

RESEARCH

Open Access



Bayesian kinetic modeling for tracer-based metabolomic data

Xu Zhang^{1*}, Ya Su², Andrew N. Lane^{3,4,5}, Arnold J. Stromberg¹, Teresa W. M. Fan^{3,4,5} and Chi Wang^{1,3,6*}

*Correspondence:
xu.zhang0131@gmail.com;
chi.wang@uky.edu

¹Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington 40536, USA

²Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond 23220, USA

³Markey Cancer Center, University of Kentucky, Lexington 40536, USA

⁴Center for Environmental and Systems Biochemistry, University of Kentucky, Lexington 40536, USA

⁵Department of Toxicology and Cancer Biology, University of Kentucky, Lexington 40536, USA

⁶Division of Cancer Biostatistics, Department of Internal Medicine, University of Kentucky, Lexington 40536, USA

Abstract

Background: Stable Isotope Resolved Metabolomics (SIRM) is a new biological approach that uses stable isotope tracers such as uniformly ^{13}C -enriched glucose ($^{13}\text{C}_6$ -Glc) to trace metabolic pathways or networks at the atomic level in complex biological systems. Non-steady-state kinetic modeling based on SIRM data uses sets of simultaneous ordinary differential equations (ODEs) to quantitatively characterize the dynamic behavior of metabolic networks. It has been increasingly used to understand the regulation of normal metabolism and dysregulation in the development of diseases. However, fitting a kinetic model is challenging because there are usually multiple sets of parameter values that fit the data equally well, especially for large-scale kinetic models. In addition, there is a lack of statistically rigorous methods to compare kinetic model parameters between different experimental groups.

Results: We propose a new Bayesian statistical framework to enhance parameter estimation and hypothesis testing for non-steady-state kinetic modeling of SIRM data. For estimating kinetic model parameters, we leverage the prior distribution not only to allow incorporation of experts' knowledge but also to provide robust parameter estimation. We also introduce a shrinkage approach for borrowing information across the ensemble of metabolites to stably estimate the variance of an individual isotopomer. In addition, we use a component-wise adaptive Metropolis algorithm with delayed rejection to perform efficient Monte Carlo sampling of the posterior distribution over high-dimensional parameter space. For comparing kinetic model parameters between experimental groups, we propose a new reparameterization method that converts the complex hypothesis testing problem into a more tractable parameter estimation problem. We also propose an inference procedure based on credible interval and credible value. Our method is freely available for academic use at <https://github.com/xuzhang0131/MCMCFlux>.

Conclusions: Our new Bayesian framework provides robust estimation of kinetic model parameters and enables rigorous comparison of model parameters between experimental groups. Simulation studies and application to a lung cancer study demonstrate that our framework performs well for non-steady-state kinetic modeling of SIRM data.

Keywords: SIRM, Kinetic modeling, Bayesian method



Background

The metabolome comprises the full set of compounds and the biochemical reactions that represent the functional phenotypes of living organisms. It changes in response to either internal environments (such as mutations or alterations in gene/protein expression) or external perturbations (such as altered nutrient supply) which occur during disease progression. Unraveling the dynamics of changes in the metabolome is a key to elucidating molecular regulations at both gene and protein levels that underlie macroscopic phenotypes such as individual variations in disease development or drug susceptibility. In recent years, a powerful approach, Stable Isotope Resolved Metabolomics (SIRM) [1–11] has been developed to decipher metabolic networks and changes in disease states, which can be linked to altered regulation at gene and protein expression levels. Based on mass spectrometry and nuclear magnetic resonance techniques, SIRM uses stable isotope tracers such as uniformly ^{13}C -enriched glucose ($^{13}\text{C}_6\text{-Glc}$) to trace metabolic pathways or networks at the atomic level [1, 7, 9, 11–16] in cells, tissues, living organisms, or even human patients, that respond to disease development or drug treatment. By determining the atomic position (isotopomers) of stable isotope incorporation into relevant metabolites, the enzymes that are involved in the altered biochemical conversion of parent tracers into these metabolites can be systematically identified along with their possible allosteric regulator(s). In cancer research, for example, SIRM studies have uncovered key reprogrammed metabolic events and regulations in cancer cells and tissues [14, 17–23].

Kinetic modeling analysis is a commonly used approach to quantitatively model metabolic pathway dynamics based on total metabolite profiles and/or tracer data. Many kinetic modeling analyses [24–26] focus on the case that the system has achieved both metabolic and isotopic steady state. Several statistical and machine learning algorithms have been developed to characterize kinetic models and estimate model parameters for such models, see Saa and Nielsen [27] and Cuperlovic-Culf [28] for reviews. However, the steady state is difficult to achieve in complex biological systems such as mammalian cells, tissues, or in whole bodies [29]. An important development is the extension of kinetic modeling analysis to non-metabolic and non-isotopic steady flux analysis based on ordinary differential equations (ODEs) that represent the kinetics of individual or groups of enzyme-catalyzed reactions within the proposed metabolic networks [30–35]. Non-steady-state kinetic models are more realistic and have been increasingly used to understand the regulation of normal metabolism, the development of disease, and to predict metabolic reactions upon genetic and environmental perturbation [36–39]. Such characterization is then necessary for the development of novel therapeutic strategies.

For statistical analyses of non-steady-state kinetic modeling, current methods [31, 40, 41] are weighted least squares-based, which seek an optimal set of parameters to minimize the difference between model and the observed data. However, due to the complexity of the model and limited number of replicates, there are usually multiple sets of parameter values that fit the data equally well [40]. Without imposing additional constraints, it is difficult to unambiguously determine all model parameters. In addition, quantifying the uncertainty in parameter estimation is also challenging because the Hessian matrix, which is required to calculate the standard error of parameter estimator, is often ill-conditioned. Further, without an accurate characterization of the standard error

of parameter estimator, it is difficult to statistically rigorously compare kinetic model parameters as well as metabolic flux between different experimental groups, e.g. treatment versus control. Consequently, Fan *et al.* [41] only compared the point estimate of a parameter between groups without providing a quantification of statistical significance such as a p-value.

Here we describe a new Bayesian statistical framework for non-steady-state kinetic modeling of SIRM data to provide accurate and robust estimation of model parameters and to allow rigorous comparison of model parameters between experimental groups. For estimation of kinetic model parameters, our Bayesian method incorporates experts' knowledge about the most likely region of model parameters by a prior distribution, which yields robust estimates of parameters and avoids retrieving scientifically unrealistic parameter values. In addition, we propose a Bayesian shrinkage prior for the variances of the isotopomer abundances to stabilize their estimation, and thus more reliably quantify the uncertainty in kinetic model parameter estimation for limited sample sizes. Further, we introduce advanced Markov chain Monte Carlo (MCMC) methods including adaptive Metropolis [42–44] and delayed rejection [45, 46] to improve the efficiency of posterior sampling under the high-dimensional situation. For comparison of kinetic parameters between experimental groups, we propose a novel reparameterization method that converts the complex hypothesis testing problem, which involves high-dimensional integrals, into a more tractable parameter estimation problem. Based on the posterior sample of that parameter, a credible interval can be easily constructed to assess the null hypothesis. We further propose a *credible value* to quantify the likelihood for the null hypothesis to hold. Simulation studies were conducted to evaluate the accuracy of the proposed method in estimating kinetic model parameters and detecting differentiated parameters between experimental groups. By applying our method, we characterized the dysregulation of purine synthesis in lung squamous cell carcinoma tissues.

Methods

In this section, we introduce a kinetic model to characterize the dynamics of biochemical processes. Based on this model, we propose a Bayesian framework to estimate the model parameters and conduct hypothesis tests comparing parameter values between experimental groups.

A Bayesian kinetic model

Biochemical processes can be described by systems of biochemical reactions, showing how the reactants are transformed and related to the products. The most common approach to model the evolution of such a system over time is dynamical system modeling via ODEs. These ODEs characterize the functional relationship between reactants at a specific time point t , where the time derivatives quantify the biochemical reaction rates between reactants. Specifically, let $\mathbf{y}_{tj} = (y_{tj1}, \dots, y_{tjn})'$ be a vector of observed concentrations of n isotopomers at time t in independent repeated sample j , $j = 1, \dots, m$, and $\boldsymbol{\mu}_t$ be a vector of mean concentrations of isotopomers at time t . We consider the following model

$$\begin{aligned}\log y_{tj} &= \log \mu_t + \delta_{tj}, \\ \frac{d\mu_t}{dt} &= f(\mu_t; \beta, t),\end{aligned}\tag{1}$$

where f is the formula of a set of pre-specified nonlinear ODEs with respect to μ_t for different isotopomers, β is a vector of unknown ODE model parameters on the logarithmic scale, and δ_{tj} is a vector of independent random error terms. The ODE model parameters represent the rate constants of the biochemical reactions which are determined by the enzyme parameters, such as the K_m (the Michaelis-Menten constant) and k_{cat} (the turnover number, the number of times each enzyme site converts substrate to product per unit time), which are our main parameters of interest. Note that because ODE parameters are always positive, we consider β , the natural logarithm of ODE parameters, for the convenience of prior specification and posterior sampling. The error term δ_{tj} describes the technical and biological variations. It has been suggested that a normally distributed additive error term on the logarithm of the data is appropriate when modeling biochemical reactions [47]. Therefore, we assume $\delta_{tj} \sim N(0, \Sigma_\delta)$, where Σ_δ is a diagonal matrix with the i th diagonal element being σ_i^2 . Let σ^2 be a vector of the σ_i^2 's.

Although our interest lies in β , the inference relies heavily on a stable estimate of σ^2 . When we assume discrete T time points are observed, the likelihood function of the model, $L(\beta, \sigma^2)$, is

$$|\Sigma_\delta|^{-\frac{mT}{2}} \exp \left\{ -\frac{1}{2} \sum_t \sum_{j=1}^m (\log y_{tj} - \log \mu_t)' \Sigma_\delta^{-1} (\log y_{tj} - \log \mu_t) \right\},$$

where μ_t is a function of β and t defined in Eq. (1), and Σ_δ contains σ^2 . This model is highly nonlinear and the parameters are embedded within the ODEs. It is challenging to use the maximum likelihood method directly to obtain model parameter estimations since taking derivatives with respect to the parameters within ODEs is not trivial, though it can be done numerically, or explicitly for known function forms such as Hill equation or the Michaelis-Menten equation). Bayesian methods are valuable alternatives to obtain posterior samples of parameters without the need to take derivatives. They also provide a natural and principled way to combine the observed data with experts' prior knowledge and information about the parameters, which is the key to achieve robust parameter estimation for a complex system with a limited sample size. Let $p(\beta)$ and $p(\sigma^2)$ be priors of β and σ^2 , respectively, and assume that they are independent with each other, the joint posterior distribution of β and σ^2 is

$$p(\beta, \sigma^2 | y) \propto L(\beta, \sigma^2) p(\beta) p(\sigma^2),$$

where $y = (y_{tj})$ is the observed data across time point $t = 1, \dots, T$ and replicate $j = 1, \dots, m$.

Prior specification

We assume that the prior distribution of β is a truncated multivariate normal distribution. Specifically, we first consider a multivariate normal distribution with mean ξ and covariance matrix Λ , where values of ξ and Λ are specified based on experts'

knowledge and using literature data and databases such as BRENDA [48]. We then further constrain the distribution within a reasonable value range between ξ_l and ξ_u , which are also specified by experimentalists, to rule out some extreme values that are not biologically possible, and thus enhances the robustness of our method under the high-dimensional situation. Note that we do not require the experts to specify very narrow ranges. The ranges are determined by the amount of information available about each enzyme. If there is extensive information about an enzyme under different conditions, a relatively narrow range will be specified, whereas for less well characterized enzymes, a wider range will be used.

When only a limited number of replicates are collected for an individual isotopomer, it is unlikely to provide a stable estimate of σ_i^2 without any regulation. We consider a Bayesian shrinkage approach that assumes a common prior for all σ_i^2 's, which regulates the variance estimate by borrowing information from the ensemble of isotopomers. This approach has been shown to provide a more stable estimate of the variance in transcriptomics studies [49]. Importantly, a stable estimate of the variance can also enhance the estimation of β . Specifically, we assume the common prior is an inverse gamma distribution with shape parameter α_* and scale parameter κ_* . The hyperparameters α_* and κ_* are empirically determined based on the sample variances of all isotopomers, i.e.

$$\alpha_* = \left(\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n s_{it}^2 \right)^2 / \left\{ \frac{1}{nT-1} \sum_{t=1}^T \sum_{i=1}^n (s_{it}^2 - \bar{s}^2)^2 \right\} + 2,$$

$$\kappa_* = \left(\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n s_{it}^2 \right) (\alpha_* - 1),$$

where $s_{it}^2 = \{ \sum_{j=1}^m (\log y_{tji} - \sum_{j=1}^m \log y_{tji} / m)^2 \} / (m - 1)$ denotes the sample variance over m replicates for each isotopomer i at time point t , $\bar{s}^2 = (\sum_{t=1}^T \sum_{i=1}^n s_{it}^2) / (nT)$ stands for the average of all sample variances across time and isotopomer.

To demonstrate how the priors impact parameter estimation, consider the logarithm of the posterior

$$\begin{aligned} \log p(\beta, \sigma^2 | \mathbf{y}) \propto & -mT \log |\Sigma_\delta| - \sum_t \sum_j (\log \mathbf{y}_{tj} - \log \boldsymbol{\mu}_t)' \Sigma_\delta^{-1} \\ & (\log \mathbf{y}_{tj} - \log \boldsymbol{\mu}_t) - (\boldsymbol{\beta} - \boldsymbol{\xi})' \Lambda^{-1} (\boldsymbol{\beta} - \boldsymbol{\xi}) \\ & + 2 \sum_i^n \left\{ (-\alpha_* - 1) \log \sigma_i^2 - \frac{\kappa_*}{\sigma_i^2} \right\}. \end{aligned}$$

The above formula is a penalized log-likelihood function from a frequentist point of view, where a penalty term $-(\boldsymbol{\beta} - \boldsymbol{\xi})' \Lambda^{-1} (\boldsymbol{\beta} - \boldsymbol{\xi})$ is introduced to penalize $\boldsymbol{\beta}$ value that is far away from its prior mean $\boldsymbol{\xi}$. Likewise, a penalty term $(-\alpha_* - 1) \log \sigma_i^2 - \kappa_* / \sigma_i^2$ is to penalize σ_i^2 values far away from the overall empirical mean. Therefore, by using the prior distribution to regularize the parameter estimation, our method reduces the chances of getting extreme or implausible estimated values for those parameters.

Inferring posterior distribution

Point estimates and credible intervals for model parameters are obtained from the posterior distribution using MCMC methods. We use the Gibbs sampling [50] to iteratively sample each element of β and σ^2 from its conditional posterior distribution given all other parameters. In the next two subsections, we describe methods to sample from those conditional posterior distributions.

Posterior sampling for kinetic parameters

The conditional posterior sample of a kinetic parameter is obtained by using MCMC because that posterior has no analytic expression due to the complexity of the ODEs. Since the dimension of parameters is high, the standard Metropolis-Hastings algorithm is inefficient and may yield low rates of acceptance and poor mixing of the chain [44]. To improve efficiency, we opted to implement advanced MCMC methods including component-wise adaptive Metropolis [42–44] and delayed rejection [45, 46].

The adaptive Metropolis [43] tunes the covariance matrix of the proposal distribution in the Metropolis-Hastings algorithm based on the past sample path of the chain and automatically adapts the proposal distribution towards the target distribution. Based on the adaptive Metropolis algorithm, our proposal distribution for the i th element of the parameter vector, β_i , at the $(l + 1)^{th}$ iteration is

$$q(\beta_i^* | \beta_i^l) = N(\beta_i^l, D_i^l),$$

where the variance of the proposal distribution is

$$D_i^l = C_{ii}^l - C_{i,-i}^l (C_{-i,-i}^l)^{-1} C_{-i,i}^l,$$

where C_{ii}^l is the (i, i) element of C^l , $C_{i,-i}^l$ is the i^{th} row of C^l after removing the i^{th} column, $C_{-i,i}^l$ is the i^{th} column of C^l after removing the i^{th} row, and $C_{-i,-i}^l$ is the submatrix of C^l leaving out the i^{th} row and the i^{th} column with

$$C^l = \begin{cases} C_0, & l \leq l_0, \\ s_d \text{cov}(\beta^1, \dots, \beta^l) + s_d \epsilon I_d, & l > l_0. \end{cases}$$

The matrix C^l is fixed at C_0 for the first l_0 iterations and adaptive to the covariance of the past sample β^1, \dots, β^l afterwards. The specifications of C_0 , s_d and ϵ are provided in Additional file 1.

The delayed rejection [45, 46] modifies the standard Metropolis-Hastings algorithm by delaying the rejection of proposed moves to improve MCMC efficiency in the Peskun sense [51]. When a candidate generated from the proposal distribution is rejected in Metropolis-Hastings, instead of advancing time and retaining the same position, a second candidate move is proposed with the acceptance probability adjusted to preserve the reversibility of the Markov chain relative to the target distribution. Specifically, when the proposed β_i^* is rejected, we will do a further Metropolis step with scaled covariance, i.e.

$$q'(\beta_i^{**} | \beta_i^l, \beta_i^*) = N(\beta_i^l, \gamma D_i^l)$$

where γ is the scale parameter for delayed rejection. A detailed description of the component-wise adaptive Metropolis algorithm with delayed rejection is provided in Additional file 1.

Posterior sampling for error variances

The prior of σ_i^2 , an inverse gamma distribution with shape parameter α_* and scale parameter κ_* , is a conjugate prior based on model (1). In Additional file 1, we show that the conditional posterior distribution of σ_i^2 given β is an inverse gamma distribution with shape parameter $\alpha_* + mT/2$ and scale parameter $\kappa_* + \sum_{t=1}^T \sum_{j=1}^m (\log y_{tji} - \log \mu_{ti})^2/2$, where μ_{ti} is the underlying concentration corresponding to the ODE in model (1) given the current value of β . The posterior sample of σ_i^2 is directly obtained from this distribution.

Comparison of kinetic model parameters between experimental groups

A primary goal of many metabolomic studies is to identify metabolic alterations in response to disease development or drug treatment by comparing kinetic model parameters between two experimental groups, e.g. cancer vs. normal or drug-treated vs. control. Let $\beta^{(1)}$ and $\beta^{(2)}$ be kinetic parameters in two experimental groups, respectively. We focus on the situation of assessing whether the value of a specific kinetic model parameter, e.g. the k^{th} parameter, is equal between experimental groups. Our hypothesis testing problem is:

$$H_0 : \beta_k^{(1)} = \beta_k^{(2)} \text{ vs. } H_1 : \beta_k^{(1)} \neq \beta_k^{(2)}.$$

A standard Bayesian hypothesis testing requires calculation of the following Bayes factor,

$$\frac{\int \dots \int p(\log y | \beta_{-k}^{(1)}, \beta_{-k}^{(2)}, \beta_k^*) p(\beta_k^*, \beta_{-k}^{(1)}, \beta_{-k}^{(2)}) d\beta_k^* d\beta_{-k}^{(1)} d\beta_{-k}^{(2)}}{\int \dots \int p(\log y | \beta^{(1)}, \beta^{(2)}) p(\beta^{(1)}, \beta^{(2)}) d\beta^{(1)} d\beta^{(2)}},$$

where $\beta_{-k}^{(1)}$ and $\beta_{-k}^{(2)}$ are vectors of parameters in $\beta^{(1)}$ and $\beta^{(2)}$ excluding $\beta_k^{(1)}$ and $\beta_k^{(2)}$, respectively, and β_k^* denotes the common value of $\beta_k^{(1)}$ and $\beta_k^{(2)}$ under H_0 . The Bayes factor involves intimidatingly high dimensional integrals that are very difficult to compute with a reasonable accuracy. To address the hypothesis testing problem while circumventing the high-dimensional integration, we propose the following reparameterization:

$$\eta_k = \beta_k^{(1)} - \beta_k^{(2)}.$$

The hypothesis testing problem then becomes

$$H_0 : \eta_k = 0 \text{ vs. } H_1 : \eta_k \neq 0.$$

Importantly, this hypothesis testing can be converted into a parameter estimation problem, where we first estimate η_k by obtaining its posterior distribution, and then check the likelihood of $\eta_k = 0$ by constructing a credible interval and further calculating the tail probability under its posterior distribution.

Specifically, we jointly model data from the two experimental groups with the reparameterization in effect:

$$\begin{aligned} \log y_{ij}^{(1)} &= \log \mu_t^{(1)} + \delta_{ij}^{(1)}, \\ \log y_{ij}^{(2)} &= \log \mu_t^{(2)} + \delta_{ij}^{(2)}, \\ \frac{d\mu_t^{(1)}}{dt} &= f(\mu_t^{(1)}; \beta_{(-k)}^{(1)}, \beta_k^{(1)}, t), \\ \frac{d\mu_t^{(2)}}{dt} &= f(\mu_t^{(2)}; \beta_{(-k)}^{(2)}, \beta_k^{(1)}, \eta_k, t), \end{aligned}$$

where variables/parameters are defined similarly as before with an additional superscript to indicate the experimental group. We assume that $\delta_{ij}^{(1)}$ and $\delta_{ij}^{(2)}$ are independent and their i th elements follow normal distributions with mean 0 and variance $\sigma_{i,(1)}^2$ and $\sigma_{i,(2)}^2$, respectively.

The prior of $\beta^{(1)}$ and $\beta_{-k}^{(2)}$ are specified based on the expert’s knowledge and the prior of each $\sigma_{i,(1)}^2, \sigma_{i,(2)}^2$ is specified by an inverse gamma distribution using the shrinkage approach. The prior of η_k is specified as a normal distribution with mean zero and variance equal to the variance of $\beta_k^{(1)}$. With some simple modification of the MCMC algorithm proposed in the last subsection, we obtain posterior samples of parameters $(\beta_{-k}^{(1)}, \beta_{-k}^{(2)}, \beta_k^{(1)}, \eta_k, \sigma_{(1)}^2, \sigma_{(2)}^2)$ with an additional sampling step for η_k . We then perform the hypothesis testing based on the posterior credible interval of η_k , where the two specific kinetic model parameters are considered as different if the 95% credible interval of η_k does not cover 0. To further quantify the level of evidence supporting the null hypothesis, we investigate the highest level of credible interval that does not cover 0 and define a *credible value*

$$p = 1 - \max_{0 \leq \alpha \leq 1, 0 \notin HDI_\alpha} \alpha,$$

where HDI_α denotes the highest density interval of level α under the posterior distribution of η_k . This credible value reflects how extreme the null value is with respect to our posterior belief about η_k . In other words, it can be used as an empirical rule regarding how likely the null hypothesis value is correct. A similar idea has recently been investigated in simpler models in clinical trial studies [52]. As we will show in the simulation studies subsection, the distribution of the above credible values is concentrated around zero when the true values for $\beta_k^{(1)}$ and $\beta_k^{(2)}$ are different, while it is much more scattered between zero and one when the true values for $\beta_k^{(1)}$ and $\beta_k^{(2)}$ are the same. Thus, the credible value has a similar behavior as the frequentist’s p-value. We consider a credible value < 0.05 as statistically significant. As the credible value is calculated based on HDI_α that can reflect deviation of η_k from zero in either positive or negative direction, our test is a two-sided test.

The proposed algorithm could be applied in other scenarios (e.g. whether there is a difference in metabolic flux between groups) when the object of interest concerns the comparison of multiple kinetic parameters, more details are given in the Discussion section.

Results

Real data analysis

We validated our proposed Bayesian framework by using published data in [41]. Fan et al. [41] employed multiplexed SIRM (i.e. multiple labeled substrates in the same

experiment) to track the metabolism of $^{13}\text{C}_6\text{-Glc}$, $D_2\text{-glycine}$, $^{13}\text{C}_2\text{-glycine}$, and $D_3\text{-serine}$ into purine nucleotides in thin slices of cancerous and matched noncancerous lung tissues freshly resected from a patient with lung squamous cell carcinoma. The data include abundances of isotopologues of serine and glycine and purine nucleotides in cancer and non-cancer tissue slices of that patient with three replicates, in both the tissue and the medium. Kinetic models were constructed and a weighted least square-based method was used to estimate model parameters. Fan *et al.* [41] considered a model with three distinct pools of cytoplasmic Ser/Gly (Fig. 1), where one pool is derived from $^{13}\text{C}_6\text{-Glc}$, the second is from exogenous $D_3\text{-Ser}$ or $D_2\text{-Gly}$, and the third is from unlabeled sources. However, no statistical confidence interval or p-value was available due to the limitation of the weighted least squares-based method used in that paper. In this subsection, we use our proposed methods to reanalyze the data and provide more rigorous statistical inference. Initial ranges of kinetic parameters (Additional File 2: Table S1) were taken from literature sources for the appropriate human isoforms [48, 53].

Kinetic parameter estimation

Following our proposed method, we obtained the posterior samples for all the β parameters in model (1) based on 5500 MCMC iterations after 7500 burn-in iterations. The convergence of the MCMC chains was examined by the Geweke test (Additional File 2: Table S2). By inserting the β values into model (1), we estimated the 24-hour concentrations of isotopologues. Fig. 2 compares those estimated log concentrations \pm posterior standard deviations with mean log concentration \pm standard deviation in the observed data for cancer and non-cancer groups, respectively. The estimated concentrations were very close to the observed ones for most isotopologues except for some low abundance ones. Likewise, the variations under the proposed model were also comparable to those presented by the observed data for most isotopologues except for some low abundance ones. These results indicate that our estimators perform well.

We next investigated the posterior mean and standard deviation over time for two fluxes, $V_{Glc-SerPool1-13C0-13C0D000}$ and $V_{Glc-SerPool1-13C1-13C3D000}$, from the cancer

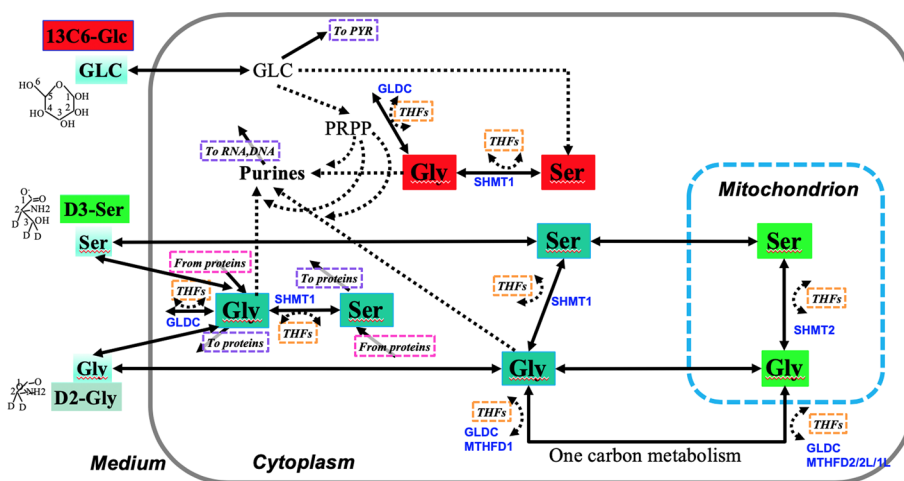


Fig. 1 Tracer labeled three-pool purine synthesis process

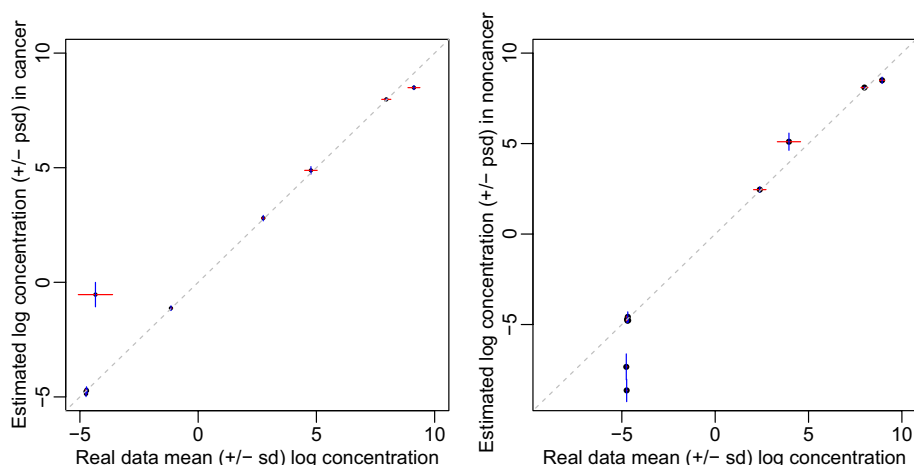


Fig. 2 Comparison of estimated and observed isotopologue concentrations in log scale in lung cancer (left panel) and non-cancer (right panel) tissues. A red horizontal line indicates the interval “mean log concentration \pm standard deviation (sd)” of observed concentrations for an isotopologue. A blue vertical line indicates the interval “estimated log concentrations \pm posterior standard deviation (psd)” of the corresponding concentration. The gray dashed line is $y = x$

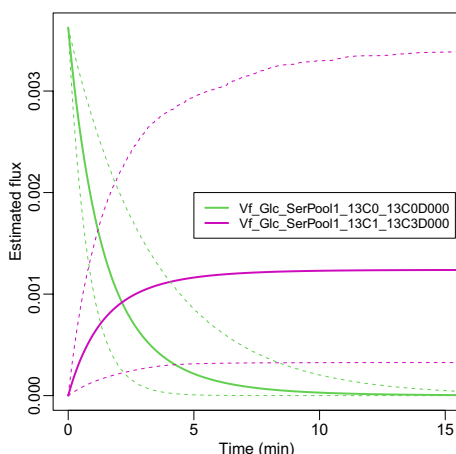


Fig. 3 Estimated flux changes over time in lung cancer tissues. Solid curves: estimated fluxes; Dashed curves: 95% credible bounds

group, as shown in Fig. 3. Based on Fig. 1, we had a non-zero concentration for labeled *Glc* in medium and unlabeled *Glc* in the cytoplasm at time zero. On the one hand, the labeled *Glc* concentration in the cytoplasm increased with time, and the forward flux from labeled *Glc* to labeled *Ser* in cytoplasm also tended to increase. On the other hand, the unlabeled *Glc*’s concentration decreased owing to the activity of forward reactions to other metabolites; similarly the forward flux from unlabeled *Glc* to unlabeled *Ser* in cytoplasm tended to decrease too, as expected for actively metabolizing cells.

Group comparison

We compared each kinetic model parameter between cancer and non-cancer tissues based on our proposed hypothesis testing procedure. Specifically, we calculated the

95% credible interval for the difference in parameter value between cancer and non-cancer tissues and checked whether it covered zero. A credible value was further calculated to quantify the significance level of the test. We obtained the posterior samples based on 2000 MCMC iterations after 5500 burn-in iterations. The convergence of the MCMC chains was examined by the Geweke test (Additional File 2: Table S3). Results are shown in Fig. 4. We found that the forward flux from glucose to serine in pool 1 ($K_{f_GLC-SerPool1}$) is significantly higher in cancer than non-cancer tissues (credible value < 0.001). This result is consistent with the result reported in [41], providing evidence for the validity of our method. Further, compared to [41], our method is able to infer the difference in kinetic model parameters in a statistically rigorous way.

Simulation studies

We conducted simulation studies to evaluate the performance of our method. Data were simulated based on Model (1) with parameters specified according to the real data from [41]. Specifically, we considered the three-pool model, which contains 608 ODEs and 52 parameters, from [41] and borrowed the kinetic parameter estimates from [41] and accordingly set the true values of β in our simulations. As to the variances of the assumed normally distributed data errors, we borrowed the variation information in the real data with three replicates in [41] and randomly generated the σ^2 values from an inverse Gamma distribution which took the shrinkage mean and variance of the real data replicates' variation as the mean and variance. These randomly generated values

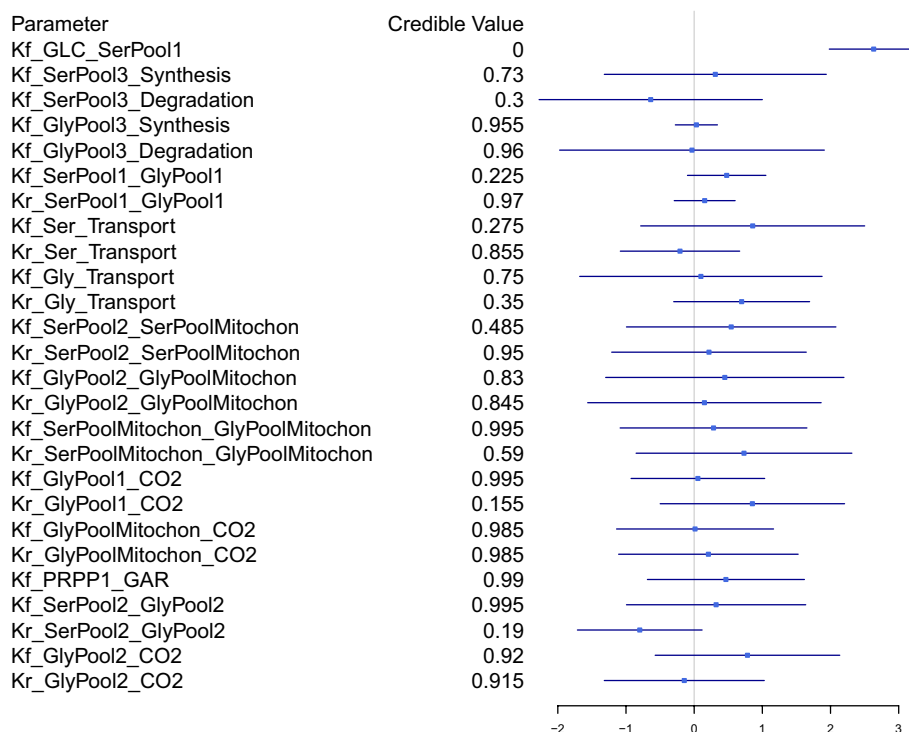


Fig. 4 Comparison of each kinetic parameter between lung cancer and non-cancer tissues. The first column lists parameter names, the second column lists credible values and the third column lists 95% credible intervals for the differences in parameter values between cancer and non-cancer tissues

were taken as the true values of the data error variances in our simulation studies. Then our simulated concentration data were generated accordingly based on the assumed normal distributed errors.

Parameter estimation

Our first set of simulations evaluated the performance of our proposed parameter estimation method. We simulated data with 3 replicates based on the cancer data from [41]. We applied our method to generate posterior samples from 15,000 MCMC iterations, where the adaptive sampling started after 3000 burn-in iterations and the delayed rejection started after another 1000 iterations. The simulations were repeated 30 times, i.e. 30 separate datasets were generated. We examined the difference between the estimated parameter value (posterior mean) and true value. Figure 5 plots the averaged difference across 30 simulation runs ± 2 se, indicated by a gray box, for each of the 52 kinetic parameters, where se is the standard error of the difference between estimated and true parameter values across 30 simulation runs. Almost all of the boxes (51 out of

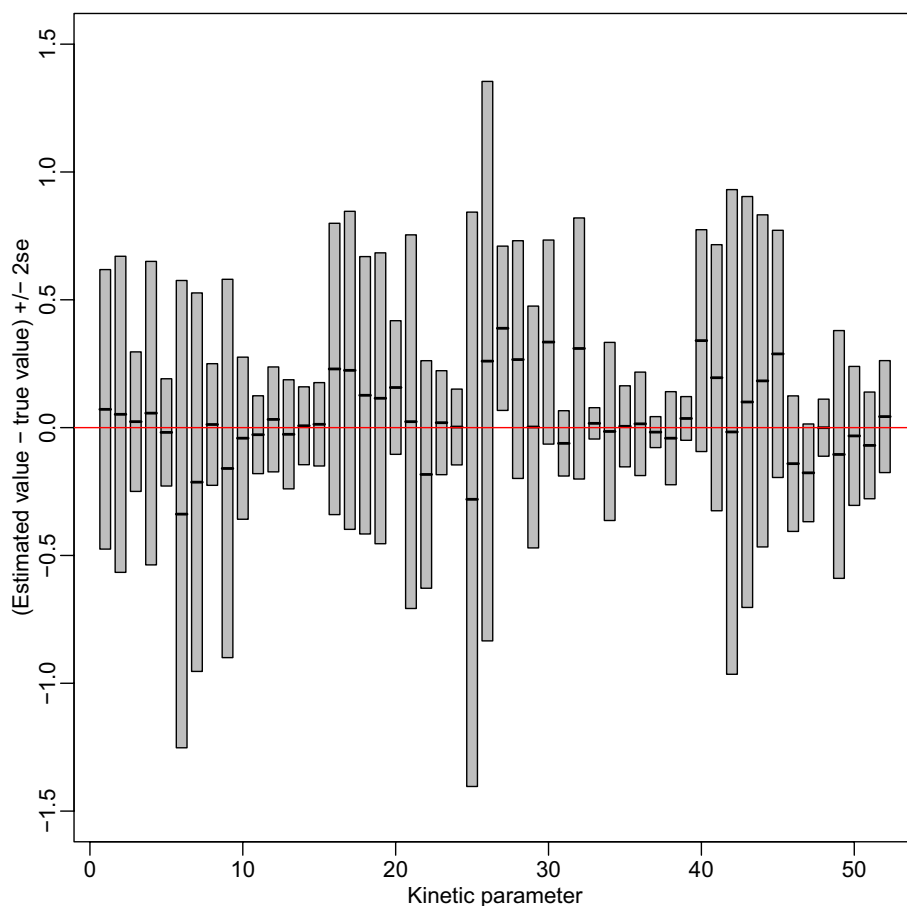


Fig. 5 Difference between estimated and true parameter values. Each box presents result for one kinetic parameter. The thick black line in the middle of a box indicates the averaged difference between estimated and true parameter values across 30 simulation runs. The top and bottom of the box indicate averaged difference ± 2 se, respectively. The horizontal red line at zero indicates the target value (no difference between estimated and true values)

52) covered zero, indicating that the true parameter value was within 2 se of the estimated values. The one exception was β_{27} , where the true value was outside 2 se. But it was still within 2.5 se of the averaged estimated value. In summary, our estimated values were close to true values relative to the standard errors of the estimated values. Therefore, the simulation results demonstrate that the performance of our estimation method was satisfying.

Group comparison

Our second set of simulations assessed the performance of the hypothesis testing method. Data were simulated following the same procedure as the first set of cancer simulations (superscripted by (1)), except that a second non-cancer experimental group (superscripted by (2)) was added. We tested the hypothesis of $\beta_{10}^{(1)} = \beta_{10}^{(2)}$ and considered the following three scenarios: $\beta_{10}^{(1)} = \beta_{10}^{(2)}$, $\beta_{10}^{(2)} = \beta_{10}^{(1)} - 0.2$, and $\beta_{10}^{(2)} = \beta_{10}^{(1)} - 0.35$.

We applied our proposed testing method, which converts the testing problem into one problem of estimating $\eta_{10} = \beta_{10}^{(1)} - \beta_{10}^{(2)}$ and rejects the null hypothesis if the 95% credible interval of η_{10} does not cover zero, or equivalently, the credible value is smaller than 0.05. For every simulation, we used 3000 burn-in steps before heading into adaptive phase, and after further 1000 steps, we added delayed rejection. The simulations were repeated 20 times (i.e. 20 datasets were generated) for each scenario and the posterior samples contained 10,000 steps in total. Under the first scenario where the null hypothesis held (Fig. 6A and B), the 95% credible interval covered 0 and thus did not reject the null hypothesis in most of the 20 simulations. The distribution of credible value over 20 simulations was dispersed with many of them having large values, also suggesting the null hypothesis was likely to hold. Under the second scenario where there was a moderate difference between $\beta_{10}^{(1)}$ and $\beta_{10}^{(2)}$ (Fig. 6C and D), the 95% credible interval did not cover 0 in most simulations and the distribution of credible value was very concentrated around small values, both suggesting the null hypothesis was unlikely to hold. Under the third scenario where the true difference was larger (Fig. 6E and F), the 95% credible interval did not cover 0 in almost all simulations and the distribution of credible values all approached to zero value, both strongly suggesting that the null hypothesis should be rejected. In summary, our proposed reparameterization method for hypothesis testing tended to not rejecting the null hypothesis under the scenario that the null hypothesis held, and rejecting the null hypothesis under the scenarios that the null hypothesis did not held. Therefore, these simulation results indicate that our hypothesis testing method was well performed.

Discussion

Our method has close points of contact with [54], where the authors developed a novel Bayesian approach for the inference of ODEs for characterizing transcription factor activities based on gene expression data. Our method extends this Bayesian idea in multiple ways. First, we propose a new reparameterization method and credible value-based procedure that enables Bayesian hypothesis testing for comparing kinetic parameters between experimental groups. Second, we introduce component-wise adaptive Metropolis and delayed rejection methods to improve the MCMC

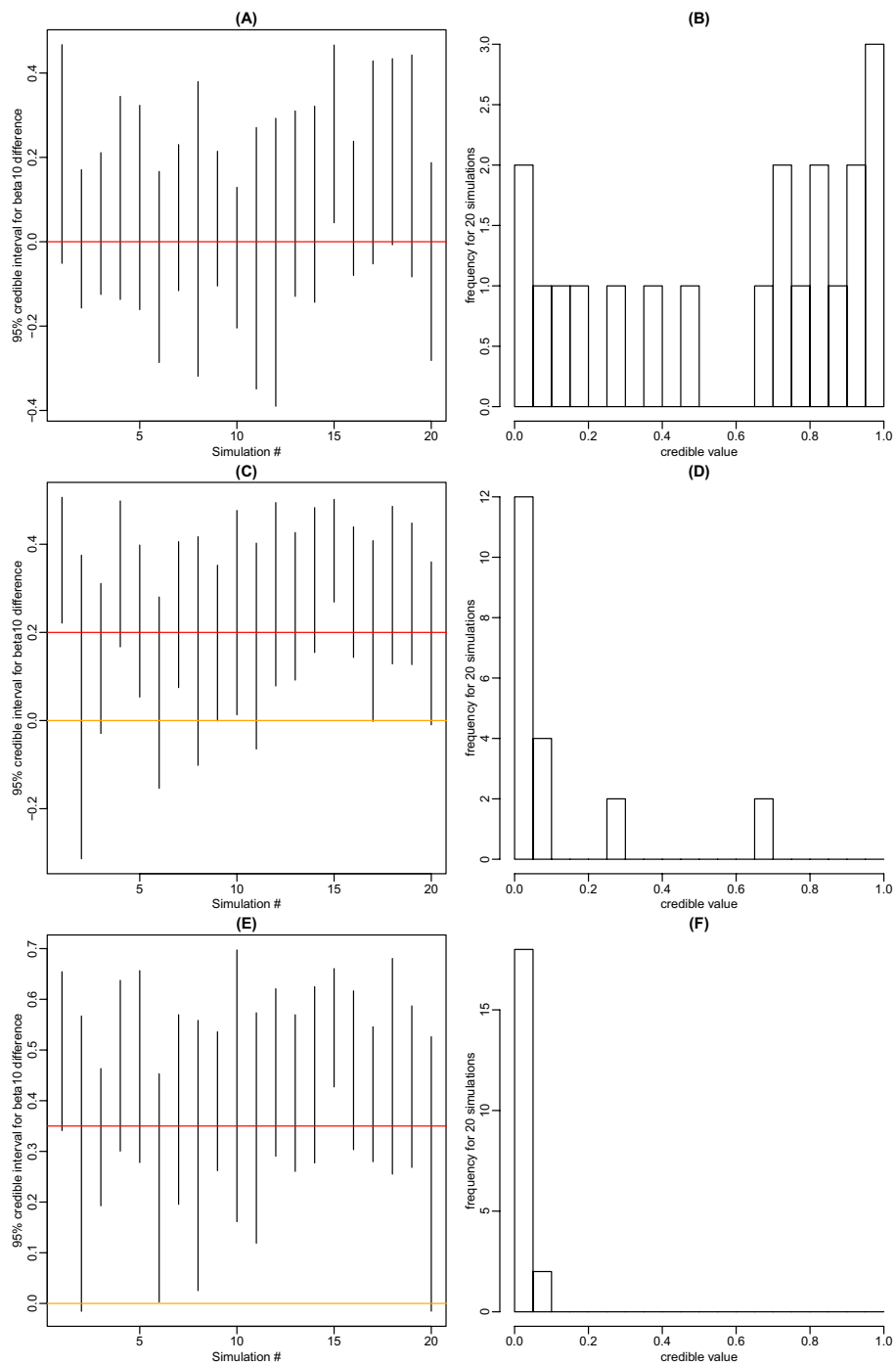


Fig. 6 The 95% credible interval of $\beta_{10}^{(1)} - \beta_{10}^{(2)}$ and distribution of the credible value for testing the hypothesis of $\beta_{10}^{(1)} = \beta_{10}^{(2)}$ under the following three scenarios: (A) and (B), $\beta_{10}^{(1)} = \beta_{10}^{(2)}$; (C) and (D), $\beta_{10}^{(1)} - \beta_{10}^{(2)} = 0.2$; and (E) and (F), $\beta_{10}^{(1)} - \beta_{10}^{(2)} = 0.35$. A red line indicates the true value of $\beta_{10}^{(1)} - \beta_{10}^{(2)}$ and an orange line at zero indicates the value of $\beta_{10}^{(1)} - \beta_{10}^{(2)}$ under the null hypothesis. Results are based on 20 simulation runs

efficiency for complex ODE model systems with high-dimensional parameters. Third, we propose a Bayesian shrinkage prior to stabilize variance parameter estimation under limited sample size.

Our hypothesis testing method for comparing kinetic model parameters can be extended to identify the between-group difference in metabolic flux, which is the net rate of flow of atoms or metabolic subunits through a metabolic pathway or network segment. Based on the kinetic model, the flux can be expressed as a function of t and β_k , where β_k is a sub-vector of β that are related to the isotopic flow through the k^{th} metabolite. To determine whether the flux stays the same between experimental groups, it is sufficient to evaluate whether $\beta_k^{(1)}$ and $\beta_k^{(2)}$ are equal, where $\beta_k^{(1)}$ and $\beta_k^{(2)}$ are values of β_k in the two groups, respectively. Consider the reparameterization that $\eta_k = \beta_k^{(2)} - \beta_k^{(1)}$, the aforementioned hypothesis testing problem becomes $H_0 : \eta_k = 0$ vs. $H_1 : \eta_k \neq 0$. We can jointly model data from the two experimental groups and obtain posterior samples of η_k . The 95% highest density credible region of η_k can then be constructed to assess whether the null hypothesis holds. A credible value can also be defined to quantify the level of evidence supporting the null hypothesis. Given the straightforward extension to comparing a vector of parameters, technical difficulties remain in constructing multi-dimensional credible regions. Some successes have been found in the bivariate case [55] or log-concave posterior distributions [56]. However, these algorithms cannot be used in our case since neither the explicit form nor the log-concavity property of the posterior distribution is known. In addition, our current testing strategy is biologically informative because each kinetic parameter in the ODEs is biologically meaningful. The pathway flux can be controlled mainly by one or several enzymes in the pathway, and the ranking of the enzymes by the significance level of their corresponding β 's helps determine the relative importance of each enzyme in the pathway for flux control (e.g. [40, 57]). This also allows testing of a specific hypothesis that a particular enzyme contributes most to flux control, which would then be a potential therapeutic target.

The credible value approach is easy to implement and practically valid based on simulation studies and real data analysis. Its theoretical performance needs further investigation. We refer to [58] and postulate that the credible value has an asymptotically uniform distribution when the null hypothesis is true, that is, the parameters are identical between cancer and non-cancer group. Further, Xie and Singh [58] demonstrated the asymptotic equivalence of posterior distribution with confidence distribution, which supports our approach of doing hypothesis testing through evaluating the “significance” of null hypothesis under any confidence distribution, and thus through credible value.

An alternative MCMC algorithm, the geometric based sampling methods (mMALA and RMHMC included) [59, 60] may be considered for our model. Those methods, shown to be more efficient than the standard Metropolis-Hastings algorithm and can effectively handle highly correlated samples [59, 60], could be more suitable for applications when the dynamic system presents partially unidentifiable structures commonly seen in biochemical networks [61]. Implementating these methods relies on a fast and accurate solver for the auxiliary sensitivity equations (Eq. (2.1) in [61]) at each sampling iteration, which remains to be explored given the large number of ODE equations and dimension of parameters in our application. This will be our interest as a future research.

Conclusions

In this paper, we have developed a Bayesian framework for non-steady-state kinetic modeling analysis of SIRM data. By using prior distributions to incorporate biological knowledge and integrate information across metabolites, our method provides robust estimation of model parameters. By introducing a new reparameterization and credible value-based inference procedure, our method allows comparison of kinetic model parameters between experimental groups in a statistically rigorous way. Real data analysis and simulation studies demonstrate that our framework performs well in practical situations.

Abbreviations

SIRM	Stable isotope resolved metabolomics
ODEs	Ordinary differential equations
MCMC	Markov chain Monte Carlo
psd	Posterior standard deviation
sd	Standard deviation
se	Standard error
$^{13}\text{C}_6\text{-Glc}$	^{13}C -enriched glucose.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05211-5>.

Additional file 1. Supplemental Methods.

Additional file 2. Supplemental Tables.

Acknowledgements

Not applicable.

Author Contributions

CW, TWMF, ANL and AJS designed the study. XZ, YS and CW derived the method. XZ implemented the method and performed simulation studies and real data analyses. TWMF and ANL provided the real data and interpreted the data analysis results. XZ, YS, TWMF, AJS, ANL and CW wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Institutes of Health [1R03CA211835, 5P20GM103436-15, 1P01CA163223-01A1], the Biostatistics and Bioinformatics and Redox Metabolism Shared Resource Facilities of the University of Kentucky Markey Cancer Center [P30CA177558] and University of Kentucky Center for Cancer and Metabolism [P20GM121327].

Availability of data and materials

Our method is freely available for academic use at <https://github.com/xuzhang0131/MCMCFlux>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 August 2022 Accepted: 24 February 2023

Published: 22 March 2023

References

1. Fan TW-M, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM. Altered regulation of metabolic pathways in human lung cancer discerned by ^{13}C stable isotope-resolved metabolomics (SIRM). *Mol Cancer*. 2009;8(1):41.

2. Lane AN, Fan TW-M, Xie Z, Moseley HN, Higashi RM. Isotopomer analysis of lipid biosynthesis by high resolution mass spectrometry and NMR. *Anal Chim Acta*. 2009;651(2):201–8.
3. Fan TW-M, Lane AN, Higashi RM, Yan J. Stable isotope resolved metabolomics of lung cancer in a SCID mouse model. *Metabolomics*. 2011;7(2):257–69.
4. Moseley HN, Lane AN, Belshoff AC, Higashi RM, Fan TW-M. A novel deconvolution method for modeling UDP-N-Acetyl-D-glucosamine biosynthetic pathways based on ¹³C mass isotopologue profiles under non-steady-state conditions. *BMC Biol*. 2011;9(1):37.
5. Le A, Lane AN, Hamaker M, Bose S, Gouw A, Barbi J, Tsukamoto T, Rojas CJ, Slusher BS, Zhang H. Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. *Cell Metab*. 2012;15(1):110–21.
6. Fan TW-M, Tan J, McKinney MM, Lane AN. Stable isotope resolved metabolomics analysis of ribonucleotide and RNA metabolism in human lung cancer cells. *Metabolomics*. 2012;8(3):517–27.
7. Fan TW-M, Lorkiewicz PK, Sellers K, Moseley HN, Higashi RM, Lane AN. Stable isotope-resolved metabolomics and applications for drug development. *Pharmacol Ther*. 2012;133(3):366–91.
8. Lorkiewicz P, Higashi RM, Lane AN, Fan TW-M. High information throughput analysis of nucleotides and their isotopically enriched isotopologues by direct-infusion FTICR-MS. *Metabolomics*. 2012;8(5):930–9.
9. Fan TW-M, Lane AN. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc*. 2016;92:18–53.
10. Fan TW-M, Warmoes MO, Sun Q, Song H, Turchan-Cholewo J, Martin JT, Mahan A, Higashi RM, Lane AN. Distinctly perturbed metabolic networks underlie differential tumor tissue damages induced by immune modulator β -glucan in a two-case ex vivo non-small-cell lung cancer study. *Molecular Case Studies*. 2016;2(4): 000893.
11. Alarcon-Barrera JC, Kostidis S, Ondo-Mendez A, Giera M. Recent advances in metabolomics analysis for early drug development. *Drug Discov Today*. 2022;27:1763–73.
12. Yuneva MO, Fan TW-M, Allen TD, Higashi RM, Ferraris DV, Tsukamoto T, Matés JM, Alonso FJ, Wang C, Seo Y. The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab*. 2012;15(2):157–70.
13. Xie H, Hanai J-I, Ren J-G, Kats L, Burgess K, Bhargava P, Signoretti S, Billiard J, Duffy KJ, Grant A, Wang X, Lorkiewicz PK, Schatzman S, Bousamra M, Lane AN, Higashi RM, Fan TW-M, Pandolfi PP, Sukhatme VP, Seth P. Targeting lactate dehydrogenase- α inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell Metab*. 2014;19(5):795–809.
14. Sellers K, Fox MP, Bousamra M, Slone SP, Higashi RM, Miller DM, Wang Y, Yan J, Yuneva MO, Deshpande R, Lane AN, Fan TW-M. Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. *J Clin Investig*. 2015;125(2):687–98.
15. Fan TW-M, Warmoes MO, Sun Q, Song H, Turchan-Cholewo J, Martin JT, Mahan A, Higashi RM, Lane AN. Distinctly perturbed metabolic networks underlie differential tumor tissue damages induced by immune modulator β -glucan in a two-case ex vivo non-small-cell lung cancer study. *Molecular Case Studies*. 2016;2(4): 000893.
16. Jung SM, Le J, Doxsey WG, Haley JA, Park G, Guertin DA, Jang C. Stable isotope tracing and metabolomics to study in vivo brown adipose tissue metabolic fluxes. In: *Brown adipose tissue*, 2022;119–130. Humana, New York, NY.
17. Lu D, Mulder H, Zhao P, Burgess SC, Jensen MV, Kamzolova S, Newgard CB, Sherry AD. ¹³C NMR isotopomer analysis reveals a connection between pyruvate cycling and glucose-stimulated insulin secretion (GSIS). *Proc Natl Acad Sci*. 2002;99(5):2708–13.
18. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, Thompson CB. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci*. 2007;104(49):19345–50.
19. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, Micarioni M, Chaneton B, Adam J, Hedley A. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*. 2011;477(7363):225–8.
20. Mullen AR, Wheaton WW, Jin ES, Chen P-H, Sullivan LB, Cheng T, Yang Y, Linehan WM, Chandel NS, DeBerardinis RJ. Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature*. 2012;481(7381):385–8.
21. Lewis CA, Parker SJ, Fiske BP, McCloskey D, Gui DY, Green CR, Vokes NI, Feist AM, Vander Heiden MG, Metallo CM. Tracing compartmentalized NADPH metabolism in the cytosol and mitochondria of mammalian cells. *Mol Cell*. 2014;55(2):253–63.
22. DeNicola GM, Chen P-H, Mullarky E, Sudderth JA, Hu Z, Wu D, Tang H, Xie Y, Asara JM, Huffman KE. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat Genet*. 2015;47(12):1475.
23. Lin P, Dai L, Crooks DR, Neckers LM, Higashi RM, Fan TW, Lane AN. NMR methods for determining lipid turnover via stable isotope resolved metabolomics. *Metabolites*. 2021;11(4):202.
24. Steuer R, Gross T, Selbig J, Blasius B. Structural kinetic modeling of metabolic networks. *Proc Natl Acad Sci*. 2006;103(32):11868–73.
25. Jamshidi N, Palsson BØ. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*. 2008;4(1):171.
26. Saa PA, Nielsen LK. Construction of feasible and accurate kinetic models of metabolism: a Bayesian approach. *Sci Rep*. 2016;6(1):1–13.
27. Saa PA, Nielsen LK. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnol Adv*. 2017;35(8):981–1003.
28. Cuperlovic-Culf M. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites*. 2018;8(1):4.
29. Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA: a resource for enzyme data and metabolic information. New York: Elsevier; 2002.

30. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W. Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments. *J Biotechnol*. 2007;129(2):249–67.
31. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol Bioeng*. 2008;99(3):686–99.
32. de Mas IM, Selivanov VA, Marin S, Roca J, Orešič M, Agius L, Cascante M. Compartmentation of glycogen metabolism revealed from ^{13}C isotopologue distributions. *BMC Syst Biol*. 2011;5(1):175.
33. Wiechert W, Nöh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr Opin Biotechnol*. 2013;24(6):979–86.
34. Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*. 2014;30(9):1333–5.
35. Foguet C, Marin S, Selivanov VA, Fanchon E, Lee W-NP, Guinovart JJ, de Atauri P, Cascante M. Hepatodyn: a dynamic model of hepatocyte metabolism that integrates ^{13}C isotopomer data. *PLoS Comput Biol*. 2016;12(4):1004899.
36. Resat H, Petzold L, Pettigrew MF. In: Iretton R, Montgomery K, Bumgarner R, Samudrala R, McDermott J, editors. *Kinetic modeling of biological systems*. Totowa: Humana Press; 2009. p. 311–35.
37. Selivanov VA, Vizán P, Mollinedo F, Fan TW-M, Lee PW, Cascante M. Edelfosine-induced metabolic changes in cancer cells that precede the overproduction of reactive oxygen species and apoptosis. *BMC Syst Biol*. 2010;4(1):135.
38. Cascante M, Selivanov V, Ramos-Montoya A. In: Fan TW-M, Lane AN, Higashi RM, editors. *Application of tracer-based metabolomics and flux analysis in targeted cancer drug design*. Totowa, NJ: Humana Press; 2012. p. 299–320.
39. Saa PA, Nielsen LK. Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. *Biotechnol Adv*. 2017;35(8):981–1003.
40. Selivanov VA, Marin S, Lee PW, Cascante M. Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. *Bioinformatics*. 2006;22(22):2806–12.
41. Fan TW-M, Bruntz RC, Yang Y, Song H, Chernyavskaya Y, Deng P, Zhang Y, Shah PP, Beverly LJ, Qi Z. De novo synthesis of serine and glycine fuels purine nucleotide biosynthesis in human lung cancer tissues. *J Biol Chem*. 2019;294(36):13464–77.
42. Haario H, Saksman E, Tamminen J. Componentwise adaptation for high dimensional MCMC. *Comput Stat*. 2005;20(2):265–73.
43. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli*. 2001;7(2):223–42.
44. Haario H, Laine M, Mira A, Saksman E. DRAM: efficient adaptive MCMC. *Stat Comput*. 2006;16(4):339–54.
45. Tierney L, Mira A. Some adaptive Monte Carlo methods for Bayesian inference. *Stat Med*. 1999;18(17–18):2507–15.
46. Mira A. On Metropolis-Hastings algorithms with delayed rejection. *Metron*. 2001;59(3–4):231–41.
47. Kreuz C, Rodriguez MB, Maiwald T, Seidl M, Blum H, Mohr L, Timmer J. An error model for protein quantification. *Bioinformatics*. 2007;23(20):2747–53.
48. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D. Brenda, the elixir core data resource in 2021: new developments and updates. *Nucleic Acids Res*. 2021;49(D1):498–508.
49. Wu H, Wang C, Wu Z. A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics*. 2013;14(2):232–43.
50. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*. 1984;6:721–41.
51. Peskun PH. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*. 1973;60(3):607–12.
52. Quan H, Zhang B, Lan Y, Luo X, Chen X. Bayesian hypothesis testing with frequentist characteristics in clinical trials. *Contemp Clin Trials*. 2019;87: 105858.
53. Wittig U, Rey M, Weidemann A, Kania R, Müller W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Res*. 2018;46(D1):656–60.
54. Rogers S, Khanin R, Girolami M. Bayesian model-based inference of transcription factor activity. *BMC Bioinform*. 2007;8(2):1–11.
55. Turkkan N, Pham-Gia T. Highest posterior density credible region and minimum area confidence region: the bivariate case. *Appl Stat*. 1997;46:131–40.
56. Pereyra M. Maximum-a-posteriori estimation with Bayesian confidence regions. *SIAM J Imag Sci*. 2017;10(1):285–302.
57. Joy MP, Elston TC, Lane AN, Macdonald JM, Cascante M. Introduction to metabolic control analysis (MCA). In: *The Handbook of Metabolomics*. Springer; 2012. p. 279–97.
58. Xie M-G, Singh K. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int Stat Rev*. 2013;81(1):3–39.
59. Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian monte Carlo methods. *J Royal Stat Soc: Series B (Stat Methodol)*. 2011;73(2):123–214.
60. Kramer A, Stathopoulos V, Girolami M, Radde N. MCMC_CLIB—an advanced MCMC sampling package for ODE models. *Bioinformatics*. 2014;30(20):2991–2.
61. Calderhead B, Girolami M. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface focus*. 2011;1(6):821–35.
62. Hug S, Raue A, Hasenauer J, Bachmann J, Klingmüller U, Timmer J, Theis F. High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. *Math Biosci*. 2013;246(2):293–304.
63. Sun RC, Fan TW-M, Deng P, Higashi RM, Lane AN, Le A-T, Scott TL, Sun Q, Warmoes MO, Yang Y. Noninvasive liquid diet delivery of stable isotopes into mouse models for deep metabolic network tracing. *Nat Commun*. 2017;8(1):1–10.
64. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
65. Chen C, Gonzalez FJ, Idle JR. LC-MS-based metabolomics in drug metabolism. *Drug Metab Rev*. 2007;39(2–3):581–97.

66. West FD, Henderson WM, Yu P, Yang J-Y, Stice SL, Smith MA. Metabolomic response of human embryonic stem cell-derived germ-like cells after exposure to steroid hormones. *Toxicol Sci.* 2012;129(1):9–20.
67. Kamburov A, Cavill R, Ebbels TM, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics.* 2011;27(20):2917–8.
68. Heijne WH, Lamers R-JA, van Bladeren PJ, Groten JP, van Nesselrooij JH, Van Ommen B. Profiles of metabolites and gene expression in rats with chemically induced hepatic necrosis. *Toxicol Pathol.* 2005;33(4):425–33.
69. Heijne WH, Jonker D, Stierum RH, van Ommen B, Groten JP. Toxicogenomic analysis of gene expression changes in rat liver after a 28-day oral benzene exposure. *Mut Res/Fund Mol Mech Mutagenesis.* 2005;575(1–2):85–101.
70. Morvan D, Demidem A. Metabolomics by proton nuclear magnetic resonance spectroscopy of the response to chloroethylnitrosourea reveals drug efficacy and tumor adaptive metabolic pathways. *Can Res.* 2007;67(5):2150–9.
71. Ho JE, Larson MG, Vasani RS, Ghorbani A, Cheng S, Rhee EP, Florez JC, Clish CB, Gerszten RE, Wang TJ. Metabolite profiles during oral glucose challenge. *Diabetes.* 2013;62(8):2689–98.
72. Kankainen M, Gopalacharyulu P, Holm L, Orešič M. MPEA—metabolite pathway enrichment analysis. *Bioinformatics.* 2011;27(13):1878–9.
73. Yoo H, Antoniewicz MR, Stephanopoulos G, Kelleher JK. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *J Biol Chem.* 2008;283(30):20621–7.
74. Wise DR, Ward PS, Shay JE, Cross JR, Gruber JJ, Sachdeva UM, Platt JM, DeMatteo RG, Simon MC, Thompson CB. Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability. *Proc Natl Acad Sci.* 2011;108(49):19611–6.
75. Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, Jewell CM, Johnson ZR, Irvine DJ, Guarente L. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature.* 2012;481(7381):380–4.
76. Yang Y, Lane AN, Ricketts CJ, Sourbier C, Wei M-H, Shuch B, Pike L, Wu M, Rouault TA, Boros LG. Metabolic reprogramming for producing energy and reducing power in fumarate hydratase null cells from hereditary leiomyomatosis renal cell carcinoma. *PLoS ONE.* 2013;8(8):72179.
77. Jiang L, Shestov AA, Swain P, Yang C, Parker SJ, Wang QA, Terada LS, Adams ND, McCabe MT, Pietrak B. Reductive carboxylation supports redox homeostasis during anchorage-independent growth. *Nature.* 2016;532(7598):255–8.
78. Voit E, Qi Z, Miller G. Steps of modeling complex biological systems. *Pharmacopsychiatry.* 2008;41(1):78–84.
79. Qi Z, Miller G, Voit E. A mathematical model of presynaptic dopamine homeostasis: implications for schizophrenia. *Pharmacopsychiatry.* 2008;41(1):89–98.
80. Bernardo J, Berger J, Dawid A, Smith A. Efficient metropolis jumping rules. In: *Bayesian Statistics*, vol. 5. New York: Oxford University Press; 1996.
81. Lane AN, Fan TW-M. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res.* 2015;43(4):2466–85.
82. Liu Y-C, Li F, Handler J, Huang CRL, Xiang Y, Neretti N, Sedivy JM, Zeller KI, Dang CV. Global regulation of nucleotide biosynthetic genes by c-Myc. *PLoS ONE.* 2008;3(7):2722.
83. Hori H, Tran P, Carrera CJ, Hori Y, Rosenbach MD, Carson DA, Nobori T. Methylthioadenosine phosphorylase cDNA transfection alters sensitivity to depletion of purine and methionine in A549 lung cancer cells. *Can Res.* 1996;56(24):5653–8.
84. Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends Biochem Sci.* 2014;39(4):199–218.
85. Wikoff WR, Grapov D, Fahrman JF, DeFelice B, Rom WN, Pass HI, Kim K, Nguyen U, Taylor SL, Gandara DR. Metabolomic markers of altered nucleotide metabolism in early stage adenocarcinoma. *Cancer Prev Res.* 2015;8(5):410–8.
86. Tedeschi PM, Vazquez A, Kerrigan JE, Bertino JR. Mitochondrial methylenetetrahydrofolate dehydrogenase (MTHFD2) overexpression is associated with tumor cell proliferation and is a novel target for drug development. *Mol Cancer Res.* 2015;13(10):1361–6.
87. Maddocks OD, Athineos D, Cheung EC, Lee P, Zhang T, van den Broek NJ, Mackay GM, Labuschagne CF, Gay D, Kruiswijk F. Modulating the therapeutic response of tumours to dietary serine and glycine starvation. *Nature.* 2017;544(7650):372–6.
88. Paone A, Marani M, Fiascarelli A, Rinaldo S, Giardina G, Contestabile R, Paiardini A, Cutruzzola F. SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. *Cell Death Dis.* 2014;5(11):1525–1525.
89. Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, Soh BS, Sun LL, Tai BC, Nga ME. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell.* 2012;148(1–2):259–72.
90. Tedeschi PM, Markert EK, Gounder M, Lin H, Dvorzhinski D, Dolfi S, Chan LL, Qiu J, DiPaola R, Hirshfield K. Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. *Cell Death Dis.* 2013;4(10):877–877.
91. DeNicola GM, Chen P-H, Mullarky E, Sudderth JA, Hu Z, Wu D, Tang H, Xie Y, Asara JM, Huffman KE. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat Genet.* 2015;47(12):1475.
92. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo H-K, Jang HG, Jha AK. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature.* 2011;476(7360):346–50.
93. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, Heffron G, Metallo CM, Muranen T, Sharfi H. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat Genet.* 2011;43(9):869–74.
94. Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 2002;30(1):47–9.
95. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell.* 1984;6:721–41.
96. Besag J, York J. Bayesian restoration of images. In: Matsunawa T, editor. *Analysis of Statistical Information* 1989. p. 491–507.

97. Gilks WR, Roberts GO, Sahu SK. Adaptive Markov chain Monte Carlo through regeneration. *J Am Stat Assoc*. 1998;93(443):1045–54.
98. Sahu SK, Zhigljavsky AA. Self-regenerative Markov chain Monte Carlo with adaptation. *Bernoulli*. 2003;9(3):395–422.
99. Liebermeister W, Uhlendorf J, Klipp E. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*. 2010;26(12):1528–34.
100. Voit EO. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge: Cambridge University Press; 2000.
101. Brooks S, Gelman A, Jones G, Meng X-L. *Handbook of Markov Chain Monte Carlo*. Boca Raton: CRC Press; 2011.
102. Kramer A, Stathopoulos V, Girolami M, Radde N. MCMC_CLIB—an advanced MCMC sampling package for ode models. *Bioinformatics*. 2014;30(20):2991–2.
103. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19(6):716–23.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

