Willis MJ, von Stosch M.

[Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models.](#)

***Computers and Chemical Engineering* 2017, 104, 366-376**

Copyright:

© 2017. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<https://doi.org/10.1016/j.compchemeng.2017.05.005>

Date deposited:

07/06/2017

Embargo release date:

17 May 2018

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models

Mark J. Willis and Moritz von Stosch
School of Chemical Engineering and Advanced Materials
University of Newcastle
Newcastle upon Tyne, NE1 7RU.
mark.willis@ncl.ac.uk, moritz.von-stosch@ncl.ac.uk

Abstract – In this work, a hybrid semi-parametric modelling framework implemented using mixed integer linear programming (MILP) is used to extract (coupled) nonlinear ordinary differential equations (ODEs) from process data. Applied to fed-batch (bio) chemical reaction systems, unknown (or partially known) system connectivity and/or reaction kinetics are represented using a multivariate rational function (MRF) superstructure. The MRF's are embedded within an ODE framework which is used to incorporate known system model characteristics. Using derivative estimation, the ODEs are decoupled and a MILP algorithm is then used to identify appropriate constitutive model terms using sparse regression. Superstructure sparsity is promoted using a L_0 – pseudo norm penalty, i.e. the cardinality of the model parameter vector, enabling the simultaneous yet decoupled identification of the parameters and model structure discrimination. Using simulated data, two case studies demonstrate a principled approach to hybrid model development, distilling unknown elements of (bio) chemical model structures from process data.

Keywords: Hybrid semi-parametric modelling; sparse regression; mixed integer linear programming; fed-batch (bio) chemical reactors

1.0 Introduction

Linear, polynomial and rational function models are used widely in many fields of science and engineering. They form the basis of many empirical models (correlations) and are constituent components within a wide-class of mechanistic model descriptions within chemical engineering and systems biology. For example, for the isothermal description of chemical reaction networks, the assumption of mass action kinetics yields a system of differential equations with polynomial terms. Similarly, gene expression networks may be formulated as rational / polynomial expressions, biological networks and catalyzed reactions as a set of differential equations with rational functional components and so on. These equations typically contain simple constitutive nonlinear terms indicating parameter and species combinations. A key objective in chemical engineering and systems biology is to determine these structural relationships using measured experimental data. By assembling (non-linear) ordinary differential equations (ODEs) describing the dynamic behaviour of the process, systems analysis, optimisation and control studies may be undertaken (Maria, 2004).

The aim of this work is to detail and demonstrate a novel approach to hybrid modelling; embedding a multivariate rational function (MRF) within a mechanistic model structure. The MRF represents a model superstructure; subsets of monomials within this superstructure being used to instantiate unknown structural relationships within the ODEs, which may be linear, polynomial or rational descriptions. As it makes little sense to use a data-based modelling methodology to represent information that is already known about the dynamic

system of interest mechanistic and data-driven modelling are combined, i.e. a general approach known as hybrid semi-parametric modelling is adopted (see, von Stosch et al, 2014). This allows the components of the system ODEs that can be modelled relatively easily from first principles to be incorporated ahead of system identification studies.

The method that is developed is based on the decomposition of the estimation problem through estimation of each of the system derivatives at all the measured time points forming an equivalent set of independent algebraic equations, see Kahrs and Marquardt (2008). Through this approach and the use of a MRF superstructure the difficult task of simultaneous structure and parameter estimation is reduced to one of just parameter estimation. The correct estimation of the parameters and hence the monomial terms within the MRF is, theoretically, sufficient to deduce the structure of the underlying mathematical model. It also allows, in principle, a choice to be made between competing mechanistic model variants while selecting the MRF structure by parameter estimation. To identify unknown model parameters within the hybrid model, regularized regression is used, see a review by Hesterberg et al. (2008). A mixed integer linear programming (MILP) framework is used to implement a novel regularization strategy which uses a L_0 – pseudo norm penalty, i.e. the cardinality of the model parameter vector. This is achieved using a set of binary variables associated with each of the parameters of the model indicating the existence, or not, of a model parameter and hence a feature within the model.

A specific advantage of the proposed approach when compared to alternative hybrid modelling methodologies presented in the literature is that additional model constraints, e.g. representing structured knowledge about a specific process, can be incorporated within the MILP using equality/inequality constraints in a straightforward manner. Such constraints may be, for example, that isothermal rate constants should be positive, that a particular monomial within the MRF should be present, etc. These constraints may be introduced iteratively as part of the model development lifecycle providing a structured approach to model development. The presented case studies demonstrate typical system specific constraints that may be applied.

2.0 Hybrid modelling

Hybrid models combine parametric and nonparametric model components. The mathematical relationships within the parametric structure are established using process knowledge, whereas the structure of the nonparametric model components is derived from experimental data. In contrast to the approach adopted in this paper, the nonparametric model is normally specified as a neural network (see, Oliveira 2004, Kahrs and Marquardt 2008, von Stosch et al, 2014). Primarily this is because neural networks are universal approximators that can arbitrarily approximate nonlinear input – output relationships. They are however, essentially ‘black-boxes’ giving limited physical insight into the data, i.e. any knowledge about the data, system or process is encoded as network weights. Furthermore, the architecture of a neural network is normally deduced by “training” hybrid models using various network structures. Their performance is then compared using information criteria such as the Akaike information criteria (AIC) proposed by (Akaike, 1974) and Bayesian Information Criteria (BIC) proposed by (Schwarz, 1978). These trial and error procedures can take a significant amount of time (von Stosch et al, 2014), as for all potential changes in the parametric or nonparametric model structure the parameter identification problem has to be solved *de novo*.

In an initiative to improve the development of hybrid model structures, Kahrs and Marquardt (2008) proposed an incremental hybrid modelling approach. First, the derivatives of the process outputs are estimated from measured data. They are then used along with Target Factor Analysis proposed by Bonvin and Rippin (1990) to infer the time-invariant model

structure, the stoichiometry. The remaining mechanistically unknown parts of the process model are then represented by nonparametric model structures which are identified separately. Finally, having established the model components, their parameters are re-identified to ensure statistical optimality. However, as with alternative hybrid modelling methods, the determination of the nonparametric model structure is again accomplished in a trial and error manner, i.e. systematically changing the structure of the nonparametric model and testing its performance. For instance, in the case of a neural network the performance for different sets of inputs (feature selection), as well as the number of hidden layers, the number of nodes in the hidden layers and the type of transfer functions would have to be evaluated. As the parametric structure may be incorrect, introducing an inductive bias, the structure of the nonparametric model would have to be identified more than once. The semi-automatic and simultaneous identification of the parameters and discrimination of the model structure would greatly reduce hybrid model development time. In the proposed approach, described in the following, all possible parametric model alternatives are contained in a superstructure.

2.1 Hybrid model structure

To develop a hybrid model it is assumed that, p , ordinary differential equations (ODEs) may be constructed as,

$$\frac{dx_j}{dt} = \bar{f}_j(\mathbf{x}) + \bar{g}_j(\mathbf{x})\hat{f}_j(\mathbf{x}), \quad (j = 1, \dots, p) \quad (1)$$

The first two functions in the differential equations are $\bar{f}_j(\mathbf{x})$ and $\bar{g}_j(\mathbf{x})$ where $\mathbf{x}$ is a vector comprising the p dependent variables, $x_1, \dots, x_p$. These functions represent the terms in the differential equation that are known, i.e. the terms that describe the known behaviour of the j^{th} variable, x_j . Specifically, $\bar{f}_j(\mathbf{x})$ represents sub-elements of the ODE model that may be derived from first-principles such as e.g. the conservation laws, namely material, momentum, impulse, population or energy balances derived for the process being considered. While the term $\bar{g}_j(\mathbf{x})$ is used to impose additional mechanistic knowledge, where available, in order to reduce and structure the space spanned by variables and the parameters of the nonparametric model, $\hat{f}_j(\mathbf{x})$, see, Thompson and Kramer (1994) and Fiedler and Schuppert, (2008). This may include structural information such as stoichiometry or yield coefficients (Chen et al, 2000; Van de Wouwer et al., 2004; Brendel and Marquardt, 2008; Georgieva and de Azevedo, 2009), or known information about the interaction between specific variables. For example, for biochemical systems, when considering the modeling of the biomass growth, or substrate consumption, the assumption that biomass is a catalyst (Psichogios and Ungar, 1992; Schubert et al., 1994; Van de Wouwer et al., 2004; Oliveira, 2004) allows the kinetic rate terms to be formulated as a product of the biomass concentration, which defines $\bar{g}_j(\mathbf{x})$.

The final function $\hat{f}_j(\mathbf{x})$ represents the unknown terms i.e. the nonparametric model component. This may represent the underlying kinetic or transport terms, which are generally much more difficult to establish for a generally valid model representation at an acceptable cost. In this work, the following MRF superstructure is used to specify $\hat{f}_j(\mathbf{x})$,

$$\hat{f}_j(\mathbf{x}) = \frac{\sum_{k=1}^{N_m} b_{j,k} m_k(\mathbf{x})}{1 + \sum_{k=2}^{N_m} b_{j,k+N_m-1} m_k(\mathbf{x})} \quad (2)$$

where $\{m_k(\mathbf{x})\}_{k=1}^{N_m}$ are a collection of N_m monomials constructed using the dependent variables and $b_{j,k}$ are model parameters. The N_m monomials are assumed to be ordered where

$m_1(\mathbf{x}) = 1$ and each of the remaining monomials are defined as $m_k(\mathbf{x}) = x_1^{v_{1k}} x_2^{v_{2k}} \dots x_p^{v_{pk}}$ where $v_{jk} \in \mathbb{Z}_{\geq 0}$ and a polynomial of degree ‘ d ’, consists of all unique combinations of monomials subject to the constraint $\sum_{j=1}^p v_{jk} \leq d$. If the parameters, $b_{j,N_m+1}, \dots, b_{j,2N_m-1} = 0$, the MRF reduces to a polynomial and if $b_{j,k} (k > 2) = 0$, a linear function. Intuitively this MRF structure has more appeal than, e.g. the use of a neural network as the nonparametric model component. It relates to empirical model structures used in (bio) chemical modelling such as the Monod and Haldane equation as well as those derived from theoretical considerations, e.g. chemical reactions, occurring in well mixed, relatively dilute, homogeneous phases – such as may be found in controlled laboratory batch and fed-batch experiments – typically obey the law of mass action kinetics. However, the MRF could be replaced with a number of other nonparametric models, such as extreme learning machines (Huang et al. 2006), structure additive regression (see, Zhu et al., 2014) or support vector machines (see, Vapnik., 2000), which would still allow simultaneous parameter identification and model discrimination via MILP because the problem is linear in the parameters. This would not be the case for classical neural network structures such as a feed-forward neural net as these are nonlinear in the parameters and would require the application of mixed integer nonlinear programming.

2.2 Estimation of the nonparametric model components

Estimating the parameters of nonlinear ODE models, such as (1) from measured data usually requires the use of iterative techniques (e.g. iterative non-linear optimisation approaches). Normally, this involves integration of many trial solutions (each set of trial ODEs has parameters supplied by the optimisation routine) until the simulation matches the measured process data. A drawback of this approach is that the numerical integration of the many trial solutions supplied by the optimiser can be highly computer resource intensive. A number of authors suggest that derivative terms may be approximated, prior to the model identification procedure, i.e. the derivative in (1) is approximated at all measured data time points, $t = 1, \dots, N$, by slopes $s(x_j)_t$. This forms an equivalent set of algebraic equations that can be solved independently from each other, i.e. the interdependencies of the ODEs have been removed.

This estimation, referred to as the inverse problem, is generally ill-posed, i.e. small errors in the concentration data can be amplified to large errors in the derivatives. The normal strategy adopted is to approximate the functional relationship between the dependent variable and time, $x_j = f(t)$. Determining an accurate (smooth) estimate of the species concentrations then allows extraction of the appropriate derivatives through differentiation (with respect to time) of the identified model. There are a number of parametric model forms that have been used to approximate data and extract the derivatives, including rational polynomials, smoothing splines, artificial neural networks etc., see, Hosten (1979); Mata-Perez and Perez-Benito (1987); Kamenski and Dimitrov (1993); Voit and Almeida (2004); Bardow and Marquardt (2004); Marquardt (2005); Burnham et al. (2008).

Successful extraction of the derivatives, recasts the overall optimisation problem, which requires numerical integration, into ‘ p ’ separate sub-problems that in general are less computer resource intensive. Furthermore, a set of linear in the parameter algebraic equations are obtained, if the denominator of the MRF is multiplied by each of the individual terms in (1) giving (where $\varepsilon_{j,t}$ represents the model prediction error of $s(x_j)_t$ at time t),

$$s(x_j)_t = \bar{f}_j(\mathbf{x})_t + \bar{g}_j(\mathbf{x})_t \sum_{k=1}^{N_m} b_{j,k} m_k(\mathbf{x})_t + \sum_{k=2}^{N_m} b_{j,k+N_m-1} m_k(\mathbf{x})_t [\bar{f}_j(\mathbf{x})_t - s(x_j)_t] + \varepsilon_{j,t} \quad (3)$$

Given the $t = 1, \dots, N$ measured data points and estimated slopes this may be written in matrix-vector format as,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon} \quad (4)$$

Where, $\mathbf{b} = [b_{j,1} \ \dots \ b_{j,r}]^T$ is a vector of model parameters, $\boldsymbol{\varepsilon} = [\varepsilon_{j,1} \ \dots \ \varepsilon_{j,N}]^T$ a vector of model prediction errors and the response vector, $\mathbf{y}$ ($N \times 1$) and matrix $\mathbf{X}$ ($N \times r$) of dependent variables is given by,

$$\mathbf{y} = \begin{bmatrix} s(x_j)_1 - \bar{f}_j(\mathbf{x})_1 \\ \vdots \\ s(x_j)_N - \bar{f}_j(\mathbf{x})_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \bar{g}_j(\mathbf{x})_1 m_1(\mathbf{x})_1 & \dots & m_{N_m}(\mathbf{x})_1 (\bar{f}_j(\mathbf{x})_1 - s(x_j)_1) \\ \vdots & \ddots & \vdots \\ \bar{g}_j(\mathbf{x})_N m_1(\mathbf{x})_N & \dots & m_{N_m}(\mathbf{x})_N (\bar{f}_j(\mathbf{x})_N - s(x_j)_N) \end{bmatrix}$$

Therefore, for each variable x_j , the unknown ODE model parameters may be determined separately using an error metric, such as the minimisation of the squared prediction error. However, the MRF superstructure will in general represent an over-parameterised model; in (2), if the degree of the numerator and denominator polynomial are assumed the same, then the total number of regression coefficients is given by, $r = (2N_m - 1)$ and all combinations of these monomials should be considered to determine the optimum nonparametric model structure. In the absence of additional constraints on the structure of the system equations a naive identification strategy, such as batch least squares, will over-fit the observed data with terms being included which model measurement noise rather than actual system dynamics. This would have negative effects on both the interpretability of the model and the portability of the model, i.e. its ability to model different instances of the system. Furthermore, as many of the monomials are comprised of similar terms, the matrix of dependent variables, $\mathbf{X}$, will be highly correlated. To develop a robust identification framework suitable to problems of practical size, an efficient sparse approximation strategy is therefore required, i.e. a method to drop irrelevant model terms.

3.0 Sparse regression

An approach to sparse regression that has gained popularity in recent years uses model parameter regularization; performing regression using the entire set of model input variables and controlling model complexity in order to improve predictive performance. To do this, an objective function is minimized that combines a measure of model prediction error and a term that penalizes model complexity. The sparse approximation metric is,

$$J(\lambda) = \|\boldsymbol{\varepsilon}\|_m + \lambda P(\hat{\mathbf{b}}) \quad (5)$$

The term $\|\boldsymbol{\varepsilon}\|_m$ is the L_m - norm of the error between the measured and predicted model output (generally specified as, $m = 2$) and the penalty term $P(\hat{\mathbf{b}})$ is a non-negative function of the estimated model parameters. The weighting (λ) is known as the (model) regularization parameter. For increasing values of λ different (sparse) solutions to (5) will be obtained (a value of $\lambda = 0$ corresponding to the non-penalized solution (where all parameters will be

present in the model) and a value of $\lambda \rightarrow \infty$ being a fully penalized solution with all model parameters at zero. The aim is to identify a value of λ that gives both the best model structure and the associated parameters. This may be achieved using cross-validation strategies using, for example, a second data set or by the use of the AIC or BIC.

The natural choice for $P(\hat{\mathbf{b}})$ is the pseudo-norm defined as the number of nonzero elements (the cardinality of the parameter vector), subsequently referred to as the L_0 - norm,

$$P(\hat{\mathbf{b}}) = \|\hat{\mathbf{b}}\|_0 = \text{card} \{ \{ \hat{b}_j | \hat{b}_j \neq 0 \} \} \quad (6)$$

3.1 Sparse regression using mixed integer linear programming

As opposed to using a least squares objective function, which is minimizing the squared error between an output and a predicted output, the measure of model prediction error used is the sum of the absolute errors, the L_1 - norm, i.e. $m = 1$ in objective function (5). This provides a popular alternative to Least Squares, L_2 - norm, because it is insensitive to outliers in the data set. Moreover, the L_1 - norm may be formulated as a linear objective function. The L_1 cost function, including the L_0 - norm regularization penalty is (7).

$$J(\lambda) = \|\boldsymbol{\varepsilon}\|_1 + \lambda \|\hat{\mathbf{b}}\|_0 \quad (7)$$

Rewriting (7) using a vector of auxiliary variables, $\mathbf{z} = (z_{j,1}, \dots, z_{j,N})^T$ and binary variables $\boldsymbol{\delta} = (\delta_{j,1}, \dots, \delta_{j,r})^T$ (where $\delta_{j,k} = 1$ if $\hat{b}_{j,k} \neq 0$ and $\delta_{j,k} = 0$ if $\hat{b}_{j,k} = 0$) gives (8). If this is minimised, subject to the constraints (9) – (12), the MILP implementation is equivalent to (7).

$$J_j(\lambda) = \sum_{t=1}^N z_{j,t} + \lambda \sum_{k=1}^r \delta_{j,k} \quad (8)$$

$$\begin{aligned} z_{j,t} &\geq \varepsilon_{j,t} \\ z_{j,t} &\geq -\varepsilon_{j,t} \end{aligned} \quad (t = 1, \dots, N) \quad (9)$$

$$L_{j,k} \delta_{j,k} \leq \hat{b}_{j,k} \leq U_{j,k} \delta_{j,k} \quad (j = 1, \dots, r) \quad (10)$$

$$\delta_{j,k} \in \{0,1\} \quad (j = 1, \dots, r) \quad (11)$$

$$z_{j,t} \geq 0 \quad (t = 1, \dots, N) \quad (12)$$

The decision variables for the MILP are **a)** the $z_{j,t}$ ($t = 1, \dots, N$) where the constraints (9) ensure the smallest possible positive values are obtained that minimize (8), **b)** the model parameters $\hat{b}_{j,k}$ ($k = 1, \dots, r$) and **c)** the binary variables, $\delta_{j,k}$ ($k = 1, \dots, r$). The $L_{j,k}$ and $U_{j,k}$ represent the upper and lower bounds on the model parameter values, $\hat{b}_{j,k}$. The constraints (10) and (12) ensure the values of $\delta_{j,k} = 1$ if $\hat{b}_{j,k} \neq 0$ and $\delta_{j,k} = 0$ if $\hat{b}_{j,k} = 0$. This is the well-known Big-M formulation (see, Griva et al., 2009) frequently used in the development of MILP models. Provided the lower ($L_{j,k}$) and upper ($U_{j,k}$) bounds are chosen to be sufficiently large a solution to the MILP will be obtained.

It may be noted that (8) uses the set of binary variables, associated with each of the parameters of the model, to perform regularization rather than the parameters themselves -

which is the case with alternative regularisation approaches such as the Least Absolute Shrinkage and Selection Operator (LASSO), Tibshirani (1994). The binary variables provide a normalised entropy measure (independent of the magnitude of the regression parameters) and are directly related to the number of parameters in the model (the cardinality of the parameter vector). This overcomes the well-known problem associated with the LASSO, where parameter shrinkage occurs as the regularisation weight is increased leading to biased estimation of model parameters (see Willis and von Stosch, 2016). Thus, the parameter identification problem is solved independently though simultaneously to the model structure discrimination problem, i.e. the problems are uncoupled. It should be noted however, the L_0 – norm is non-convex and discontinuous and solution of (1) using the L_0 – norm penalty is known to be NP-hard. For instance, when implemented as a best subset regression problem it does not scale to problem sizes where, $r > \sim 30 - 40$. However, to promote an efficient search strategy, the MILP may be formulated as a smooth, constrained problem allowing the cardinality constrained MILP to be efficiently solved for problems of practical size.

3.1.1 Efficient L_0 - norm regularization

The MILP (8) – (12) is non-smooth and therefore not as easy to solve for as, say, a Linear Program (LP). One way to overcome this difficulty is to introduce slack variables into the problem. The non-smooth, MILP can be cast into the following equivalent smooth, constrained problem which is more amenable to solution,

$$J_j(\lambda) = \sum_{t=1}^N z_{j,t} + \lambda s \quad (13)$$

$$\sum_{k=1}^r \delta_{j,k} - s \leq 0 \quad (14)$$

$$s \in \mathbb{Z}_{\geq 0}, 0 \leq s \leq N$$

The additional constraints (14) ensure that the slack variable, s , (which is an integer and an additional decision variable for the MILP) provides a smooth penalty that represents the constraint violation. At the optimum, the slack variable, s , will be equal to the value of, $\sum_{k=1}^r \delta_{j,k}$ if the constraints are satisfied. This type of formularization has been successfully used in the process control literature (see, Oliveira and Biegler, 1994; Kerrigan and Maciejowski, 2000; Richards, 2013) for the development of constrained model predictive control algorithms. Essentially, it allows the MILP to ‘soften’, i.e. violate the constraints, if no alternative solution can be found thereby promoting a more efficient search. For increasing numbers of regression parameters, r , the solution of this problem has been found to be much more computational efficient than the original problem. It seems to steer the branch-and-bound algorithm towards using a breadth-first search strategy, which can be expected to be more efficient than a depth-first search, for cases where few terms are contained in the model.

4.0 Model structure selection

The choice of the optimal regularization parameter (λ) is an important issue and this may be achieved using a model validation strategy. The standard approach is to use cross validation determining the optimal λ by finding the minimum of the error on a test data set. However, choosing the regularization parameter in this manner can be computationally intensive.

Alternative approaches are to use information criteria such as AIC and BIC. It is known that AIC-based methods are not consistent for model selection as irrelevant model parameters tend to be selected, see Shao (1993). Therefore, in this work the BIC criterion is used which may be described by,

$$BIC = -2 \ln(lik) + \ln(N)df$$

Where '*lik*' is the maximum value of the likelihood function of the model, *df* is the number of parameters (degree of freedom) of the model. Using the **least absolute deviation** (LAD) cost function the BIC cost function is,

$$BIC(\lambda) = N \ln \left(\frac{\|\varepsilon\|_1}{N} \right) + \ln(N)df \quad (15)$$

An optimal model structure corresponds to the regularization parameter λ that minimizes (15). Therefore, cost function (9) or (13) may be minimised for a range of λ values in order to determine the regularization path (or landscape) and the corresponding values of (15) are calculated to determine the optimal model structure. For LASSO, Zou, Hastie, and Tibshirani (2007) prove that the number of non-zero coefficients within the model is an unbiased estimator of the model degree of freedom, *df*; for the MILP regularization strategy this is the sum of the binary variables associated with each model parameter.

5.0 Refining the hybrid model structure

Rearranging and re-parameterization of nonlinear models to one that is linear in a new parameter set e.g. methods such as Lineweaver-Burk linearization, has been criticised many times. The re-parametrization alters the relationship between the dependent and independent variables leading to an error in the variables regression problem which can lead to biased parameter estimation, see Norton (1986), Cornish-Bowden (2002). In this work, to mitigate the potential bias in estimation of the parameters the smoothed derivatives and corresponding smoothed measurements are used. However, having established the model components through sparse regularization of the decoupled ODEs, the final model structure is re-identified to ensure statistical optimality of the model parameter estimates. Instead of the decoupling method described previously, a full simultaneous optimisation is therefore performed by numerically integrating the ODEs using the initial conditions (and any inlet flow profiles of the original experiments). Parameter estimation requires the use of iterative optimisation techniques (e.g. gradient descent methods) to determine the kinetic rate constants. This involves repeated numerical solution of the ODEs using many trial parameter sets until the simulation closely matches the experimental data. A schematic of the kinetic fitting strategy is shown in Fig. 1.

Fig. 1. Kinetic fitting of an ODE model. In this work, this was implemented in MATLAB. The ODEs are integrated (using ODE45) and the species concentration obtained using an initial estimate of the unknown model parameters are compared to the measured species concentration data. A nonlinear optimizer (fmincon) then iteratively adjusts model parameters to obtain a model that approximates the data as close as possible. Where appropriate, constraints are applied e.g. bounds on the values of rate constants etc.

The integrated trial solutions are compared with the measured, non-smoothed data using the cost function (5), with $m = 2$, i.e. the squared error between the measured and predicted species concentrations. Note that because the structure of the hybrid model is now assumed to be known, a penalty term in the cost function is no longer required.

6.0 Case studies

To demonstrate the hybrid semi-parametric modelling strategy in this section of the paper two case studies are presented. The first example, considers the estimation of a hybrid model of a fed-batch bioreactor, where Monod and Haldane kinetics are considered. The second is an application to the identification of the simulated dynamics of fed-batch reactor, where the (assumed unknown) reaction rate terms are modelled using the assumption of mass action kinetics (a set of differential equations with a sparse subset of polynomial terms). Both examples incorporate data smoothing and derivative estimation and highlight the iterative steps used in the development of a hybrid model as well as the application of equality/inequality constraints within the MILP framework. **Though the kinetics of the two examples are rather simple, they serve as to demonstrate that the proposed methodology can correctly identify the “true” underlying model structure.**

In all the results presented, **a)** the derivatives are estimated using smoothing splines, ‘eye-balling’ the approximation and adjusting the spline’s smoothing parameter until an accurate (smooth) approximation is obtained. The estimated derivatives are then obtained through analytical differentiation of the resulting approximation. **b)** the MILP is solved using the function ‘intlinprog’ with default settings in MATLAB.

6.1 Case study one

The development of a model describing biomass growth on substrate uptake is considered. A set of assumed unknown ODEs representing the material balances for biomass (X) and substrate (S) were numerically solved to generate simulated experimental data over a period of 15 hours. Two fed-batch experiments were used to generate the data. In the first fed-batch experiment, the initial conditions are $X_0 = 1 \text{ g.litre}^{-1}$, $S_0 = 0.5 \text{ g.litre}^{-1}$, $V_0 = 5 \text{ litre}$ and in the second, $X_0 = 1 \text{ g.litre}^{-1}$, $S_0 = 2 \text{ g.litre}^{-1}$, $V_0 = 5 \text{ litre}$. In order to provide a rich data set for the purposes of system identification the two fed-batch experiments were operated using different substrate feeding strategies. For the first experiment, the initial 6.9 hr were carried out in batch mode and subsequently an exponential feeding strategy was used,

$$F = 0.043 \exp(0.6(t - 6.9)) .$$

In the second experiment, a linear feeding strategy was used,

$$F = 0.01 + 0.035t$$

Noisy measurements of the concentrations of the species within the reactor (5% Gaussian noise was added to the measurements) – as well as non-noisy inlet flowrates and inlet concentrations were sampled with a total of 100 samples being collected over the two batch experiments. The objectives are to, firstly, develop a hybrid model that provides satisfactory predictions of the dynamic behaviour of the system and secondly, use the model to interpret the MRF parameters as a structural model of the underlying chemical reactions, i.e. gain an understanding of the kinetics.

Two versions of assumed unknown kinetics describing biomass growth on substrate are considered. First, specific biomass growth rate was described by Monod kinetics ($K_S = 0.5 \text{ g.litre}^{-1}$, $\mu_{max} = 0.4 \text{ hr}^{-1}$) with the specific substrate uptake being proportional using a yield coefficient ($Y = 1.2$). Secondly, the specific biomass growth rate was specified by Haldane kinetics ($\mu_{max} = 0.4 \text{ hr}^{-1}$, $K_S = 0.5 \text{ g.litre}^{-1}$, $K_I = 1/75 \text{ g}^2.\text{litre}^{-2}$) with the specific substrate uptake rate again being proportional by the yield coefficient ($Y = 1.2$).

6.1.1 Hybrid model structure

The first step in developing a hybrid model is to formulate a set of equations that includes as much process knowledge as possible to describe the temporal behaviour of the system. In fed-batch (bio) chemical reaction systems, the behaviour of process design variables, such as flowrates and inlet species concentrations, are often relatively easily quantifiable components of the system model, whereas, it is more difficult to specify the structure of a set of rate equations that satisfactorily account for the change in species concentrations due to the (bio) chemical reactions. Assuming a homogeneous, well mixed, constant density system the system equations may be written as,

$$\begin{aligned}\frac{dX}{dt} &= -\frac{F}{V}X + X\mu \\ \frac{dS}{dt} &= \frac{F}{V}(S_f - S) - YX\mu \\ \frac{dV}{dt} &= F\end{aligned}$$

In these ODEs, X and S are the biomass (g.litre^{-1}) and substrate (g.litre^{-1}) concentrations, S_f is the inlet substrate concentration (g.litre^{-1}), F (litre.hr^{-1}) is the inlet flowrate (a single inlet is assumed for simplicity) and V (e.g. litre) of the reaction mixture within the vessel at time, t (hr). μ is the specific rate of biomass growth (hr^{-1}), and Y the yield coefficient associated with substrate consumption. Referring to (1), the parametric model structure therefore defines,

$$\bar{f}_1(\mathbf{x}) = -\frac{F}{V}X, \quad \bar{g}_1(\mathbf{x}) = X, \quad \bar{f}_2(\mathbf{x}) = \frac{F}{V}(S_{in} - S), \quad \bar{g}_2(\mathbf{x}) = -YX$$

The non-parametric model component is therefore the model for the specific biomass growth rate, i.e. $\mu = \hat{f}_j(\mathbf{x})$. In other words, due to the incorporation of structural knowledge hybrid model identification becomes one of just estimating the functional form of the specific biomass growth rate model. Given the structure of the ODEs this is typically achieved using the differential equation describing the rate of change of biomass with respect to time (Tholudur and Ramirez, 1996; Tobajas and Garcia-Calvo, 2000) where,

$$\frac{dX}{dt} + \frac{F}{V}X = \frac{1}{XV} \frac{d(XV)}{dt} = \hat{f}_1(\mathbf{x}) \quad (16)$$

However, given the non-parametric component of the model is a MRF structure, the derivative of (XV) with respect to time may also be approximated giving (where XV augments the terms in numerator of the MRF),

$$\frac{d(XV)}{dt} = XV\hat{f}_1(\mathbf{x}) \quad (17)$$

This formularization of the identification and discrimination problem is possible as the nonparametric component of the hybrid model is a MRF as opposed to e.g. a neural network.

6.1.2 Estimation of the derivatives

To estimate the derivatives a cubic smoothing spline was used to approximate the term XV using the MATLAB function `csaps` ($P = 0.9$). The smoothing spline was then analytically differentiated with respect to time and the differential evaluated at the instances at which concentration measurements were available. An estimate of the specific biomass growth rate was then obtained by dividing the resulting derivative estimates by the term, XV . The first and last five estimated derivative points were removed from each set, since it is known that at the start and end points the quality of the estimated derivatives is poor (Brendel et al., 2006; Kahrs and Marquardt, 2008).

The actual (taken from the assumed unknown, ODEs) and estimated specific growth rates for the two fed batch reactions with Monod kinetics are shown in Fig 2a while the Haldane kinetics are shown in Fig2b. It may be noted that the estimates differ considerably in comparison to the actual derivatives. This was despite the fact that the feeding regimes were designed to yield slow variations in the substrate concentration, where the aim was to obtain smooth rates of change of biomass and substrate concentrations that differentiate the impact of measurement noise. Fig 2c and 2d show the derivative of XV with respect to time which are considerably more accurate apart from at the end of the second fed-batch experiment. During this phase, the underlying system dynamics, make the smooth approximation of the derivatives using splines more challenging. However, a key observation from Fig. 2 is that the variation in the estimate of the specific growth rate is caused by the division of the derivative of XV by the term, XV . In light of this, the relative merits of the use of (16) or (17) to estimate the MRF model parameters is considered in the next section of the paper.

(a) Actual specific growth rate (red dots) obtained from the simulation and the derivatives (solid black line) obtained by analytically differentiating the fitted cubic spline to the measured data (Monod kinetics).

(b) Actual specific growth (red dots) obtained from the simulation and the derivatives (solid black line) obtained by analytically differentiating the fitted cubic spline to the measured XV (Haldane kinetics).

(c) Actual XV derivatives (red dots) obtained from the simulation and the derivatives (solid black line) obtained by analytically differentiating the fitted cubic spline to the measured XV (Monod kinetics).

(d) Actual XV derivatives (red dots) obtained from the simulation and the derivatives (solid black line) obtained by analytically differentiating the fitted cubic spline to the measured XV (Haldane kinetics).

Fig. 2. The actual and estimated derivatives (specific growth rate for a) and b) and the derivative of XV for c) and d)). The actual derivatives are taken from the simulated set of ODEs used to generate the experimental data. They are shown in order to highlight the accuracy of the derivative approximation.

6.1.3 Estimation of the MRF parameters using the experimental data

A MRF of degree, $d = 3$ was specified, corresponding to a MRF superstructure containing seven unknown terms and 5040 possible structures from which any feasible subset would normally have to be evaluated manually. To perform model regularization, λ was increased from an initial value of 0.1 to a final value of λ_{max} over fifty log-spaced intervals. The value of λ_{max} is taken to be the sum of the absolute values of the estimated derivatives (the value of the LAD cost function when all MRF parameters are zero). The upper and lower bounds on the parameter values were specified as, $(L_{j,k}, U_{j,k}) = (-5, 5)$. To ensure a feasible MRF structure that may be used for the purposes of prediction, a constraint was applied to ensure that at least one coefficient within the numerator of the MRF must be non-zero, i.e. $b_{j,1} \vee b_{j,2}, \dots, \vee b_{j,N_m} \neq 0$. Using the set of binary variables associated with each of the model parameters, the appropriate constraint is, $\delta_{j,1} + \delta_{j,2} + \dots + \delta_{j,N_m} \geq 1$. Furthermore, given the structure of (bio) chemical rate expressions, the constraints that the coefficients in the denominator of the MRF are always positive were applied. This constraint was imposed by tightening the lower parameter bound constraint on the parameter values within the denominator of the MRF, $L_{j,k} (k = N_m + 1, \dots, r) = 0$.

The L_0 -norm regularization landscape for the parameter values using (17) are shown in Fig. 3a and 3c. Fig. 3b and 3d shows the corresponding BIC value and model degrees of freedom (DoF). For the data extracted from the simulation using Monod kinetics (Fig. 3a and 3b), four distinct changes can be observed in model parameter profiles and the minimum BIC values indicate the best performing model structure, which is initially obtained with a regularization weight of $\lambda = 279$, was,

$$\mu = \frac{0.789S_1}{1 + 1.9381S_1}$$

This corresponds to an estimate of the K_S value of 0.516 g.litre⁻¹ and an estimate of the μ_{max} value of 0.407 hr⁻¹. In comparison to the actual values used to generate the data (0.5 g.litre⁻¹ and 0.4 hr⁻¹ respectively) the identified values only exhibit minor discrepancies (bias). Using

(16), the following estimate of the specific growth rate was obtained (the model structure corresponding to the minimum BIC),

$$\mu = \frac{0.8201S_1}{1 + 2.0615S_1}$$

This corresponds to an estimate of the K_S value of 0.48 g.litre⁻¹ and an estimate of the μ_{max} value of 0.397 hr⁻¹, which exhibit a larger parameter bias. This would be expected as the errors in the derivative estimation propagate to errors in the numerator of the MRF given the linear re-parameterization (3).

With regard to the regularization landscape for the data extracted from the simulation based on the Haldane kinetics more changes in the parameter values may be observed when compared to the Monod model (perhaps because the underlying MRF being estimated is slightly more complex). The best model structure (with a minimum BIC value) obtained was ($\lambda = 128$),

$$\mu = \frac{0.5763S_1}{1 + 1.2116S_1 + 0.0357S_1^2}$$

The respective kinetic parameter values are $K_S = 0.8254$ g.litre⁻¹, $\mu_{max} = 0.4757$ hr⁻¹ and $K_I = 1/33.89$ g².litre⁻². In comparison to the actual values ($K_S = 0.5$ g.litre⁻¹, $\mu_{max} = 0.4$ hr⁻¹, $K_I = 1/75$ g².litre⁻²), all parameters show a significant bias although the greatest deviation can be observed for K_I . Using (16), the following estimate of the specific growth rate was obtained (the model structure corresponding to the minimum BIC),

$$\mu = \frac{0.6644S_1}{1 + 1.4607S_1 + 0.0392S_1^2}$$

Here, the respective kinetic parameter values are $K_S = 0.6846$ g.litre⁻¹, $\mu_{max} = 0.4549$ hr⁻¹ and $K_I = 1/37.36$ g².litre⁻² (again while the correct structure has been obtained, the parameters demonstrate a significant bias). This parameter estimation bias (using either of the derivative estimation techniques) may be due to **a**) the data only containing a short period in which substrate inhibition occurred (samples 58 to 79 see Figure 2b and 2d) or **b**) the variation/ approximation errors in the estimated derivatives during substrate inhibited conditions masking the true underlying function, i.e. inhibition. Improved estimation accuracy may have been obtained by simultaneous estimation of the substrate balance equation using suitably defined constraints. However, sufficient structural information has been obtained to estimate the model parameters using kinetic fitting via a nonlinear optimiser as described in section 5. Performing this kinetic fitting for the simulated system where the underlying data was generated using the Haldane kinetics gave the following estimate of specific growth rate (initial parameter estimates were specified as $K_S = 0.8254$ g.litre⁻¹, $\mu_{max} = 0.4757$ hr⁻¹ and $K_I = 1/33.89$ g².litre⁻²),

$$\mu = \frac{0.4079S_1}{0.5455 + S_1 + 1/64.7S_1^2}$$

Though the parameter values are closer to the actual values than before, still a significant deviation to the actual value can be observed in case of K_I . This seems to indicate that the bias in the parameters is due to the data containing a limited period where substrate inhibition occurred.

(a) Model regularisation using the L_0 norm (Monod kinetics)

(b) BIC Criterion and the number of model parameters (DoF) (Monod kinetics)

(c) Model regularisation using the L_0 norm (Haldane kinetics)

(d) BIC Criterion and the number of model parameters (DoF) (Haldane kinetics)

Fig. 3 Model regularization using the L_0 norm. (a) and (c), the coefficient profiles as the regularization parameter is increased. (b) and (d) the BIC criterion and the number of model parameters (DoF) as a function of the regularization parameter .

6.2 Case study two

The reaction system contains five chemical species, $x_1, \dots, x_5$. A set of assumed unknown ODEs were numerically solved to generate simulated experimental data. A single fed-batch experiment was used, with the initial reactor volume set to $V = 5$ litre. The initial conditions of each of the chemical species was set to zero, the physical interpretation of this is that the reactor is initially charged with an inert solvent. An inlet feed stream is then used to dose reactant x_1 where,

$$t > 0, F = 1 \text{ litre.min}^{-1}, [x_{1,in}] = 10 \text{ mol.litre}^{-1}$$

$$t > 4, F = 0 \text{ litre.min}^{-1}$$

$$t > 12, F = 1 \text{ litre.min}^{-1}, [x_{1,in}] = 10 \text{ mol.litre}^{-1}$$

Noisy measurements of the concentrations of the species within the reactor (5% Gaussian noise was then added to the measurements) – as well as non-noisy inlet flowrates and inlet concentrations were sampled at an interval of 15 seconds and a total of 100 samples were collected. The objectives are to, firstly, develop a hybrid model that provides satisfactory

predictions of the dynamic behaviour of the system and secondly, use the model to interpret the MRF parameters as a structural model of the underlying chemical reactions, i.e. gain an understanding of the network of chemical reactions.

6.2.1 Specification of the hybrid model structure

Assuming a homogeneous, well mixed, constant density system the component balances for each of the chemical species ($p = 5$) were specified as,

$$\frac{dx_j}{dt} = \frac{F}{V}(x_{j,in} - x_j) + \frac{\bar{g}_j(\mathbf{x}) \sum_{k=1}^{N_m} b_{j,k} m_k(\mathbf{x})}{1 + \sum_{k=2}^{N_m} b_{j,k} + N_m - 1} m_k(\mathbf{x}), (j = 1, \dots, p)$$

$$\frac{dV}{dt} = F$$

In these ODEs, F (litre.min⁻¹) is the inlet flowrate (a single inlet is assumed for simplicity) and V is the volume (litre) of the reaction mixture within the vessel at time, t (min). The x_j represent each of the p species concentrations and $x_{j,in}$ represents the species concentrations in the inlet stream. The first term in the model therefore describes the effect of net inflow of material in the fed-batch system, referring to (1) this is specified $\bar{f}_j(\mathbf{x}) = \frac{F}{V}(x_{j,in} - x_j)$, $\bar{g}_j(\mathbf{x}) = 1$ with the nonparametric model component being used to represent the (assumed unknown) kinetic rate terms. Note that in these ODEs, $x_{2,in}$, $x_{3,in}$, $x_{4,in}$, $x_{5,in}$ are always zero as only species x_1 is dosed into the reactor.

6.2.2 Estimation of the rate of change of species concentrations with respect to time

The derivative of each of the chemical species were estimated (fitting a spline to the concentration data using the MATLAB function csaps with $P = 0.9$; differentiating the spline with respect to time; and evaluating the derivative of the spline at the sample instances). As an example, the approximation (for species, x_1) is shown in Fig. 4. It may be observed that the approximated derivatives closely match the actual derivatives; however there are discrepancies at the start of the fed-batch run and at the times corresponding to a change in the dosing regime in the fed-batch reactor ($t = 0, 4, 12$). As with case study one, the first and last five estimated derivative points were removed from the data set.

(a) Measured (noisy) concentration data for species x_1 (red dots). The smoothed value (solid black line) of $[x_1]$ is obtained using the cubic smoothing spline function in MATLAB, with a smoothing parameter, $P = 0.9$.

(b) Actual derivatives (red dots) obtained from the simulation and the derivatives (solid black line) obtained by analytically differentiating the fitted cubic spline.

Fig. 4. Generating a smooth approximation of the species concentration (a) and using the resulting cubic smoothing spline to approximate the derivatives, shown in (b). In (b) the actual derivatives taken from the simulated set of ODEs are also shown in order to highlight the accuracy of the approximation.

6.2.3 Estimation of the MRF parameters using the experimental data

For each of the species, a MRF of degree, $d = 2$ was specified, corresponding to a MRF superstructure containing 41 unknown terms which are combinations of each of the five chemical species, **41! possible model structures**. To perform model regularization, λ was increased from an initial value of 0.1 to a final value of λ_{max} over fifty log-spaced intervals. The value of λ_{max} is taken to be the sum of the absolute values of the estimated derivatives (the value of the LAD cost function when all MRF parameters are zero). The upper and lower bounds on the parameter values were initially specified as, $(L_{j,k}, U_{j,k}) = (-5, 5)$. As with case study one, constraints were specified to ensure that at least one coefficient within the numerator of the MRF must be non-zero and that the coefficients in the denominator of the MRF are positive. In addition, to be physically consistent, isothermal rate constants should be positive values. However, the sign of the constant multiplying a monomial within any differential equation will also be dictated by the stoichiometry. Therefore, when developing a model for the j^{th} ODE, if a particular monomial $m_k(\mathbf{x})$ contains species x_j , indicating that the species ‘reacts’ the associated parameter value $b_{j,k}$ should be negative. If the monomial does not contain x_j , indicating that species x_j is formed from other species the value of $b_{j,k}$ should be positive. This constraint may be applied within the MILP by appropriate specification / adjustment of the upper and lower estimated parameter bounds.

Fig. 5. Model regularization using the L_0 norm. (a) The coefficient profiles as the regularization parameter is increased. (b) The BIC criterion and the number of model parameters (DoF) as a function of the regularization parameter.

Fig. 5 shows the regularization landscape using the L_0 – norm penalty (for species x_1). The optimal model according to the BIC criterion contained two terms. Derivative estimation, followed by model parameter regularization was repeated for each of the species (figures not shown for sake of brevity) which resulted in the following hybrid model structure (the MRF components corresponding to the minimum BIC for each species populating the model structure),

$$\begin{aligned}
\frac{dx_1}{dt} &= \frac{F}{V}(x_{1,in} - x_1) - 0.1244x_1^2 - 0.34x_1 \\
\frac{dx_2}{dt} &= \frac{F}{V}(x_{2,in} - x_2) + 0.0980x_1^2 \\
\frac{dx_3}{dt} &= \frac{F}{V}(x_{3,in} - x_3) + 0.1015x_1 - 0.1537x_3 \\
\frac{dx_4}{dt} &= \frac{F}{V}(x_{4,in} - x_4) + 0.1381x_3 - 0.2212x_1x_4 \\
\frac{dx_5}{dt} &= \frac{F}{V}(x_{5,in} - x_5) + 0.2578x_1x_4 \\
\frac{dV}{dt} &= F
\end{aligned}$$

The first point to note with this structure is that it consists of sparse polynomial terms, i.e. no denominator terms within the MRF superstructure are identified. Secondly, it should be possible to infer a consistent topology of the chemical reaction network by considering each of the terms in the hybrid model structure. Thus, through inspection of the first ODE it may be noted that it comprises two variables x_1 and x_1^2 with negative constants, implying that species x_1 reacts in two separate reactions. The second ODE contains the single term x_1^2 with a positive constant. According to the law of mass action kinetics this would suggest the reaction, $2x_1 \rightarrow x_2$, although there is a discrepancy between the numerical values of the estimated parameters. The third and fourth ODEs indicate that the following two reactions occur, $x_1 \rightarrow x_3 \rightarrow x_4$. Within the fourth and fifth ODEs there are the terms, x_1x_4 indicating that the following reaction may also occur, $x_4 + x_1 \rightarrow x_5$. However, if this were the case, to have a consistent set of ODEs the monomial x_1x_4 should also appear within the first ODE.

The use of MILP with appropriate equality constraints offers the possibility of repairing such model deficiencies once they have been established. Therefore, to develop a consistent set of ODEs describing a feasible chemical reaction network, a further regularized regression was performed for species x_1 only, specifying that the binary variable associated with the monomial x_1x_4 was equal to one, i.e. forcing the structural component, x_1x_4 to be part of the model. This gave the following ODE,

$$\frac{dx_1}{dt} = \frac{F}{V}(x_{1,in} - x_1) - 0.1833x_1^2 - 0.2106x_1 - 0.3244x_1x_4$$

The addition of this additional structural component decreased the mean square obtained during training to a value of 0.0642. The error obtained when two components were in the model was 0.0663, the more parsimonious model is obviously favoured during the regularisation process because of the use of the L_0 norm penalty. In order to refine the parameters of the final model structure they were re-identified to ensure statistical optimality of the model parameter estimates (as described in section 5). The estimated rates constants obtained were, $\hat{k} = [0.103 \quad 0.206 \quad 0.152 \quad 0.273]$ which are in close agreement to those used in the simulation to generate the data, $k = [0.1 \quad 0.2 \quad 0.15 \quad 0.25]$ using the reaction network; $2x_1 \rightarrow x_2$, $x_1 \rightarrow x_3 \rightarrow x_4$ and $x_4 + x_1 \rightarrow x_5$ (correctly deduced during model parameter regularization).

7.0 Discussion and conclusions

This work has focused on the development of a hybrid semi-parametric modelling strategy using a MRF superstructure to describe the unknown ODE model components (the nonparametric model terms). The framework uses derivatives that are estimated from measured process output data (for reaction systems this is concentration data), it therefore falls within the realm of the so called incremental approach to system identification (Brendel

et al., 2006; Kahrs and Marquardt, 2008). However, as opposed to the trial and error testing of different **parametric and nonparametric model structures**, using a MILP framework it has been demonstrated that it is possible to identify both the structure and model parameters simultaneously using sparse regression. This has the potential to greatly reduce hybrid model development time when compared to methods that capture the optimal model structure using a trial and error approach.

The sparse regression method used in this work minimised the L_1 - norm of the model residuals, which is robust to outliers, and used as a model penalty the cardinality of the model parameter vector, i.e. the L_0 - norm of the parameters. Implemented using MILP, the cardinality of the model parameter vector is determined using the sum of binary variables, where each binary variable is determined such that for a value of zero means the associated parameter value is zero, and a value of one indicates a non-zero parameter. This offers a novel regularisation strategy, which decouples the parameter estimation and structure identification problems; and as demonstrated in (Willis and von Stosch, 2016) eliminates the effect of parameter shrinkage associated with techniques such as the LASSO. **As further work, it would be interesting to apply this model regularisation strategy to the determination of the optimal model structure of alternative empirical modelling methods such as structure additive regression, extreme learning machines and support vector machines. For example, extreme learning machines are single layer feedforward neural networks, whose hidden layer nodes are randomly parameterised and the output weights are obtained via a least squares regression method (typically the Moore Penrose pseudo-inverse is used). This allows fast network learning but can result in large networks which may have poor generalisation properties. Regularisation using the L_0 - norm penalty would allow the optimal network structure to be determined. Similarly, for support vector regression it may be possible to use the penalty to choose the optimal number of support vectors.**

The case studies presented in this paper were used to demonstrate that the suggested approach provides a principled hybrid semi-parametric model development strategy. The hybrid model structure allows known process model structures to be incorporated in advance of system identification studies which can significantly reduce the system identification problem, as demonstrated in case study one. Furthermore, as the nonparametric model component is a MRF it has intuitive appeal when compared to e.g. the use of a neural network as the nonparametric model component. It relates to empirical model structures used in (bio) chemical modelling such as the Monod and Haldane equation as well as those derived from theoretical considerations, such as mass action kinetics, Michaelis-Menten kinetics etc. Furthermore, using MILP constraints may be systematically incorporated into the identification procedure. For example, in case study two this was used to ensure that parameter values, which have the interpretation of isothermal rate constants, have the appropriate sign, e.g. negative for a reactant species, positive for a product species. It should be noted however, that in alternative applications the constraint set could be expanded. For instance, a useful application would be to utilise structural constraints, which can be defined using the binary variables, to identify a common model structure using experimental data that have been gathered under different operating conditions, e.g. different temperatures or pH values. This would allow the estimation of different parameter values for the same model structure, as may be expected if experimental conditions changed.

A potential disadvantage **of the approach arises from the transfer of the ODEs into an equivalent set of algebraic equations, as the interdependencies of the ODEs are removed** – something that is common with all methods that adopt the incremental approach to model identification. This was demonstrated in case study 2, where the initial model produced a set of ODEs that were not physically consistent. Manual interrogation of the structural

components of the ODEs enabled the identification of structural deficiencies highlighting the fact that through the use of MILP framework specification of appropriate equality constraints offers the possibility of repairing such model deficiencies once they have been established.

Finally, it cannot be over-emphasised that the successful transformation of the ODEs into an equivalent set of algebraic equations relies on accurate derivative estimation as well as a rich and appropriately designed data set. As a result of the current trend in the process industries to install online process analysers, such as spectroscopic devices, an almost continuous signal of species measurements (high sampling frequency) may become available. This should allow a more accurate estimation of the derivatives and in turn, add to the integrity of the decoupled approach to hybrid model development. However, if the time derivative suffer from significant discontinuities (step disturbance, etc.) their approximation may not avoid substantial error. It therefore would be interesting to investigate alternative decoupling approaches that avoid derivative estimation.

8.0 References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE T. Automat. Contr.* 19, 716-723.
- Bardow, A. and Marquardt, W. (2004), *Chemical Engineering Science*. 2004; 59(13): 2673–2684.
- Bonvin, D., and Rippin, D.W.T. (1990) Target factor analysis for the identification of stoichiometric models. *Chem. Eng. Sci.*, 45(12), 3417-3426.
- Brendel, M., and Marquardt, W. (2008). Experimental design for the identification of hybrid reaction models from transient data. *Chemical Engineering Journal*, 141, 264–277.
- Brendel, M., Bonvin, D. and Marquardt, W. (2006) Incremental identification of kinetic models for homogeneous reaction systems, *Chem. Eng. Sci.*, 61, 5404-5420
- Burnham, S.C., Searson, D.P., Willis, M.J. and Wright, A.R. (2008) Inference of chemical reaction networks, *Chem. Eng. Sci.*, 63(4), 862-873.
- Chen, L., Bernard, O., Bastin, G., and Angelov, P. (2000). Hybrid modelling of biotechnological processes using neural networks. *Control Engineering Practice*, 8, 821–827.
- Cornish-Bowden, A. (2002) Statistical analysis of enzyme kinetic data. In Eisenthal, R. and Danson, M. (Eds.) *Enzyme Assays*, Oxford University Press, New York.
- Fiedler, B., and Schuppert, A. (2008). Local identification of scalar hybrid models with tree structure. *IMA Journal of Applied Mathematics*, 73, 449–476.
- Georgieva, P., and de Azevedo, S. (2009). Computational intelligence techniques for bioprocess modelling, supervision and control, Volume 218. Berlin/Heidelberg:Springer.
- Griva, I., Nash, S.G. and Sofer, A. (2009) *Linear and Nonlinear Optimization*, Society for Industrial and Applied Mathematics, Philadelphia.
- Hesterberg, T, Choi, N.H, Meier, L and Fraley, C. (2008) Least angle and ℓ_1 penalized regression: a review, *Statistical surveys*, 2, 61-93
- Hosten, L.H. (1979) A comparative study of short cut procedures for parameter estimation in differential equations, *Comput. Chem. Eng.*, 3, 117–126.
- Huang, G.B., Zhu, Q.Y, Siew, C.K. (2006) *Extreme Learning Machines: theory and applications*, *Neurocomputing*, 70 (1), 489-501
- Kahrs, O. and Marquardt, W. (2008) Incremental identification of hybrid process models. *Comput. Chem. Eng.*, 32(4-5): 694-705.
- Kamenski, D.I. and Dimitrov, S.D. (1993) Parameter estimation in differential equations by applications of rational functions, *Comput. Chem. Eng.*, 17 (7), 643–651.

- Kerrigan, E. and Maciejowski, J. (2000) Soft constraints and exact penalty functions in model predictive control, in Control 2000 Conference, Cambridge.
- Maria, G. (2004) A review of algorithms and trends in kinetic model identification, *Chemical and Biochemical Engineering Quarterly*, 18 (3), 195–222.
- Marquardt, W. (2005) Model based experimental analysis of kinetic phenomena in multi-phase reactive systems, *Chem. Eng. Res. And Des.*, 83, 561 – 573.
- Mata-Perez, F. and Perez-Benito, J.F. (1987) The kinetic law for autocatalytic reactions, *Journal of Chemical Education* 64, 11, 925–927.
- Norton, J.P. (1986) An introduction to identification, Academic Press, London.
- Oliveira N.M.C and Biegler, L.T. (1994) Constraint handling and stability properties of model predictive control, *AIChE Journal*, 40, 1138–1155.
- Oliveira, R. (2004) Combining first principles modelling and artificial neural networks: a general framework. *Comput. Chem. Eng.*, 28, 755-766.
- Psichogios, D. C., and Ungar, L. H. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, 38, 1499–1511.
- Richards, A. (2013) Fast Model Predictive Control with Soft Constraints, European Control Conference (ECC)
- Schubert, J., Simutis, R., Dors, M., Havlik, I., and Luebbert, A. (1994). Hybrid modelling of yeast production processes—Combination of a priori knowledge on different levels of sophistication. *Chemical Engineering and Technology*, 17, 10–20.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6, 461-464.
- Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.* 88, 486-494.
- Tholudur, A., and Ramirez, W. F. (1996) Optimization of fed-batch bioreactors using neural network parameter function models. *Biotechnology Progress*, 12, 302–309.
- Thompson, M. L., and Kramer, M. A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40, 1328–1340.
- Tibshirani, R. (1994) Regression shrinkage and selection via the lasso. Technical report, University of Toronto.
- Tobajas, M. and Garcia-Calvo, E. (2000) Comparison of analysis methods for determination of the kinetic parameters in batch cultures, *World Journal of Microbiology and Biotechnology* 16: 845.
- Vapnik, V.N., (2000) The nature of statistical learning theory, second edition, Springer-Verlag, New York.
- Van de Wouwer, A., Renotte, C., and Bogaerts, P. (2004). Biological reaction modeling using radial basis function networks. . *Comput. Chem. Eng.*, 28, 2157–2164.
- Voit E.O. and Almeida J. (2004) Decoupling dynamical systems for pathway identification from metabolic profiles, *Bioinformatics*, 20(11): 1670-1681.
- von Stosch, M., Oliveira, R. Peres, J. and de Azevedo, S.F. (2014) Hybrid semi-parametric modeling in process systems engineering: Past, present and future, *Comp. Chem. Eng.*, 60, 86 – 101
- Willis and von Stosch (2016) Inference of Chemical Reaction Networks using Mixed Integer Linear Programming, *Comp. Chem. Eng.*, 90, 31-43
- Zou, H. Hastie, T. and Tibshirani, R. (2007) On the “degrees of freedom” of the lasso *Ann. Statist.* 35, 5, 2173-2192.
- Zou, H., Yao, F. and Zhang, HH. (2014) Structured functional additive regression in reproducing kernel hilbert spaces, *Royal Statistical Society, Series B, Statistical Methodology*, 76, 3, 581-603