

A diagnostic procedure for improving the structure of approximated kinetic models

Marco Quaglio¹, Eric S. Fraga¹, and Federico Galvanin¹

¹*Department of Chemical Engineering, University College London (UCL), Torrington Place, WC1E 7JE London, United Kingdom*

corresponding author email: f.galvanin@ucl.ac.uk

Abstract

Kinetic models of chemical and biochemical phenomena are frequently built from simplifying assumptions. Whenever a model is falsified by data, its mathematical structure should be modified embracing the available experimental evidence. A framework based on maximum likelihood inference is illustrated in this work for diagnosing model misspecification and improving the structure of approximated models. In the proposed framework, statistical evidence provides a measure to justify a modification of the model structure, namely a reduction of complexity through the removal of irrelevant parameters and/or an increase of complexity through the replacement of relevant parameters with more complex state-dependent expressions. A tailored Lagrange multipliers test is proposed to support the scientist in the improvement of parametric models when an increase in model complexity is required.

keywords: process-model mismatch, diagnosis, information, Lagrange multipliers

1 Introduction

The accurate description of kinetic phenomena requires the construction of systems of differential and algebraic equations where a high number of parameters and state variables may be involved. The complexity associated with kinetic phenomena frequently leads to the construction of model structures characterised by some degree of approximation. Whenever an approximated kinetic model is falsified by observations, its mathematical structure shall be modified embracing the available experimental evidence. Nonetheless, the selection of an appropriate update for the model equations is generally not obvious, making the modelling of kinetic phenomena a time and resource intensive task.

In standard model building frameworks, the modelling activity typically begins with the construction of a set of competing model structures (Asprey and Macchietto, 2000). The parameters involved in the available models are estimated by fitting the available experimental data (Bard, 1974). One may then compare the relative performance of the available models using an information criterion, e.g. the Akaike information criterion (Akaike, 1974) or the

Bayesian information criterion (Neath and Cavanaugh, 2012). Nevertheless, information criteria provide a relative measure of merit to compare alternative model structures but do not directly quantify the model performance in representing the distribution of the data. One may check model adequacy with a statistical test on the goodness-of-fit (Silvey, 1975). The goodness-of-fit test may inform on the presence of some modelling error, namely over-fitting or under-fitting, but it does not provide detailed information on how to improve a model structure. Hence, models which fail the goodness-of-fit test are typically rejected. This modelling approach heavily relies on the presence of experienced researchers for proposing a reasonable set of candidate models. Furthermore, it is important to notice that all the candidate models may fail the goodness-of-fit test, i.e. none of the proposed models may be adequate for representing the experimental observations.

The solution space of potentially valid model structures may be large and an exhaustive search may be impossible to perform. Genetic programming was proposed as a mean of exploring effectively vast solution spaces (Banzhaf et al., 2015). Applications of genetic programming to structural equation modelling are also available in the literature (Florin M. et al., 2004; Gandomi and Alavi, 2011; Xiao-lei Yuan et al., 2008). However, genetic approaches rely on the construction and identification of a substantial number of model structures. The estimation of parameters in a high number of kinetic models may be impractical, especially if the models are nonlinear in the parameters (Florin M. et al., 2004; Transtrum et al., 2010) and affected by problems of identifiability, i.e. measured model responses may be poorly sensitive to a change in the value of some parameters and/or parameters may be affected by extreme correlation (López Cárdenas et al., 2015). Due to the difficulty of implementing physical awareness and identifiability constraints in genetic algorithms, the scientist may prefer to devote efforts on improving the predictive performance of an approximated kinetic model structure that is already available.

Statistical and numerical tools have been proposed in the literature as computationally tractable means for improving the performance of misspecified parametric models. When the model is over-fitting, one may employ known regularisation techniques to achieve a better trade-off between the bias and the variance of the model predictions (Barz et al., 2016). Popular regularisation techniques are the Tikhonov regularisation (Bardow, 2008; Hansen, 2005; Johansen, 1997), the truncated singular value decomposition (Hansen, 2005; López Cárdenas et al., 2015) and the parameter subset selection (Barz et al., 2013; López Cárdenas et al., 2015).

When the model is under-fitting, a significant process-model mismatch is observed. Approaches have been proposed in the literature to improve the predictive performance of under-fitting models. Some approaches aim at modelling the observed mismatch as an unknown disturbance (Galvanin et al., 2011, 2012; Hewing et al., 2017). These methods however do not provide a diagnosis for the observed process-model mismatch and the identification of the model components that require revision typically relies solely on human intuition. The importance of diagnosing process-model mismatch is recognised in the literature on process monitoring (Badwe et al., 2009; Wang et al., 2012), but the problem is considered only in the context of linear, black-box models for control applications. An approach for diagnosing process-model mismatch in phenomenological models was proposed by Meneghetti et al. (2014), where a latent variable model is used to detect differences between process and model in the distribution of some auxiliary variables. These auxiliary variables represent

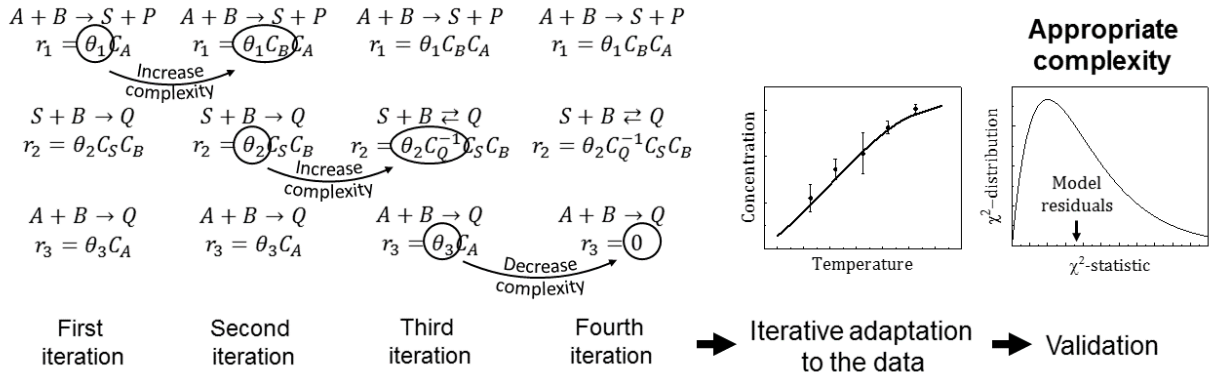


Figure 1: Tuning of complexity in parametric models. Model parameters (θ) that are relevant to fit the data are substituted with more complex expressions and parameters that are irrelevant for model fitting are removed from the model. The model structure is trimmed until an appropriate level of complexity is achieved.

user-defined combinations of state variables and parameters appearing in the model. However, defining opportune auxiliary variables may not be obvious and the diagnosis is sensitive to their definition.

In this manuscript, a model building framework based on maximum likelihood inference is proposed where the structure of an available kinetic model is iteratively refined until an appropriate level of complexity is achieved. A qualitative sketch is given in Figure 1 to illustrate the proposed approach. In the framework, *statistical evidence* provides an index to the scientist to justify changes in the model structure. Whenever over-fitting is detected, model complexity is reduced by removing irrelevant free parameters from the model structure. A Wald test (Wald, 1943) is employed to detect which parameters are unnecessary for fitting the data. If under-fitting is detected, model complexity is increased by substituting relevant free parameters with state-dependent functions. A tailored Lagrange multipliers test (Silvey, 1959) is introduced in this manuscript to support the improvement of approximated model structures towards a higher level of complexity. A model modification index (MMI) is defined as a function of the Lagrange multipliers statistic and is proposed as a heuristic measure of model misspecification. Parameters with the highest associated MMI are those that are expected to improve the model fitting quality the most should they be replaced with state-dependent quantities.

The present manuscript is structured as follows. A brief overview of the statistical tests employed in the proposed framework is given in Section 2. The proposed modelling approach is illustrated in Section 3, where particular emphasis is given to the description of the proposed Lagrange multipliers test. In Section 4, the proposed test is demonstrated on simulated case studies with a baker's yeast growth model and a glucose-insulin interaction model. A discussion on the MMI distribution is given in Section 5.

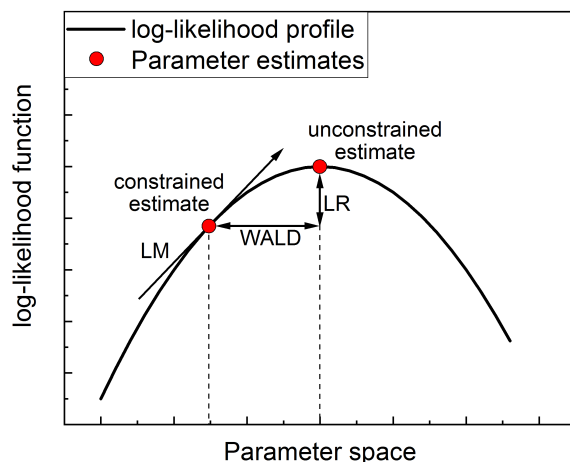


Figure 2: A graphical interpretation of the Likelihood Ratio (LR), Wald and Lagrange Multiplier (LM) statistics for hypothesis testing in model identification frameworks based on maximum likelihood inference.

2 Hypothesis tests in likelihood-based inference

Whenever a kinetic model structure is proposed, the estimation of its kinetic coefficients requires the fitting of experimental data. The construction of increasingly complex parametric models for describing kinetic phenomena has led to a significant reliance on maximum likelihood inference both for parameter estimation and hypothesis testing (Bard, 1974; Walter and Pronzato, 1997). In likelihood-based inference, model parameters are estimated by maximising the likelihood function (or equivalently the log-likelihood function) and the modelling hypotheses are checked using a goodness-of-fit test (Silvey, 1975). A two-tailed goodness-of-fit test based on χ^2 can inform on the appropriateness of the model in representing the data. Nonetheless, whenever over-fitting or under-fitting is detected by a failed goodness-of-fit test, no information is obtained on how the model structure can be improved.

More targeted tests can be performed to support the refinement of parametric models in a likelihood-based modelling framework. More specifically, the likelihood ratio test, the Wald test and the Lagrange multipliers test (Buse, 1982) are regularly applied for structural equation modelling in many applied sciences including psychometrics and econometrics (Anselin, 1988; Chou and Bentler, 1990; Engle, 1984; Green et al., 1999). These tests were proposed to evaluate the generic null hypothesis that model parameters satisfy certain constraints. As an example, one may use the aforementioned tests to challenge the hypothesis that some parameters equal zero. If there is not sufficient evidence from the data for disproving this hypothesis, the considered parameters should be fixed to zero and treated as constants.

The aforementioned tests are asymptotically equivalent and assume the same null hypothesis, but they differ significantly in the construction of their test statistics (Chandra and Joshi, 1983). A graphical interpretation of the statistics for the different tests is given in Figure 2. The Wald statistic is a measure of the distance between the unconstrained and

the constrained maximum likelihood estimates in the parameter space (Wald, 1943). The Lagrange multipliers statistic is a function of the log-likelihood gradient at the constrained estimate (Bera and Biliias, 2001; Rao, 1948; Silvey, 1959). The statistic used in the likelihood ratio test quantifies the distance between the constrained and the unconstrained estimates in terms of log-likelihood values (Wilks, 1938). Depending on the specific case, one statistic may be significantly more convenient to compute than the others (Engle, 1984). In fact, while the likelihood ratio test requires both the constrained and the unconstrained estimates to be computed, the Wald test and Lagrange multipliers test require respectively only the unconstrained and the constrained estimate.

Whenever a proposed kinetic model is falsified for over-fitting, one may regard its parameter estimates as an unconstrained estimate. A Wald test may then inform on which parameters can be constrained to zero and removed from the model structure. When under-fitting is detected, a change in the model structure may be required. An approach for increasing model complexity consists of regarding the proposed model as a constrained instance of one or multiple alternative superstructures (Breusch and Pagan, 1980; Engle, 1982). A Lagrange multipliers test is then performed to challenge the constrained model against the more complex alternatives. A limitation of this approach is that the definition of appropriate superstructures relies entirely on the intuition of the modeller.

In the following section, a model building framework is proposed where the model structure is iteratively trimmed until an appropriate level of complexity is achieved. A tailored Lagrange multipliers test is used in the framework to diagnose model misspecification in conditions of under-fitting. The test does not require the definition of alternative model structures or the definition of superstructures. Instead, the test aims at disproving the null hypothesis that a given model parameter is a state-independent constant. A failed test on a parameter under diagnosis suggests that a significant improvement on the fitting quality is expected if that parameter were replaced a more complex expression of the state variables.

3 Proposed methodology

A setup is assumed to be available to perform experiments on and collect samples from a physical system of interest. A set \mathbf{u} of N_u control variables can be manipulated in the setup and a set \mathbf{y} of N_y physical quantities can be sampled from the setup during the experiments. Samples are affected by multivariate Gaussian measurement noise with covariance Σ_y [$N_y \times N_y$]. It is also assumed that a dataset $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ consisting of N samples of \mathbf{y} is collected for identifying a model for the process under study. Let φ_i denote the set of experimental conditions adopted for the collection of the i -th sample in Y .

A framework for kinetic model building is introduced in Figure 3. The framework begins with the construction of an approximated model structure M . Model M may be either an empirical input-output relationship or a set of phenomenological equations derived from mechanistic insight on the system. Without loss of generality, in the following it is assumed that the scientist proposes a model in the form of a set of differential and algebraic equations

$$M : \begin{cases} \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) = \mathbf{0} \\ \hat{\mathbf{y}} = \mathbf{h}(\mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) \end{cases} \quad (1)$$

where $\hat{\mathbf{y}}$ is an $N_y \times 1$ array of model predictions for the sampled physical quantities \mathbf{y} , \mathbf{x} is an $N_x \times 1$ array of state variables and $\dot{\mathbf{x}}$ is an $N_x \times 1$ array of time derivatives for the state variables. The variable time is denoted as t . \mathbf{f} and \mathbf{h} are respectively an $N_f \times 1$ and an $N_y \times 1$ array of functions. The identification of model M involves the estimation of a set of N_θ parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{N_\theta}]^T$. It is assumed that model M satisfies the requirement for practical identifiability given the available dataset Y (Raue et al., 2009). In different words, the model parameters $\boldsymbol{\theta}$ can be uniquely estimated by fitting the dataset Y .

The framework then involves the following sequential steps:

1. *Parameter estimation.* The model parameters are fitted to the available dataset using a maximum likelihood approach (Bard, 1974).
2. *Goodness-of-fit test.* The adequacy of the model in representing the dataset is assessed with a two-tailed test on the goodness-of-fit (Silvey, 1975). A two-tailed test is employed to detect modeling errors either when model residuals are too small or too large compared with the level of measurement noise present in the system. The test has three possible outcomes:
 - (a) *Passed.* The model is not falsified and its complexity is adequate for representing the dataset. There is no evidence for justifying a change in the model structure.
 - (b) *Failed for over-fitting.* The model is too complex for representing the process. If over-fitting is detected, one shall proceed by performing a Wald test (Wald, 1943) for parameter significance. Unnecessary parameters are removed from the model structure and the procedure is repeated from step 1.
 - (c) *Failed for under-fitting.* The model is not sufficiently complex to capture the underlying mechanisms of the physical system. A tailored Lagrange multipliers test (Silvey, 1959) is proposed in this work to challenge the hypothesis that a given parameter is a state-independent quantity. This Lagrange multipliers statistic is used to compute a heuristic Model Modification Index (MMI), which is proposed as a measure of model misspecification. The MMI quantifies the expected improvement on the model fitting quality should a given parameter be replaced with a state-dependent function. The scientist shall select appropriate functions of the state variables to replace the parameters with the highest MMI. After an appropriate modification of the model structure, the procedure is repeated from step 1.

The illustrated procedure is further detailed in the following subsections assuming that the goodness-of-fit test is failed for under-fitting. Particular emphasis is given to the description of the Lagrange multipliers test, which is proposed to diagnose model descriptive limits and inform the scientist on possible model components that should be considered for revision.

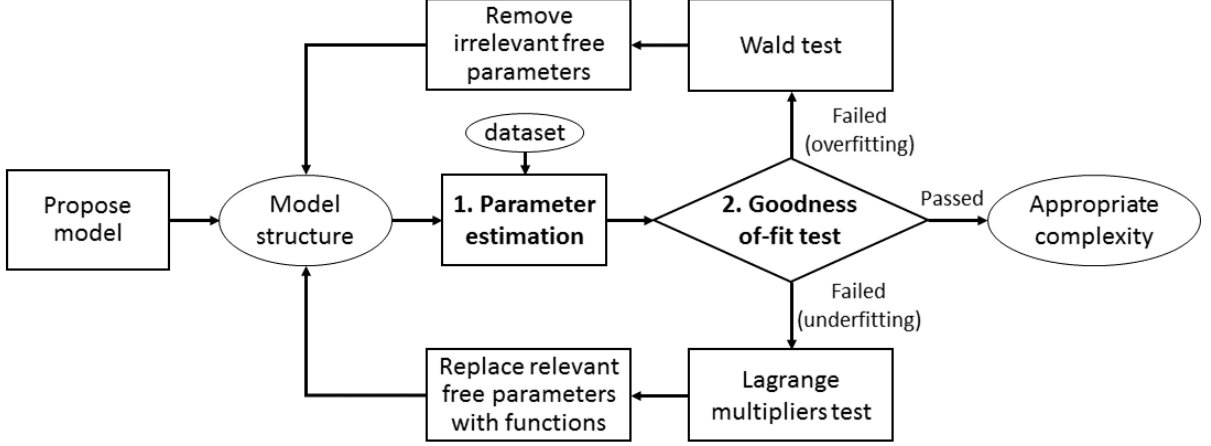


Figure 3: Proposed framework for kinetic model building. In the proposed approach, statistical tests are performed to diagnose model misspecification and to support the scientist in the improvement of misspecified model structures.

3.1 Parameter estimation

Model parameters $\boldsymbol{\theta}$ are estimated with a maximum likelihood approach (Bard, 1974). The log-likelihood function Φ is given by:

$$\begin{aligned} \Phi(\boldsymbol{\theta}|Y) = & -\frac{N}{2}[N_y \ln(2\pi) + \ln(\det(\boldsymbol{\Sigma}_y))] \\ & - \frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta_1, \dots, \theta_{N_\theta})]^T \boldsymbol{\Sigma}_y^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta_1, \dots, \theta_{N_\theta})] \end{aligned} \quad (2)$$

where $\hat{\mathbf{y}}_i$ denotes the model prediction for the i -th sample in the dataset, i.e. \mathbf{y}_i . The maximum likelihood estimate $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \dots, \hat{\theta}_{N_\theta}]^T$ is computed by maximising the unconstrained log-likelihood function

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \Phi(\boldsymbol{\theta}|Y) \quad (3)$$

The maximum likelihood estimate satisfies the unconstrained likelihood equations

$$\nabla \Phi(\hat{\boldsymbol{\theta}}|Y) = \mathbf{0} \quad (4)$$

where the symbol ∇ denotes the gradient operator in the parameter space.

3.2 Goodness-of-fit test

Once the model parameters are fitted to the available dataset, the adequacy of the model is checked with a two-tailed χ^2 test. Under the hypothesis of the proposed model being *exact*, the sum of normalised squared residuals χ_Y^2 is distributed as a χ^2 distribution with degree of freedom $N \cdot N_y - N_\theta$:

$$\chi_Y^2 = \sum_{i=1}^N [\mathbf{y}_i - \hat{\mathbf{y}}_i(\hat{\boldsymbol{\theta}})]^T \boldsymbol{\Sigma}_y^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\hat{\boldsymbol{\theta}})] \sim \chi_{N \cdot N_y - N_\theta}^2 \quad (5)$$

In this work, a two-tailed χ^2 test with 90% of significance is used. If the statistic χ_Y^2 lies between the 5% and the 95% percentiles of the χ^2 distribution, the model is considered as an adequate representation of the physical system. Whenever χ_Y^2 is below the 5% percentile, the model is falsified for over-fitting. If χ_Y^2 is above the 95% percentile, the model is falsified for under-fitting.

3.3 Lagrange multipliers test

When the model is under-fitting, a significant discrepancy between experimental observations and model predictions is observed. It is assumed that the reduction of the discrepancy (and eventually its elimination to the limit of measurement noise) requires the replacement of a certain model parameter with an opportune function of the state variables. A statistical test is proposed to diagnose model misspecification by challenging the hypothesis that a given parameter θ_i is a state-independent constant. The proposed test aims at diagnosing whether it is appropriate to assume a specific model component as a free parameter or whether a significant improvement in the model fitting quality is expected should that parameter be replaced with a function of the state variables. Without loss of generality, the test is detailed assuming that the parameter under diagnosis is the first parameter, i.e. $\theta_i = \theta_1$. The competing hypotheses under test are:

Null hypothesis H_0 : θ_1 and $\theta_j \forall j \neq 1$ are all state-independent constants.

Alternative hypothesis H_a : θ_1 is a state-dependent function and $\theta_j \forall j \neq 1$ are state-independent constants.

The parameter estimation problem is formulated under the assumptions that θ_1 is a function g of the experimental conditions, i.e. $\theta_1 = g(\boldsymbol{\varphi})$, and $\theta_j \forall j \neq 1$ are fixed coefficients. One shall notice that no assumption on the functional form of g is required to perform the test. The $N \times 1$ parameter array $\boldsymbol{\theta}_d$ (subscript d stands for diagnosis) is defined as $\boldsymbol{\theta}_d = [\theta_{1,1}, \dots, \theta_{1,N}]^T$ where the i -th element in the array represents the value of g at the experimental conditions $\boldsymbol{\varphi}_i$, i.e. $\theta_{1,i} = g(\boldsymbol{\varphi}_i) \forall i = 1, \dots, N$. The log-likelihood function Φ_d is constructed under parametrisation $\boldsymbol{\theta}_d$:

$$\begin{aligned} \Phi_d(\boldsymbol{\theta}_d|Y) = & -\frac{N}{2}[N_y \ln(2\pi) + \ln(\det(\boldsymbol{\Sigma}_y))] \\ & - \frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta_{1,i}, \hat{\theta}_2, \dots, \hat{\theta}_{N_\theta})]^T \boldsymbol{\Sigma}_y^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta_{1,i}, \hat{\theta}_2, \dots, \hat{\theta}_{N_\theta})] \end{aligned} \quad (6)$$

In (6), the i -th element in the sum is a function of parameter $\theta_{1,i}$ only. The other model parameters are set equal to their maximum likelihood value and treated as fixed constants in the test, i.e. $\theta_j = \hat{\theta}_j \forall j \neq 1$. The set of $N - 1$ functions \mathbf{s} is defined as

$$\mathbf{s} = [\theta_{1,1} - \theta_{1,2}, \dots, \theta_{1,N-1} - \theta_{1,N}]^T \quad (7)$$

The null and alternative hypotheses are then formalised mathematically as the presence/absence of an $N - 1$ set of constraints for the functions \mathbf{s} as follows

$$\begin{aligned} H_0 : \quad & \mathbf{s} = \mathbf{0} \\ H_a : \quad & \mathbf{s} \neq \mathbf{0} \end{aligned} \quad (8)$$

Notice that the imposition of constraints $\mathbf{s} = \mathbf{0}$ is equivalent to assuming that g is a constant function that is independent from the experimental conditions φ . The constrained maximum likelihood estimate $\hat{\boldsymbol{\theta}}_d = [\hat{\theta}_{1,1}, \dots, \hat{\theta}_{1,N}]^T$ is obtained by maximising the log-likelihood function Φ_d under constraints $\mathbf{s} = \mathbf{0}$.

$$\begin{aligned} \hat{\boldsymbol{\theta}}_d &= \arg \max_{\boldsymbol{\theta}_d} \Phi_d(\boldsymbol{\theta}_d|Y) \\ \text{s.t. } \mathbf{s} &= \mathbf{0} \end{aligned} \quad (9)$$

Under constraints $\mathbf{s} = \mathbf{0}$ all the elements in $\hat{\boldsymbol{\theta}}_d$ are equal to the unconstrained maximum likelihood estimate for parameter θ_1 , i.e. $\hat{\theta}_{1,i} = \hat{\theta}_1 \forall i = 1, \dots, N$. The constrained maximum likelihood estimate $\hat{\boldsymbol{\theta}}_d$ also satisfies the set of constrained maximum likelihood equations

$$\begin{aligned} \nabla \Phi_d(\hat{\boldsymbol{\theta}}_d|Y) + \nabla \mathbf{s} \hat{\boldsymbol{\alpha}} &= \mathbf{0} \\ \mathbf{s} &= \mathbf{0} \end{aligned} \quad (10)$$

where $\hat{\boldsymbol{\alpha}}$ is the $N - 1 \times 1$ array of Lagrange multipliers associated to the constraints. As demonstrated by Aitchison and Silvey (1958) and Silvey (1959), under the null hypothesis being true, the Lagrange multipliers statistic ξ_1 is asymptotically distributed as a χ^2 distribution with degree of freedom equal to the number of constraints (i.e. $N - 1$) as shown in the following equation

$$\xi_1 = \hat{\boldsymbol{\alpha}}^T \nabla \mathbf{s}^T \mathbf{H}_d^{-1} \nabla \mathbf{s} \hat{\boldsymbol{\alpha}} \sim \chi_{N-1}^2 \quad (11)$$

In (11), \mathbf{H}_d represents the $N \times N$ expected Fisher information matrix for the model under parametrisation $\boldsymbol{\theta}_d$, which is approximated by the following expression under null hypothesis conditions

$$\mathbf{H}_d = \sum_{i=1}^N \nabla \hat{\mathbf{y}}_i(\hat{\theta}_{1,i}) \boldsymbol{\Sigma}_y^{-1} \nabla \hat{\mathbf{y}}_i(\hat{\theta}_{1,i})^T \quad (12)$$

Notice that the solution of the constrained maximum likelihood equations (10) is not required to compute the statistic ξ_1 . In fact, ξ_1 may be directly computed as a function of the log-likelihood gradient evaluated setting $\boldsymbol{\theta}_d = \hat{\boldsymbol{\theta}}_d$ as follows

$$\xi_1 = \nabla \Phi_d(\hat{\boldsymbol{\theta}}_d|Y)^T \mathbf{H}_d^{-1} \nabla \Phi_d(\hat{\boldsymbol{\theta}}_d|Y) \sim \chi_{N-1}^2 \quad (13)$$

and does not require the evaluation of the Lagrange multipliers $\hat{\boldsymbol{\alpha}}$ (Rao, 1948). In this work, the Lagrange multipliers statistic is computed according to the expression in (13).

The illustrated approach for constructing the statistic ξ_1 , associated with parameter θ_1 , is repeated for all model parameters obtaining the set of statistics $\xi_i \forall i = 1, \dots, N_\theta$. A heuristic measure of model misspecification, namely a Model Modification Index (MMI), is defined as

$$\text{MMI}_i = \frac{\xi_i}{\chi_{N-1}^2(95\%)} \quad \forall i = 1, \dots, N_\theta \quad (14)$$

The MMI represents a ratio between a Lagrange multipliers statistic and the 95% percentile of the χ^2 distribution with degree of freedom $N - 1$. A MMI larger than 1 indicates that the null hypothesis is falsified by a χ^2 test with 95% of significance. If $\text{MMI}_i > 1$,

one shall expect a significant improvement in the model fitting quality if parameter θ_i were replaced with a state-dependent function. Conversely, if $\text{MMI}_i < 1$ there is no statistical evidence for justifying the substitution of parameter θ_i with a more complex function of the states. The MMI quantifies the expected rate of increase in the log-likelihood function associated with an infinitesimal relaxation of the constraints $\mathbf{s} = \mathbf{0}$. Hence, if the MMI is above 1 for more than one parameter, one shall expect a more significant improvement in the model fitting quality if the parameters with the highest MMI were replaced with functions.

4 Case studies and results

In this section, simulated case studies are proposed to demonstrate the use of the Lagrange multipliers test proposed in Section 3.3 on the diagnosis of model misspecification in under-fitting models. It is shown how the MMI can be used to detect promising parameters that one may consider to replace with functions to improve the model performance. Two case studies are presented: a baker’s yeast growth model (Asprey and Macchietto, 2000) and a more comprehensive case study on a glucose-insulin interaction model (Bergman et al., 1981). The models considered in the two case studies differ in *i*) number of equations N_f *ii*) number of measurable states N_y and *iii*) number of parameters N_θ . Furthermore, in contrast to the baker’s yeast growth model, a discontinuity is present in the glucose-insulin interaction model. The numerical results presented in this case study were obtained with Python 3.5 (Python Core Team, 2018).

4.1 Baker’s yeast growth model

The considered system is a cultivation of yeast in a fed-batch bioreactor. The aim of the scientist is to model the dynamics of yeast concentration $x_1(t)$ [g L⁻¹] and substrate concentration $x_2(t)$ [g L⁻¹] as a function of two system inputs, namely the dilution factor u_1 [h⁻¹] and the substrate concentration in the feed u_2 [g L⁻¹]. The system kinetics are assumed to be described by the following set of differential and algebraic equations:

$$\frac{dx_1}{dt} = (r - u_1 - \theta_4)x_1 \quad (15)$$

$$\frac{dx_2}{dt} = -\frac{rx_1}{\theta_3} + u_1(u_2 - x_2) \quad (16)$$

$$r = \frac{\theta_1 x_2}{\theta_2 x_1 + x_2} \quad (17)$$

where the rate r follows a Cantois-type kinetic law. The system model involves a set of $N_\theta = 4$ parameters $\boldsymbol{\theta}$ whose values are $\boldsymbol{\theta}^* = [0.310, 0.180, 0.550, 0.050]^T$. An array $\mathbf{y} = [x_1, x_2]^T$ of system states can be sampled in the experiments. Measurements of x_1 and x_2 are corrupted by uncorrelated Gaussian measurement noise with standard deviation 0.05 g L⁻¹.

A full factorial experimental design with four dynamic experiments is assumed with two levels for the dilution factor, i.e. $u_1 = \{0.05, 0.20\}$ h⁻¹, and two levels for the substrate concentration in the feed, i.e. $u_2 = \{5.0, 35.0\}$ g L⁻¹. In each experiment, 7 samples of \mathbf{y} are collected at sampling times $t_s = \{3.0, 6.0, 9.0, 12.0, 15.0, 18.0, 21.0\}$ h. The initial conditions

for the differential variables are the same in all the experiments, i.e. $x_1(0) = 1.0 \text{ g L}^{-1}$ and $x_2(0) = 0.01 \text{ g L}^{-1}$. The dataset is generated in-silico by integrating the system model and adding random noise to the responses $x_1(t)$ and $x_2(t) \forall t \in t_s$.

It is assumed that the scientist does not know the functional form of the system model and proposes an approximated model structure. The approximated model includes equations (15) and (16) with a Monod-type kinetic law:

$$r = \frac{\theta_1 x_2}{\theta_2 + x_2} \quad (18)$$

Table 1: Baker’s yeast system. Goodness-of-fit test and model modification index for all model parameters. Results are presented both for the system model and for the approximated model structure.

Model structure	Goodness-of-fit test				MMIs associated to $[\theta_1, \theta_2, \theta_3, \theta_4]$
	$\chi^2(5\%)$	χ_Y^2	$\chi^2(95\%)$	Outcome	
System model	36.44	61.32	69.83	Passed	[0.67, 0.74, 0.77, 0.77]
Approximated model	36.44	2210.37	69.83	Failed for under-fitting	[16.98, 47.08, 11.90, 18.58]

The approximated model and the system model differ in the functional form of the rate expression. More specifically, the element $\theta_2 x_1$ appearing at the denominator in (17) is modelled as a state independent parameter, i.e. θ_2 , in the denominator of the approximated rate law (18).

When the system model is used to fit the dataset, the sum of squared residuals is $\chi_Y^2 = 61.32$, which lies within the acceptable range assumed for the two-tailed goodness-of-fit test with 90% of significance, i.e. $36.44 < \chi_Y^2 < 69.83$. The test suggests that there is no evidence for modifying the model structure. The MMIs associated with the system model are reported in Table 1 and plotted in the radar chart in Figure 4a for visualisation purposes. All the MMIs associated with the parameters in the system model are below 1. Hence, there is no evidence to justify the replacement of any parameter with a function of the state variables.

The parameter set involved in the approximated model is estimated by fitting the dataset. As one can see from Table 1, the approximated model is falsified by the goodness-of-fit test. More specifically, a sum of squared residuals $\chi_Y^2 = 2210.37$ larger than the χ^2 value at 95% of significance highlights the presence of under-fitting. The complexity of the approximated model shall be increased through the substitution of some parameter with an opportune function of the state-variables. The MMIs associated with the model parameters are plotted in Figure 4b. The MMI is larger than 1 for all model parameters. Hence, an improvement in the fitting quality is expected if any of the model parameters were replaced with some state-dependent function. Nonetheless, the highest MMI is associated to θ_2 , i.e. $\text{MMI}_2 = 47.08$, meaning that the most significant improvement in the fitting quality is expected if θ_2 were substituted with an appropriate state-dependent function. The scientist may then focus on choosing an opportune state-dependent function to replace parameter θ_2 in the approximated

model. The approximated model becomes indistinguishable from the system model if θ_2 were replaced with the functional form $\theta_2 x_1$.

The estimation of the model parameters and the computation of the MMIs required only few seconds of CPU time.

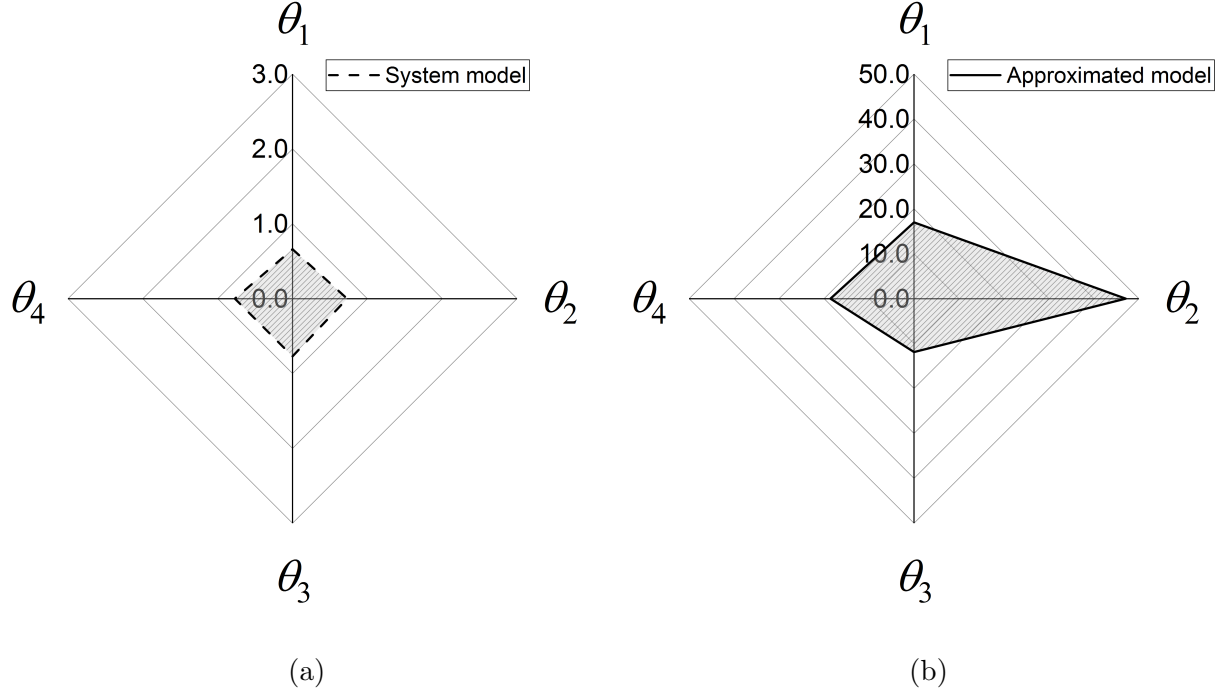


Figure 4: Baker's yeast system. Model Modification Index associated with the model parameters of (a) the system model and (b) the approximated model.

4.2 Glucose-insulin interaction model

In this simulated case study, it is assumed that the aim of the modeller is to accurately describe the dynamic interaction of plasma glucose concentration $G(t)$ [mg dL^{-1}] and plasma insulin concentration $I(t)$ [$\mu\text{U mL}^{-1}$] in a healthy test subject with basal glucose concentration $G_b = 93.0 \text{ mg dL}^{-1}$. The system dynamics are described by the following set of equations (Bergman et al., 1981):

$$\frac{dG}{dt} = -\theta_1(G - G_b) - \theta_2 X G \quad (19)$$

$$\frac{dX}{dt} = -\theta_3 X + I \quad (20)$$

$$\text{IDR} = \begin{cases} \theta_4(G - \theta_5)t & \text{if } G - \theta_5 > 0 \\ 0 & \text{if } G - \theta_5 \leq 0 \end{cases} \quad (21)$$

$$\frac{dI}{dt} = \text{IDR} - \theta_6 I \quad (22)$$

In (19-22), IDR represents the insulin delivery rate as a function of glucose concentration in plasma (Toffolo et al., 1980) and $X(t)$ [$\mu\text{U min mL}^{-1}$] represents the insulin action term associated with the remote insulin receptor (Bergman et al., 1979, 1981; Insel et al., 1975; Zeleznik and Roth, 1978). The system model involves a set $\boldsymbol{\theta}$ of $N_\theta = 6$ parameters. The values of the system parameters associated with the test subject are $\boldsymbol{\theta}^* = [2.96 \cdot 10^{-2}, 6.51 \cdot 10^{-6}, 1.86 \cdot 10^{-2}, 5.36 \cdot 10^{-3}, 9.09 \cdot 10^1, 2.3 \cdot 10^{-1}]$.

It is assumed that G , I and X can be measured from the patient during an intravenous glucose tolerance test (IVGTT). The experimental design for the IVGTT is the same design adopted by Bergman et al. (1981), where 23 samples are collected from the test subject in the course of a 182.0 min assay.

The physiologist proposes an approximated model for the system which involves equations (20-22) and the following differential equation:

$$\frac{dG}{dt} = -\theta_1(G - G_b) - \theta_2 X \quad (23)$$

The system model and the approximated model differ in the functional form of the equation describing glucose concentration in plasma. More specifically, the nonlinear term $-\theta_2 X G$ appearing in the system equation (19) is modelled as a linear term, i.e. $-\theta_2 X$, in the approximated model equation (23).

Table 2: Glucose-Insulin interaction system. Summary of cases considered in the study.

Case ID	IVGTT number	Measured variables	Measurement noise
A	1	G, I	low
B	1	G, I, X	low
C	2	G, I	low
D	1	G, I	high

The model identification approach proposed in Section 3 is applied in different conditions to assess the sensitivity of the MMIs to a change in the experimental design (i.e. different sets of measured state variables and different initial conditions of the test subject) and to a change in the level of measurement noise in the system. The cases are summarised in Table 2 and further described in the following.

Case A. A single IVGTT is performed at initial conditions $G(0) = 298.0 \text{ mg dL}^{-1}$, $I(0) = 333.0 \text{ } \mu\text{U mL}^{-1}$, $X(0) = 0.0 \text{ } \mu\text{U min mL}^{-1}$. The sample includes measurements for G and I , i.e. $\mathbf{y} = [G, I]^T$. A low level of uncorrelated, Gaussian system noise is assumed with standard deviations 1.0 mg dL^{-1} for measurements of G and $1.5 \text{ } \mu\text{U mL}^{-1}$ for measurements of I .

Case B. Same as Case A, but measuring also the insulin action X , i.e. $\mathbf{y} = [G, I, X]^T$. Measurement noise for X is characterised by a standard deviation of $10.0 \text{ } \mu\text{U min mL}^{-1}$.

Case C. Same as Case A, but with an additional IVGTT performed at initial conditions $G(0) = 276.0 \text{ mg dL}^{-1}$, $I(0) = 69.0 \text{ } \mu\text{U mL}^{-1}$, $X(0) = 0.0 \text{ } \mu\text{U min mL}^{-1}$.

Case D. Same as Case A, but assuming high system noise with standard deviations 5.0 mg dL^{-1} for G and $7.5 \text{ } \mu\text{U mL}^{-1}$ for I .

Table 3: Glucose-insulin interaction system. Goodness-of-fit test and model modification index for all model parameters. Results are presented both for the system model and for the approximated model structure in the different considered cases.

Case ID	Model structure	Goodness-of-fit test				MMIs associated to $[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$
		$\chi^2(5\%)$	χ_Y^2	$\chi^2(95\%)$	Outcome	
A	System model	26.51	42.05	55.57	Passed	[0.61, 0.61, 0.61, 0.70, 0.66, 0.80]
	Approximated model	26.51	97.8	55.57	Failed for under-fitting	[2.20, 2.19, 2.20, 1.18, 1.17, 1.40]
B	System model	45.57	64.34	82.57	Passed	[0.69, 0.59, 0.64, 0.80, 0.79, 0.71]
	Approximated model	45.74	128.75	82.52	Failed for under-fitting	[2.17, 2.15, 0.68, 0.85, 0.86, 0.76]
C	System model	65.62	85.99	108.64	Passed	[0.81, 0.80, 0.79, 0.69, 0.69, 0.74]
	Approximated model	65.62	334.71	108.64	Failed for under-fitting	[4.23, 4.27, 4.27, 2.39, 1.91, 2.63]
D	System model	26.51	36.96	55.57	Passed	[0.45, 0.45, 0.46, 0.72, 0.65, 0.64]
	Approximated model	26.51	49.17	55.75	Passed	[0.81, 0.83, 0.84, 0.77, 0.73, 0.91]

For all the illustrated cases, the parameters of both system model and approximated model are estimated and a goodness-of-fit test is performed. Numerical results for the goodness-of-fit test and computed MMIs are reported in Table 3. As one can see from Table 3, in all cases, the system model passes the goodness-of-fit test and its associated MMIs are always below 1, suggesting that there is no evidence to justify a change in the model structure. The approximated model is falsified for under-fitting in Cases A-C. In Case D, the approximated model is not falsified due to an excessive level of system noise. The MMIs associated with the system model are plotted in the radar charts in Figure 5 (dotted lines) together with the MMIs associated with the approximated model (solid lines) for a visual comparison.

The MMIs associated with the approximated model in Case A are plotted in Figure 5a (solid line). As one can see from Figure 5a, all the MMIs associated with the approximated model are higher than 1. The MMIs associated with θ_1 , θ_2 and θ_3 are the largest with a value around 2.20. The analysis suggests that the most significant improvement in the model fitting quality may be achieved by replacing any of these parameters with some appropriate state-dependent quantity. The discrepancy between approximated and system model structures

vanishes if θ_1 were replaced with $\theta_1 + \theta_2 X(G - 1)/(G - G_b)$ or if θ_2 were replaced with $\theta_2 G$. A significant improvement on the fitting quality can also be achieved through the substitution of θ_3 . In fact, by replacing θ_3 with a function, it is possible to modify the behaviour of variable X in order to compensate for the absence of state G in the addend $-\theta_2 X$ in (23).

A change in equation (20) has the potential of improving the fitting quality for variables G and I without causing a degradation in the fitting quality for X . In fact, variable X is not observed in Case A. The observed under-fitting vanishes if parameter θ_3 were replaced with the function

$$\theta_1 + \theta_3 - \frac{GI}{X} + \frac{I}{X} - \frac{\theta_1 G_b}{G} + \frac{\theta_2 X}{G} \quad (24)$$

In Case B, measurements of X are included in the log-likelihood function. The MMIs associated with the approximated model in Case B are plotted in Figure 5b (solid line). In Case B, only the MMI of parameters θ_1 and θ_2 are above 1. The Lagrange multipliers test does not suggest an alteration of parameters $\theta_3 - \theta_6$. Parameters $\theta_3 - \theta_6$ are involved in the correctly specified equations (20) and (22), and their replacement with functions would result in a degradation of the fitting quality for the observed states X and I .

In Case C, the inclusion in the log-likelihood function of an additional IVGTT causes an increase of all MMIs with respect to Case A. The MMIs associated with the approximated kinetic model in Case C are plotted in Figure 5c (solid line). As in Case A, also in Case C the state X is not observed and the Lagrange multipliers test suggests that a major model improvement may be achieved through the substitution of any parameter in the range $\theta_1 - \theta_3$ with some function of the states. A less significant improvement is expected from an alteration of parameters $\theta_4 - \theta_6$. As in Case A, a change in the correctly specified equations (21) and (22) may result in an improvement of the fitting quality for variable G , but it would cause a degradation in the fitting quality for variable I .

In Case D, the approximated model does not fail the goodness-of-fit test. The high system noise in Case D prevents the falsification of the modelling hypothesis and there is no evidence to modify the model structure. The MMIs in Case D, plotted in Figure 5d (solid line) are all below 1, suggesting that no parameter should be altered.

In all the considered cases, the estimation of parameters and the computation of the MMIs required few seconds of CPU time.

5 The MMI distribution

The Lagrange multipliers statistic ξ is asymptotically distributed as a χ^2 distribution under the null hypothesis being true. An analysis is proposed in this section to evaluate the effect of using a finite dataset on the MMIs distribution using both the system model, i.e. when the null hypothesis is true, and the approximated model.

The MMIs associated with the system model were evaluated for 30 different choices of the random seed used to generate the in-silico experimental data. When the system model is adopted to fit the data, the null hypothesis H_0 holds and the MMIs are expected to follow a normalised χ^2 distribution. In this work, the MMI is computed as the ratio between the Lagrange multipliers statistic ξ and $\chi^2(95\%)$ as in (14). Hence, under the null hypothesis

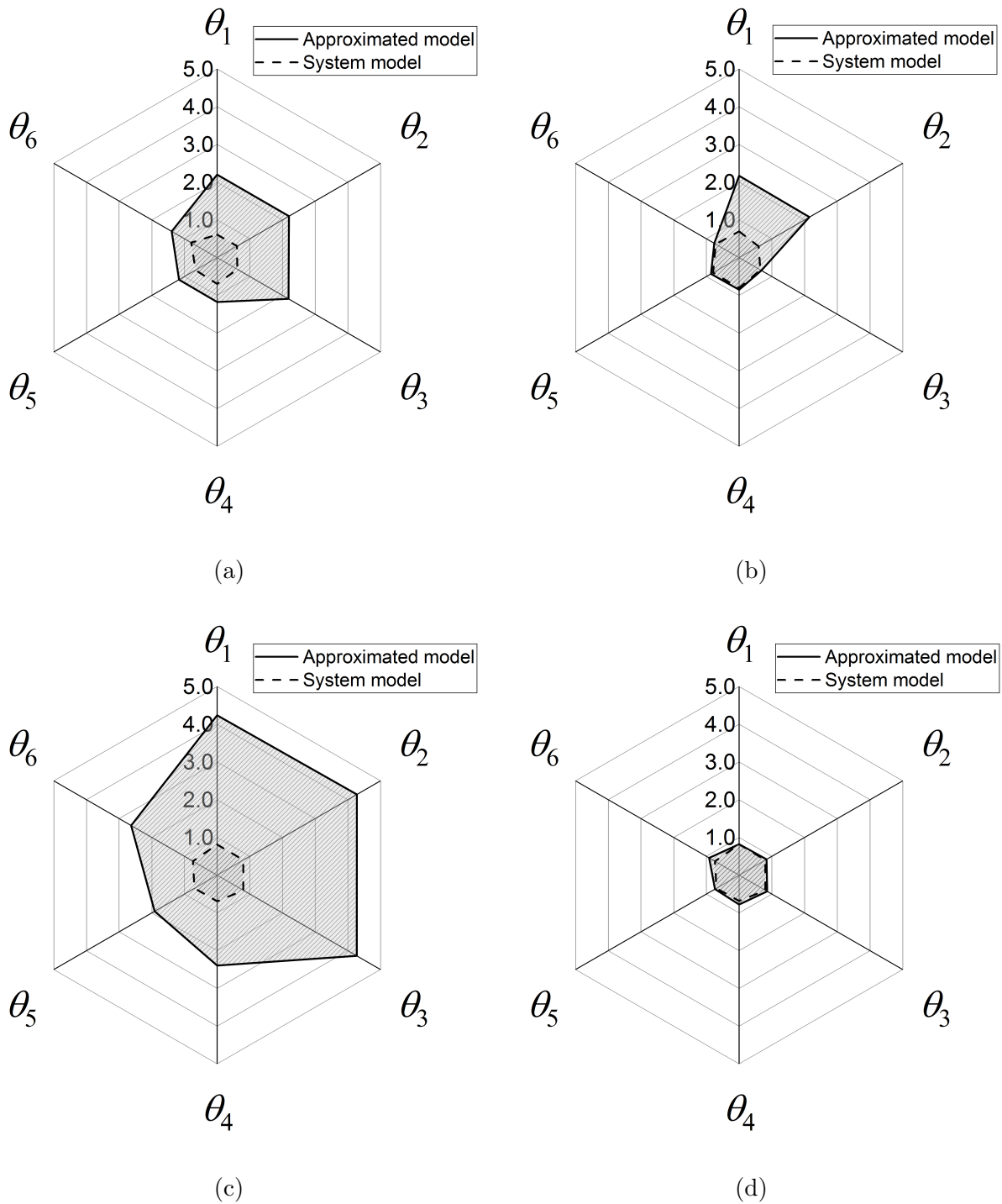


Figure 5: Glucose-Insulin interaction system. Model Modification Indexes for all model parameters: (a) Case A: One performed IVGTT; G, I observed variables; low system noise. (b) Case B: One performed IVGTT; G, I, X observed variables; low system noise. (c) Case C: Two performed IVGTTs; G, I observed variables; low system noise. (d) Case D: One performed IVGTT; G, I observed variables; high system noise. For all Cases, MMIs are plotted for the approximated model (solid line) and for the system model (dotted line).

being true, the MMI is expected to be below 1 for around 95% of the seeds if the number of samples in the dataset is large. The distribution of the MMIs computed with the system model of baker’s yeast is plotted in Figure 6a. The distribution of the MMIs associated with the system model of glucose-insulin interaction (Case A) is given in Figure 6b. In the plots, solid dots highlight instances where the goodness-of-fit test failed to reject the system model. The empty dots refer to instances where the system model is falsified for under-fitting, i.e. empty dots highlight a *Type I* error associated with the goodness-of-fit test (Silvey, 1975).

As one can see from the plots in Figure 6, the MMIs computed with the system models exceed 1 only in a limited number of cases. In the baker’s yeast case, the dataset consisted of $N \cdot N_y = 54$ while in the glucose-insulin case, the dataset included $N \cdot N_y = 46$ measured values. The analysis suggests that with such datasets, the actual MMI distribution is well approximated by the asymptotic χ^2 distribution. Nonetheless, it is recognised that with smaller datasets, a model diagnosis based on the illustrated tests may lead to misleading results. In such cases, one may prefer to use a more conservative reference value in the tests and/or apply opportune finite sample corrections to the considered statistics (Cordeiro and Cribari-Neto, 2014).

A functional form for the distribution of the MMIs under the null hypothesis being false is not available (Silvey, 1959). A further analysis was conducted to assess the sensitivity of the MMIs to the randomness of measurements when the null hypothesis is not satisfied. The MMIs were computed with the approximated models adopting different random seeds. The distribution of the MMIs is plotted in Figure 7a for the approximated model of baker’s yeast and in Figure 7b for the approximated model of glucose-insulin interaction (Case A). In Figure 7, solid dots highlight instances where the goodness-of-fit test failed to reject the approximated model, i.e. solid dots indicate a *Type II* error associated with the goodness-of-fit test (Silvey, 1975).

As one can see from Figure 7a, in the baker’s yeast case the approximated model was falsified for all the choices of the random seed. More specifically, the sum of squared residuals χ_Y^2 in the range 2020.55-2404.20 is significantly larger than the assumed threshold value $\chi^2(95\%) = 69.83$. In all instances, the largest MMI is the one associated with θ_2 . Therefore, the model misspecification diagnosis in the baker’s yeast case is not affected by the randomness in the experimental data.

In the glucose-insulin interaction system, the discrepancy between approximated model and system model in representing the data is not substantial¹. The sum of squared residuals χ_Y^2 is in the range 47.41-101.96 while the 95% reference value is $\chi^2(95\%) = 55.57$. For some choices of the random seed the sum of squared residual is within the range of acceptance considered in the goodness-of-fit test and the approximated model is not falsified. In most instances, the MMIs associated with parameters θ_1 - θ_3 are higher than the MMIs associated with θ_4 - θ_6 (see Figure 7b). Therefore, the misspecification diagnosis is not significantly affected by the random seed. Nonetheless, it is recognised that a low level of system noise is a fundamental requirement for diagnosing model misspecification and increase model complexity. When the system noise increases, it becomes increasingly difficult to falsify an inexact model from the distribution of its normalised residuals, especially if the number of available

¹This is compatible with the observations of Bergman et al. (1979) who demonstrated that both model structures achieve a comparable performance in the fitting of real experimental data.

samples is small. Similarly, in the presence of high system noise, the Lagrange multipliers test tends to favour a constrained model and it becomes increasingly challenging to reject a false null hypothesis with a finite dataset.

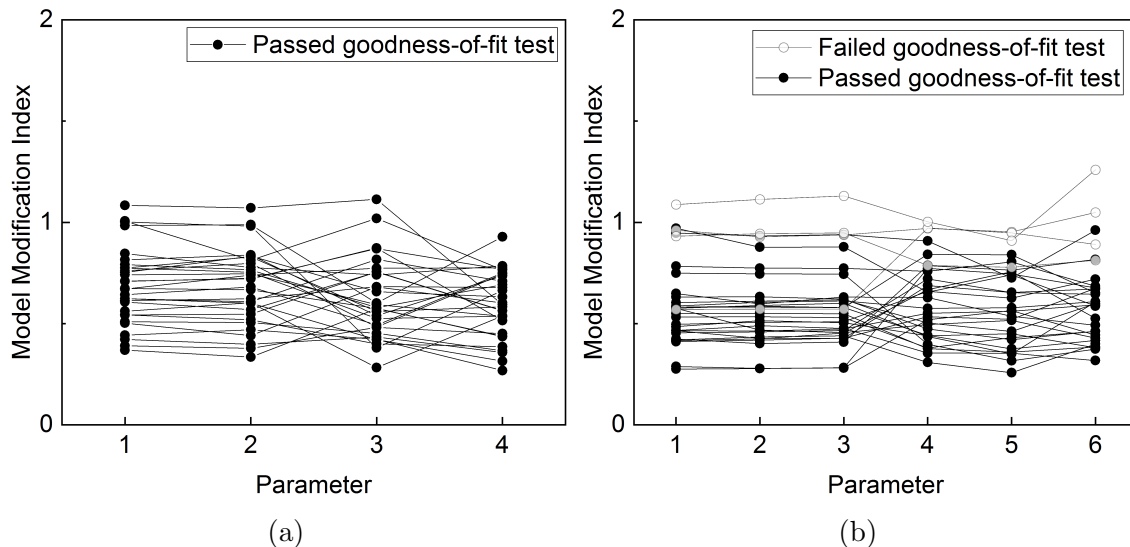


Figure 6: Model Modification Indexes associated with the system model for different choices of the random seed used to generate the data in-silico: (a) Baker’s yeast system; (b) Glucose-Insulin interaction system (Case A). Empty dots indicate indexes computed in cases where the system model failed the goodness-of-fit test.

6 Conclusion

A diagnostic procedure based on maximum likelihood inference is illustrated in this manuscript to support the improvement of approximated kinetic model structures. In the proposed model building framework, modifications in the model structure are justified using experimental evidence. When the model is over-fitting, model parameters that are *irrelevant* for representing the data are removed from the model structure. A Wald test is used to determine which model parameters one shall omit from the model. When the model is in conditions of under-fitting, *relevant* model parameters are replaced with more complex state-dependent functions. A tailored Lagrange multipliers test is proposed in this work to determine which model parameters one shall consider to substitute with more complex functions of the state variables.

The proposed Lagrange multipliers test does not require the definition of alternative model structures or superstructures. In fact, the test aims at disproving the null hypothesis that a given model parameter is a state-independent constant. A model modification index (MMI) is introduced as a function of a Lagrange multipliers statistic. Parameters with the highest MMI are those that are expected to improve the model fitting quality the most if they were replaced with more complex functions of the state variables. When the MMI is below unity there is scarce evidence for justifying an alteration of the parameter. The test is

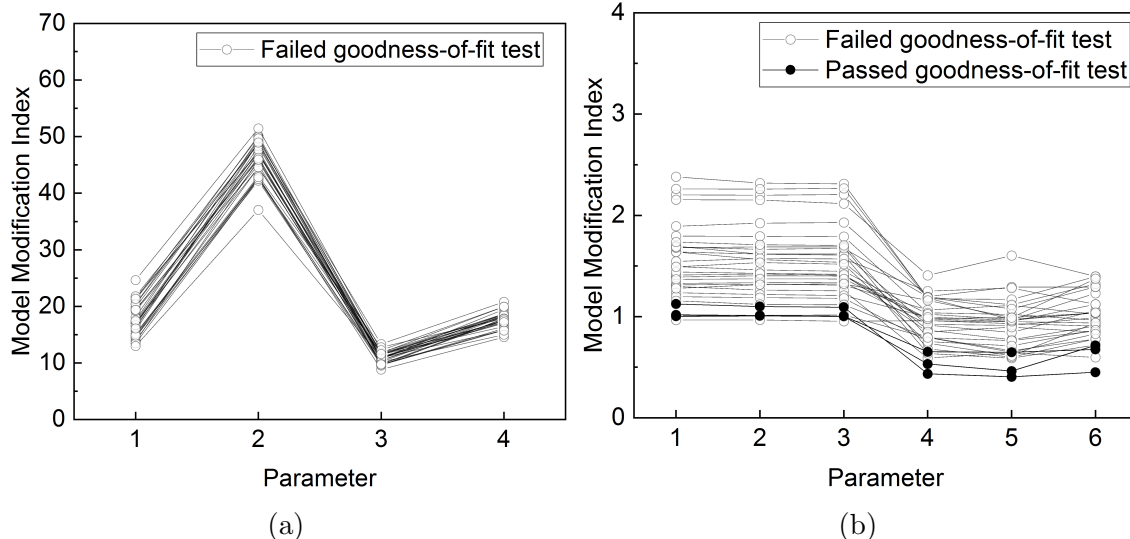


Figure 7: Model Modification Indexes associated with the approximated model for different choices of the random seed used to generate the data in-silico: (a) Baker's yeast system; (b) Glucose-Insulin interaction system (Case A). Solid dots indicate indexes computed in cases where the approximated model passed the goodness-of-fit test.

demonstrated on a number of simulated cases with a baker's yeast growth model and with a model of glucose-insulin interaction. It is shown that, in the presence of moderate system noise, the MMIs correctly highlight the parameters that are primarily associated with model misspecification. When the system noise increases, the falsification of an incorrect modelling hypothesis for under-fitting becomes increasingly challenging and a decrease in the MMIs is observed. When the system noise is excessive, the falsification of an incorrectly specified model structure with a finite dataset may be impractical and the computed MMIs decrease below unity, suggesting that there is no evidence to justify an increase in model complexity.

In future research activities, additional tests will be developed to support the scientist in the selection of appropriate functional forms to replace critical model parameters in under-fitting models. The proposed framework will offer a basis to support the development of experimental design criteria for model diagnosis and model structure improvement. The possibility of implementing the proposed approach into automated or supervised search-based model building algorithms will also be considered. The illustrated modelling framework will also be validated on further case studies with real experimental data.

Acknowledgements

This work was supported by the Hugh Walter Stern PhD Scholarship, University College London.

References

- Aitchison, J., Silvey, S. D., 1958. Maximum-Likelihood Estimation of Parameters Subject to Restraints. *The Annals of Mathematical Statistics* 29 (3), 813–828.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716–723.
- Anselin, L., 1988. Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis* 20 (1), 1–17.
- Asprey, S. P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering* 24 (2), 1261–1267.
- Badwe, A. S., Gudi, R. D., Patwardhan, R. S., Shah, S. L., Patwardhan, S. C., 2009. Detection of model-plant mismatch in MPC applications. *Journal of Process Control* 19 (8), 1305–1313.
- Banzhaf, W., Nordin, P., Keller, R. E., Francone, F. D., 2015. *Genetic Programming: An Introduction*. Morgan Kaufmann.
- Bard, Y., 1974. *Nonlinear Parameter Estimation*. Academic Press, New York.
- Bardow, A., 2008. Optimal experimental design of ill-posed problems: The METER approach. *Computers & Chemical Engineering* 32 (1), 115–124.
- Barz, T., López Cárdenas, D. C., Arellano-Garcia, H., Wozny, G., 2013. Experimental evaluation of an approach to online redesign of experiments for parameter determination. *AIChE Journal* 59 (6), 1981–1995.
- Barz, T., López Cárdenas, D. C., Cruz Bournazou, M. N., Körkel, S., Walter, S. F., 2016. Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. *Computers & Chemical Engineering* 94, 104–116.
- Bera, A. K., Biliyas, Y., 2001. Rao’s score, Neyman’s $C(\alpha)$ and Silvey’s LM tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference* 97 (1), 9–44.
- Bergman, R. N., Ider, Y. Z., Bowden, C. R., Cobelli, C., 1979. Quantitative estimation of insulin sensitivity. *The American Journal of Physiology* 236 (6), E667–677.
- Bergman, R. N., Phillips, L. S., Cobelli, C., 1981. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *The Journal of Clinical Investigation* 68 (6), 1456–1467.
- Breusch, T. S., Pagan, A. R., 1980. The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies* 47 (1), 239–253.

- Buse, A., 1982. The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note. *The American Statistician* 36 (3), 153–157.
- Chandra, T. K., Joshi, S. N., 1983. Comparison of the Likelihood Ratio, Rao’s and Wald’s Tests and a Conjecture of C. R. Rao. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 45 (2), 226–246.
- Chou, C. P., Bentler, P. M., 1990. Model Modification in Covariance Structure Modeling: A Comparison among Likelihood Ratio, Lagrange Multiplier, and Wald Tests. *Multivariate Behavioral Research* 25 (1), 115–136.
- Cordeiro, G. M., Cribari-Neto, F., 2014. Likelihood-Based Inference and Finite-Sample Corrections: A Brief Overview. In: Cordeiro, G. M., Cribari-Neto, F. (Eds.), *An Introduction to Bartlett Correction and Bias Reduction*. SpringerBriefs in Statistics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–11.
- Engle, R. F., 1982. A general approach to lagrange multiplier model diagnostics. *Journal of Econometrics* 20 (1), 83–104.
- Engle, R. F., 1984. Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: *Handbook of Econometrics*. Vol. 2. Elsevier, pp. 775–826.
- Florin M., M., Witzak, M., Korbicz, J., 2004. A novel genetic programming approach to nonlinear system modelling: application to the DAMADICS benchmark problem. *Engineering Applications of Artificial Intelligence* 17 (4), 363–370.
- Galvanin, F., Barolo, M., Pannocchia, G., Bezzo, F., 2011. A disturbance estimation approach for online model-based redesign of experiments in the presence of systematic errors. *Computer Aided Chemical Engineering* 29, 467–471.
- Galvanin, F., Barolo, M., Pannocchia, G., Bezzo, F., 2012. Online model-based redesign of experiments with erratic models: A disturbance estimation approach. *Computers & Chemical Engineering* 42, 138–151.
- Gandomi, A. H., Alavi, A. H., 2011. Multi-stage genetic programming: A new strategy to nonlinear system modeling. *Information Sciences* 181 (23), 5227–5239.
- Green, S. B., Thompson, M. S., Poirier, J., 1999. Exploratory analyses to improve model fit: Errors due to misspecification and a strategy to reduce their occurrence. *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1), 113–126.
- Hansen, P. C., 2005. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM.
- Hewing, L., Kabzan, J., Zeilinger, M. N., 2017. Cautious Model Predictive Control using Gaussian Process Regression. arXiv:1705.10702 [cs, math].
- Insel, P. A., Liljenquist, J. E., Tobin, J. D., Sherwin, R. S., Watkins, P., Andres, R., Berman, M., 1975. Insulin Control of Glucose Metabolism in Man. *Journal of Clinical Investigation* 55 (5), 1057–1066.

- Johansen, T. A., 1997. On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica* 33 (3), 441–446.
- López Cárdenas, D. C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering* 77, 24–42.
- Meneghetti, N., Facco, P., Bezzo, F., Barolo, M., 2014. A Methodology to Diagnose Process/Model Mismatch in First-Principles Models. *Industrial & Engineering Chemistry Research* 53 (36), 14002–14013.
- Neath, A. A., Cavanaugh, J. E., 2012. The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics* 4 (2), 199–203.
- Python Core Team, 2018. Python: A dynamic, open source programming language. <https://www.python.org/>.
- Rao, C. R., 1948. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 44 (1), 50–57.
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25 (15), 1923–1929.
- Silvey, S. D., 1959. The Lagrangian Multiplier Test. *The Annals of Mathematical Statistics* 30 (2), 389–407.
- Silvey, S. D., 1975. *Statistical Inference*. CRC Press.
- Toffolo, G., Bergman, R. N., Finegood, D. T., Bowden, C. R., Cobelli, C., 1980. Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. *Diabetes* 29 (12), 979–990.
- Transtrum, M. K., Machta, B. B., Sethna, J. P., 2010. Why are nonlinear fits so challenging? *Physical Review Letters* 104 (6).
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54 (3), 426–482.
- Walter, E., Pronzato, L., 1997. *Identification of Parametric Models from Experimental Data*. Springer-Verlag, London.
- Wang, H., Xie, L., Song, Z., 2012. A Review for Model Plant Mismatch Measures in Process Monitoring. *Chinese Journal of Chemical Engineering* 20 (6), 1039–1046.
- Wilks, S. S., 1938. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* 9 (1), 60–62.

Xiao-lei Yuan, Yan Bai, Ling Dong, 2008. Identification of linear time-invariant, nonlinear and time varying dynamic systems using genetic programming. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). pp. 56–61.

Zelevnik, A. J., Roth, J., 1978. Demonstration of the insulin receptor in vivo in rabbits and its possible role as a reservoir for the plasma hormone. *The Journal of Clinical Investigation* 61 (5), 1363–1374.