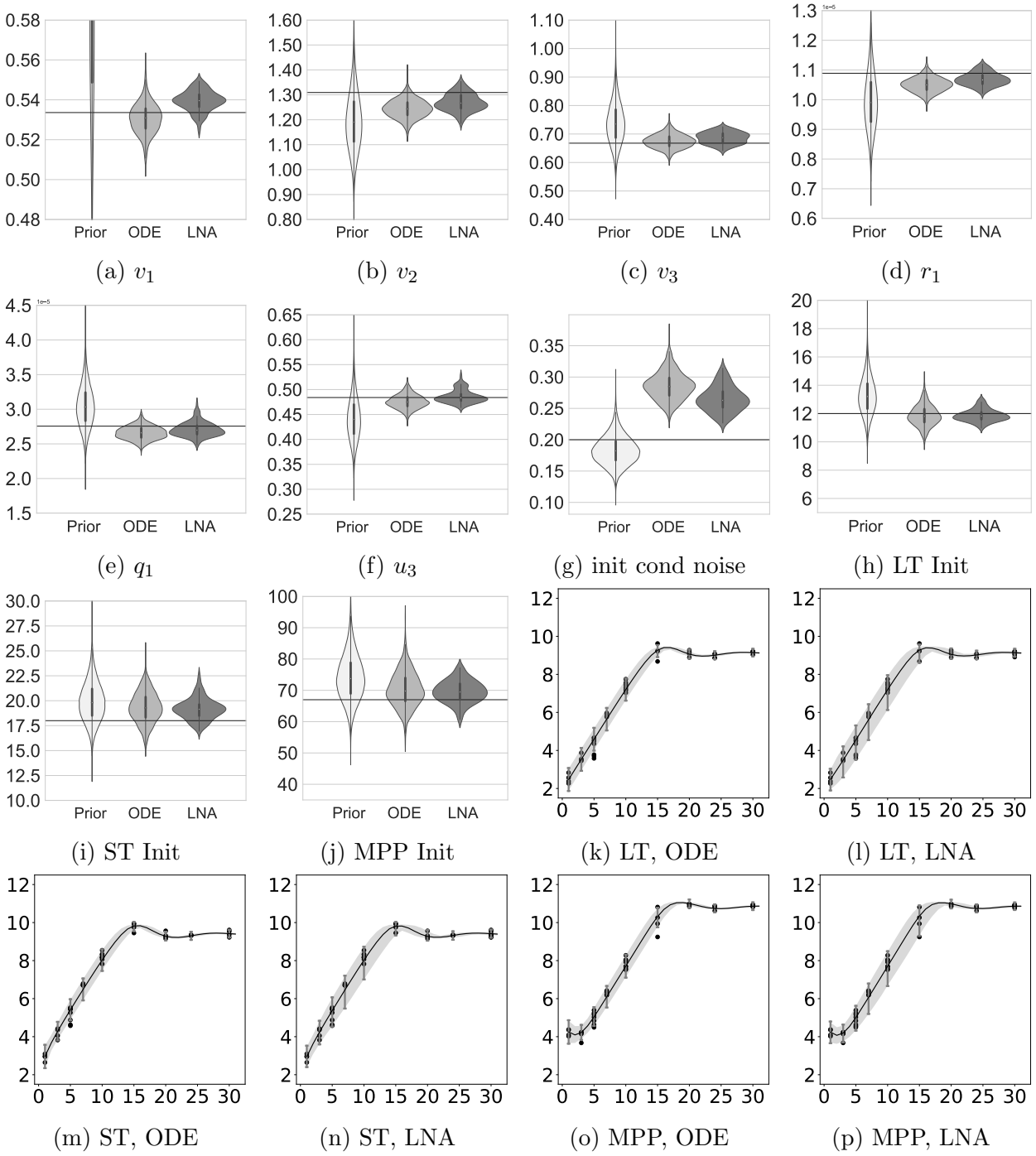
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.14: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.1$, $\delta_{tech}^{real} = 0.1$. Fix the technical measurement standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



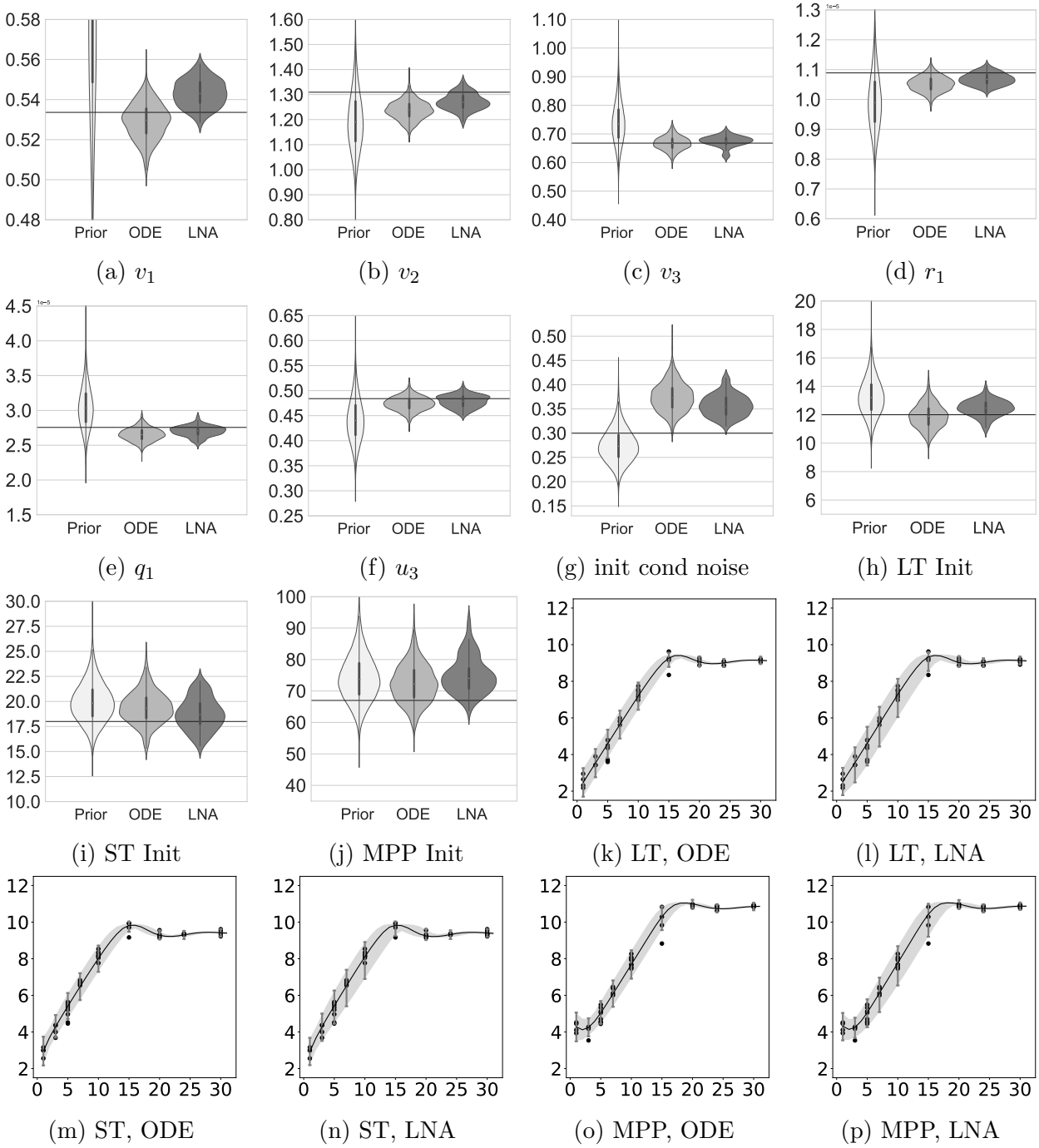
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.15: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.2$, $\delta_{tech}^{real} = 0.1$. Fix the technical measurement standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



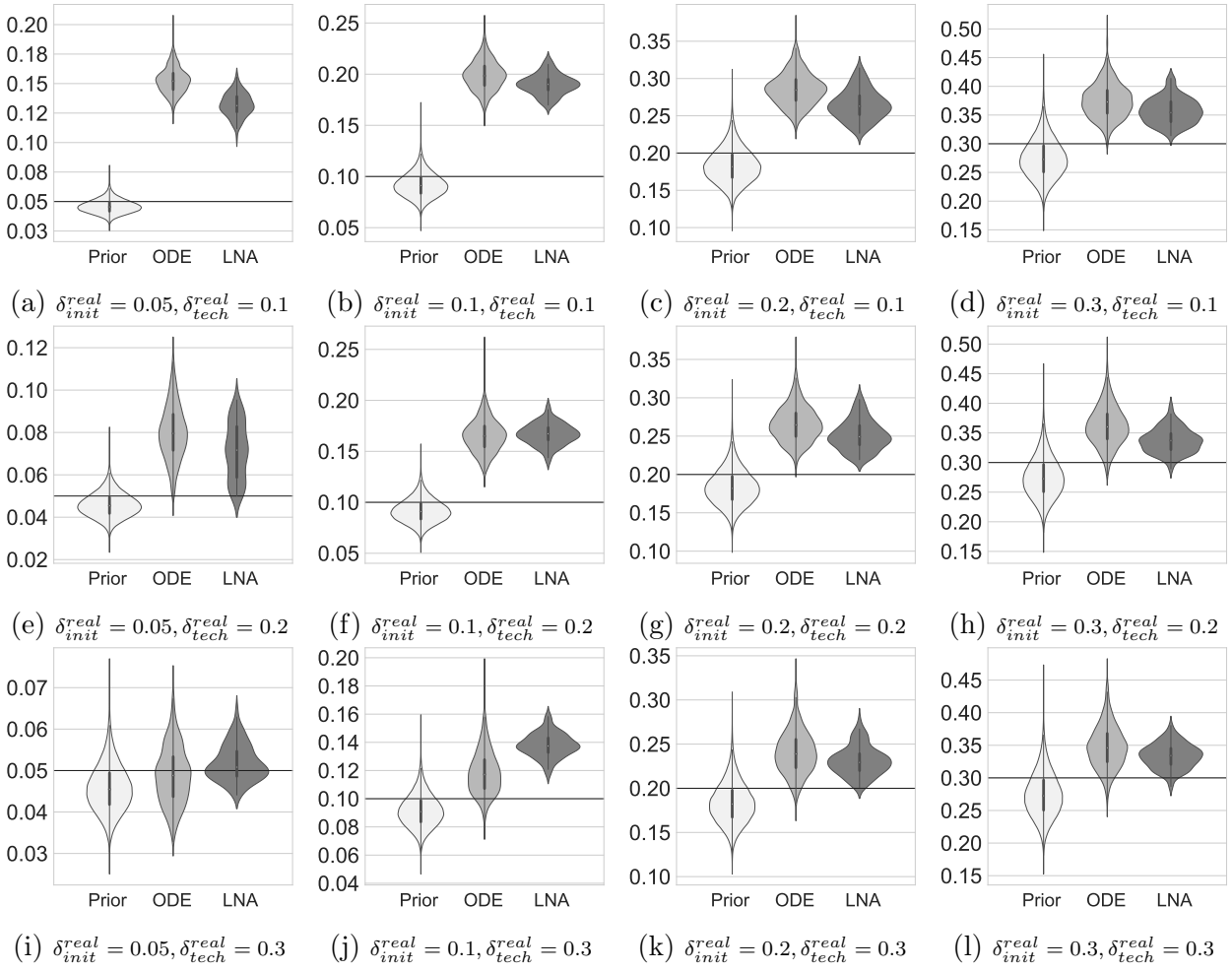
(a) v_1 (b) v_2 (c) v_3 (d) r_1

(e) q_1 (f) u_3 (g) init cond noise (h) LT Init

(i) ST Init (j) MPP Init (k) LT, ODE (l) LT, LNA

(m) ST, ODE (n) ST, LNA (o) MPP, ODE (p) MPP, LNA

Figure B.16: Fit LNA model to partially observed synthetic dataset. Ground truth initial condition and technical measurement standard deviations are $\delta_{init}^{real} = 0.3$, $\delta_{tech}^{real} = 0.1$. Fix the technical measurement standard deviation at ground truth and infer other parameters. Subfigures (a-j) are prior posterior violin plots for parameters. Subfigures (k,m,o) are dynamic uncertainty plots for LT, ST, MPP data fitted by the ODE model. Subfigures (l,n,p) are dynamic uncertainty plots for LT, ST, MPP data fitted by the LNA model.



(a) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.1$ (b) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.1$ (c) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.1$ (d) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.1$

(e) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.2$ (f) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.2$ (g) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.2$ (h) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.2$

(i) $\delta_{init}^{real} = 0.05, \delta_{tech}^{real} = 0.3$ (j) $\delta_{init}^{real} = 0.1, \delta_{tech}^{real} = 0.3$ (k) $\delta_{init}^{real} = 0.2, \delta_{tech}^{real} = 0.3$ (l) $\delta_{init}^{real} = 0.3, \delta_{tech}^{real} = 0.3$

Figure B.17: Prior and posterior violin plots for initial condition standard deviation term δ_{init} for simulation study 4.4 that fix technical measurement standard deviation δ_{tech} at ground truth value and infer all other parameters when the ground truth for standard deviation terms δ_{init}^{real} and δ_{tech}^{real} are in relatively large scales.

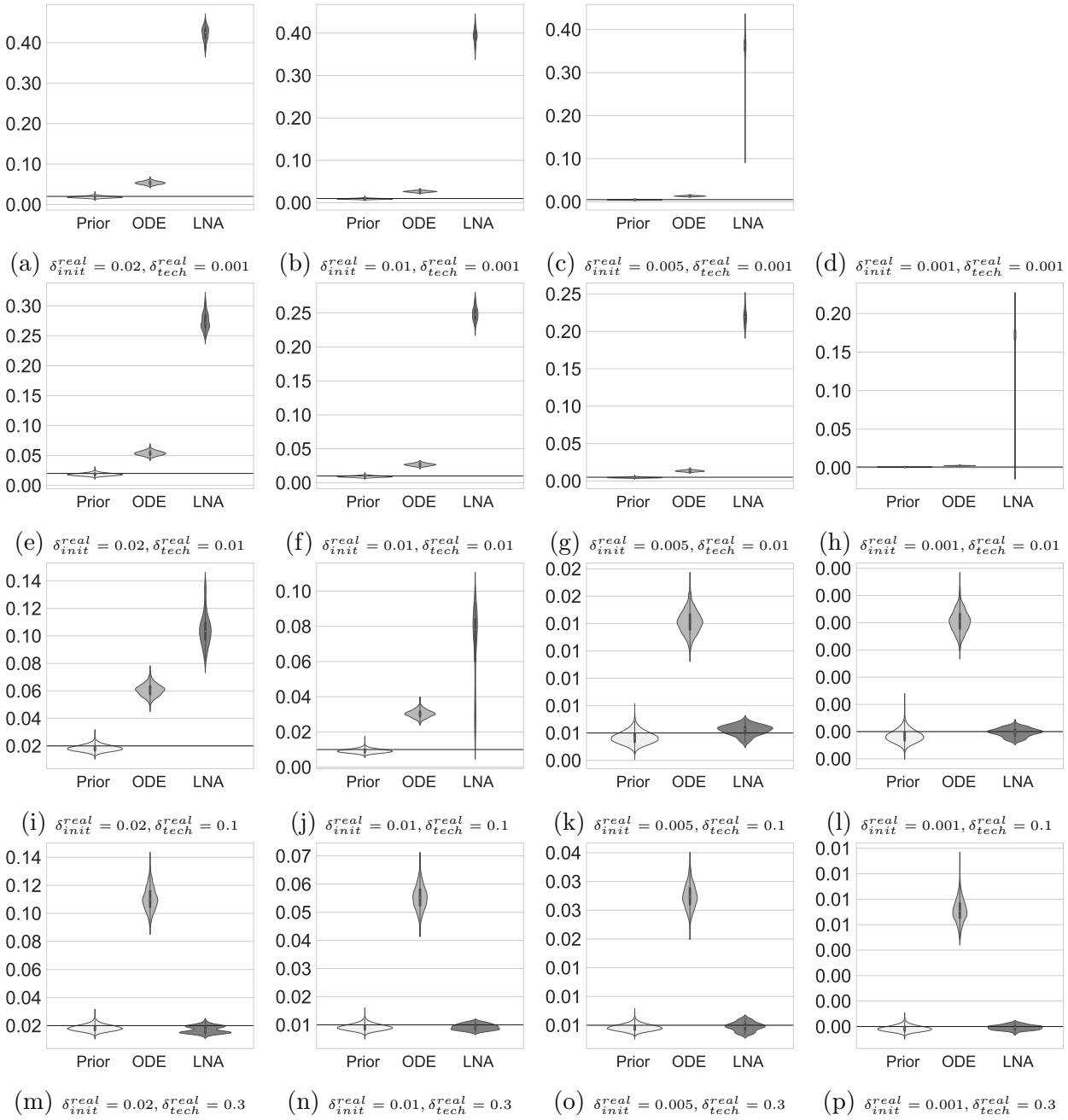


Figure B.18: Prior and posterior violin plots for initial condition standard deviation term δ_{init} for simulation study 4.4 that fix technical measurement standard deviation δ_{tech} at ground truth value and infer all other parameters when the ground truth for standard deviation terms δ_{init}^{real} and δ_{tech}^{real} are in relatively small scales.