THE UNIVERSITY OF WARWICK

Original citation:
Evans, N. D. et al. Structural identifiability of surface binding reactions involving heterogeneous analyte: application to surface plasmon resonance experiments. Automatica. Forthcoming

Permanent WRAP url:
http://wrap.warwick.ac.uk/49183

Publisher's statement:

A note on versions:
The version presented here is a working paper or pre-print that may be later published elsewhere. If a published version is known of, the above WRAP url will contain details on finding it.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
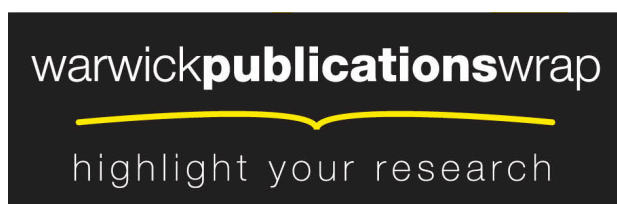
warwickpublicationswrap

highlight your research

http://go.warwick.ac.uk/lib-publications

Structural identifiability of surface binding reactions involving heterogeneous analyte: application to surface plasmon resonance experiments

N. D. Evans [a,1], H. A. J. Moyse [b], D. Lowe [c,d], D. Briggs [d] R. Higgins [e] D. Mitchell [c], D. Zehnder [c,e], M. J. Chappell [a]

[a] *School of Engineering, University of Warwick, Coventry, CV4 7AL UK*

[b] *MOAC Doctoral Training Center, University of Warwick, Coventry, CV4 7AL UK*

[c] *Clinical Sciences Research Laboratory, University of Warwick, Clifford Bridge Road, Coventry CV2 2DX UK*

[d] *Histocompatibility Laboratory, NHS Blood and Transplant, Birmingham B15 2SG UK*

[e] *Transplant Unit, University Hospitals Coventry and Warwickshire NHS Trust, Clifford Bridge Road, Coventry CV2 2DX UK*

Abstract

Binding affinities are useful measures of target interaction and have an important role in understanding biochemical reactions that involve binding mechanisms. Surface plasmon resonance (SPR) provides convenient real-time measurement of the reaction that enables subsequent estimation of the reaction constants necessary to determine binding affinity. Three models are considered for application to SPR experiments—the well mixed Langmuir model and two models that represent the binding reaction in the presence of transport effects. One of these models, the effective rate constant approximation, can be derived from the other by applying a quasi-steady state assumption. Uniqueness of the reaction constants with respect to SPR measurements is considered via a structural identifiability analysis. It is shown that the models are structurally unidentifiable unless the sample concentration is known. The models are also considered for analytes with heterogeneity in the binding kinetics. This heterogeneity further confounds the identifiability of key parameters necessary for reliable estimation of the binding affinity.

Key words: Biomedical systems, structural identifiability, surface plasmon resonance, surface-volume reactions, binding affinity, system identification

1 Introduction

Surface plasmon resonance (SPR) provides convenient real-time measurement of the biophysical parameters of noncovalent interactions between a free flowing analyte and an immobilised ligand. These measurements subsequently allow estimation of reaction constants of association and dissociation from the target protein on the SPR chip, permitting estimation of the relevant binding affinity (the ratio of the dissociation and association constants). Binding affinities are useful measures of target interaction and have an important role in understanding

biochemical reactions that involve binding mechanisms, in particular for antibody with target protein interaction.

A number of authors have considered the problem of determining the rate constants for a binding reaction in which one of the reactants is immobilised on a sensor surface. One approach is to model the reaction as occurring on the sensor surface via a reacting boundary condition on a transport equation for the analyte (Myszka et al., 1998; Edwards et al., 1999). Another approach has been to include a receptor layer in the model, to allow for the dextran layer in which the receptor is immobilised, and to also consider analyte diffusion into the layer to reach binding sites (Edwards, 2001). Typically the full partial differential equations framework is used to consider the appropriateness of ordinary differential equation (ODE)

based models that are more easily applied to experimental data to yield estimates for the binding constants (see, for example, Edwards (2001)). The simplest ODE model is the rapid mixing model that considers the reaction, after a brief transient period, as the same as two reactants in a well-mixed volume. However, the measured interaction between analyte and receptor is a combination of the binding reaction and transport effects arising from the flow and diffusive processes (Myszka et al., 1998). Therefore, if the rapid mixing model is not appropriate then systematic errors in the estimates for the rate constants can arise (Chaiken et al., 1992). Myszka et al. (1998) propose a two-compartment model that includes transport effects while modelling the binding process using Langmuir-type mass action kinetics. Under a quasi-steady state (QSS) assumption this two-compartment model can be reduced to a single differential equation for the bound analyte. This QSS model is the *effective rate constant approximation*. Edwards (2001) derived the effective rate constant approximation directly from consideration of the fluid dynamics of the analyte in the flow and the receptor layer, and the subsequent binding. This model has been shown to provide a good approximation under certain conditions (Edwards, 2001).

In order to estimate the (unknown) model parameters from SPR data it is necessary to include in the model an output structure, which corresponds to the function of the model variables that is to be compared with data. Before actually collecting experimental data it is necessary to test for uniqueness of the unknown parameters with respect to this output structure, since estimates for unidentifiable parameters are effectively meaningless. Such a structural identifiability analysis (Bellman and Åström, 1970) assesses whether the model output contains enough information to determine all of the model parameters uniquely (Jacquez, 1996), and relates only to the structure of the model and output. For linear systems there are many well-established techniques for performing a structural identifiability analysis (for further details, and details of nonlinear approaches, see the tutorial by Godfrey and DiStefano III (1987) and other works in the same volume and the book by Walter (1982)).

For nonlinear systems, such as those involving binding kinetics, greater care has to be taken over the choice of technique, since the assumption that any external input is persistently exciting becomes more restrictive, and computational tractability becomes a greater issue. For a single analytic input (including zero), perhaps the easiest approach conceptually is the Taylor series approach introduced by Pohjanpalo (1978), which uses the uniqueness of the coefficients of a Taylor series expansion of the output about a known time point to determine uniqueness of the unknown parameters. Since the list of coefficients is infinite it is necessary to ensure that sufficient terms have been considered for a full analysis, though some work has been done for rational systems to give an upper bound on the number of terms required

(Margaria et al., 2001). Without an appropriate upper bound the approach only provides a sufficient condition for identifiability. Computational complexity arises because determining the coefficients requires repeated differentiation of the output and evaluation at the known time point. Other approaches that are suitable for systems without input (or single analytic input) are based on: the existence of a smooth transformation between locally observable models with identical outputs (Evans et al., 2002), differential algebra (Ljung and Glad, 1994; Saccomani et al., 2003), or on polynomial realisation theory (Němcová, 2010).

In this paper a differential equation for the output and its derivatives will be obtained that permits a direct analysis of the uniqueness of the parameters. This approach is similar in nature to that taken by Denis-Vidal et al. (2001). Essentially this is an elimination problem in converting from a state space description of the model to an input-output one. If the underlying system is locally observable then one way to achieve this is to apply a coordinate transformation that yields the observable normal form (Isidori, 1995), in which the last state equation is the required input-output map. However, a number of authors have sought to exploit the polynomial nature of many system models and so use algebraic techniques to perform the state to output conversion.

A rich area of research in structural identifiability has considered the identifiability problem via determination of a characteristic set of a differential ideal defining the system (see Ollivier (1990); Glad and Ljung (1990); Ljung and Glad (1994); Audoly et al. (2001); Saccomani et al. (2003) for example). With a suitable ranking on the variables the identifiability can be determined directly from terms in the characteristic set (Glad, 1992). Glad (1991) considered the problem of conversion from state space description to input-output form via characteristic sets, which has also been the topic of study by others (Saccomani et al., 2001; Bearup et al., 2011). For systems described by characteristic sets in input-output form Glad (1992) showed that a canonical parameterisation of the characteristic set is uniquely identifiable. In effect the canonical parameterisation (with corresponding characteristic set) provides an exhaustive summary for the model (Saccomani et al., 2001).

Forsman (1991) considered the application of methods from constructive algebra to the problem of state elimination to derive input-output relations. The approaches developed involved the generation of a suitable (differential) ideal and the determination of a suitable Gröbner Basis. The ideal required is generated from successive Lie derivatives of the output that, for observable systems, defines the coordinate transformation into Observable Normal Form. Andersson (1994) considered a similar ideal and the determination of a suitable Gröbner Basis to tackle the identifiability problem. Using this approach the ideals are considered with the parameter in-

cluded as indeterminates and the form of the resulting Gröbner Basis enables determination of the identifiability of the model.

For nonlinear systems the problem of computational intractability means that different approaches should be viewed as complementary, rather than mutually exclusive.

2 Compartmental models

Three core model types are applied to the problem of estimating kinetic rate constants from SPR data: the first model is the well-mixed Langmuir model, with the second model being the two-compartment model proposed by Myszka et al. (1998) that includes transport effects, while the third model, the effective rate constant approximation, can be derived from the second via a quasi-steady state assumption (or directly from consideration of the fluid dynamics of the analyte in the flow and subsequent binding (Edwards, 2001)).

Since the (Bio-Rad) SPR platform permits multiple analytes (up to six) across multiple immobilised ligands (again up to six) to be run in a single experiment the models are each applied to the interaction spots (up to 36 in total), with some parameter values possibly being common across certain channels. In addition, the possibility of the analyte being heterogeneous with respect to the binding kinetics is also considered.

2.1 Langmuir model

The homogeneous analyte form of the Langmuir model is given by:

$$\dot{B}(t) = k_a I(t) \left(R - B(t) \right) - k_d B(t) \qquad (1)$$

where $B(t)$ is the average bound (area) concentration, k_a and k_d are the association and dissociation rate constants, and R is the maximum (area) density of bound analyte possible at the interaction site. Finally, $I(t)$ is the inlet concentration of analyte in the channel and is given by

$$I(t) = \begin{cases} 0 & t \notin [t_s, t_f] \\ C_T & t \in [t_s, t_f] \end{cases} \qquad (2)$$

where C_T is the analyte sample concentration (in the flow), t_s is the start time and t_f the finish time for the association phase (the interval $[t_s, t_f]$). Therefore, the dissociation phase corresponds to $t > t_f$.

The output structure for the model consists of the measurements of bound analyte and is given by

$$y(t) = \alpha B(t), \qquad (3)$$

where α is the conversion factor from the units of $B(t)$ to the response units (RU) of the sensorgrams (1 RU = 10^{-3} ng/mm^2).

Two forms of heterogeneity in the analyte are considered: the analyte consists of a heterogeneous mixture (for example a polyclonal antibody sample), or a homogeneous mixture in which the analyte exhibits heterogeneity in its binding kinetics (such as that resulting from different binding domains on an IgM sample, for example).

For the heterogeneous mixture version of the Langmuir model it is assumed that the analyte consists of a number of different binding species with (unknown) inlet sample concentrations of C_{Ti} (so that $C_T = \sum_{i=1}^{m} C_{Ti}$). The modified model is then given by the following:

$$\dot{B}_i(t) = k_{ai} I_i(t) \left(R - \sum_{i=1}^{m} B_i(t) \right) - k_{di} B_i(t) \qquad (4)$$

where $B_i(t)$ is the average bound (area) concentration for analyte species i ($i = 1, \ldots, m$), k_{ai} and k_{di} are the association and dissociation rate constants, and

$$I_i(t) = \begin{cases} 0 & t \notin [t_s, t_f] \\ C_{Ti} & t \in [t_s, t_f] \end{cases} \qquad (5)$$

for $i = 1, \ldots, m$. In this case, assuming that the binding species are indistinguishable at the receptor, the output structure is given by the following:

$$y_i(t) = \alpha \sum_{i=1}^{m} B_i(t). \qquad (6)$$

For the heterogeneous analyte version of the Langmuir model it is assumed that the analyte is able to bind in a number of ways. Thus the modified model in this case is similar to (4) except that the inlet concentration, $I(t)$, is the same for all binding types:

$$\dot{B}_i(t) = k_{ai} I(t) \left(R - \sum_{i=1}^{m} B_i(t) \right) - k_{di} B_i(t) \qquad (7)$$

where $B_i(t)$ is the average bound (area) concentration for binding mechanism i; k_{ai} and k_{di} are the association and dissociation rate constants, and $I(t)$ is given by (2). In this case, the output structure is also given by (6).
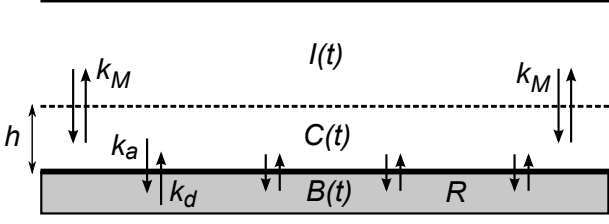
Fig. 1. Schematic of two-state Langmuir binding model that includes transport effects (Myszka et al., 1998).

2.2 Langmuir with transport model

The model proposed by Myszka et al. (1998) (see Fig 1) is given by the following:

$$h\dot{C} = -k_aC(R - B) + k_dB + k_M(I - C)$$
$$\dot{B} = k_aC(R - B) - k_dB \qquad (8)$$

where: $C(t)$ is the (volume) concentration of analyte at the surface of the channel; $h = V/S$, where V is the volume in contact with the surface from which binding takes place and S is the surface area; and k_M is the transport coefficient describing diffusive movement of analyte between the flow and the volume in contact with the surface. The output is again given by (3).

Note that, Myszka et al. (1998) report that for the range of parameter values associated with Biacore experiments the solution to (8) and (3) is insensitive to the value of h. This remark suggests a lack of structural (and/or numerical) identifiability for h that is investigated in subsequent sections.

The heterogeneous mixture version of the model is given by the following:

$$h\dot{C_i} = -k_{ai}C_i\left(R - \sum_{i=1}^{m} B_i\right) + k_{di}B_i + k_M(I_i - C_i)$$
$$\dot{B_i} = k_{ai}C_i\left(R - \sum_{i=1}^{m} B_i\right) - k_{di}B_i \qquad (9)$$

for $i = 1, \ldots, m$. Note that the different binding species are assumed to differ only in their binding characteristics. The inflows, $I_i(t)$, are given by (5) and the output is once again given by (6).

The heterogeneous analyte version of the model is given by the following:

$$h\dot{C} = \sum_{i=1}^{m}\left[-k_{ai}C\left[R - \sum_{i=1}^{m} B_i\right] + k_{di}B_i\right] + k_M(I - C)$$
$$\dot{B_i} = k_{ai}C\left(R - \sum_{i=1}^{m} B_i\right) - k_{di}B_i \qquad (10)$$

for $i = 1, \ldots, m$. Again note that the different binding species are assumed to differ only in their binding characteristics. The inflow, $I(t)$, is given by (2) and the output is once again given by (6).

2.3 Effective rate constant approximation

Applying a quasi-steady state assumption to (8) consists of solving $\dot{C}(t) = 0$ for $C(t)$ to yield

$$C(t) = \frac{k_dB(t) + k_MI(t)}{k_a(R - B(t)) + k_M}$$

and then substituting this into the equation for $\dot{B}(t)$ to give

$$\dot{B}(t) = \frac{k_aI(t)(R - B(t)) - k_dB(t)}{1 + (k_a/k_M)(R - B(t))}. \qquad (11)$$

The inflow, $I(t)$, and output, $y(t)$ for the quasi-steady state version of the model are still given by (2) and (3). Equation (11) is the effective rate constant approximation derived by Edwards (2001), who showed that this equation gives a good approximation to a full fluid dynamics model up to $O(\text{Da}^2)$ where

$$k_aR/k_M = \text{Da } h_d.$$

Da is the *Damköhler number*, which is the ratio of the reaction velocity to diffusion velocity in the diffusive boundary layer, and h_d is a positive constant that incorporates effects of the receptor layer (see Edwards (2001) for details).

Notice from (11), as discussed by Myszka et al. (1998), that the parameter h has been eliminated from the model. Therefore in the (quasi) steady state situation this parameter has no bearing on the observation.

The heterogeneous mixture version of the model is given by the following:

$$\dot{B_i}(t) = \frac{k_{ai}I_i(t)\left[R - \sum_{i=1}^{m} B_i(t)\right] - k_{di}B_i(t)}{1 + (k_{ai}/k_M)\left[R - \sum_{i=1}^{m} B_i(t)\right]} \qquad (12)$$

for $i = 1, \ldots, m$. The inflows, $I_i(t)$, and output, $y(t)$, are once again given by (5) and (6).

The heterogeneous analyte version of the model is given by the following:

$$\dot{B_i}(t) = \frac{k_{ai}I(t)\left[R - \sum_{i=1}^{m} B_i(t)\right] - k_{di}B_i(t)}{1 + (k_{ai}/k_M)\left[R - \sum_{i=1}^{m} B_i(t)\right]} \qquad (13)$$

for $i = 1, \ldots, m$. The inflow, $I(t)$, and output, $y(t)$, are once again given by (2) and (6).

3 Structural identifiability

In this section the uniqueness of the unknown parameters in a general nonlinear model is considered with respect to the (single) output. More precisely, let $\boldsymbol{p} \in \Omega \subset \mathbb{R}^r$ denote a vector comprising the unknown parameters in the model, which belongs to an open set of admissible vectors. To make the parameter dependence of the model output more explicit it is written $y(t, \boldsymbol{p})$.

Two parameter vectors $\boldsymbol{p}, \overline{\boldsymbol{p}} \in \Omega$ are *indistinguishable*, written $\boldsymbol{p} \sim \overline{\boldsymbol{p}}$, if they give rise to identical outputs:

$$y(t, \boldsymbol{p}) = y(t, \overline{\boldsymbol{p}}) \qquad \text{for all } t \geq 0.$$

For generic $\boldsymbol{p} \in \Omega$, the parameter p_i is *locally identifiable* if there is a neighbourhood, N, of $\boldsymbol{p}$ such that

$$\overline{\boldsymbol{p}} \in N, \;\; \boldsymbol{p} \sim \overline{\boldsymbol{p}} \quad \text{implies that} \quad \overline{p}_i = p_i.$$

In particular, if $N = \Omega$ in the above definition then p_i is *globally identifiable*, otherwise it is *nonuniquely (locally) identifiable*. Notice that, for a given output, a locally identifiable parameter can take any of a distinct (countable) set of values. If there does not exist a suitable neighbourhood N then p_i is *unidentifiable* and, for a given output, can take an (uncountably) infinite set of values.

A system model is *structurally globally identifiable* (SGI) if all parameters are globally identifiable; it is *structurally locally identifiable* (SLI) if all parameters are locally identifiable and at least one is nonuniquely identifiable; and the model is *structurally unidentifiable* (SU) if at least one parameter is unidentifiable.

Following a similar approach to that taken by Denis-Vidal et al. (2001), suppose that a rational equation can be generated for the output that only involves $y(t, \boldsymbol{p})$ and its derivatives. Then any indistinguishable parameter vector, $\overline{\boldsymbol{p}}$, must also satisfy the same equation with $\boldsymbol{p}$ replaced by $\overline{\boldsymbol{p}}$. Therefore, for $y(t, \boldsymbol{p})$ and its derivatives, a polynomial of the following form can be generated:

$$\sum_{k=1}^{l} c_k(\boldsymbol{p}, \overline{\boldsymbol{p}}) \phi_k(y(t, \boldsymbol{p}), y'(t, \boldsymbol{p}), y''(t, \boldsymbol{p}), \dots) = 0 \quad (14)$$

for all $t \geq 0$, where $\phi_k(y(t, \boldsymbol{p}), y'(t, \boldsymbol{p}), y''(t, \boldsymbol{p}), \dots)$ is a monomial in $y(t, \boldsymbol{p})$ and its derivatives. If the $\phi_k(\cdot)$ are linearly independent then it must be the case that

$$c_k(\boldsymbol{p}, \overline{\boldsymbol{p}}) = 0 \qquad k = 1, \dots, l$$

and so the relationship between $\overline{\boldsymbol{p}}$ and $\boldsymbol{p}$ can be determined. If the only solution is $\overline{\boldsymbol{p}} = \boldsymbol{p}$ then the model is SGI, if there is a set of distinct solutions then the model is SLI and it is SU otherwise.

3.1 Observable Normal Form

Suppose that the model description for the system is of the following state space form:

$$\begin{aligned}
\dot{\boldsymbol{x}}(t, \boldsymbol{p}) &= \boldsymbol{f}(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}), \quad \boldsymbol{x}(0, \boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}), \\
y(t, \boldsymbol{p}) &= h(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}),
\end{aligned} \quad (15)$$

where $\boldsymbol{p} \in \Omega$, $\boldsymbol{f}(\cdot, \cdot)$ and $h(\cdot, \cdot)$ are rational functions (i.e., fractions of polynomials) in both $\boldsymbol{x}$ and $\boldsymbol{p}$, and the initial condition $\boldsymbol{x}_0(\cdot)$ is a rational function in $\boldsymbol{p}$. For all $\boldsymbol{p} \in \Omega$, denote by $M(\boldsymbol{p})$ the largest connected open subset of $\mathbb{R}^n$ containing $\boldsymbol{x}_0(\boldsymbol{p})$ such that both $\boldsymbol{f}(\cdot, \boldsymbol{p})$ and $h(\cdot, \boldsymbol{p})$ are well-defined on $M(\boldsymbol{p})$.

For $\lambda \in C^\infty(M(\boldsymbol{p}))$, the Lie derivative of λ along the vector field $\boldsymbol{f^p} = \boldsymbol{f}(\cdot, \boldsymbol{p})$ is the smooth function given by

$$\mathrm{L}_{\boldsymbol{f^p}} \lambda(\boldsymbol{x}) = \frac{\partial \lambda}{\partial \boldsymbol{x}}(\boldsymbol{x}) \boldsymbol{f^p}(\boldsymbol{x}).$$

Define a vector field $\boldsymbol{H_p}(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}, \boldsymbol{p}), \dots, \mu_n(\boldsymbol{x}, \boldsymbol{p}))^{\mathrm{T}}$, where

$$\begin{aligned}
\mu_1(\boldsymbol{x}, \boldsymbol{p}) &= h(\boldsymbol{x}, \boldsymbol{p}) \\
\mu_{i+1}(\boldsymbol{x}, \boldsymbol{p}) &= \mathrm{L}_{\boldsymbol{f^p}} \mu_i(\boldsymbol{x}, \boldsymbol{p}) \qquad i = 1, \dots, n.
\end{aligned}$$

If the Jacobian matrix of $\boldsymbol{H_p}$ (with respect to $\boldsymbol{x}$), evaluated at $\boldsymbol{x}_0(\boldsymbol{p})$, has full rank then (15) satisfies the *Observability Rank Condition* (ORC) and is locally weakly observable (Hermann and Krener, 1977). In particular, $\boldsymbol{z} = \boldsymbol{H_p}(\boldsymbol{x})$ is a coordinate transformation into the Observable Normal Form:

$$\begin{aligned}
\dot{z}_i(t, \boldsymbol{p}) &= z_{i+1}(t, \boldsymbol{p}) \qquad i = 1, \dots, n-1 \\
\dot{z}_n(t, \boldsymbol{p}) &= \mu_{n+1}\left(\boldsymbol{H_p}^{-1}(\boldsymbol{z}(t, \boldsymbol{p})), \boldsymbol{p}\right) \qquad (16) \\
y(t, \boldsymbol{p}) &= z_1(t, \boldsymbol{p}).
\end{aligned}$$

The components of the state, $z_i(t, \boldsymbol{p})$, correspond to the output, $y(t, \boldsymbol{p})$, or its derivatives, and the final state equation gives an output equation. Thus the output satisfies an nth order (ordinary) differential equation (with suitable initial conditions).

For two indistinguishable parameter vectors, $\boldsymbol{p} \sim \overline{\boldsymbol{p}}$, the required polynomial output equation (14) can be constructed from $\dot{z}_n(t, \boldsymbol{p}) - \dot{z}_n(t, \overline{\boldsymbol{p}})$. Note that the monomials are in the output and its derivatives up to order $n-1$ ($\phi_k(y(t, \boldsymbol{p}), y'(t, \boldsymbol{p}), \dots, y^{(n-1)}(t, \boldsymbol{p}))$). These monomial terms come from the right-hand side of the equation for $\dot{z}_n$. Thus if the monomials were not linearly independent then the output would also satisfy an $(n-1)^{\text{th}}$ order differential equation (with the same initial conditions)—a contradiction, unless the particular system initial condition gives rise to redundancy in the generating equations ($\mu_i(\boldsymbol{x}, \boldsymbol{p})$ for $i = 1, \dots, n+1$).

The output equation with associated initial conditions uniquely determines the solution for the output of the original state space model. Thus indistinguishability of two parameter vectors for a given model requires the solution of the coefficient terms, $c_k(\boldsymbol{p}, \overline{\boldsymbol{p}}) = 0$, and the corresponding initial conditions:

$$y(0, \boldsymbol{p}) = y(0, \overline{\boldsymbol{p}}), \quad y^{(k)}(0, \boldsymbol{p}) = y^{(k)}(0, \overline{\boldsymbol{p}})$$

for $k = 1, \ldots, n-1$.

3.2 Algebraic approach

Notice that the state space system in (15) can be rewritten in the form

$$\begin{aligned} \dot{\boldsymbol{x}}(t, \boldsymbol{p}) - \boldsymbol{f}(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}) &= 0, \\ y(t, \boldsymbol{p}) - h(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}) &= 0, \end{aligned} \tag{17}$$

with $\boldsymbol{x}(0, \boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p})$. Since $\boldsymbol{f}$ and h are rational the left-hand sides of (17) can be rewritten as polynomials and the system considered to correspond to the differential ideal generated by these polynomials. A differential polynomial in y (and its derivatives) only in this ideal corresponds to an output equation for the system.

Glad (1991) used (differential) characteristic sets to derive input-output representations for polynomial state space systems, while Forsman (1991) considered the use of Gröbner Bases for the ideal generated by successive Lie derivatives of $h(\cdot, \boldsymbol{p})$ along $\boldsymbol{f}(\cdot, \boldsymbol{p})$, thus providing an algebraic method for deriving the Observable Normal Form.

Consider the following list of rational functions (i.e., terms have been collected and written over a common denominator):

$$y(t, \boldsymbol{p}) - h(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}), \quad \dot{y}(t, \boldsymbol{p}) - \mu_2(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}),$$
$$\ldots \quad , y^{(n)}(t, \boldsymbol{p}) - \mu_{n+1}(\boldsymbol{x}(t, \boldsymbol{p}), \boldsymbol{p}).$$

Let $\mathcal{L}^n$ denote the ideal generated by the numerators of these functions. Then a Gröbner basis for $\mathcal{L}^n$ with a ranking in which the states, x_i, are higher than the output, y, contains an output relation for (17). Let $\mathcal{I}_j$ denote the contraction of the ideal $\mathcal{L}^j$ to the polynomial ring in the indeterminates $y, \dot{y}, \ldots, y^{(j)}$. If n is the smallest value for which $\mathcal{I}_n \neq 0$ then $\mathcal{I}_n$ is a principal ideal (see Theorem 5.3 in Forsman (1991)). In this case there are no other fundamentally different equations of minimal order describing the output relation.

Forsman (1991) showed that a polynomial system satisfying the ORC for generic initial conditions is algebraically observable; and it is known that a system in state space form of dimension n is algebraically observable if and only if there is no output equation of order fewer than n in y (Glad, 1990). As seen previously, if the system satisfies the ORC then the output is the unique solution of a differential equation of order n, and this is the generator of the required principal ideal. However, since the identifiability analysis is performed with respect to particular initial conditions it must first be determined that the generating equations for the ideal $\mathcal{L}^n$ do not change when $\boldsymbol{x}$ is restricted to the submanifold described by the solution, $\boldsymbol{x}(t, \boldsymbol{p})$, of the system equations starting at $\boldsymbol{x}_0(\boldsymbol{p})$.

Saccomani et al. (2003) considered the problem of particular initial conditions for controlled systems that may give rise to a change in the characteristic set defining the system. For these types of systems it was shown that an accessibility condition ensures that the characteristic set does not change, and hence an exhaustive summary can be determined as for generic initial conditions.

Forsman (1991) provides three ways in which Gröbner Bases can be used to determine the output equation (that he implemented in MAPLE):

(1) Find the contraction of $\mathcal{L}^n$ to the polynomial ring in the indeterminate $y^{(n)}$ with coefficients in a field extension that includes $y, \dot{y}, \ldots, y^{(n-1)}$.
(2) Compute a Gröbner basis for $\mathcal{L}^n$ with a ranking in which the states, x_i, are higher than $y^{(n)}$, and $y^{(k)}$ is ranked higher than $y^{(k-1)}$ for $k = 1, \ldots, n$.
(3) Compute a Gröbner basis for $\mathcal{L}^n$ with a ranking in which the states, x_i, are higher than $y^{(n)}$, and the remaining $y^{(k)}$ are treated as parameters.

3.3 Combining experiments

The ProteOn XPR36 (Bio-Rad) SPR platform enables up to 36 simultaneous experiments (interactions) to be performed in a single physical experiment, corresponding to a maximum of six analytes and six ligands. Parameters can vary across both ligand and analyte, or across only one of these, or be constant across all lanes and channels. For an individual interaction between ligand i and analyte j partition the vector of unknown parameters as follows:

$$\boldsymbol{p} = \left(\boldsymbol{p}_0^T, \boldsymbol{p}_i^T, \boldsymbol{p}_j^T, \boldsymbol{p}_{ij}^T \right)^T$$

where $\boldsymbol{p}_0$ is the vector of parameters constant across all ligands and analytes; $\boldsymbol{p}_i$ is the vector of parameters constant across the analytes; $\boldsymbol{p}_j$ is the vector of parameters constant across the ligands; and $\boldsymbol{p}_{ij}$ are the parameters that vary across all ligands and analytes.

A structural identifiability analysis of an individual interaction, between ligand i and analyte j, gives rise to

a set of relations between indistinguishable parameter vectors:
$$\mathcal{S}_{ij}(\boldsymbol{p}) = \{\overline{\boldsymbol{p}} \in \Omega : \overline{\boldsymbol{p}} \sim \boldsymbol{p}\}.$$
Thus to combine individual analyses $\overline{\boldsymbol{p}}_0$ must be the same across all $\mathcal{S}_{ij}$; $\overline{\boldsymbol{p}}_i$ must be the same across $\mathcal{S}_{ij}$, $j = 1, \ldots, 6$; and $\overline{\boldsymbol{p}}_j$ must be the same across $\mathcal{S}_{ij}$, $i = 1, \ldots, 6$. These constraints may give rise to additional relations between the parameters for the global experiment.

4 Results

4.1 Langmuir model

For the Langmuir model (1) there are four unknown parameters, $\boldsymbol{p} = (k_a, k_d, C_T, R)^T$, for the single model output. The output equation is given by (see Appendix A.2)
$$\dot{y}(t, \boldsymbol{p}) + (k_a C_T + k_d) \, y(t, \boldsymbol{p}) - \alpha k_a C_T R = 0$$
in the association phase and
$$\dot{y}(t, \boldsymbol{p}) + k_d y(t, \boldsymbol{p}) = 0$$
in the dissociation phase. From the latter equation it is immediately evident that the dissociation rate constant k_d is globally identifiable.

Now considering the association phase equation, if $\boldsymbol{p}$ is generic and $\overline{\boldsymbol{p}}$ denotes an indistinguishable parameter vector, then the following relations must hold:
$$\overline{k}_a \overline{C}_T = k_a C_T, \quad \overline{R} = R, \quad \text{and} \quad \overline{k}_d = k_d.$$
Therefore the parameters k_d and R are globally identifiable, but the parameters k_a and C_T are unidentifiable (and hence the model is SU). An important consequence of this result is that the binding affinity $A = k_d/k_a$ is unidentifiable (i.e., not uniquely determined by the model output). However, if the sample concentration, C_T, is known then $\overline{C}_T = C_T$ and so k_a becomes globally identifiable and the model is SGI.

Considering initial conditions for the output equations does not provide any further relations between $\boldsymbol{p}$ and $\overline{\boldsymbol{p}}$. Even with multiple experiments taking place on the same chip and combining results the model does not become SGI, unless the sample concentration is known *a priori*.

For the heterogeneous mixture version of the Langmuir model (4) ($m = 2$) there are 7 unknown parameters,
$$\boldsymbol{p} = (k_{a1}, k_{a2}, k_{d1}, k_{d2}, C_{T1}, C_{T2}, R)^T,$$
for the single model output. For parameter vectors $\boldsymbol{p}$ and $\overline{\boldsymbol{p}}$ to give the same output it must be the case that (see Appendix A.3)
$$\overline{R} = R, \quad \overline{k}_{d1} = k_{d1}, \quad \overline{k}_{d2} = k_{d2}, \quad \overline{C}_{T1}\overline{k}_{a1} = C_{T1}k_{a1},$$
$$\text{and} \quad \overline{C}_{T2}\overline{k}_{a2} = C_{T2}k_{a2}$$
or
$$\overline{R} = R, \quad \overline{k}_{d1} = k_{d2}, \quad \overline{k}_{d2} = k_{d1}, \quad \overline{C}_{T1}\overline{k}_{a1} = C_{T2}k_{a2},$$
$$\text{and} \quad \overline{C}_{T2}\overline{k}_{a2} = C_{T1}k_{a1}.$$

Therefore the parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable, and the parameters C_{T1}, C_{T2}, k_{a1} and k_{a2} are unidentifiable (but the combinations $C_{T1}k_{a1}$ and $C_{T2}k_{a2}$ are locally identifiable). Hence the model is SU. No further information is provided by the initial conditions.

If the total sample concentration is known, so that in addition it is known that $\overline{C}_{T1} + \overline{C}_{T2} = C_{T1} + C_{T2}$, the model remains structurally unidentifiable. However, if the individual concentrations, C_{T1} and C_{T2} are known, then the model is structurally locally identifiable. The main implication of these findings is that the binding affinities $A_i = k_{di}/k_{ai}$ are also unidentifiable, unless both concentrations, C_{T1} and C_{T2} are known.

For the heterogeneous analyte version of the Langmuir model (7) ($m = 2$) there are 6 unknown parameters,
$$\boldsymbol{p} = (k_{a1}, k_{a2}, k_{d1}, k_{d2}, C_T, R)^T,$$
for the single model output. For parameter vectors $\boldsymbol{p}$ and $\overline{\boldsymbol{p}}$ to give the same output it must be the case that (see Appendix A.4)
$$\overline{R} = R, \ \overline{k}_{d1} = k_{d1}, \ \overline{k}_{d2} = k_{d2}, \ \overline{C}_T\overline{k}_{a1} = C_T k_{a1},$$
$$\text{and} \quad \overline{C}_T\overline{k}_{a2} = C_T k_{a2},$$
or
$$\overline{R} = R, \ \overline{k}_{d1} = k_{d2}, \ \overline{k}_{d2} = k_{d1}, \ \overline{C}_T\overline{k}_{a1} = C_T k_{a2},$$
$$\text{and} \quad \overline{C}_T\overline{k}_{a2} = C_T k_{a1}$$

Therefore parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable, and the parameters C_T, k_{a1} and k_{a2} are unidentifiable (but the combinations $C_T k_{a1}$ and $C_T k_{a2}$ are locally identifiable). Hence the model is SU.

If the sample concentration is known, so that in addition it is known that $\overline{C}_T = C_T$, the model becomes SLI. The binding affinities $A_i = k_{di}/k_{ai}$ are locally identifiable.

4.2 Langmuir with transport model

The same structural identifiability approach was applied to the three forms of the Langmuir with transport model.

For the homogeneous version of the Langmuir with mass transport model (8) there are 6 unknown parameters (k_a, k_d, C_T, R, k_M, and h). The parameters R and k_d are globally identifiable while the parameters C_T, k_a, h, and k_M are unidentifiable. Hence the model is SU.

If the sample concentration, C_T, were not known *a priori* then it would not be possible to uniquely determine it or the binding affinity, $A = k_d/k_a$. However, if the sample concentration is known *a priori*, so that $\overline{C}_T = C_T$, then all of the remaining rate constants are globally identifiable and the model becomes SGI.

For the heterogeneous mixture ($m = 2$) version of the Langmuir with transport model (9) there are 9 unknown parameters (k_{ai}, k_{di}, C_{Ti}, R, k_M, and h for $i = 1, 2$). The parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable and the remaining parameters, C_{T1}, C_{T2}, k_{a1}, k_{a2}, h, and k_M, are unidentifiable. Hence the model is SU.

If the total sample concentration is known, so that in addition it is known that $\overline{C}_{T1} + \overline{C}_{T2} = C_{T1} + C_{T2}$, the model becomes SLI, with R and k_M being globally identifiable and the remaining parameters being locally (nonuniquely) identifiable. However, if the individual concentrations, C_{T1} and C_{T2} are known, then the model is SGI. Therefore, the binding affinities $A_i = k_{di}/k_{ai}$ are also unidentifiable, unless the total sample concentration, $C_{T1} + C_{T2}$ is known, when they become locally identifiable.

For the heterogeneous analyte ($m = 2$) version of the Langmuir with transport model (10) there are 8 unknown parameters (k_{ai}, k_{di}, C_T, R, k_M, and h for $i = 1, 2$). The parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable and the remaining parameters, C_T, k_{a1}, k_{a2}, h, and k_M, are unidentifiable. Hence the model is SU.

If the sample concentration is known ($\overline{C}_T = C_T$) then the model becomes SLI. The binding affinities are unidentifiable if the sample concentration is not known and locally identifiable if it is known.

4.3 ERC model

The same structural identifiability approach was also applied to the three forms of the ERC model. For the homogeneous form (11) there are 5 unknown parameters (k_a, k_d, C_T, R, and k_M). The parameters R and k_d are globally identifiable, but the individual parameters C_T, k_a, and k_M are all unidentifiable. Hence the model is structurally unidentifiable and it is not possible to estimate the unknown concentration, C_T, or the binding affinity, $A = k_d/k_a$, reliably. If the sample concentration, C_T, is known *a priori*, then all of the rate constants are

globally identifiable and the model becomes structurally globally identifiable.

For the heterogeneous mixture version of the ERC model (12) there are 8 unknown parameters (k_{ai}, k_{di}, C_{Ti}, R, and k_M for $i = 1, 2$). The parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable and the remaining parameters, C_{T1}, C_{T2}, k_{a1}, k_{a2}, and k_M, are unidentifiable. Hence the model is SU.

If the total sample concentration is known, so that in addition it is known that $\overline{C}_{T1} + \overline{C}_{T2} = C_{T1} + C_{T2}$, the model becomes SLI, with R and k_M being globally identifiable and the remaining parameters being locally (nonuniquely) identifiable. However, if the individual concentrations, C_{T1} and C_{T2} are known, then the model is SGI. The main implication of these findings is that the binding affinities $A_i = k_{di}/k_{ai}$ are also unidentifiable, unless the total sample concentration, $C_{T1} + C_{T2}$ is known, when they become locally identifiable.

For the heterogeneous analyte version of the ERC model (13) there are 7 unknown parameters (k_{ai}, k_{di}, C_T, R, and k_M for $i = 1, 2$). The parameter R is globally identifiable, k_{d1} and k_{d2} are locally identifiable and the remaining parameters, C_T, k_{a1}, k_{a2}, and k_M, are unidentifiable. Hence the model is SU.

If the sample concentration is known ($\overline{C}_T = C_T$) the model becomes SLI, with R and k_M being globally identifiable and the remaining parameters being locally (nonuniquely) identifiable. The binding affinities $A_i = k_{di}/k_{ai}$ are also unidentifiable, unless the sample concentration is known, when they become locally identifiable.

4.4 Summary

A summary of the identifiability analysis results from the previous sections is presented in Table 1. It is noticeable from these results that the heterogeneous mixture version of the standard Langmuir model is "more unidentifiable" than the other candidate structures (more precisely, in the identifiability analysis there is a two dimensional manifold of solutions for $\overline{p}$ such that $p \sim \overline{p}$, compared with one dimensional manifolds for the other unidentifiable cases). It is also clear from the structural identifiability analyses that it is necessary to be able to determine the total sample concentration in order to obtain identifiable models. When the total sample concentration is known then the associated binding affinities are locally identifiable (uniquely in the homogeneous case and nonuniquely in the heterogeneous cases), except for the Langmuir model, but this just corresponds to interchanging of the indices $i = 1$ and $i = 2$.

Table 1
Summary of the structural identifiability of all models considered (SGI: structurally globally identifiable; SLI: structurally locally identifiable; SU: structurally unidentifiable). In addition, the identifiability of the binding affinities is also shown (GI: globally identifiable; LI: locally identifiable; U: unidentifiable).

Model	C_T unknown	C_T (only) known	C_{T1}, C_{T2} known
Langmuir (homogeneous analyte)	SU (A:U)	SGI (A:GI)	-
Langmuir (mixed analyte)	SU (A_i:U)	SU (A_i:U)	SLI (A_i:LI)
Langmuir (heterogeneous analyte)	SU (A_i:U)	SLI (A_i:LI)	-
Transport (homogeneous analyte)	SU (A:U)	SGI (A:GI)	-
Transport (mixed analyte)	SU (A_i:U)	SLI (A_i:LI)	SGI (A_i:GI)
Transport (heterogeneous analyte)	SU (A_i:U)	SLI (A_i:LI)	-
ERC (homogeneous analyte)	SU (A:U)	SGI (A:GI)	-
ERC (mixed analyte)	SU (A_i:U)	SLI (A_i:LI)	SGI (A_i:GI)
ERC (heterogeneous analyte)	SU (A_i:U)	SLI (A_i:LI)	-

5   Conclusions

Three models typically used to estimate the kinetic rate constants in binding reactions from SPR experiments have been considered. The first model is the well mixed case and consists of Langmuir binding kinetics. The second model is a two-compartment system proposed by Myszka et al. (1998) that includes mass transport effects in a Langmuir type reaction scheme. This model gives rise to the third model, the effective rate constant approximation, under a quasi-steady state assumption. The third model has been previously shown to be a good approximation to a full fluid dynamics model for low Damköhler number (Edwards, 2001).

The Langmuir model has been shown to be structurally unidentifiable unless the total sample concentration, C_T, is known, in which case the homogeneous form of the model becomes structurally globally identifiable and the heterogeneous analyte form becomes structurally locally identifiable. However, the heterogeneous mixture form of the model remains unidentifiable unless the concentrations of the individual components of the analyte are known, in which case the model only becomes structurally locally identifiable. Thus if the analyte exhibits heterogeneity then the binding affinities are at best, when the sample concentration is known, locally (nonuniquely) identifiable.

The full two-compartment transport model has been shown to be structurally globally identifiable in the homogeneous case and structurally locally identifiable in the heterogeneous cases provided the total sample concentration, C_T, is known. Otherwise the model is unidentifiable, in which case certain parameters can not be uniquely determined from the output (and hence ultimately from sensorgrams). These unidentifiable parameters include the sample concentration and the association rate constant meaning that only a concentration corrected binding affinity, $k_d/(k_a C_T)$, can be uniquely

determined. Similar results have been produced for the effective rate constant approximation.

Myszka et al. (1998) report that for parameters relevant to Biacore experiments the solution to the two-compartment transport model is insensitive to the value of h. It has been confirmed that the model is structurally unidentifiable even with both association and dissociation phases considered and that h is one of the unidentifiable parameters. Therefore, the output of the model to be compared with experimental data is independent of the value of h, provided the product hC_T remains fixed in the homogeneous and heterogeneous analyte cases, and hC_{Ti} in the mixture case (with C_{Ti} interchangeable). In particular, it is necessary to know the sample concentration, C_T, or one of the other unidentifiable parameters (h, k_a, and k_M) for the unknown parameters to be uniquely determined by the output. Therefore assigning an arbitrary value to h (as suggested by Myszka et al. (1998)) makes the model artificially structurally globally identifiable.

The results obtained in this paper now enable the investigation of antibody-antigen binding kinetics using SPR and characterises the uniqueness of the estimates for binding affinity when applying one of the key model structures. This is the first time that a formal analysis of the uniqueness of these estimates has been performed for SPR experiments.

References

D.G. Myszka, X. He, M. Dembo, T.A. Morton, and B. Goldstein.   Extending the range of rate constants available from BIACORE: Interpreting mass transport-influenced binding data. *Biophys. J.*, 75: 583–594, 1998.

D.A. Edwards, B. Goldstein, and D.S. Cohen. Transport effects on surface-volume biological reactions:. *J. Math. Biol.*, 39:533–561, 1999.

D.A. Edwards. The effect of a receptor layer on the measurement of rate constants. *B. Math. Biol.*, 63: 301–327, 2001.

I. Chaiken, S. Roseé, and R. Karlsson. Analysis of macromolecularinteractions using immobilzed ligands. *Anal. Biochem.*, 201:197–210, 1992.

R. Bellman and K.J. Åström. On structural identifiability. *Mathematical Biosciences*, 7:329–339, 1970.

J.A. Jacquez. *Compartmental Analysis in Biology and Medicine.* BioMedware, Ann Arbor, MI, 1996.

K.R. Godfrey and J.J. DiStefano III. Identifiability of model parameters. In E. Walter, editor, *Identifiability of Parametric Models*, chapter 1, pages 1–20. Pergamon Press, 1987.

E. Walter. *Identifiability of State Space Models.* Springer-Verlag, 1982.

H. Pohjanpalo. System identifiability based on the power series expansion of the solution. *Math. Biosci.*, 41: 21–33, 1978.

G. Margaria, E. Riccomagno, M.J. Chappell, and H.P. Wynn. Differential algebra methods for the study of the structural identifiability of rational polynomial state-space models in the biosciences. *Mathematical Biosciences*, 174:1–26, 2001.

N.D. Evans, M.J. Chapman, M.J. Chappell, and K.R. Godfrey. Identifiability of uncontrolled nonlinear rational systems. *Automatica*, 38:1799–1805, 2002.

L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30:265–276, 1994.

M.P. Saccomani, S. Audoly, and L. D'Angiò. Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, 39:619–632, 2003.

J. Němcová. Structural identifiability of polynomial and rational systems. *Math. Biosci.*, 223:83–96, 2010.

L. Denis-Vidal, G. Joly-Blanchard, and C. Noiret. Some effective approachesto check identifiability of uncontrolled nonlinear rational systems. *Math. Comput. Simulat.*, 57:35–44, 2001.

A. Isidori. *Nonlinear Control Systems.* Springer, 1995.

F. Ollivier. *Le probléme de lidentifiabilité structurelle globale: étude théorique, méthodes effectives et bornes de complexité.* Thèse de doctorat en science, Ecole polytechnique, Paris, France, 1990.

S.T. Glad and L. Ljung. Model structure identifiability and persistence of excitation. In *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 3236–3240, 1990.

S. Audoly, G. Bellu, L. D'Angiò, M.P. Saccomani, and C. Cobelli. Global identifiability of nonlinear models of biological systems. *IEEE Transactions on Biomedical Engineering*, 48:55–65, 2001.

S.T. Glad. Nonlinear input output relations and identifiability. In *Proceedings of the 31st Conference on Decision and Control*, 1992.

S.T. Glad. Input output representations from characteristic sets. In *Proceedlngs of the 30th Conference on Decision and Control*, pages 726–730, Brighton, England, December 1991.

M.P. Saccomani, S. Audoly, G. Bellu, and L. D'Angiò. A new differential algebra algorithm to test identifiability of nonlinear systems with given initial conditions. In *Proceedings of the 40th IEEE Conference on Decision and Control*, 2001.

D. Bearup, N.D. Evans, and M.J. Chappell. The input-output relationship approach to structural identifiability analysis. Submitted to Computer Methods and Programs in Biomedicine, 2011.

K. Forsman. *Constructive Commutative Algebra in Nonlinear Control Theory.* PhD thesis, Linköping Institute of Technology, 1991. Number 261.

T. Andersson. *Concepts and Algorithms for Non-linear System Identifiability.* PhD thesis, Linköping University, 1994. Number 448.

R. Hermann and A.J. Krener. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, AC-22:728–740, 1977.

S.T. Glad. Differential algebraic modelling of nonlinear systems. In MA Kaashoek, JH van Schuppen, and ACM Ran, editors, *Realization and Modelling in System Theory, Proceedings of the International Symposium MTNS-89*, volume I, pages 97–105. Birkhäuser, 1990.

A   Structural identifiability analyses

Due to the computational complexities involved in performing a structural identifiability analysis a symbolic algebra system such as MAPLE (Waterloo Maple Inc) or MATHEMATICA (Wolfram Research) is an invaluable tool. All analyses in this section were performed using MAPLE (version 15.01) on a dual core (2×2.66 GHz) Windows XP PC with 4 GB memory. Results were confirmed using MATHEMATICA (version 8.01) by generating the Observable Normal Form of the model.

A.1   Generating output equation in MAPLE

Based on the MAPLE code proposed by Forsman (1991) the following procedures are used to determine the output equation for state space systems. The procedures use short command names and so the following packages are loaded:

```
with(LinearAlgebra): with(Groebner):
```

The following procedure, lieDer, determines the Lie derivative of H along F:

```
lieDer := proc (H, F)
 local N, V, vars;
 N := nops(F);
 vars := [seq(x[t], t = 1 .. N)]:
 V:=map( (a, b) -> diff(b, a), vars, H):
 DotProduct(Vector(F),Vector(V),conjugate=false)
end:
```

while the next procedure, `listLieDer`, determines the list of the first `N` Lie derivatives of `H` along `F`:

```
listLieDer := proc (H, F, N)
 local L, i, tmp;
 L := [y[0]-H];
 tmp := H;
 for i to N do
  tmp := lieDer(tmp, F);
  L := [op(L), y[i]-tmp]
 od;
end:
```

The three approaches to determining the output equation using Gröbner Bases proposed by Forsman (1991) are implemented by the following three procedures:

```
outptEqn1 := proc (F, H)
 local N, L, vars;
 N := nops(F);
 L := listLieDer(H, F, N);
 L := map(expand, L);
 vars := [seq(x[t], t = 1 .. N), y[N]];
 UnivariatePolynomial(y[N], L, vars)
end:
outptEqn2 := proc (F, H)
 local N, G, L, vars;
 N := nops(F);
 L := listLieDer(H, F, N);
 vars:=[seq(x[t],t=1..N),y[N-i]$i=0..N];
 G := Basis(L, plex(op(vars)));
 G[1]
end:
outptEqn3 := proc (F, H)
 local N, G, L, vars;
 N := nops(F);
 L := listLieDer(H, F, N);
 vars := [seq(x[t], t = 1 .. N), y[N]];
 G := Basis(L, plex(op(vars)));
 G[1]
end:
```

A.2 Langmuir (Homogeneous analyte)

The identifiability analysis is performed by first defining the model and determining the output equation:

```
F := [ k[a]*C[T]*(R-x[1])-k[d]*x[1] ] :
H := alpha*x[1]:
out := outptEqn1( F , H );
```

In this case the first method proposed by Forsman (1991) has been used (the other methods generate the same output equation). Next the coefficients of the output equation are extracted, though note that the coefficient of the highest derivative term should be 1 (for polynomial examples):

```
p := collect(out,[y[0],y[1]],'distributed'):
uA := { coeffs(p,[y[0],y[1]]) };
outD := eval(out,C[T]=0):
p := collect(outD,[y[0],y[1]],'distributed'):
uD := { coeffs(p,[y[0],y[1]]) };
u := union( uA , uD );
```

The coefficients for a single interaction correspond to those (`uA`) of the output equation in the association phase (`out`) and those (`uD`) of the output equation in the dissociation phase (`outD`), which is obtained by setting $C_T = 0$.

Now suppose that there exists an indistinguishable parameter vector with elements `kb[d]`, `kb[a]`, `Cb[T]`, and `Rb`. Since it is indistinguishable from the original parameter vector it must give rise to the same output equation, but with the coefficients in terms of the new vector. However, these coefficients are unique for a given output and this is used to determine the relations between the elements of the two parameter vectors:

```
ub:=eval(u,
    [k[d]=kb[d],k[a]=kb[a],C[T]=Cb[T],R=Rb]);
eqns := convert(u, list)-convert(ub, list);
solve(eqns, [kb[d], kb[a], Cb[T], Rb]);
```

Since this example is 1 dimensional and the initial condition for the association phase does not correspond to a steady state the results hold for the particular initial condition of interest. The initial condition for the dissociation phase is an unknown parameter and therefore generic.

A.3 Langmuir (Heterogeneous mixture)

For the heterogeneous mixture version of the Langmuir model (4) the model is written as follows:

```
F := [ k[a1]*C[T1]*(R-x[1]-x[2])-k[d1]*x[1],
    k[a2]*C[T2]*(R-x[1]-x[2])-k[d2]*x[2] ] :
H := alpha*(x[1]+x[2]):
out := outptEqn2( F , H );
```

In this case the second method proposed by Forsman (1991) has been used (the other methods generate the same output equation). The output equation is second order and given by

$$
\begin{aligned}
y_2 &+ (k_{d1} + k_{a1}C_{T1} + k_{a2}C_{T2} + k_{d2}) \, y_1 \\
&+ (k_{d2}k_{a1}C_{T1} + k_{d1}k_{a2}C_{T2} + k_{d1}k_{d2}) \, y_0 \\
&- \alpha \, (k_{d1}k_{a2}C_{T2}R + k_{d2}k_{a1}C_{T1}R) = 0
\end{aligned}
$$

where $y_0 = y(t, \boldsymbol{p})$ and $y_i = y^{(i)}(t, \boldsymbol{p})$. Since this is a polynomial state space system it is algebraically observable (and satisfies the ORC for generic initial conditions). The identifiability analysis proceeds as in the previous

section, except that, in this case, the output equation in the dissociation phase is obtained by setting $C_{Ti} = 0$ ($i = 1, 2$).

Since the model consists of two distinct phases to ensure the validity of the analysis for the particular experiments of interest it must first be confirmed that the generators of the ideal $\mathcal{L}^n$ do not change for the states corresponding to particular initial conditions for the experiment. For the association phase the initial conditions are given by

$$B_1(t_s) = B_2(t_s) = 0.$$

Since all dependencies between state variables have been included in the system equations and the association phase initial conditions do not belong to an invariant submanifold let b_{10} and b_{20} denote the state variables at time $\tau > t_s$:

$$B_1(\tau) = b_{10}, \quad \text{and} \quad B_2(\tau) = b_{20}.$$

Then the system can be regarded as starting from the generic initial conditions b_{10} and b_{20}, and so the generating equations do not change. For the dissociation phase take $\tau = t_f$.

A.4   Langmuir (Heterogeneous analyte)

For the heterogeneous analyte version of the Langmuir model (7) the output equation is second order and is given by the following

$$
\begin{aligned}
y_2 + (C_T (k_{a1} + k_{a2}) + k_{d1} + k_{d2}) y_1 \\
+ (C_T (k_{d1}k_{a2} + k_{d2}k_{a1}) + k_{d1}k_{d2}) y_0 \\
- \alpha C_T R (k_{d1}k_{a2} + k_{d2}k_{a1}) = 0
\end{aligned}
$$

where $y_0 = y(t, \boldsymbol{p})$ and $y_i = y^{(i)}(t, \boldsymbol{p})$. Since this is a polynomial state space system it is algebraically observable (and satisfies the ORC for generic initial conditions). The identifiability analysis proceeds as in the previous sections.

A.5   Langmuir with transport (homogeneous analyte)

For the Langmuir with transport model (8) the output equation is second order and is given by the following

$$
\begin{aligned}
\alpha h(y_0 - \alpha R)y_2 + \alpha(k_M + 2k_a R)y_0 y_1 - k_a y_0^2 y_1 \\
+ \alpha k_M (k_d + k_a C_T)y_0^2 - \alpha^2 R(k_a R + k_d h + k_M)y_1 \\
- \alpha h y_1^2 - \alpha^2 k_M R(2k_a C_T + k_d)y_0 + \alpha^3 k_a k_M C_T R^2 = 0
\end{aligned}
$$

where $y_0 = y(t, \boldsymbol{p})$ and $y_i = y^{(i)}(t, \boldsymbol{p})$. Since this is a polynomial state space system it is algebraically observable (and satisfies the ORC for generic initial conditions).

Notice that the coefficient of $y^{(2)}(t, \boldsymbol{p})$ is not 1 and so, in order to perform the structural identifiability analysis, the equation is divided by $\alpha h(y_0 - \alpha R)$ (since $y_0 \neq \alpha R$). Then the equation $\alpha(y^{(2)}(t, \boldsymbol{p}) - y^{(2)}(t, \overline{\boldsymbol{p}}))$ is formed, where $\overline{\boldsymbol{p}}$ represents an indistinguishable parameter vector (to $\boldsymbol{p}$), and expressed as a polynomial. The coefficients of this polynomial must be 0 (otherwise the output satisfies an equation of order less than 2) and these are solved to determine the required relations between the elements of $\boldsymbol{p}$ and $\overline{\boldsymbol{p}}$.

A.6   Langmuir with transport (heterogeneous mixture)

The output equation generated in MAPLE for the heterogeneous analyte version of the Langmuir model (9) is very large and the computation involved necessitates the use of a computer algebra system. Since this output equation is of fourth order and the system is polynomial it is algebraically observable. Proceeding as in the previous examples generates the set of the coefficients, $c_k(\boldsymbol{p}, \overline{\boldsymbol{p}})$, that must be zero, which contains 387 elements, even though there are only 9 parameters. Solving all coefficients simultaneously is a computationally intensive task. However, if the first 25 coefficients are solved and the solution applied to the remaining ones, then this is the solution to all coefficients.

A.7   Langmuir with transport (heterogeneous analyte)

The output equation generated in MAPLE for the heterogeneous ligand version of the Langmuir model (10) is also quite large and the computation proceeds as in the previous examples, though the set of coefficients, $c_k(\boldsymbol{p}, \overline{\boldsymbol{p}})$, that must be zero, contains 39 elements.

A.8   ERC approximation (homogeneous analyte)

The ERC model (11) is defined as follows:

```
F := [(k[a]*C[T]*(R-x[1])-k[d]*x[1])/
    (1 + (k[a]/k[M])*(R-x[1]))  ] :
H := alpha*x[1]:
```

Although this model consists of a single variable, and thus the output equation is readily obtained, an analysis is performed in MAPLE to illustrate the approach for a rational system. In this case the numerators from the list of Lie derivatives is used to define the ideal $\mathcal{L}_f^n$, from which the output equation is obtained:

```
LF1 := numer(listLieDer(H, F, 1)):
out := UnivariatePolynomial(y[1],
    map(expand, LF1), [x[1], y[1]])
```

The output equation is given by:

$$
\begin{aligned}
(\alpha (k_M + k_a R) - k_a y_0) y_1 + k_M \alpha (k_a C_T + k_d) y_0 \\
- k_M \alpha^2 k_a C_T R = 0.
\end{aligned}
$$

It is seen from the Lie derivatives that the ORC is satisfied for generic initial conditions and for the model-specific $B(t) = 0$. Therefore the analysis proceeds as before.

A.9   ERC approximation (heterogeneous mixture)

The heterogeneous mixture version of the Langmuir with transport model (12) gives rise to a second order differential equation for the output that is quite large and the analysis proceeds as in previous examples.

A.10   ERC approximation (heterogeneous analyte)

The heterogeneous analyte version of the Langmuir with transport model (13) gives rise to a second order differential equation for the output that is quite large and the analysis proceeds as in previous examples.