Parameter Estimation Using Divide-and-Conquer Methods for Differential Equation Models

Seongho Kim[*]

Biostatistics Core, Karmanos Cancer Institute, Department of Oncology, School of Medicine, Wayne State University, USA

In systems biology, a key topic is the elucidation of the dynamic behavior of biological processes that are made up of complex biochemical networks. Statistical modeling is an important to capture the dynamics of biochemical networks such as metabolic networks, signal transduction pathways, and gene regulatory networks. These biochemical models have a set of parameters that represent the physical properties of the systems, such as kinetic constants and reaction rates. In general, the development of these models requires two steps: model structure construction and parameter estimation. The models are often constructed with time derivative expressions, such as ordinary differential equations (ODEs), to describe the change of certain quantities of interest over time [1,2]. The model parameters are then estimated by simulating the actual processes obtained from experimental analyses [3–5]. However, because the differential equation model has many uncertain parameters and limited measurement data, parameter estimation is a major bottleneck in the development of useful biochemical models [6,7].

Optimization algorithms cannot deal with the high dimensionality of search space due to calculation complexity. One way to circumvent this difficulty is to simplify complicated systems biology models using model order reduction methods. Model order reduction methods reduce the number of states and parameters of dynamical systems that are defined by ODEs [8]. Lumping is one model order reduction method in which the original states of the model are lumped or merged to a reduced number of pseudo-states, resulting in a fewer equations and parameters but with effectively the same or similar input-output behavior. Proper lumping is a special case of lumping where each of the original states contributes to only one of the pseudostates of the reduced system thereby forming groups that retain a clear physical interpretation. With these methods, the reduced systems include less information, but are supposed to retain the basic features or properties of the original models. Although computational expense is saved, it is highly likely that the simplification loses critical information, especially if there is excessive simplification. Another strategy is to use divide-and-conquer methods, which decompose a large network of interest into smaller sub-networks [9,10]. For example, Voit and Almeida [11] developed an approach to transforming the problem into several sets of decoupled algebraic equations, being processed efficiently in

[*]**Corresponding author:** Seongho Kim, Biostatistics Core, Karmanos Cancer Institute, Department of Oncology, School of Medicine, Wayne State University, Detroit, MI 48201, USA, Tel: +1-313-576-8653; kimse@karmanos.org.

parallel or sequentially, in large genetic network models. Kimura et al. [12] employed a cooperative co-evolutionary algorithm with a decomposition strategy to handle large S-system models with noisy time-series data. When there are no closed loops, Koh et al. [13] decomposed the network into small, independent sub-networks and estimated the parameters for each sub-network separately under the assumption that signals or mass flow in one direction. van Riel and Sontag [14] proposed a different approach to utilizing the modular structure of biochemical networks, providing the time courses of the intra-modular components that interact with neighboring modules. Those divide-and-conquer strategies, however, are not suitable for complex networks consisting of multiple closed or feedback loops, because dividing closed loops can change their intrinsic regulatory structures, greatly altering their dynamic features and the sensitivity of search parameters. Recently, to handle this difficulty, Maeda et al. [5] employed flux module decomposition that separates a complex, large-scale dynamic model into multiple flux modules without destroying its basic control architectures. However, it assumes that all parameters are necessary without accounting for differences in uncertainty of parameters.

To circumvent the aforementioned issues, we propose a divide-and-conquer approach to avoiding unnecessary information loss while estimating high-dimensional parameters efficiently. To do this, we first divide a large complete system into sub-systems so that each subsystem has a smaller, manageable number of differential equations. Then we estimate parameters for each sub-system, followed by refinement of the estimates through communication among subsystems. The success of the proposed algorithm depends on how the complete system is divided into small sub-systems.

We illustrate our proposed approaches with a simple three-compartment model. Its system of ordinary differential equations (ODEs) is as follows:

$$\frac{dx(t)}{dt} = -Ka \cdot x(t);$$

$$\frac{dy(t)}{dt} = Ka \cdot x(t) - Kb \cdot y(t) + Kb \cdot z(t);$$

$$\frac{dz(t)}{dt} = Kb \cdot y(t) - Kb \cdot z(t) - Kc \cdot z(t)$$

Where (Ka,Kb,Kc) are the parameters to estimate (i.e., Ka,Kb,Kc are the absorption rate, the distribution rate, the elimination rate constants, respectively); and $(x(t), y(t), z(t))_{t=0} = (0, 0, 0)$. Its graphical representation is shown in Figure 1a. Using this model, we investigated the performance of the proposed approach in a simulation study. We generated 100 simulations and estimated the parameters using 1) a conventional approach (ONE) and 2) a divide-and-conquer approach (DAQ). The brief schematic representation of DAQ can be seen in Figure 1b.

As for DAQ, the parameters (Ka, Kb) are first estimated given Kc and then the parameters (Kb, Kc) are estimated given Ka. This procedure was repeated until convergence. Table 1 displays the results of 100 simulation studies with mean squared errors (MSEs) and estimates' bias by three levels of measurement errors. The performance of DAQ is comparable to that of ONE, and, in some cases, the biases of DAQ are smaller than these of ONE in Table 1.

It is worth noting that, as the whole model is divided into smaller models, the computation expense decreases, but the information loss increases. For this reason, it is important to ensure that the decomposition is optimal, and future work will further need to find out the relationship between the decomposition and the information loss. Overall, as shown in the limited simulation study, the proposed approach preserves important properties of the original model and thereby increases the quality of the biochemical networks due to the property that the proposed approach does not depend on simplification. Furthermore, the proposed parameter estimation approach can be easily applied to other high-dimensional data such as genomics, transcriptomics, proteomics, and metabolomics. Therefore, the proposed work will benefit for many types of high-dimensional studies.

Acknowledgments

References

1. Abdullah A, Deris S, Anwar S, Arjunan SN. An evolutionary firefly algorithm for the estimation of nonlinear biological model parameters. PLoS One. 2013; 8:e56310. [PubMed: 23469172]

2. Lillacci G, Khammash M. Parameter estimation and model selection in computational biology. PLoS Comput Biol. 2010; 6:e100069.

3. Chou IC1, Voit EO. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci. 2009; 219:57–83. [PubMed: 19327372]

4. Tashkova K, Korosec P, Silc J, Todorovski L, Dzeroski S. Parameter estimation with bio-inspired meta-heuristic optimization: modelling the dynamics of endocytosis. BMC Syst Biol. 2011; 5:159–185. [PubMed: 21989196]

5. Maeda K, Minamida H, Yoshida K, Kurata H. Flus module decomposition for parameter estimation in a multiple-feedback loop model of biochemical networks. Bioprocess and Biosystems Engineering. 2013; 36:333–344. [PubMed: 22820677]

6. Kotte O, Heinemann M. A divide-and-conquer approach to analyze underdetermined biochemical models. Bioinformatics. 2009; 25:519–525. [PubMed: 19126574]

7. van Rie NA. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Brief Bioinform. 2006; 7:364–374. [PubMed: 17107967]

8. Okino MS, Mavrovouniotis ML. Simplification of Mathematical Models of Chemical Reaction Systems. Chem Rev. 1998; 98:391–408. [PubMed: 11848905]

9. Kim S, Kim S. A divide-and-conquer approach in applying EM for large recursive models with incomplete categorical data. Computational Statistics & Data Analysis. 2006; 50:611–641.

10. Kim S, Kim S. A note on collapsibility in recursive graphical models of contingency tables. Scandinavian Journal of Statistics. 2006; 33:575–590.

11. Voit EO, Almeida J. Decoupling dynamical systems for pathway identification from metabolic profiles. Bioinformatics. 2004; 20:1670–1681. [PubMed: 14988125]

12. Kimura S1, Ide K, Kashihara A, Kano M, Hatakeyama M, et al. Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. Bioinformatics. 2005; 21:1154–1163. [PubMed: 15514004]

13. Koh G, Teong HF, Clément MV, Hsu D, Thiagarajan PS. A decompositional approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk. Bioinformatics. 2006; 22:e271–e280. [PubMed: 16873482]

14. van Rie NA, Sontag ED. Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. Syst Biol (Stevenage). 2006; 153:263–274. [PubMed: 16986628]
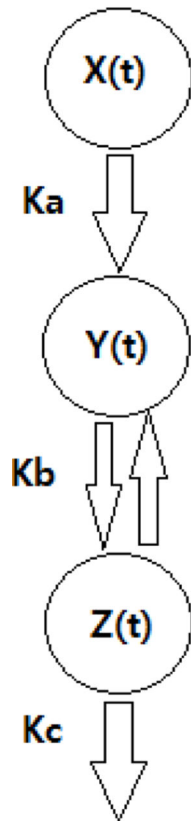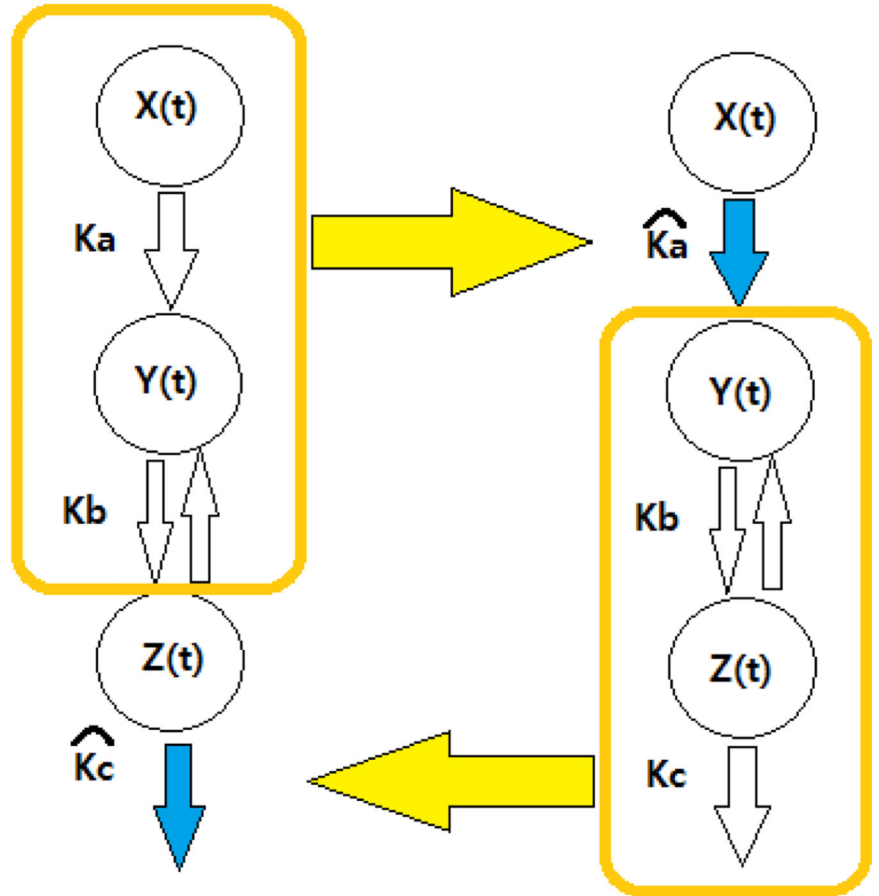
(a) One-step

(b) Divide-and-conquer

Figure 1.
Graphical representation of three compartmental models.

Table 1

Results of 100 simulations of ONE and DAQ.

	TRUE	σ=0.1 DAQ	σ=0.1 ONE	σ=0.5 DAQ	σ=0.5 ONE	σ=1.0 DAQ	σ=1.0 ONE
		Mean	Mean	Mean	Mean	Mean	Mean
		(SD)	(SD)	(SD)	(SD)	(SD)	(SD)
MSE		0.009	0.009	0.216	0.212	0.875	0.859
		−0.004	−0.004	−0.098	−0.096	−0.365	−0.358
		Bias	Bias	Bias	Bias	Bias	Bias
		(SD)	(SD)	(SD)	(SD)	(SD)	(SD)
Ka	log(0.8)	0.003	0.001	0.025	0.012	0.048	0.022
		−0.006	−0.007	−0.043	−0.045	−0.084	−0.09
Kb	log(10)	0.005	0.001	0.045	0.034	0.104	0.073
		−0.011	−0.021	−0.073	−0.067	−0.134	−0.218
Kc	log(8)	−0.002	−0.001	−0.021	−0.026	−0.038	−0.047
		−0.006	−0.014	−0.044	−0.044	−0.086	−0.123