Journal of Theoretical Biology ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

On linear models and parameter identifiability in experimental biological systems

Q1 Timothy O. Lamberton [a], Nicholas D. Condon [b], Jennifer L. Stow [b], Nicholas A. Hamilton [a,b,*]

[a] Division of Genomics & Computational Biology, Institute for Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia
[b] Division of Molecular Cell Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

HIGHLIGHTS

- Methods are developed to design optimal experiments to ensure model parameter identifiability.
- We give sufficient conditions to identify parameters in linear, diagonalizable ODE models of experiments.
- A framework for proving identifiability in linked models is demonstrated.
- Applications to experimentally common situations and measurement protocols are shown.

ARTICLE INFO

Article history:
Received 10 September 2013
Received in revised form
19 May 2014
Accepted 20 May 2014

Keywords:
Modeling
Experimental design
Parameter identifiability
Ordinary differential equations

ABSTRACT

A key problem in the biological sciences is to be able to reliably estimate model parameters from experimental data. This is the well-known problem of parameter identifiability. Here, methods are developed for biologists and other modelers to design optimal experiments to ensure parameter identifiability at a structural level. The main results of the paper are to provide a general methodology for extracting parameters of linear models from an experimentally measured scalar function – the transfer function – and a framework for the identifiability analysis of complex model structures using linked models. Linked models are composed by letting the output of one model become the input to another model which is then experimentally measured. The linked model framework is shown to be applicable to designing experiments to identify the measured sub-model and recover the input from the unmeasured sub-model, even in cases that the unmeasured sub-model is not identifiable. Applications for a set of common model features are demonstrated, and the results combined in an example application to a real-world experimental system. These applications emphasize the insight into answering "where to measure" and "which experimental scheme" questions provided by both the parameter extraction methodology and the linked model framework. The aim is to demonstrate the tools' usefulness in guiding experimental design to maximize parameter information obtained, based on the model structure.

1. Introduction and background

Q5 Q4 Linear systems of ordinary differential equations (ODEs) are a widely applied tool used for modeling of biological systems to generate hypotheses for experimental testing. Linear models are used in fields such as molecular cell biology to investigate endocytic traffic and sorting (Ghosh et al., 1994; Sheff et al., 2002; Henry and Sheff, 2008) and phosphoinositide transformation in endocytosis (Belward et al., 2011), and systems biology to study physiology (Cobelli and DiStefano, 1980), gene expression (Crampin, 2006) and regulatory networks (de Jong, 2002). The rapid advance of technologies, such as fluorescent microscope imaging technologies which allow imaging of the behavior of multiple proteins at multiple locations in live cells to be quantified and analyzed, have given access to large amounts of data which can be used for robust model development and validation (Hamilton, 2009), and has already lead to breakthroughs in understanding major diseases such as Parkinson's Disease (Follett et al., 2014). The rise in the use of mathematical modeling in biological fields brings with it pitfalls which are to be avoided. Often such models include large numbers of parameters which are impractical to measure individually, and must be inferred

Q3 * Corresponding author at: Division of Genomics & Computational Biology, Institute for Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia. Tel.: +61 7 3346 2033; fax: +61 7 3346 2101.
E-mail address: n.hamilton@uq.edu.au (N.A. Hamilton).

http://dx.doi.org/10.1016/j.jtbi.2014.05.028
0022-5193/

from their effect on the system, using data sets which can be incomplete and noisy (Ghosh et al., 1994; Sheff et al., 2002; Voit et al., 2006; Henry and Sheff, 2008; Berthoumieux et al., 2011). The problem of parameter estimation for linear and non-linear ODE-based models has been extensively studied from a variety of perspectives, for example, applied mathematics (Liu et al., 2011) and systems biology (Jaqaman and Danuser, 2006; Ashyraliyev et al., 2009). Complicating the process of parameter estimation is the idea of parameter **identifiability**, that is, whether different combinations of parameter values lead to indistinguishable model output, both in terms of inherent model structure (**structural parameter identifiability**) and due to experimental noise (**practical parameter identifiability**) (Bellman and Åstrom, 1970; Cobelli and DiStefano, 1980; Holmberg, 1982; Chou and Voit, 2009; Nikerel et al., 2009; Chen et al., 2010; Berthoumieux et al., 2012).

Approaches to dealing with identifiability have been developed for a variety of modeling situations. The structural identifiability problem (defined precisely below) is more fundamental than the practical identifiability problem, in that it represents a best-case scenario for any practical identifiability analysis, and also more analytically tractable, with well-defined approaches for linear (Bellman and Åstrom, 1970; Cobelli and DiStefano, 1980) and many non-linear models (Pohjanpalo, 1978; Jaqaman and Danuser, 2006; Bellu et al., 2007; Nemcova, 2010; Chis et al., 2011). The problem of practical identifiability, which considers whether experimental noise will allow parameters that may be structurally identifiable to be resolved to a level of certainty, is generally approached by developing confidence intervals for estimated parameters and considering the parameter space of an objective function near solutions, either using quadratic approximations or numerical methods (Cobelli and DiStefano, 1980; Raue et al., 2009, Berthoumieux et al., 2012). Bayesian frameworks have also be used to represent parameter uncertainty and model sensitivity to parameter values (Liepe et al., 2013). Practical identifiability of biological systems is a difficult problem, as illustrated by a study (Gutenkunst et al., 2007) where it was found that biological systems show degrees of sensitivity distributed fairly evenly over multiple orders of magnitude to changes in different independent combinations of parameters, with highly sensitive parameters being difficult to constrain with even large amounts of data.

There are intuitive gaps in the underlying causes of both structural and practical identifiability. Although practical identifiability is fundamentally a problem of experimental noise, understanding model sensitivity to parameter variation in most cases involves performing complex analytical and often numerical calculations. In the case of structural identifiability, many of the analytic approaches allow the straight-forward computation of the binary identifiability/non-identifiability of model parameters on a case-by-case basis; however it is often the case that these approaches miss a broad intuition of what underlying mechanics of models allows or does not allow the determination of the parameters.

A related issue to parameter identifiability is experimental design optimization, that is, how to maximize information obtained from measurements. An effort has been made (Liepe et al., 2013) to address this problem using a Bayseian framework as a general computational method for optimizing experiments. An analytical approach to both parameter identifiability and experimental design problems could provide several advantages over computational methods, by providing fine-grained feedback on how parameters affect outputs under a wide range of conditions, and lead to intuitive rule-of-thumb guidance for experimental design.

Here, theorems applicable to broad classes of biological structural parameter identifiability problems are proved. In Section 2.1, new methods are developed which consider the underlying mechanics of models and how they are captured by experimental design, allowing the development of measurement schemes to efficiently extract parameter information. These methods are demonstrated with applications to a range of commonly used models (see figures) to design experimental schemes to optimize the identification of model parameters. In Section 2.2 a framework for constructing complex models by considering the outputs of components of simple model classes as inputs to other models to create linked models is outlined. By considering the problem of identifying the combined model, it is shown that under certain conditions the component models can effectively be distinguished, and structural identifiability results for the individual models when not linked also hold for the corresponding components of the combined model. It is also shown that even when components of the models are "hidden" from measurement much can often be inferred about the hidden parts from the measurable components. This linked model framework is also applied to a range of commonly used model configurations, which are in turn used to analyze a realistic biological system.

The specific focus of the paper is on the broad family of mathematical models, **linear systems of ordinary differential equations (ODEs)**. Cobelli and DiStefano (1980) and Raue et al. (2011) give good overviews of modeling using linear systems of ODEs in the context of physiological systems and gene signaling networks respectively, fields where linear models have been widely and successfully applied. Following a similar framework, here we will consider models that have the following linear general form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) \tag{1}$$

for a model state vector $\mathbf{x}$ with n components. The initial state of the model is specified by the length n vector $\mathbf{x}(0)=\mathbf{b}$ and the model structure by the constant $n \times n$ matrix $\mathbf{A}$. The vector of m model observables is given by

$$\mathbf{y}(t_i) = \mathbf{C}\mathbf{x}(t_i) + \boldsymbol{\varepsilon}_i \tag{2}$$

where $\boldsymbol{\varepsilon}_i$ is a length m vector of measurement noise assumed to be normally distributed. $\mathbf{C}$ is an $m \times n$ non-negative matrix with each row defining the combinations of model locations measured in a corresponding experimental data set.

The model structure $\mathbf{A}$ and the initial state $\mathbf{b}$ are functions on the parameters $\boldsymbol{\theta}$ of the model. $\mathbf{A}$ is dependent on a set of constant parameters $\mathbf{k}$ (**rate constants**) that denote the rates at which the model components interact and together with an independent initial state $\mathbf{b}$ make up the parameters $\boldsymbol{\theta}$. For example, in cell protein trafficking models (Ghosh et al., 1994; Sheff et al., 2002; Henry and Sheff, 2008) the model parameters consist of **rate constants** at which a protein of interest is trafficked between **compartments** (such as endosomes) in the cell, as well as initial concentrations of the protein in the compartments. In enzyme reaction models such as those used by Raue et al. (2011), the model parameters are the rates at which the presence of enzymes inhibit or accelerate the production of other enzymes, as well as initial concentrations. The model states represent the concentration of protein in the intracellular compartments, or the enzyme concentrations in a solution, for example within a cell cytoplasm, or an organism's blood stream. An often used experimental set-up (Bellman and Åstrom 1970; Cobelli and DiStefano, 1980; Henry and Sheff, 2008) is a **pulse** or **tracer** input at a single location, which is then measured over time at other locations. In terms of Eqs. (1) and (2) this corresponds to an initial state $\mathbf{b}=b_i\mathbf{e}_i^{\mathrm{T}}$ for the pulse location i, where $\mathbf{e}_i$ is a standard basis vector (1 in the ith component and zeros elsewhere), and b_i denotes the size of the input. Another example set-up is a **fill/release** method, where a

Please cite this article as: Lamberton, T.O., et al., On linear models and parameter identifiability in experimental biological systems. J. Theor. Biol. (2014), http://dx.doi.org/10.1016/j.jtbi.2014.05.028