Modeling uncertain biomass composition in genome-scale metabolic models with Flexible Nets

Teresa Joven[1][0009−0008−7534−359X], Jorge Lázaro[1][0009−0009−5732−9150], and Jorge Júlvez[1][0000−0002−7093−228X]

Department of Computer Science and Systems Engineering,
University of Zaragoza, Zaragoza, Spain
joven@unizar.es, jorgelazaro@unizar.es, julvez@unizar.es

Abstract. Genome-scale Metabolic Models (GEMs) are mathematical representations of an organism's metabolism that describe mass-balanced relationships between metabolites using gene-protein-reaction associations. These models are used to analyze the metabolic fluxes through reactions involved in the metabolism, with the most commonly used method being Flux Balance Analysis (FBA). The usefulness of GEMs is limited by the presence of uncertain parameters, which can lead to poor predictions. In order to model uncertain biomass composition, a particular class of Flexible Nets (FNs), called ENDI, is proposed. The impact of uncertain biomass composition on the growth rate of the organism can be assessed straightforwardly by a linear programming problem.

Keywords: Performance evaluation · Non-deterministic models · Metabolic models · Flexible nets · Uncertain parameters

1 Introduction

Genome-scale metabolic models (GEMs) are mathematical representations of the metabolism of an organism that describe a whole set of mass-balanced relationships between the metabolites of the organism using gene-protein-reaction associations resulting from the annotation process and experimental information. A key element of a GEM is the biomass reaction, whose flux represents the growth rate of the organism. Determining exactly the biomass reaction of a GEM is a challenging task [1] since it is a fake reaction that, taking into account the biomass compositions, specifies the metabolites consumed and produced by the cellular growth together with their corresponding stoichiometric weights. For simplicity, these weights are *exact* real numbers in most modelling formalisms. Nevertheless, their actual values are usually uncertain. The effect of biomass composition on the accuracy of the model has been investigated in several works, e.g. [2], which show how the distribution of fluxes in the metabolism depends on the biomass composition. This paper proposes the modelling formalism of Flexible Nets (FNs) to model GEMs with uncertain stoichiometry and, in particular, with uncertain biomass composition.

2 Flexible Nets for GEMs

A GEM can be represented as a constraint-based model (CBM) with a tuple $C_B = \{R, M, S, lb, ub\}$ where R is the set of reactions, M is the set of metabolites, $S \in \mathbb{R}^{|M|\times|R|}$ is the stoichiometric matrix, and lb, $ub : R \to \mathbb{R}$ are lower and upper flux bounds of the reactions. Each reaction is associated with a set of reactant metabolites and a set of product metabolites (one of these sets can be empty).

The stoichiometric matrix S accounts for all the stoichiometric weights of the reactions, i.e. $S[m, r]$ is the stoichiometric weight of metabolite $m \in M$ for reaction $r \in R$. Thus, if $S[m, r] < 0$ then m is consumed when r occurs; if $S[m, r] > 0$ then m is produced when r occurs; and if $S[m, r] = 0$ then m is neither consumed nor produced when r occurs.

Given that most GEMs lack of kinetic information and just report flux for their reactions, the FNs [3] will be constrained here to event nets with default intensities, or equivalently, fluxes that are independent of the metabolic concentrations.

Definition 1 (Event net with default intensities (ENDI)) *An event net with default intensities is a tuple* $N_V = \{P, T, V, E_V, A, B, J, K\}$ *where* (P, T, V, E_V) *is a tripartite graph determining the net structure,* (A, B) *are matrices determining the stoichiometry, and* (J, K) *are matrices constraining potential intensities of the net.*

An ENDI has three types of vertices: places (set P), transitions (set T), and event handlers (set V). Each place, $p \in P$, is depicted as a circle and models a metabolite. Each transition, $t \in T$, is depicted as a rectangle and models a reaction. Each event handler, $v \in V$, is depicted as a dot and models the stoichiometry of a reaction. The vertices of the net are connected by the edges in E_V. Each pair of vertices can be connected by at most one edge. The set E_V is partitioned into two sets E_V^P (arc from a place to a handler or vice versa) and E_V^T (edge connecting a transition and a handler). Direct connections among places and transitions are not allowed. The matrices A and B model the potential ways in which the concentrations of metabolites change when reactions occur.

The ENDI in Figure 1 has two places, $P = \{p_a, p_k\}$, three transitions, $T = \{t_1, t_2, t_3\}$, and three event handlers $V = \{v_1, v_2, v_3\}$. The set of arcs is $E_V^P = \{(v_1, p_a), (v_2, p_k), (p_a, v_3), (p_k, v_3)\}$ and the set of edges is $E_V^T = \{(t_1, v_1), (t_2, v_2), (t_3, v_3)\}$.

In an ENDI, the flux of reactions is given by a vector $\lambda \in \mathbb{R}_{\geq 0}^{|R|}$, i.e. $\lambda[t]$ is the flux of reaction t, and the speed at which the amounts of metabolites change is given by a vector $\Delta m_\tau \in \mathbb{R}_{\geq 0}^{|E_V|}$, i.e., $\Delta m_\tau[(v, p)]$ is the speed at which v produces metabolites in p, and $\Delta m_\tau[(p, v)]$ be the speed at which v consumes metabolites from p.

In general, the relationship between reaction fluxes, λ, and speeds at which metabolite change is given by (see Definition 1):

$$A\Delta m_\tau \leq B\lambda \tag{1}$$

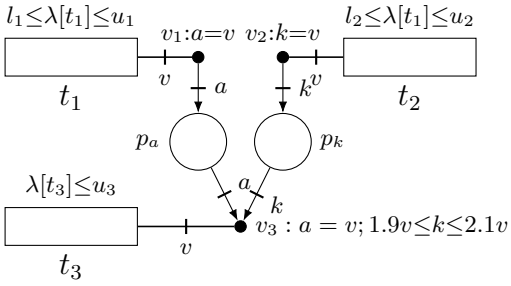

Fig. 1: Event net with default intensities (ENDI) and uncertain stoichiometry.

In order to account for uncertain fluxes, all the potential values of λ are assumed to be constrained by:

$$J\lambda \leq K \tag{2}$$

where J and K are real matrices of appropriate size. Inequality 2 allows the modeller not only to establish individual lower and upper flux bounds as in usual CBMs, but also to state linear constraints among fluxes of different reactions.

3 Modelling uncertain biomass composition

A general expression for the biomass reaction, r_g, is given by:

$$r_g : w_{r1}M_{r1} + ... + w_{rk}M_{rk} \rightarrow w_{p1}M_{p1} + ... + w_{pk'}M_{pk'} \tag{3}$$

where $\{M_{r1}, ..., M_{rk}\}$ represent the set of reactants, $\{w_{r1}, ..., w_{rk}\}$ are their corresponding coefficients, $\{M_{p1}, ..., M_{pk'}\}$ represent the set of products and their coefficients are $\{w_{p1}, ..., w_{pk'}\}$. In CBMs, these stoichiometric coefficients are *sharp* real numbers, this makes it difficult to model partially known reactions. An appealing feature of ENDIs is that they can accommodate uncertain stoichiometric coefficients.

The ENDI in Figure 1 models three reactions:

$$R_1 : \emptyset \rightarrow p_a$$
$$R_2 : \emptyset \rightarrow p_k$$
$$R_3 : p_a + kp_k \rightarrow \emptyset$$

where the stoichiometric weight k is uncertain but known to be in the interval $[1.9, 2.1]$. This uncertainty is modelled by the inequalities $1.9v \leq k \leq 2.1v$ associated with v_3.

The fluxes of the reactions R_1, R_2, and R_3 are uncertain and constrained to the intervals $[l_1, u_1]$, $[l_2, u_2]$, and $[0, u_3]$ respectively. This is modelled by the inequalities above the reactions.

4 Steady state bounds

Similarly to FBA, it will be assumed that in the steady state both, the fluxes of reactions and the amounts of metabolites keep constant. This assumption leads to the following constraint that must be satisfied in the steady state:

$$Z\Delta m_\tau = 0 \tag{4}$$

where Z is a matrix with rows indexed by P, columns indexed by E_V^P, and such that $Z[p_i,(p_i,v_k)] = -1 \ \forall \ (p_i,v_k) \in E_V^P$, $Z[p_i,(v_k,p_i)] = 1 \ \forall \ (v_k,p_i) \in E_V^P$ and the rest of the elements in Z are 0.

Equation (4) together with equations (1) and (2) can be used to establish a set of necessary conditions for the state variables, λ and Δm_τ, in the steady state.

Proposition 1 *Let N_V be an ENDI, all the potential steady states $(\lambda, \Delta m_\tau)$ belong to SS_{N_V} where:*

$$\begin{aligned} SS_{N_V} = \{(\lambda,&\Delta m_\tau)| \\ &A\Delta m_\tau \leq B\lambda \\ &Z\Delta m_\tau = 0 \\ &J\lambda \leq K\} \end{aligned} \tag{5}$$

Lower and upper bounds for a function of interest can be computed by adding an appropriate objective function to the constraints in (5). For instance, an upper bound for the flux of the biomass reaction, r_g, i.e. an upper bound for the growth rate, can be obtained by solving the following linear programming problem (LPP):

$$\begin{aligned} max \ &\lambda[t_g] \ \text{ subject to} \\ &A\Delta m_\tau \leq B\lambda \\ &Z\Delta m_\tau = 0 \\ &J\lambda \leq K \end{aligned} \tag{6}$$

where t_g is the reaction modelling biomass composition.

References

1. Bernstein, D. B., Sulheim, S., and Almaas, E., Segrè, D.: Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. Genome Biology **22**(64), 1–22 (2021). https://doi.org/10.1186/s13059-021-02289-z
2. Müller, S., Széliová, D., Zanghellini, J.: Elementary vectors and autocatalytic sets for resource allocation in next-generation models of cellular growth. PLoS Computational Biology **18**(2), e1009843 (2022). https://doi.org/10.1371/journal.pcbi.1009843
3. Júlvez, J., Oliver, S. G.: Steady state analysis of flexible nets. IEEE Transactions on Automatic Control **65**(6), 2510-2525, (2019). https://doi.org/10.1109/TAC.2019.2931836