The input–output relationship approach to structural identifiability analysis

Daniel J. Bearup [a,*], *Neil D. Evans* [b], *Michael J. Chappell* [b]

[a] *Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK*
[b] *School of Engineering, University of Warwick, Coventry CV4 7AL, UK*

ABSTRACT

Analysis of the identifiability of a given model system is an essential prerequisite to the determination of model parameters from physical data. However, the tools available for the analysis of non-linear systems can be limited both in applicability and by computational intractability for any but the simplest of models. The input–output relation of a model summarises the input–output structure of the whole system and as such provides the potential for an alternative approach to this analysis. However for this approach to be valid it is necessary to determine whether the monomials of a differential polynomial are linearly independent. A simple test for this property is presented in this work. The derivation and analysis of this relation can be implemented symbolically within *Maple*. These techniques are applied to analyse classical models from biomedical systems modelling and those of enzyme catalysed reaction schemes.

© 2012 Elsevier Ireland Ltd. Open access under CC BY license.

1. Introduction

Structural identifiability methods test whether the parameters of a parameterised model can be uniquely determined (or otherwise) from perfect, continuous and noise free, observations. Determining structural identifiability is essential if parameters are to be estimated from real experimental data. For linear models a variety of techniques can be employed for this analysis, see examples in [1]. Methods by which non-linear models can be analysed are rather more limited, with only the Taylor series approach [2]; similarity transformation approach [3]; differential algebra approaches (see, for example, [4,5]); and other related approaches [6,7], available. Furthermore, each of these techniques has certain weaknesses when applied to non-linear systems; which are frequently of particular interest in biomedical systems modelling.

For linear models strict upper limits on the number of Taylor series coefficients required to determine the possible solutions are known [8]. However, for non-linear systems only a loose upper limit has been determined [9] and, as such, it is typically difficult to prove that a given model is unidentifiable using this technique. Furthermore, the complexity of higher order Taylor series coefficients often renders this approach computationally intractable. The generalisation of the similarity transform approach to non-linear models provides a relatively straightforward test for unidentifiability but is more difficult to use to prove local or global identifiability [10].

The original differential algebra approach [4] requires analysis of characteristic sets, the calculation of which can be computationally expensive. Analysis of the input–output relationship [5] is less computationally expensive, but it is necessary to establish the linear independence of the monomials of this relationship before analysis can proceed. Some implementations of this approach fail to check this requirement, while others make use of an appropriate Wronskian calculation [11]. Furthermore, the use of numerical solutions at certain stages of the analysis in some implementations

* *Corresponding author*. Tel.: +44 116 252 5060; fax: +44 116 223 1093.
  E-mail address: djb69@le.ac.uk (D.J. Bearup).

renders the results somewhat non-generic thereby limiting their applicability [12,13].

For any of these techniques the use of computational packages is typically necessary. For simple models it may be possible to implement them by hand; however the majority of models prove too complex for such an approach. Instead these methods are typically implemented using a symbolic computer algebra package. The use of such packages to solve complex equations is computationally expensive, consequently it is desirable to use the simplest equations possible when determining identifiability. This is the chief advantage of the input–output relationship approach over the Taylor series approach in that it typically produces relatively simple equations to solve for the model parameters.

In this work an implementation of the input–output approach in *Maple* using the Rosenfeld–Gröbner algorithm is presented. A criterion for the linear independence of the monomials of the relation analysed is introduced. The input–output relationship approach is first applied to a classical model from biomedical systems modelling. The results obtained agree with those published using alternative techniques. The approach is then applied to three enzyme reaction models. Study of such systems is a prerequisite to construction of metabolic pathway models, which have applications to drug development and design [14]. However, typically they prove too complex for analysis using other techniques.

2. Structural identifiability

A parameterised state space model, $\Sigma(\mathbf{p})$, can be defined as follows [15]:

$$\mathbf{x}^{(1)}(t, \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) + u(t)\mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}), \tag{1}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p}), \tag{2}$$

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{h}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}), \tag{3}$$

where $\mathbf{x}(t, \mathbf{p}) = (x_1(t, \mathbf{p}), \ldots, x_n(t, \mathbf{p}))^T$ is the state vector which lies in a connected open subset $M(\mathbf{p}) \subset \mathbb{R}^n$. Note that $\mathbf{x}(t, \mathbf{p})$ and $\mathbf{y}(t, \mathbf{p})$ are also dependent on $u(t)$, due to Eq. (1). This dependence is suppressed in the notation in the interests of brevity. To maximise flexibility Lagrange's notation for a derivative with respect to time, $dx/dt = x^{(1)}$, is used.

The vector of model outputs, $\mathbf{y}(t, u; \mathbf{p}) \in \mathbb{R}^r$, comprises the combination of elements of the state vector which are measured experimentally. Let $\mathbf{p} = (p_1, \ldots, p_q)^T$ be a vector of unknown model parameters which lies in some open set, $\Omega \subset \mathbb{R}^q$, of feasible values. The input $u(t) \in U$, the set of admissible controls, is assumed to be analytic for $t \geq 0$ and rational in $\mathbf{p}$. The functions $\mathbf{f}(\cdot, \mathbf{p})$ and $\mathbf{g}(\cdot, \mathbf{p})$ which determine the time dependent state transitions, and $\mathbf{h}(\cdot, \mathbf{p})$ which determines the model outputs, are similarly analytic on $M(\mathbf{p})$ and rational with respect to $\mathbf{p}$ [15].

Structural identifiability is concerned with whether the parameter vector is uniquely determined by the resulting input–output structure. For a model Σ a parameter vector $\overline{\mathbf{p}}$ is **indistinguishable** from $\mathbf{p}$, denoted $\mathbf{p} \sim \overline{\mathbf{p}}$, if, for all inputs, u, they give rise to identical model outputs, that is:

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{y}(t, \overline{\mathbf{p}}), \quad \text{for all } t \geq 0. \tag{4}$$

Following the definitions presented by Hattersley et al. [16], for generic $\mathbf{p} \in \Omega$ (that is, for all $p \in \Omega$ except for a subset of a closed set of Lebesgue measure zero) a parameter p_i is **locally identifiable** (LI) if there exists a neighbourhood of points around $\mathbf{p}$, $N(\mathbf{p})$, such that if $\overline{\mathbf{p}} \in N(\mathbf{p})$, $\mathbf{p} \sim \overline{\mathbf{p}}$ implies that $\overline{p}_i = p_i$. If no such $N(\mathbf{p})$ exists for p_i it is **unidentifiable**. If $N(\mathbf{p}) = \Omega$ for p_i then it is **globally identifiable** (GI). A model is unidentifiable if any parameter is unidentifiable. It is **structurally locally identifiable** (SLI) if all parameters are LI and at least one is not GI. It is **structurally globally identifiable** (SGI) if all parameters are GI.

3. The input–output relationship approach

The input–output relationship approach derives from the differential algebra approaches developed by Ljung and Glad [4]. The differential equations and equilibrium relations defining the system model are the generators of a radical differential ideal [17] (note that this requires the system to be stated in polynomial terms. However, in general a rational system can be rearranged to a polynomial system as described by Margaria et al. [9]. As such this approach is more broadly applicable to rational as well as polynomial systems). This differential ideal can be decomposed into an intersection of differential ideals using the Rosenfeld–Gröbner algorithm [18] corresponding to the general and singular solutions of the differential equations [19]. Given an appropriate choice of ranking each ideal contains expressions in derivatives of the input and output functions only [20]. This (possibly vector) function is the input–output relation.

The examples presented in this work are all uncontrolled and have a single output, $y(t, \mathbf{p})$. Consequently, for these applications, the input–output relation is a scalar function in $y(t, \mathbf{p})$ and its derivatives. The following analysis can be readily generalised to the more complex controlled case with multiple outputs.

The input–output relation, denoted $R((y^{(i)}(t, \mathbf{p}))_{i=0}^m, \mathbf{p})$ where m is the order of the highest derivative of $y(t, \mathbf{p})$ that appears, can be formally considered a differential polynomial over the differential ring $\mathbb{K}\{y\}$. As for a normal polynomial, a differential polynomial is a linear sum of monomials in this case arising from a differential ring. Thus a monomial in this work is a product of y and its derivatives. For a formal definition of these concepts see [17] or more recently [18].

Note that the input–output relation is one of the generators of an ideal, consequently it must equal zero. However, the coefficients of this differential polynomial are generically nonzero as they are linear combinations of the elements of the parameter vector, $\mathbf{p}$. Hence the monomials of the input–output relation are linearly dependent. It can, however, be rearranged to give the highest order derivative of $y(t, \mathbf{p})$ as a (possibly rational) function of strictly lower order derivatives as follows:

$$y(t, \mathbf{p})^{(m)} = \hat{R}((y^{(i)}(t, \mathbf{p}))_{i=0}^{m-1}, \mathbf{p}). \tag{5}$$

This rearranged expression can be used for structural identifiability analysis as follows. Suppose that two indistinguishable parameter vectors, $\mathbf{p}$ and $\overline{\mathbf{p}}$ satisfy Eq. (4). Since $\mathbf{f}$, $\mathbf{g}$, $\mathbf{h}$ and $\mathbf{u}$ are analytic, by induction, they also satisfy:

$$y^{(k)}(t,\mathbf{p}) = y^{(k)}(t,\overline{\mathbf{p}}), \quad \text{for all } t \geq 0, k \in \mathbb{N}. \tag{6}$$

Thus for indistinguishable parameter vectors $\mathbf{p}$ and $\overline{\mathbf{p}}$:

$$y(t,\mathbf{p})^{(m)} - \hat{R}((y^{(i)}(t,\mathbf{p}))_{i=0}^{m-1},\mathbf{p}) = y(t,\mathbf{p})^{(m)} - \hat{R}((y^{(i)}(t,\mathbf{p}))_{i=0}^{m-1},\overline{\mathbf{p}}), \tag{7}$$

from which the following can be obtained:

$$y(t,\mathbf{p})^{(m))} - y(t,\mathbf{p})^{(m))} = 0 = \hat{R}((y^{(i)})_{i=0}^{m-1},\overline{\mathbf{p}}) - \hat{R}_\rho((y^{(i)})_{i=0}^{m-1},\mathbf{p}). \tag{8}$$

The right hand side of this expression can now be rearranged into the form of a differential polynomial. In obtaining Eq. (8) one monomial has been eliminated from the expression, consequently the remaining monomials may be linearly independent. If they are linearly independent then the coefficients must all be zero and the solution set for $\overline{\mathbf{p}}$ in terms of $\mathbf{p}$ may be determined. A single solution implies that the model is SGI; countably many solutions that it is SLI; and uncountably many that the model is unidentifiable.

Thus it is essential to determine whether the monomials of the differential polynomial obtained from Eq. (8) are linearly independent. If they can be rewritten as univariate polynomials then the Fundamental Theorem of Algebra guarantees this property. However, where this is not the case an alternative criterion is needed; this is discussed in the following section.

The derivation of input–output relations and their subsequent analysis were implemented in *Maple* [21] as follows:

Step 1: The differential equations and equilibrium relations describing the model are stated as the generators, {**G**}, of a differential ideal over the field of rational numbers extended by the unknown kinetic parameters and initial conditions.
Step 2: An autoreduced and orthonomic subset, {**G**₁}, of these expressions, with respect to an appropriate ranking (see the following section), is selected. The ideal generated by this subset is decomposed using the `Rosenfeld_Groebner` algorithm [20] and an elimination ranking such that u and $\mathbf{y}$ are ranked below the state variables (the use of an elimination ranking ensures that the input–output relation will be an element of the ideal generated [18]).
Step 3: Any singular solutions are analysed to determine to which system state they correspond. The denominators of generators of the non-singular ideal are incorporated into {**G**₁} and the resulting ideals are decomposed again using the `Rosenfeld_Groebner` algorithm. If the resulting solutions correspond to the singular solutions already obtained then these denominators are generically non-zero.
Step 4: The input–output relation is extracted from the non-singular ideal and rearranged to the form given by Eq. (5). An alternative unknown parameter vector, $\overline{\mathbf{p}}$, is introduced and the elimination described by Eq. (8) is undertaken. The resulting expression is rearranged into a polynomial form. The coefficients of the monomials of this expression are solved using the `solve` function in *Maple*, for the alternative

unknown parameter vector to determine the identifiability structure of the dynamic section of the model.
Step 5: Where the initial conditions cannot all be incorporated directly into the differential ideal the remaining generators of the non-singular ideal are used to establish additional constraints on the alternative parameter vector. Each model variable is described by one of these generators; consequently it is possible to introduce the unknown initial condition by considering the appropriate generator at $t=0$. The alternative unknown parameter vector, $\overline{\mathbf{p}}$, is introduced subject to the constraints introduced at Step 4 and an equation in the unknown parameters is constructed by elimination of terms of the input or output. The resulting equations are then solved for the unknown parameters as in Step 4.

3.1. *Criterion for linear independence*

The linear independence of the monomials, $M_i(t)$, of a differential polynomial can be checked by computation of a Wronskian [11]:

$$W(t) = \left| \begin{pmatrix} M_1 & \cdots & M_N \\ M_1^{(1)} & \cdots & M_N^{(1)} \\ & \ddots & \\ M_1^{(N-1)} & \cdots & M_N^{(N-1)} \end{pmatrix} \right|. \tag{9}$$

If there exists a time point at which the Wronskian is non-zero, then the monomials are linearly independent [22]. Typically components of the output, $y_i(t)$, will be included amongst the monomials, $M_i(t)$. As such using this approach is generally at least as computationally complex as the calculation of $N-1$ Taylor series coefficients, where N is the number of monomials in the input–output relation. As noted previously calculation of large numbers of Taylor series coefficients is undesirable; thus an alternative approach is needed.

For a more formal introduction to the concepts used in this section see [17] or more recently [18]. A ranking on a differential ring, $\mathbb{K}\{y\}$, is defined to be a total ordering on the ring which is compatible with the differentiations over the alphabet; that is, the ranking is preserved by differentiation and higher order derivatives are ranked higher than lower order derivatives. For a differential polynomial, P, in $\mathbb{K}\{y\}$, and a ranking, K, the **leader**, L, of P is the largest derivative with respect to the ranking which appears in P. Let d be the degree of L in P. The **initial**, I_P, of P is the coefficient of L^d in P. The **separant**, S_P, of P is its partial derivative with respect to its leader L ($S_P = \partial P/\partial L$). The differential polynomial Q is **partially reduced** with respect to P if no monomial in Q contains a proper derivative of L. It is **reduced** with respect to P if it is partially reduced with respect to P and its degree in L is less than d.

A set of polynomials A is **triangular** if its elements have different leaders; **autoreduced** if each element is reduced with respect to every other element; and **orthonomic** if the initials and separants of its elements are in $\mathbb{K}$ as described in [18] and [20].

If a set of generators is autoreduced and orthonomic then the ideal generated by them is prime [18] and the
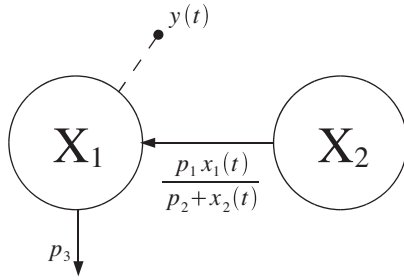
Fig. 1 – A nonlinear two compartment model of a batch reactor. The observed compartment is denoted by a dashed line ending in a bullet.

corresponding variety is irreducible. As such the generators cannot be factorised and must have a single solution. Thus if a single monomial of a generator is eliminated the resulting expression can only be identically zero if the coefficients of its monomials are zero; or equivalently the monomials are linearly independent. If instead the leader is eliminated, as proposed in Section 3, it is necessary to check that the initial of the leader is generically non-zero. If this is not the case the elimination process would require division by zero. This issue is addressed by adding the initial of the leader to the set of generators of the ideal. If the decomposition of this alternative ideal is different to that obtained from the original set of generators then the initial is generically non-zero. The conclusions above are otherwise unchanged.

4. Examples

Biomedical systems models vary widely in complexity, from relatively simple two or three compartment models to detailed reconstructions of metabolic pathways. The difficulty of identifiability analysis tends to be related to the structural complexity of the model considered. Nonetheless identifiability analysis of even relatively simple models can be computationally intractable if an inappropriate technique is used. In the following sections we consider four illustrative examples and apply the approach introduced in the previous sections to analyse their structural identifiability.

The first is a classical two compartment models that has been previously analysed using techniques similar to that presented here. It is used principally to demonstrate that this approach produces the same results as these previous analyses and to illustrate some of the subtleties of its implementation. The remaining three examples are models of enzyme catalysed reactions, the principal constituent of metabolic pathways. Such models are relatively unanalysed from a structural identifiability perspective, largely due to the complexity of the models required to fully capture their dynamic behaviour.

higher than one. Since, in practice, very few high order derivatives will be used in the following sections derivatives will, from now on, be denoted using Newton's notation, $dx/dt = \dot{x}$.

4.1. Batch reactor model

Populations of microorganisms are both a cause of disease and a potential tool in the production of pharmaceuticals. Consequently, developing models of population growth are of particular interest in biomedicine. A relatively simple example of such a model, Fig. 1, was described by Holmberg [23] and analysed by Chappell and Godfrey [24] and Evans et al. [10]. This model considers the interaction between the microbial population, represented by the first compartment, and a nutrient pool, represented by the second compartment. Conversion of nutrients to population is assumed to follow Michaelis–Menten type kinetics with respect to nutrients and linear kinetics with respect to population. This conversion process is not necessarily one to one necessitating the inclusion of a yield coefficient, denoted p_4. Population mortality is modelled as a linear population dependent process.

The model state vector, denoted $\mathbf{x}(t, \mathbf{p})$, describes the microbial population and concentration of available nutrients; initial levels are unknown parameters. The state vector ($\mathbf{x}$), initial conditions ($\mathbf{x}(0)$), and the unknown parameter vector ($\mathbf{p}$) are given by:

$$\mathbf{x}(t, \mathbf{p}) = (x_1(t, \mathbf{p}), x_2(t, \mathbf{p}))^T, \tag{10}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0 = (p_5, p_6)^T, \tag{11}$$

$$\mathbf{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T \in \mathbb{R}_+^6. \tag{12}$$

The state equations, which describe the transitions between the compartments of the model, are given by:

$$\dot{x}_1(t, \mathbf{p}) = \frac{p_1 x_1(t) x_2(t)}{p_2 + x_2(t)} - p_3 x_1(t), \tag{13a}$$

$$\dot{x}_2(t, \mathbf{p}) = \frac{-p_1 x_1(t) x_2(t)}{p_4(p_2 + x_2(t))}. \tag{13b}$$

A standard tool for structural identifiability analysis is the Taylor series approach introduced by Pohjanpalo [2]. This approach utilises the uniqueness of the coefficients in a Taylor series expansion of the output, usually about $t = 0$. For this system the first three Taylor series coefficients are:

$$y(0) = p_5, \tag{14}$$

$$\dot{y}(0) = \frac{p_5(p_1 p_6 - p_3(p_2 + p_6))}{p_2 + p_6}, \tag{15}$$

$$\ddot{y}(0) = \frac{p_5(p_1 p_6(2p_3 p_4(p_2 + p_6)^2 + p_1 p_2 p_5) - (p_2 + p_6)(p_1^2 p_6^2 p_4 + p_3^2 p_4(p_2 + p_6)^2))}{p_4(p_2 + p_6)^3}. \tag{16}$$

The use of Lagrange's notation in differential polynomials can be unclear when the derivative in question is of degree

Calculation of the first nine Taylor series coefficients takes about a second of computational time in *Maple* on an Intel®

E8500 (2×3.16 GHz, 3 GB RAM) chipset. An alternative parameter vector:

$$\overline{\mathbf{p}} = (\overline{p_1}, \overline{p_2}, \overline{p_3}, \overline{p_4}, \overline{p_5}, \overline{p_6})^T, \tag{17}$$

is introduced which is assumed to be indistinguishable from $\mathbf{p}$. Given this assumption the Taylor series coefficients for each of the parameter vectors must be the same, thus simultaneous equations in the two parameter vectors can be created. Solving these equations for the alternative parameter vector in *Maple* is also inexpensive on this system (approximately fifty seconds of computational time). The following set of relations is obtained:

$$\left\{ \begin{array}{l} \overline{p_1} = p_1, \overline{p_3} = p_3, \overline{p_5} = p_5, \\ \overline{p_2 p_4} = p_2 p_4, \ \ p_2 \overline{p_6} = \overline{p_2} p_6 \end{array} \right\} \tag{18}$$

From these relations p_1, p_3 and p_5 are globally identifiable. The upper limit on the number of Taylor series coefficients required, derived by Margaria et al. [9], is nine for this system. Consequently it is possible to conclude that this system is unidentifiable. Nonetheless this highlights a key weakness of the Taylor series approach. While it can be used to determine if a nonlinear model is globally identifiable it cannot always be used to determine if such a model is unidentifiable in cases where this upper limit on the number of coefficients required is not known. The input–output relationship approach is now applied to the same problem. The model equations are first restated below in the form of polynomial generators of a differential ideal, with explicit dependence on t and $\mathbf{p}$ omitted for brevity:

$$\dot{x}_1 + p_3 x_1 - p_1 x_1 x_2 x_3, \tag{19a}$$

$$p_4 \dot{x}_2 + p_1 x_1 x_2 x_3, \tag{19b}$$

$$p_4 \dot{x}_3 - p_1 x_1 x_2 x_3^3, \tag{19c}$$

$$y - x_1. \tag{19d}$$

Note that it is necessary to reduce the rational system given by Eq. (13) to a polynomial system. However it is not sufficient to simply multiply out by the denominators of the right hand sides of Eq. (13) since the resulting system cannot be made orthonomic. Instead an additional state variable $x_3 = 1/(p_2 + x_2)$ must be introduced. The derivative of this new state variable is derived as follows:

$$\dot{x}_3 = \frac{-\dot{x}_2}{(p_2 + x_2)} = -x_3^2 \dot{x}_2 = \frac{p_1 x_1 x_2 x_3^3}{p_4}. \tag{20}$$

Combined with the initial condition $x_3(0) = 1/(p_2 + p_6)$ this expression preserves the relationship between x_3 and x_2.

Eq. (19d) simply requires that x_1 always equals y, consequently we can eliminate it by making this substitution

throughout Eq. (19). The remaining generators form an orthonomic and autoreduced set with respect to any ranking. With an elimination ranking the Rosenfeld–Gröbner algorithm, implemented in *Maple*, decomposes this ideal into five ideals each of which corresponds to a particular system state. The first four are generated by the following expressions:

$$\{y, \dot{x}_2, \dot{x}_3\}, \{x_2, \dot{x}_3, \dot{y} + p_3 y\}, \{x_3, \dot{x}_2, \dot{y} + p_3 y\},$$
$$\{x_2 x_3 - 1, p_4 \dot{x}_2 + p_1 y, \dot{y} - (p_1 - p_3)y\}. \tag{21}$$

Each of these ideals is a reduction of the complete system in some sense. For example in the first three of these ideals the first generator requires that one of the model variables is zero for all t. In the fourth ideal the first generator, in combination with the definition of x_3, requires that $x_2/(p_2 + x_2) = 1$, a constraint on the second model variable which does not arise in the complete system. Ideals of this type, which do not describe the full model system, will be referred to as singular ideals. The system states to which they correspond will be referred to as singular system states.

These ideals can typically be excluded from the structural identifiability analysis because they violate the initial conditions of the model used. For the first two ideals here this is straightforward, if y or x_2 must be zero for all time, then the initial conditions for these variables must be equal to zero. Thus p_5 or p_6 must be zero and the parameter vector would not lie in $\mathbb{R}_+^6$ as required by Eq. (12). The third and fourth ideals, in combination with the definition of x_3, require that $x_2 = \infty$ for all time. This constraint requires that $p_6 = \infty$ and again the parameter vector with not lie in $\mathbb{R}_+^6$. Note that these singular system states can only arise in experiments designed to produce them.

The fifth ideal is generated by the following expressions:

$$x_2 = \frac{((p_1 - p_3)y - \dot{y})(\dot{y} + p_3 y)^2}{p_1 p_4 (\ddot{y}y - \dot{y}^2)}, \tag{22a}$$

$$x_3 = \frac{p_4 (\ddot{y}y - \dot{y}^2)}{y(\dot{y} + p_3 y)(\dot{y} + (p_1 - p_3)y)}, \tag{22b}$$

$$\dddot{y} = \frac{\ddot{y}^2 y(3\dot{y} + (3p_3 - p_1)y) + 2\ddot{y}\dot{y}y((p_3 - p_1)\dot{y} + 2p_3(p_3 - p_1)y) + (2p_1 - 3p_3)\dot{y}^4 + 3p_3(p_1 - p_3)\dot{y}^3 y}{y(\dot{y} + p_3 y)(\dot{y} + (p_1 - p_3)y)}, \tag{22c}$$

and does describe the full model system. Note that the first two generators have been rearranged to give state variables in terms of functions of y and its derivatives, while the third, the input–output relation, has been rearranged to give the highest order derivative of y in terms of lower order derivatives of y. This last rearrangement introduces the denominator:

$$y(\dot{y} + p_3 y)(\dot{y} + (p_1 - p_3)y), \tag{23}$$

This rearrangement is clearly invalid when this denominator is zero. To check that this is not always the case each factor of the denominator was individually added to the set of generators of the differential ideal, i.e. the set of equations which define the structure of the model. When the Rosenfeld–Gröbner algorithm is applied to these augmented

sets of generators, it produces one, three or two ideals respectively. Each of the ideals obtained in this way corresponds to one of the singular system states, Eq. (21), obtained above. Thus these augmented ideals describe model states which can be ignored for the purpose of this analysis. Hence the denominator can be assumed to be non-zero somewhere and so, by the analyticity of the model equations, it is generically non-zero.

With this constraint it is possible to eliminate the leader of the input–output relation, $\ddot{y}$, as described in Section 3. In particular a second version of the input–output relation is created by substitution of the alternative unknown parameter vector, Eq. (17). As argued in Section 3, the leaders of these two expressions are equivalent and can be eliminated by subtraction of one of these expressions from the other. The result is a rational expression which can be multiplied out to obtain a differential polynomial. Since this ideal is generated by an orthonomic and autoreduced set of generators the monomials of this differential polynomial are linearly independent as shown in Section 3.1. As this differential polynomial must always equal zero, and its six monomials are linearly independent, the coefficients of these monomials must be zero. Applying this condition they produce the following set of simultaneous equations in the alternative parameters:

$$3(p_3 - \overline{p_3}) - 2(p_1 - \overline{p_1}) = 0, \tag{24}$$

$$3(p_3 - \overline{p_3})(p_3 + \overline{p_3}) - 2(p_1 - \overline{p_1})(p_3 + \overline{p_3}) - (p_1 + \overline{p_1})(p_3 - \overline{p_3}) = 0, \tag{25}$$

$$3p_3\overline{p_3}(p_1 - \overline{p_1} - p_3 + \overline{p_3}) - p_1\overline{p_3}^2 + \overline{p_1}p_3^2 - p_1\overline{p_1}(p_3 - \overline{p_3}) = 0, \tag{26}$$

Note that three equations have been omitted, each differs from one of the equations above by an integer factor. The simultaneous equations can be solved in *Maple* for the alternative parameter vector, $\overline{\mathbf{p}}$, to obtain:

$$\{\overline{p_1} = p_1, \overline{p_3} = p_3\}. \tag{27}$$

Note that no information regarding the identifiability of the initial conditions, p_5 and p_6 has been obtained. This is because the initial conditions are not described as part of the structure of the dynamic portion of the model. Furthermore the addition of the state variable x_3 has concealed the influence of the parameters p_2 and p_4. To determine whether these parameters are identifiable the remaining elements of the differential ideal are used at $t = 0$. First, given that $y = x_1$ it is clear that:

$$\overline{p_5} = y(0, \overline{\mathbf{p}}) = y(0, \mathbf{p}) = p_5,$$
$$\overline{p_5} = p_5. \tag{28}$$

Note that the right hand side of Eq. (22a) contains one instance of the unknown parameter p_4 and is otherwise

entirely a function of the two identifiable parameters and derivatives of y, $F_1(y, \dot{y}, \ddot{y}, p_1, p_3)$. Thus:

$$\overline{p_4}x_2(0, \overline{\mathbf{p}}) = F_1(y(0), \dot{y}(0), \ddot{y}(0), p_1, p_3) = p_4x_2(0, \mathbf{p}),$$

and hence:

$$\overline{p_4p_6} = p_4p_6. \tag{29}$$

Finally note that the denominator of the RHS of Eq. (22a) and the numerator of the RHS of Eq. (22b) differ only by a factor of p_1. Multiplying these functions and reducing to simplest terms yields:

$$x_2x_3 = \frac{\dot{y} + p_3y}{yp_1}, \tag{30}$$

where again the RHS is a function only of derivatives of y and the identifiable parameters, $F_2(y, \dot{y}, p_1, p_3)$. Evaluating at $t = 0$ produces:

$$x_2(0, \overline{\mathbf{p}})x_3(0, \overline{\mathbf{p}}) = F_2(y(0), \dot{y}(0), p_1, p_3) = x_2(0, \mathbf{p})x_3(0, \mathbf{p}),$$

and hence:

$$\frac{\overline{p_6}}{\overline{p_2} + \overline{p_6}} = \frac{p_6}{p_2 + p_6}.$$

Multiplying out the denominators and eliminating the $p_6\overline{p_6}$ which appears on both sides yields:

$$p_2\overline{p_6} = \overline{p_2}p_6. \tag{31}$$

Note that by substitution of a rearrangement of this expression into Eq. (29) the following expression can be obtained:

$$\overline{p_2p_4} = p_2p_4. \tag{32}$$

Thus the final set of relations between the parameter vectors obtained is:

$$\{\overline{p_1} = p_1, \overline{p_3} = p3, \overline{p_5} = p_5, \overline{p_2p_4} = p_2p_4, p_2\overline{p_6} = \overline{p_2}p_6\} \tag{33}$$

which is equivalent to that obtained using Taylor series analysis. Thus the input–output approach yields the same result as the Taylor series approach in this case and agrees with the result obtained by Evans et al. [10]. This analysis takes less than a second of computational time within *Maple* on the system described above. Note however that, at present, this analysis requires more manual manipulation than the equivalent Taylor series analysis.

4.2. Single substrate enzyme catalysed fully reversible reaction mechanism

The previous model considered is relatively well known and understood. The focus of the remaining examples is enzyme catalysed reactions which have not been extensively studied with regard to identifiability. These reactions are usually studied experimentally in isolation, but are the building blocks of metabolic pathways. The first, relatively simple, example considers the classical Michaelis–Menten type reaction with a single substrate. Enzyme, E, and substrate, S, bind reversibly

forming a complex, ES. This complex breaks down reversibly releasing enzyme and product P. Experimentally the initial conditions can typically be controlled and it is relatively common, when considering a transient timescale, to be able to measure the complex concentration; this case is considered here. The reaction scheme below describes the desired mechanism [25]:

$$
\left.
\begin{array}{l}
E + S \underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES \\[2mm]
ES \underset{r_2}{\overset{k_2}{\rightleftharpoons}} E + P \\[2mm]
y = ES
\end{array}
\right\}
\tag{34}
$$

Note that this system is uncontrolled and proceeds simply from the initial conditions chosen. The concentrations of the reaction species constitute the model state vector, denoted $\mathbf{x}(t, \mathbf{p})$. The state vector, the initial conditions used, and the unknown parameter vector are given respectively by:

$$
\mathbf{x}(t, \mathbf{p}) = (E, S, ES, P)^T,
\tag{35}
$$

$$
\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0 = (E_0, S_0, 0, 0)^T,
\tag{36}
$$

$$
\mathbf{p} = (k_1, k_2, r_1, r_2)^T \in \mathbb{R}_+^4.
\tag{37}
$$

The derivatives of the elements of the state vector are given by:

$$
\dot{E} = -k_1 E \cdot S + (r_1 + k_2) ES - r_2 E \cdot P,
\tag{38a}
$$

$$
\dot{S} = -k_1 E \cdot S + r_1 ES,
\tag{38b}
$$

$$
\dot{ES} = k_1 E \cdot S - (r_1 + k_2) ES + r_2 E \cdot P,
\tag{38c}
$$

$$
\dot{P} = k_2 ES - r_2 E \cdot P.
\tag{38d}
$$

In addition the following conservation relationships can be derived from the initial conditions, Eq. (36), above:

$$
E_0 = E + ES,
\tag{38e}
$$

$$
S_0 = S + ES + P.
\tag{38f}
$$

A Taylor series analysis of this model was undertaken in the same way as those described in previous sections. The details are omitted in the interests of brevity. It suffices to state that this analysis demonstrates that if two parameter vectors $\mathbf{p}$ and $\bar{\mathbf{p}}$ are indistinguishable then they are equal. As such this model is SGI. These calculations were completed in under a second of computational time on the system described above.

Applying the input–output approach it is first necessary to rearrange the system equations to obtain an orthonomic and autoreduced set of generators. This is achieved by eliminating E and S using the conservation relations, Eqs. (38e) and (38f),

and replacing ES with y since this is the observation used. The resulting generators are given by:

$$
\dot{y} - k_1(E_0 - y)(S_0 - y - P) + (r_1 + k_2)y - r_2(E_0 - y)P,
\tag{39a}
$$

$$
\dot{P} - k_2 y + r_2(E_0 - y)P.
\tag{39b}
$$

The leaders of these generators are $y^{(1)}(t)$ and $P^{(1)}(t)$ respectively. Each leader appears in only one of the generators and no proper derivatives of the leaders appear in either generator, as such the generators are autoreduced. In addition since the coefficient of both leaders is 1, the generators are also orthonomic. Thus the criterion in Section 3.1 is met.

When applied to Eq. (39) the Rosenfeld–Gröbner algorithm produces a single ideal generated by the following expressions:

$$
P = \frac{\dot{y} - k_1 y^2 + (k_1(E_0 + S_0) + (k_2 + r_1))y - E_0 S_0 k_1}{(y - E_0(k_1 - r_2))},
\tag{40a}
$$

$$
\ddot{y} = \frac{E_0^2(k_1 k_2 + r_1 r_2 + (E_0 + 3S_0)k_1 r_2)y - E_0(2(k_1 k_2 + r_1 r_2) + 3(E_0 + S_0)k_1 r_2)y^2 + (k_1 k_2 + r_1 r_2 + (3E_0 + S_0)k_1 r_2)y^3 - k_1 r_2 y^4 + \dot{y}((k_1 + r_2)y^2 + E_0(E_0(k_1 + r_2) + k_2 + r_1) - 2E_0(k_1 + r_2)y) + \dot{y}^2 - E_0^3 S_0 k_1 r_2}{y - E_0},
\tag{40b}
$$

the latter of which is the input–output relation. Note that the generators have been arranged into a rational rather than differential polynomial form. In contrast to the batch reactor model, in this case there are no singular system states so it is possible to proceed immediately to analysis of the input–output relationship.

Note that it is still necessary to ensure that the denominator of the input–output relation, Eq. (40b):

$$
y - E_0,
\tag{41}
$$

is not always zero. This is relatively simple in this case, since if $y = ES = E_0$ then by Eq. (38e), $E = ES - E_0 = 0$, a contradiction of the initial conditions, Eq. (36). Thus the denominator is generically non-zero as previously argued. If the denominator is added to Eq. (39) and the Rosenfeld–Gröbner algorithm is applied to this augmented set of equations the empty set is output. This indicates that this augmented set of equations has that no solution, i.e. that they describe a physically impossible scenario. This difference from the batch reactor example arises precisely because there can be no singular states of this system.

With the constraint that the denominator is generically non-zero it is possible to eliminate the leader of the output relation, $\ddot{y}$, as previously described. In particular a second version of the output relation is created by substitution of an alternative unknown parameter vector given by:

$$
\bar{\mathbf{p}} = (\overline{k_1}, \overline{k_2}, \overline{r_1}, \overline{r_2})^T,
\tag{42}
$$

into Eq. (40b). As argued in Section 3, the leaders of these two expressions are equivalent and can be eliminated by subtraction of one of these expressions from the other. The resulting rational expression can then be multiplied out to obtain a differential polynomial. The eight monomials of this polynomial

are linearly independent since the criterion in Section 3.1 is met. Their coefficients are given below:

$$k_1 r_2 - \overline{k_1} \overline{r_2}, \tag{43}$$

$$E_0^3 S_0 (k_1 r_2 - \overline{k_1} \overline{r_2}), \tag{44}$$

$$(\overline{r_2} - r_2) + (\overline{k_1} - k_1), \tag{45}$$

$$2E_0 ((k_1 - \overline{k_1}) + (r_2 - \overline{r_2})), \tag{46}$$

$$E_0^2 ((\overline{k_1} - k_1) + (\overline{r_2} - r_2)) + E_0 ((\overline{k_2} - k_2) + (\overline{r_1} - r_1)), \tag{47}$$

$$(\overline{k_1 k_2} - k_1 k_2) + (3E_0 + S_0)(\overline{k_1 r_2} - k_1 r_2) + (\overline{r_1 r_2} - r_1 r_2), \tag{48}$$

$$3E_0(S_0 + E_0)(k_1 k_2 - \overline{k_1} \overline{r_2}) + 2E_0(k_1 k_2 + r_1 r_2 - \overline{k_1 k_2} - \overline{r_1 r_2}), \tag{49}$$

$$3E_0^2(S_0 + E_0)(\overline{k_1 r_2} - k_1 r_2) + E_0^2(\overline{k_1 k_2} + \overline{r_1 r_2} - k_1 k_2 - r_1 r_2). \tag{50}$$

These coefficients were solved in *Maple*, equal to zero, for the alternative parameters, $\overline{\mathbf{p}}$, to obtain the following solutions:

$$\{\overline{k_1} = k_1, \overline{k_2} = k_2, \overline{r_1} = r_1, \overline{r_2} = r_2\}, \tag{51}$$

$$\{\overline{k_1} = r_2, \overline{k_2} = r_1, \overline{r_1} = k_2, \overline{r_2} = k_1\}. \tag{52}$$

At first glance this solution differs from that obtained by the Taylor series approach. This is because the initial concentrations of *ES* and *P* cannot be incorporated directly into the differential ideal. In analysing the batch reactor model the remaining generators of the ideal yielded additional information about the identifiability of the initial conditions of the system. In this case the remaining generator, Eq. (40a), allows the neglected initial conditions to be used. Since $P(0) = 0$ and $y(0) = ES(0) = 0$, this equation becomes:

$$0 = -\dot{y}(0) + E_0 S_0 r_1. \tag{53}$$

Introducing the alternative parameter vector, $\overline{\mathbf{p}}$, and eliminating the leader yields:

$$0 = E_0 S_0 (r_1 - \overline{r_1}). \tag{54}$$

The only possible solution for this expression is $\overline{r_1} = r_1$. Only one of the solutions above, Eq. (51), satisfies this additional expression, and thus this model is SGI. This analysis required a similar amount of computational time to that undertaken using the Taylor series approach.

4.3. Two substrate enzyme catalysed reaction mechanism

The two substrate form of the Michaelis–Menten type enzyme catalysed reaction mechanism is considered next. Enzyme, E, and substrates, S_1 and S_2, bind sequentially and reversibly forming complexes, ES_1 and $ES_1 S_2$, the latter then breaks down

irreversibly releasing enzyme and product, P. The reaction scheme below describes the desired mechanism [25]:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 + S_2 &\underset{r_2}{\overset{k_2}{\rightleftharpoons}} ES_1 S_2 \\ ES_1 S_2 &\overset{k_3}{\rightarrow} E + P \\ y &= P \end{aligned}\right\} \tag{55}$$

Note that this system is uncontrolled and proceeds simply from the initial conditions chosen. The concentrations of the reaction species constitute the model state vector, denoted $\mathbf{x}(t, \mathbf{p})$. The state vector, the initial conditions used, and the unknown parameter vector are given respectively by:

$$\mathbf{x}(t, \mathbf{p}) = (E, S_1, S_2, ES_1, ES_1 S_2, P)^T, \tag{56}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0 = (E_0, S_{10}, S_{20}, 0, 0, 0)^T, \tag{57}$$

$$\mathbf{p} = (k_1, k_2, k_3, r_1, r_2)^T \in \mathbb{R}_+^5. \tag{58}$$

The derivatives of the state vector are given by:

$$\dot{E} = -k_1 E \cdot S_1 + r_1 ES_1 + k_3 ES_1 S_2, \tag{59a}$$

$$\dot{S}_1 = -k_1 E \cdot S_1 + r_1 ES_1, \tag{59b}$$

$$\dot{S}_2 = -k_2 ES_1 \cdot S_2 + r_2 ES_1 S_2, \tag{59c}$$

$$\dot{ES}_1 = k_1 E \cdot S_1 - r_1 ES_1 - k_2 ES_1 \cdot S_2 + r_2 ES_1 S_2, \tag{59d}$$

$$\dot{ES_1 S_2} = k_2 ES_1 \cdot S_2 - (r_2 + k_3) ES_1 S_2, \tag{59e}$$

$$\dot{P} = k_3 ES_1 S_2. \tag{59f}$$

In addition the following conservation relations can be derived from the initial conditions, Eq. (57):

$$E_0 = E + ES_1 + ES_1 S_2, \tag{59g}$$

$$S_{10} = S_1 + ES_1 + ES_1 S_2 + P, \tag{59h}$$

$$S_{20} = S_2 + ES_1 S_2 + P, \tag{59i}$$

A Taylor series analysis of this model could not be completed on the system described above due to the computational cost of deriving and solving the necessary Taylor series coefficients. However, an analysis using the input–output relationship approach was possible. The model equations can be reduced to an orthonomic and autoreduced set in the same way as the single substrate case, see below:

$$\dot{ES}_1 + r_1 ES_1 + k_2 ES_1(S_{20} - ES_1 S_2 - y) - r_2 ES_1 S_2$$
$$- k_1(E_0 - ES_1 - ES_1 S_2)(S_{10} - ES_1 - ES_1 S_2 - y), \tag{60a}$$

$$\dot{ES_1 S_2} - k_2 ES_1(S_{20} - ES_1 S_2 - y) + (r_2 + k_3) ES_1 S_2, \tag{60b}$$

$$\dot{y} - k_3 ES_1 S_2. \tag{60c}$$

The differential ideal generated by these expressions was decomposed into two ideals by application of the Rosenfeld–Gröbner algorithm. The first ideal corresponds to a singular model state and is generated by the following expressions:

$$\left\{ \begin{array}{l} y - S_{20}, \quad ES_1 S_2, \\ \dot{ES}_1 + (k_1 E_0 + r_1 - k_1 ES_1)ES_1 + k_1(ES_1 - E_0)(S_{10} - S_{20}). \end{array} \right\} \tag{61}$$

It is immediately clear from the first generator of this ideal, $y - S_{20} = 0$ or equivalently $y = S_{20}$, that this model state violates the initial conditions, Eq. (57), and as such can be ignored for the purposes of the identifiability analysis. Nonetheless the physical interpretation of this model state is of some interest. If $y = P = S_{20}$ then, from Eq. (59i), the concentration of the second substrate is zero. Thus this ideal describes a model state in which the second substrate has been consumed. However the third generator of the ideal describes a dynamic process, binding and release of the first substrate to and from the enzyme. This model state could potentially be used experimentally, with an appropriate choice of initial conditions and an alternative observation, to estimate the kinetic parameters of this subsystem.

The second ideal corresponds to the complete model state. The generators of this ideal, and in particular the input–output relation are very long and are omitted in the interests of brevity. Several of these generators contain the denominator:

$$\dot{y} - k_3(S_{20} - y). \tag{62}$$

If this equation is added to the generators of the differential ideal and the resulting ideal decomposed using the Rosenfeld–Gröbner algorithm a single ideal is obtained equivalent to that given by Eq. (61). As such Eq. (62) is generically non-zero. With this constraint it is possible to eliminate the leader of the output relation, $\ddot{y}$, by substitution of an alternative unknown parameter vector given by:

$$\overline{\mathbf{p}} = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{r_1}, \overline{r_2})^T, \tag{63}$$

and obtain a differential polynomial as previously described. The criterion stated in Section 3.1 being met, the 34 monomials of this polynomial are linearly independent. The coefficients of the polynomial were solved equal to zero for the alternative parameter vector, $\overline{\mathbf{p}}$, to obtain the following four solutions:

$$\{\overline{k_1} = \overline{k_2} = 0\}, \{\overline{k_2} = \overline{k_3} = 0\}, \{\overline{k_1} = \overline{k_3} = 0\}, \{\overline{k_1} = k_1,$$
$$\overline{k_2} = k_2, \overline{k_3} = k_3, \overline{r_1} = r_1, \overline{r_2} = r_2\}. \tag{64}$$

The first three solutions do not lie within the set of feasible parameter values, and as such they are rejected. Thus a unique solution remains, $\overline{\mathbf{p}} = \mathbf{p}$, showing that with this observation the model is SGI. These calculations required less than two seconds of computational time on the system described above.

Note however that relationships need only be found between the five parameters and the five alternative parameters. Thus most of the thirty-four equations solved are likely to provide no significant information. A subset of the equations that yielded the same solution was found as follows:

Step 1: The parent set of equations were ordered according to their complexity, using Maple functions.
Step 2: The first, and thus simplest, equation was added to the subset of equations to be solved. It was then solved for one of the alternative parameters, $\overline{p_i}$, that appeared in this equation.
Step 3: The relation obtained at Step 2 was used to eliminate $\overline{p_i}$ from the parent set of equations, reducing the number of non-trivial equations remaining. This reduced set of equations was then taken as the new parent set, and the process was returned to Step 1.

This process yielded the following five coefficients:

$$k_3^2 \overline{k_3}^{-2}(k_2 \overline{k_1} - \overline{k_2} k_1), \tag{65}$$

$$k_3^2 \overline{k_3}^{-2}(\overline{k_2 k_3}(k_2 - k_1) + k_2 k_3(\overline{k_1} - \overline{k_2})), \tag{66}$$

$$k_2 k_3 \overline{k_2 k_3}((k_2 \overline{k_3} - k_3 \overline{k_2}) + 2(k_1 \overline{k_3} - k_3 \overline{k_1})), \tag{67}$$

$$k_2 k_3^3 \overline{k_2 k_3}^3 S_{20}^2 (\overline{k_3} - k_3 + E_0 S_{10}(\overline{k_1} - k_1) + \overline{r_1} - r_1$$
$$+ S_{20}(\overline{k_2} - k_2) + \overline{r_2} - r_2), \tag{68}$$

$$k_2 \overline{k_2}(S_{20}(k_1 k_2 \overline{k_3}^3 - k_3^3 \overline{k_1 k_2}) + k_3^2 \overline{k_3}^2(k_2 - \overline{k_2} + 2(k_1 - \overline{k_1}))$$
$$+ k_3 \overline{k_3}(2(k_1 r_2 \overline{k_3} - k_3 \overline{k_1 r_2}) + (k_1 k_2 \overline{k_3} - k_3 \overline{k_1 k_2})(E_0 + S_{10} + 2S_{20})), \tag{69}$$

which were then solved as a group for the alternative parameter vector. The resulting solutions were identical to those obtained by solving the entire system of equations, Eq. (64).

4.4. Inhibited single substrate enzyme catalysed reaction mechanism

The final model is a return to a single substrate mechanism. However a second species, *I*, is added to the reaction which inhibits the reaction process by binding to the enzyme to form an unreactive complex, *EI*. In this case the final product release step is assumed to be irreversible as it was for the two substrate reaction. The mechanism is described by the reaction scheme below [25]:

$$\left. \begin{array}{l} E + S \underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES \\ E + I \underset{r_i}{\overset{k_i}{\rightleftharpoons}} EI \\ ES \overset{k_2}{\rightarrow} E + P \\ y = P \end{array} \right\} \tag{70}$$

The concentrations of the reaction species constitute the model state vector, denoted $\mathbf{x}(t, \mathbf{p})$. The state vector, the initial conditions used, and the unknown parameter vector are given by:

$$\mathbf{x}(t, \mathbf{p}) = (E, S, I, ES, EI, P)^T, \tag{71}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0 = (E_0, S_0, I_0, 0, 0, 0)^T, \tag{72}$$

$$\mathbf{p} = (k_1, k_i, k_2, r_1, r_i)^T \in \mathbb{R}_+^5. \tag{73}$$

The derivatives of the state vector are given by:

$$\dot{E} = -E \cdot (k_1 S + k_i I) + (r_1 + k_2)ES + r_i EI, \tag{74a}$$

$$\dot{S} = -k_1 E \cdot S + r_1 ES, \tag{74b}$$

$$\dot{I} = -k_i E \cdot I + r_i EI, \tag{74c}$$

$$\dot{ES} = k_1 E \cdot S - (r_1 + k_2)ES, \tag{74d}$$

$$\dot{EI} = k_i E \cdot I - r_i EI, \tag{74e}$$

$$\dot{P} = k_2 ES. \tag{74f}$$

In addition the following conservation relations can be found using the initial conditions, Eq. (72):

$$E_0 = E + ES + EI, \tag{74g}$$

$$S_0 = S + ES + P, \tag{74h}$$

$$I_0 = I + EI. \tag{74i}$$

An orthonomic and autoreduced set of generators was obtained as previously described and is given below:

$$\dot{ES} - k_1(E_0 - ES - EI)(S_0 - ES - y) + (r_1 + k_2)ES, \tag{75a}$$

$$\dot{EI} - k_i(E_0 - ES - EI)(I_0 - EI) + r_i EI, \tag{75b}$$

$$\dot{y} - k_2 ES, \tag{75c}$$

The differential ideal generated by this set of the generators was decomposed into two ideals using the Rosenfeld–Gröbner algorithm. The first ideal once again corresponded to a singular model state and contained the generator $y = S_0$ which violates the initial conditions, Eq. (72). Consequently this model state could be ignored for identifiability analysis. Note however that as in the previous example the physical interpretation of this model state is potentially of interest. In particular this model state corresponds to a case where substrate has been exhausted and the binding of enzyme to inhibitor can be observed in isolation. Thus parameters of this interaction might be estimated from this system if a suitable observation could be designed.

The second, non-singular, ideal is again generated by very long expressions which are omitted for brevity. The denominator of the input–output relation was shown to be generically non-zero as previously described in Section 4.3. With this

constraint the leader of the relation, $\dddot{y}$, was eliminated by substitution of the following alternative unknown parameter vector:

$$\overline{\mathbf{p}} = (\overline{k_1}, \overline{k_i}, \overline{k_2}, \overline{r_1}, \overline{r_i})^T, \tag{76}$$

to obtain a differential polynomial as previously described. The criterion stated in Section 3.1 being met, the 28 monomials of this polynomial are linearly independent. The following five coefficients of the polynomial, obtained by the method outlined in the previous section:

$$k_2 \overline{k_2}(k_1 \overline{k_i} - k_i \overline{k_1}), \tag{77}$$

$$k_1 k_2^2 \overline{k_1 k_2}^2 (k_1 - \overline{k_1}), \tag{78}$$

$$k_1 k_2^2 \overline{k_1 k_2}^2 (k_1 r_i - \overline{k_1} \overline{r_i}), \tag{79}$$

$$E_0 k_1 k_2^2 \overline{k_1 k_2}^2 (k_1 k_2 r_i + \overline{k_1 k_2} \overline{r_i}), \tag{80}$$

$$k_1 k_2 \overline{k_1 k_i}(\overline{k_2} + \overline{r_1}) - k_1 k_i \overline{k_1 k_2}(k_2 + r_1) + k_1 \overline{k_1}(k_1 r_i \overline{k_2} - k_2 \overline{k_1} \overline{r_i}), \tag{81}$$

are sufficient to establish that the model is SGI. Four solutions were found for the alternative parameter vector:

$$\{\overline{k_1} = \overline{k_2} = 0\}, \{\overline{k_1} = \overline{k_i} = 0\}, \{\overline{k_1} = \overline{k_i}, \overline{k_2} = 0, \overline{r_1} = \overline{r_i}\},$$
$$\{\overline{k_1} = k_1, \overline{k_i} = k_i, \overline{k_2} = k_2, \overline{r_1} = r_1, \overline{r_i} = r_i\}. \tag{82}$$

However, the first three do not lie within the set of feasible parameter values and, as such, are rejected. Thus a unique solution remains, $\overline{\mathbf{p}} = \mathbf{p}$, showing that with this observation the model is SGI. Note that once again identical results were obtained by solving all coefficients of the polynomial equal to zero. This analysis is, of course, based on the assumption that a non-zero input of the inhibitor is provided, if it is not, an alternative model should be used. Less than two seconds of computational time were required to complete this analysis on the system described previously. Note that a Taylor series analysis of this system again proved computationally intractable.

5.   Conclusion

Models of biological systems will typically include some non-linearity either due to binding kinetics, in enzyme catalysed reactions or otherwise, or due to the use of rate limiting approximations. As previously noted the methods by which the structural identifiability of such models can be analysed are limited. In this work two of the four, relatively simple, models analysed could not be completely analysed using one of the primary approaches for such analysis, the Taylor series approach, due to the computational complexity of the analysis. However a variant of the differential algebra approaches, based on the input–output relation, could be used in these cases.

This technique has previously been used by other authors, see for example [11] and [12], but its application has been

limited by the need to ensure linear independence of the monomials of the input–output relation. Previously the only way to check this condition has been computation of a Wronskian which is often at least as computationally complex as a Taylor series analysis of the same problem. By contrast, the criterion for linear independence presented in this work is easy to check by hand.

Furthermore, while other implementations of this technique have required the use of numerical solutions, in this work a symbolic implementation, in *Maple*, proved straightforward. While numerical solutions are easier to obtain, they are non-generic, and thus are applicable only to the cases which are directly considered. As such this demonstration that symbolic solutions can be obtained using this technique represents significant progress.

However for some systems, typically fully reversible enzyme reactions such as that considered in Section 4.2, generation of the input–output relationship via the Rosenfeld–Gröbner algorithm proves computationally intractable. Development of alternative algorithms to handle these systems constitutes a potential area of possible future work. In addition, note that the input–output relation can be used to generate the Taylor series coefficients of the model output. As such further analysis of this relation may provide a means to determine a strict upper limit on the number of coefficients required for a complete Taylor series analysis of a given model.

Conflict of interest

The authors have no conflicts of interest to disclose.

Acknowledgements

REFERENCES

[1] E. Walter, Identifiability of State Space Models, with Applications to Transformation Systems, Springer, Berlin, Heidelberg and New York, 1982.

[2] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, Mathematical Biosciences 41 (1978) 21–33.

[3] S. Vajda, K. Godfrey, H. Rabitz, Similarity transformation approach to identifiability analysis of nonlinear compartmental models, Mathematical Biosciences 93 (1989) 217–248.

[4] L. Ljung, T. Glad, On global identifiability for arbitrary model parametrizations, Automatica 30 (1994) 265–276.

[5] S. Audoly, G. Bellu, L. D'Angiò, M. Saccomani, C. Cobelli, Global identifiability of nonlinear models of biological systems, IEEE Transactions on Biomedical Engineering 48 (2001) 55–65.

[6] J. Nemcova, Structural identifiability of polynomial and rational systems, Mathematical Biosciences 223 (2010) 83–96.

[7] J.W.T. Yates, N.D. Evans, M.J. Chappell, Structural identifiability analysis via symmetries of differential equations, Automatica 45 (2009) 2585–2591.

[8] S. Vajda, Structural identifiability of linear, bilinear, polynomial and rational systems, in: L. Ljung, A. Titli (Eds.), Proceedings of the 9th IFAC World Congress, 1984, p. 107.

[9] G. Margaria, E. Riccomagno, M. Chappell, H. Wynn, Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences, Mathematical Biosciences 174 (2001) 1–26.

[10] N. Evans, M. Chapman, M. Chappell, K. Godfrey, Identifiability of uncontrolled nonlinear rational systems, Automatica 38 (2002) 1799–1805.

[11] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, Some effective approaches to check the identifiability of uncontrolled nonlinear systems, Mathematics and Computers in Simulation 57 (2001) 35–44.

[12] G. Bellu, M. Saccomani, S. Audoly, L. D'Angiò, DAISY: a new software tool to test global identifiability of biological and physiological systems, Computer Methods and Programs in Biomedicine (2007) 52–61.

[13] A. Sedoglavic, A probabilistic algorithm to test local algebraic in polynomial time, Journal of Symbolic Computation 33 (2002) 735–755.

[14] D.B. Kell, Systems biology, metabolic modelling and metabolomics in drug discovery and development, Drug discovery today 11 (2006) 1085–1092.

[15] N. Evans, M. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Mathematical Biosciences 168 (2000) 137–159.

[16] J.G. Hattersley, J. Pérez-Velázquez, M.J. Chappell, D. Bearup, D. Roper, C. Dowson, T. Bugg, N.D. Evans, Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis, Computer Methods and Programs in Biomedicine 104 (2011) 70–80.

[17] J. Ritt, Differential Algebra, 1st edition, AMS Colloquium Publications, New York, 1950.

[18] F. Boulier, D. Lazard, F. Ollivier, M. Petitot, Representation for the radical of a finitely generated differential ideal, in: in: ISSAC'95: Proceedings of the International Symposium on Symbolic and Algebraic Computation, 1995, pp. 158–166.

[19] E. Hubert, Étude algébrique et algorithmique des singularités des équations différentielles implicites, Ph.D. thesis, Institut National Polytechnique de Grenoble, 1997.

[20] F. Boulier, D. Lazard, F. Ollivier, M. Petitot, Computing representations for radicals of finitely generated differential ideals, Applicable Algebra in Engineering, Communication and Computing 20 (2009) 73–121.

[21] M.B. Monagan, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter, J. McCarron, P. DeMarco, Maple 10 Programming Guide, Maplesoft, Waterloo ON, Canada, 2005.

[22] M. Krusemeyer, Why does the Wronskian work, American Mathematical Monthly 95 (1988) 46–49.

[23] A. Holmberg, On the practical identifiability of microbial growth models incorporating Michaelis–Menten type nonlinearities, Mathematical Biosciences 62 (1982) 23–43.

[24] M.J. Chappell, K.R. Godfrey, Structural identifiability of the parameters of a nonlinear batch reactor model, Mathematical Biosciences 108 (1992) 241–251.

[25] M. Dixon, E.C. Webb, Enzymes, 3rd edition, Longman Group Ltd., New York, 1979.