**State Substitution: A novel method for parameter identification of large-scale bio-chemical kinetic models**

Paulius Rasiukas1, Chris O’Malley2, Mark J. Willis3

1 School of Engineering Newcastle University, p.rasiukas@ncl.ac.uk*;* 2 School of Engineering Newcastle University, chris.o'malley@ncl.ac.uk*;* 3 School of Engineering Newcastle University, mark.willis@ncl.ac.uk

**Highlights**

* + - A new method to identify unknown parameters in large scale dynamic models.
* State substitution using a non-parametric model allows decoupled equations to be solved independently.
* A computationally efficient and accurate parameter identification method when compared to alternative techniques.

**1.0 Introduction**

The ability to identify the parameters of models of complex large-scale (bio) chemical systems is critical in order to develop an understanding of a system as well as to use the model as a basis for process control or process optimization. Established methods of parameter identification normally simultaneously solve the entire set of nonlinear ordinary differential equations (ODEs) that describe the system to determine the model parameters. Generally, the ODEs are numerically integrated and the model parameters obtained using an optimisation algorithm that minimises the difference between the model predictions and the measured process data. However, this is known to be a non-trivial task, often requiring significant solution times and being prone to sub-optimal parameter identification. In this contribution we propose a new method for the identification of unknown parameters of large scale (bio)chemical kinetic models.

**2.0 Methods**

We decouple the set of ordinary differential equations (ODEs) describing kinetic phenomena into separate sub-systems, whose unknown model parameters can be identified separately. Each ODE is considered a separate sub-system and state substitution is used to replace any coupled model components, i.e. the additional states that would normally be obtained as a result of the numerical integration of the entire system model. The state substitution method we employ relies on the use of measured state variables that are regressed against time using a suitable non-parametric model, e.g. polynomial, spline, neural network. The nonparametric model is then used to replace the states of any of the coupled components of each ODE. This allows numerical integration of each ODE separately and the use of traditional parameter identification techniques to identify any unknown parameters of each sub-model. In a final stage of our method, we then integrate the entire set of ODEs using a global optimisation algorithm using the numerical values of the model parameters obtained from the optimisation of each sub-system as the initial guess. This not only reduces computational time significantly compared to current simultaneous solution methods it ensures statistical optimality of the final parameter values.

To demonstrate the method, it is applied to a biochemical process model described in [1], and compared with a Latin hypercube induced multi-start method, which has been reported to be an efficient method for model calibration in systems biology models [2,3]. The model is shown below where it may be observed that there are 15 unknown kinetic parameters, $p1,…,p16$, and each state $x1,…,x6$, is assumed measured,

|  |  |
| --- | --- |
| $$\frac{d}{dt}\left[\begin{matrix}x1\\x2\\x3\\x4\\x5\\x6\end{matrix}\right] = \left[\begin{matrix}1&-1&0&0\\-1/p8&0&-1&0\\p9/p8&0&p9&0\\-1/p12&0&0&0\\p13/p12&0&0&0\\0&0&0&1\end{matrix}\right]\left[\begin{matrix}f1\\f2\\f3\\f4\end{matrix}\right]$$ | $$f\_{i} = u\_{i}x1 , i=1,4$$$$u\_{1} = p1\frac{x2}{(p2+x2)}\frac{x4}{(p3+x4)}$$$$u\_{2} = p4\frac{1}{(p1-p5x3)}\frac{1}{(p1-p6x5)}\frac{p7}{\left(p7+x4\right)}$$$$u\_{3} = p10\frac{x2}{(p11+x2)}$$$$u\_{4} = p14 +p15\frac{u1}{p16 +u1}$$ |

In the results that follow, the non-parametric model used for state substitution was a cubic spline and each decoupled ODE was integrated using the MATLAB function *ode45*. The optimisation algorithm used to determine the unknown model parameters was the MATLAB function *lsqnonlin* via repeated integration each decoupled sub-component of the overall kinetic model. Simultaneous optimisation of the full kinetic model used MATLAB’s genetic algorithm routine.

**3.0 Results and discussion**

A significant number of comparative studies have been made via adjustment of the number of data measurements assumed available and the level of measurement error (measurement noise). Table 1 indicates a typical scenario comparing the estimated parameter values to the actual (known) model parameters.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | $$p1$$ | $$p2$$ | $$p3$$ | $$p4$$ | $$p5$$ | $$p6$$ | $$p7$$ | $$p8$$ | $$p9$$ | $$p10$$ | $$p11$$ | $$p12$$ | $$p13$$ | $$p14$$ | $$p15$$ | $$p16$$ |
| (1) | 1.09 | 1.1 | 0.3 | 0.09 | 0.01 | 0.06 | 0.02 | 0.11 | 1.8 | 1.7 | 19 | 0.38 | 0.85 | 3.5 | 25.7 | 0.02 |
| (2) | 1.09 | 1.06 | 0.24 | 0.05 | 0.03 | 0 | 0.40 | 0.12 | 1.82 | 4.07 | 53.9 | 0.41 | 0.86 | 2.92 | 26.80 | 0.02 |
| (3) | 1.12 | 0.91 | 0.27 | 0.11 | 0.01 | 0 | 0.50 | 0.12 | 1.83 | 0.79 | 1.94 | 0.405 | 0.82 | 2.87 | 25.73 | 0.01 |
| (4) | 4.31 | 0.01 | 7.93 | 4.05 | 0.10 | 0.80 | 286.8 | 0.34 | 1.85 | 18.5 | 60.1 | 0.68 | 0.49 | 3.86 | 19.45 | 0.01 |

**Table 1.** Parameter estimates obtained assuming, $x1,…,x6$, are measured every 0.3 h (total number of measurements of each state being 12) with a measurement error of 10% (assumed to be white noise). In the table, (1) Actual values, (2) State – substitution, (3) Multi-start with Latin Hypercube experimental design (4) Simultaneous solution and optimization of the full model.

It may be observed that the parameter estimates obtained using the state-substitution method (2) and the use of a multi-start strategy with Latin Hypercube experimental design to obtain the initial parameter values of each run of the algorithm (3) are of comparable accuracy and are generally significantly superior to that of the simultaneous solution (4). While algorithm run-time will be machine dependent, as an indication of the typical performance of the methods; the state-substitution method (2) took 2371s to find a solution, whereas the multi-start method (3) took 5778s to find a solution while the simultaneous solution (4) ran for 255h and was still not able to converge to reasonable solution.

**4.0 Conclusions**

A new method for the identification of unknown parameters of large scale (bio)chemical kinetic models was been proposed. For the example given, this provided a fast and reliable parameter estimation procedure. The full paper will present a more detailed assessment of the method using additional bench-mark models [4]. Furthermore, we will use Monte Carlo simulation to provide a statistical perspective on our run-time results and parameter estimates and assess the impact that unmeasured species states have on the identification procedure.

**References**

[1] Saraiva, I., VandeWouwer, A. and Hantson, A.L. (2015) 'Parameter identification of a dynamic model of CHO cell cultures: an experimental case study', *Bioprocess BiosystEng*, 38(11), pp. 2231-48.

[2] Raue, A., Schilling, M., Bachmann, J., A.L. (2013) 'Lessons learned from quantitative dynamical modeling in systems biology', *PLoS One*, 8(9), p. e74335.

[3] Degasperi, A., Fey, D. and Kholodenko, B.N. (2017) 'Performance of objective functions and optimisation procedures for parameter estimation in system biology models', *NPJ Syst Biol Appl*, 3, p. 20.

[4] Hass, H., Loos, C., Alvarez, E.R., A.L. (2019) ‘Benchmark problems for dynamic modeling of Intracellular processes’ Bioinformatics.