On structural identifiability analysis of cascaded linear time varying systems in dynamic isotope experiments

Weilu Lin*, Mingzhi Huang, Zejian Wang*, Ju Chu, Yingping Zhuang, Siliang Zhang

State Key Laboratory of Bioreactor Engineering
East China University of Science and Technology
130 Meilong Road, Shanghai 200237, P.R.China
Tel. +86-21-64253337; Fax: +86-21-64253702
Corresponding authors: W.Lin, wllin@ecust.edu.cn; Z.Wang, wangzejian@ecust.edu.cn

Abstract

The dynamic 13C labelling experiment, as an emerging experimental technique, can be utilized to investigate the intracellular metabolism under changing physiological environment. A well known *in silico* analysis in metabolic flux analysis is the structural identifiability analysis. It comes from the fact that some enrichment measurement sets cannot uniquely elucidate all intracellular fluxes. To the best of our knowledge, the structural identifiability analysis of dynamic isotope experiments is not available in the literature. In this work, it is shown that if one measurement plan makes the dynamic isotopic fractions balance equations structurally identifiable then for any arbitrary small time interval the plan also makes the equations structurally identifiable. Based on the fact, in order to resolve the local structural identifiablity problem of the dynamic isotopic fractions balance equations approximated with piecewise affine intracellular fluxes, one should check the local structural identifiablity for the corresponding cascaded linear time invariant system at each sampling point with the approach proposed in our earlier work (Lin *et al.*, Math Biosci. 2018; 300:122-129). Two simulated metabolic networks are adopted to demonstrate the utility of the proposed method.
Keywords: Dynamic isotope experiments, Local structural identifiability, Taylor series expansion, Dynamic cumomer balance equations, Linear parameter varying systems

1. Introduction

Isotope experiments[1] have been widely utilized to qualitatively and quantitatively elucidate intracellular fluxes. Although valuable biological insights can be obtained through qualitative isotope experiments[2], the quantitative isotope experiments are more popular in strain engineering[3] and fermentation processes optimization[4,5]. Traditionally, to quantify intracellular fluxes, cell culture has to be cultivated under steady state intracellular metabolic condition. However, in practical, the cell culture is cultivated under changing physiological environments. Not only concentrations of key metabolites such as carbon and nitrogen sources are varying with respect to time, but also pH and temperature are varying as well. This may result in the non-steady state intracellular metabolic condition. The emergence of dynamic isotope experimental technique[6,7] is essential to investigate the cell metabolism under changing physiological environment.

The framework of the dynamic isotope experiment is composed of three steps. Firstly, the dynamic fraction balance equations for a metabolic network that can be simulated for certain tracer configuration and certain intracellular fluxes and pools are proposed. Secondly, the experiment is conducted and the measurements with GC/MS or NMR are obtained at each sampling point. Finally, the experimental fractions are fitted with the proposed network.

In our recent work[8], in order to make the computed isotopic fractions stable at the first step, the stability issue of the dynamic isotope experiment is

investigated. However, at the last step of the experiment, the structural identifiability analysis issue is arose. It comes from the fact that some enrichment measurement sets can not uniquely elucidate all intracellular fluxes. The goal of the structural identifiability analysis of isotope experiments is to conduct *in silico* analysis to verify that if the proposed enrichment measurement sets can *uniquely* elucidate all intracellular fluxes. The flux identifiability analysis issue for steady state isotope experiments has been investigated by some researchers in the last two decades[9,10,11]. The structural identifiability issue for isotopically non-stationary 13C labeling experiments is investigated recently[12]. However, to the best of our knowledge, the investigation of the structural identifiability analysis of dynamic isotope experiments is not available in the literature.

In this work, it is shown that if one measurement plan makes the dynamic isotope experiment of a metabolic network structurally identifiable then for any arbitrary small time interval the plan also makes the network identifiable. The cascaded linear time varying systems in dynamic isotope experiments transfers to the cascaded linear time invariant systems at each sampling point. For the cascaded linear time invariant systems, only one single time point, for example, the initial time point, is necessary to achieve the structural identifiability analysis[12]. Based on the fact, in order to resolve the local structural identifiablity problem of the dynamic balance equations of isotopic fractions approximated with piecewise affine fluxes, one can check the local structural identifiablity with the Taylor series approach[13] for the

corresponding cascaded linear time invariant system at each sampling point.

The rest of the paper is organized as follows. Section 2 introduces the formulation of cascaded linear dynamic models in dynamic isotope experiments. In section 3, the method to conduct local structural identifiability for the cascaded linear time varying systems approximated with the piecewise affine intracellular fluxes is proposed. The method to obtain the optimal measurements for dynamic isotope experiments is introduced in section 4. The structural identifiability analysis for two simulated networks is conducted in section 5. Section 6 concludes the work with discussion.

2. Cascaded linear time varying models in isotope experiments

This section introduces the formulation of cascaded linear dynamic models in dynamic isotope experiments. Method to estimate the piecewise affine approximation of the intracellular fluxes is introduced[6,14] in the section.

Under the metabolic unsteady state condition, the metabolic network balance equations can be established as follows,

$$\frac{dc}{dt} = Nv \tag{1}$$

where N represents the structure of the metabolic network, v is the vector denoting time varying external fluxes and intracellular fluxes, c denotes the time varying concentrations of the intracellular metabolic pools.

Wiechert and his coworkers[15] propose cumomer labeling systems to decompose isotopomer labeling systems into cascaded linear dynamic systems

for isotopically non-stationary 13C labelling experiments. For dynamic isotope experiments, after the transformation, one can also represent the dynamic cumomer balance equations as cascaded linear time varying systems as follows,

$$bdiag([c_1 I_{i,1}, \cdots, c_L I_{i,L}])\frac{dy_i}{dt} = H_i(v)y_i + d_i(y_{i'}, v, tracers) \tag{2}$$

such that

$$i' < i \; for \; i = 1, ..., N_c$$

where y_i is the time varying i-cumomer fractions with the i^{th} weight, H_i is the square coefficient matrix, d_i is the vector that is dependent on tracers and previous cumomer fractions, N_c is the maximum weight of cumomer fractions. c_i is the concentration of the i^{th} pool. $I_{i,l}$ is the identity matrix for the l^{th} metabolic pool for the i-cumomer fractions, L is the total number of intracellular metabolites. $bdiag(.)$ stands for block diagonalization operator. The change of the external fluxes, for example, the rate of the feeding stream, can result in the dynamics of cumomer fractions. The initial values of cumomer fractions, $y_{i,0}$, for $i = 1, \cdots, N_c$, can be determined from the cell culture state before the dynamic isotope experiments. The Eq. 2 for the metabolic network reactions in Fig. 3 are listed in Table 3. From the result, one can immediately observe:

1. The i^{th} diagonal tuples of H_i are the negative values of the summation of the influxes to the i^{th} pool.

2. The summation of the tuples for each row of $[H_i \; d_i]$ is zero.

Since the number of samples is finite, the piecewise affine approximation of the intracellular fluxes is adopted[6,14] to approximately estimate the intracellular fluxes and concentrations of the intracellular pools. Assume that the cell culture is at the metabolic steady-state state at t_0, such that $C(t_0)$ is measured and $v(t_0)$ is determined from the steady state isotope experiment. For sampling time, t_1 to t_K, one can define a first order polynomial function, P_1, for $v(t)$ as follows

$$v(t) = P_1(v(t_{k-1}), b_k, t) \; if \; t \in [t_{k-1} \; t_k] \tag{3}$$

for $k = 1, \cdots, K$. b_k is the slopes of the piecewise affine approximate of the intracellular fluxes for the k^{th} sampling time interval.

From Eq.1, one can obtain

$$\frac{dc}{dt} = N_1 v_d + N_2 v_f + N_3 v_{ext} \tag{4}$$

where v_d is the dependent flux vector, v_f is the free flux vector, v_{ext} is the measured external flux vector. N_1, N_2 and N_3 are composed of corresponding columns of N for v_d, v_f and v_{ext}, respectively. Since the v_{ext} is measured, b_k is partially known such that

$$b_k = \begin{bmatrix} b_{k,v_d} \\ b_{k,v_f} \\ b_{k,v_{ext}} \end{bmatrix} \tag{5}$$

One can define

$$\bar{b}_k = \begin{bmatrix} b_{k,v_d} \\ b_{k,v_f} \end{bmatrix} \tag{6}$$

as unknown parameters for $k = 1, \cdots, K$.

$v(t)$ and $c(t)$ can be defined as

$$v(t) = P_1(v(t_{k-1}), b_k, t) \tag{7}$$

$$c(t) = P_2(v(t_{k-1}), b_k, t, c(t_{k-1})) \tag{8}$$

if $t \in [t_{k-1}\ t_k]$ for $k = 1, \cdots, K$. P_1 and P_2 are the first order polynomial function and the second order polynomial function, respectively.

Eq. 2 can be transformed as follows

$$\frac{dy_i}{dt} = \bar{H}_i(P_1, P_2, t)y_i + \bar{d}_i(y_{i'}, P_1, P_2, tracers, t) \tag{9}$$

for $i = 1, \cdots, N_c$. $\bar{H}_i$ and $\bar{d}_i$ are coefficient matrices. Note that in Eq. 9 the intracellular fluxes are piecewise affine approximated in $t \in [t_{k-1}\ t_k]$ for $k = 1, \cdots, K$.

Assume certain cumomer fractions are measured at $[t_1, \cdots, t_K]$, one can estimate $\bar{b}_1$ to $\bar{b}_K$ with the following objective function

$$\sum_{k=1}^{K} \|y_{i,c}(t_k) - y_{i,m}(t_k)\|^2 \tag{10}$$

where $y_{i,c}(t_k)$ and $y_{i,m}(t_k)$ are calculated and measured cumomer fractions at t_k, respectively, for $k = 1, \cdots, K$ and $i = 1, \cdots, N_c$.

Assume that certain measurements at each sampling time are obtained for the linear time varying compartmental system, the goal of the dynamic isotope experiments is *to uniquely estimate the piecewise dynamic of intracellular fluxes and the concentrations of the intracellular pools in Eq. 9 for cell culture under certain cultivation condition.*

Remark 1. In isotope experiments, the mass distribution vectors (MDVs) are measured instead of cumomer fractions. The MDV is the measured mass spectra for a certain fragment of metabolite molecular with a mass spectrometer which can be represented as certain linear combination of isotopomer fractions. Since there is a one to one relationship between isotopomer fractions and cumomer fractions, the MDVs can also be regarded as a certain linear combination of cumomer fractions. The relationship of MDVs, isotopomer fractions and cumomer fractions is shown in Fig. 1.

Remark 2. One can regard Eq. 9 as linear parameter varying systems[16] which is a special class of nonlinear system. They can be regarded as an extension of linear time invariant systems where the model parameters are time varying.

3. Structural identifiability for dynamic isotope experiments

For certain initial conditions on cumomer fractions, intracellular fluxes, concentrations of intracellular metabolites, measured external fluxes and tracer composition, there exists the unique time trajectories of the measured fractions, for example, MDVs, and those of the concentrations of the

intracellular pools through solving Eq. 1 and Eq. 2[17]. However, as shown in Fig. 2, given certain tracers and measured external fluxes, if one assumes that there is an inverse system to map the trajectories of measured fractions back to the trajectories of intracellular fluxes and those of concentrations of intracellular metabolites, these trajectories could map back to several different sets of trajectories instead of one unique set of trajectories. The main contribution of the work is to propose a way to check the local structural identifiability of the approximated model for dynamic isotope experiments with piecewise affine fluxes.

3.1. Structural identifiability for cascaded linear time varying dynamic systems

The structural identifiablity condition for the dynamic cumomer balance equations can be stated as follows,

Theorem 1. One measurement plan makes the cascaded linear time varying system structurally identifiable if and only if it makes the dynamic system structurally identifiable for any arbitrary small time interval of the experiment.

Proof. The sufficient part of the theorem is straightforward, while the necessary part of the theorem can be proved by contradiction.

Corollary 1. If and only if one measurement plan makes the resulting dynamic system structurally identifiable at every sampling time point of the dynamic isotope experiment then the plan makes the corresponding approximated cascaded linear time varying dynamic systems with piecewise affine

approximated fluxes structurally identifiable.

Proof. Theorem 1 indicates that the cascaded linear time varying dynamic system is structurally identifiable if and only if that the measurement plan makes the system structurally identifiable at arbitrary time point. At an arbitrary time point, the cascaded linear time varying systems turn out to be the cascaded linear time invariant systems. As our earlier work[12] indicates that only one single time point, for example, the initial time point, is necessary to achieve the structural identifiability analysis of the cascaded linear time invariant systems. Therefore, if the measurement plan makes the system structurally identifiable at every sampling point, then based on the fact that there is an one to one relationship between the piecewise affine approximated fluxes and the approximated pool sizes (Eq. 7 and Eq. 8) and there is an unique line represented the piecewise affine flux between two successive sampling points, one can prove the sufficient part of Corollary 1. The necessary part of the corollary can be proved by contradiction.

3.2. Structural identifiability for cascaded linear time varying dynamic systems with piecewise affine approximated fluxes

From Corollary 1, the structural identifiability analysis of the approximated cascaded linear time varying dynamic systems with piecewise affine fluxes can be conducted through the structural identifiability analysis of the corresponding cascaded linear time invariant dynamic systems at every sampling point. In this section, we extend the structural identifiability result for isotopically non-stationary 13C labeling experiments[12] to the cascaded

linear time varying dynamic system for dynamic isotope experiments at one sampling time point during the cultivation process.

[**Lemma 2**] Suppose that there is a vectorial function, $\hat{z}(t,p)$, where p is the parameter vector such that

$$\frac{dz}{dt} = f(z,t,p) \tag{11}$$

$$\hat{z} = g(z,p) \tag{12}$$

For $t = t_0$, if $\bar{\hat{Z}}_\infty(t_0,p) = [\hat{z}^T(t_0,p), \dot{\hat{z}}^T(t_0,p), \cdots, \hat{z}^{(\infty)T}(t_0,p)]^T$ is continuously differentiable with respect to p, then $\hat{z}(t,p)$ is locally structurally identifiable at p^* if $\frac{\partial \bar{\hat{Z}}_\infty(t_0,p)}{\partial p}|_{p^*}$ has full column rank.

Proof: Refer to Proposition 1 in Pohjanpalo's work[13].

The implication of Lemma 2 is that if an upper bound of the time derivatives exists for the $\bar{\hat{Z}}_\infty(t_0,p)$, the local structural identifiability of Eq. 11 and Eq. 12 can be determined at a single time point, t_0.

[**Theorem 2**] For dynamic cumomer balance equations of dynamic isotope experiments, Eq. 2, can be transformed to the equations as follows

$$\frac{dy_i}{dt} = \tilde{H}_i(v)y_i + \tilde{d}_i(y_{i'}, v, tracers) \tag{13}$$

such that

$$i' < i \; for \; i = 1, ..., N_c$$

Then, at each sampling time point (t_j), $k_{i,max} = n_{y_i}$ such that for any $k > k_{i,max}$, $y_i^{(k)}(t_j,v,c)$ is a linear combination of $y_i(t_j,v,c), \dot{y}_i(t_j,v,c), \cdots, y_i^{(k_{max})}(t_j,v,c)$

where n_{y_i} is the dimension of y_i.

Proof: At one sampling time point, $t = t_j$, one can obtain the extended derivatives of the measured fractions as follows,

$$\begin{bmatrix} y_i \\ \dot{y}_i \\ \vdots \\ y_i^{(k)} \end{bmatrix}_{t=t_j} = \begin{bmatrix} I \\ \tilde{H}_i(t_j) \\ \vdots \\ \tilde{H}_i^k(t_j) \end{bmatrix} y_{i,t=t_j} + \begin{bmatrix} 0 \\ I \\ \vdots \\ \tilde{H}_i^{(k-1)}(t_j) \end{bmatrix} \tilde{d}_{i,t=t_j} \tag{14}$$

From the Cayley-Hamilton theorem[17], at each t_j, one has

$$\tilde{H}_i^{n_{y_i}}(t_j) + \beta_{(n_{y_i}-1)}\tilde{H}_i^{(n_{y_i}-1)}(t_j) + \cdots + \beta_0 I = 0$$

such that

$$det(\tilde{H}_i(t_j) - sI) = s^{n_{y_i}} + \beta_{(n_{y_i}-1)}s^{n_{y_i}-1} + \cdots + \beta_0.$$

Therefore, if $k > n_{y_i}$, $y_i^{(k)}(t_j)$ is the linear combination of $y_i(t_j)$, $\cdots$, $y_i^{(n_{y_i})}(t_j)$. Since t_j can be an arbitrary time point, one proves the Theorem 2.

In isotope experiments, the fraction measurements, for example, MDVs, can be represented as the linear combination of cumomer fractions. Suppose for an experimental setup, for each sampling point, there are totally N_m such measurements. One can check that if, given measured fractions, the metabolic network is locally structurally identifiable with the following theorem.

[**Theorem 3**] Suppose that there are N_m weighted measurements for each

sampling time, $j = 1, \cdots, J$, which can be represented as the following

$$\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_{N_m} \end{bmatrix}_{t_j} = \begin{bmatrix} \tilde{D}_{11} \cdots \tilde{D}_{1N_c} \\ \vdots \ddots \vdots \\ \tilde{D}_{N_m1} \cdots \tilde{D}_{N_mN_c} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{N_c} \end{bmatrix}_{t_j} \tag{15}$$

. For, $j = 1, \cdots, J$, the above equation can be presented as the following

$$\hat{\mathscr{Y}}_{t_j} = \tilde{\mathscr{D}} \mathscr{Y}_{t_j} \tag{16}$$

such that $\hat{\mathscr{Y}}_{t_j} = [(\hat{y}_1)^T, ..., (\hat{y}_{N_m})^T]^T_{t_j}$ and $\mathscr{Y} = [(y_1)^T, ..., (y_{N_c})^T]^T_{t_j}$. Furthermore, if one defines

$$\bar{\hat{Y}}(t_j, v, c) = \begin{bmatrix} \hat{\mathscr{Y}} \\ \hat{\mathscr{Y}}^{(1)} \\ \vdots \\ \hat{\mathscr{Y}}^{(k_{max})} \end{bmatrix}_{t_j} \tag{17}$$

such that $k_{max} = max(n_{y_1}, \cdots, n_{y_{N_c}})$. The dynamic cumomer balance equation for dynamic isotope experiments, Eq. 2, is locally structurally identifiable at $p^*(t_j)$, $p(t_j) = \bar{b}_j$, if the $\frac{\partial \bar{\hat{Y}}(t_j,p)}{\partial p}|_{p^*}$ is full column rank.

Proof: From Theorem 2, for $k > k_{max}$, $\hat{\mathscr{Y}}^{(k)}(t_j)$ is the linear combination of $\bar{\hat{Y}}(t_j, p)$. Since $\bar{\hat{Y}}(t_j, p)$ is continuously differentiable with respect to p, through Lemma 2, one can conclude that Eq. 2 is locally structurally identifiable at $t = t_j$ if the $\frac{\partial \bar{\hat{Y}}(t_j,p)}{\partial p}|_{p^*}$ has full column rank. The algorithm to calculate $\frac{\partial \bar{\hat{Y}}(t_j,p)}{\partial p}$ is introduced in Appendix A.

Remark 1: The structural identifiability condition is not directly related

to the local optimality condition of the nonlinear least squares problem with the objective function as defined in Eq. 10. In this work, a necessary and sufficient condition similar to Theorem 3 in[12] can be proposed as well for the cascaded linear time varying dynamic systems to satisfy the local optimality condition of the nonlinear least squares problem.

Remark 2: Although in this work the dynamic cumomer balance equations are utilized for the proposed structural identification analysis, the approach can be readily extended to other balance equations[18,19,20] that can be formulated in cascaded linear dynamic equations.

4. Method to obtain the optimal measurement sets

The numerical rank of the Jacobian matrices is difficult to determine due to the numerical overflow of double data type. Isermann *et al.*[9] adopt integer arithmetic to determine the rank of them. The algebraic calculation of the ranks of these matrices requires symbolic computation which is computationally very expensive for large matrices. Furthermore, when the dimension of the system increases the coefficients of the Jacobian matrices can result in overflow. In our recent work[12], multiprecision computing toolbox for *Matlab*[21] is adopted to avoid overflow. The approach can be introduced as follows,

I. If the smallest singular values of the Jacobian matrices are zeros, the rational Jacobian matrices are either algebraic rank deficient or that the free parameters are the zeros of the matrices.

II. The very small smallest singular values are due to either numerical errors introduced by round off errors or the free parameters are close to the zeros of the rational Jacobian matrices. However, the algebraic ranks of the Jacobian matrices should be full. In this work, if the smallest singular value is less than the predefined threshold, the Jacobian matrix is regarded as not full column rank. The threshold can be defined as the product of maximum size of the Jacobian and $eps(\sigma_{max})$ as in the *rank* function of $Matlab$ where σ_{max} is the largest singular value of the Jacobian.

In the section, the optimal measurement sets issue for cascaded linear dynamic systems is elucidated. One can state the problem as follows,

Given the cascaded linear time varying dynamic systems for a metabolic network (Eq. 2) and the measured fractions (Eq. 15), what are the optimal measured fractions needed to uniquely identify the trajectories of intracellular fluxes and the concentrations of metabolite pools of the metabolic network around the neighbourhood of the nominal values of them?

The measured fractions, Eq. 15, is error free. The optimal measurements mean the minimal number of measurements. The optimal measurements sets can then be determined as follows:

Step I. Assuming that measured fractions are obtained through Eq. 15, calculate the Jacobian of the matrix as shown in Eq. 17 with respect to the nominal free parameters, $v^*(t_j)$ and $c^*(t_j)$, at each sampling point for $j = 1, \cdots, J$.

Step II. Assume that only one measured fractions is needed. Gener-

ate all possible measurement sets and the corresponding Jacobian matrices, $J_1^*(t_j), ..., J_{N_m}^*(t_j)$ for $j = 1, \cdots, J$.

Step III. Assume that all measured fractions are needed. Construct the following matrix

$$\hat{J}_{N_m}^*(t_j) = \begin{bmatrix} J_1^* \\ \vdots \\ J_{N_m}^* \end{bmatrix}_{t_j} \tag{18}$$

for $j = 1, \cdots, J$. If every $\hat{J}_{N_m}^*(t_j)$ has full column rank, go to step IV. Otherwise stop the algorithm.

Step IV. Start from the first combination of the measured fractions with $k = 1$, construct the following matrix

$$\hat{J}_k^*(t_j) = \begin{bmatrix} J_{n_1}^* \\ \vdots \\ J_{n_k}^* \end{bmatrix}_{t_j} \tag{19}$$

where $\hat{J}_k^*(t_j)$ is the Jacobian matrix of the selected measurement with maximum k measurements at the j^{th} sampling time. $J_{n_1}^*(t_j), \cdots, J_{n_k}^*(t_j)$ are the corresponding Jacobian matrices in the current combination of the measured fractions at the j^{th} sampling time. If every $\hat{J}_k^*(t_j)$ has full column rank, record the measurements as one of optimal measurements sets. Otherwise, go to the beginning of step IV to calculate the next $\hat{J}_k^*(t_j)$ for $j = 1, \cdots, J$. The identifiability analysis stops when all combinations with maximum k

measurements have been tested.

Step V. If any optimal measurements set has been found, then stop. Otherwise, set $k = k + 1$ and go back to step IV.

5. Simulation case studies

5.1. A simple simulation case

A simple simulation case[19] is adopted in this section to demonstrate the utility of the proposed algorithms. The metabolic network as shown in Fig. 3 has three intracellular metabolite pools with six fluxes. v_1 is the external flux with assigned value. The atom transition information is listed in Table 1. The dynamic metabolic network balance equations and the dynamic cumomer balance equations are listed in Table 2 and Table 3, respectively. To simplify the calculation, we assume that $N_m = N_c$ and $\hat{y}_i = y_i$ for $i = 1, \cdots, N_c$ in Eq. 15 in this section. The metabolic balance equation of the intercellular pools is

$$\frac{dc}{dt} = Nv$$

where

$$N = \begin{bmatrix} 1 & -1 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix}$$

At initial time, $t_0 = 0$, cell culture is at steady state, v_1 to v_6, are equal to 100, 110, 50, 20, 20, and 80 $mmol/10^{10}cell/hr$, respectively. v_1 and v_6 are external fluxes. The concentration of the metabolites B, C and D are 0.012, 0.0071 and 0.0160 $mmol/10^{10}cell$, respectively. The molar fraction of

the A molecular labeled with ^{13}C at the second carbon position is 0.5 while the rest of the A molecular is not labeled. The initial conditions for Eq. 9 can be obtained through solving the corresponding steady state cumomer balance equations. The cell culture is at steady state at the initial condition, $t_0 = 0$,, therefore, with D as the optimal measurement set, the system is structurally identifiable with the chosen parameters[9].

The corresponding approximated dynamic cascaded linear time varying system with piecewise affine approximated fluxes is simulated with the b_k (Table 4) for $k = 1, \cdots, 4$ for sampling time at $t_1 = 4$, $t_2 = 6$, $t_3 = 8$, and $t_4 = 10$ hours, respectively. The resulting $c(t)$ and $v(t)$ are shown in Fig. 4 which indicates the system is under intracellular unsteady metabolic state. The equations to describe the sizes of pools and the cumomer fractions at $\triangle\ t \in (0\ \ \triangle\ t_k]$ during the k^{th} sampling interval such that $\triangle\ t_k = t_k - t_{k-1}$ are shown in the supplementary materials.

For this case, the external fluxes, v_1 and v_6 are measured. Therefore, $\bar{b}_k = [b_{k,v_2}\ b_{k,v_3}\ b_{k,v_4}\ b_{k,v_5}]^T$. The results of the local structural identifiability analysis at nominal free parameters for $t_1 = 4$, $t_2 = 6$, $t_3 = 8$, and $t_4 = 10$ hours are shown in Table 5 and Table 6. As the results show, the optimal measurement sets are B, C and D, respectively. It means that if all cumomer fractions of any one of B, C and D are measured the free parameters of the piecewise affine approximated fluxes for every sampling interval can be uniquely estimated.

5.2. A simulated TCA cycle case

The simulated TCA cycle as shown in Fig. 5 is chosen as the case study in this section. There are six metabolite pools in the network. With totally 14 fluxes, there are 9 intercellular fluxes and 5 external fluxes. The atom transition information is listed in Table 7. The dynamic metabolic balance equations and the dynamic cumomer balance equations of the network are listed in the supplementary materials. To simplify the calculation, we assume that $N_m = N_c$ and $\hat{y}_i = y_i$ for $i = 1, \cdots, N_c$ in Eq. 15 in this section. The metabolic balance equation of the intercellular pools is

$$\frac{dc}{dt} = Nv$$

where

$$N = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & -1 \end{bmatrix}$$

At initial time, $t_0 = 0$, cell culture is at steady state, v_1 to v_{14}, are equal to 101, 50, 10.4, 10.3, 50, 60.2, 20.0, 40, 20.1, 50, 90.5, 50, 40, and 50 $mmol/10^{10}cell/hr$, respectively. v_5, v_8 and v_{10} are free fluxes. The concentrations of the metabolites $PYR, ACCOA, CIT, AKG, SUC, OAA$ and $CO2$ are 0.0156, 0.0031, 0.0248, 0.0095, 0.0037, 0.0164 and 0.0135 $mmol/10^{10}cell$, respectively. At the initial time ($t = 0$), the external influx to α-ketoglutaric acid pool is composed of tracers labeled with ^{13}C at the second carbon posi-

tion with molar fraction of 0.5, while the rest of the flux is composed of tracer with ^{13}C at all carbon positions. The influx to pyruvate pool is composed of tracers labeled with ^{13}C at all carbon positions with molar fraction of 0.5, while the rest of the flux is composed of pyruvate with ^{13}C at the first carbon positions. The initial conditions for Eq. 9 can be obtained through solving the corresponding steady state cumomer balance equations. The steady state structural identifiability analysis shows that at the initial condition, $t_0 = 0$, with AKG or SUC or OAA as the optimal measurement set, the system is structurally identifiable with the chosen parameters[9].

The corresponding approximated dynamic cascaded linear time varying system with piecewise affine approximated fluxes is simulated with the b_k as shown in the supplementary materials for $k = 1, \cdots, 4$ for sampling time at $t_1 = 4$, $t_2 = 6$, $t_3 = 8$, and $t_4 = 10$ hours, respectively. The resulting $c(t)$ and $v(t)$ are shown in Fig. 6 which indicates the system is under intracellular unsteady metabolic state.

For this case, the external fluxes, v_1, v_2, and v_{12} to v_{14} are measured. Therefore, $\bar{b}_k = [b_{k,v_3}\ b_{k,v_4}\ \cdots\ b_{k,v_{10}}\ b_{k,v_{11}}]^T$. The results of the local structural identifiability analysis at nominal free parameters for $t_1 = 4$, $t_2 = 6$, $t_3 = 8$, and $t_4 = 10$ hours are shown in Table 8 to Table 11. As the results show, the optimal measurement sets are PYR, CIT and SUC or CIT, SUC and OAA. It means that if all cumomer fractions of PYR, CIT and SUC or CIT or SUC or OAA are measured the free parameters of the piecewise affine approximated fluxes for every sampling interval can be uniquely estimated.

6. Conclusions

The dynamic isotope experiments make the traditional isotope experiments applicable to quantify intracellular fluxes of the cell culture under the changing physiological environment, for example, fed batch culture. The piecewise affine approximated fluxes for every sampling interval are adopted to approximate the corresponding cascaded linear time varying system.

A well known issue in isotope experiments is structural identifiability analysis. It comes from the fact that some enrichment measurement sets can not uniquely elucidate all intracellular fluxes. The goal of the structural identifiability analysis of isotope experiments is to conduct *in silico* analysis to verify that if the proposed enrichment measurement sets can *uniquely* elucidate all intracellular fluxes. However, to the best of our knowledge, the structural identifiability analysis of dynamic isotope experiments is not available in the literature.

It is shown that in order to make the resulting linear parameter varying system[16] locally structurally identifiable, one should check the local structural identifiability of the cascaded linear time varying system at every sampling time. At a specific sampling time, the cascaded linear time varying system is transferred to the cascaded linear time invariant system. As our earlier work[12] indicates that only one single time point, for example, the initial time point, is necessary to achieve the structural identifiability analysis of the cascaded linear time invariant systems. Therefore, the Taylor series approach[13] can be utilized to conduct the local structural identifiability of

the resulting system at each sample time. The method to obtain the optimal measurements for dynamic isotope experiments is also introduced in the work. The structural identifiability analysis for two simulated networks is conducted to demonstrate the utility of the proposed method.

Our in silico analysis experience shows that the variation of physiological conditions, for example, both the variation of the intracellular fluxes and/or the configuration of tracers affect the result of the structural identifiability analysis. Furthermore, as our earlier work[12] shown, the confidence interval approach[22] can not be a sufficient condition to justify the local optimum of the nonlinear LS problem in the isotopically non-stationary 13C labelling experiments. Therefore, we argue that the local structural identifiability analysis is essential for the design of experiments step[23] in the dynamic isotope experiments. For example, one should check the structural identifiability before conducting the statistical analysis when chooses the optimal tracer plan.

Acknowledgement

Authors thank financial support by National Key Research and Development Program of China (2019YFA0904300) and Open Funding Project of the State Key Laboratory of Bioreactor Engineering (F200-B-01G).

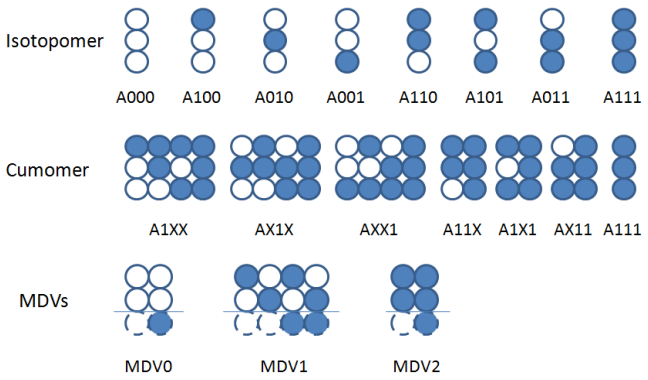

Figure 1: The relationship between isotopomer and cumomer fractions for molecular A with three carbon atoms and MDV fractions for the fragment of molecular A with two carbon atoms.

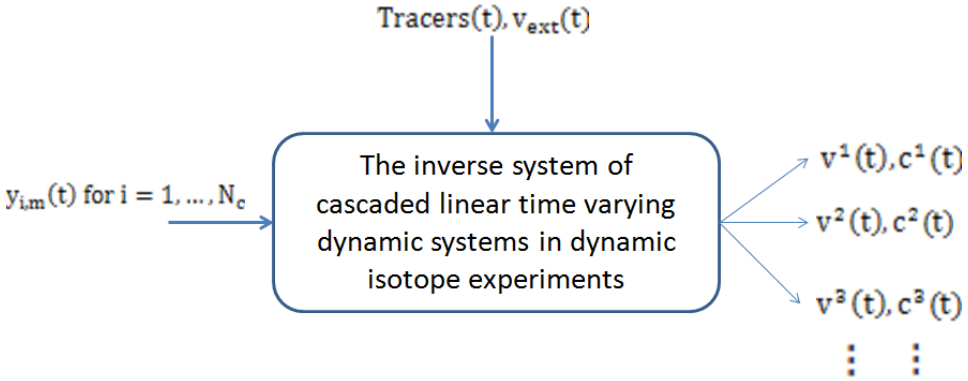

Figure 2: The structural identifiability issue in dynamic isotope experiments.

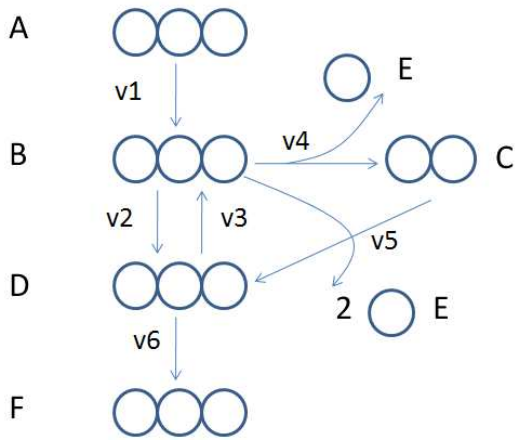

Figure 3: A simple metabolic network.

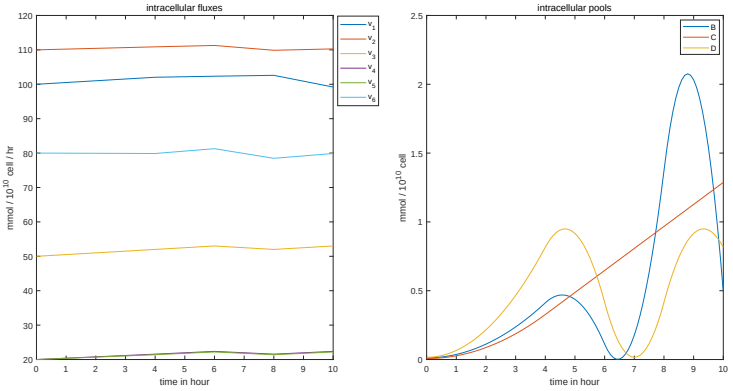

Figure 4: The simulated intracellular fluxes and sizes of pools for the example as shown in Fig. 3.

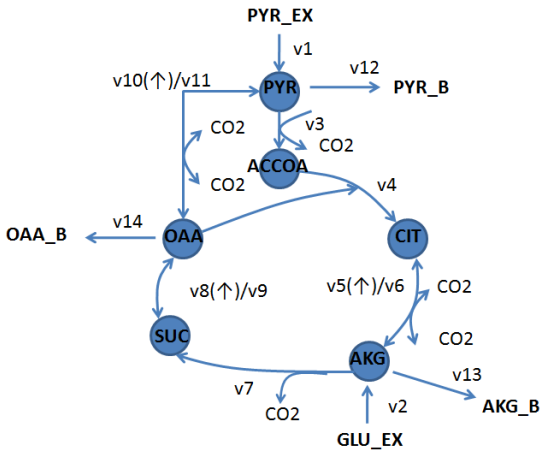

Figure 5: A simulated TCA cycle example.

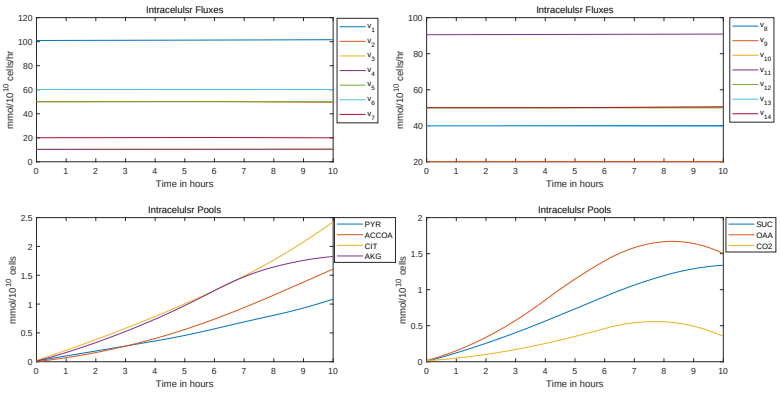

Figure 6: The simulated intracellular fluxes and sizes of pools for the example as shown in Fig. 5.

Table 1: Stoichiometry and atom transitions for the metabolic network reactions in Fig. 3.

	Reactions	Atom Transformations
1	$A \rightarrow B$	$abc \rightarrow abc$
2	$B \rightarrow D$	$abc \rightarrow abc$
3	$D \rightarrow B$	$abc \rightarrow abc$
4	$B \rightarrow C + E$	$abc \rightarrow bc + a$
5	$B + C \rightarrow D + E + E$	$abc + de \rightarrow bcd + a + e$
6	$D \rightarrow F$	$abc \rightarrow abc$

Table 2: The dynamic metabolic network balance equations for the metabolic network reactions in Fig. 3.

	Pools	Dynamic metabolic network balance equations
1	B	$\frac{dc_B}{dt} = (v_1 + v_3) - (v_2 + v_4 + v_5)$
2	C	$\frac{dc_C}{dt} = v_4 - v_5$
3	D	$\frac{dc_D}{dt} = (v_2 + v_5) - (v_3 + v_6)$

Table 3: The dynamic cumomer balance equations for the metabolic network reactions in Fig. 3.

	Cumomer Fractions	Dynamic Cumomer Balance Equations
1	B_{1xx}	$c_B \frac{dB_{1xx}}{dt} = -(v_1+v_3)B1xx + v_1A1xx + v_3D1xx$
2	B_{x1x}	$c_B \frac{dB_{x1x}}{dt} = -(v_1+v_3)Bx1x + v_1Ax1x + v_3Dx1x$
3	B_{xx1}	$c_B \frac{dB_{xx1}}{dt} = -(v_1+v_3)Bxx1 + v_1Axx1 + v_3Dxx1$
4	C_{1x}	$c_C \frac{dC_{1x}}{dt} = -v_4C1x + v_4Bx1x$
5	C_{x1}	$c_C \frac{dC_{x1}}{dt} = -v_4Cx1 + v_4Bxx1$
6	D_{1xx}	$c_D \frac{dD_{1xx}}{dt} = -(v_2+v_5)D1xx + v_2B1xx + v_5Bx1x$
7	D_{x1x}	$c_D \frac{dD_{x1x}}{dt} = -(v_2+v_5)Dx1x + v_2Bx1x + v_5Bxx1$
8	D_{xx1}	$c_D \frac{dD_{xx1}}{dt} = -(v_2+v_5)Dxx1 + v_2Bxx1 + v_5C1x$
9	B_{11x}	$c_B \frac{dB_{11x}}{dt} = -(v_1+v_3)B11x + v_1A11x + v_3D11x$
10	B_{1x1}	$c_B \frac{dB_{1x1}}{dt} = -(v_1+v_3)B1x1 + v_1A1x1 + v_3D1x1$
11	B_{x11}	$c_B \frac{dB_{x11}}{dt} = -(v_1+v_3)Bx11 + v_1Ax11 + v_3Dx11$
12	C_{11}	$c_C \frac{dC_{11}}{dt} = -v_4C11 + v_4Bx11$
13	D_{11x}	$c_D \frac{dD_{11x}}{dt} = -(v_2+v_5)D11x + v_2B11x + v_5Bx11$
14	D_{1x1}	$c_D \frac{dD_{1x1}}{dt} = -(v_2+v_5)D1x1 + v_2B1x1 + v_5Bx1xC1x$
15	D_{x11}	$c_D \frac{dD_{x11}}{dt} = -(v_2+v_5)Dx11 + v_2Bx11 + v_5Bxx1C1x$
16	B_{111}	$c_B \frac{dB_{111}}{dt} = -(v_1+v_3)B111 + v_1A111 + v_3D111$
17	D_{111}	$c_D \frac{dD_{111}}{dt} = -(v_2+v_5)D111 + v_2B111 + v_5Bx11C1x$

Table 4: The slopes ($mmol/10^{10}cell/hr^2$) of the piecewise affine approximate of the fluxes for each sampling time interval for example shown in Fig. 3.

	v_1	v_2	v_3	v_4	v_5	v_6
b_1	0.51	0.22	0.5	0.39	0.35	-0.03
b_2	0.15	0.20	0.5	0.40	0.40	0.70
b_3	0.13	-0.70	-0.5	-0.40	-0.40	-1.40
b_4	-1.70	0.20	0.5	0.40	0.40	0.70

Table 5: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 3 (Continuous).

	$t_1 = 4$ hours				$t_2 = 6$ hours			
	σ_{max}	σ_{min}	threshold	rank	σ_{max}	σ_{min}	threshold	rank
$\frac{\partial Y_B}{\partial p}$	1.48×10^{18}	4.76×10^{12}	1.2×10^4	4	7.71×10^{20}	3.41×10^{15}	6.2×10^6	4
$\frac{\partial Y_C}{\partial p}$	2.74×10^{17}	2.76×10^{11}	7.4×10^2	4	1.90×10^{19}	7.09×10^{12}	9.4×10^4	4
$\frac{\partial Y_D}{\partial p}$	8.27×10^{17}	2.83×10^{12}	6.0×10^3	4	1.92×10^{20}	2.25×10^{14}	1.5×10^6	4

Table 6: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 3.

	$t_3 = 8$ hours				$t_4 = 10$ hours			
	σ_{max}	σ_{min}	threshold	rank	σ_{max}	σ_{min}	threshold	rank
$\frac{\partial Y_B}{\partial p}$	3.46×10^{15}	1.46×10^{10}	2.3×10^{1}	4	7.07×10^{16}	6.49×10^{10}	3.8×10^{2}	4
$\frac{\partial Y_C}{\partial p}$	6.59×10^{14}	1.59×10^{8}	2.9	4	2.81×10^{15}	4.13×10^{8}	1.2×10^{1}	4
$\frac{\partial Y_D}{\partial p}$	3.26×10^{16}	1.82×10^{11}	1.9×10^{2}	4	6.98×10^{16}	1.92×10^{10}	3.8×10^{2}	4

Table 7: Stoichiometry and atom transitions for the metabolic network reactions in Fig. 5.

	Reactions	Atom Transformations
v_1	$PRY_{EX} \rightarrow PYR$	$abc \rightarrow abc$
v_2	$GLU_{EX} \rightarrow AKG$	$abced \rightarrow abced$
v_3	$PYR \rightarrow ACCOA + CO_2$	$abc \rightarrow bc + a$
v_4	$ACCOA + OAA \rightarrow CIT$	$ab + cdef \rightarrow fedbac$
v_6/v_5	$CIT \leftrightarrow AKG + CO_2$	$abcdef \leftrightarrow abcde + f$
v_7	$AKG \rightarrow 0.5SUC + 0.5SUC + CO_2$	$abcde \rightarrow 0.5abcd + 0.5dcba + e$
v_8/v_9	$SUC \leftrightarrow OAA$	$abcd \leftrightarrow abcd$
v_{10}/v_{11}	$OAA \leftrightarrow PYR + CO_2$	$abcd \rightarrow abc + d$

Table 8: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 5 at t=4 hours.

	σ_{max}	σ_{min}	threshold	rank
$PYR/CIT/AKG$	2.667×10^{107}	1.126×10^{95}	1.042×10^{95}	9
$PYR/CIT/SUC$	2.668×10^{106}	1.982×10^{95}	8.732×10^{94}	9
$CIT/SUC/OAA$	9.116×10^{106}	3.102×10^{95}	4.834×10^{94}	9

Table 9: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 5 at t=6 hours.

	σ_{max}	σ_{min}	threshold	rank
$PYR/CIT/SUC$	6.688×10^{98}	2.401×10^{87}	3.253×10^{86}	9
$CIT/SUC/OAA$	2.200×10^{98}	5.666×10^{87}	9.003×10^{85}	9

Table 10: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 5 at t=8 hours.

	σ_{max}	σ_{min}	threshold	rank
$PYR/CIT/AKG$	1.631×10^{93}	4.111×10^{81}	7.404×10^{80}	9
$PYR/CIT/SUC$	1.632×10^{93}	9.219×10^{81}	6.204×10^{80}	9
$ACCOA/CIT/SUC$	7.921×10^{91}	6.899×10^{79}	3.666×10^{79}	9
$CIT/SUC/OAA$	6.753×10^{92}	2.737×10^{82}	3.434×10^{80}	9
$CIT/AKG/OAA$	6.723×10^{92}	5.639×10^{81}	4.034×10^{80}	9

Table 11: The local structural identifiability analysis results of the cumomer balance equations of the metabolic network reactions in Fig. 5 at t=10 hours.

	σ_{max}	σ_{min}	threshold	rank
$PYR/ACCOA/CIT$	1.907×10^{89}	1.401×10^{77}	6.349×10^{76}	9
$PYR/CIT/AKG$	1.907×10^{89}	4.8710×10^{78}	9.038×10^{76}	9
$PYR/CIT/SUC$	1.913×10^{89}	3.518×10^{78}	7.573×10^{76}	9
$PYR/CIT/OAA$	2.370×10^{89}	2.655×10^{77}	7.573×10^{76}	9
$ACCOA/AKG/SUC$	1.547×10^{88}	3.318×10^{75}	2.928×10^{75}	9
$ACCOA/CIT/SUC$	1.600×10^{88}	1.402×10^{77}	8.951×10^{75}	9
$CIT/SUC/OAA$	1.416×10^{89}	7.026×10^{78}	8.385×10^{76}	9
$CIT/AKG/OAA$	1.408×10^{89}	6.503×10^{78}	9.849×10^{76}	9
$CIT/AKG/SUC$	1.540×10^{88}	7.744×10^{76}	6.156×10^{75}	9

References

1. Jang C, Chen L, Rabinowitz JD. Metabolomics and isotope tracing. *Cell* 2018;173:822–837.

2. Buescher JM, Antoniewicz M, Boros LG, Burgess SC, Brunengraber H, Clish CB, DeBerardinis RJ, Feron O, Frezza C, Ghesquiere B, Gottlieb E, Hiller K, Jones RG, Kamphorst JJ, Kibbey RG, Kimmelman AC, Locasale JW, Lunt SY, Maddocks OD, Malloy C, Metallo CM, Meuillet EJ, Munger J, Noh K, Rabinowitz JD, Ralser M, Sauer U, Stephanopoulos G, St-Pierre J, Tennant DA, Wittmann C, Heiden MGV, Vazquez A, Vousden K, Young JD, Zamboni N, Fendt S. A roadmap for interpreting 13c metabolite labeling patterns from cells. *Current Opinion in Biotechnology* 2015;34:189–201.

3. Blank LM, Kuepfer L. Metabolic flux distributions: genetic information,

computational predictions, and experimental validation. *Applied and Environmental Microbiology* 2010;86:1243–1255.

4. Nie Y, Huang M, Lu J, Qian J, Lin W, Chu J, Zhuang Y, Zhang S. Impacts of high β-galactosidase expression on central metabolism ofrecombinant pichia pastoris gs115 using glucose as sole carbonsource via 13c metabolic flux analysis. *Journal of Biotechnology* 2014;187:124–134.

5. Liu P, Huang M, Guo M, Qian J, Lin W, Chu J, Zhuang Y, Zhang S. Combined 13c assisted metabolomics and metaboic flux analysis reveals the impacts of glutamate on the central metabolism of high β galactosidase producing pichia pastoris. *Bioresources and bioprocessing* 2016;3:47–57.

6. Leighty R, Antoniewicz M. Dynamic metabolic flux analysis (dmfa): A framework for determining fluxes metabolic non-steady state. *Metab Eng* 2011;13:745–755.

7. Antoniewicz MR. Dynamic metabolic flux analysis - tools for probing transient states of metabolic networks. *Current Opinion in Biotechnology* 2013;24:973–978.

8. Lin W, Wang Z, Huang M, Chu J, Zhuang Y, Zhang S. On stability analysis of cascaded linear time varying systems in dynamic isotope experiments. *AIChE J* 2020;:In press.

9. Isermann N, Wiechert W. Metabolic isotopomer labeling systems. part ii: structural flux identifiability analysis. *Mathematical Biosciences* 2003;183:175–214.

10. Rantanen A, Mielikainen T, Rousu J, Maaheimo H, Ukkonen E. Planning optimal measurements of isotopomer distributions for estimation of metabolic fluxes. *Bioinformatics* 2006;22:1198–1206.

11. Chang Y, Suthers PF, Maranas CD. Identification of optimal measurement sets for complete flux elucidation in metabolic flux analysis experiments. *Biotechnology and Bioengineering* 2008;100:1039–1049.

12. Lin W, Wang Z, Huang M, Zhuang Y, Zhang S. On structural identifiability analysis of the cascaded linear dynamic systems in isotopically non-stationary 13c labelling experiments. *Mathematical Biosciences* 2018;300:122–129.

13. Pohjanpalo H. System identifiability based on power series expansion of solution. *Mathematical Biosciences* 1978;41:21–33.

14. Abate A, Hillen RC, Wahl SA. Piecewise affine approximations of fluxes and enzyme kinetics from in vivo 13c labeling experiments. *Int J of Robust and Nonlinear Control* 2012;22:1120–1139.

15. Wiechert W, Wurzel M. Metabolic isotopomer labeling systems part i: global dynamic behavior. *Mathematical Biosciences* 2001;169:173–205.

16. Toth R. Modeling and identification of linear parameter varying systems. Springer-Verlag Berlin Heidelberg; 2010.

17. Rugh WJ. Linear System Theory. Second ed.; Prentice-Hall, Inc.; 1996.

18. Sriram G, Shanks J. Improvement in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and boolean function mapping. *Metab Eng* 2004;6:116–132.

19. Antoniewicz M, Kelleher JK, Stephanopoulos G. Elementary metabolic units (emu): a novel framework for modeling isotopic distributions. *Metab Eng* 2007;9:68–86.

20. Lin W, Huang M, Wang Z, Zhuang Y, Zhang S. Modeling steady state intercellular isotopic distributions with isotopomer decomposition units. *Computer Chem Eng* 2019;121:248–264.

21. Holoborodko P. Multiprecision Computing Toolbox for MATLAB. Advanpix LLC.; 2015.

22. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab Eng* 2006;8:324–337.

23. Millard P, Sokol S, Letisse F, Portais J. Isodesign: A software for optimizing the design of 13c-metabolic flux analysis experiments. *Biotechnology and Bioengineering* 2014;111:202–208.