Computer Methods and Programs in Biomedicine 143 (2017) 129–135

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

DISTING: A web application for fast algorithmic computation of alternative indistinguishable linear compartmental models

Natalie R. Davidson[a,1], Keith R. Godfrey[b,*], Faisal Alquaddoomi[a], David Nola[a], Joseph J. DiStefano III[a]

[a] Biocybernetics Laboratory, Computer Science Department, University of California Los Angeles, CA 90095, U.S.A.
[b] School of Engineering, University of Warwick, Coventry, CV4 7AL, U.K.

ARTICLE INFO

Article history:
Revised 20 February 2017
Accepted 27 February 2017

Keywords:
Compartmental modeling
Model selection
Parallel computing
Structural identifiability
Symbolic algebra

ABSTRACT

Background and objectives: We describe and illustrate use of DISTING, a novel web application for computing alternative structurally identifiable linear compartmental models that are input-output indistinguishable from a postulated linear compartmental model. Several computer packages are available for analysing the structural identifiability of such models, but DISTING is the first to be made available for assessing indistinguishability.

Methods: The computational algorithms embedded in DISTING are based on advanced versions of established geometric and algebraic properties of linear compartmental models, embedded in a user-friendly graphic model user interface. Novel computational tools greatly speed up the overall procedure. These include algorithms for Jacobian matrix reduction, submatrix rank reduction, and parallelization of candidate rank computations in symbolic matrix analysis.

Results: The application of DISTING to three postulated models with respectively two, three and four compartments is given. The 2-compartment example is used to illustrate the indistinguishability problem; the original (unidentifiable) model is found to have two structurally identifiable models that are indistinguishable from it. The 3-compartment example has three structurally identifiable indistinguishable models. It is found from DISTING that the four-compartment example has five structurally identifiable models indistinguishable from the original postulated model. This example shows that care is needed when dealing with models that have two or more compartments which are neither perturbed nor observed, because the numbering of these compartments may be arbitrary.

Conclusions: DISTING is universally and freely available via the Internet. It is easy to use and circumvents tedious and complicated algebraic analysis previously done by hand.

1. Introduction

A structured mathematical model based on limited experimental data is often used to predict model or system attributes that cannot be measured directly. This often engenders two fundamental analysis problems: model structural identifiability (SI) and model structural indistinguishability. The word 'structural' in both of these indicates that perfect input-output data are assumed; neither problem is concerned with parameter estimation accuracy using real data. SI analysis addresses whether particular model *parameters* or combinations of them are quantifiable in principle from the input-output I-O data. Many techniques are available for analysing the SI of both linear and nonlinear models, and several computer packages are available for facilitating the analysis. Examples of packages that test for *global* (i.e. unique) structural identifiability are given in [1–3].

By contrast, the less familiar model indistinguishability problem addresses whether there are *other models* within the same class (e.g., the class of compartmental models), but *with different structures* that may equally well fit the input-output (I-O) data, again assumed perfect. Techniques for analysing structural indistinguishability among Linear Compartmental (LC) models were first established more than 25 years ago, for LC models in general [4] and more specifically for pharmacokinetic models [5]. For LC models, indistinguishability analysis involves constructing every other LC model that can give rise to identical input-output behaviour through a suitable choice of its unknown parameters.

* Corresponding author.
E-mail address: k.godfrey@warwick.ac.uk (K.R. Godfrey).
[1] Current address: Tri-Institutional Program in Computational Biology and Medicine, Cornell University, Ithaca, New York, Weill Cornell Medical College, New York, New York, Memorial Sloan-Kettering Cancer Center, New York, New York, U.S.A.

http://dx.doi.org/10.1016/j.cmpb.2017.02.025
0169-2607/

Techniques available in 1990 for analysing structural identifiability and indistinguishability of LC models were compared and contrasted in [6]. Techniques for analysing indistinguishability between *nonlinear* compartmental models have been established [7,8], but by the very nature of the problem, the analysis has to be much more restrictive, confined to whether a single candidate nonlinear (NL) model is indistinguishable from a postulated structured NL model.

Establishing model uniqueness – or *model disambiguation* – is a major multifaceted problem in quantitative science, with methods and solutions available only for the very limited classes of models described above. The indistinguishablity problem is often ignored or unknown by modelers, which could lead to incorrect models and false predictions. Algebraically, the problem is computationally complex, involving structured symbolic (non-numerical) analysis of typically large matrices (as we describe in Section 2). Until now, no software tools have been made available to address this problem.

The present paper describes and exemplifies the use of DISTING, a freely available web application (app) for structural indistinguishability analysis of LC models. The DISTING app was partially motivated by substantially increased interest in model selection methods within the systems biology and systems pharmacology communities. Recent developments in computational matrix analysis and parallel computing algorithms – for facilitating and speeding up the symbolic computations needed for indistinguishability analysis – and more user-friendly web interface design also motivated its creation.

The computational algorithms embedded in DISTING are based on the geometric and algebraic properties of LC models described in [4], embedded in a graphical user interface. They are advanced versions of the algorithms proposed in 1991 by Zhang and coworkers [9] for implementing these properties. No software based on the work in [9] was offered or made available at the time and, as far as we are aware, none has been made available since.

The remainder of the paper is organised as follows. The techniques used in DISTING are described in Section 2 and three examples of its use are given in Section 3. Conclusions are drawn in Section 4.

2. Algorithms used in DISTING

2.1. Overall summary

Linear compartmental models can be described in the standard state space formulation involving ordinary differential equations (ODEs) and algebraic output equations:

$$\dot{\mathbf{x}} = K\mathbf{x} + B\mathbf{u}, \quad \mathbf{y} = C\mathbf{x} \tag{1}$$

with corresponding input-output Laplace transfer function matrix:

$$H(s, \boldsymbol{p}) = C(\boldsymbol{p})[sI - K(\boldsymbol{p})]^{-1}B(\boldsymbol{p}) \equiv \left[H_{ij}(s, \boldsymbol{p})\right] = \begin{pmatrix} H_{11} & H_{12} & \dots & H_{1j} \\ H_{21} & H_{22} & \dots & H_{2j} \\ \dots & \dots & \dots & \dots \\ H_{i1} & H_{i2} & \dots & H_{ij} \end{pmatrix} \tag{2}$$

The parameter vector of rate constants k_{ij} of compartmental-structure system matrix K, b_{ij} of input structure matrix B and c_{ij} of output structure matrix C is denoted $\boldsymbol{p}$. The transfer function for the single experiment with input at node j and output at node i is given by:

$$H_{ij}(s, \boldsymbol{p}) = \frac{\beta_n^{ij} s^{n-1} + \beta_{n-1}^{ij} s^{n-2} + \beta_{n-2}^{ij} s^{n-3} + \dots + \beta_2^{ij} s + \beta_1^{ij}}{s^n + \alpha_n^{ij} s^{n-1} + \dots + \alpha_2^{ij} s + \alpha_1^{ij}} \tag{3}$$

The α's and β's are algebraic functions of the components of k_{ij}, b_{ij} and c_{ij} and are called *structural invariants* or *moment invariants*.

The postulated linear compartmental model can be formally realized algorithmically by matrices K, B and C. Zhang and coworkers [9] described a method to distinguish this model from all alternative models with the same input-output transfer function, i.e. all I-O indistinguishable models. The procedure in [9] begins by generating all possible candidate models. Infeasible candidates are then removed from consideration during various stages of analysis, using geometric and algebraic properties previously developed in [4] that must apply to all indistinguishable models.

The geometric rules check graphical connectivity within each model. Each must be the same in the postulated and candidate models for possible indistinguishability. If any of the rules is not obeyed for a candidate model, it is removed from further consideration. The remaining candidate models are possibly, but not guaranteed to be, structurally indistinguishable from the postulated model. The subsequent algebraic analysis is much more time and resource intensive, so the geometric analysis is done first – to eliminate as many models as possible. It is very effective in removing candidate models, thus reducing the effort spent on computationally-intensive tasks.

DISTING implements the methods outlined in [4] and [9] algorithmically, augmented with several computational tools that greatly speed up the overall procedure. These include a novel simplification approach and parallel processing algorithms for matrix analysis. The geometric rules are described in Section 2.2, while the algebraic analysis and the computational refinements used therein are described in Section 2.3.

2.2. Geometric rules

The following geometric rules [4], deduced only from input-output knowledge about a postulated model [10], apply to all models that are candidates for indistinguishability:

Rule 1: The length of the shortest path from a perturbed compartment to an observed compartment must remain the same.
Rule 2: The number of compartments with a path to a given observed compartment must remain the same.
Rule 3: The number of compartments that can be reached from a perturbed compartment must remain the same.
Rule 4: The number of traps must remain the same [11]. A trap is defined as a strongly connected set of compartments with no flows out of this set to other compartments or to the environment. A strongly connected set of compartments is one in which there exists a path from any compartment of the set to any other.

These rules are computationally inexpensive to verify, and as a consequence, the focus has not been placed in DISTING on optimising this part of the overall procedure.

2.3. Algebraic analysis

The algebraic analysis is done in two stages. First, the transfer functions H_{ij} between each input and each output of the remaining candidate models is computed to determine whether they are of the same format as those of the postulated model (Eq. (3)). Those that are not are eliminated. Second, the Jacobian matrix of the coefficients of the transfer function with respect to each of the parameters p of the remaining candidate models is computed, and any that have a different rank from the postulated model are eliminated.

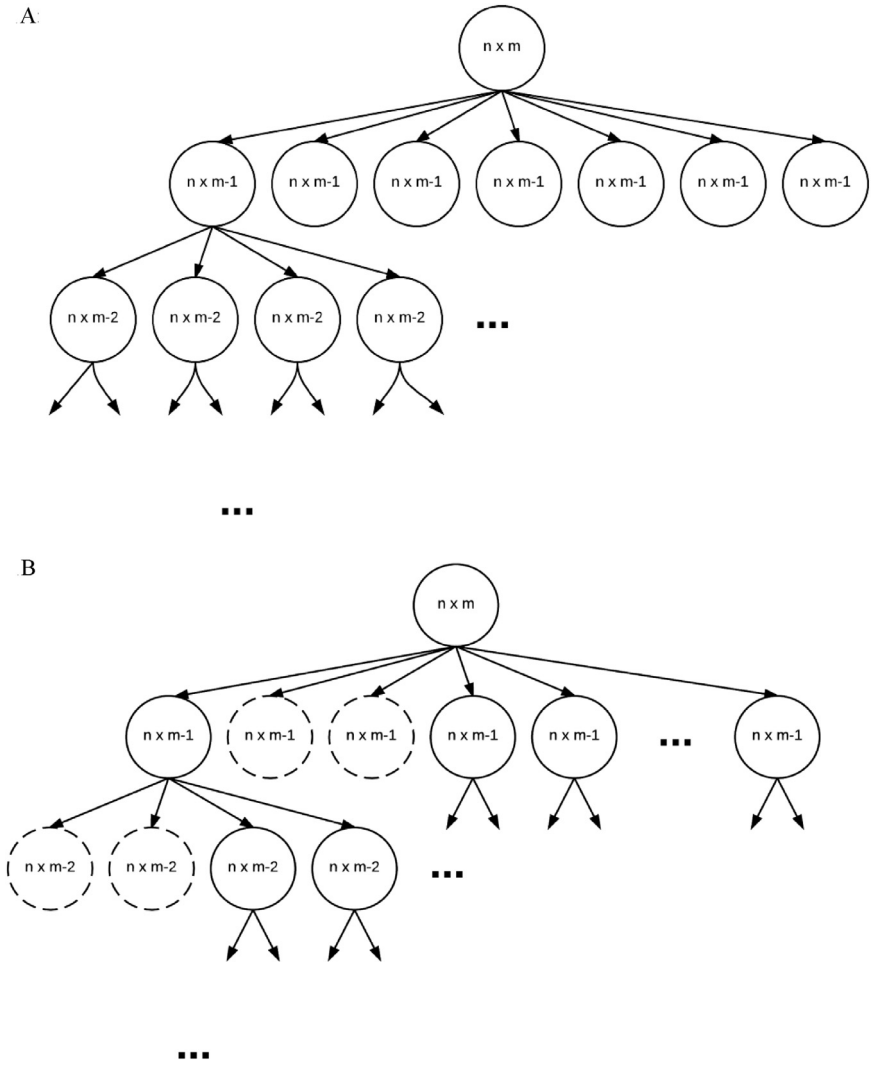

Fig. 1. (A). Tree of all sub-matrices of the full Jacobian matrix (B). Tree of all reduced sub-matrices of *J*.

The Jacobian is given by

$$J(\boldsymbol{p}) = \begin{pmatrix} \frac{\delta\beta_1^{11}}{\delta p_1} & \frac{\delta\beta_1^{11}}{\delta p_2} & \cdots & \frac{\delta\beta_1^{11}}{\delta p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta\alpha_1^{11}}{\delta p_1} & \frac{\delta\alpha_1^{11}}{\delta p_2} & \cdots & \frac{\delta\alpha_1^{11}}{\delta p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta\beta_1^{ij}}{\delta p_1} & \frac{\delta\beta_1^{ij}}{\delta p_2} & \cdots & \frac{\delta\beta_1^{ij}}{\delta p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta\alpha_1^{ij}}{\delta p_1} & \frac{\delta\alpha_1^{ij}}{\delta p_2} & \cdots & \frac{\delta\alpha_1^{ij}}{\delta p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta\beta_1^{rm}}{\delta p_1} & \frac{\delta\beta_1^{rm}}{\delta p_2} & \cdots & \frac{\delta\beta_1^{rm}}{\delta p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta\alpha_1^{rm}}{\delta p_1} & \frac{\delta\alpha_1^{rm}}{\delta p_2} & \cdots & \frac{\delta\alpha_1^{rm}}{\delta p_n} \end{pmatrix} \quad (4)$$

Rank comparisons of all candidate Jacobian matrices are the most computationally intensive tasks in the whole procedure, due to the number of sub-matrices for each candidate model. To accomplish these computations, DISTING implements two novel approaches for matrix simplification and rank determination, as described below. Both were mentioned as possibilities but not discussed further in [9].

2.3.1. Jacobian matrix reduction

All transfer function components in Eq. (2) have the same denominator, i.e. the characteristic equation of the model is unique. These generate linearly dependent rows of the Jacobian matrix *J*, one for each candidate model. All but one of the rows associated with these α coefficients is removed before computing the rank of *J*. This not only reduces the dimension of the full Jacobian matrix, but it also means that considerably fewer sub-matrix rank calculations are needed. This is particularly valuable if there are several input-output pairings.

2.3.2. Sub-matrix rank reduction

A full rank matrix has only full rank sub-matrices and, for the algorithm, this means that guaranteed full rank sub-matrices can be used systematically to select/delete models in advance, so reducing the number of rank calculations needed. Sub-matrix rank calculations are done recursively and stored in a tree structure. The root node is the original matrix and its children are the sub-matrices with one fewer row, as shown in Fig. 1A.

Using the tree structure makes comparing sub-matrices computationally easier to generate and analyse. The data structure used to keep track of retained and eliminated sub-matrices are sets, ensuring that none are duplicated. The tree is traversed and

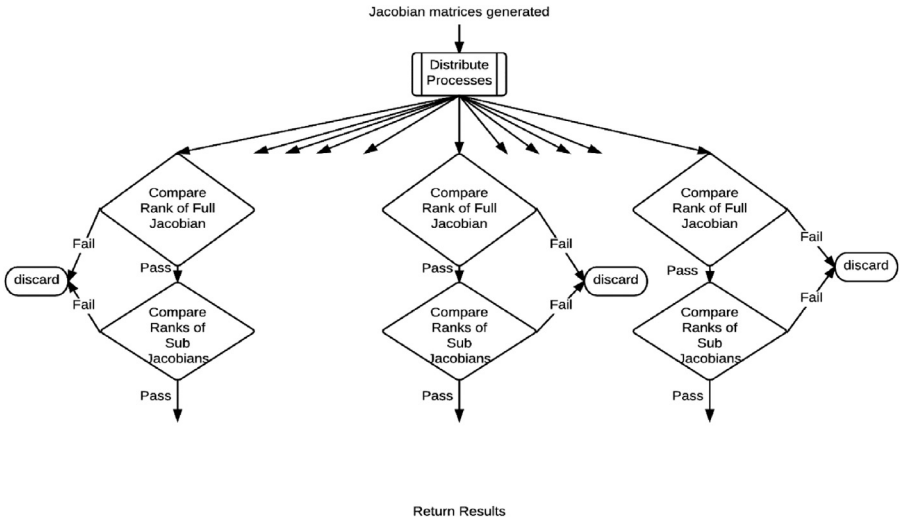

Fig. 2. Flow chart of distributed rank calculations.

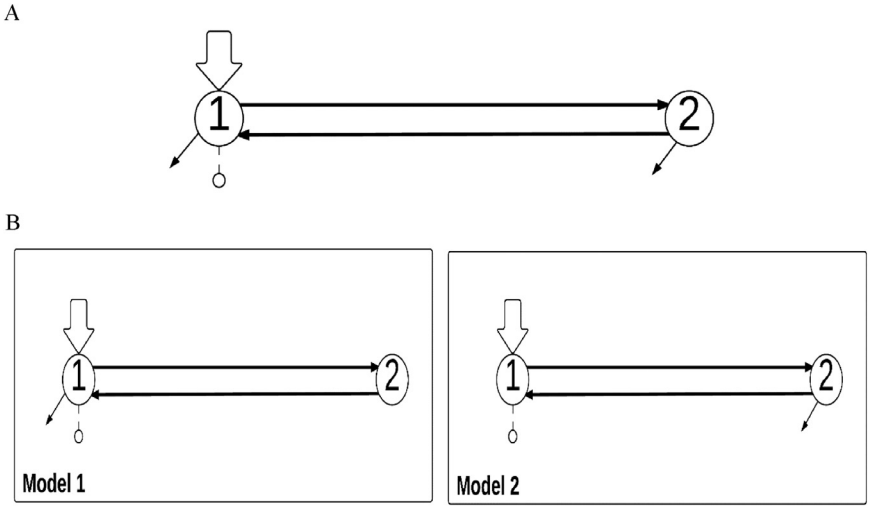

Fig. 3. (A). Postulated 2-compartment model (Model 0). (B) Indistinguishable 2-compartment models.

sub-matrix rank is calculated at each node; if a parent node has full rank, then the rank calculation of children nodes is omitted.

An example showing reduced rank calculations is shown in Fig. 1B. The solid line nodes are not full rank matrices, while the dashed-line nodes are. The dashed-line nodes have no children because their sub-matrices also have full rank, so no redundant rank calculation is done. The matrices and their ranks are stored in a lookup table, so that no matrix with the same value in different candidates is reused.

2.3.3. Parallelization of candidate rank computations

The Jacobian matrix and sub-matrix reduction algorithms eliminate many of the possible candidate models, but for those remaining, the rank computations are intensive. Computation time is minimized in DISTING using parallel computations, implemented with Python Library software *CELERY* [12]. A flow chart of the distributed computations is shown in Fig. 2.

3. Examples

3.1. Two-compartment example

To familiarise readers with the problem being tackled, the first example deals with finding models that are indistinguishable from a postulated fully-parameterized two-compartment model with input into and observation of Compartment 1 only, exchange in both directions between the compartments and elimination to the environment from both compartments (4 rate constants). The postulated model structure (Model 0) shown in Fig. 3A is entered into DISTING. The transfer function matrix of this model has three moment invariants and, because there are four unknown rate constants, the postulated model is structurally unidentifiable from this I-O experiment – see, for example, [6,13,14].

DISTING produces Models 1 and 2 shown in Fig. 3B as indistinguishable from the postulated model. Both alternative models have only one elimination to the environment. DISTING only produces structurally identifiable models, not necessarily globally (uniquely) SI. In this case, separate analysis shows that both are both globally identifiable [6,13,14].

3.2. Three-compartment example

The second example is drawn from [15], with postulated model structure (Model 0) shown in Fig. 4A; note that there is one trap, Compartment 3.

DISTING gives alternative Models 1, 2 and 3 in Fig. 4B as indistinguishable from the postulated model. All three have one trap. Model 1 has Compartments 2 and 3 forming a trap; Model 2

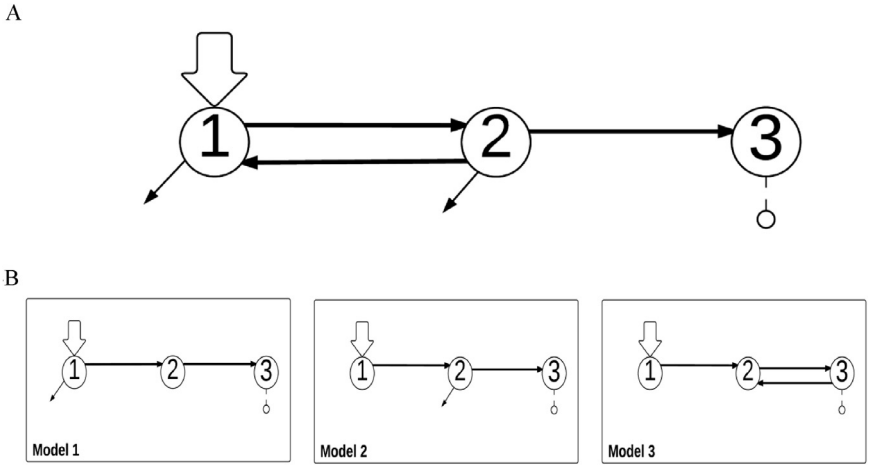

Fig. 4. (A) Postulated 3-compartment model (Model 0). (B) Indistinguishable 3-compartment models.

has Compartment 3 as a trap; and Model 3 has Compartments 1, 2 and 3 forming a trap. Analysis in [6] shows that the postulated model in Fig. 4A has three non-zero moment invariants and five rate constants, so it is structurally unidentifiable. In contrast, the three indistinguishable models in Fig. 4B – each with two fewer parameters – are all locally structurally identifiable, with two sets of solutions for the three rate constants [6]. Again, we note that DISTING only produces structurally identifiable (SI) models, not necessarily globally (uniquely) SI.

This indistinguishability analysis is, however, incomplete, because it is based on purely symbolic algebra manipulations of the structural model, with no regard to numerical values of its rate constants. For this example, there is another LC model – also with 3 rate constants – that *may be* indistinguishable from the postulated model, depending on the values of the 3 rate constants [6]. This alternative model has a loop, with rate constants from Compartment 1 to 2, 2 to 3 and 3 back to 1 – all with no elimination to the environment (so that all three compartments form a trap). When any LC model has three or more compartments in a loop, it is possible for there to be complex conjugate eigenvalues in model solutions – see [16], Section 4.1.6. This alternative model can have either 3 real eigenvalues or 1 real eigenvalue and a complex conjugate pair of eigenvalues, depending on the rate constant values [6]. On the other hand, if there is a complex conjugate pair of eigenvalues, it cannot be indistinguishable from Models 1, 2 and 3 of Fig. 4B, all of which can only have real eigenvalues – like the postulated model – for any (positive) values of the rate constants. Candidate models with complex eigenvalues in their solutions cannot be fully analysed with DISTING.

3.3. Four-compartment example

A simple drug-metabolite-metabolism exchange model with 7 rate constants, drawn from [4], is shown in Fig. 5A. As with the previous examples, this is called Model 0. Compartments 1 and 2 represent drug exchange between two organs (e.g. blood and liver) and Compartments 3 and 4 metabolite exchange between the same organs, with unidirectional biotransformation of drug to metabolite only in the peripheral organ (e.g. liver metabolism). Input of drug is to Compartment 1, while Compartments 1 and 3 are observed (measurement of drug concentration in Compartment 1 and metabolite concentration in Compartment 3). In the postulated model, elimination to the environment is from all four compartments. An example of the use of this model has been to describe the dynamic behaviour of bromosulfothalein [17,18].

DISTING provides the ten alternative models shown in Fig. 5B as indistinguishable from the postulated model, all with six rate constants. However, some care is needed in interpreting this result, because a subtle symmetry exists in the original and the candidate models generated: neither of the peripheral compartments 2 or 4 is externally perturbed or explicitly observed, so their numbering is arbitrary (i.e. 2 could be numbered 4 and vice-versa). Recognising this, there are only five indistinguishable models, with pairs of Models 1 and 2, 3 and 10, 4 and 9, 5 and 8, and 6 and 7 being the same except for peripheral compartment numbering.

Structural identifiability analysis in [4] shows that the postulated model has six moment invariants and is therefore unidentifiable. Alternative models 1 to 10, however, each have 6 rate constants and all are structurally identifiable. Models 3 (10), 5 (8) and 6 (7) have some rate constants globally identifiable and others locally identifiable; Model 1 (2) has all rate constants locally identifiable; while Model 4 (9) has all rate constants globally identifiable [4].

4. Results

The application of DISTING has been illustrated with three examples, in which the original (postulated) models have had respectively two, three and four compartments. The 2-compartment example was used to illustrate the indistinguishability problem; the original (unidentifiable) model was found to have two structurally identifiable models that are indistinguishable from it.

From the DISTING analysis, the 3-compartment example had three structurally identifiable indistinguishable models. However, this example showed that DISTING does not necessarily provide a complete indistinguishability analysis, because it is based on purely symbolic algebra manipulations of the postulated structural model, with no regard to numerical values of its rate constants. It does not, for example, distinguish between models having complex versus real eigenvalue model solutions. Additional data and/or numerical analysis are needed to fully analyse such models.

DISTING indicated that the original (postulated) model in the 4-compartment example apparently has ten structurally identifiable models indistinguishable from the original postulated model. This example showed that care is needed when dealing with models that have two or more compartments which have neither direct input or direct observation, because the numbering of these compartments may be arbitrary. Once this was taken into account in this example, the ten indistinguishable models were found to

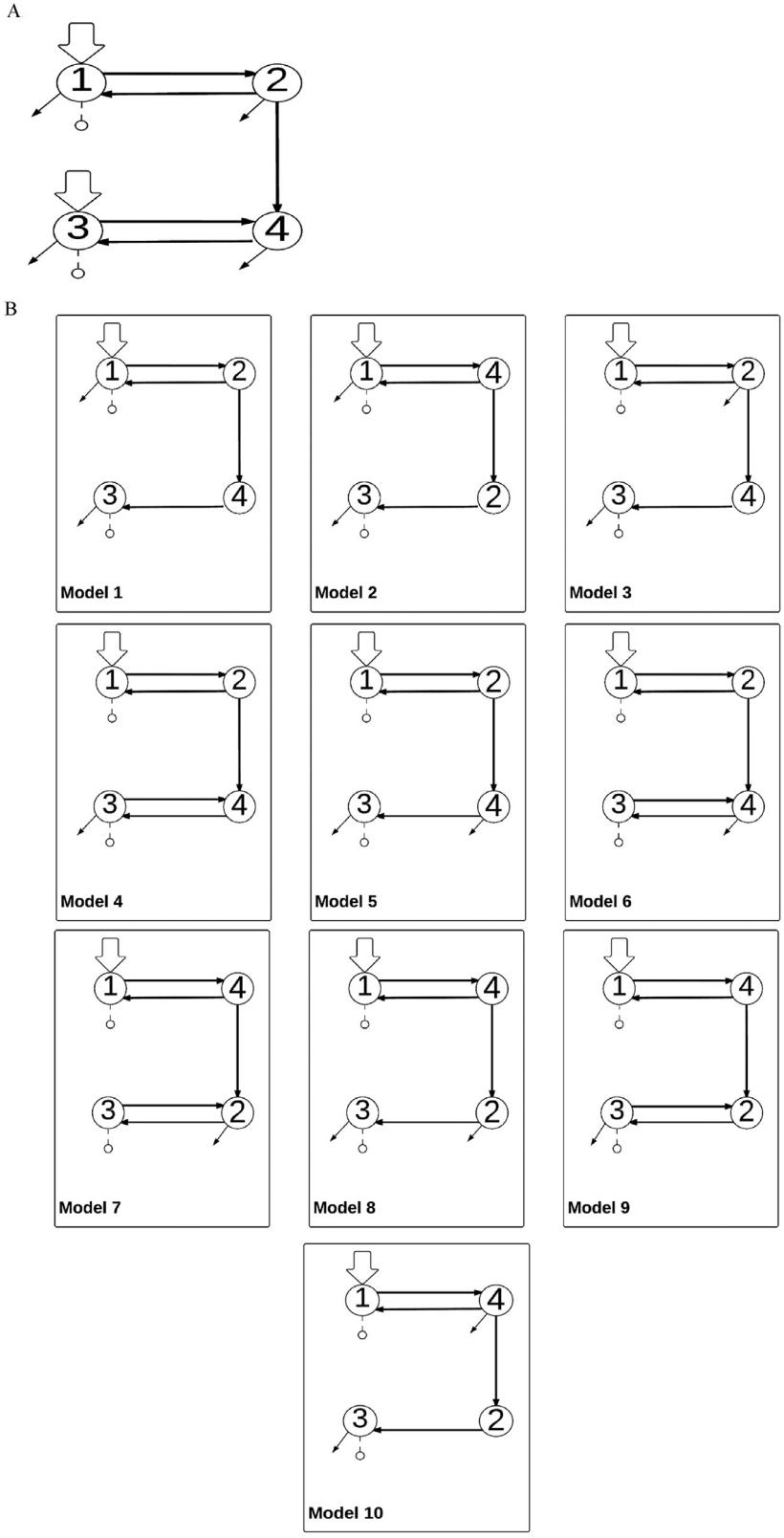

Fig. 5. (A) Postulated 4-compartment model (Model 0). (B) Indistinguishable 4-compartment models, 5 of which are redundant (see text).

consist of five independent pairs of structurally identifiable indistinguishable models. For models with two or more unperturbed and unobserved compartments, such a check can be readily completed once DISTING has produced the indistinguishable model candidates.

5. Discussion

We have described and illustrated DISTING, a new software tool implemented as a Web application, universally and freely available for finding structurally identifiable linear

compartmental models that are input-output (I-O) indistinguishable from a postulated model. DISTING is easy to use and circumvents tedious and complicated algebraic analysis previously done by hand. Runtimes on a (10 year old) Linux server at UCLA were less than 10 s for Examples 1 and 2, and just over a minute for Example 3. (Runtimes typically vary because the server also runs other on-demand apps, controlled by a queuing system developed using the open source continuous integration tool Jenkins [http://jenkins.io].).

Our web application is designed to make model indistinguishability analysis more accessible to modelers with a variety of skill-levels. It provides alternative I-O equivalent models, in both matrix form and compartmental diagram form, with the same number of rate constants as there are moment invariants in the model transfer function, so that all models produced are structurally identifiable, even when the postulated model is not. DISTING does not specifically check structural identifiability (SI) properties of the models produced. They are SI, but may be locally or globally SI. This can be assessed separately, e.g. using COMBOS [3] (http://biocyb1.cs.ucla.edu/combos), resident on the same Linux server running DISTING. The postulated model (Model 0) and the alternative models provided by DISTING can be downloaded in png format, using current versions of Google Chrome, Firefox or Safari Web browsers.

DISTING is accessed at http://biocyb1.cs.ucla.edu/DISTING. The structure of postulated Model 0 is entered by checking entries in an *adjacency* (connection) matrix A in which element $a_{i,j}$ denotes a link from Compartment j to Compartment i and element $a_{i,i}$ denotes an elimination from Compartment i to the environment. The locations of compartmental inputs and observations (measurements) are entered by checking appropriate entries in vectors R and M. Note that the adjacency matrix A is NOT the same as the system matrix K (Eq. (1)).

References

[1] G. Beliu, M. Saccomani, S. Audoly, L. D'Angio, A new software tool to test global identifiability of biological and physiological systems, *Comput. Methods Programs Biomed.* 88 (2007) 52–61.

[2] O. Chis, J.R. Banga, E. Balsa-Canto, GenSSI: a software toolbox for structural identifiability analysis of biological models, Bioinformatics 27 (2011) 2610–2611.

[3] N. Meshkat, C.E.-z. Kuo, J. DiStefano III, On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: a novel Web implementation, PLOS ONE 9 (2014) e110261.

[4] M.J. Chapman, K.R. Godfrey, A methodology for compartmental model indistinguishability, *Math. Biosci.* 96 (1989) 141–164.

[5] K.R. Godfrey, M.J. Chapman, The problem of indistinguishability in pharmacokinetics, *J. Pharmacokinetics Biopharm.* 17 (1989) 229–267.

[6] K.R. Godfrey, M.J. Chapman, Identifiability and indistinguishability of linear compartmental models, *Math. Comput. Simul.* 32 (1990) 273–295.

[7] M.J. Chapman, K.R. Godfrey, S. Vajda, Indistinguishability for a class of nonlinear compartmental models, *Math. Biosci.* 119 (1994) 77–95.

[8] N.D. Evans, M.J. Chappell, M.J. Chapman, K.R. Godfrey, Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems, Automatica 40 (2004) 1947–1953.

[9] Ll-Q. Zhang, J.C. Collins, P.H. King, Indistinguishability and identifiability analysis of linear compartmental models, *Math. Biosci.* 103 (1991) 77–95.

[10] M.J. Chapman, K.R. Godfrey, On structural equivalence and identifiability constraint ordering, in: E. Walter (Ed.), Identifiability of parametric models, Pergamon, New York, 1987, pp. 32–41.

[11] D.M. Foster, J.A. Jacquez, Multiple zeros for eigenvalues and the multiplicity of traps of a linear compartmental system, *Math. Biosci.* 26 (1975) 89–97.

[12] Celery Distributed Task Queue software, http://www.celeryproject.org/.

[13] K.R. Godfrey, J.J. DiStefano III, Identifiability of model parameters, in: E. Walter (Ed.), Identifiability of parametric models, Pergamon, New York, 1987, pp. 1–20.

[14] J.J. DiStefano III, Dynamic systems biology modeling and simulation, Academic Press, 2013.

[15] K.R. Godfrey, M.J. Chapman, A systematic approach to indistinguishability analysis of compartmental models, in: J. Eisenfeld, D.S. Levine (Eds.), Biomedical modelling and simulation, J.G. Baltzer AG Scientific Publishing, 1989, pp. 91–93.

[16] K.R. Godfrey, Compartmental models and their application, Academic Press, 1983.

[17] M. Milanese, G.P. Molino, Structural identifiability of compartmental models and pathophysiological information from the kinetics of drugs, *Math. Biosci.* 26 (1975) 175–190.

[18] M. Milanese, G. Belforte, Structural problems in identification, state estimation and aggregation, *Large Scale Syst.* 2 (1981) 97–104.