Submitted, 2019 American Control Conference (ACC)
http://www.cds.caltech.edu/~murray/preprints/sm19-acc_s.pdf

Cell-Free Extract Data Variability Reduction in the Presence of Structural Non-Identifiability

Vipul Singhal[1] and Richard M. Murray[2]

***Abstract*— The bottom up design of genetic circuits to control cellular behavior is one of the central objectives within Synthetic Biology. Performing design iterations on these circuits *in vivo* is often a time consuming process, which has led to *E. coli* cell extracts to be used as simplified circuit prototyping environments. Cell extracts, however, display large batch-to-batch variability in gene expression. In this paper, we develop the theoretical groundwork for a model based calibration methodology for correcting this variability. We also look at the interaction of this methodology with the phenomenon of parameter (structural) non-identifiability, which occurs when the parameter identification inverse problem has multiple solutions. In particular, we show that under certain consistency conditions on the sets of output-indistinguishable parameters, data variability reduction can still be performed, and when the parameter sets have a certain structural feature called covariation, our methodology may be modified in a particular way to still achieve the desired variability reduction.**

I. INTRODUCTION

Cell-free extracts have been proposed as a potential tool for the rapid prototyping of genetic circuits in synthetic biology [1]. One of the challenges in the development of this technology is that there is significant variability across different batches of extracts, which limits our ability to reliably generalize the results of any one extract. Takahashi et al. [2] showed large variation in the constitutive expression of a fluorescent protein between batches, and Hu et al. [3] showed that the variability in expression could be mapped to variability in the parameter estimates. Interestingly, Garamella et al. [4] showed minimal variability in constitutive gene expression between four extract batches. However, such reproducibility has not been demonstrated in other labs, and furthermore, Garamella et al. did not demonstrate the lack of variability in the behavior of more complex circuits.

In this paper, we develop a framework for the computational reduction of extract variability. We begin by defining some notation (Section II) and framing the variability reduction in terms of the so called *data correction problem* (Section III). We then define the *calibration-correction method*, named after a similar method developed to correct wind tunnel variability [5], which solves this problem (Section III). Next, we show that under certain consistency conditions, the presence of parameter non-identifiability does not hinder our methodology (Section IV). We also show that these consistency conditions may be violated when the non-identifiability possesses a certain structural feature, and end with a modification to the methodology that addresses this phenomenon (Section V).

This work was supported by the SBIR-STTR grant W911NF-16-P-0003 and the AFOSR grant FA9550-14-1-0060

[1]Vipul Singhal is with the Computation and Neural Systems option, California Institute of Technology, USA; and the Agency for Science, Technology and Research, Singapore. vsinghal@caltech.edu

[2]Richard M. Murray is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena CA 91125, USA. murray@cds.caltech.edu

II. NOTATION AND PRELIMINARY IDEAS

A. Systems, Experiments, Models and Parameters

We consider systems $\mathcal{S} = (\mathcal{E}, \mathcal{C})$ described as a combination of an extract $\mathcal{E}$ and a circuit $\mathcal{C}$, and define an experiment $\mathcal{H} = (\mathcal{S}, x_0, \overline{y})$ to be the execution of a system under initial conditions x_0 and output measurements $\overline{y}$, where the bar denotes the assumption that experimental data is reflects the "ground truth". Time dependent inputs may be included without significant change to the results derived in this paper, and are suppressed for simplicity.

The parameter vector θ of a model M associated with a given experiment will be partitioned into *extract specific parameter* (ESP) coordinates $e \in \mathbb{R}^{q_E}$, and *circuit specific parameter* (CSP) coordinates $c \in \mathbb{R}^{q_C}$. We do not restrict these parameters to be in the positive orthant, since any positive parameters may be log transformed.

The partition of $\theta = (e, c)$ into ESPs and CSPs may be made using the following guidelines: ESPs are parameters associated primarily with species that are present in the system regardless of the the circuit implemented. Examples include the total concentration of transcriptional or translational machinery, or elongation rates. CSPs are parameters associated with species that may no longer exist in the system when the circuit is changed. Examples include transcription factor dimerization rates.

Experiments are modeled using initialized parametrized models with the equations of the general form

$$\begin{aligned} \dot{x} &= f(x, \theta), \\ y(\theta, x_0) &= h(x, \theta), \qquad x(0) = x_0(\theta). \end{aligned} \tag{1}$$

Here $x, x_0 \in \mathbb{R}^n_+$, the solutions are assumed to exist for all $t \geq 0$, the parameter vector symbol is $\theta = (e, c) \in \Omega \subseteq \mathbb{R}^{q_E + q_C}$, where Ω is the set of all parameter values of interest. The output is denoted $y(\theta, x_0) \in \mathbb{R}^r$. The functions f and h are assumed to be analytic vector fields with respect to x in some neighborhood of any attainable x, and time dependence of the vector fields can be modeled by including t in the state variables [6]. We will use the shorthand $y(\theta, x_0) = M(\theta, x_0)$ to refer to the model in Equation (1), and will often suppress arguments such as x_0 for brevity. We will sometimes replace (θ) with (e,c), as in stating $y(e,c) = M(e,c)$ or just $M(e,c)$. We will use the hat symbol ($\hat{}$) to denote an estimated parameter value ($\hat{\theta}$, for instance), or a simulated model trajectory, $\hat{y}$. The tilde ($\tilde{}$) over parameter symbols is reserved for miscellaneous purposes, particularly in proofs.

B. The Model Universe and Model Correctness Assumptions

Our analytical results will be stated and proved in a virtual *model universe*, where artificial data $\overline{y}$ are generated using nominal models $\overline{M}$ with known nominal parameter values $\overline{\theta}$. I.e., in the model universe, we identify $\mathcal{H} = (\mathcal{S}, x_0, \overline{y})$ with $\overline{y} = \overline{M}(\overline{\theta}, x_0)$.

We will also make a *model correctness* assumption, denoted $M = \overline{M}$, which states that the models we use to estimate parameters from the data are the very models used to generate the data. The model correctness assumption allows us to isolate the interaction of non-identifiability with our method. Issues associated with model correctness or the use of approximate models (that often arise due to model order reduction) form an interesting direction for future work. Furthermore, note that we will always use single points to specify nominal parameter values, even when we can only identify sets of parameter values from the output trajectories. In this paper, we will use the model universe assumption to refer to both assumptions.

C. Parameter Non-Identifiability

In this section, we follow Walter and Lecourtier [6] in defining the notion of structural non-identifiability.

Definition 1 (Output-Indistinguishability)**.** Let $M(\theta_A)$ be a parametrized model, and let $M(\theta_B)$ be a model with the same structure. $M(\theta_A)$ and $M(\theta_B)$ are said to be *output-indistinguishable* if

$$\begin{aligned} &\theta_A,\ \theta_B \in \Omega, \\ &y(\theta_A, x_0) = y(\theta_B, x_0) \quad \forall t \geq 0,\ \forall x_0 \in \mathbb{R}^n_+. \end{aligned} \tag{2}$$

Definition 2 (Structural Global Identifiability (parameter))**.** The i^{th} coordinate of θ_A, denoted $\theta_{A,i}$, is *structurally globally identifiable* (SGI) if for almost any $\theta_A \in \Omega$, Equation (2) has a unique solution for $\theta_{B,i}$.

This means that for an SGI coordinate, output indistinguishable trajectories almost always lead to a unique estimate of the coordinate.

Definition 3 (Structural Global Identifiability (model))**.** The model $M(\theta)$ is called *structurally globally identifiable* (SGI) if all its parameters θ_i, for $i = 1, 2, \dots, q_E + q_P$, are SGI.

In the absence of global structural identifiability, multiple points in the parameter space give rise to the same output behavior. In biological applications, this situation tends to be common due to a limited number of measurements and a large number of state variables [7]. Our main goal is to demonstrate that it is not always necessary to achieve global identifiability for every parameter to achieve a modeling objective such as ours. To this end, we shall consider models with non-SGI parameters, and thus allow e and c to exist in *sets of output-indistinguishable parameters*, denoted by E and C respectively.

III. THE CALIBRATION-CORRECTION METHODOLOGY

In this section, we give formal definitions of the data correction problem, the parameter identification operation and the calibration-correction method. We begin by defining two extracts, the *reference extract* ($\mathcal{E}_1$), and a *candidate extract* ($\mathcal{E}_2$). Let $\mathcal{H}_{i,\text{cal}}$ (resp. $\mathcal{H}_{i,\text{test}}$) be an experiment performed with a *calibration circuit* $\mathcal{C}_{\text{cal}}$ (resp. *test circuit* $\mathcal{C}_{\text{test}}$) in the extract $\mathcal{E}_i$. Let M_{cal} and M_{test} be the corresponding models, and assume that the reaction mechanisms defining them are modeled at the same level of detail. For example, in this paper, we consider a constitutive gene expression circuit as $\mathcal{C}_{\text{cal}}$ and a tetR mediated repression circuit as $\mathcal{C}_{\text{test}}$. In both these circuits, gene expression is modeled with the simple enzymatic reaction,

$$\text{dna} + \text{enz} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{dna:enz} \xrightarrow{k_c} \text{dna} + \text{enz} + \text{prot}. \tag{3}$$

This is crucial because it allows for parameters estimated from one model to be used in the other. With these definitions, we may state the data correction problem (Figure 1).

Definition 4 (Data Correction Problem)**.** Consider the test circuit experiments $\mathcal{H}_{i,\text{test}} = ((\mathcal{E}_i, \mathcal{C}_{\text{test}}), x_{0,\text{test}}, \overline{y}_{i,\text{test}})$, $i = 1, 2$. Assume that we may choose the experiments $\mathcal{H}_{i,\text{cal}}$, $i = 1, 2$ and collect the corresponding data $\overline{y}_{1,\text{cal}}$ and $\overline{y}_{2,\text{cal}}$. Furthermore, assume that we may pick the models M_{cal} and M_{test}, as long as they are at the same level of modeling. Solving the *data correction problem* (DCP) involves taking the tuple $(M_{\text{cal}}, M_{\text{test}}, \overline{y}_{1,\text{cal}}, \overline{y}_{2,\text{cal}}, \overline{y}_{2,\text{test}})$ and returning a trajectory $\hat{y}_{1,\text{test}}$, such that $\hat{y}_{1,\text{test}} = \overline{y}_{1,\text{test}}$.

Remark 1. In general, the DCP will only be solvable in the model universe, where $\overline{y}_{i,\text{cal}} \triangleq \overline{M}_{\text{cal}}(\overline{e}_i, \overline{c}_{\text{cal}})$ and $\overline{y}_{i,\text{test}} \triangleq \overline{M}_{\text{test}}(\overline{e}_i, \overline{c}_{\text{test}})$, $i = 1, 2$. ◇

Remark 2. With real data, the equality $\hat{y}_{1,\text{test}} = \overline{y}_{1,\text{test}}$ must be replaced with the approximate equality $\hat{y}_{1,\text{test}} \approx \overline{y}_{1,\text{test}}$ defined in some sense. For instance, we may require that $d(\overline{y}_{1,\text{test}}, \hat{y}_{1,\text{test}}) < \frac{1}{2} d(\overline{y}_{1,\text{test}}, \overline{y}_{2,\text{test}})$, where d is an appropriate metric. ◇

Next, define the set valued parameter identification operator that will be used for studying the effect of non-identifiability on our methodology.

Definition 5 (Parameter Identification)**.** Let the set Γ_θ be the set of all pairs $(y, M(\theta))$ for which there exists a parameter $\hat{\theta} \in \Omega$ such that $y = M(\hat{\theta})$. Let $\mathcal{P}(\Omega)$ be the power set of Ω. We define the *parameter identification* of the θ coordinates of the model M as an operation $\text{ID}_\theta : \Gamma_\theta \to \mathcal{P}(\Omega)$, with $\text{ID}_\theta(y, M(\theta)) = \{\hat{\theta} \in \Omega \mid y = M(\hat{\theta})\}$

In the definition above we have explicitly included θ as a subscript to ID and Γ. This is useful because we also

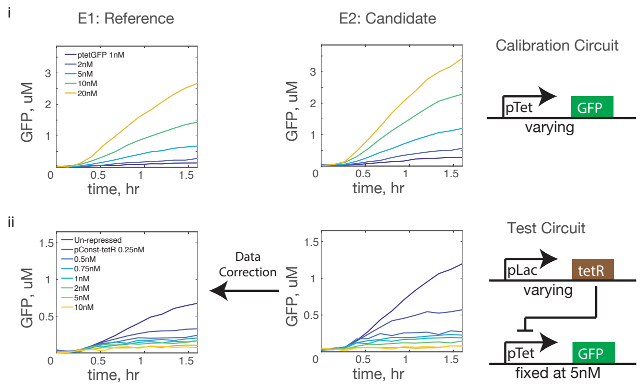

Fig. 1. The data correction problem involves the transformation of the behavior of a *test* circuit (ii, a tetR repression circuit here), from a *candidate* extract to a *reference* extract. We have the freedom to design and implement a set of *calibration* experiments ($\mathcal{H}_{i,\text{cal}}$, $i = 1, 2$) on the two extracts (i, constitutive expression of GFP here), and collect the resulting data ($\overline{y}_{1,\text{cal}}$ and $\overline{y}_{2,\text{cal}}$).

define a *conditional* version of ID, as described in Remark 3 below.

Remark 3. We define two modifications to the use of the ID_θ operator. First, we allow for the identification of a subset of parameter coordinates, such as c, with values for the remaining parameter coordinates fixed, $e = \tilde{e}$. We call this the conditional ID operator, and use the notation $\text{ID}_{c|e=\tilde{e}}(y, M(e,c))$, which will often be abbreviated to $\text{ID}_c(y, M(\tilde{e}, c))$, to describe it. The domain of the modified operator is $\Gamma_{c|e=\tilde{e}} = \{(y, M) \mid \exists c : y = M(\tilde{e}, c)\}$, or Γ_c for short. Similarly, the codomain can be either $\mathcal{P}(\mathbb{R}^{q_c})$ or $\mathcal{P}(\text{proj}_c\, \Omega)$, where proj_c denotes the projection operator from the full coordinate space to the CSP coordinates, c.

A second method of identifying values for some subset of parameter coordinates (say c once again) is to identify values over all the parameter coordinates, and then to project the resulting set down to the coordinates of interest, as in $\text{proj}_c\, \text{ID}_\theta(y, M)$, where $\theta = (e, c)$. ⋄

Next, we define the calibration-correction method as a sequence of steps involving parameter identification and prediction. Along with stating each step of the method in terms of single parameter points or trajectories, we also give descriptions of the sets of all such points and trajectories. The definitions of these sets allow for the investigation of how structural non-identifiability of parameters affects this method's ability to solve the DCP.

Briefly, the method involves performing calibration experiments to find ESPs corresponding to each extract by fitting M_cal to $\overline{y}_{i,\text{cal}}$ for $i = 1, 2$. Next, the ESPs for the candidate extracts are fixed in the test model, and the CSPs for this model are estimated using corresponding data. Finally, a corrected trajectory is generated by simulating the test model using these CSPs, along with the ESPs for the reference extract, $\mathcal{E}_1$.

Definition 6 (The Calibration-Correction Method). Consider the DCP in the model universe. We define the *calibration-correction method* as a sequence of steps that takes the tuple $(M_\text{cal}, M_\text{test}, \overline{y}_{1,\text{cal}}, \overline{y}_{2,\text{cal}}, \overline{y}_{2,\text{test}})$ and returns a prediction $\hat{y}_{1,\text{test}}$. The steps are:

1) *Calibration Step.* Find $\hat{e}_{1,\text{cal}}$ and $\hat{e}_{2,\text{cal}}$ such that $(\hat{e}_{1,\text{cal}}, \hat{e}_{2,\text{cal}}, \hat{c}_\text{cal})$ satisfies $\overline{y}_{1,\text{cal}} = M_\text{cal}(\hat{e}_{1,\text{cal}}, \hat{c}_\text{cal})$ and $\overline{y}_{2,\text{cal}} = M_\text{cal}(\hat{e}_{2,\text{cal}}, \hat{c}_\text{cal})$ for some $\hat{c}_\text{cal}$. The sets of all such ESP points are found by projecting the set
$$\tilde{\Theta}_\text{cal} \triangleq \left\{ (e_1, e_2, c) \,\middle|\, \overline{y}_{i,\text{cal}} = M_\text{cal}(e_i, c),\ i = 1, 2 \right\},$$
onto the corresponding ESP coordinates:
$$E_{i,\text{cal}} \triangleq \text{proj}_{e_i} \tilde{\Theta}_\text{cal}, \qquad i = 1, 2. \tag{4}$$

2) *Correction Step One.* Identify $\hat{c}_{2,\text{test}}$ such that $\overline{y}_{2,\text{test}} = M_\text{test}(\hat{e}_{2,\text{cal}}, \hat{c}_{2,\text{test}})$. The set of all such points is
$$C'_{2,\text{test}} \triangleq \bigcup_{\hat{e} \in E_{2,\text{cal}}} \text{ID}_{c|e=\hat{e}} \left(\overline{y}_{2,\text{test}}, M_\text{test}(e, c) \right). \tag{5}$$

3) *Correction Step Two.* Generate the prediction $\hat{y}_{1,\text{test}} \triangleq M_\text{test}(\hat{e}_{1,\text{cal}}, \hat{c}_{2,\text{test}})$. Note that the set of all such predictions is
$$Y_1 \triangleq \bigcup_{\hat{e} \in E_{1,\text{cal}}} \bigcup_{\hat{c} \in C'_{2,\text{test}}} \hat{y}_1(\hat{e}, \hat{c}), \tag{6}$$
with individual predictions $\hat{y}_1(\hat{e}, \hat{c}) \triangleq M_\text{test}(\hat{e}, \hat{c})$.

Remark 4. The version of the calibration step defined above is straightforward to implement computationally. It involved a single estimation step, followed by projections. We also give an equivalent (equivalence established by the Lemma in Appendix VI-A) definition that allows for the estimation of the parameters of the two extracts separately. Start by estimating the joint ESP-CSP sets for individual extracts, $\Theta_{i,\text{cal}} \triangleq \text{ID}_\theta \left(\overline{y}_{i,\text{cal}}, M_\text{cal}(\theta) \right)$, $i = 1, 2$, and then compute the set of CSPs where these agree, $C_\text{cal} \triangleq \text{proj}_c\, \Theta_{1,\text{cal}} \cap \text{proj}_c\, \Theta_{2,\text{cal}}$. Finally, the ESP sets are generated by restricting the $\Theta_{i,\text{cal}}$ by C_cal and projecting the result into the ESP coordinates, $E_{i,\text{cal}} \triangleq \left\{ e \mid \exists c \in C_\text{cal} : (e, c) \in \Theta_{i,\text{cal}} \right\}$, $i = 1, 2$.

The fact that the sets $\Theta_{i,\text{cal}}$, $i = 1, 2$, are estimated separately can be useful in cases where the dimension of the spaces e and c live in (i.e., q_E and q_C) are large enough that estimating $\tilde{\Theta}_\text{cal} \in \mathbb{R}^{2q_E + q_C}$ might be much more difficult than estimating $\Theta_{i,\text{cal}} \in \mathbb{R}^{q_E + q_C}$. The trade-off here is that intersections and restrictions of sets represented by point clouds can be computationally difficult. ⋄

Remark 5. Note that the set $C'_{2,\text{test}}$ is a subset of the larger set $C_{2,\text{test}} \triangleq \text{proj}_c\, \text{ID}_\theta(\overline{y}_{2,\text{test}}, M_\text{test})$. Indeed, $C'_{2,\text{test}}$ is obtained from $C_{2,\text{test}}$ by only keeping the points whose corresponding e coordinate values were in the calibration set $E_{2,\text{cal}}$. We use $C'_{2,\text{test}}$ because in the first correction step, we identify c only after fixing the value of e to an arbitrary point within $E_{2,\text{cal}}$. ⋄

Remark 6. We can define two *failure conditions* for the calibration-correction method that will be useful in deriving the main theoretical results of this chapter. Both the conditions must be avoided for the calibration-correction method to solve the DCP.

The first condition (FC1) occurs if a parameter identification step is attempted when no parameter exists such that the model fits the data. This means that the data-model pair (y, M) under consideration is not in the domain, Γ, of the operator ID. For example, in the first correction step, if $\hat{e}_{2,\text{cal}}$ is such that there is no $\tilde{c}$ that satisfies $\overline{y}_{2,\text{test}} = M_{\text{test}}(\hat{e}_{2,\text{cal}}, \tilde{c})$, then the parameter estimation step fails at this point. In terms of Equation (5), this failure condition occurs if it occurs for *any* point e in $E_{2,\text{cal}}$.

The second failure condition (FC2) occurs if correction step two is able to produce a trajectory not equal to the true trajectory, i.e., $\hat{y}_{1,\text{test}} \neq \overline{y}_{1,\text{test}}$. In terms of the set Y_1 defined in Equation (6), this means that Y_1 contains at least one element that is not equal to $\overline{y}_{1,\text{test}}$. ⋄

IV. IDENTIFIABILITY CONDITIONS

In this section, we show that the SGI property is not necessary for the calibration-correction method to solve the DCP. This will be stated as a corollary of the main result of this section (Theorem 1), which gives conditions on the sets of parameters obtained during the calibration-correction method such that the method solves the DCP.

The results rely on two insights. First, the set of output indistinguishable parameters is an equivalence class with respect to the problem of fitting model output trajectories to data [8]. In terms of implementation, this means that an *arbitrary* point may be picked from this set, and the model output will fit the data trajectories at this point. Second, the calibration-correction method involves only fitting and predicting the output trajectories, and not the full state trajectories. This allows us to consider the possibility of treating the sets of parameters obtained during the calibration step and the first correction step as equivalence classes with respect to the prediction step (correction step two) of the method. Indeed, we derive conditions under which we may pick arbitrary points from the sets $E_{2,\text{cal}}$, $C_{2,\text{test}}$ and $E_{1,\text{cal}}$ and still have the method solve the DCP.

Theorem 1 (Parameter consistency)**.** *Consider the DCP in the model universe. Furthermore, consider the calibration-correction method, and the sets $\tilde{\Theta}_{\text{cal}}$, $E_{1,\text{cal}}$, $E_{2,\text{cal}}$ and $C'_{2,\text{test}}$. Define $\Theta_{i,\text{test}} \triangleq \text{ID}_\theta\left(\overline{y}_{i,\text{test}}, \overline{M}_{\text{test}}(\theta)\right)$ for $i = 1, 2$. Then, the conditions*

$$\tilde{\Theta}_{\text{cal}} \neq \emptyset, \tag{7}$$

$$E_{2,\text{cal}} \subseteq \text{proj}_e\, \Theta_{2,\text{test}}, \tag{8}$$

$$E_{1,\text{cal}} \times C'_{2,\text{test}} \subseteq \Theta_{1,\text{test}}, \tag{9}$$

are necessary and sufficient for the calibration-correction method to solve the DCP.

Proof. We note that solving the DCP using the calibration-correction method simply involves avoiding the failure conditions FC1 and FC2 described in Remark 6. Avoiding FC1 wherever it may occur ensures that the method can be implemented in the first place, and avoiding FC2 means that the method returns only the desired result. Thus, we must show that the conditions (7-9) are necessary and sufficient for avoiding FC1 and FC2.

The necessity of condition (7) follows from the fact that if $\tilde{\Theta}_{\text{cal}} = \emptyset$, then there does not exist a vector (e_1, e_2, c) such that $\overline{y}_{i,\text{cal}} = M_{\text{cal}}(e_i, c)$ for $i = 1, 2$, leading to FC1 at the calibration step. We note in passing that in the model universe, condition (7) always holds.

Next, we prove the necessity of $E_{2,\text{cal}} \subseteq E_{2,\text{test}} \triangleq \text{proj}_e\, \Theta_{2,\text{test}}$. Assume that there exists an $\tilde{e} \in E_{2,\text{cal}}$ such that $\tilde{e} \notin E_{2,\text{test}}$. Thus, there does not exist a $\tilde{c}$ such that $M_{\text{test}}((\tilde{e}, \tilde{c})) = \overline{y}_{2,\text{test}}$. Since the operator $\text{ID}_{c|e=\tilde{e}}$ is only defined on the set $\{(y, M) \mid \exists c : M((\tilde{e}, c)) = y\}$, we see that the map $\text{ID}_{c|e=\tilde{e}}(\overline{y}_{2,\text{test}}, M_{\text{test}}(e, c))$ is not well defined, leading to FC1 at the first correction step.

We prove the necessity of condition (9) as follows. Assume that there exists a $(\tilde{e}, \tilde{c}) \in E_{1,\text{cal}} \times C'_{2,\text{test}}$ such that $(\tilde{e}, \tilde{c}) \notin \Theta_{1,\text{test}}$. Since we use points $\hat{e} \in E_{1,\text{cal}}$ and $\hat{c} \in C'_{2,\text{test}}$ to generate the prediction $\hat{y}_{1,\text{test}}$ in the second correction step, it is possible that $\hat{e} = \tilde{e}$ and $\hat{c} = \tilde{c}$. Since $\Theta_{1,\text{test}}$ is the set of all points (e, c) that give the correct trajectory $\overline{y}_{1,\text{test}}$, we have the possibility that $\hat{y}_{1,\text{test}} \neq \overline{y}_{1,\text{test}}$, giving us FC2.

Finally, sufficiency is a simple consequence of the fact that conditions (7-9) address both the points in the method where FC1 could be met, and the point in the method where FC2 could occur. Explicitly, condition (7) allows the calibration step to avoid FC1, condition (8) allows correction step one to avoid FC1, since it implies that for all $\tilde{e} \in E_{2,\text{cal}}$, there exists a $\tilde{c}$ such that $(\tilde{e}, \tilde{c}) \in \Theta_{2,\text{test}}$. Condition (9) enables correction step two to avoid FC2, since it implies that for all $\tilde{e} \in E_{1,\text{cal}}$ and for all $\tilde{c} \in C'_{2,\text{test}}$ we have that $\overline{y}_{1,\text{test}} = M_{\text{test}}(\tilde{e}, \tilde{c})$, implying that the set of all possible predicted trajectories only has the correct trajectory in it, $Y_1 = \{\overline{y}_{1,\text{test}}\}$. □

Remark 7. We can give some physical interpretations of the conditions (7-9). To do this, we first note that condition (9) implies (see Lemma 3 in Appendix VI-B)

$$E_{1,\text{cal}} \subseteq \text{proj}_e\, \Theta_{1,\text{test}}, \tag{10}$$

$$C'_{2,\text{test}} \subseteq C'_{1,\text{test}}, \tag{11}$$

where $C'_{1,\text{test}}$ is defined in a similar way to $C'_{2,\text{test}}$.

Condition (7) and (10) may be interpreted to mean that the calibration experiments must be more informative about the ESPs than the test circuit experiments. This follows from the fact that the sets of output-indistinguishable ESPs obtained from the calibration step are subsets of the corresponding sets from the test circuits, $\text{proj}_e\, \Theta_{i,\text{test}}$.

Condition (11) says that the CSP sets for the test circuit, if estimated by first fixing the ESPs to values obtained at the calibration stage, must agree. Agreement here is defined to be unidirectional, with one set being a subset of another. This is only because the correction being performed is from the candidate extract to the reference extract. If bidirectional correction (Corollary 2, below) were required, then we would have equality in condition (11).

Finally, condition (9) says that the ESP and CSP coordinates in the set $\Theta_{1,\text{test}}$ can only *covary* outside $E_{1,\text{cal}} \times C'_{2,\text{test}}$, i.e., all the points within this set must belong to $\Theta_{1,\text{test}}$. Covariation is defined in Section V. ◇

Next, we state a few corollaries of the theorem.

Corollary 1 (SGI Sufficiency)**.** *SGI models are sufficient for the calibration-correction method to solve the DCP in the model universe.*

Proof. Since the models are SGI, the nominal model universe parameters uniquely fit the model to the data, and therefore the sets in conditions (7-9) only have single entries. Therefore, these conditions are trivially satisfied:

$$\begin{aligned}
&\tilde{\Theta}_{\text{cal}} = \{(\bar{e}_1, \bar{e}_2, \bar{c}_{\text{cal}})\} \neq \emptyset, \\
&E_{2,\text{cal}} = \{\bar{e}_2\} \subseteq \text{proj}_e\{(\bar{e}_2, \bar{c}_{\text{test}})\} = \text{proj}_e\, \Theta_{2,\text{test}}, \\
&E_{1,\text{cal}} \times C'_{2,\text{test}} = \{\bar{e}_1\} \times \{\bar{c}_{\text{test}}\} \subseteq \{(\bar{e}_1, \bar{c}_{\text{test}})\} = \Theta_{1,\text{test}}.
\end{aligned}$$

□

Corollary 2 (Bidirectional Correction)**.** *To be able to correct the test data from either extract to the other requires that:*

$$\begin{aligned}
\tilde{\Theta}_{\text{cal}} &\neq \emptyset, \\
E_{i,\text{cal}} &\subseteq \text{proj}_e\, \Theta_{i,\text{test}}, \qquad i = 1, 2, \\
E_{1,\text{cal}} \times C'_{2,\text{test}} &\subseteq \Theta_{1,\text{test}}, \\
E_{2,\text{cal}} \times C'_{1,\text{test}} &\subseteq \Theta_{2,\text{test}}.
\end{aligned}$$

Proof. The proof is a simple union of the sets of conditions implied by Theorem 1 for each direction of correction. □

Remark 8. We note that the condition $C'_{2,\text{test}} \subseteq C'_{1,\text{test}}$ discussed in Remark 7 gets transformed into $C'_{2,\text{test}} = C'_{1,\text{test}}$. ◇

Next we discuss the case of correcting the calibration data itself. This will be important in the next section when we examine the effect of a phenomenon called parameter covariation on the calibration-correction method. There, we will prove that a modified version of the method is able to solve the problem at least for this case, even in the presence of parameter covariation.

Corollary 3 ('Test = Calib' Case)**.** *Consider the DCP for the case where $\overline{y}_{i,\text{test}} = \overline{y}_{i,\text{cal}}$ and $\overline{M}_{\text{test}} = \overline{M}_{\text{cal}}$ for $i = 1, 2$. Furthermore, define $\Theta_{i,\text{cal}} \triangleq \text{ID}_\theta\left(\overline{y}_{i,\text{cal}}, M_{\text{cal}}(\theta)\right)$ for $i = 1, 2$, and*

$$C'_{2,\text{cal}} \triangleq \bigcup_{\tilde{e} \in E_{2,\text{cal}}} \text{ID}_c\left(\overline{y}_{2,\text{cal}}, M_{\text{cal}}(\tilde{e}, c)\right). \tag{12}$$

Then, the conditions

$$\tilde{\Theta}_{\text{cal}} \neq \emptyset, \tag{13}$$

$$E_{2,\text{cal}} \subseteq \text{proj}_e\, \Theta_{2,\text{cal}}, \tag{14}$$

$$E_{1,\text{cal}} \times C'_{2,\text{cal}} \subseteq \Theta_{1,\text{cal}}, \tag{15}$$

are necessary and sufficient for the calibration correction method to solve this problem.

Proof. Simply specialize Theorem 1 to this case. □

V. PARAMETER COVARIATION

In this section, we describe parameter covariation (Figure 2), and show that it causes the calibration correction method to fail. We then discuss an improvement to the method that addresses this issue. We start by defining a device that will be useful for taking slices of parameter sets.

Definition 7 (Cutting Plane)**.** Consider the space of parameters $\mathbb{R}^q$, the vector $\theta \in \mathbb{R}^q$ partitioned into two sets of coordinates $\theta = (\theta_a, \theta_b) \in \mathbb{R}^{q_a} \times \mathbb{R}^{q_b}$ and the subspaces $A \triangleq \mathbb{R}^{q_a} \times \{0\}$ and $B \triangleq \{0\} \times \mathbb{R}^{q_b}$ corresponding to the θ_a and θ_b coordinates respectively. Let $\tilde{\theta}_a \in A$. Then, we denote the *cutting plane* generated by shifting the origin of B to $(\tilde{\theta}_a, 0)$ with the notation $\text{cut}_{\theta_b}(\tilde{\theta}_a)$.

Definition 8 (Parameter Covariation)**.** Consider the space of parameters $\mathbb{R}^q$ and the vector $\theta \in \mathbb{R}^q$ partitioned into two sets of coordinates $\theta = (\theta_a, \theta_b) \in \mathbb{R}^{q_a} \times \mathbb{R}^{q_b}$. Consider some set of parameters $\Theta \subseteq \mathbb{R}^q$. If there exist $\tilde{\theta}_{a1}, \tilde{\theta}_{a2} \in \text{proj}_{\theta_a} \Theta$ such that $\text{proj}_{\theta_b}\left(\Theta \cap \text{cut}_{\theta_b}(\tilde{\theta}_{a1})\right) \neq \text{proj}_{\theta_b}\left(\Theta \cap \text{cut}_{\theta_b}(\tilde{\theta}_{a2})\right)$, then Θ is said to have *parameter covariation* of its θ_b coordinates with respect to its θ_a coordinates.

Remark 9. We will often abbreviate parameter covariation to just covariation, and say that parameter coordinates can *covary*. ◇

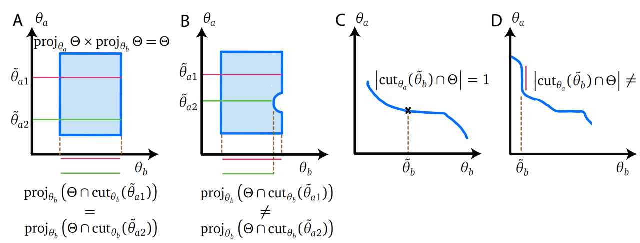

Fig. 2. Parameter covariation. (A) A Cartesian product condition is equivalent to a set not having covariation (Lemma 1). (B) The definition of covariation illustrated. (C) Thin covariation in the θ_a coordinates with respect to the θ_b coordinates. (D) The set in blue does not display thin covariation.

Lemma 1. *Let $\theta = (\theta_a, \theta_b) \in \Theta \subseteq \mathbb{R}^q$ be a partition of the coordinates of $\mathbb{R}^q$. Then, the set Θ has covariation of its θ_b coordinates with respect to its θ_a coordinates if and only if $\text{proj}_{\theta_a} \Theta \times \text{proj}_{\theta_b} \Theta \neq \Theta$.*

Proof. First, we prove the ($\Rightarrow$) direction. Covariation implies that for some $\theta_{a1}, \theta_{a2} \in \text{proj}_{\theta_a} \Theta$ there exists a point $\tilde{\theta}_b \in \text{proj}_{\theta_b} \Theta$ such that

$$\tilde{\theta}_b \in \left(\text{proj}_{\theta_b}\left(\Theta \cap \text{cut}_{\theta_b}(\tilde{\theta}_{a1})\right) \triangle \left(\text{proj}_{\theta_b}\left(\Theta \cap \text{cut}_{\theta_b}(\tilde{\theta}_{a2})\right)\right), \tag{16}$$

where $\triangle$ is the symmetric difference set operation. It further implies that there exists a point $\tilde{\theta}_a \in \{\tilde{\theta}_{a1}, \tilde{\theta}_{a2}\} \subseteq \text{proj}_{\theta_a} \Theta$ such that $(\tilde{\theta}_a, \tilde{\theta}_b) \notin \Theta$. Thus, $\text{proj}_{\theta_a} \Theta \times \text{proj}_{\theta_b} \Theta \neq \Theta$.

Next, we prove the ($\Leftarrow$) direction. Let $(\tilde{\theta}_{a1}, \tilde{\theta}_b) \in \mathrm{proj}_{\theta_a} \Theta \times \mathrm{proj}_{\theta_b} \Theta$ be such that $(\tilde{\theta}_{a1}, \tilde{\theta}_b) \notin \Theta$. Since $\tilde{\theta}_b \in \mathrm{proj}_{\theta_b} \Theta$, there exists a $\tilde{\theta}_{a2} \in \mathrm{proj}_{\theta_a} \Theta$ such that $(\tilde{\theta}_{a2}, \tilde{\theta}_b) \in \Theta$. Thus we have $\tilde{\theta}_b \in \mathrm{proj}_{\theta_b} \left(\Theta \cap \mathrm{cut}_{\theta_b}(\tilde{\theta}_{a2})\right)$ but $\tilde{\theta}_b \notin \mathrm{proj}_{\theta_b} \left(\Theta \cap \mathrm{cut}_{\theta_b}(\tilde{\theta}_{a1})\right)$, which proves the assertion. $\square$

Corollary 4. *The set Θ has covariation of its θ_b coordinates with respect to its θ_a coordinates if and only if it has covariation of its θ_a coordinates with respect to its θ_b coordinates.*

Proof. Using Lemma 1, along with a version of it where the roles of θ_a and θ_b are swapped, leads to this result. $\square$

Remark 10. This equivalence will allow us to refer to sets having covariation with respect to a given partition, such as (e, c). $\diamond$

Next, we show that in the presence of covariation, the calibration-correction method is unable to solve the DCP even in the 'Test = Calib' case of Corollary 3. In particular, we will assume that the restriction of $\Theta_{1,\mathrm{cal}}$ to $E_{1,\mathrm{cal}} \times \mathrm{proj}_c \Theta_{2,\mathrm{cal}}$ has covariation with respect to the (e, c) partition.

Proposition 1. *Consider the 'Test = Calib' case of the DCP. Assume the conditions*

$$\tilde{\Theta}_{\mathrm{cal}} \neq \emptyset, \tag{17}$$

$$C'_{2,\mathrm{cal}} \subseteq \mathrm{proj}_c \Theta_{1,\mathrm{cal}}, \tag{18}$$

$$E_{i,\mathrm{cal}} \subseteq \mathrm{proj}_e \Theta_{i,\mathrm{cal}}, \qquad i = 1, 2, \tag{19}$$

hold, but the set

$$\Theta'_{1,\mathrm{cal}} \triangleq \Theta_{1,\mathrm{cal}} \cap \left(E_{1,\mathrm{cal}} \times \mathrm{proj}_c \Theta_{2,\mathrm{cal}}\right) \tag{20}$$

has covariation with respect to the (e, c) partition. Then, the calibration-correction method fails to solve this problem.

Proof. Condition (18), along with the fact that for the 'Test = Calib' case, $C'_{2,\mathrm{cal}} = \mathrm{proj}_c \Theta_{2,\mathrm{cal}}$, implies that $\mathrm{proj}_c \Theta'_{1,\mathrm{cal}} = C'_{2,\mathrm{cal}}$. Condition (19) implies $\mathrm{proj}_e \Theta'_{1,\mathrm{cal}} = E_{1,\mathrm{cal}}$. Covariation implies that $\mathrm{proj}_e \Theta'_{1,\mathrm{cal}} \times \mathrm{proj}_c \Theta'_{1,\mathrm{cal}} \neq \Theta'_{1,\mathrm{cal}}$. Thus, the proper subset relation $\Theta'_{1,\mathrm{cal}} \subsetneq E_{1,\mathrm{cal}} \times C'_{2,\mathrm{cal}}$ holds, and therefore there exists $(\tilde{e}, \tilde{c}) \in E_{1,\mathrm{cal}} \times C'_{2,\mathrm{cal}}$ such that $(\tilde{e}, \tilde{c}) \notin \Theta'_{1,\mathrm{cal}} \subseteq \Theta_{1,\mathrm{cal}}$. This implies that $E_{1,\mathrm{cal}} \times C'_{2,\mathrm{cal}} \nsubseteq \Theta_{1,\mathrm{cal}}$, which violates condition (15). $\square$

Next, we show that for a specific type of covariation, which we call *thin* covariation, a modified version of the calibration-correction method is able to solve the DCP for the 'Test = Calib' case.

Definition 9 (Thin Covariation)**.** Let $\Theta \subset \mathbb{R}^q$ be a set of parameters and let $(\theta_a, \theta_b) \in \mathbb{R}^q$ be a partition of the coordinates of $\mathbb{R}^q$. If Θ covaries with respect to this partition and if for all $\tilde{\theta}_b \in \mathrm{proj}_{\theta_b} \Theta$, we have $\left|\mathrm{cut}_{\theta_a}(\tilde{\theta}_b) \cap \Theta\right| = 1$, then we say that the covariation of the θ_a coordinates of Θ is thin with respect to the θ_b coordinates.

Remark 11. We note that if $\Theta \triangleq \mathrm{ID}_\theta(\overline{y}, M(\theta))$, then the condition that for all $\tilde{\theta}_b \in \mathrm{proj}_{\theta_b} \Theta$, we have

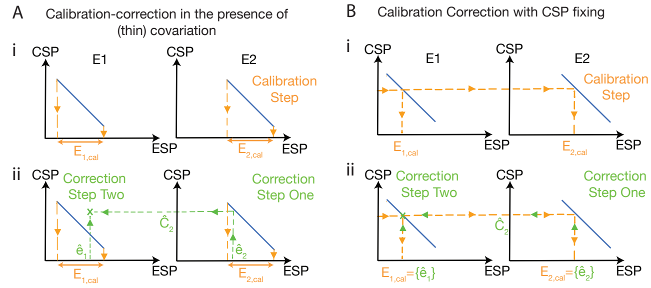

Fig. 3. (A) A schematic description of how thin covariation between the ESP-CSP coordinates in the estimated joint parameter sets can cause calibration-correction to fail at correcting even the calibration data ('Test = Calib' special case). The blue lines in all the plots are the joint ESP-CSP sets of all the parameter values that fit the calibration model to data. (B) How the CSP fixing modification (Definition 10) to the calibration step helps solve this issue. The ESP sets estimated at the calibration step are now generated by first intersecting the parameter sets (blue lines) with a line parallel to the ESP axis ('cutting plane' parallel to the ESP subspace in higher dimensions) centered at an arbitrary CSP value that can be attained, and secondly projecting these intersections to the ESP coordinates for both extracts.

$\left|\mathrm{cut}_{\theta_a}(\tilde{\theta}_b) \cap \Theta\right| = 1$ is equivalent to the θ_a coordinates of the model $M(\theta_a, \theta_b)$ being SGI for each fixed θ_b. $\diamond$

Remark 11 says that this type of covariation is essentially a statement about the some coordinates being conditionally structurally globally identifiable, despite covarying with respect to the remaining coordinates.

Definition 10 (CSP Fixing)**.** Consider the sets $\Theta_{i,\mathrm{cal}} \triangleq \mathrm{ID}_\theta\left(\overline{y}_{i,\mathrm{cal}}, M_{\mathrm{cal}}(\theta)\right)$, $i = 1, 2$ and let $\tilde{c} \in \mathrm{proj}_c \Theta_{1,\mathrm{cal}} \cap \mathrm{proj}_c \Theta_{2,\mathrm{cal}}$. Then, we define CSP fixing as a modification to the calibration step in which the sets $E_{i,\mathrm{cal}} \triangleq \mathrm{proj}_e\left(\mathrm{cut}_e(\tilde{c}) \cap \Theta_{i,\mathrm{cal}}\right)$ for $i = 1, 2$.

Proposition 2. *Consider the sets $\Theta_{i,\mathrm{cal}} \triangleq \mathrm{ID}_\theta\left(\overline{y}_{i,\mathrm{cal}}, M_{\mathrm{cal}}(\theta)\right)$ for $i = 1, 2$, and the partition $\theta = (e, c)$. Assume that the $\Theta_{i,\mathrm{cal}}$ have thin covariation in their c coordinates with respect to their e coordinates. Then, the calibration-correction method with CSP fixing is able to solve the DCP for the 'Test = Calib' case of Corollary 3.*

Proof. Let $\tilde{c} \in \mathrm{proj}_c \Theta_{1,\mathrm{cal}} \cap \mathrm{proj}_c \Theta_{2,\mathrm{cal}}$ and $\tilde{e}_2 \in E_{2,\mathrm{cal}} \triangleq \mathrm{proj}_e\left(\mathrm{cut}_e(\tilde{c}) \cap \Theta_{2,\mathrm{cal}}\right)$. We note that the sets $\mathrm{proj}_c\left(\mathrm{cut}_c(\tilde{e}_2) \cap \Theta_{2,\mathrm{cal}}\right) = \mathrm{ID}_c(\overline{y}_{2,\mathrm{cal}}, M_{\mathrm{cal}}(\tilde{e}_2, c))$ are equal by definition. Now, pick an arbitrary point $\tilde{c}' \in \mathrm{proj}_c\left(\mathrm{cut}_c(\tilde{e}_2) \cap \Theta_{2,\mathrm{cal}}\right)$. It follows that $\tilde{c}' = \tilde{c}$ from the fact that $\tilde{c} \in \mathrm{proj}_c\left(\mathrm{cut}_c(\tilde{e}_2) \cap \Theta_{2,\mathrm{cal}}\right)$ and that the element in $\left|\mathrm{cut}_{\theta_a}(\tilde{\theta}_b) \cap \Theta\right| = 1$ is unique. Thus, the only possible CSP value that can be returned by the first correction step is $\tilde{c}$.

Next, we look at the second correction step. Pick an arbitrary $\tilde{e}_1 \in E_{1,\mathrm{cal}} \triangleq \mathrm{proj}_e\left(\mathrm{cut}_e(\tilde{c}) \cap \Theta_{1,\mathrm{cal}}\right)$. Since the point $(\tilde{e}_1, \tilde{c}) \in \Theta_{1,\mathrm{cal}}$, we have that $\overline{y}_{1,\mathrm{cal}} = \hat{y}_{1,\mathrm{cal}} \triangleq M(\tilde{e}_1, \tilde{c})$, and FC2 is avoided. $\square$

VI. DISCUSSION

The framework presented in this work is not limited to cell extracts, and is readily generalizable to correct *process* behavior between different *environments*. Examples include correcting circuit behavior between cell strains or between *in vitro* and *in vivo* environments, or even between wind tunnels [5].

Extensions include generalizing condition 9 and CSP fixing to a prescription of how part models with parameter non-identifiability can be combined to predict the behavior of an entire system. Indeed, dealing with covariation by partially identifying parameters might be used for experiment design. We may also generalize these results to the case when there is noise in the data, and notions of practical identifiability [7] must be considered.

APPENDIX

A. Equivalence of the Two Definitions of the Calibration Step

In this section, we prove two identities that establish the equivalence of the two definitions of the calibration step given in Definition 6 and Remark 4.

Lemma 2. *Let $\tilde{\Theta}_{\text{cal}}$, $\Theta_{1,\text{cal}}$ and $\Theta_{2,\text{cal}}$ be as defined in Definition 6 and Remark 4. Then, the identities*

$$\operatorname{proj}_c \tilde{\Theta}_{\text{cal}} \equiv \operatorname{proj}_c \Theta_{1,\text{cal}} \cap \operatorname{proj}_c \Theta_{2,\text{cal}}, \tag{21}$$

$$\operatorname{proj}_{e_i} \tilde{\Theta}_{\text{cal}} \equiv \Big\{ e \;\Big|\; \exists c \in \big(\operatorname{proj}_c \Theta_{1,\text{cal}} \cap \operatorname{proj}_c \Theta_{2,\text{cal}} \big) \\ \textit{s.t. } (e,c) \in \Theta_{i,\text{cal}} \Big\}, \qquad i = 1, 2, \tag{22}$$

hold.

Proof. First, we prove (21) using a series of equivalences. Let $\tilde{c} \in \operatorname{proj}_c \tilde{\Theta}_{\text{cal}}$. This is equivalent to

$$\begin{aligned} &\exists e_1, e_2 : (e_1, e_2, \tilde{c}) \in \tilde{\Theta}_{\text{cal}} \\ &\Longleftrightarrow \exists e_1, e_2 : \overline{y}_{i,\text{cal}} = M_{\text{cal}}(e_i, \tilde{c}), \qquad i = 1,2 \\ &\Longleftrightarrow (e_i, \tilde{c}) \in \Theta_{i,\text{cal}}, \qquad i = 1,2 \\ &\Longleftrightarrow \tilde{c} \in \operatorname{proj}_c \Theta_{1,\text{cal}} \cap \operatorname{proj}_c \Theta_{2,\text{cal}}, \end{aligned}$$

which proves the assertion.

Next, we prove (22) for e_1 by showing that the left and right hand sides are subsets of each other. The proof for the e_2 case is similar. Denote the set on the left hand side with L, and the one on the right with R. Let $\tilde{e}_1 \in L = \operatorname{proj}_{e_1} \tilde{\Theta}_{\text{cal}}$. Then, $\exists \tilde{e}_2, \tilde{c}$ such that $(\tilde{e}_1, \tilde{e}_2, \tilde{c}) \in \tilde{\Theta}_{\text{cal}}$, which implies $\tilde{c} \in \operatorname{proj}_c \tilde{\Theta}_{\text{cal}}$ and $\overline{y}_{1,\text{cal}} = M_{\text{cal}}(\tilde{e}_1, \tilde{c})$. By the identity (21), we have that $\tilde{c} \in \operatorname{proj}_c \Theta_{1,\text{cal}} \cap \operatorname{proj}_c \Theta_{2,\text{cal}}$ and $(\tilde{e}_1, \tilde{c}) \in \Theta_{1,\text{cal}}$, which shows that $L \subseteq R$.

We conclude the proof by showing that $R \subseteq L$. Let $\tilde{e}_1 \in R$, which means that there exists a $\tilde{c} \in \operatorname{proj}_c \Theta_{1,\text{cal}} \cap \operatorname{proj}_c \Theta_{2,\text{cal}}$ such that $\overline{y}_{1,\text{cal}} = M_{\text{cal}}(\tilde{e}_1, \tilde{c})$. Furthermore, since $\tilde{c} \in \operatorname{proj}_c \Theta_{2,\text{cal}}$, there also exists an $\tilde{e}_2$ such that $\overline{y}_{2,\text{cal}} = M_{\text{cal}}(\tilde{e}_2, \tilde{c})$. Together these imply that $(\tilde{e}_1, \tilde{e}_2, \tilde{c}) \in \tilde{\Theta}_{\text{cal}}$, which gives $\tilde{e}_1 \in \operatorname{proj}_{e_1} \tilde{\Theta}_{\text{cal}}$, proving the assertion. □

B. Equivalence of the Two CSP Subset Conditions Given in Remark 7

The Cartesian product condition given in Equation (9) implies two further conditions, which we state in Lemma 3 below.

Lemma 3. *Condition* (9)*, which states that $E_{1,\text{cal}} \times C'_{2,\text{test}} \subseteq \Theta_{1,\text{test}}$, implies that*

$$E_{1,\text{cal}} \subseteq \operatorname{proj}_e \Theta_{1,\text{test}}, \tag{23}$$

$$C'_{2,\text{test}} \subseteq C'_{1,\text{test}}, \tag{24}$$

where $C'_{1,\text{test}}$ is defined in a similar way to $C'_{2,\text{test}}$,

$$C'_{1,\text{test}} \triangleq \bigcup_{\hat{e} \in E_{1,\text{cal}}} \operatorname{ID}_{c|e=\hat{e}} \Big(\overline{y}_{1,\text{test}}, M_{\text{test}}\big(e, c\big) \Big).$$

Proof. Condition (23) follows simply by applying the proj_e operator to both sides of condition (9). To prove condition (24), we note that condition (9) implies that for an arbitrary $\tilde{c} \in C'_{2,\text{test}}$, we have that for all $\tilde{e} \in E_{1,\text{cal}}$, the model fits the data, $\overline{y}_{1,\text{test}} = M_{\text{test}}(\tilde{e}, \tilde{c})$. This in turn implies that

$$\tilde{c} \in \bigcup_{\hat{e} \in E_{1,\text{cal}}} \operatorname{ID}_{c|e=\hat{e}} \Big(\overline{y}_{1,\text{test}}, M_{\text{test}}\big(e, c\big) \Big) = C'_{1,\text{test}}.$$

Thus, $C'_{2,\text{test}} \subseteq C'_{1,\text{test}}$. □

ACKNOWLEDGMENT

The authors would like to thank Samuel Clamons, Wolfgang Halter, William Poole, Anandh Swaminathan and Andras Gyorgy for useful discussions.

REFERENCES

[1] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl, "Rapid cell-free forward engineering of novel genetic ring oscillators," *eLife*, vol. 4, p. e09771, Oct. 2015.

[2] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. J. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, and J. B. Lucks, "Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription–Translation (TX-TL) Systems," *ACS Synthetic Biology*, Mar. 2014.

[3] C. Y. Hu, J. D. Varner, and J. B. Lucks, "Generating Effective Models and Parameters for RNA Genetic Circuits," *ACS Synthetic Biology*, June 2015.

[4] J. Garamella, R. Marshall, M. Rustad, and V. Noireaux, "The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology," *ACS Synthetic Biology*, vol. 5, pp. 344–355, Apr. 2016.

[5] R. S. Swanson and C. L. Gillis, "Wind-Tunnel Calibration and Correction Procedures for Three-Dimensional Models," *NACA Wartime Reports*, vol. L4E31, Oct. 1944.

[6] E. Walter and Y. Lecourtier, "Global approaches to identifiability testing for linear and nonlinear state space models," *Mathematics and Computers in Simulation*, vol. 24, pp. 472–482, Dec. 1982.

[7] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, pp. 1923–1929, Aug. 2009.

[8] S. Vajda, "Structural equivalence of linear systems and compartmental models," *Mathematical Biosciences*, vol. 55, pp. 39–64, July 1981.