Bulletin of Mathematical Biology (2019) 81:2133–2175
https://doi.org/10.1007/s11538-019-00594-0

Society for Mathematical Biology

Identifiability from a Few Species for a Class of Biochemical Reaction Networks

Gabriela Jeronimo[1,2,3] · **Mercedes Pérez Millán**[1,2,3] · **Pablo Solernó**[1,2]

Received: 20 November 2018 / Accepted: 4 March 2019 / Published online: 22 March 2019

Abstract
Under mass-action kinetics, biochemical reaction networks give rise to polynomial autonomous dynamical systems whose parameters are often difficult to estimate. We deal in this paper with the problem of identifying the kinetic parameters of a class of biochemical networks which are abundant, such as multisite phosphorylation systems and phosphorylation cascades (for example, MAPK cascades). For any system of this class, we explicitly exhibit a single species for each connected component of the associated digraph such that the successive total derivatives of its concentration allow us to identify all the parameters occurring in the component. The number of derivatives needed is bounded essentially by the length of the corresponding connected component of the digraph. Moreover, in the particular case of the cascades, we show that the parameters can be identified from a bounded number of successive derivatives of the last product of the last layer. This theoretical result induces also a heuristic interpolation-based identifiability procedure to recover the values of the rate constants from exact measurements.

Keywords Chemical reaction networks · Mass-action kinetics · Identifiability · MAPK cascade

Gabriela Jeronimo, Mercedes Pérez Millán and Pablo Solernó contributed equally to this work.

Partially supported by UBACYT 20020170100048BA (MPM), UBACYT 20020160100039BA (GJ, PS), CONICET PIP 11220150100483 (MPM), CONICET PIP 11220130100527CO (GJ), CONICET P-UE 22920170100037CO (GJ, MPM, PS), and ANPCyT PICT 2016-0398 (MPM), Argentina.

✉ Mercedes Pérez Millán
mpmillan@dm.uba.ar

Extended author information available on the last page of the article

1 Introduction

Parameter identifiability in a system of ordinary differential equations mainly addresses the question of deciding whether the system parameters can be uniquely determined from data (see for instance Walter and Pronzato 1997; DiStefano 2014, Chapter 10). Since the pioneering paper (Bellman and Åström 1970), this problem has been broadly studied for general systems under different perspectives, including Taylor series and generating series approaches and differential algebra-based approaches. More details can be found in Pohjanpalo (1978), Ollivier (1990), Ljung and Glad (1994), Sedoglavic (2002), Xia and Moog (2003), Saccomani et al. (2003), Bellu et al. (2007), Meshkat et al. (2009), Chis et al. (2011a), Raue et al. (2014) and Hong et al. (2018a). Also, a variety of software tools for identifiability have been developed that work for general classes of models (e.g., polynomial or rational), such as DAISY (Bellu et al. 2007), COMBOS (Meshkat et al. 2014), GenSSI (Ligon et al. 2017) and SIAN (Hong et al. 2018b).

In this paper, we address the identifiability problem for a specific infinite class of models. Our aim is to obtain general statements about all the models in the class (see Walch and Eisenberg 2016; Brouwer et al. 2017 for prior results of this sort but for different classes of models). More precisely, we consider a particular class of systems of equations arising from biochemical reaction networks under mass-action kinetics, which induces polynomial autonomous systems of differential equations. In this framework, in Craciun and Pantea (2008), the authors describe necessary and sufficient conditions for the unique identifiability of the reaction rate constants (the parameters) of a chemical reaction network. Following their approach, we provide in this work sufficient conditions for uniquely identifying all the rate constants of a certain family of biochemical reaction networks from a reduced set of variables (see Definition 3). Unlike other authors (Anguelova et al. 2012), we do not consider all the possible minimal sets of variables allowing parameter identifiability, but we only focus on certain biologically relevant sets.

The family of networks we deal with is abundant in the literature. One example is the multisite phosphorylation system which describes the phosphorylation of a protein in L sites by a kinase(Y)/phosphatase($\tilde{Y}$) pair in a sequential and distributive mechanism (Deshaies and Ferrell 2001). The substrate S_i is the phosphoform obtained from the unphosphorylated substrate S_0 by attaching i phosphate groups to it. Each phosphoform can accept (via an enzymatic reaction involving Y) or lose (via a reaction involving the phosphatase $\tilde{Y}$) at most one phosphate (the mechanism is "distributive"), and there is a specific order to be followed for attaching and removing the phosphate groups (the phosphorylation is "sequential").

Example 1 The reactions in the L-site sequential phosphorylation/dephosphorylation network are represented by the following labeled digraph:

$$\begin{aligned} Y + S_0 &\underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \xrightarrow{c_1} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} \dots \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \xrightarrow{c_L} Y + S_L \\ \tilde{Y} + S_L &\underset{\tilde{b}_L}{\overset{\tilde{a}_L}{\rightleftarrows}} V_L \xrightarrow{\tilde{c}_L} \dots \xrightarrow{\tilde{c}_2} \tilde{Y} + S_1 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \xrightarrow{\tilde{c}_1} \tilde{Y} + S_0, \end{aligned}$$

where $U_1, \ldots, U_L, V_1, \ldots, V_L$ are intermediate enzyme-substrate species. The mass-action dynamical system for this network is [see identity (1) in Sect. 2.1]:

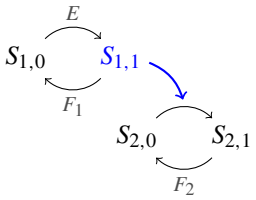
$$\begin{aligned}
\dot{s}_0 &= -a_1 y s_0 + b_1 u_1 + \tilde{c}_1 v_1,\\
\dot{s}_L &= c_L u_L - \tilde{a}_L \tilde{y} s_L + \tilde{b}_L v_L,\\
\dot{u}_i &= a_i y s_{i-1} - (b_i + c_i) u_i,\ 1 \le i \le L,\\
\dot{v}_i &= \tilde{a}_i \tilde{y} s_i - (\tilde{b}_i + \tilde{c}_i) v_i,\ 1 \le i \le L,\\
\dot{y} &= i = 1\sum^{L} - a_i y s_{i-1} + (b_i + c_i) u_i,\\
\dot{\tilde{y}} &= i = 1\sum^{L} - \tilde{a}_i \tilde{y} s_i + (\tilde{b}_i + \tilde{c}_i) v_i,\\
\dot{s}_i &= c_i u_i - a_{i+1} y s_i + b_{i+1} u_{i+1} + \tilde{c}_{i+1} v_{i+1} - \tilde{a}_i \tilde{y} s_i + \tilde{b}_i v_i,\ 1 \le i \le L-1,
\end{aligned}$$

where lower-case letters represent the time-varying concentration of the corresponding chemical species. Here, the derivative with respect to time is represented with a dot over the corresponding variable.

As a consequence of Theorem 1 proved below, all the constants in the first connected component can be identified, in the sense of Definition 3, from the successive total derivatives of s_L up to order $\max\{2, 2L-1\}$ and all the constants in the second connected component can be identified from the successive total derivatives of s_0 up to the same order. Moreover, as proved in Proposition 3, all the constants in the whole network can be identified from the successive total derivatives of s_L up to order $\max\{2, 2L-1\}$.

Another example of major biological importance is phosphorylation cascades, such as the mitogen-activated protein kinase (MAPK) cascade (Catozzi et al. 2016; Huang and Ferrell 1996; Kholodenko 2000; Shaul and Seger 2007). This cascade plays an essential role in signal transduction by modulating gene transcription in response to changes in the cellular environment. MAPK cascades participate in a number of diseases including chronic inflammation and cancer (Davis 2000; Kyriakis and Avruch 2001; Pearson et al. 2001; Schaeffer and Weber 1999; Zarubin and Han 2005) as they control key cellular functions (Hornberg et al. 2005; Pearson et al. 2001; Widmann et al. 1999). We depict in the following example the two-layer signaling cascade.

Example 2 Consider the graph associated with the two-layer simple phosphorylation cascade where the simplified diagram and the corresponding reactions are, respectively:



$$E + S_{1,0} \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \xrightarrow{c_1} E + S_{1,1}$$

$$F_1 + S_{1,1} \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \xrightarrow{\tilde{c}_1} F_1 + S_{1,0}$$

$$S_{1,1} + S_{2,0} \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \xrightarrow{c_2} S_{1,1} + S_{2,1}$$

$$F_2 + S_{2,1} \underset{\tilde{b}_2}{\overset{\tilde{a}_2}{\rightleftarrows}} V_2 \xrightarrow{\tilde{c}_2} F_2 + S_{2,0}.$$

The corresponding mass-action dynamical system is [see (1)]:

$$\begin{aligned}
\dot{s}_{1,0} &= -a_1 e s_{1,0} + b_1 u_1 + \tilde{c}_1 v_1,\\
\dot{s}_{1,1} &= c_1 u_1 - \tilde{a}_1 f_1 s_{1,1} + \tilde{b}_1 v_1 - a_2 s_{1,1} s_{2,0} + (b_2 + c_2) u_2,\\
\dot{e} = -\dot{u_1} &= -a_1 e s_{1,0} + (b_1 + c_1) u_1,\\
\dot{f}_1 = -\dot{v}_1 &= -\tilde{a}_1 f_1 s_{1,1} + (\tilde{b}_1 + \tilde{c}_1) v_1,\\
\dot{s}_{2,0} &= -a_2 s_{1,1} s_{2,0} + b_2 u_2 + \tilde{c}_2 v_2,\\
\dot{s}_{2,1} &= c_2 u_2 - \tilde{a}_2 f_2 s_{2,1} + \tilde{b}_2 v_2,\\
\dot{u}_2 &= a_2 s_{1,1} s_{2,0} - (b_2 + c_2) u_2,\\
\dot{f}_2 = -\dot{v}_2 &= -\tilde{a}_2 f_2 s_{2,1} + (\tilde{b}_2 + \tilde{c}_2) v_2.
\end{aligned}$$

We prove in Theorem 2 that all the parameters in a signaling cascade system can be identified from a single variable: the last product of the last layer ($S_{2,1}$ in the cascade presented in Example 2). This species is usually an output of interest for this type of cascades (Aoki et al. 2011; Chen et al. 2009; Hagen et al. 2013; Lin et al. 2009).

The organization of the paper is as follows. The next section provides introductory material on chemical reaction networks, mass-action kinetics equations and identifiability. Section 3 deals with the general assumptions required by the biochemical reaction networks we consider along the paper. In Sects. 4 and 5 we analyze the identifiability for sequential phosphorylation/dephosphorylation networks and phosphorylation cascades, respectively. We illustrate these results in Sect. 5.2 with a procedure to determine, from (noise-free) data, the 30 rate constants in the three-layer MAPK cascade, which relies on a heuristic to choose points to specialize the variables and solve for the rate constants. Finally, we include a section "Appendix" with the complete proofs of the results stated in the paper.

2 Preliminaries and Basic Notions

2.1 Chemical Reaction Systems

We briefly recall the basic setup of chemical reaction networks and how they give rise to autonomous dynamical systems under mass-action kinetics.

Given a set of s chemical species (denoted by capital letters), a *chemical reaction network* on this set of species is a finite directed graph whose vertices are indicated by complexes (non negative integer linear combinations of the species) and whose edges are labeled by parameters (positive reaction rate constants). The labeled digraph is denoted $G = (\mathcal{V}, \mathcal{R}, \mathbf{k})$, with vertex set $\mathcal{V}$, edge set $\mathcal{R}$ and edge labels $\mathbf{k} \in \mathbb{R}_{>0}^{\#\mathcal{R}}$. If $(y, y') \in \mathcal{R}$, we note $y \to y'$. The complexes determine vectors in $\mathbb{Z}_{\geq 0}^{s}$ (the coefficients of the linear combinations) according to the stoichiometry of the species they consist of. We identify each complex with its corresponding vector and also with the formal linear combination of species specified by its coordinates.

We present a basic example that illustrates how a chemical reaction network gives rise to a dynamical system. This example represents a classical mechanism of enzymatic reactions, usually known as the futile cycle (Huang and Ferrell 1996; Kholodenko 2000; Wang and Sontag 2008):

Example 3 Consider the following graph

$$E + S_0 \underset{b}{\overset{a}{\rightleftarrows}} U \xrightarrow{c} E + S_1 \qquad F + S_1 \underset{\tilde{b}}{\overset{\tilde{a}}{\rightleftarrows}} V \xrightarrow{\tilde{c}} F + S_0.$$

The $s = 6$ variables U, V, S_0, S_1, E, F denote the chemical species. The source and the product of each reaction (i.e. the vertices) are the complexes (non negative linear combinations of the species). Finally, the edge labels in $\mathbf{k} = (a, b, c, \tilde{a}, \tilde{b}, \tilde{c})$ are the reaction rate constants describing how concentrations of the six species change in time as the reactions occur.

The first three complexes give rise to the vectors (0, 0, 1, 0, 1, 0), (1, 0, 0, 0, 0, 0) and (0, 0, 0, 1, 1, 0), while those in the second ones are (0, 0, 0, 1, 0, 1), (0, 1, 0, 0, 0, 0), and (0, 0, 1, 0, 0, 1).

A chemical reaction network G as above under the assumption of *mass-action kinetics* induces a polynomial dynamical system in the following way. Suppose that the species are $X_1, \ldots, X_s$ and their respective concentrations are denoted by $x_1, \ldots, x_s$ (denoted by small letters). We write $k_{yy'}$ for the reaction rate of each reaction $y \to y'$ in $\mathcal{R}$. We introduce the following chemical reaction dynamical system:

$$\dot{\mathbf{x}} = \left(\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}x_s}{\mathrm{d}t}\right) = \sum_{y \to y'} k_{yy'} \mathbf{x}^y (y' - y), \tag{1}$$

where $\mathbf{x} := (x_1, \ldots, x_s)$ and $\mathbf{x}^y := x_1^{y_1} \cdots x_s^{y_s}$ if $y = (y_1, \ldots, y_s)$. The right-hand side of each differential equation $\dot{x}_i$ is a polynomial $f_i(\mathbf{x}, \mathbf{k})$, in the variables $x_1, \ldots, x_s$ with coefficients depending on the parameters $\mathbf{k} := (k_{yy'})_{(y,y') \in \mathcal{R}}$.

For instance, in Example 3 this induced dynamical system is:

$$\begin{aligned}
\dot{u} &= aes_0 - (b+c)u,\\
\dot{v} &= \tilde{a} f s_1 - (\tilde{b}+\tilde{c})v,\\
\dot{s}_0 &= -aes_0 + bu + \tilde{c}v,\\
\dot{s}_1 &= -\tilde{a} f s_1 + \tilde{b}v + cu,\\
\dot{e} &= -aes_0 + (b+c)u,\\
\dot{f} &= -\tilde{a} f s_1 + (\tilde{b}+\tilde{c})v.
\end{aligned} \tag{2}$$

2.2 Identifiability in Chemical Reaction Systems

Among all the different (not always equivalent) notions of identifiability in differential equations and control theory, we have chosen to work from the one introduced in Craciun and Pantea (2008) since it seems specially well suited to the dynamical biochemical systems we consider here (see, for instance, Chis et al. 2011a; Raue et al. 2014 for a survey on the state of the art).

One of the main differences in the various approaches to identifiability is an assumption on the number of experiments that can be conducted with the same parameter values but different initial conditions: a single-experiment approach assumes the experiment is performed only once with some (often generic) initial condition (see, for example, DiStefano 2014, Chapter 10), whereas the multi-experiment approach we adopt in this paper assumes that it is allowed to perform as many experiments as needed with the same parameter values but different initial conditions.

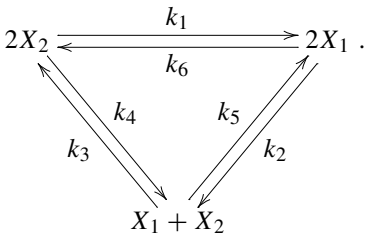
Definition 1 Let $G = (\mathcal{V}, \mathcal{R}, \mathbf{k})$ be a chemical reaction network with s species. Its associated reaction system (1) is called *identifiable* if the map $\Phi : \mathbb{R}^{\#\mathcal{R}}_{>0} \to \mathbb{R}[\mathbf{x}]^s$,

$$\Phi(\mathbf{k}) = \sum_{y \to y'} k_{yy'} \mathbf{x}^y (y' - y),$$

is injective (here $\mathbf{k} = (k_{yy'})_{(y,y') \in \mathcal{R}}$ and $\mathbb{R}[\mathbf{x}]$ is the polynomial ring in the variables $x_1, \ldots, x_s$).

Example 4 In Example 3 [see the corresponding differential equation system (2)], the domain of the map Φ is $\mathbb{R}^6_{>0}$, the target space is $\mathbb{R}[u, v, s_0, s_1, e, f]^6$ and the coordinate functions are the right-hand sides of the differential equations in (2). It is clear that Φ is injective and therefore, the reaction system is identifiable: the right-hand sides of $\dot{s}_0$ and $\dot{s}_1$ determine the six constants $\mathbf{k} = (a, b, c, \tilde{a}, \tilde{b}, \tilde{c})$.

Example 5 (see Craciun and Pantea 2008, Section 2, Fig. 1) Consider the following graph



Here $s = 2$, $\#\mathcal{R} = 6$ and the associated dynamical system is

$$\begin{aligned} \dot{x}_1 &= \ (2k_1 + k_4)x_2^2 - (k_2 + 2k_6)x_1^2 + (k_5 - k_3)x_1x_2 \\ \dot{x}_2 &= -(2k_1 + k_4)x_2^2 + (k_2 + 2k_6)x_1^2 + (k_3 - k_5)x_1x_2 \end{aligned} . \tag{3}$$

Clearly, the map Φ is not injective: parameters $\mathbf{k} \in \mathbb{R}^6_{>0}$ define the same polynomials under Φ if and only if the linear forms $2k_1 + k_4$, $k_2 + 2k_6$ and $k_5 - k_3$ take the same values when evaluated at $\mathbf{k}$. For instance, $\Phi(1, 1, 1, 1, 1, 1) = \Phi(1, 1, 2, 1, 2, 1) = (3x_2^2 - 3x_1^2, -3x_2^2 + 3x_1^2)$. Therefore, the system (3) is not identifiable.

Definition 2 For a chemical reaction network G, we introduce the *total derivative* (or *Lie derivative*) associated to the induced differential equations system as follows: given a differentiable function $\varphi : \mathbb{R}^s \to \mathbb{R}$, its total derivative $\dot{\varphi}$ is defined as

$$\dot{\varphi} := \sum_{i=1}^{s} \frac{\partial \varphi}{\partial x_i} \frac{dx_i}{dt} = \sum_{i=1}^{s} \frac{\partial \varphi}{\partial x_i} \sum_{y \to y'} k_{yy'} \mathbf{x}^y (y'_i - y_i),$$

where each partial derivative $\dfrac{\mathrm{d}x_i}{\mathrm{d}t}$ is replaced according to system (1). For an integer $\ell \geq 1$, we denote by $\varphi^{(\ell)}$ the ℓth iteration of the total derivative of φ (in particular $\varphi^{(1)} = \dot{\varphi}$).

For instance, for the network given in Example 3, its associated dynamical system (2) and the function $\varphi = u^4 + v$, we have

$$\dot{\varphi} = 4u^3(aes_0 - (b + c)u) + \tilde{a} f s_1 - (\tilde{b} + \tilde{c})v.$$

Note that for a differentiable function $\varphi : \mathbb{R}^s \to \mathbb{R}$, the total derivative $\varphi^{(\ell)}$ can be regarded as a function depending on the $(s + \#\mathcal{R})$-variables $\mathbf{x}, \mathbf{k}$.

Definition 3 Let $G = (\mathcal{V}, \mathcal{R}, \mathbf{k})$ be a chemical reaction network with s species. We say its associated reaction system (1) is *identifiable from the variables* $x_{i_1}, \ldots, x_{i_t}$ if there exists a positive integer D such that the following injectivity condition holds: if $\mathbf{k}^*, \mathbf{k}^{**} \in \mathbb{R}^{\#\mathcal{R}}_{>0}$ verify

$$x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k}^*) = x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k}^{**}),$$

for all $1 \leq \ell \leq D$, $1 \leq j \leq t$, then $\mathbf{k}^* = \mathbf{k}^{**}$.

The introduction of the Lie derivative in identifiability is a usual and quite natural approach suitable adapted to our purposes (see, for instance, Chis et al. 2011a). Among other works following this approach, Sedoglavic (2002), Chiş et al. (2011b) and Anguelova et al. (2012) also include a discussion about the number of derivatives needed for the proposed identifiability analysis.

Definitions 1 and 3 are related in the obvious way:

Proposition 1 *A chemical reaction system in the variables* $\mathbf{x} = x_1, \ldots, x_s$ *is identifiable in the sense of Definition 1 if and only it is identifiable from the variables* $x_1, \ldots, x_s$ *in the sense of Definition 3.*

Proof First we observe that the identity $\Phi = \dot{x}_1 \times \dot{x}_2 \times \cdots \times \dot{x}_s$ holds as functions of the argument $\mathbf{k}$. Thus, if Φ is injective, the condition of Definition 3 is satisfied for the variables $x_1, \ldots, x_s$ and the integer $D = 1$. Conversely, suppose that the chemical reaction system is identifiable from the variables $x_1, \ldots, x_s$ using a certain number D of successive total derivatives. Then the function Φ is necessarily injective in the arguments $\mathbf{k}$: if it is not the case, there exist $\mathbf{k}^* \neq \mathbf{k}^{**}$ such that $\dot{x}_i(\mathbf{x}, \mathbf{k}^*) = \dot{x}_i(\mathbf{x}, \mathbf{k}^{**})$ as functions of the variables $\mathbf{x}$ for all $i = 1, \ldots, s$. Since the values of $\mathbf{k}^*, \mathbf{k}^{**}$ are constants with respect to the total derivative we conclude that $x_i^{(\ell)}(\mathbf{x}, \mathbf{k}^*) = x_i^{(\ell)}(\mathbf{x}, \mathbf{k}^{**})$ for all $\ell \in \mathbb{N}$ and all $1 \leq i \leq s$, arriving at a contradiction. □

Example 6 Consider the graph

$$X_1 + X_2 \xrightarrow{k_1} X_3 \xrightarrow{k_2} X_4.$$

and its associated system

$$\dot{x}_1 = -k_1 x_1 x_2, \quad \dot{x}_2 = -k_1 x_1 x_2, \quad \dot{x}_3 = k_1 x_1 x_2 - k_2 x_3, \quad \dot{x}_4 = k_2 x_3.$$

The system is identifiable in the sense of Definition 1. Following Definition 3, the system is identifiable from the single variable x_3 with one derivative (i.e. in this case $D = 1$ in Definition 3). It is also identifiable from the variable x_4, but its total derivative of second order is needed in order to determine all the parameters (i.e. $D = 2$ for this variable). On the other hand, the system is not identifiable from the set of variables $\{x_1, x_2\}$, since the constant k_2 does not appear in any of the successive total derivatives of x_1 nor x_2.

For technical reasons, we need to slightly generalize the notion of identifiability introduced in Definition 3. The following definition is related to the notion of identifiability of parameter combinations (Boulier 2007; Meshkat et al. 2009):

Definition 4 Let $G = (\mathcal{V}, \mathcal{R}, \mathbf{k})$ be a chemical reaction network. Let $p \in \mathbb{N}$ and $\psi : \mathbb{R}_{>0}^{\#\mathcal{R}} \to \mathbb{R}^p$ be a map from the space of parameters in an affine space $\mathbb{R}^p$. We say that the map ψ is *identifiable from the variables* $x_{i_1}, \ldots, x_{i_t}$ if there exists a positive integer D such that the following injectivity condition holds: if $\mathbf{k}^*, \mathbf{k}^{**} \in \mathbb{R}_{>0}^{\#\mathcal{R}}$ verify

$$x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k}^*) = x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k}^{**}),$$

for all $1 \leq \ell \leq D$, $1 \leq j \leq t$, then $\psi(\mathbf{k}^*) = \psi(\mathbf{k}^{**})$.

Roughly speaking, Definition 4 says that the value of the function ψ is uniquely determined by the values of the successive derivatives $x_{i_j}^{(\ell)}$.

Observe that the notion of identifiability of a system from the variables $x_{i_1}, \ldots, x_{i_t}$ as it is defined in Definition 3 can be translated in the sense of Definition 4 as the identifiability of the function $\psi : \mathbb{R}_{>0}^{\#\mathcal{R}} \to \mathbb{R}^{\#\mathcal{R}}$, $\psi(\mathbf{k}) = \mathbf{k}$.

For instance, in the (non identifiable) Example 5, the function $\psi : \mathbb{R}_{>0}^{6} \to \mathbb{R}^3$, defined as $\psi(\mathbf{k}) := (2k_1 + k_4, k_2 + 2k_6, k_5 - k_3)$, is identifiable from x_1 (or x_2, or both variables). In this case we say simply that the constants $2k_1 + k_4,\ k_2 + 2k_6,\ k_5 - k_3$ can be identified from x_1.

This notion will be useful along the paper. We will typically consider very simple functions ψ whose coordinates are either the rate constants or the sum of all the rate constants leaving from one complex.

3 Assumptions on the Biochemical Reaction Networks

We will analyze the identifiability problem for a specific kind of chemical reaction networks. We start by describing the assumptions on the networks we will consider in the sequel.

First, we assume that the "building blocks" of the network have the following shape:

$$X_1 + X_2 \underset{b}{\overset{a}{\rightleftarrows}} U \xrightarrow{c} X_1 + X_3,$$

where U is a species that only participates in those three reactions along all the network. We call U an intermediate species, and we say that species X_1 acts as an *enzyme*, species X_2 acts as a *substrate* and species X_3 acts as a *product*.

Definition 5 We say an intermediate species U *reacts to* the non-intermediate species X_1 if there exists another non-intermediate species X_2 such that the reaction $U \to X_1 + X_2$ exists. We say the non-intermediate species X_1 *reacts with* the non-intermediate species X_2 if there exists an intermediate species U such that the reaction $X_1 + X_2 \to U$ exists.

Example 7 (Example 2 continued) Species U_1, V_1, U_2, V_2 are the intermediate species. E and F act as enzymes. $S_{1,0}$ acts as a substrate in the first connected component and as a product in the second one. Species $S_{2,0}$ and $S_{2,1}$ also act as both substrates and products (in the third and fourth connected components). Finally, $S_{1,1}$ acts as a product in the first connected component, as a substrate in the second one, and as an enzyme in the third one.

We make the following assumption concerning the structure of the network:

Assumption 1

1. Each connected component of the graph is of the following form:

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} \dots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \overset{c_L}{\to} Y + S_L,$$

   where there is a unique enzyme Y acting on all the reactions of the connected component.
2. The intermediate species U_j appearing in the entire network are all different.
3. The non-intermediate species S_j in each connected component are all different, but they may also appear in other connected components.
4. Each complex lies in a unique connected component of the network.

Although the above assumption seems restrictive, it is satisfied by many networks such as the multisite phosphorylation system described in Example 1, the phosphorylation cascades as the one described in Example 2 and also the network in Example 3. As we observed before, in Examples 1 and 3 each species plays a unique role but in Example 2 the species $S_{1,1}$ acts alternatively as a product (in the first connected component), as a substrate (in the second one) and as an enzyme (in the third one).

For an intermediate species U, we call

$$\mathscr{S}_U = \{S : S \text{ acts as a substrate or a product in the connected component determined by } U\}.$$

For instance, in Example 2 we have $\mathscr{S}_{U_1} = \mathscr{S}_{V_1} = \{S_{1,0}, S_{1,1}\}$ and $\mathscr{S}_{U_2} = \mathscr{S}_{V_2} = \{S_{2,0}, S_{2,1}\}$.

We finish our assumptions on the kind of graphs we consider with a slightly technical condition.

Assumption 2 There is a partition of the species of the graph, that is, a decomposition into nonempty disjoint subsets:

$$\mathscr{S} = \mathscr{S}^{(0)} \bigsqcup \mathscr{S}^{(1)} \bigsqcup \cdots \bigsqcup \mathscr{S}^{(M)},$$

where $M \geq 2$, $\sqcup$ denotes the disjoint union, $\mathscr{S}^{(0)}$ is the set of intermediate species and given an intermediate species U with Y acting as an enzyme in the corresponding connected component, there exists $\alpha \geq 1$ with $\mathscr{S}_U \subseteq \mathscr{S}^{(\alpha)}$ and $Y \notin \mathscr{S}^{(\alpha)}$.

Remark 1 Under Assumption 1, the new condition imposed on the graph by Assumption 2 implies the following fact: if X_1 reacts with X_2, then there exists $\alpha \neq \beta$ such that $X_1 \in \mathscr{S}^{(\alpha)}$ and $X_2 \in \mathscr{S}^{(\beta)}$. In particular, if S_i and S_j are two substrates or products in the same connected component, the complex $S_i + S_j$ is not present in the network.

Example 8 In Example 2 we can consider the following partition $\mathscr{S}^{(0)} = \{U_1, V_1, U_2, V_2\}$, $\mathscr{S}^{(1)} = \{S_{1,0}, S_{1,1}\}$, $\mathscr{S}^{(2)} = \{S_{2,0}, S_{2,1}\}$, $\mathscr{S}^{(3)} = \{E\}$, $\mathscr{S}^{(4)} = \{F_1\}$, $\mathscr{S}^{(5)} = \{F_2\}$.

However it is not the unique possible partition: for instance, another choice could be $\mathscr{S}^{(0)}$, $\mathscr{S}^{(1)}$ and $\mathscr{S}^{(2)}$ as before, but $\mathscr{S}^{(3)}$, $\mathscr{S}^{(4)}$ and $\mathscr{S}^{(5)}$ are replaced by the single set $\{E, F_1, F_2\}$.

4 Identifiability in Connected Components

This section is devoted to dealing with the identifiability problem for chemical reaction networks satisfying the assumptions stated in Sect. 3. Our aim is to show that all reaction constants of the network can be identified from the successive derivatives of the variables in a certain family of non-intermediates.

In order to do this, we choose a suitable subset of variables and estimate the maximum number of successive derivatives of them that we need to identify all the reaction constants. Namely, we choose variables $x_{i_1}, \ldots, x_{i_t}$ and determine a number D_j of successive derivatives of x_{i_j}, for $1 \leq j \leq t$, so that the injectivity condition in Definition 3 holds for $D = \max\{D_j\}$.

Since the derivatives $x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k})$ are polynomials in the variables $\mathbf{x}$ with coefficients that are polynomials in the reaction rate constants $\mathbf{k}$, showing that the parameters $\mathbf{k}$ are identifiable from $x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k})$ for $1 \leq \ell \leq D_j$, $1 \leq j \leq t$, is the same as showing that they are uniquely determined by the coefficients of the polynomials $x_{i_j}^{(\ell)}(\mathbf{x}, \mathbf{k})$. Thus, our strategy to proving identifiability will be to locate suitable subsets of monomials in the derivatives $x_{i_j}^{(\ell)}$ that enable us to prove that the values of all the reaction constants can be uniquely determined from their corresponding coefficients.

4.1 Identifying the Constants in One Connected Component from One Variable

The aim of this section is to show that all the reaction constants in a connected component

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \xrightarrow{c_1} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \xrightarrow{c_2} \ldots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \xrightarrow{c_L} Y + S_L \tag{4}$$

of a network satisfying the assumptions stated in Sect. 3 are identifiable from a limited number of successive derivatives of the variable s_L representing the concentration of the last product.

We start by showing that all the constants c_L, a_L, b_L, and, for $1 \leq j \leq L-1$, a_j and $b_j + c_j$ can be identified (in the sense of Definition 4) simply from the first three derivatives of this variable. Then, we proceed to identify recursively all the constants c_j (and consequently, also the constants b_j) for $j = L-1, \ldots, 1$, from higher-order derivatives of s_L. The main result of this section is the following:

Proposition 2 *All the constants in a connected component*

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \xrightarrow{c_1} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \xrightarrow{c_2} \ldots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \xrightarrow{c_L} Y + S_L$$

Table 1 The constants in the connected component (4) can be identified from s_L

Derivative	Monomial	Coefficient (up to sign)	Identified constant
$\dot{s}_L$	u_L	c_L	c_L
$\ddot{s}_L$	ys_{L-1}	$c_L a_L$	a_L
	u_L	$c_L(b_L + c_L)$	b_L
$s_L^{(3)}$	$ys_{j-1}s_{L-1}$	$c_L a_L a_j$	a_j $(1 \le j \le L-1)$
	$u_j s_{L-1}$	$c_L a_L (b_j + c_j)$	$b_j + c_j$ $(1 \le j \le L-1)$
$s_L^{(2k+1)}$	$y^k u_{L-k}$	$c_{L-k} \prod_{j=0}^{k-1} c_{L-j} a_{L-j}$	c_{L-k}, b_{L-k} $(1 \le k \le L-1)$

This table shows the monomials to be considered (column 2) in each of the successive derivatives of s_L (column 1). For each monomial, taking into account the constants already identified, the corresponding coefficient (column 3) enables us to identify the constant appearing in the last column

of a network satisfying the assumptions in Sect. 3 can be identified from $s_L^{(\ell)}$ *with* $1 \le \ell \le \max\{2, 2L-1\}$.

The strategy in the proof of this result consists in the exact computation of the coefficients of certain distinguished monomials in the successive derivatives of s_L. This explicit computation enables us to achieve the identifiability of all the constants of the connected component by means of a recursive procedure that we summarize in Table 1. For a complete proof, see Proposition 4 in "Appendix A."

We illustrate the procedure underlying the proof of the previous statement with a simple example.

Example 9 Consider the network

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \xrightarrow{c_1} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \xrightarrow{c_2} Y + S_2$$
$$Z + S_2 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} W \xrightarrow{\tilde{c}_1} Z + S_3$$

According to Proposition 2, all the constants in the first connected component can be identified from $s_2^{(\ell)}$ with $1 \le \ell \le 3$. In fact, if we call $K_1 = b_1 + c_1$, $K_2 = b_2 + c_2$ and $\tilde{K}_1 = \tilde{b}_1 + \tilde{c}_1$,

$$\dot{s}_2 = -\tilde{a}_1 s_2 z + \boxed{c_2} u_2 + \tilde{b}_1 w,$$

$$\ddot{s}_2 = -\tilde{a}_1[\dot{s}_2 z + s_2(\underbrace{-\tilde{a}_1 s_2 z + \tilde{K}_1 w}_{\dot{z}})] + c_2[\underbrace{\boxed{a_2} s_1 y - \boxed{K_2} u_2}_{\dot{u}_2}] + \tilde{b}_1[\underbrace{\tilde{a}_1 s_2 z - \tilde{K}_1 w}_{\dot{w}}],$$

$$\begin{aligned} s_2^{(3)} = &-\tilde{a}_1[\ddot{s}_2 z + 2\dot{s}_2\dot{z} + s_2(-\tilde{a}_1(\dot{s}_2 z + s_2\dot{z}) + \tilde{K}_1\dot{w})] \\ &+ c_2[a_2(\underbrace{(-a_2 s_1 y + \boxed{c_1} u_1 + b_2 u_2)}_{\dot{s}_1} y + s_1(\underbrace{-\boxed{a_1} s_0 y + \boxed{K_1} u_1 - a_2 s_2 y + K_2 u_2}_{\dot{y}})) - K_2 \dot{u}_2] \\ &+ \tilde{b}_1[\tilde{a}_1(\dot{s}_2 z + s_2\dot{z}) - \tilde{K}_1\dot{w}], \end{aligned}$$

where the constants c_2, a_2, K_2 (thus, also $b_2 = K_2 - c_2$), a_1, K_1 and c_1 (thus, also $b_1 = K_1 - c_1$) are identified in Table 1.

A direct consequence of Proposition 2 is the following theorem:

Theorem 1 *If a chemical reaction network satisfying the assumptions in Sect. 3 consists of N connected components*

$$\begin{aligned}
&Y_1 + S_{1,0} \underset{b_{1,1}}{\overset{a_{1,1}}{\rightleftarrows}} U_{1,1} \overset{c_{1,1}}{\to} Y_1 + S_{1,1} \underset{b_{1,2}}{\overset{a_{1,2}}{\rightleftarrows}} U_{1,2} \overset{c_{1,2}}{\to} \dots \\
&\dots Y_1 + S_{1,L_1-1} \underset{b_{1,L_1}}{\overset{a_{1,L_1}}{\rightleftarrows}} U_{1,L_1} \overset{c_{1,L_1}}{\to} Y_1 + S_{1,L_1} \\
&\vdots \\
&Y_N + S_{N,0} \underset{b_{N,1}}{\overset{a_{N,1}}{\rightleftarrows}} U_{N,1} \overset{c_{N,1}}{\to} Y_N + S_{N,1} \underset{b_{N,2}}{\overset{a_{N,2}}{\rightleftarrows}} U_{N,2} \overset{c_{N,2}}{\to} \dots \\
&\dots Y_N + S_{N,L_N-1} \underset{b_{N,L_N}}{\overset{a_{N,L_N}}{\rightleftarrows}} U_{N,L_N} \overset{c_{N,L_N}}{\to} Y_N + S_{N,L_N},
\end{aligned}$$

then the associated system is identifiable from the variables $s_{1,L_1}, \dots, s_{N,L_N}$ corresponding to the last products of each connected component of the network. Moreover, for every $1 \le i \le N$, the order of derivation needed for the variable s_{i,L_i} is at most $\max\{2, 2L_i - 1\}$.

4.2 Identifying the Constants in Two Connected Components from One Variable

In this subsection, we analyze the identifiability problem for a subclass of the networks we have been considering. More precisely, we consider networks containing pairs of connected components of the following type:

$$\begin{aligned}
&Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} \dots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \overset{c_L}{\to} Y + S_L, \\
&\widetilde{Y} + S_L \underset{\tilde{b}_L}{\overset{\tilde{a}_L}{\rightleftarrows}} V_L \overset{\tilde{c}_L}{\to} \widetilde{Y} + S_{L-1} \underset{\tilde{b}_{L-1}}{\overset{\tilde{a}_{L-1}}{\rightleftarrows}} V_{L-1} \overset{\tilde{c}_{L-1}}{\to} \dots \widetilde{Y} + S_1 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \overset{\tilde{c}_1}{\to} \widetilde{Y} + S_0.
\end{aligned} \tag{5}$$

As before, we work under the assumptions made in Sect. 3.

By Proposition 2, we know that all the constants in the first connected component in (5) can be identified from a certain number of successive derivatives of s_L. Using the specific structure of the second component, we can prove that the same derivatives also enable the identification of the reaction rate constants of that component.

We first prove that the constants $\tilde{a}_L, \tilde{b}_L, \tilde{c}_L$, and, for $1 \le j \le L-1$, $\tilde{a}_j$ and $\tilde{b}_j + \tilde{c}_j$ can be identified from $\dot{s}_L$ and $\ddot{s}_L$ and, then, by means of a recursive explicit computation of coefficients of a family of distinguished monomials in higher-order derivatives of s_L, we show how to successively identify the constants $\tilde{b}_j$ for $j = L-1, \dots, 1$, and, consequently, also the constants $\tilde{c}_j$. In this way, we deduce:

Table 2 The constants in the two connected components in (5) can be identified from s_L

Derivative	Monomial	Coefficient (up to sign)	Identified constant
$\dot{s}_L$	u_L	c_L	c_L
	$\tilde{y}s_L$	$\tilde{a}_L$	$\tilde{a}_L$
	v_L	$\tilde{b}_L$	$\tilde{b}_L$
$\ddot{s}_L$	ys_{L-1}	$c_L a_L$	a_L
	u_L	$c_L(b_L + c_L)$	b_L
	v_L	$\tilde{b}_L(\tilde{b}_L + \tilde{c}_L)$	$\tilde{c}_L$
	$\tilde{y}s_j s_L$	$\tilde{a}_L \tilde{a}_j$	$\tilde{a}_j$ $(1 \le j \le L-1)$
	$v_j s_L$	$\tilde{a}_L(\tilde{b}_j + \tilde{c}_j)$	$\tilde{b}_j + \tilde{c}_j$ $(1 \le j \le L-1)$
$s_L^{(3)}$	$ys_{j-1}s_{L-1}$	$c_L a_L a_j$	a_j $(1 \le j \le L-1)$
	$u_j s_{L-1}$	$c_L a_L(b_j + c_j)$	$b_j + c_j$ $(1 \le j \le L-1)$
$s_L^{(2k+1)}$	$y^k u_{L-k}$	$c_{L-k}\prod_{j=0}^{k-1} c_{L-j}a_{L-j}$	c_{L-k}, b_{L-k}
$k = 1, \ldots, L-1$	$y^k v_{L-k}$	$\tilde{b}_{L-k}\prod_{j=0}^{k-1} c_{L-j}a_{L-j}$	$\tilde{b}_{L-k}, \tilde{c}_{L-k}$

This table shows the monomials to be considered (column 2) in each of the successive derivatives of s_L (column 1). For each monomial, taking into account the constants already identified, the corresponding coefficient (column 3) enables us to identify the constant appearing in the last column

Proposition 3 *Given a chemical reaction network satisfying the assumptions in Sect. 3, all the constants in two connected components of the type*

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} \ldots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \overset{c_L}{\to} Y + S_L,$$
$$\widetilde{Y} + S_L \underset{\tilde{b}_L}{\overset{\tilde{a}_L}{\rightleftarrows}} V_L \overset{\tilde{c}_L}{\to} \widetilde{Y} + S_{L-1} \underset{\tilde{b}_{L-1}}{\overset{\tilde{a}_{L-1}}{\rightleftarrows}} V_{L-1} \overset{\tilde{c}_{L-1}}{\to} \ldots \widetilde{Y} + S_1 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \overset{\tilde{c}_1}{\to} \widetilde{Y} + S_0$$

can be identified from $s_L^{(\ell)}$ *with* $1 \le \ell \le \max\{2, 2L-1\}$.

We summarize the identifiability procedure underlying the proof of the previous proposition in Table 2, and we also illustrate the result in Example 10. For a complete proof, see Proposition 5 in "Appendix A."

Example 10 Consider the network

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} Y + S_2$$
$$\tilde{Y} + S_2 \underset{\tilde{b}_2}{\overset{\tilde{a}_2}{\rightleftarrows}} V_2 \overset{\tilde{c}_2}{\to} \tilde{Y} + S_1 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \overset{\tilde{c}_1}{\to} \tilde{Y} + S_0$$

According to Proposition 3, all the constants in the two connected components can be identified from $s_2^{(\ell)}$ with $1 \le \ell \le 3$. In fact, if we call $K_1 = b_1 + c_1$, $K_2 = b_2 + c_2$, $\tilde{K}_1 = \tilde{b}_1 + \tilde{c}_1$ and $\tilde{K}_2 = \tilde{b}_2 + \tilde{c}_2$:

$$
\begin{aligned}
\dot{s}_2 &= -\boxed{\tilde{a}_2} s_2\tilde{y} + \boxed{c_2} u_2 + \boxed{\tilde{b}_2} v_2,\\
\ddot{s}_2 &= -\tilde{a}_2[\dot{s}_2\tilde{y} + s_2(\underbrace{-\boxed{\tilde{a}_1} s_1\tilde{y} + \boxed{\tilde{K}_1} v_1 - \tilde{a}_2 s_2\tilde{y} + \boxed{\tilde{K}_2} v_2}_{\dot{\tilde{y}}})]\\
&\quad + c_2[\underbrace{\boxed{a_2} s_1 y - \boxed{K_2} u_2}_{\dot{u}_2}] + \tilde{b}_2[\underbrace{\tilde{a}_2 s_2\tilde{y} - \tilde{K}_2 v_2}_{\dot{v}_2}],\\
s_2^{(3)} &= -\tilde{a}_2[\ddot{s}_2\tilde{y} + 2\dot{s}_2\dot{\tilde{y}} + s_2(-\tilde{a}_1(\dot{s}_1\tilde{y} + s_1\dot{\tilde{y}})\\
&\quad + \tilde{K}_1(\tilde{a}_1 s_1\tilde{y} - \tilde{K}_1 v_1) - \tilde{a}_2(\dot{s}_2\tilde{y} + s_2\dot{\tilde{y}}) + \tilde{K}_2\dot{v}_2)]\\
&\quad + c_2[a_2(\underbrace{(-a_2 s_1 y - \tilde{a}_1 s_1\tilde{y} + \boxed{c_1} u_1 + b_2 u_2 + \tilde{c}_2 v_2 + \boxed{\tilde{b}_1} v_1)}_{\dot{s}_1} y\\
&\quad + s_1(\underbrace{-\boxed{a_1} s_0 y + \boxed{K_1} u_1 - a_2 s_2 y + K_2 u_2}_{\dot{y}})) - K_2\dot{u}_2]\\
&\quad + \tilde{b}_2[\tilde{a}_2(\dot{s}_2\tilde{y} + s_2\dot{\tilde{y}}) - \tilde{K}_2\dot{v}_2].
\end{aligned}
$$

Here, the constants c_2, $\tilde{a}_2$, $\tilde{b}_2$, a_2, K_2 (then, b_2), $\tilde{c}_2$, $\tilde{a}_1$, $\tilde{K}_1$, a_1, K_1, c_1 (then, b_1) and $\tilde{b}_1$ (then, $\tilde{c}_1$) are identified in Table 2.

A direct consequence of Proposition 3 is the following corollary:

Corollary 1 *If a chemical reaction network satisfying the assumptions in Sect. 3 consists of $2N$ connected components of the shape*

$$
\begin{aligned}
&Y_1 + S_{1,0} \underset{b_{1,1}}{\overset{a_{1,1}}{\rightleftarrows}} U_{1,1} \overset{c_{1,1}}{\to} Y_1 + S_{1,1} \underset{b_{1,2}}{\overset{a_{1,2}}{\rightleftarrows}} U_{1,2} \overset{c_{1,2}}{\to} \dots\\
&\dots Y_1 + S_{1,L_1-1} \underset{b_{1,L_1}}{\overset{a_{1,L_1}}{\rightleftarrows}} U_{1,L_1} \overset{c_{1,L_1}}{\to} Y_1 + S_{1,L_1}\\
&\widetilde{Y}_1 + S_{1,L_1} \underset{\tilde{b}_{1,L_1}}{\overset{\tilde{a}_{1,L_1}}{\rightleftarrows}} V_{1,L_1} \overset{\tilde{c}_{1,L_1}}{\to} \widetilde{Y}_1 + S_{1,L_1-1} \underset{\tilde{b}_{1,L_1-1}}{\overset{\tilde{a}_{1,L_1-1}}{\rightleftarrows}} V_{1,L_1-1} \overset{\tilde{c}_{1,L_1-1}}{\to} \dots\\
&\dots \widetilde{Y}_1 + S_{1,1} \underset{\tilde{b}_{1,1}}{\overset{\tilde{a}_{1,1}}{\rightleftarrows}} V_{1,1} \overset{\tilde{c}_{1,1}}{\to} \widetilde{Y}_1 + S_{1,0}\\
&\vdots\\
&Y_N + S_{N,0} \underset{b_{N,1}}{\overset{a_{N,1}}{\rightleftarrows}} U_{N,1} \overset{c_{N,1}}{\to} Y_N + S_{N,1} \underset{b_{N,2}}{\overset{a_{N,2}}{\rightleftarrows}} U_{N,2} \overset{c_{N,2}}{\to} \dots\\
&\dots Y_N + S_{N,L_N-1} \underset{b_{N,L_N}}{\overset{a_{N,L_N}}{\rightleftarrows}} U_{N,L_N} \overset{c_{N,L_N}}{\to} Y_N + S_{N,L_N}
\end{aligned}
$$

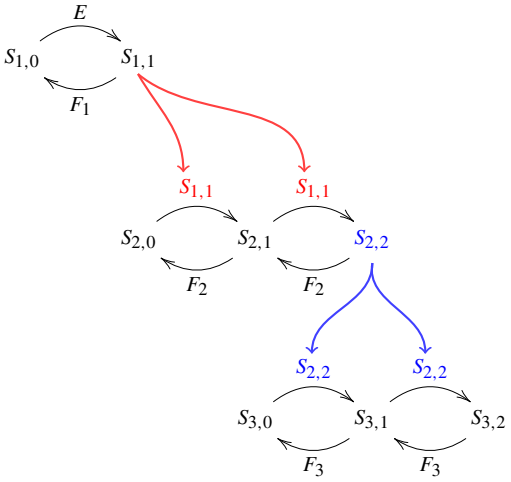
$$\widetilde{Y}_N + S_{N,L_N} \underset{\tilde{b}_{N,L_N}}{\overset{\tilde{a}_{N,L_N}}{\rightleftarrows}} V_{N,L_N} \overset{\tilde{c}_{N,L_N}}{\to} \widetilde{Y}_N + S_{N,L_N-1} \underset{\tilde{b}_{N,L_N-1}}{\overset{\tilde{a}_{N,L_N-1}}{\rightleftarrows}} V_{N,L_N-1} \overset{\tilde{c}_{N,L_N-1}}{\to} \dots$$

$$\dots \widetilde{Y}_N + S_{N,1} \underset{\tilde{b}_{N,1}}{\overset{\tilde{a}_{N,1}}{\rightleftarrows}} V_{N,1} \overset{\tilde{c}_{N,1}}{\to} \widetilde{Y}_N + S_{N,0}$$

then the associated system is identifiable from the variables $s_{1,L_1}, \dots, s_{N,L_N}$. *Moreover, for every* $1 \leq i \leq N$, *the order of derivation needed for the variable* s_{i,L_i} *is at most* $\max\{2, 2L_i - 1\}$.

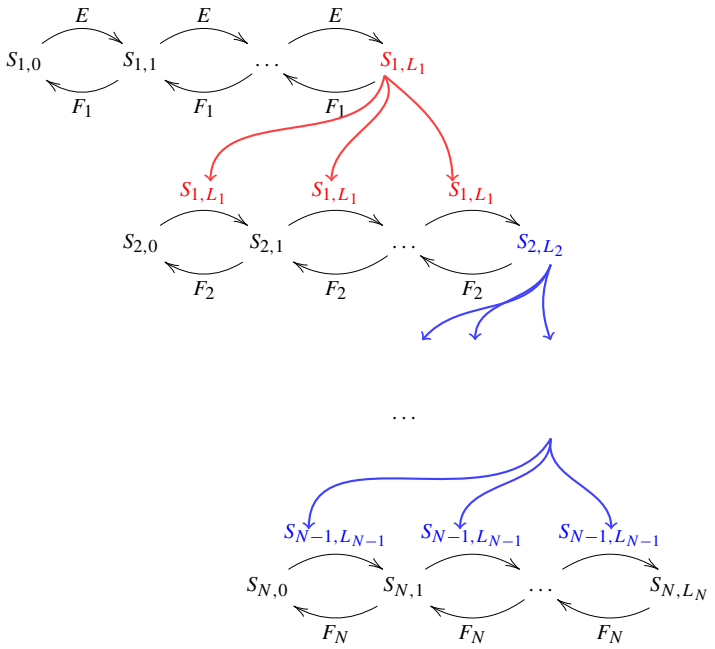
5 Identifying the Cascade

We will consider in this section networks that are called *cascades*. Signaling cascades are biochemical networks of major biological importance as they participate in a number of several diseases and also control key cellular functions (Davis 2000; Kyriakis and Avruch 2001; Pearson et al. 2001; Schaeffer and Weber 1999; Widmann et al. 1999; Zarubin and Han 2005). The mitogen-activated protein kinase (MAPK) cascade is a network present in all eukaryotic cells and one of the most extensively modeled signaling systems (Hornberg et al. 2005; Huang and Ferrell 1996; Qiao et al. 2007). A schematic representation of the network is the following



where $S_{1,0}$ represents the kinase MAPKKK, and $S_{1,1}$ represents the activated form MAPKKK*. $S_{2,0}$, $S_{2,1}$ and $S_{2,2}$ stand for MAPKK, MAPKK-P and MAPKK-PP, respectively. And finally, $S_{3,0}$, $S_{3,1}$ and $S_{3,2}$ stand for MAPK, MAPK-P and MAPK-PP, respectively. F_1 represents the enzyme that deactivates MAPKKK*, and F_2 and F_3 represent the corresponding phosphatase of each layer.

More generally, cascades consist of $N \geq 1$ layers and are represented by the following scheme:



One important feature of cascades is that the enzyme on the first connected component of a certain layer is the last product of the first component of the previous layer. For instance, S_{1,L_1} is the enzyme on the second layer and so on. The corresponding reaction network for the N-layer cascade is the following

$$E + S_{1,0} \underset{b_{1,1}}{\overset{a_{1,1}}{\rightleftarrows}} U_{1,1} \overset{c_{1,1}}{\to} E + S_{1,1} \underset{b_{1,2}}{\overset{a_{1,2}}{\rightleftarrows}} U_{1,2} \overset{c_{1,2}}{\to} \dots$$

$$\dots E + S_{1,L_1-1} \underset{b_{1,L_1}}{\overset{a_{1,L_1}}{\rightleftarrows}} U_{1,L_1} \overset{c_{1,L_1}}{\to} E + S_{1,L_1}$$

$$F_1 + S_{1,L_1} \underset{\tilde{b}_{1,L_1}}{\overset{\tilde{a}_{1,L_1}}{\rightleftarrows}} V_{1,L_1} \overset{\tilde{c}_{1,L_1}}{\to} F_1 + S_{1,L_1-1} \underset{\tilde{b}_{1,L_1-1}}{\overset{\tilde{a}_{1,L_1-1}}{\rightleftarrows}} V_{1,L_1-1} \overset{\tilde{c}_{1,L_1-1}}{\to} \dots$$

$$\dots F_1 + S_{1,1} \underset{\tilde{b}_{1,1}}{\overset{\tilde{a}_{1,1}}{\rightleftarrows}} V_{1,1} \overset{\tilde{c}_{1,1}}{\to} F_1 + S_{1,0}$$

$$S_{1,L_1} + S_{2,0} \underset{b_{2,1}}{\overset{a_{2,1}}{\rightleftarrows}} U_{2,1} \overset{c_{2,1}}{\to} S_{1,L_1} + S_{2,1} \underset{b_{2,2}}{\overset{a_{2,2}}{\rightleftarrows}} U_{2,2} \overset{c_{2,2}}{\to} \dots$$

$$\dots S_{1,L_1} + S_{2,L_2-1} \underset{b_{2,L_2}}{\overset{a_{2,L_2}}{\rightleftarrows}} U_{2,L_2} \overset{c_{2,L_2}}{\to} S_{1,L_1} + S_{2,L_2}$$

$$F_2 + S_{2,L_2} \underset{\tilde{b}_{2,L_2}}{\overset{\tilde{a}_{2,L_2}}{\rightleftarrows}} V_{2,L_2} \overset{\tilde{c}_{2,L_2}}{\to} F_2 + S_{2,L_2-1} \underset{\tilde{b}_{2,L_2-1}}{\overset{\tilde{a}_{2,L_2-1}}{\rightleftarrows}} V_{2,L_2-1} \overset{\tilde{c}_{2,L_2-1}}{\to} \dots$$

$$
\begin{aligned}
&\ldots F_2 + S_{2,1} \underset{\tilde{b}_{2,1}}{\overset{\tilde{a}_{2,1}}{\rightleftarrows}} V_{2,1} \overset{\tilde{c}_{2,1}}{\to} F_2 + S_{2,0} \\
&\vdots \\
&S_{N-1,L_{N-1}} + S_{N,0} \underset{b_{N,1}}{\overset{a_{N,1}}{\rightleftarrows}} U_{N,1} \overset{c_{N,1}}{\to} \ldots \\
&\ldots S_{N-1,L_{N-1}} + S_{N,L_N-1} \underset{b_{N,L_N}}{\overset{a_{N,L_N}}{\rightleftarrows}} U_{N,L_N} \overset{c_{n,L_n}}{\to} S_{N-1,L_{N-1}} + S_{N,L_N} \\
&F_N + S_{N,L_N} \underset{\tilde{b}_{N,L_N}}{\overset{\tilde{a}_{N,L_N}}{\rightleftarrows}} V_{N,L_N} \overset{\tilde{c}_{N,L_N}}{\to} F_N + S_{N,L_N-1} \underset{\tilde{b}_{N,L_N-1}}{\overset{\tilde{a}_{N,L_N-1}}{\rightleftarrows}} V_{N,L_N-1} \overset{\tilde{c}_{N,L_N-1}}{\to} \ldots \\
&\ldots F_N + S_{N,1} \underset{\tilde{b}_{N,1}}{\overset{\tilde{a}_{N,1}}{\rightleftarrows}} V_{N,1} \overset{\tilde{c}_{N,1}}{\to} F_N + S_{N,0}. \qquad (6)
\end{aligned}
$$

We will assume $F_i \neq F_j$ if $i \neq j$ and consider the following partition of the non-intermediate species, which satisfies Assumption 2:

$$\mathscr{S} = \mathscr{S}^{(1)} \bigsqcup \mathscr{S}^{(2)} \bigsqcup \cdots \bigsqcup \mathscr{S}^{(2N+1)},$$

with $\mathscr{S}^{(m)} = \{S_{m,0}, \ldots, S_{m,L_m}\}$ and $\mathscr{S}^{(N+m)} = \{F_m\}$, for $1 \leq m \leq N$, and $\mathscr{S}^{(2N+1)} = \{E\}$.

As our running example for this section, we will consider the two-layer cascade with 18 reactions.

Example 11

$$
\begin{aligned}
&E + S_{1,0} \underset{b_{1,1}}{\overset{a_{1,1}}{\rightleftarrows}} U_{1,1} \overset{c_{1,1}}{\to} E + S_{1,1} \\
&F_1 + S_{1,1} \underset{\tilde{b}_{1,1}}{\overset{\tilde{a}_{1,1}}{\rightleftarrows}} V_{1,1} \overset{\tilde{c}_{1,1}}{\to} F_1 + S_{1,0} \\
&S_{1,1} + S_{2,0} \underset{b_{2,1}}{\overset{a_{2,1}}{\rightleftarrows}} U_{2,1} \overset{c_{2,1}}{\to} S_{1,1} + S_{2,1} \underset{b_{2,2}}{\overset{a_{2,2}}{\rightleftarrows}} U_{2,2} \overset{c_{2,2}}{\to} S_{1,1} + S_{2,2} \\
&F_2 + S_{2,2} \underset{\tilde{b}_{2,2}}{\overset{\tilde{a}_{2,2}}{\rightleftarrows}} V_{2,2} \overset{\tilde{c}_{2,2}}{\to} F_2 + S_{2,1} \underset{\tilde{b}_{2,1}}{\overset{\tilde{a}_{2,1}}{\rightleftarrows}} V_{2,1} \overset{\tilde{c}_{2,1}}{\to} F_2 + S_{2,0}.
\end{aligned}
$$

The first layer consists of two connected components. The first component consists of one modification performed by the enzyme E on the substrate $S_{1,0}$, which is transformed into the product $S_{1,1}$. On the second connected component, the enzyme F_1 performs the reverse modification on the substrate $S_{1,1}$. The second layer is similar. For this network we have $\mathscr{S}^{(1)} = \{S_{1,0}, S_{1,1}\}$, $\mathscr{S}^{(2)} = \{S_{2,0}, S_{2,1}, S_{2,2}\}$, $\mathscr{S}^{(3)} = \{F_1\}$, $\mathscr{S}^{(4)} = \{F_2\}$ and $\mathscr{S}^{(5)} = \{E\}$.

5.1 Identifiability of Constants in a General Cascade

The aim of this section is to show that all the constants in the cascades introduced in (6) can be identified from successive derivatives of the variable corresponding to the last product of the last layer, S_{N,L_N}. In order to prove this, we relate the derivatives of the last product of a given layer of the cascade with the derivatives of the last product of the layer immediately above.

To shorten notation, we will denote $K_{m,j} = b_{m,j} + c_{m,j}$ and $\tilde{K}_{m,j} = \tilde{b}_{m,j} + \tilde{c}_{m,j}$ for every $1 \leq m \leq N$, $1 \leq j \leq L_m$. Also, for unifying purposes, we set $S_{0,L_0} := E$.

For $1 \leq n \leq N$, consider the variable s_{n,L_n} corresponding to the last product of the nth layer of the cascade. We have that

$$\begin{aligned}\dot{s}_{n,L_n} = {} & c_{n,L_n} u_{n,L_n} - \tilde{a}_{n,L_n} s_{n,L_n} f_n + \tilde{b}_{n,L_n} v_{n,L_n} \\ & - \sum_{j=1}^{L_{n+1}} a_{n+1,j} s_{n,L_n} s_{n+1,j-1} + \sum_{j=1}^{L_{n+1}} K_{n+1,j} u_{n+1,j}\end{aligned}$$

and, for $n = N$, only the three first terms appear in the derivative, *i.e.* $a_{N+1,j} = 0$, $K_{N+1,j} = 0$ for all j. The second derivative of s_{n,L_n} is

$$\begin{aligned}\ddot{s}_{n,L_n} = {} & c_{n,L_n}(a_{n,L_n} s_{n-1,L_{n-1}} s_{n,L_n-1} - K_{n,L_n} u_{n,L_n}) - \tilde{a}_{n,L_n}(\dot{s}_{n,L_n} f_n + s_{n,L_n} \dot{f}_n) \\ & + \tilde{b}_{n,L_n}(\tilde{a}_{n,L_n} s_{n,L_n} f_n - \tilde{K}_{n,L_n} v_{n,L_n}) - \sum_{j=1}^{L_{n+1}} a_{n+1,j}(\dot{s}_{n,L_n} s_{n+1,j-1} + s_{n,L_n} \dot{s}_{n+1,j-1}) \\ & + \sum_{j=1}^{L_{n+1}} K_{n+1,j}(a_{n+1,j} s_{n,L_n} s_{n+1,j-1} - K_{n+1,j} u_{n+1,j}).\end{aligned}$$

We can see that the variable $s_{n-1,L_{n-1}}$ corresponding to the last product of the $(n-1)$th layer appears in the second derivative of s_{n,L_n}. More precisely, from the above expression, it follows easily that it only appears in the term $c_{n,L_n} a_{n,L_n} s_{n-1,L_{n-1}} s_{n,L_n-1}$, since $S_{n-1,L_{n-1}}$ does not react with or to F_n or $S_{n+1,j}$ for any j. Thus, two differentiation steps enable us to "jump" from one layer of the cascade to the layer immediately above. Inductively, the idea is that, for $m < n$, by taking $2(n-m)$ derivatives of s_{n,L_n} we will reach the mth layer; that is, the variable s_{m,L_m} will appear and so, the successive derivatives of s_{m,L_m} will appear in higher-order derivatives of s_{n,L_n}.

Now, by the results in Sect. 4.2 for the case of two connected components of the form (5), we can identify all constants in the mth layer of the cascade by looking at the coefficients of certain monomials of the derivatives of s_{m,L_m}. Then, our previous considerations will imply that those constants can be identified from successive derivatives of s_{n,L_n} as well. In order to ensure that this can be achieved, we prove that certain monomials effectively appear in the derivatives of s_{n,L_n} and compute their coefficients (see Proposition 6 in "Appendix A" for a precise statement and its proof).

When considering the last product of the last layer of the cascade, we obtain our main result:

Theorem 2 *All the constants in the network* (6) *can be identified from* $s^{(\ell)}_{N,L_N}$ *with* $1 \leq \ell \leq \max\{2N;\, 2(N-m+L_m)-1,\ 1 \leq m \leq N\}$.

We now summarize the identifiability procedure which proves the previous theorem. The procedure obtains recursively, for $m = N, N-1, \ldots, 1$, the values of the constants $a_{m,j}, \tilde{a}_{m,j}, b_{m,j}, \tilde{b}_{m,j}, c_{m,j}$ and $\tilde{c}_{m,j}$, for $1 \leq j \leq L_m$, from the successive derivatives of s_{N,L_N}, according to Table 3.

In order to shorten notation, let $\mathcal{P}_N := 1$, $\mathcal{C}_N := 1$, $\mathcal{K}_N := 0$ and, for $1 \leq m \leq N-1$, $\mathcal{P}_m := \prod_{i=m+1}^{N} s_{i,L_i-1}$, $\mathcal{C}_m := \prod_{i=m+1}^{N} c_{i,L_i} a_{i,L_i}$ and $\mathcal{K}_m := \sum_{i=m+1}^{N} K_{i,L_i}$.

Example 12 (Example 11 continued) Here we find the monomials relevant for identifiability in the two-layer cascade. We highlight with blue boxes the constants that we are identifying in each derivative. We moreover highlight with green boxes the monomials that we used to identify $b_{1,1}$ and $\tilde{c}_{1,1}$ from $K_{1,1}$ and $\tilde{K}_{1,1}$, respectively (see rows 5 and 6 in Table 3).

$$
\begin{aligned}
\dot{s}_{2,2} &= -\boxed{\tilde{a}_{2,2}} f_2 s_{2,2} + \boxed{\tilde{b}_{2,2}} v_{2,2} + \boxed{c_{2,2}} u_{2,2} \\
\ddot{s}_{2,2} &= -\tilde{a}_{2,2}[(-\tilde{a}_{2,2} f_2 s_{2,2} + \tilde{K}_{2,2} v_{2,2} - \boxed{\tilde{a}_{2,1}} f_2 s_{2,1} + \boxed{\tilde{K}_{2,1}} v_{2,1}) s_{2,2} + f_2 \dot{s}_{2,2}] \\
&\quad + \tilde{b}_{2,2}(\tilde{a}_{2,2} f_2 s_{2,2} - \boxed{\tilde{K}_{2,2}} v_{2,2}) + c_{2,2}(\boxed{a_{2,2}} s_{1,1} s_{2,1} - \boxed{K_{2,2}} u_{2,2}) \\
s^{(3)}_{2,2} &= \sum_{h+i \leq 2} \beta_{f_2,h,i} f_2^{(h)} s^{(i)}_{2,2} + \delta_{v_{2,2}} v_{2,2} + c_{2,2}(a_{2,2}[(\boxed{c_{1,1}} u_{1,1} - \boxed{\tilde{a}_{1,1}} f_1 s_{1,1} \\
&\quad + \boxed{\tilde{b}_{1,1}} v_{1,1} - \boxed{a_{2,1}} s_{1,1} s_{2,0} + \boxed{K_{2,1}} u_{2,1} - a_{2,2} s_{1,1} s_{2,1} + K_{2,2} u_{2,2}) s_{2,1} \\
&\quad + s_{1,1}(\boxed{c_{2,1}} u_{2,1} - a_{2,2} s_{1,1} s_{2,1} + b_{2,2} u_{2,2} + \tilde{c}_{2,2} v_{2,2} - \tilde{a}_{2,1} f_2 s_{2,1} + \boxed{\tilde{b}_{2,1}} v_{2,1})] \\
&\quad - K_{2,2}(a_{2,2} s_{1,1} s_{2,1} - K_{2,2} u_{2,2})) \\
s^{(4)}_{2,2} &= \sum_{h+i \leq 3} \beta_{f_2,h,i} f_2^{(h)} s^{(i)}_{2,2} + \delta_{v_{2,2}} v_{2,2} + c_{2,2}(a_{2,2}[(c_{1,1}(\boxed{a_{1,1}} e s_{1,0} - \boxed{K_{1,1}} \boxed{u_{1,1}}) \\
&\quad - \tilde{a}_{1,1}(\dot{f}_1 s_{1,1} + f_1 \dot{s}_{1,1}) + \tilde{b}_{1,1}(\tilde{a}_{1,1} f_1 s_{1,1} - \boxed{\tilde{K}_{1,1}} \boxed{v_{1,1}}) - a_{2,1}(\dot{s}_{1,1} s_{2,0} + s_{1,1} \dot{s}_{2,0}) \\
&\quad + K_{2,1} \dot{u}_{2,1} - a_{2,2}(\dot{s}_{1,1} s_{2,1} + s_{1,1} \dot{s}_{2,1}) + K_{2,2} \dot{u}_{2,2}) s_{2,1} + 2 \dot{s}_{1,1} \dot{s}_{2,1} + s_{1,1} \ddot{s}_{2,1}] \\
&\quad - K_{2,2}(a_{2,2}[(c_{1,1} \boxed{u_{1,1}} - \tilde{a}_{1,1} f_1 s_{1,1} + \tilde{b}_{1,1} \boxed{v_{1,1}} - a_{2,1} s_{1,1} s_{2,0} + K_{2,1} u_{2,1} \\
&\quad - a_{2,2} s_{1,1} s_{2,1} + K_{2,2} u_{2,2}) s_{2,1} + s_{1,1} \dot{s}_{2,1}] - K_{2,2} \dot{u}_{2,2})
\end{aligned}
$$

5.2 An Example of How to Obtain the Rate Constants from Data

Here, we will illustrate our previous theoretical identifiability results in a specific example, showing how they can be used as a guidance in experimental design for practical parameter identification from observable data.

Table 3 The constants in the cascade can be identified from s_{N,L_N}

Derivative	Monomial	Coefficient (up to sign)	Constant
$s^{(2(N-m)+1)}_{N,L_N}$	$u_{m,L_m}\mathcal{P}_m$	$c_{m,L_m}\mathcal{C}_m$	c_{m,L_m}
	$f_m s_{m,L_m}\mathcal{P}_m$	$\tilde{a}_{m,L_m}\mathcal{C}_m$	$\tilde{a}_{m,L_m}$
	$v_{m,L_m}\mathcal{P}_m$	$\tilde{b}_{m,L_m}\mathcal{C}_m$	$\tilde{b}_{m,L_m}$
$s^{(2(N-m)+2)}_{N,L_N}$	$s_{m-1,L_{m-1}}s_{m,L_m-1}\mathcal{P}_m$	$c_{m,L_m}a_{m,L_m}\mathcal{C}_m$	a_{m,L_m}
	$u_{m,L_m}\mathcal{P}_m$	$c_{m,L_m}(K_{m,L_m}+\mathcal{K}_m)\mathcal{C}_m$	b_{m,L_m}
	$v_{m,L_m}\mathcal{P}_m$	$\tilde{b}_{m,L_m}(\tilde{K}_{m,L_m}+\mathcal{K}_m)\mathcal{C}_m$	$\tilde{c}_{m,L_m}$
	$f_m s_{m,j}s_{m,L_m}\mathcal{P}_m$	$\tilde{a}_{m,j}\tilde{a}_{m,L_m}\mathcal{C}_m$	$\tilde{a}_{m,j}$ $(1\le j\le L_m-1)$
	$v_{m,j}s_{m,L_m}\mathcal{P}_m$	$\tilde{K}_{m,j}\tilde{a}_{m,L_m}\mathcal{C}_m$	$\tilde{K}_{m,j}$ $(1\le j\le L_m-1)$
$s^{(2(N-m)+3)}_{N,L_N}$	$s_{m-1,L_{m-1}}s_{m,j-1}s_{m,L_m-1}\mathcal{P}_m$	$c_{m,L_m}a_{m,L_m}a_{m,j}\mathcal{C}_m$	$a_{m,j}$ $(1\le j\le L_m-1)$
	$u_{m,j}s_{m,L_m-1}\mathcal{P}_m$	$c_{m,L_m}a_{m,L_m}K_{m,j}\mathcal{C}_m$	$K_{m,j}$ $(1\le j\le L_m-1)$
$s^{(2(N-m)+2k+1)}_{N,L_N}$	$s^k_{m-1,L_{m-1}}u_{m,L_m-k}\mathcal{P}_m$	$c_{m,L_m-k}\prod_{j=L_m-k+1}^{L_m}c_{m,j}a_{m,j}\,\mathcal{C}_m$	c_{m,L_m-k}, b_{m,L_m-k}
$1\le k\le L_m-1$	$s^k_{m-1,L_{m-1}}v_{m,L_m-k}\mathcal{P}_m$	$\tilde{b}_{m,L_m-k}\prod_{j=L_m-k+1}^{L_m}c_{m,j}a_{m,j}\,\mathcal{C}_m$	$\tilde{b}_{m,L_m-k}$, $\tilde{c}_{m,L_m-k}$

This table shows the monomials to be considered (column 2) in each of the successive derivatives of s_{N,L_N} (column 1). For each monomial, taking into account the constants already identified, the corresponding coefficient (column 3) enables us to identify the constant appearing in the last column

The three-layer cascade with $L_1 = 1, L_2 = L_3 = 2$ represents the well-known MAPK signaling cascade with $s_{3,2}$ representing the concentration of the doubly phosphorylated kinase MAPK-PP (Catozzi et al. 2016; Huang and Ferrell 1996; Kholodenko 2000; Shaul and Seger 2007). Consider, then, the cascade (6) for $N = 3$ and $L_1 = 1, L_2 = L_3 = 2$, whose schematic representation is introduced at the beginning of Sect. 5. In this case, we have 22 species concentrations $\mathbf{x}$ and 30 rate constants $\mathbf{k}$ which can be identified from $s_{3,2}^{(\ell)}$, $1 \leq \ell \leq 6$, by Theorem 2. According to Definition 3, this means that if we consider the polynomial system

$$s_{3,2}^{(\ell)}(\mathbf{x}, \mathbf{k}) = p_\ell(\mathbf{x}, \mathbf{k}) \tag{7}$$

for the corresponding polynomials p_ℓ obtained from (1) by computing the successive total derivatives of $s_{3,2}$, the function that maps the vector of rate constants $\mathbf{k}$ to the coefficients of the polynomials p_ℓ's (considered as polynomials in the species concentrations $\mathbf{x}$) is injective. This means that all the rate constants can be recovered from noise-free data by a suitable interpolation procedure: if we evaluate these polynomials at "sufficiently many" points $\mathbf{x} \in \mathbb{R}^{22}$, we may reconstruct the coefficients and, consequently, determine uniquely the values of the rate constants.

However, it is not clear which $\mathbf{x} \in \mathbb{R}^{22}$ are suitable for identifying the parameters of the system, nor how many of them are enough for this purpose. We give here a heuristic to choose a list of $\mathbf{x} \in \mathbb{R}^{22}$ based on the monomials in the second column of Table 4, which is the adapted version of Table 3 for this particular case. This heuristic can be used as an aid to design experiments to obtain the rate constants values. Each initial state $\mathbf{x} \in \mathbb{R}^{22}$ is in correspondence with a different experiment.

In order to recover the value of the 30 rate constants in this case, we propose the following algorithm:

Step 1. Consider $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{30} \in \mathbb{R}^{22}$ defined as follows: for the ith monomial in Table 4, consider $\mathbf{x}_i \in \mathbb{R}^{22}$ where all the coordinates are 0 except for those coordinates corresponding to variables that divide the monomial, which are equal to 1. For example, for the monomial $u_{1,1}s_{2,1}s_{3,1}$, all the coordinates of the associated point are equal to 0, except for the three coordinates corresponding to $u_{1,1}$, $s_{2,1}$ and $s_{3,1}$ that are equal to 1.

Step 2. For each $i \in \{1, \ldots, 30\}$, obtain the value $s_{3,2}^{(\ell)}(\mathbf{x}_i, \mathbf{k})$ for the order ℓ that corresponds to the ith monomial in Table 4. Ideally, these values should be obtained experimentally, for instance considering $\mathbf{x}_i$ the initial state at time $t = 0$.

Step 3. Construct a (nonlinear) polynomial equation system from (7), of 30 equations in the 30 unknowns $\mathbf{k}$, by evaluating the right-hand sides at $\mathbf{x}_1, \ldots, \mathbf{x}_{30}$ and replacing the left-hand sides with the values obtained in the previous step.

Step 4. Solve the polynomial system in the unknowns $\mathbf{k}$.

A vague explanation of why this heuristic works is that each monomial in Table 4 incorporates a new variable that comes paired with the new rate constant to be identified. Further research is needed to find a rigorous proof for this conjecture.

We implemented the algorithm above by reconstructing the values of the left-hand sides of (7) with the rate constants in the third column of Table S2 in the Supporting

Information of Qiao et al. (2007). We used Maple (2014) to solve the system of equations and successfully obtained the following values (in a few seconds using a standard desktop computer).

$$
\begin{aligned}
&a_{1,1} = 337.2299998, \quad a_{2,1} = 1226.000001, \quad a_{2,2} = 3383.7,\\
&a_{3,1} = 229.5699981, \quad a_{3,2} = 3388.7, \quad \tilde{a}_{1,1} = 1841.000002, \quad \tilde{a}_{2,1} = 2960.300016,\\
&\tilde{a}_{2,2} = 1956.8, \quad \tilde{a}_{3,1} = 297.0, \quad \tilde{a}_{3,2} = 974.7,\\
&b_{1,1} = 261.1000013, \quad b_{2,1} = 623.1700002, \quad b_{2,2} = 605.3100002,\\
&b_{3,1} = 694.13, \quad b_{3,2} = 485.3499999, \quad \tilde{b}_{1,1} = 198.47, \quad \tilde{b}_{2,1} = 163.0,\\
&\tilde{b}_{2,2} = 48.804, \quad \tilde{b}_{3,1} = 301.09, \quad \tilde{b}_{3,2} = 587.45,\\
&c_{1,1} = 146.07, \quad c_{2,1} = 420.0000001, \quad c_{2,2} = 214.65, \quad c_{3,1} = 43.658,\\
&c_{3,2} = 65.732, \quad \tilde{c}_{1,1} = 338.4400021, \quad \tilde{c}_{2,1} = 668.2000111,\\
&\tilde{c}_{2,2} = 67.97000003, \quad \tilde{c}_{3,1} = 31.743, \quad \tilde{c}_{3,2} = 175.91.
\end{aligned}
$$

The same three-layer cascade may be completely identified also by means of the result stated in Theorem 1: in this case the rate constants in each connected component can be identified from $s_{1,1}^{(\ell_1)}$, $s_{1,0}^{(\ell_2)}$, $s_{2,2}^{(\ell_3)}$, $s_{2,0}^{(\ell_4)}$, $s_{3,2}^{(\ell_5)}$, and $s_{3,0}^{(\ell_6)}$, respectively, for $1 \leq \ell_1, \ell_2 \leq 2$ and $1 \leq \ell_3, \ell_4, \ell_5, \ell_6 \leq 3$. By Corollary 1 we can also identify the constants from $s_{1,1}^{(\ell_1)}$, $s_{2,2}^{(\ell_3)}$ and $s_{3,2}^{(\ell_5)}$, for $1 \leq \ell_1 \leq 2$ and $1 \leq \ell_3, \ell_5 \leq 3$. We adapted the procedure above and implemented it in Maple, and we obtained the same rate constants as before.

Throughout the article, we assume that one can use noise-free data in order to recover the rate constants values. Nevertheless, there are certain numerical errors that appear at Step 4, when the polynomial system in the unknowns $\mathbf{k}$ is solved. If we moreover implement the algorithm with numerical approximations of the total derivatives, more numerical errors are bound to occur. The major drawback of considering the last two approaches, based on Theorem 1 or Corollary 1, is that more species have to be measured. However, the value that has to be numerically estimated corresponds to a derivative of order at most three, which can be approximated more accurately and with fewer time measurements than those values of higher-order derivatives.

The Maple code for both procedures can be found at http://cms.dm.uba.ar/Members/mpmillan/identifiability.

6 Discussion and Further Work

The main contribution of this paper has been to prove that all the rate constants in several well-known chemical reaction networks that are abundant in the literature can be identified from a reduced set of kinetic variables. The work here extends previous results by Craciun and Pantea (2008) and avoids computationally expensive procedures such as differential elimination and Gröbner basis (Bellu et al. 2007; Boulier 2007; Meshkat et al. 2009).

We should point out that we assumed that there is a special partition of the set of chemical species and that every connected component of the chemical reaction network

Table 4 The constants in the three-layer cascade with 30 constants can be identified from $s_{3,2}$

Derivative	Monomial	Coefficient (up to sign)	Constant
$\dot{s}_{3,2}$	$u_{3,2}$	$c_{3,2}$	$c_{3,2}$
	$f_3 s_{3,2}$	$\tilde{a}_{3,2}$	$\tilde{a}_{3,2}$
	$v_{3,2}$	$\tilde{b}_{3,2}$	$\tilde{b}_{3,2}$
$\ddot{s}_{3,2}$	$s_{2,2}s_{3,1}$	$c_{3,2}a_{3,2}$	$a_{3,2}$
	$u_{3,2}$	$c_{3,2}K_{3,2}$	$b_{3,2}$
	$v_{3,2}$	$\tilde{b}_{3,2}\tilde{K}_{3,2}$	$\tilde{c}_{3,2}$
	$f_3 s_{3,1}s_{3,2}$	$\tilde{a}_{3,1}\tilde{a}_{3,2}$	$\tilde{a}_{3,1}$
	$v_{3,1}s_{3,2}$	$\tilde{K}_{3,1}\tilde{a}_{3,2}$	$\tilde{K}_{3,1}$
$s^{(3)}_{3,2}$	$s_{2,2}s_{3,0}s_{3,1}$	$c_{3,2}a_{3,2}a_{3,1}$	$a_{3,1}$
	$u_{3,1}s_{3,1}$	$c_{3,2}a_{3,2}K_{3,1}$	$K_{3,1}$
	$s_{2,2}u_{3,1}$	$c_{3,1}c_{3,2}a_{3,2}$	$c_{3,1}, b_{3,1}$
	$v_{3,1}s_{2,2}$	$\tilde{b}_{3,1}c_{3,2}a_{3,2}$	$\tilde{b}_{3,1}, \tilde{c}_{3,1}$
	$u_{2,2}s_{3,1}$	$c_{2,2}c_{3,2}a_{3,2}$	$c_{2,2}$
	$f_2 s_{2,2}s_{3,1}$	$\tilde{a}_{2,2}c_{3,2}a_{3,2}$	$\tilde{a}_{2,2}$
	$v_{2,2}s_{3,1}$	$\tilde{b}_{2,2}c_{3,2}a_{3,2}$	$\tilde{b}_{2,2}$
$s^{(4)}_{3,2}$	$s_{1,1}s_{2,1}s_{3,1}$	$c_{2,2}a_{2,2}c_{3,2}a_{3,2}$	$a_{2,2}$
	$u_{2,2}s_{3,1}$	$c_{2,2}(K_{2,2}+K_{3,2})\mathcal{C}_2$	$b_{2,2}$
	$v_{2,2}s_{3,1}$	$\tilde{b}_{2,2}(\tilde{K}_{2,2}+K_{3,2})\mathcal{C}_2$	$\tilde{c}_{2,2}$
	$f_2 s_{2,1}s_{2,2}s_{3,1}$	$\tilde{a}_{2,1}\tilde{a}_{2,2}\mathcal{C}_2$	$\tilde{a}_{2,1}$
	$v_{2,1}s_{2,2}s_{3,1}$	$\tilde{K}_{2,1}\tilde{a}_{2,2}\mathcal{C}_2$	$\tilde{K}_{2,1}$
$s^{(5)}_{3,2}$	$u_{1,1}s_{2,1}s_{3,1}$	$c_{1,1}\mathcal{C}_1$	$c_{1,1}$
	$f_1 s_{1,1}s_{2,1}s_{3,1}$	$\tilde{a}_{1,1}\mathcal{C}_1$	$\tilde{a}_{1,1}$
	$v_{1,1}s_{2,1}s_{3,1}$	$\tilde{b}_{1,1}\mathcal{C}_1$	$\tilde{b}_{1,1}$
	$s_{1,1}s_{2,0}s_{2,1}s_{3,1}$	$c_{2,2}a_{2,2}a_{2,1}\mathcal{C}_2$	$a_{2,1}$
	$u_{2,1}s_{2,1}s_{3,1}$	$c_{2,2}a_{2,2}K_{2,1}\mathcal{C}_2$	$K_{2,1}$
	$s_{1,1}u_{2,1}s_{3,1}$	$c_{2,1}c_{2,2}a_{2,2}a_{3,2}c_{3,2}$	$c_{2,1}, b_{2,1}$
	$s_{1,1}v_{2,1}s_{3,1}$	$\tilde{b}_{2,1}\mathcal{C}_1$	$\tilde{b}_{2,1}$,$\tilde{c}_{2,1}$
$s^{(6)}_{3,2}$	$es_{1,0}s_{2,1}s_{3,1}$	$c_{1,1}a_{1,1}\mathcal{C}_1$	$a_{1,1}$
	$u_{1,1}s_{2,1}s_{3,1}$	$c_{1,1}(K_{1,1}+K_{2,2}+K_{3,2})\mathcal{C}_1$	$b_{1,1}$
	$v_{1,1}s_{2,1}s_{3,1}$	$\tilde{b}_{1,1}(\tilde{K}_{1,1}+K_{2,2}+K_{3,2})\mathcal{C}_1$	$\tilde{c}_{1,1}$

This table shows the monomials to be considered (column 2) in each of the successive derivatives of $s_{3,2}$ (column 1). For each monomial, taking into account the constants already identified, the corresponding coefficient (column 3) enables us to identify the constant appearing in the last column. Here, we consider $\mathcal{C}_1 = c_{2,2}a_{2,2}c_{3,2}a_{3,2}$ and $\mathcal{C}_2 = c_{3,2}a_{3,2}$

has a particular shape (see Sect. 3). Both assumptions are natural when modeling multisite phosphorylation systems and signaling cascades (Wang and Sontag 2008; Huang and Ferrell 1996). We have then shown, in Sect. 4, how to identify the rate constants in every connected component, or two related connected components, from a single species. In Sect. 5 we have moreover proved that all the rate constants in signaling cascade networks can be identified from only one species: the last product of the first component of the last layer. Additionally, we have presented in Sect. 5.2 an example showing how to compute the values of the rate constants from noise-free data according to our theoretical results in the previous sections. The procedure is based on a heuristic to choose the right input data; it would be of great interest to find a formal proof for establishing a good set of sufficient data for any network of the class considered in this paper.

We expect that the techniques used in this paper could be applied for identifiability from a few variables to a number of modifications of the networks we have considered here. For instance, it would be interesting to introduce more intermediate complexes within different reactions. Another potential adaptation is relaxing the assumption $F_i \neq F_j$ for $i \neq j$ in the cascade network, and allowing for repetition of these enzymes. Both modifications are natural extensions of the networks we have analyzed, and we conjecture that similar results can be obtained. We moreover would like to apply our techniques to more general but hence well structured networks such as MESSI networks (Pérez Millán and Dickenstein 2018). Another future research direction is to characterize which other variables can be considered to identify the rate constants of either a whole connected component or the entire biochemical network.

Acknowledgements The authors wish to thank the anonymous referees for their thoughtful comments which helped to improve the manuscript.

A Proofs

Throughout this "Appendix," we maintain the notation and assumptions introduced in Sects. 2 and 3.

Before stating and proving our results, we introduce some further notation and formulas we will use in our analysis. We consider an autonomous dynamical system

$$\dot{\mathbf{x}} = \sum_{y \to y'} k_{yy'} \mathbf{x}^y (y' - y), \tag{8}$$

arising from a chemical reaction network satisfying the assumptions stated in Sect. 3.

For a non-intermediate species X, let

$$\mathscr{Z}_X = \{Z : Z \text{ reacts with } X\} \quad \text{and} \quad \mathscr{W}_X = \{W : W \text{ reacts to } X\}. \tag{9}$$

By the shape of the networks we consider, $\mathscr{Z}_X$ is a set of non-intermediate species and $\mathscr{W}_X$ is a set of intermediate species. From (8), we then have that

$$\dot{x} = -\sum_{Z\in\mathscr{Z}_X} \mu_z x z + \sum_{W\in\mathscr{W}_X} \eta_w w, \tag{10}$$

for suitable non negative real numbers μ_z and η_w. For $\ell \geq 2$, Leibniz rule implies that

$$x^{(\ell)} = -\sum_{Z\in\mathscr{Z}_X} \mu_z \sum_{h+i=\ell-1} \binom{\ell-1}{h} x^{(h)} z^{(i)} + \sum_{W\in\mathscr{W}_X} \eta_w w^{(\ell-1)}. \tag{11}$$

If $W \in \mathscr{W}_X$ is involved in a block of reactions

$$Z_{w,1} + Z_{w,2} \underset{b_w}{\overset{a_w}{\rightleftarrows}} W \xrightarrow{c_w} Z_{w,3} + X,$$

then, according to (8), the differential equation $\dot{w} = a_w z_{w,1} z_{w,2} - K_w w$, with $K_w = b_w + c_w$, is satisfied, and

$$w^{(\ell-1)} = \sum_{h+i\leq\ell-2} (-K_w)^{\ell-2-h-i} a_w \binom{h+i}{h} z_{w,1}^{(h)} z_{w,2}^{(i)} + (-K_w)^{\ell-1} w. \tag{12}$$

By separating the cases where $X \in \{Z_{w,1}, Z_{w,2}\}$ and $X \notin \{Z_{w,1}, Z_{w,2}\}$, we can simplify:

$$x^{(\ell)} = \sum_{Z\in\mathscr{Z}_X} \sum_{h+i\leq\ell-1} \beta_{z,h,i}\, x^{(h)} z^{(i)} + \sum_{\substack{W\in\mathscr{W}_X \\ X\notin\{Z_{w,1},Z_{w,2}\}}} \sum_{h+i\leq\ell-2} \gamma_{w,h,i}\, z_{w,1}^{(h)} z_{w,2}^{(i)} + \sum_{w\in\mathscr{W}_X} \delta_w\, w, \tag{13}$$

for suitable real numbers $\beta_{z,h,i}$, $\gamma_{w,h,i}$ and δ_w that depend on ℓ and the reaction rate constants.

From the previous formulas interpreted as polynomials in the variables x, z, w, we deduce straightforwardly.

Lemma 1 *For a reaction network satisfying the assumptions of Sect. 3, we have:*

1. *The constant monomial does not appear in any derivative of any species.*
2. *The only monomials of degree* 1 *appearing in a derivative* $x^{(\ell)}$, $\ell \geq 1$, *for a non-intermediate species* X, *are the monomials* w *corresponding to* $W \in \mathscr{W}_X$, *that is, those that appear in* $\dot{x}$.

A.1 Proofs of Sect. 4.1: Identifying the Constants in One Connected Component from One Variable

Here we give the proofs of our identifiability result for a connected component of the type:

$$Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} \dots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \overset{c_L}{\to} Y + S_L \tag{14}$$

We maintain the hypotheses and notations introduced in Sect. 3 and previously in this "Appendix."

Lemma 2 *Given a connected component as in* (14)*, the constants a_L, b_L and c_L can be identified from $\dot{s}_L$ and $\ddot{s}_L$, and, if $L > 1$, the constants a_j and $K_j := b_j + c_j$, for $1 \le j \le L-1$, can be identified from $\dot{s}_L, \ddot{s}_L$ and $s_L^{(3)}$.*

Proof Following (10), we have

$$\dot{s}_L = -\sum_{Z \in \mathscr{Z}_L} \mu_z s_L z + \sum_{W \in \mathscr{W}_L} \eta_w w \tag{15}$$

where, using the notation in (9), $\mathscr{Z}_L := \mathscr{Z}_{S_L}$ and $\mathscr{W}_L := \mathscr{W}_{S_L}$. By separating the term corresponding to $U_L \in \mathscr{W}_L$, we obtain

$$\dot{s}_L = -\sum_{Z \in \mathscr{Z}_L} \mu_z s_L z + c_L u_L + \sum_{W \in \mathscr{W}_L^*} \eta_w w$$

where $\mathscr{W}_L^* := \mathscr{W}_L \backslash \{U_L\}$. Then, we can identify c_L from $\dot{s}_L$ as the coefficient of the monomial u_L.

Consider now

$$\ddot{s}_L = -\sum_{Z \in \mathscr{Z}_L} \mu_z [\dot{s}_L z + s_L \dot{z}] + c_L (a_L y s_{L-1} - K_L u_L) + \sum_{W \in \mathscr{W}_L^*} \eta_w [a_w z_{w,1} z_{w,2} - K_w w].$$

From this expression, since $c_L \neq 0$, we can identify a_L and K_L from the coefficients of the monomials $y s_{L-1}$ and u_L (which only appear in $\ddot{s}_L$ from the derivative $\dot{u}_L$) and, as we know c_L, we can also identify b_L. If $L = 1$ we have identified all the constants.

If $L > 1$, consider the third derivative

$$\begin{aligned} s_L^{(3)} = &\sum_{Z \in \mathscr{Z}_L} \sum_{h+i \le 2} \beta_{z,h,i}\, s_L^{(h)} z^{(i)} + c_L a_L [\dot{y} s_{L-1} + y \dot{s}_{L-1}] - c_L K_L (a_L y s_{L-1} - K_L u_L) \\ &+ \sum_{\substack{W \in \mathscr{W}_L^* \\ S_L \notin \{Z_{w,1}, Z_{w,2}\}}} \gamma_{w,1} [\dot{z}_{w,1} z_{w,2} + z_{w,1} \dot{z}_{w,2}] + \gamma_{w,0} z_{w,1} z_{w,2} + \delta_w w. \end{aligned} \tag{16}$$

The constants a_j and K_j, for $1 \leq j \leq L-1$, appear in $\dot{y} = \sum_{1\leq j\leq L}(-a_j y s_{j-1} + K_j u_j) + \cdots$ as the coefficients (up to sign) of the monomials $y s_{j-1}$ and u_j, respectively. Then, they appear in the expression (16) from the product $\dot{y} s_{L-1}$ in the coefficients of the monomials $y s_{j-1} s_{L-1}$ and $u_j s_{L-1}$, for $1 \leq j \leq L-1$. We will now look for these monomials in the whole expression (16) and show that they come only from the product $\dot{y} s_{L-1}$.

As $Y \notin \mathscr{Z}_L$ and, for every $Z \in \mathscr{Z}_L$, by Assumption 2, we have $Z \neq S_l$ for all $0 \leq l \leq L-1$, the monomials $y s_{j-1} s_{L-1}$ and $u_j s_{L-1}$, for $1 \leq j \leq L-1$, do not appear in $s_L^{(h)} z$, for $0 \leq h \leq 2$. Also, it is clear that they do not appear in $s_L z^{(i)}$, for $0 \leq i \leq 2$. On the other hand, every monomial of degree 3 that appears in a product of two derivatives of order 1 is a multiple of an intermediate; so, $y s_{j-1} s_{L-1}$ does not appear in $\dot{s}_L \dot{z}$ and, by Lemma 1, $u_j s_{L-1}$ does not appear either since no derivative contains a constant term or the degree one monomial s_{L-1}.

Now, consider $W \in \mathscr{W}_L^*$ such that $S_L \notin \{Z_{w,1}, Z_{w,2}\}$, and the corresponding block of reactions $Z_{w,1} + Z_{w,2} \rightleftarrows W \to Z_{w,3} + S_L$. Since $U_j \notin \mathscr{W}_L^*$ for every $1 \leq j \leq L$, then $Z_{w,1} + Z_{w,2} \neq Y + S_{j-1}$. Also, by Assumption 2, $Z_{w,1} + Z_{w,2} \neq S_l + S_{L-1}$ for every $0 \leq l \leq L$. Every monomial in $\dot{z}_{w,1} z_{w,2}$ is either of the form $w_0 z_{w,2}$ for an intermediate W_0 that reacts to $Z_{w,1}$ or of the form $z_0 z_{w,1} z_{w,2}$ for a non-intermediate Z_0 that reacts with $Z_{w,1}$. If $z_0 z_{w,1} z_{w,2} = y s_{j-1} s_{L-1}$, it follows that $Z_{w,1} + Z_{w,2} \in \{Y + S_{j-1}, Y + S_{L-1}, S_{j-1} + S_{L-1}\}$, leading to a contradiction. If $w_0 z_{w,2} = u_j s_{L-1}$, then $Z_{w,2} = S_{L-1}$ and U_j reacts to $Z_{w,1}$, meaning that $Z_{w,1} \in \{Y, S_{j-1}, S_j\}$, which is not possible.

Finally, the monomial $y s_{j-1} s_{L-1}$ does not appear in $y \dot{s}_{L-1}$ since, by Assumption 2, S_{j-1} does not react with S_{L-1} for every j.

We conclude that, for $1 \leq j \leq L-1$, the coefficients in $s_L^{(3)}$ of the monomials $y s_{j-1} s_{L-1}$ and $u_j s_{L-1}$ are $-c_L a_L a_j$ and $c_L a_L K_j$, respectively. As we have already identified c_L and a_L, these coefficients enable us to identify a_j and $K_j = b_j + c_j$, for $1 \leq j \leq L-1$. □

We show now some auxiliary results concerning the behavior of monomials appearing in the successive derivatives of some variables and their relations with the reaction network. They will allow us to prove Lemma 5, the key recursive tool to show the identifiability results of Sect. 4.1.

Lemma 3 *If $\prod_{i=1}^m z_i$, with $m \geq 2$, is a monomial of $x^{(\ell)}$ where Z_i is a non-intermediate species for every i, then there exist $1 \leq i_1 < i_2 \leq m$ such that Z_{i_1} reacts with Z_{i_2}.*

Proof If $\ell = 1$, this is true. Assume $\ell \geq 2$. Recalling that $x^{(\ell)} = \sum_v \frac{\partial x^{(\ell-1)}}{\partial v} \dot{v}$ (where the sum runs over all variables v representing non-intermediates or intermediate species), it follows that $\prod_{i=1}^m z_i$ is a monomial in $\frac{\partial x^{(\ell-1)}}{\partial v} \dot{v}$ for some variable v. Since every monomial appearing in $\dot{v}$ is either a single intermediate or a product of two non-intermediate species that react together, the result follows. □

Corollary 2 *If X is a non-intermediate species, no derivative $x^{(\ell)}$ for $\ell \geq 1$ contains a monomial which is a pure power of degree $m \geq 2$ of a variable corresponding to a non-intermediate species.*

Lemma 4 *Given an intermediate species U and non-intermediate species X and Y such that $Y \neq X$, if a monomial $y^r u$, $r \geq 0$, appears in $x^{(\ell)}$ for some $\ell \geq 1$, then either U reacts to X or $\ell \geq 2$, the network contains a block of reactions*

$$Y + Z_w \rightleftarrows W \to \widetilde{Z}_w + X, \tag{17}$$

where $Z_w \neq X$, and a monomial $y^t u$ with $t < r$ appears in $z_w^{(i)}$ for some $i \leq \ell - 2$. If, in addition, Y acts as an enzyme in all the reactions of the connected component determined by U, then $X \in \mathscr{S}_U$ and the block of reactions in (17) is $Y + Z_w \rightleftarrows W \to Y + X$, and it is contained in the connected component determined by U.

Moreover, if U does not react to X and ℓ is the smallest integer such that a monomial $y^r u$ appears in $x^{(\ell)}$, then $r \geq 1$, $\ell \geq 2$, and the monomial $y^{r-1}u$ appears in $z_w^{(i)}$ for some $i \leq \ell - 2$.

Proof We prove the first part by induction on r. If $r = 0$, then u appears in $x^{(\ell)}$ for some $\ell \geq 1$; by Lemma 1 (2), this is equivalent to the fact that U reacts to X. In particular, if $Y \neq X$ acts as an enzyme in the connected component determined by U, then $X \in \mathscr{S}_U$.

Now, if $r \geq 1$, since no monomial $y^r u$ with $r \geq 1$ appears in $\dot{x}$, it follows that $\ell \geq 2$. Then, by identity (13), the monomial $y^r u$ can only appear in a product of derivatives of two species, and by Lemma 1 and Corollary 2, one of these species must be Y and the corresponding order of derivation must be zero.

If $y^r u$ appears in a product $x^{(h)} z^{(i)}$ for some $Z \in \mathscr{Z}_X$ and $h + i \leq \ell - 1$, as $X \neq Y$, then $Z = Y$ and $y^{r-1}u$ appears in $x^{(h)}$; then, the result follows by the inductive hypothesis.

Finally, if $y^r u$ appears in a product $z_{w,1}^{(h)} z_{w,2}^{(i)}$ with $h + i \leq \ell - 2$ for some $W \in \mathscr{W}_X$ such that $X \notin \{Z_{w,1}, Z_{w,2}\}$, again by Lemma 1 and Corollary 2, we may assume that $Y = Z_{w,1}$ and $y^{r-1}u$ appears in $z_{w,2}^{(i)}$. Since $X \neq Z_{w,2}$ and W reacts to X, we must have $Y + Z_{w,2} \rightleftarrows W \to \widetilde{Z}_w + X$ for some species $\widetilde{Z}_w$, that is, a block of reactions as in (17). By the induction hypothesis applied to the non-intermediate $Z_{w,2} \neq Y$, if Y acts as an enzyme in the connected component determined by U, it follows that $Z_{w,2} \in \mathscr{S}_U$. Then, $Y + Z_{w,2}$ is a complex in the connected component determined by U, where Y acts as an enzyme. As $X \neq Y$, necessarily $\widetilde{Z}_w = Y$ and $X \in \mathscr{S}_U$.

To see that the last statement of the lemma holds, note that if U does not react to X and a monomial $y^r u$ appears in a derivative $x^{(\ell)}$, then $r \geq 1$ and $\ell \geq 2$ and, by assuming ℓ minimal, the only possibility in the above reasoning is the last one. □

Now, we are able to prove the key lemma for the proof of our main result on the identifiability of constants in a single connected component. We keep our previous notation and assumptions.

For technical reasons, we define the empty product of factors α_i as $\prod_{i=0}^{-1} \alpha_i = 1$.

Lemma 5 *Given a connected component as in (14), with $L \geq 1$, let $1 \leq n \leq L$ and $0 \leq k \leq n - 1$ be fixed. If ℓ is minimum such that $y^r u_{n-k}$ is a monomial of $s_n^{(\ell)}$ for*

some $r \geq 0$, then $\ell = 2k+1$, $r = k$ and the coefficient of $y^k u_{n-k}$ in $s_n^{(2k+1)}$ is

$$c_{n-k} \prod_{j=0}^{k-1} a_{n-j}\, c_{n-j}.$$

Proof For $k = 0$, first notice that, for all $1 \leq n \leq L$, as U_n reacts to S_n, then u_n appears in $\dot{s}_n$ and so, $\ell = 1$, $r = 0$, and the coefficient of u_n is c_n, as we wanted to prove.

We follow the proof by induction on n.

If $n = 1$, the only possibility is $k = 0$, which we have already proven.

Assume now $n \geq 2$, and let $k \geq 1$. If a monomial $y^r u_{n-k}$ appears in $s_n^{(\ell)}$ and considering ℓ minimal, as U_{n-k} does not react to S_n, by Lemma 4 applied to $U := U_{n-k}$ and $X := S_n$, the network contains a block of reactions $Y + Z_w \rightleftarrows W \to Y + S_n$, and the monomial $y^{r-1}u_{n-k}$ appears in $z_w^{(i)}$ for some $i \leq \ell - 2$. This block of reactions is necessarily $Y + S_{n-1} \rightleftarrows U_n \to Y + S_n$ and so, $y^{r-1}u_{n-k}$ appears in $s_{n-1}^{(i)}$ for some $i \leq \ell - 2$. Moreover, by formula (13) applied to $x = s_n$, the only terms contributing to the monomial $y^r u_{n-k}$ come from products $y s_{n-1}^{(i)}$ with $i \leq \ell - 2$. Since $y^{r-1}u_{n-k} = y^{r-1}u_{(n-1)-(k-1)}$, by the induction hypothesis, $i \geq 2(k-1)+1 = 2k-1$; then, $\ell - 2 \geq 2k - 1$ or, equivalently, $\ell \geq 2k + 1$.

Consider now formula (13) for $s_n^{(2k+1)}$. The only product of derivatives where a monomial $y^r u_{n-k}$ may appear is $y s_{n-1}^{(2k-1)}$, since $i \leq 2k - 1$ for all derivatives $s_{n-1}^{(i)}$ involved. Then, the coefficient of $y^r u_{n-k}$ in $s_n^{(2k+1)}$ equals $\gamma_{u_n,0,2k-1}$ multiplied by the coefficient of $y^{r-1}u_{n-k}$ in $s_{n-1}^{(2k-1)}$. By the induction hypothesis, a monomial $y^{r-1}u_{n-k}$ appears with nonzero coefficient in $s_{n-1}^{(2k-1)}$ if and only if $r - 1 = k - 1$, that is $r = k$, and the corresponding coefficient is $c_{n-1-(k-1)} \prod_{j=0}^{k-2} a_{n-1-j} c_{n-1-j}$. To determine $\gamma_{u_n,0,2k-1}$ note that, by formula (12) applied to u_n, the product $y s_{n-1}^{(2k-1)}$ appears in $u_n^{(2k)}$ multiplied by a_n and, by formula (11), $u_n^{(2k)}$ appears in $s_n^{(2k+1)}$ multiplied by c_n; then, $\gamma_{u_n,0,2k-1} = c_n a_n$.

Summarizing, the monomial $y^k u_{n-k}$ appears with nonzero coefficient in $s_n^{(2k+1)}$; hence, $\ell = 2k + 1$. Moreover, it is the only monomial of the form $y^r u_{n-k}$ effectively appearing in $s_n^{(2k+1)}$, and its corresponding coefficient is $c_n a_n c_{n-k} \prod_{j=0}^{k-2} a_{n-1-j}$ $c_{n-1-j} = c_{n-k} \prod_{j=0}^{k-1} a_{n-j}\, c_{n-j}$. □

Remark 2 An interesting fact is that the previous lemmas also hold for networks where not all the reactions are enzymatic. By this we mean that the blocks of reactions are of the form:

$$X_1 + X_2 \underset{b}{\overset{a}{\rightleftarrows}} U \xrightarrow{c} X_3 + X_4,$$

with $X_1 \neq X_2$, $X_3 \neq X_4$ but not necessarily $\{X_1, X_2\} \cap \{X_3, X_4\} \neq \emptyset$.

Combining Lemmas 2 and 5, we may now prove the main result of Sect. 4.1 (Proposition 2 in the main text):

Proposition 4 *All the constants in a connected component as* (14) *of a network satisfying the assumptions in Sect.* 3 *can be identified from* $s_L^{(\ell)}$ *with* $1 \le \ell \le \max\{2, 2L-1\}$.

Proof By Lemma 2, we can identify a_L, b_L and c_L from $\dot{s}_L$ and $\ddot{s}_L$, which implies the statement of the proposition for $L = 1$.

For $L \ge 2$, again by Lemma 2, we can also identify a_j and $K_j = b_j + c_j$, for $1 \le j \le L-1$, from $s_L^{(3)}$. In order to identify all the constants, we need to "separate" b_j from c_j for $1 \le j \le L-1$. We do this by identifying the constants c_{L-k} recursively, for $k = 1, \ldots, L-1$, from the successive derivatives of s_L.

Let $k \ge 1$ and assume c_{L-j} has been identified, for $0 \le j < k$. By Lemma 5, the coefficient of the monomial $y^k u_{L-k}$ in $s_L^{(2k+1)}$ is $c_{L-k} \prod_{j=0}^{k-1} a_{L-j}\, c_{L-j}$. As a_{L-j} and c_{L-j} for $0 \le j \le L-1$ are known, from this coefficient we identify c_{L-k}. □

A.2 Proofs of Sect. 4.2: Identifying the Constants in Two Connected Components from One Variable

The case of two connected components of the type

$$\begin{aligned} &Y + S_0 \underset{b_1}{\overset{a_1}{\rightleftarrows}} U_1 \overset{c_1}{\to} Y + S_1 \underset{b_2}{\overset{a_2}{\rightleftarrows}} U_2 \overset{c_2}{\to} \ldots Y + S_{L-1} \underset{b_L}{\overset{a_L}{\rightleftarrows}} U_L \overset{c_L}{\to} Y + S_L, \\ &\widetilde{Y} + S_L \underset{\tilde{b}_L}{\overset{\tilde{a}_L}{\rightleftarrows}} V_L \overset{\tilde{c}_L}{\to} \widetilde{Y} + S_{L-1} \underset{\tilde{b}_{L-1}}{\overset{\tilde{a}_{L-1}}{\rightleftarrows}} V_{L-1} \overset{\tilde{c}_{L-1}}{\to} \ldots \widetilde{Y} + S_1 \underset{\tilde{b}_1}{\overset{\tilde{a}_1}{\rightleftarrows}} V_1 \overset{\tilde{c}_1}{\to} \widetilde{Y} + S_0 \end{aligned} \tag{18}$$

considered in the paper runs in a similar way than the one connected component case. The first result concerning this class of networks is in the spirit of Lemma 2.

Lemma 6 *Given two connected components as in* (18)*, the constants* $\tilde{a}_L$, $\tilde{b}_L$, $\tilde{c}_L$, *and* $\tilde{a}_j$, $\tilde{K}_j := \tilde{b}_j + \tilde{c}_j$, *for* $1 \le j \le L-1$, *can be identified from* $\dot{s}_L$ *and* $\ddot{s}_L$.

Proof Consider the formula for $\dot{s}_L$ given in (15). Separating the terms corresponding to $\widetilde{Y} \in \mathscr{Z}_L$ and $V_L \in \mathscr{W}_L$, and writing $\mathscr{Z}_L^\times := \mathscr{Z}_L \backslash \{\widetilde{Y}\}$ and $\mathscr{W}_L^\times := \mathscr{W}_L \backslash \{V_L\}$, we obtain

$$\dot{s}_L = -\sum_{Z \in \mathscr{Z}_L^\times} \mu_z s_L z + \sum_{W \in \mathscr{W}_L^\times} \eta_w w - \tilde{a}_L \tilde{y} s_L + \tilde{b}_L v_L.$$

Then, we can identify $\tilde{a}_L$ and $\tilde{b}_L$ as the coefficients (up to sign) of the monomials $\tilde{y} s_L$ and v_L in $\dot{s}_L$.

Consider now

$$\begin{aligned} \ddot{s}_L = &-\sum_{Z \in \mathscr{Z}_L^\times} \mu_z [\dot{s}_L z + s_L \dot{z}] + \sum_{W \in \mathscr{W}_L^\times} \eta_w [a_w z_{w,1} z_{w,2} - K_w w] \\ &- \tilde{a}_L [\dot{\tilde{y}} s_L + \tilde{y} \dot{s}_L] + \tilde{b}_L [\tilde{a}_L \tilde{y} s_L - \tilde{K}_L v_L]. \end{aligned} \tag{19}$$

From the coefficient of v_L in $\ddot{s}_L$, we can identify $\tilde{K}_L$ and, therefore, $\tilde{c}_L$, since we have already identified $\tilde{b}_L$. The constants $\tilde{a}_j$ and $\tilde{K}_j$, for $1 \leq j \leq L-1$, appear in the derivative

$$\dot{\tilde{y}} = \sum_{1 \leq j \leq L} (-\tilde{a}_j \tilde{y} s_j + \tilde{K}_j v_j) + \cdots,$$

then, they appear in the expression (19) from the product $\dot{\tilde{y}} s_L$ in the coefficients of the monomials $\tilde{y} s_j s_L$ and $v_j s_L$, respectively. By Assumption 2, $S_j \notin \mathscr{Z}_L$ for every $1 \leq j \leq L-1$; hence, the monomials $\tilde{y} s_j s_L$ do not come from any other term in (19). Also, it is immediate that the monomials $v_j s_L$ only come from $\dot{\tilde{y}} s_L$. Then, the coefficients of $\tilde{y} s_j s_L$ and $v_j s_L$ in $\ddot{s}_L$ are $\tilde{a}_L \tilde{a}_j$ and $-\tilde{a}_L \tilde{K}_j$, respectively, and enable us to identify $\tilde{a}_j$ and $\tilde{K}_j$, for $1 \leq j \leq L-1$, since $\tilde{a}_L \neq 0$. □

In order to establish a statement extending Lemma 5 to this new setting, we need a previous technical lemma (a suitable analogue of Lemma 4):

Lemma 7 *Given an intermediate species V and a non-intermediate species Y that acts as an enzyme in a connected component where the set of substrates and products is $\mathscr{S}_V$, if X is a non-intermediate species such that $X \in \mathscr{S}^{(\alpha)}$ for some $\alpha \geq 1$ and $Y, \widetilde{Y} \notin \mathscr{S}^{(\alpha)}$, where $\widetilde{Y}$ is the enzyme in the connected component determined by V, and the monomial $y^r v$ appears in $x^{(\ell)}$ for some $r \geq 0$ and $\ell \geq 1$, then $X \in \mathscr{S}_V$.*

Moreover, either V reacts to X or $r \geq 1$, $\ell \geq 2$ and a monomial $y^t v$ with $t < r$ appears in $z_w^{(i)}$, for some $i \leq \ell - 2$, for a species Z_w involved in a block of reactions $Y + Z_w \rightleftarrows W \to Y + X$. If $r \geq 1$ and ℓ is minimal, then $t = r - 1$.

Proof First, note that $X \neq Y$ and $X \neq \widetilde{Y}$, because of the assumption that $X \in \mathscr{S}^{(\alpha)}$ and $Y, \widetilde{Y} \notin \mathscr{S}^{(\alpha)}$. We proceed by induction on $r \in \mathbb{N}_0$.

If $r = 0$, by Lemma 1 (2), V reacts to X. As X is not the enzyme $\widetilde{Y}$, then $X \in \mathscr{S}_V$.

For $r \geq 1$, since $X \neq Y$, Lemma 4 states that either V reacts to X (which we have already considered) or the network contains a block of reactions $Y + Z_w \rightleftarrows W \to \widetilde{Z}_w + X$, where $Z_w \neq X$, and a monomial $y^t v$ with $t < r$ appears in $z_w^{(i)}$ for some $i \leq \ell - 2$ (furthermore, $t = r - 1$ if ℓ is minimal). In the latter case, $\widetilde{Z}_w$ acts as an enzyme in the connected component determined by W and $X \in \mathscr{S}_W$, which implies that $\mathscr{S}_W \subset \mathscr{S}^{(\alpha)}$. If $\widetilde{Z}_w = Z_w$, then $Y \in \mathscr{S}_W$, contradicting the assumption that $Y \notin \mathscr{S}^{(\alpha)}$; therefore, $\widetilde{Z}_w = Y$, and $Z_w \in \mathscr{S}^{(\alpha)}$. By the induction hypothesis, $Z_w \in \mathscr{S}_V$. As $\mathscr{S}_V$ is the set of substrates and products in a connected component where Y acts as an enzyme, the complex $Y + Z_w$ lies in that component, and so, $X \in \mathscr{S}_V$. □

We are now able to prove the result that will play the key role in order to give a recursive argument to identify all the constants in suitable pairs of connected components:

Lemma 8 *Given two connected components as in (18) with $L \geq 1$, let $1 \leq n \leq L$ and $0 \leq k \leq n-1$ be fixed. If ℓ is minimum such that $y^r v_{n-k}$ is a monomial of $s_n^{(\ell)}$ for*

some $r \geq 0$, then $\ell = 2k+1$, $r = k$ and the coefficient of $y^k v_{n-k}$ in $s_n^{(2k+1)}$ is

$$\tilde{b}_{n-k} \prod_{j=0}^{k-1} a_{n-j} c_{n-j}.$$

Proof For $k = 0$, and all $1 \leq n \leq L$, v_n appears in $\dot{s}_n$ (since V_n reacts to S_n) with coefficient $\tilde{b}_n$ and so, $r = 0$ and $\ell = 1$. We now proceed by induction on n.

If $n = 1$, the only possibility is $k = 0$, which has already been considered.

For $n \geq 2$, let $k \geq 1$. By Assumption 2, there exists $\alpha \geq 1$ such that $S_j \in \mathscr{S}^{(\alpha)}$ for every $0 \leq j \leq L$, and $Y, \widetilde{Y} \notin \mathscr{S}^{(\alpha)}$. If the monomial $y^r v_{n-k}$ appears in a derivative of s_n and ℓ is the minimum derivation order where it appears, as V_{n-k} does not react to S_n, by Lemma 7, $r \geq 1$, $\ell \geq 2$ and the monomial $y^{r-1} v_{n-k}$ appears in $z_w^{(i)}$, for some $i \leq \ell - 2$, for a species Z_w in a block of reactions $Y + Z_w \rightleftarrows W \to Y + S_n$. Then, $W = U_n$ and $Z_w = S_{n-1}$; so, $y^{r-1} v_{n-k}$ appears in $s_{n-1}^{(i)}$ for some $i \leq \ell - 2$. By the induction hypothesis, we have that $i \geq 2k-1$; therefore, $\ell \geq 2k+1$.

Now, following *mutatis mutandis* the proof of Lemma 5, we deduce that the coefficient of the monomial $y^k v_{n-k}$ in $s_n^{(2k+1)}$ is equal to $c_n a_n$ multiplied by the coefficient of $y^{k-1} v_{n-k}$ in $s_{n-1}^{(2k-1)}$, and we conclude by applying the induction hypothesis. □

Similarly as in the previous subsection, from Lemmas 6 and 8 we deduce the following identifiability result for two connected components that extends Proposition 4 and constitutes the main result in Sect. 4.2 (Proposition 3 in the main text):

Proposition 5 *Given a chemical reaction network satisfying the assumptions in Sect. 3, all the constants in two connected components as in (18) can be identified from $s_L^{(\ell)}$ with $1 \leq \ell \leq \max\{2, 2L-1\}$.*

Proof The result holds for $L = 1$, since by Lemmas 2 and 6, we can identify $a_L, b_L, c_L, \tilde{a}_L, \tilde{b}_L$ and $\tilde{c}_L$ from $\dot{s}_L$ and $\ddot{s}_L$.

Assume now $L \geq 2$. By Proposition 4, all the constants a_j, b_j and c_j, for $1 \leq j \leq L$, can be identified from $s_L^{(\ell)}$ with $1 \leq \ell \leq \max\{2, 2L-1\}$. It remains to show that we can also identify $\tilde{a}_j, \tilde{b}_j$ and $\tilde{c}_j$, for $1 \leq j \leq L$.

By Lemma 6, the constants $\tilde{a}_L$, $\tilde{b}_L$, $\tilde{c}_L$ and $\tilde{a}_j$ and $\tilde{K}_j = \tilde{b}_j + \tilde{c}_j$, for $1 \leq j \leq L-1$, are identifiable from $\dot{s}_L$ and $\ddot{s}_L$. We just need to "separate" $\tilde{b}_j$ and $\tilde{c}_j$ for $1 \leq j \leq L-1$. Due to Lemma 8, this can be done by identifying $\tilde{b}_{L-k}$ recursively, for $k = 1, \ldots, L-1$, from the coefficients of the monomials $y^k v_{L-k}$ in $s_L^{(2k+1)}$. □

A.3 Proofs of Sect. 5: Identifying the Cascade

The following two auxiliary technical lemmas will be used in subsequent arguments concerning the identifiability in the cascade.

Lemma 9 *If $\prod_{j=1}^{M} z_j$, with Z_j non-intermediate species for all j, is a monomial of $x^{(\ell)}$ for a non-intermediate species $X \in \mathscr{S}^{(\alpha)}$ and $\ell \geq 1$, then there exists $1 \leq j_1, j_2 \leq M$*

such that $Z_{j_1} \in \mathscr{S}^{(\alpha)}$ and $Z_{j_2} \in \mathscr{S}^{(\beta)}$ for some β such that the network contains a complex $X + Z$ with $Z \in \mathscr{S}^{(\beta)}$.

Proof For $\ell = 1$ the result is true, since the only products of non-intermediate species appearing in $\dot{x}$ are of the form xz for a species Z that reacts with X. Assume the lemma holds for derivatives of order $1 \leq h \leq \ell - 1$ of non-intermediate species.

By equation (13), if the monomial appears in $x^{(h)}z^{(i)}$ for some $h + i \leq \ell - 1$ and $h > 0$, by Lemma 1(1), there is a monomial $\prod_{l=1}^{M'} z_{j_l}$ in $x^{(h)}$ with $1 \leq h \leq \ell - 1$, and the induction hypothesis gives the result. Assume now $h = 0$ and $x = z_M$. If the monomial appears in $xz^{(i)}$ with $i \leq \ell - 1$, either $i = 0$, and the monomial is xz with X and Z reacting together, which implies the statement, or $1 \leq i \leq \ell - 1$ and the monomial $\prod_{j=1}^{M-1} z_j$ appears in $z^{(i)}$. If the latter holds, $Z_{j_1} = X \in \mathscr{S}^{(\alpha)}$ and, by the induction hypothesis applied to $Z \in \mathscr{S}^{(\beta)}$ for some β and $1 \leq i \leq \ell - 2$, there exists j_2 such that $Z_{j_2} \in \mathscr{S}^{(\beta)}$.

If the product appears in $z_{w,1}^{(h)} z_{w,2}^{(i)}$, for some $h + i \leq \ell - 2$, coming from a block of reactions $Z_{w,1} + Z_{w,2} \rightleftarrows W \to Z_{w,3} + X$ with $X \notin \{Z_{w,1}, Z_{w,2}\}$, then the enzyme is $Z_{w,3}$ and, assuming $Z_{w,1} = Z_{w,3}$, it follows that $Z_{w,2}$ and X lie in $\mathscr{S}_W$. Since $X \in \mathscr{S}^{(\alpha)}$, then $\mathscr{S}_W \subset \mathscr{S}^{(\alpha)}$; in particular, $Z_{w,2} \in \mathscr{S}^{(\alpha)}$. On the other hand, $Z_{w,1} = Z_{w,3} \in \mathscr{S}^{(\beta)}$ for some $\beta \neq \alpha$. If $i = 0$, there exists $1 \leq j_1 \leq M$ such that $Z_{j_1} = Z_{w,2} \in \mathscr{S}^{(\alpha)}$ and, if $i \geq 1$, by the induction hypothesis applied to $Z_{w,2} \in \mathscr{S}^{(\alpha)}$ and the factor of the monomial appearing in $z_{w,2}^{(i)}$, there exists j_1 such that $Z_{j_1} \in \mathscr{S}^{(\alpha)}$. Similarly, if $h = 0$, there exists $1 \leq j_2 \leq M$ such that $Z_{j_2} = Z_{w,1} \in \mathscr{S}^{(\beta)}$ and, if $h \geq 1$, by the induction hypothesis applied to $Z_{w,1} \in \mathscr{S}^{(\beta)}$, there exists j_2 such that $Z_{j_2} \in \mathscr{S}^{(\beta)}$. □

Lemma 10 *If $u \prod_{i=1}^{M} z_i$, with $M \geq 1$, is a monomial of $x^{(\ell)}$ for U an intermediate species and X, Z_i non-intermediate species for all i, either there exist $1 \leq i_1 < i_2 \leq M$ such that Z_{i_1} reacts with Z_{i_2} or there exist $1 \leq i_0 \leq M$ and a species V that reacts with Z_{i_0} such that U reacts to a complex containing V.*

Proof Note that there are no monomials of this type in $\dot{x}$; thus, $\ell \geq 2$. For $\ell = 2$, the only monomials in $\ddot{x}$ that are multiples of an intermediate and non-intermediates are:

- uz, for an intermediate species U that reacts to X and a non-intermediate Z that reacts with X. In this case, the statement holds with $V = X$ and $Z_{i_0} = Z$;
- ux, for an intermediate species U that reacts to a non-intermediate species Z reacting with X. The statement holds with $V = Z$ and $Z_{i_0} = X$.

For $\ell > 2$, recalling that $x^{(\ell)} = \sum_v \frac{\partial x^{(\ell-1)}}{\partial v} \dot{v}$ (where the sum runs over all variables v representing non-intermediates or intermediate species), it follows that $u \prod_{i=1}^{M} z_i$ is a monomial in $\frac{\partial x^{(\ell-1)}}{\partial v} \dot{v}$ for some variable v. Every monomial in $\dot{v}$ is either a single intermediate or a product of two non-intermediate species in a reaction. In the second case, the result follows. Now, if $\prod_{i=1}^{M} z_i$ is a monomial of $\frac{\partial x^{(\ell-1)}}{\partial v}$ and u is a monomial of $\dot{v}$, we have that $v \prod_{i=1}^{M} z_i$ is a monomial of $x^{(\ell-1)}$ and one of the following possibilities for V:

- $V = U$; then, $u \prod_{i=1}^{M} z_i$ is a monomial of $x^{(\ell-1)}$ and the result follows by induction.

– V is a non-intermediate species such that U reacts to a complex containing V. By Lemma 3, there are two variables in $v\prod_{i=1}^{M} z_i$ that react together. If none of these variables is v, there exist $1 \le i_1, i_2 \le M$ such that Z_{i_1} and Z_{i_2} react together; otherwise, there exists $1 \le i_0 \le M$ such that V reacts with Z_{i_0}. □

We follow here the notations introduced in Sect. 5, more precisely, in the general cascade (6). We also set $S_{0,L_0} := E$.

For $1 \le n \le N$, we have

$$\dot{s}_{n,L_n} = c_{n,L_n}u_{n,L_n} - \tilde{a}_{n,L_n}s_{n,L_n}f_n + \tilde{b}_{n,L_n}v_{n,L_n} - \sum_{j=1}^{L_{n+1}} a_{n+1,j}s_{n,L_n}s_{n+1,j-1} + \sum_{j=1}^{L_{n+1}} K_{n+1,j}u_{n+1,j}$$

and, for $n = N$, only the three first terms appear in the derivative, *i.e.* $a_{N+1,j} = 0$, $K_{N+1,j} = 0$ for all j.

For $\ell \ge 2$, by Eq. (13):

$$\begin{aligned} s^{(\ell)}_{n,L_n} = &\sum_{h+i\le\ell-1} \beta_{f_n,h,i}\, s^{(h)}_{n,L_n} f^{(i)}_n + \sum_{j=1}^{L_{n+1}} \sum_{h+i\le\ell-1} \beta_{s_{n+1,j-1},h,i}\, s^{(h)}_{n,L_n} s^{(i)}_{n+1,j-1} \\ &+ \sum_{h+i\le\ell-2} \gamma_{u_{n,L_n},h,i}\, s^{(h)}_{n-1,L_{n-1}} s^{(i)}_{n,L_n-1} + \delta_{u_{n,L_n}} u_{n,L_n} + \delta_{v_{n,L_n}} v_{n,L_n} + \sum_{j=1}^{L_{n+1}} \delta_{u_{n+1,j}} u_{n+1,j} \end{aligned} \tag{20}$$

where

$$\beta_{f_n,h,i} = \begin{cases} -\binom{\ell-1}{h}\tilde{a}_{n,L_n} & \text{if } h+i = \ell-1 \\ \tilde{a}_{n,L_n}\tilde{b}_{n,L_n}\binom{h+i}{h}(-\tilde{K}_{n,L_n})^{\ell-2-h-i} & \text{if } h+i \le \ell-2 \end{cases},$$

$$\beta_{s_{n+1,j-1},h,i} = \begin{cases} -\binom{\ell-1}{h}a_{n+1,j} & \text{if } h+i = \ell-1 \\ -\binom{h+i}{h}a_{n+1,j}(-K_{n+1,j})^{\ell-1-h-i} & \text{if } h+i \le \ell-2 \end{cases},$$

$$\gamma_{u_{n,L_n},h,i} = c_{n,L_n}a_{n,L_n}\binom{h+i}{h}(-K_{n,L_n})^{\ell-2-h-i} \quad \text{for } 0 \le h+i \le \ell-2,$$

$$\delta_{u_{n,L_n}} = c_{n,L_n}(-K_{n,L_n})^{\ell-1}, \quad \delta_{v_{n,L_n}} = \tilde{b}_{n,L_n}(-\tilde{K}_{n,L_n})^{\ell-1},$$

$$\delta_{u_{n+1,j}} = (-1)^{\ell-1}K^{\ell}_{n+1,j} \quad \text{for } 0 \le h+i \le \ell-2$$

According to formula (20), every monomial of $s^{(\ell)}_{n,L_n}$ is either an intermediate species that appears in $\dot{s}_{n,L_n}$, or it appears as a monomial in one of the products:

(a) $s^{(h)}_{n,L_n} f^{(i)}_n$ for $h+i \le \ell-1$,
(b) $s^{(h)}_{n,L_n} s^{(i)}_{n+1,j-1}$ $(1 \le j \le L_{n+1})$ for $h+i \le \ell-1$,
(c) $s^{(h)}_{n-1,L_{n-1}} s^{(i)}_{n,L_n-1}$ for $h+i \le \ell-2$.

The following three technical lemmas describe how the coefficients of some distinguished monomials change recursively after differentiation. These results allow us

to obtain Proposition 6 and hence, the identifiability result about the cascade stated in Sect. 5 (Theorem 2 in the main text).

Lemma 11 *Let $\mathcal{M} = \prod_{j=1}^{M} z_j$ be a monomial of $s_{n-1,L_{n-1}}^{(\ell_0)}$ which is not a monomial of any derivative of $s_{n-1,L_{n-1}}$ of lower order and only involves variables corresponding to species in $\mathscr{S}^{(k)}$, $\mathscr{S}^{(N+k)}$, for $1 \le k \le n-1$, and $\mathscr{S}^{(2N+1)}$. Assume that:*

- *$\mathcal{M}$ is square free and does not involve two disjoint pairs of variables corresponding to species that react together;*
- *if $s_{n-1,L_{n-1}}$ divides $\mathcal{M}$, for every $1 \le j_1, j_2 \le M$ such that Z_{j_1} and Z_{j_2} react together, $Z_{j_1} = s_{n-1,L_{n-1}}$ or $Z_{j_2} = s_{n-1,L_{n-1}}$.*

Then, $\widehat{\mathcal{M}} := s_{n,L_n-1}\mathcal{M}$ is a monomial of $s_{n,L_n}^{(\ell_0+2)}$ and of no lower-order derivative of s_{n,L_n}. Moreover, if $C_{\mathcal{M}}$ is the coefficient of $\mathcal{M}$ in $s_{n-1,L_{n-1}}^{(\ell_0)}$, the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+2)}$ is $c_{n,L_n}a_{n,L_n}C_{\mathcal{M}}$.

Proof Assume $\widehat{\mathcal{M}}$ is a monomial of $s_{n,L_n}^{(\ell)}$ for some $\ell \ge 1$. Then, it is a monomial of one of the products in cases (a), (b) or (c) stated above. We will show that it can only appear in case (c) with $i = 0$.

In cases (a) or (b), we must have $i > 0$, since the variables f_n and $s_{n+1,j-1}$ do not divide $\widehat{\mathcal{M}}$. Then, a factor of $\widehat{\mathcal{M}}$ is a monomial of a derivative $f_n^{(i)}$ or $s_{n+1,j-1}^{(i)}$ of positive order and, by Lemma 9, it contains a variable in $\mathscr{S}^{(N+n)}$ or $\mathscr{S}^{(n+1)}$, contradicting the assumption on the variables involved in $\mathcal{M}$. It follows that $\widehat{\mathcal{M}}$ is a monomial in a product in (c).

Assume that $i \ge 1$. If $h = 0$, then $s_{n-1,L_{n-1}}$ divides $\mathcal{M}$ and $\widetilde{\mathcal{M}} := s_{n,L_n-1} \cdot \frac{\mathcal{M}}{s_{n-1,L_{n-1}}}$ is a monomial of $s_{n,L_n-1}^{(i)}$. Due to Lemma 3, $\widetilde{\mathcal{M}}$ contains two variables corresponding to species that react together. By the second assumption of the lemma and the fact that S_{n,L_n-1} only reacts with $S_{n-1,L_{n-1}}$ or F_n (and f_n does not divide $\widehat{\mathcal{M}}$), one of these variables must be $s_{n-1,L_{n-1}}$; but $s_{n-1,L_{n-1}}$ does not divide $\widetilde{\mathcal{M}}$, since it is square free. If $h \ge 1$ and $\widehat{\mathcal{M}} = \mathcal{M}_1 \cdot \mathcal{M}_2$, where $\mathcal{M}_1$ is a monomial in $s_{n-1,L_{n-1}}^{(h)}$ and $\mathcal{M}_2$ is a monomial in $s_{n,L_n-1}^{(i)}$, by Lemma 3, each of the monomials $\mathcal{M}_1$ and $\mathcal{M}_2$ contains two variables corresponding to species that react together. One of these variables must be s_{n,L_n-1}, because $\mathcal{M}$ does not contain two pairs of variables corresponding to species that react together. Since S_{n,L_n-1} only reacts with $S_{n-1,L_{n-1}}$ or F_n, this is only possible in the case where $s_{n-1,L_{n-1}}$ divides $\mathcal{M}$, but then $s_{n-1,L_{n-1}}$ does not divide $\frac{\mathcal{M}}{s_{n-1,L_{n-1}}}$ and it does not contain two variables corresponding to species that react together.

Then, necessarily $i = 0$ and $\mathcal{M}$ is a monomial of $s_{n-1,L_{n-1}}^{(h)}$ for $h \le \ell - 2$. This implies that $\ell \ge \ell_0 + 2$.

Finally, let us show that $\widehat{\mathcal{M}}$ effectively appears in $s_{n,L_n}^{(\ell_0+2)}$ and compute its coefficient. Considering formula (20) for $\ell = \ell_0+2$, by our previous arguments, we have that $\widehat{\mathcal{M}} = s_{n,L_n-1}\mathcal{M}$ can only arise from a product $s_{n-1,L_{n-1}}^{(h)} s_{n,L_n-1}$ when $\mathcal{M}$ is a monomial of $s_{n-1,L_{n-1}}^{(h)}$ and $h \le \ell_0$. By the minimality of ℓ_0, the only possibility is that $h = \ell_0$; moreover, if $C_{\mathcal{M}}$ is the coefficient of $\mathcal{M}$ in $s_{n-1,L_{n-1}}^{(\ell_0)}$, the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+2)}$ is $\gamma_{u_{n,L_n},\ell_0,0}C_{\mathcal{M}} = c_{n,L_n}a_{n,L_n}C_{\mathcal{M}}$. □

Lemma 12 *Let $u\,\mathcal{M}$ be a monomial of $s_{n-1,L_{n-1}}^{(\ell_0)}$ which is not a monomial of any derivative of $s_{n-1,L_{n-1}}$ of lower order, where U is an intermediate species and $\mathcal{M}$ only involves variables corresponding to species in $\mathscr{S}^{(k)}$, for $1 \le k \le n-1$, and $\mathscr{S}^{(2N+1)}$. Assume that $\mathcal{M}$ does not involve two variables corresponding to species that react together and $s_{n-1,L_{n-1}}$ does not divide $\mathcal{M}$.*

Then, $\widehat{\mathcal{M}} := s_{n,L_n-1} u\,\mathcal{M}$ is a monomial of $s_{n,L_n}^{(\ell_0+2)}$ and of no lower-order derivative of s_{n,L_n}. Moreover, if $C_{\mathcal{M}}$ is the coefficient of $u\,\mathcal{M}$ in $s_{n-1,L_{n-1}}^{(\ell_0)}$, the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+2)}$ is $c_{n,L_n}a_{n,L_n}C_{\mathcal{M}}$. In addition, if $\widetilde{C}_{\mathcal{M}}$ is the coefficient of $u\,\mathcal{M}$ in $s_{n-1,L_{n-1}}^{(\ell_0+1)}$, the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+3)}$ is $c_{n,L_n}a_{n,L_n}(\widetilde{C}_{\mathcal{M}} - K_{n,L_n}C_{\mathcal{M}})$.

Proof Assume $\widehat{\mathcal{M}}$ is a monomial of $s_{n,L_n}^{(\ell)}$ and consider the three cases (a), (b) and (c) listed above. We will show that it can only appear in case (c) with $i = 0$.

If $\widehat{\mathcal{M}}$ appears from a product of type (a), (b), or (c) with $h \ge 1$ and $i \ge 1$, there is a factor of $\widehat{\mathcal{M}}$ not involving intermediate species which is a monomial of a derivative of positive order of a non-intermediate species and, by Lemma 3, this factor involves two variables of species that react together. But $\mathcal{M}$ does not contain two variables of species reacting together; in addition, the only species in $\mathscr{S}^{(k)}$, for $1 \le k \le n-1$, that reacts with S_{n,L_n-1} is $S_{n-1,L_{n-1}}$, and $s_{n-1,L_{n-1}}$ does not divide $\mathcal{M}$.

On the other hand, $\widehat{\mathcal{M}}$ cannot appear from cases (a) or (b) with $h = 0$ or $i = 0$, since none of the variables s_{n,L_n}, f_n or $s_{n+1,j-1}$, for $1 \le j \le L_{n+1}$, divides $\widehat{\mathcal{M}}$. Finally, the assumption that $s_{n-1,L_{n-1}}$ does not divide $\mathcal{M}$ implies that the monomial cannot appear in case (c) with $h = 0$.

We conclude that $\widehat{\mathcal{M}}$ only appears as a monomial in $s_{n-1,L_{n-1}}^{(h)} s_{n,L_n-1}$ for $1 \le h \le \ell - 2$, that is, when $u\mathcal{M}$ is a monomial of $s_{n-1,L_{n-1}}^{(h)}$. Then, $\ell \ge \ell_0 + 2$.

The computation of the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+2)}$ follows as in the proof of Lemma 11.

Finally, let us obtain the coefficient of $\widehat{\mathcal{M}}$ in $s_{n,L_n}^{(\ell_0+3)}$. As shown before, in formula (20) the monomial $\widehat{\mathcal{M}}$ may only appear from terms of the form (c) with $i = 0$ and $1 \le h \le \ell_0 + 1$ such that $u\mathcal{M}$ is a monomial of $s_{n-1,L_{n-1}}^{(h)}$. By the minimality of ℓ_0, the only possible values of h are ℓ_0 and $\ell_0 + 1$; thus, the corresponding coefficient is $\gamma_{u_n,L_n,\ell_0-1,0}\widetilde{C}_{\mathcal{M}} + \gamma_{u_n,L_n,\ell_0-2,0}C_{\mathcal{M}} = c_{n,L_n}a_{n,L_n}\widetilde{C}_{\mathcal{M}} + c_{n,L_n}a_{n,L_n}(-K_{n,L_n})C_{\mathcal{M}} = c_{n,L_n}a_{n,L_n}(\widetilde{C}_{\mathcal{M}} - K_{n,L_n}C_{\mathcal{M}})$. □

Lemma 13 *For $1 \le l \le L_m - 1$,*

$$\mathcal{M}_{n,s_{m,l}} = s_{m,l}\, f_m\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1} \quad \text{and} \quad \mathcal{M}_{n,v_{m,l}} = v_{m,l}\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$$

are monomials of $s_{n,L_n}^{(2(n-m+1))}$ for every $n \ge m+1$, and they are not monomials of any derivative of s_{n,L_n} of lower order. The corresponding coefficients are, respectively,

$$\tilde{a}_{m,l}\,\tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i}a_{i,L_i} \quad \text{and} \quad \tilde{K}_{m,l}\,\tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i}a_{i,L_i}.$$

Proof For $n = m+1$, we must show that, for every $1 \le l \le L_m - 1$,

$$\mathcal{M}_{m+1,s_{m,l}} = s_{m,l} f_m s_{m,L_m} s_{m+1,L_{m+1}-1} \text{ and } \mathcal{M}_{m+1,v_{m,l}} = v_{m,l} s_{m,L_m} s_{m+1,L_{m+1}-1}$$

are monomials of $s^{(4)}_{m+1,L_{m+1}}$ and of no lower-order derivative of $s_{m+1,L_{m+1}}$.

It is easy to see that none of the required monomials appears in $\dot{s}_{m+1,L_{m+1}}$ or $\ddot{s}_{m+1,L_{m+1}}$, because these derivatives do not contain monomials of degree 4 and the monomials that are multiples of intermediates have degree at most 2 (see the proof of Lemma 10).

Consider now the expression of $s^{(\ell)}_{m+1,L_{m+1}}$ following (20), with $\ell \ge 3$.

The monomials $\mathcal{M}_{m+1,s_{m,l}}$ and $\mathcal{M}_{m+1,v_{m,l}}$ do not arise from products of type (a) or (b) with $h = 0$ or $i = 0$, since they are not multiples of $s_{m+1,L_{m+1}}$, f_{m+1} or $s_{m+2,j-1}$. Taking into account that every monomial in a first-order derivative of a non-intermediate is either a multiple of the non-intermediate or an intermediate that reacts to it, we have that the monomials do not appear either from products of type (a) or (b) with $h = 1$ or $i = 1$. As $h + i \le \ell - 1$ in products of type (a) or (b), we deduce that $\mathcal{M}_{m+1,s_{m,l}}$ and $\mathcal{M}_{m+1,v_{m,l}}$ do not appear in these products for $\ell = 3$ nor $\ell = 4$.

In products of type (c), if $h + i \le 1$, there are no monomials of degree 4, and those that are multiples of an intermediate have degree at most 2.

We conclude that $\mathcal{M}_{m+1,s_{m,l}}$ and $\mathcal{M}_{m+1,v_{m,l}}$ are not monomials of $s^{(3)}_{m+1,L_{m+1}}$ and that they may only appear in $s^{(4)}_{m+1,L_{m+1}}$ from products of type (c) with $h + i = 2$.

- $h = 0$, $i = 2$. By looking at the expansion of $s^{(2)}_{m+1,L_{m+1}-1}$, we deduce that $s_{m,l} f_m s_{m+1,L_{m+1}-1}$ and $v_{m,l} s_{m+1,L_{m+1}-1}$, for $l < L_m$, are not monomials of this derivative.
- $h = i = 1$: The monomials $\mathcal{M}_{m+1,s_{m,l}}$ do not appear in this product because the only variable involved that reacts with $S_{m+1,L_{m+1}-1}$ is S_{m,L_m} and the monomials $s_{m,l} f_m$ do not appear in $\dot{s}_{m,L_m}$ for $l < L_m$. The monomials $\mathcal{M}_{m+1,v_{m,l}}$ do not appear since $v_{m,l}$ does not react to s_{m,L_m} or $s_{m+1,L_{m+1}-1}$ for $l < L_m$.
- $h = 2$, $i = 0$: As in the proof of Lemma 6, it follows that $s_{m,l} f_m s_{m,L_m}$ and $v_{m,l} s_{m,L_m}$ are monomials of $s^{(2)}_{m,L_m}$ with respective coefficients $\tilde{a}_{m,l}\tilde{a}_{L_m}$ and $\tilde{K}_{m,l}\tilde{a}_{m,L_m}$.

Therefore, $\mathcal{M}_{m+1,s_{m,l}}$ and $\mathcal{M}_{m+1,v_{m,l}}$ effectively appear in $s^{(4)}_{m+1,L_{m+1}}$; more precisely, they arise from the product $\gamma_{u_{m+1},L_{m+1},2,0} s^{(2)}_{m,L_m} s_{m+1,L_{m+1}-1}$. The corresponding coefficients can be obtained from the fact that $\gamma_{u_{m+1},L_{m+1},2,0} = c_{m+1,L_{m+1}}.a_{m+1,L_{m+1}}$.

Let $n > m+1$ and assume the monomials $\mathcal{M}_{n-1,s_{m,l}}$ and $\mathcal{M}_{n-1,v_{m,l}}$ appear in $s^{(2(n-m))}_{n-1,L_{n-1}}$ and in no derivative of $s_{n-1,L_{n-1}}$ of a lower order.

Let $1 \le l \le L_m - 1$. Consider first $\mathcal{M}_{n,s_{m,l}}$, which is a product of non-intermediates. If it appears in a derivative $s^{(\ell)}_{n,L_n}$, it arises from a product in case (a), (b) or (c) listed previously.

Since $\mathcal{M}_{n,s_{m,l}}$ does not contain any variable corresponding to a species in $\mathscr{S}^{(N+n)} = \{F_n\}$ or $\mathscr{S}^{(n+1)} = \{S_{n+1,j}, 0 \le j \le L_{n+1}\}$, by Lemma 9, it cannot appear from cases (a) or (b). Then, it is a monomial in a product $s^{(h)}_{n-1,L_{n-1}} s^{(i)}_{n,L_n-1}$ for $h + i \le \ell - 2$. If

$i > 0$, the factor $\mathcal{M}_1$ of $\mathcal{M}_{n,s_{m,l}}$ which is a monomial in $s^{(i)}_{n,L_n-1}$ contains a variable in $\mathscr{S}^{(n)}$, namely s_{n,L_n-1}, and another variable in a set $\mathscr{S}^{(k)}$ that contains a species reacting with S_{n,L_n-1}. Since the only species that react with S_{n,L_n-1} are $S_{n-1,L_{n-1}}$ and F_n, it follows that $\mathcal{M}_1$ contains a variable in $\mathscr{S}^{(n-1)}$. Now, $\mathcal{M}_{n,s_{m,l}}/\mathcal{M}_1$ is a monomial in $s^{(h)}_{n-1,L_{n-1}-1}$; therefore, it also contains a variable in $\mathscr{S}^{(n-1)}$. But, since $n > m+1$, the only factor of $\mathcal{M}_{n,s_{m,l}}$ in $\mathscr{S}^{(n-1)}$ is $s_{n-1,L_{n-1}-1}$, leading to a contradiction. We conclude that $i = 0$ and $\mathcal{M}_{n,s_{m,l}}$ appears as a monomial in $s^{(h)}_{n-1,L_{n-1}} s_{n,L_n-1}$, namely $\mathcal{M}_{n,s_{m,l}} = \mathcal{M}_{n-1,s_{m,l}} s_{n,L_n-1}$ with $\mathcal{M}_{n-1,s_{m,l}}$ a monomial in $s^{(h)}_{n-1,L_{n-1}}$ for $h \leq \ell - 2$. Then $\ell \geq 2(n-m+1)$.

Now, consider $\mathcal{M}_{n,v_{m,l}}$ and assume it is a monomial of $s^{(\ell)}_{n,L_n}$. As, for $n > m+1$, none of the variables s_{n,L_n}, $s_{n+1,j-1}$, f_n or $s_{n-1,L_{n-1}}$ divides $\mathcal{M}_{n,v_{m,l}}$, this monomial cannot arise from cases (a) or (b) with either $i = 0$ or $h = 0$, nor from (c) with $h = 0$. If it arises from cases (a), (b) or (c) with $h \geq 1$ and $i \geq 1$, then $\mathcal{M}_{n,v_{m,l}} = \mathcal{M}_1\mathcal{M}_2$ with $\mathcal{M}_1$ and $\mathcal{M}_2$ monomials appearing in derivatives of positive order of non-intermediate species. Assume $v_{m,l}$ divides $\mathcal{M}_1$. Then, $\mathcal{M}_2$ is a product of non-intermediates; by Lemma 3, it contains the only two variables of $\mathcal{M}_{n,v_{m,l}}$, s_{m,L_m} and $s_{m+1,L_{m+1}-1}$, corresponding to species that react together. On the other hand, $\mathcal{M}_1 = v_{m,l}\mathcal{M}$, where $\mathcal{M}$ is not constant since $V_{m,l}$ does not react to S_{n,L_n}, F_n, $S_{n+1,j-1}$, $S_{n-1,L_{n-1}}$ nor S_{n,L_n-1} (so, $v_{m,l}$ is not a monomial in a derivative of s_{n,L_n}, $f_n, s_{n+1,j-1}, s_{n-1,L_{n-1}}$ nor s_{n,L_n-1}). By Lemma 10, taking into account that $V_{m,l}$ is only involved in the reactions $F_m + S_{m,l} \rightleftarrows V_{m,l} \to F_m + S_{m,l-1}$, we have that $\mathcal{M}$ contains either two variables corresponding to species that react together or it contains one variable that reacts with F_m, $S_{m,l-1}$ or $S_{m,l}$. But none of these possibilities happen.

We conclude that $\mathcal{M}_{n,v_{m,l}}$ arises from (c) with $i = 0$ and it appears in $s^{(h)}_{n-1,L_{n-1}} s_{n,L_n-1}$ for $h \leq \ell - 2$, that is, $\mathcal{M}_{n-1,v_{m,l}}$ is a monomial of $s^{(h)}_{n-1,L_{n-1}}$ for $h \leq \ell - 2$. Then $\ell \geq 2(n-m+1)$.

The fact that the monomials effectively appear in $s^{2(n-m+1)}_{n,L_n}$ and the computation of their coefficients follow similarly as in the proof of Lemma 11. □

From the previous lemmas and the results for the case of a single layer proved in Proposition 3, we obtain the following proposition that leads to our identifiability result for the cascade (see Table 3). The highlighted constant in each case is the one we will identify from the corresponding coefficient.

Proposition 6 *For network (6), for every $n \geq m$, the following monomials $\mathcal{M}$ appear in $s^{(\ell)}_{n,L_n}$ with coefficient $\pm C_{\mathcal{M}}$ for the stated value ℓ, and they do not appear in any derivative of s_{n,L_n} of lower order:*

1. $\mathcal{M} = f_m\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = \tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+1$;
2. *for* $0 \leq k \leq L_m - 1$, $\mathcal{M} = s^k_{m-1,L_{m-1}} u_{m,L_m-k} \prod_{i=m+1}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = c_{m,L_m-k} \left(\prod_{j=0}^{k-1} a_{m,L_m-j}\, c_{m,L_m-j}\right)\left(\prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}\right)$, $\ell = 2(n-m)+2k+1$;
3. *for* $0 \leq k \leq L_m - 1$, $\mathcal{M} = s^k_{m-1,L_{m-1}} v_{m,L_m-k} \prod_{i=m+1}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = \tilde{b}_{m,L_m-k} \left(\prod_{j=0}^{k-1} a_{m,L_m-j}\, c_{m,L_m-j}\right)\left(\prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}\right)$, $\ell = 2(n-m)+2k+1$;

4. $\mathcal{M} = s_{m-1,L_{m-1}} \prod_{i=m}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = c_{m,L_m} a_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+2$;
5. *for* $1 \le j \le L_m - 1$, $\mathcal{M} = s_{m,j}\, f_m\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = \tilde{a}_{m,j} \tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+2$;
6. *for* $1 \le j \le L_m - 1$, $\mathcal{M} = v_{m,j}\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = \tilde{K}_{m,j} \tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+2$;
7. *for* $1 \le j \le L_m - 1$, $\mathcal{M} = s_{m,j-1} s_{m-1,L_{m-1}} \prod_{i=m}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = a_{m,j} \prod_{i=m}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+3$;
8. *for* $1 \le j \le L_m - 1$, $\mathcal{M} = u_{m,j} \prod_{i=m}^{n} s_{i,L_i-1}$, $C_{\mathcal{M}} = K_{m,j} \prod_{i=m}^{n} c_{i,L_i} a_{i,L_i}$, $\ell = 2(n-m)+3$.

Furthermore, the monomials $u_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$ *and* $v_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$ *(c.f. items 2 and 3 with* $k = 0$*) appear in* $s_{n,L_n}^{(2(n-m)+2)}$ *with coefficients* $-\left(K_{m,L_m} + \sum_{i=m+1}^{n} K_{i,L_i}\right) c_{m,L_m} \left(\prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}\right)$ *and* $-\left(\tilde{K}_{m,L_m} + \sum_{i=m+1}^{n} K_{i,L_i}\right) \tilde{b}_{m,L_m} \left(\prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}\right)$*, respectively.*

Proof Fix m with $1 \le m \le N$. We prove the proposition inductively for $n \ge m$.

The case $n = m$ is considered in Sect. 4.2.

Let $n \ge m+1$. Items 5 and 6 are proved in Lemma 13. For the remaining monomials, assuming the statement holds for $n-1$, we deduce that it is also true for n by applying Lemma 11 (for items 1, 4 and 7) and Lemma 12 (for items 2, 3, 8 and the last statement of the proposition).

We present a complete proof in the first two cases. The induction step for the monomials of the remaining items follows similarly.

1. Consider $\mathcal{M}_1 = f_m\, s_{m,L_m} \prod_{i=m+1}^{n-1} s_{i,L_i-1}$. By the inductive assumption, this monomial appears in $s_{n-1,L_{n-1}}^{(2(n-1-m)+1)}$ with coefficient $C_{\mathcal{M}_1} = \tilde{a}_{m,L_m} \prod_{i=m+1}^{n-1} c_{i,L_i} a_{i,L_i}$, and in no derivative of $s_{n-1,L_{n-1}}$ of a lower order. Let us show that $\mathcal{M}_1$ satisfies the assumptions of Lemma 11. First, note that $\mathcal{M}_1$ is square free and only involves variables corresponding to species in $\mathscr{S}^{(k)}$, for $m \le k \le n-1$, and $\mathscr{S}^{(N+m)}$. In addition, since two species S_{i,L_i-1}, S_{j,L_j-1}, for $m \le i, j \le n-1$, do not react together and F_m does not react with S_{i,L_i-1} for $m+1 \le i \le n-1$, then $\mathcal{M}_1$ does not contain two disjoint pairs of variables corresponding to species that react together. Finally, we have that $s_{n-1,L_{n-1}}$ divides $\mathcal{M}_1$ only when $n = m+1$, and in this case, $\mathcal{M}_1 = f_{n-1} s_{n-1,L_{n-1}}$, which clearly satisfies the assumptions of the lemma. Therefore, by Lemma 11, we conclude that $s_{n,L_n-1}\mathcal{M}_1 = f_m\, s_{m,L_m} \prod_{i=m+1}^{n} s_{i,L_i-1}$ is a monomial of $s_{n,L_n}^{(2(n-1-m)+1+2)} = s_{n,L_n}^{(2(n-m)+1)}$ and of no lower-order derivative of s_{n,L_n}, and its corresponding coefficient is $c_{n,L_n} a_{n,L_n} C_{\mathcal{M}_1} = \tilde{a}_{m,L_m} \prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}$.
2. For a fixed k, with $0 \le k \le L_m - 1$, the monomial $\mathcal{M}$ can be written as $\mathcal{M} = s_{n,L_n-1} u \mathcal{M}_2$, where $u := u_{m,L_m-k}$ is a variable corresponding to an intermediate species and $\mathcal{M}_2 := s_{m-1,L_{m-1}}^{k} \prod_{i=m+1}^{n-1} s_{i,L_i-1}$. By the induction assumption, we have that $u\mathcal{M}_2$ is a monomial of $s_{n-1,L_{n-1}}^{(2(n-1-m)+2k+1)}$, with coef-

ficient $C_{\mathcal{M}_2} = c_{m,L_m-k}\Big(\prod_{j=0}^{k-1} a_{m,L_m-j}\, c_{m,L_m-j}\Big)\Big(\prod_{i=m+1}^{n-1} c_{i,L_i} a_{i,L_i}\Big)$, and it does not appear in any lower-order derivative of $s_{n-1,L_{n-1}}$. Let us show that $\mathcal{M}_2$ satisfies the assumptions of Lemma 12. It is clear that $\mathcal{M}_2$ only involves variables in $\mathscr{S}^{(i)}$ for $i \leq n-1$ and $\mathscr{S}^{(2N+1)}$ and that $s_{n-1,L_{n-1}}$ does not divide $\mathcal{M}_2$, since $n \neq m$. Also, since two species S_{i,L_i-1} and S_{j,L_j-1}, for $m \leq i, j \leq n-1$, do not react together and $S_{m-1,Lm-1}$ does not react with $S_{i,L_{i-1}}$ for $i \geq m+1$, it follows that $\mathcal{M}_2$ does not involve two variables corresponding to species that react together. Then, by Lemma 12, we conclude that $s_{n,L_n-1} u \mathcal{M}_2 = s^k_{m-1,L_{m-1}} u_{m,L_m-k} \prod_{i=m+1}^{n} s_{i,L_i-1}$ is a monomial of $s^{(2(n-m)+2k+1)}_{n,L_n}$ and of no lower-order derivative of s_{n,L_n}, and its coefficient is $c_{n,L_n} a_{n,L_n} C_{\mathcal{M}_2} = c_{m,L_m-k}\Big(\prod_{j=0}^{k-1} a_{m,L_m-j}\, c_{m,L_m-j}\Big)\Big(\prod_{i=m+1}^{n} c_{i,L_i} a_{i,L_i}\Big)$. □

References

Anguelova M, Karlsson J, Jirstrand M (2012) Minimal output sets for identifiability. Math Biosci 239:139–153

Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. Proc Natl Acad Sci USA 108(31):12675–12680

Bellman R, Åström K (1970) On structural identifiability. Math Biosci 7(3):329–339

Bellu G, Saccomani MP, Audoly S, D'Angìo L (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. Comput Methods Programs Biomed 88:52–61

Boulier F (2007) Differential elimination and biological modelling. Radon Ser Comput Appl Math 2:111–139

Brouwer AF, Meza R, Eisenberg MC (2017) A systematic approach to determining the identifiability of multistage carcinogenesis models. Risk Anal 37(7):1375–1387

Catozzi S, Di-Bella JP, Ventura A, Sepulchre JA (2016) Signaling cascades transmit information downstream and upstream but unlikely simultaneously. BMC Syst Biol 16(1):1–20

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK (2009) Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5:239

Chis O-T, Banga JR, Balsa-Canto E (2011a) Structural identifiability of systems biology models: a critical comparison of methods. PLoS ONE 6(11):e27755

Chiş O, Banga JR, Balsa-Canto E (2011b) GenSSI: a software toolbox for structural identifiability analysis of biological models. Bioinformatics 27(18):2610–2611

Craciun G, Pantea C (2008) Identifiability of chemical reaction networks. J Math Chem 44:244–259

Davis RJ (2000) Signal transduction by the JNK group of MAP kinases. Cell 103:239–252

Deshaies RJ, Ferrell JE (2001) Multisite phosphorylation and the countdown to S phase. Cell 107(7):819–822

DiStefano JJ III (2014) Dynamic systems biology modeling and simulation. Elsevier, London (2014)

Hagen DR, White JK, Tidor B (2013) Convergence in parameters and predictions using computational experimental design. Interface Focus 3:20130008

Hong H, Ovchinnikov A, Pogudin G, Yap C (2018a) Global identification of differential models. Preprint. URL arXiv:1801.08112

Hong H, Ovchinnikov A, Pogudin G, Yap C (2018b) SIAN: software for structural identifiability analysis of ODE models. To appear in Bioinformatics

Hornberg JJ, Binder B, Bruggeman FJ, Schoeber B, Heinrich R, Westerhoff HV (2005) Control of MAPK signalling: from complexity to what really matters. Oncogene 24:5533–5542

Huang C-YF, Ferrell JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93(19):10078–10083

Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. Eur J Biochem 267:1583–1588

Kyriakis JM, Avruch J (2001) Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. Physiol Rev 81(2):807–869
Ligon T, Fröhlich F, Chiş O, Banga J, Balsa-Canto E, Hasenauer J (2017) GenSSI 2.0: multiexperiment structural identifiability analysis of SBML models. Bioinformatics 34(8):1421–1423
Lin J, Harding A, Giurisato E, Shaw AS (2009) KSR1 modulates the sensitivity of mitogen-activated protein kinase pathway activation in T cells without altering fundamental system outputs. Mol Cell Biol 29:2082–2091
Ljung L, Glad T (1994) On global identifiability of arbitrary model parameterizations. Automatica 30:265–276
Maple 18 (2014) Maplesoft, a division of Waterloo Maple Inc., Waterloo, Ontario
Meshkat N, Eisenberg M, DiStefano JJ III (2009) An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases. Math Biosci 222:61–72
Meshkat N, Kuo C, DiStefano J (2014) On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: a novel web implementation. PLoS ONE 9(10):e110261
Ollivier F (1990) Le problème de l'identifiabilité structurelle globale: approche théorique, méthodes effectives et bornes de complexité. Thèse de Doctorat en Sciences, École Polytechnique, Paris, France
Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, Berman K, Cobb MH (2001) Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocr Rev 22:153–183
Pérez Millán M, Dickenstein A (2018) The structure of MESSI biological systems. SIAM J Appl Dyn Syst 17(2):1650–1682
Pohjanpalo H (1978) System identifiability based on power-series expansion of solution. Math Biosci 41:21–33
Qiao L, Nachbar RB, Kevrekidis IG, Shvartsman SY (2007) Bistability and oscillations in the Huang–Ferrell model of MAPK signaling. PLoS Comput Biol 3(9):1819–1826
Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J (2014) Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30(10):1440–1448
Saccomani MP, Audoly S, D'Angìo L (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica 39(4):619–632
Schaeffer HJ, Weber MJ (1999) Mitogen-activated protein kinases: specific messages from ubiquitous messengers. Mol Cell Biol 19:2435–2444
Sedoglavic A (2002) A probabilistic algorithm to test local algebraic observability in polynomial time. J Symbolic Comput 33:735–755
Shaul YD, Seger R (2007) The MEK/ERK cascade: from signaling specificity to diverse functions. Biochim Biophys Acta 1773(8):1213–1226
Walch OJ, Eisenberg MC (2016) Parameter identifiability and identifiable combinations in generalized Hodgkin–Huxley models. Neurocomputing 199:137–143
Walter E, Pronzato L (1997) Identification of parametric models from experimental data. Springer, Masson
Wang L, Sontag E (2008) On the number of steady states in a multiple futile cycle. J Math Biol 57(1):29–52
Widmann C, Gibson S, Jarpe MB, Johnson GL (1999) Mitogen-activated protein kinase conservation of a three-kinase module from yeast to human. Physiol Rev 79:143–180
Xia X, Moog CH (2003) Identifiability of nonlinear systems with applications to hiv/aids models. IEEE Trans Automat Contr 48:330–336
Zarubin T, Han J (2005) Activation and signaling of the p38 MAP kinase pathway. Cell Res 15:11–18

Affiliations

Gabriela Jeronimo[1,2,3] · Mercedes Pérez Millán[1,2,3] · Pablo Solernó[1,2]

Gabriela Jeronimo
jeronimo@dm.uba.ar

Pablo Solernó
psolerno@dm.uba.ar

1 Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Matemática, Ciudad Universitaria, Pab. I, C1428EGA Buenos Aires, Argentina

2 Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto de Investigaciones Matemáticas "Luis A. Santaló" (IMAS), Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pab. I, C1428EGA Buenos Aires, Argentina

3 Universidad de Buenos Aires, Ciclo Básico Común, Departamento de Ciencias Exactas, Buenos Aires, Argentina