Cell Population Balance and Hybrid Modeling of Population Dynamics for a Single Gene with Feedback

Michail Stamatakis[1]

Department of Chemical Engineering, University College London, University College London, Torrington Place, London WC1E 7JE, United Kingdom

Abstract

Latest research on biological systems is steadily shifting from isolated single cells to entire cell populations. The latter are inherently heterogeneous, and their modeling requires approaches that explicitly account for this property. A comprehensive such approach is the cell population balance (CPB), which, however, is computationally expensive and becomes intractable for multi-variable models.

In this work, we demonstrate the use of model-reduction to efficiently simulate cell population heterogeneity in a genetic network of a single gene with feedback. Starting from a 4-species model we use singular perturbation analysis to derive a single equation for the intracellular protein concentration. We subsequently incorporate this equation to a hybrid model consisting of a CPB for the cell volume, and a continuum equation for the protein concentration. We finally compare the results obtained with the hybrid model with those of the full CPB, demonstrating the accuracy and computational efficiency of the hybrid methodology.

Keywords

heterogeneity; cell population balance; hybrid model; kinetic Monte Carlo; singular perturbation; model reduction

[1] Corresponding author: m.stamatakis@ucl.ac.uk

Introduction

Heterogeneity in cell populations is a ubiquitous phenomenon that is observed even in the case of clonal cells (Avery, 2006; Davidson and Surette, 2008). This inherent property of cell populations stems from numerous sources, in particular, the random occurrence of intracellular reactions, the differences in the cell growth rates, the temporal variability in the transitions through the different phases of the cell cycle as well as in the division times, and finally, the stochasticity in the partitioning of mother cell contents to the two daughters. As a result, studying the behavior of single cells is not enough to characterize the behavior of the population, thereby motivating the development of experimental and computational approaches that can capture cell population heterogeneity (Fernandes et al., 2011).

Thus, from a modeling perspective, the aforementioned sources of heterogeneity, with the exception of the random reaction occurrence, were explicitly taken into account in the deterministic cell population balance (CPB) framework developed by Tsuchiya et al. (1966), and Fredrickson et al. (1967). These models describe the temporal evolution of the distribution of the population over a state variable. The latter can include the cell mass (Sinko and Streifer, 1971; Subramanian and Ramkrishna, 1971), the age and/or volume of the cell (Bell and Anderson, 1967; Sinko and Streifer, 1967; Bell, 1968; Anderson et al., 1969; Sinko and Streifer, 1969), and also the molar contents of intracellular components (Ataai and Shuler, 1985; Mantzaris et al., 2001a; Mantzaris et al., 2001b; Mantzaris et al., 2001c; Henson et al., 2002). More recently Fredrickson and Mantzaris (2002) and Fredrickson (2003) expanded the CPB formulation to account for the transitions between the different phases of the cell cycle. The CPB framework has been successfully used to model cell cycle dynamics (Faraday and Kirkby, 1992; Liu et al., 2007), capture cell population heterogeneity (Mantzaris, 2005b; Mantzaris, 2005a), predict and control the dynamics of fermentation processes in batch or continuous bioreactors (Godin et al., 1999; Mantzaris et al., 1999; Zhu et al., 2000; Mantzaris and Daoutidis, 2004; Sharifian and Fanaei, 2009), study aggregation dynamics in suspension cultures (Kolewe et al., 2012), as well as investigate in vitro cell proliferation patterns (Fadda et al., 2012b; Fadda et al., 2012a).

The deterministic CPB framework just discussed does not include stochasticity in intracellular reaction occurrence. This source of randomness is however important in the spatial scales on the order of a cell, because of the small number of molecules involved in biochemical reactions. To account for such effects, Mantzaris (2007) proposed a computational algorithm that incorporates a Langevin-type description of stochastic reaction dynamics in a CPB. This description is applicable to the mesoscopic scale, in which the copy numbers of reacting molecules is on the order of thousands. More recently, Stamatakis and Zygourakis (2010) formulated a cell population master equation and a computational framework that take into account all the major sources of heterogeneity, including random reaction occurrence and DNA duplication events. Reaction events are treated as discrete point processes, making the framework generally applicable to any scale, including the microscopic one, in which the copy numbers of molecules are low. Stamatakis and Zygourakis (2011) subsequently used this approach to model cell population heterogeneity in an artificial *lac* operon genetic network.

Though it is desirable to have a comprehensive CPB model that includes all sources of heterogeneity, the computational expense involved in the solution thereof becomes prohibitively high for multi-variable models. To tackle this challenge, methods such as proper orthogonal

decomposition have been used by Zhang et al. (2003) to construct reduced nonlinear models that capture the dynamics of full CPB simulations. Moreover, hierarchical two-tier approaches that decompose fast and slow processes have been proposed (Immanuel and Doyle III, 2003; Pinto et al., 2007). Moving-boundary approaches have also been developed (Kavousanakis et al., 2009) to overcome the issue that the state space boundaries in which the cell population distribution is confined are not known *a priori*. As an alternative to solving the actual CPB equations, computational methods that employ kinetic Monte Carlo (KMC) simulation have also been developed (Smith and Matsoukas, 1998; Mantzaris, 2006).

Finally, Stamatakis (2010) investigated a set of assumptions that allowed for the transformation of a CPB to hybrid models involving ensemble or continuum equations. These assumptions call for equal partitioning of the contents of the mother cell to the two daughters, an (intensive) reaction rate that does not depend on the volume of the cell, growth and division rates that depend on the volume of the cell but not on the intracellular contents, and a partitioning probability density function in which intensive variables do not appear (in essence the concentrations of the mother and daughter cell species are assumed to be the same). Under vegetative growth of bacterial populations these conditions are expected to hold true to a reasonably good approximation. The transformation of the CPB to hybrid models essentially unifies the aforementioned three modeling approaches (CPB, ensemble and continuum models) and can be the starting point in the development of novel and highly efficient computational methods for the solution of multi-variable CPBs.

In the current work, we compare a CPB and a hybrid model capturing population heterogeneity for a genetic network consisting of a single gene with feedback. We begin our investigations by considering two variants of a 4-species model, the first variant written for extensive quantities (species contents) and the second for intensive quantities (concentrations). We use singular perturbation analysis (SPA) to derive reduced 2-variable systems of equations for the cell volume and intracellular protein levels. We subsequently incorporate the single cell equations into two different cell population frameworks: (i) a 2-variable CPB equation, and (ii) a hybrid model consisting of an ensemble equation and a single-variable CPB. We finally compare the cell population distributions obtained by the two frameworks and analyze the accuracy and efficiency of the hybrid approach.

Model Development

Reaction Network and Single Cell Models

We investigate the dynamics of a single gene that exerts feedback on its own expression by employing a reaction network developed by Mantzaris (2006). According to this model, the protein expressed by the gene can dimerize and bind to the operator controlling the expression. The model incorporates protein production from the bare as well as the bound state of the operator. Thus, the reactions expressing the interactions for this genetic network are formulated as follows:

The bare operator can initiate the transcription of the protein Y:

$$\mathrm{O} \xrightarrow{k_0} \mathrm{O} + \mathrm{Y} \tag{1}$$

Two protein molecules can associate, giving rise to dimer species Z:

$$\mathrm{Y}+\mathrm{Y} \underset{\beta\cdot k_d}{\overset{k_d}{\rightleftarrows}} \mathrm{Z} \tag{2}$$

The dimer can subsequently bind to the operator:

$$\mathrm{O}+\mathrm{Z} \underset{\alpha\cdot k_r}{\overset{k_r}{\rightleftarrows}} \mathrm{OZ} \tag{3}$$

The bound state can initiate transcription of Y with a different rate:

$$\mathrm{OZ} \xrightarrow{k_1} \mathrm{OZ}+\mathrm{Y} \tag{4}$$

Note that for $k_1 > k_0$ the feedback is positive since the production and dimerization of Y facilitates further production thereof; in other words the network exhibits autocatalytic architecture. For $k_1 < k_0$ the protein acts as a repressor, thereby exerting negative feedback on its own production. We assume that the monomeric form of the protein is subject to degradation whereas the dimer is a stable species; hence, we only take into account the degradation of Y.

$$\mathrm{Y} \xrightarrow{\lambda} \varnothing \tag{5}$$

We further assume the total DNA concentration to be constant:

$$[O]_T \cdot V = o + oz \tag{6}$$

This assumption means that operator DNA has to be produced continuously as the cell volume increases during growth. This is a major assumption of the deterministic model, since, in reality, DNA is produced at a specific point of the cell cycle. The issue stems from the very nature of the deterministic model itself, which is valid in the thermodynamic limit of large numbers of molecules.

Consequently, the *extensive* model (pertaining to molecular contents) with the single cell growth model is written as:

$$\frac{\mathrm{d}o}{\mathrm{dt}} = -\frac{k_r}{V}\cdot o\cdot z + \alpha\cdot k_r\cdot oz + g\cdot o \tag{7}$$

$$\frac{\mathrm{d}y}{\mathrm{dt}} = k_0\cdot o + k_1\cdot oz - 2\cdot\frac{k_d}{V}\cdot y^2 + 2\cdot\beta\cdot k_d\cdot z - \lambda\cdot y \tag{8}$$

$$\frac{\mathrm{d}z}{\mathrm{dt}} = \frac{k_d}{V}\cdot y^2 - \beta\cdot k_d\cdot z - \frac{k_r}{V}\cdot o\cdot z + \alpha\cdot k_r\cdot oz \tag{9}$$

$$\frac{\mathrm{d}V}{\mathrm{dt}} = g\cdot V \tag{10}$$

The last equation expresses volume increase due to cell growth and is only valid up to the point that the cell divides. A reduced version of this extensive model will be later incorporated into the

CPB equation. On the other hand, a hybrid model utilizes equations written for intensive quantities, namely concentrations (7-9) (this is a key point for the derivation of the hybrid model, as discussed in Stamatakis, 2010). Thus, the intensive model accounting for the dilution due to cell growth will be written as:

$$\frac{\mathrm{d}[O]}{\mathrm{dt}} = -k_r \cdot [O] \cdot [Z] + \alpha \cdot k_r \cdot [OZ] \tag{11}$$

$$\frac{\mathrm{d}[Y]}{\mathrm{dt}} = k_0 \cdot [O] + k_1 \cdot [OZ] - 2 \cdot k_d \cdot [Y]^2 + 2 \cdot \beta \cdot k_d \cdot [Z] - \lambda \cdot [Y] - g \cdot [Y] \tag{12}$$

$$\frac{\mathrm{d}[Z]}{\mathrm{dt}} = k_d \cdot [Y]^2 - \beta \cdot k_d \cdot [Z] - k_r \cdot [O] \cdot [Z] + \alpha \cdot k_r \cdot [OZ] - g \cdot [Z] \tag{13}$$

$$\frac{\mathrm{d}V}{\mathrm{dt}} = g \cdot V \tag{14}$$

In the following, we will use SPA to reduce these models, thereby deriving one equation for the intracellular contents and a second equation for the cell volume. This single cell model will subsequently be incorporated into the CPB framework.

Reduction of Single Cell Models

The reduction procedure we are about to apply is based on SPA using the techniques established by Schauer and Heinrich (1983). As part of this methodology, we introduce asymptotic expansions for the variables expressing the contents (extensive) or concentrations (intensive) of the participating species. By assuming that certain reactions in our system are fast, we derive a set of quasi-equilibrium equations for the species participating therein ($\mathcal{O}\left(\varepsilon^{-1}\right)$ equations). Subsequently, we write a set of equations expressing the slow dynamics ($\mathcal{O}(1)$ equations). By taking linear combinations of these equations and making use of the quasi-equilibrium conditions, we end up with a closed system of equations for the content (or concentration) of a single species and the volume of the cell. The linear combinations just noted essentially express a coordinate transformation, in line with equations (22)-(25) of Schauer and Heinrich (1983). A slightly different methodology also invoking SPA, but without relying on asymptotic expansions, has been developed by Gerdtzen et al. (2004). In this methodology, a set of algebraic quasi-equilibrium conditions is derived for the fast timescale, which is coupled with a set of ordinary differential equations (ODEs) for the slow dynamics, thereby obtaining a system of differential algebraic equations (DAEs). By differentiating the algebraic equations, the DAE system is converted into an ODE system, and the latter is subsequently reduced to a lower-order description by means of a coordinate transformation. In both methodologies, the transformation is based on identifying the nullspace of the stoichiometric matrix of the fast reactions.

Extensive 2-Variable Model

To construct the reduced models we assume that the kinetics of protein dimerization and operator-protein association are fast, and therefore reactions (2) and (3) are in quasi-equilibrium. Thus, we make the following scaling assumptions:

$$k_d = \frac{\kappa_d}{\varepsilon} \tag{15}$$

$$k_r = \frac{\kappa_r}{\varepsilon} \tag{16}$$

The first step in our analysis is to introduce the following expansions:

$$y = y^{(0)} + \varepsilon \cdot y^{(1)} + \varepsilon^2 \cdot y^{(2)} + \ldots \tag{17}$$

$$o = o^{(0)} + \varepsilon \cdot o^{(1)} + \varepsilon^2 \cdot o^{(2)} + \ldots \tag{18}$$

$$z = z^{(0)} + \varepsilon \cdot z^{(1)} + \varepsilon^2 \cdot z^{(2)} + \ldots \tag{19}$$

The equation for V does not contain ε so we do not have to introduce an expansion for V. Using the standard perturbation procedure we separate terms of the same order with respect to ε and we obtain the following expressions:

$$\mathcal{O}\left(\varepsilon^{-1}\right)$$

From equations (6), (7) and (15-18):

$$o^{(0)} = \frac{\alpha \cdot [O]_T \cdot V^2}{z^{(0)} + \alpha \cdot V} \tag{20}$$

Moreover, from equations (6), (9), (15-17) and (19):

$$z^{(0)} = \frac{1}{\beta \cdot V} \cdot \left(y^{(0)}\right)^2 \tag{21}$$

Thus, from the above two equations we have an expression for the operator contents with respect to the protein contents and the volume:

$$o^{(0)} = \frac{\alpha \cdot \beta \cdot [O]_T \cdot V^3}{\left(y^{(0)}\right)^2 + \alpha \cdot \beta \cdot V^2} \tag{22}$$

Furthermore:

$$\mathcal{O}(1)$$

From equations (7-9), and (15-19):

$$\frac{\mathrm{d}\,o^{(0)}}{\mathrm{dt}} = -\frac{\kappa_r}{V}\cdot\left(o^{(0)}\cdot z^{(1)} + o^{(1)}\cdot z^{(0)}\right) - \alpha\cdot\kappa_r\cdot o^{(1)} + g\cdot o^{(0)} \tag{23}$$

$$\frac{\mathrm{d}\,y^{(0)}}{\mathrm{dt}} = k_0\cdot o^{(0)} + k_1\cdot\left([O]_T\cdot V - o^{(0)}\right) - 2\cdot\frac{\kappa_d}{V}\cdot 2\cdot y^{(0)}\cdot y^{(1)} + 2\cdot\beta\cdot\kappa_d\cdot z^{(1)} - \lambda\cdot y^{(0)} \tag{24}$$

$$\frac{\mathrm{d}\,z^{(0)}}{\mathrm{dt}} = \frac{\kappa_d}{V}\cdot 2\cdot y^{(0)}\cdot y^{(1)} - \beta\cdot\kappa_d\cdot z^{(1)} - \frac{\kappa_r}{V}\cdot\left(o^{(0)}\cdot z^{(1)} + o^{(1)}\cdot z^{(0)}\right) - \alpha\cdot\kappa_r\cdot o^{(1)} \tag{25}$$

We need to eliminate the ephemeral (fast) species O and Z. In the resulting equation, terms containing only $y^{(0)}$ must appear. Thus, by subtracting (23) from (25):

$$\frac{\mathrm{d}\,z^{(0)}}{\mathrm{dt}} - \frac{\mathrm{d}\,o^{(0)}}{\mathrm{dt}} = \frac{\kappa_d}{V}\cdot 2\cdot y^{(0)}\cdot y^{(1)} - \beta\cdot\kappa_d\cdot z^{(1)} - g\cdot o^{(0)} \tag{26}$$

The $y^{(1)}$ and $z^{(1)}$ terms of the above equation can be eliminated with the use of equation (24). Thus, from (24) + 2·(26):

$$\frac{\mathrm{d}\,y^{(0)}}{\mathrm{dt}} + 2\cdot\frac{\mathrm{d}\,z^{(0)}}{\mathrm{dt}} - 2\cdot\frac{\mathrm{d}\,o^{(0)}}{\mathrm{dt}} = k_0\cdot o^{(0)} + k_1\cdot\left([O]_T\cdot V - o^{(0)}\right) - 2\cdot g\cdot o^{(0)} - \lambda\cdot y^{(0)} \tag{27}$$

In the above equation, $o^{(0)}$ can be substituted by an expression of $y^{(0)}$ and V (equation 22). Furthermore, the derivatives $\mathrm{d}o^{(0)}/\mathrm{dt}$ and $\mathrm{d}z^{(0)}/\mathrm{dt}$ can be calculated in terms of $\mathrm{d}y^{(0)}/\mathrm{dt}$ from (20) and (21):

$$\frac{\mathrm{d}\,o^{(0)}}{\mathrm{dt}} = -\frac{\alpha\cdot\beta\cdot[O]_T\cdot V^3}{\left(\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2\right)^2}\cdot 2\cdot y^{(0)}\cdot\left(\frac{\mathrm{d}\,y^{(0)}}{\mathrm{dt}} + \left(3\cdot\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2\right)\cdot g\right) \tag{28}$$

$$\frac{\mathrm{d}\,z^{(0)}}{\mathrm{dt}} = \frac{2\cdot y^{(0)}}{\beta\cdot V}\cdot\left(\frac{\mathrm{d}\,y^{0}}{\mathrm{dt}} - \frac{y^{(0)}}{2}\cdot g\right) \tag{29}$$

To simplify notation, let us define the following functions:

$$\Phi\left(y^{(0)}\right) = 1 + 2\cdot\frac{2\cdot y^{(0)}}{\beta\cdot V} + 2\cdot\frac{\alpha\cdot\beta\cdot[O]_T\cdot V^3}{\left(\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2\right)^2}\cdot 2\cdot y^{(0)} \tag{30}$$

$$\Theta\left(y^{(0)}\right) = -2\cdot g\cdot\frac{\left(y^{(0)}\right)^2}{\beta\cdot V} - 2\cdot\alpha\cdot\beta\cdot[O]_T\cdot V^2\cdot\frac{3\cdot\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2}{\left(\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2\right)^2}\cdot g\cdot V \tag{31}$$

Subsequently, the reduced extensive model takes the form (using equations 27, 30):

$$\Phi\left(y^{(0)}\right)\cdot\frac{\mathrm{d}\,y^{(0)}}{\mathrm{dt}}+\Theta\left(y^{(0)}\right) = \left(k_0-k_1-2\cdot g\right)\cdot\frac{\alpha\cdot\beta\cdot[O]_T\cdot V^3}{\left(y^{(0)}\right)^2+\alpha\cdot\beta\cdot V^2}+k_1\cdot[O]_T\cdot V-\lambda\cdot y^{(0)} \tag{32}$$

Intensive 2-Variable Model

The procedure just used can also be applied to derive the reduced intensive model from equations (11-14). We assume that (15, 16) hold. Let us then introduce the expansions:

$$[Y] = [Y]^{(0)}+\varepsilon\cdot[Y]^{(1)}+\varepsilon^2\cdot[Y]^{(2)}+\ldots \tag{33}$$

$$[O] = [O]^{(0)}+\varepsilon\cdot[O]^{(1)}+\varepsilon^2\cdot[O]^{(2)}+\ldots \tag{34}$$

$$[Z] = [Z]^{(0)}+\varepsilon\cdot[Z]^{(1)}+\varepsilon^2\cdot[Z]^{(2)}+\ldots \tag{35}$$

Therefore, the equations we obtain for the different orders of ε are as follows:

$\mathcal{O}\left(\varepsilon^{-1}\right)$

From equations (6), (7), (34), (35) we derive the expression for the operator concentration in quasi-equilibrium:

$$[O]^{(0)} = \frac{\alpha\cdot[O]_T}{[Z]^{(0)}+\alpha} \tag{36}$$

Moreover, from equations (6), (9), (33), (35), (36) the dimer concentration is expressed as:

$$[Z]^{(0)} = \frac{1}{\beta}\cdot\left([Y]^{(0)}\right)^2 \tag{37}$$

The equations on the zeroth order of ε are as follows:

$\mathcal{O}(1)$

$$\frac{\mathrm{d}[O]^{(0)}}{\mathrm{dt}} = -\kappa_r\cdot\left([O]^{(0)}\cdot[Z]^{(1)}+[O]^{(1)}\cdot[Z]^{(0)}\right)-\alpha\cdot\kappa_r\cdot[O]^{(1)} \tag{38}$$

$$\begin{aligned}\frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} = {} & k_0\cdot[O]^{(0)}+k_1\cdot\left([O]_T-[O]^{(0)}\right)-2\cdot\kappa_d\cdot 2\cdot[Y]^{(0)}\cdot[Y]^{(1)}+2\cdot\beta\cdot\kappa_d\cdot[Z]^{(1)} \\ & -\lambda\cdot[Y]^{(0)}-g\cdot[Y]^{(0)}\end{aligned} \tag{39}$$

$$\frac{\mathrm{d}[Z]^{(0)}}{\mathrm{dt}} = \kappa_d \cdot 2 \cdot [Y]^{(0)} \cdot [Y]^{(1)} - \beta \cdot \kappa_d \cdot [Z]^{(1)} - \kappa_r \cdot \left([O]^{(0)} \cdot [Z]^{(1)} + [O]^{(1)} \cdot [Z]^{(0)}\right) - \alpha \cdot \kappa_r \cdot [O]^{(1)} - g \cdot [Z]^{(0)} \tag{40}$$

In a similar way as previously done for the reduction of the extensive model, we take the difference between equations (40) − (38):

$$\frac{\mathrm{d}[Z]^{(0)}}{\mathrm{dt}} - \frac{\mathrm{d}[O]^{(0)}}{\mathrm{dt}} = \kappa_d \cdot 2 \cdot [Y]^{(0)} \cdot [Y]^{(1)} - \beta \cdot \kappa_d \cdot [Z]^{(1)} - g \cdot [Z]^{(0)} \tag{41}$$

Moreover, from (39) + 2·(41):

$$\frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} + 2 \cdot \frac{\mathrm{d}[Z]^{(0)}}{\mathrm{dt}} - 2 \cdot \frac{\mathrm{d}[O]^{(0)}}{\mathrm{dt}} = k_0 \cdot [O]^{(0)} + k_1 \cdot \left([O]_T - [O]^{(0)}\right) - \lambda \cdot [Y]^{(0)} - g \cdot [Y]^{(0)} - 2 \cdot g \cdot [Z]^{(0)} \tag{42}$$

Then, the derivatives of the concentrations $[O]^{(0)}$ and $[Z]^{(0)}$ are expressed with respect to $[Y]^{(0)}$ and the temporal derivative thereof:

$$\frac{\mathrm{d}[O]^{(0)}}{\mathrm{dt}} = -\frac{\alpha \cdot \beta \cdot [O]_T}{\left(\left([Y]^{(0)}\right)^2 + \alpha \cdot \beta\right)^2} \cdot 2 \cdot [Y]^{(0)} \cdot \frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} \tag{43}$$

$$\frac{\mathrm{d}[Z]^{(0)}}{\mathrm{dt}} = \frac{2 \cdot [Y]^{(0)}}{\beta} \cdot \frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} \tag{44}$$

By defining:

$$\Psi\left([Y]^0\right) = 1 + 2 \cdot \frac{2 \cdot [Y]^{(0)}}{\beta} + 2 \cdot \frac{\alpha \cdot \beta \cdot [O]_T}{\left(\left([Y]^{(0)}\right)^2 + \alpha \cdot \beta\right)^2} \cdot 2 \cdot [Y]^{(0)} \tag{45}$$

the reduced intensive model is written as:

$$\Psi\left([Y]^{(0)}\right) \cdot \frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} = (k_0 - k_1) \cdot \frac{\alpha \cdot \beta \cdot [O]_T}{\left([Y]^{(0)}\right)^2 + \alpha \cdot \beta} + k_1 \cdot [O]_T - (\lambda + g) \cdot [Y]^{(0)} - \frac{2 \cdot g}{\beta} \cdot \left([Y]^{(0)}\right)^2 \tag{46}$$

The above equation is equivalent to the reduced extensive model (equation 32) as shown in the Appendix.

Note that instead of performing detailed SPA we could have casually set the derivatives of [*O*] and [Y] equal to zero in equations (11) and (13), thereby obtaining:

$$[O] = \frac{\alpha \cdot [O]_T}{[Z] + \alpha} \tag{47}$$

$$[Z] = \frac{1}{\left(\beta + \frac{g}{k_d}\right)} \cdot [Y]^2 \tag{48}$$

Equation (36) previously obtained through SPA is similar to (47). However, equation (37) is different than its counterpart (48), since the different orders of magnitude are not separated in the latter analysis. Finally, the intensive model would be derived from equation (12) as:

$$\frac{\mathrm{d}[Y]}{\mathrm{dt}} = (k_0 - k_1) \cdot \frac{\alpha \cdot \left(\beta + \frac{g}{k_d}\right) \cdot [O]_T}{[Y]^2 + \alpha \cdot \left(\beta + \frac{g}{k_d}\right)} + k_1 \cdot [O]_T - (\lambda + g) \cdot [Y] - \frac{2 \cdot g}{\left(\beta + \frac{g}{k_d}\right)} \cdot [Y]^2 \tag{49}$$

Note that the latter equation (49) does not have the time modulation term $\Psi\left([Y]^{(0)}\right)$ appearing in equation (46). In the results section we will compare and contrast the two reduced intensive models (46) and (49), demonstrating that SPA is essential for a faithful and accurate representation of the original dynamics.

Cell Population Balance and Hybrid Model

The population balance equation describes the temporal evolution of the distribution of cells over the possible states. The latter comprise the pairs (y, V) where y is the number of moles (or molecules) of Y and V the volume of the cell. The CPB is formulated for the normalized number density function $n(y,V,t)$: the quantity $n(y,V,t) \cdot dy \cdot dV$ represents the fraction of cells that exist between states $[y, y+dy) \times [V, V+dV)$. Thus, the population balance is written as (Stamatakis, 2010):

$$\begin{aligned} &\frac{\partial n(y,V,t)}{\partial t} + \frac{\partial}{\partial y}\left[r(y,V) \cdot n(y,V,t)\right] + \frac{\partial}{\partial V}\left[(g \cdot V) \cdot n(y,V,t)\right] + \Gamma(V) \cdot n(y,V,t) \\ &= 2 \cdot \int_0^\infty \Gamma(W) \cdot p(V \mid W) \cdot n(y,W,t)\, dW - n(y,V,t) \cdot \int_0^\infty \Gamma(W) \cdot n(x,W,t)\, dW \end{aligned} \tag{50}$$

where the advective terms correspond to intracellular reaction and cell growth, the last term on the left hand side describes loss of cells from state (y,V) due to division, the first term on the right hand side denotes influx to state (y,V) from the division of cells with higher volumes and intracellular contents, and the last term of the right hand side takes care of the normalization of $n(y,V,t)$ so that the latter always integrates to unity.

In the above equation (50) the reaction rate is obtained from the reduced extensive model:

$$r(y,V) = \frac{\mathrm{d}\,y}{\mathrm{dt}} = \frac{1}{\Phi(y)}\cdot\left[\left(k_0 - k_1 - 2\cdot g\right)\cdot\frac{\alpha\cdot\beta\cdot[O]_T\cdot V^3}{y^2+\alpha\cdot\beta\cdot V^2} + k_1\cdot[O]_T\cdot V - \lambda\cdot y - \Theta(y)\right] \tag{51}$$

and the division rate and partitioning probability density function are respectively:

$$\Gamma(V) = \left(\frac{V}{V_{div}}\right)^{n_{div}} \tag{52}$$

$$p(V\,|\,W) = \frac{1}{W}\cdot\frac{\Gamma(2\cdot q)}{\left(\Gamma(q)\right)^2}\cdot\left(\frac{V}{W}\right)^{q-1}\cdot\left(1-\frac{V}{W}\right)^{q-1} \tag{53}$$

where Γ denotes the gamma function and thus, p is a symmetric beta distribution.

Stamatakis (2010) showed that under certain assumptions, which are satisfied by equations (50-53) the CPB model is equivalent to the following hybrid model:

$$\begin{aligned}&\frac{\partial n_V(V,t)}{\partial t} + g\cdot\frac{\partial}{\partial V}\left[n_V(V,t)\right] + \Gamma(V)\cdot n(V,t)\\ &= 2\int_0^\infty \Gamma(W)\cdot p(V\,|\,W)\cdot n_V(W,t)\,dW - n_V(V,t)\cdot\int_0^\infty \Gamma(W)\cdot n_V(W,t)\,dW\end{aligned} \tag{54}$$

$$\frac{\partial n_C(C,t)}{\partial t} + \frac{\partial}{\partial C}\left[\hat{r}(C)\cdot n_C(V,t)\right] = 0 \tag{55}$$

where the concentration is $C = y / V$, function $\hat{r}$ denotes the intensive reaction rate:

$$\hat{r}(C) = \frac{1}{\Psi(C)}\cdot\left[\left(k_0 - k_1\right)\cdot\frac{\alpha\cdot\beta\cdot[O]_T}{C^2+\alpha\cdot\beta} + k_1\cdot[O]_T - (\lambda+g)\cdot C - \frac{2\cdot g}{\beta}\cdot C^2\right] \tag{56}$$

and the number density appearing in equation (50) is:

$$n(y,V,t) = n_V(V,t)\cdot n_C(C,t)\cdot\frac{1}{V} \tag{57}$$

provided that the initial condition $n(y,V,0)$ factorizes into a volume-dependent and a concentration-dependent part. This means that instead of simulating the full CPB (50) for y and V, one can simulate (i) the CPB for the volume (54) and (ii) an ensemble of realizations of the intensive model with random initial conditions, which would capture the dynamics of the advection equation (55).

Results

Validation and Characterization of the Reduced Intensive Models

When reducing the intensive model, we noted that instead of performing detailed SPA we can casually set the derivatives of the ephemeral species to zero, thereby deriving equation (49) for d[*Y*]/dt. A comparison of the two reduced intensive models (46) and (49) with the full model (equations 11-13) is shown in Figure 1a for the parameter set of Table 1. Evidently, the "casually derived" equation (49) does not correctly capture the timescale of the course toward the steady state. The model derived through SPA, however, (equation 46) exhibits a satisfactory agreement with the full model. More importantly, the results of the two latter models become identical for $\varepsilon \to 0$, as shown in Figure 1b, which portrays the error norm between the transients obtained by each of the approximate models and the full model. This error norm was computed as:

$$Error_{\substack{Reduced \\ Intensive}} = \sqrt{\int_{t_{ini}}^{t_{fin}} \left(y_{\substack{approx \\ model}}(t) - y_{\substack{full \\ model}}(t) \right)^2 dt} \tag{58}$$

Figure 1b shows that the error of the "casually derived" equation (49) stays fixed at high values, whereas the error of the singular perturbation-derived model drops linearly with ε as expected, since we carried out the analysis up to order 1. These results demonstrate that in order to perform a meaningful reduction that faithfully captures the kinetics of the original multi-dimensional model, one has to properly separate terms of the same order of magnitude through a perturbation-type analysis. Arbitrarily setting the derivatives of the fast (ephemeral) species equal to zero will yield a reduced model that exhibits the same steady states as the original model, but may have a completely different temporal behavior.

Having validated the SPA-derived model we proceed to investigate its behavior by computing the bifurcation diagram of equation (46), with respect to the dimer-operator dissociation constant α for the parameters of Table 1, for which the protein exerts positive feedback on its own production (Figure 2a). This parameter was chosen since it quantifies the strength of the feedback that protein Y exerts on its own production. Thus, Figure 2a portrays the steady state solutions of equation (46) with respect to this dissociation constant, and it is observed that the positive feedback generates a sigmoidal response: for strong dimer-operator association (low values of the dissociation constant α) the positive feedback exerted by the dimer results in the gene being expressed at a maximal rate. On the other hand, for weak dimer-operator association (high values of α) only leak expression is observed. For intermediate values of α, bistability is observed. In this region, two stable steady states coexist and a sufficiently large perturbation from the low state of leak protein expression can bring the system to the higher state of strong expression.

Contrary to what was just discussed, Figure 2b portrays the case where the protein acts as a repressor ($k_0 > k_1$ in reactions 1 and 4), thereby exerting a negative feedback on its own production. In this case, higher values of the dimer-operator dissociation constant (α) result in weaker repression, and thus, higher steady state concentrations of the protein Y. No bistability is observed in this case: the steady state concentration [*Y*] is a monotonically increasing function of the dissociation constant α.

Cell Population Simulations

For the interesting case in which the system exhibits bistability, we simulate the dynamics of a cell population, using the two frameworks discussed in section "Cell Population Balance and Hybrid Model". These simulations are performed for positive feedback architecture with a value of the dissociation constant α for which two stable steady states coexist. For the simulation of the full CPB model (equation 50) and the CPB part of the hybrid model (equation 54) KMC algorithms were used (Smith and Matsoukas, 1998; Mantzaris, 2006; Stamatakis and Zygourakis, 2010; Stamatakis and Zygourakis, 2011).

We first discuss a simple calculation that demonstrates the principle of the hybrid methodology. Figure 3 shows the trajectories of each cell in the population, obtained by a KMC simulation thereof (the stochastic analogue of equation 50) for the parameter values of Table 1. The simulation is initialized with two cells, having molecular contents and volumes such that one's concentration lies above the unstable branch of the bifurcation diagram (Figure 2) and the other's lies below. Figure 3a shows the evolution of the contents, revealing that all offspring of the first cell approach a high state and all offspring of the second approach a low state (discontinuities correspond to division events). There is a significant amount of heterogeneity and the trajectories seem rather random; however, after calculating the concentrations of the species of interest (Figure 3b), one makes the startling observation that the concentrations follow simple trends. In fact, the graph shown in Figure 3b, could have been calculated exactly just by solving the reduced intensive model (46).

Now let us treat a more complicated case, in which we initiate the simulation with a large population of cells having a distribution of sizes and protein contents. For this simulation, we use the following initial condition, which factorizes with respect to volume and concentration:

$$n(y,V,0) = \mathcal{G}_{\mu_V,\sigma_V}(V) \cdot \mathcal{G}_{\mu_C,\sigma_C}\left(\frac{y}{V}\right) \cdot \frac{1}{V} \tag{59}$$

where $\mathcal{G}_{\mu,\sigma}$ is a Gaussian distribution with mean μ and standard deviation σ, truncated within the interval $[0, \infty)$. The functional form of this distribution is:

$$\mathcal{G}_{\mu,\sigma}(t) = \frac{G_{\mu,\sigma}(t)}{\int_0^\infty G_{\mu,\sigma}(t)\,dt} \cdot H(t) \tag{60}$$

where $G_{\mu,\sigma}$ is the Gaussian distribution and H the Heaviside function.

Snapshots of the evolution of the normalized number density are shown in Figure 4. The first column of plots corresponds to simulating the hybrid system that consists of the ensemble model for the concentrations and the CPB for the volume (equations 54-57). The second column corresponds to the cell population Monte Carlo simulation where a maximum number of 100000 cells were used in a constant number simulation scheme, (see Smith and Matsoukas, 1998). The perfect agreement between the two simulations demonstrates computationally the equivalence of the two models, as theoretically predicted by Stamatakis (2010).

It is interesting to observe the dynamical behavior of the number density as revealed by the simulation. For the parameter set of this simulation, the intensive single cell model exhibits two stable steady states and one unstable. The dashed line in each of the panels of Figure 4 corresponds to a separatrix created by the unstable steady state which separates the two stable ones. Thus, the evolution of the initial distribution is a composition of two components. The first component is dictated by the growth and division dynamics, as shown in panel (b), in which some cells have already divided and some are still growing, thereby generating two distinct modes in the number density. The second component is dictated by the reaction dynamics which tend to drive the number density away from the separatrix into two elongated delta peaks as shown in panels (e, g). Note that finally the number density reaches a stationary distribution of the form:

$$n(y,V,\infty)=n_V(V,\infty)\cdot\left(\omega\cdot\delta\left(\frac{y}{V}-C_{low}\right)+(1-\omega)\cdot\delta\left(\frac{y}{V}-C_{high}\right)\right) \tag{61}$$

where ω is the integral of the initial number density over the region below the separatrix, and C_{low} and C_{high} are the low and high steady states of the intensive model for the concentration (see Figure 2).

Accuracy and Computational Efficiency of Hybrid Model

To demonstrate the accuracy and efficiency of the hybrid model, we compare the results obtained by explicitly simulating the full CPB (equation 50) with those of the hybrid model (equations 54-57) for the parameters and initial conditions of Figure 4. For solving the CPB equations (50) and (54) the constant-number KMC framework was used (Smith and Matsoukas, 1998; Mantzaris, 2006; Stamatakis and Zygourakis, 2010; Stamatakis and Zygourakis, 2011). Moreover, the solution of the ensemble equation (55) was evaluated by simulating the dynamics of each one of the cells of an ensemble. For the hybrid model, a highly accurate solution-of-reference was obtained by solving each of the components of the model (equations 54 and 55) for $8\cdot10^5$ cells. Subsequently, simulations of the full CPB model were carried out for variable population sizes, in order to perform error analysis and evaluate the computational efficiency of the hybrid model. The error was calculated as follows:

$$Error=\sum_{i=0}^{N_t}\sqrt{\int\int\left(n_{FullCPB}(y,V,i\cdot\Delta t)-n_{accurate}(y,V,i\cdot\Delta t)\right)^2 dy\,dV} \tag{62}$$

The results of these studies are shown in Figure 5a, in which the error of the full CPB model is plotted against the population size. The observed trend is in agreement with the inverse square root law, which is typically encountered in Monte Carlo algorithms. In addition, Figure 5b shows the ratio between the computational times of the full CPB model for the various population sizes, versus the time required to obtain the most accurate solution from the hybrid model. Note that it takes about the same time to solve the hybrid model for $N_{population}$ = $8\cdot10^5$, and the full CPB model for $5\cdot10^4$ cells. The error of the latter, however is still high (above unity), as highlighted in Figure 5a. By extrapolating the power-law fits of Figure 5 we deduce that a simulation of the full

CPB model with $N_{population} = 8 \cdot 10^5$, would take about 100 times more CPU time than the hybrid model, and would reduce the error to approximately 0.3. These results underscore the high accuracy and computational efficiency of the hybrid modeling methodology introduced by Stamatakis (2010).

Conclusions

The ever-increasing interest in investigating cell population heterogeneity necessitates the development of novel and efficient approaches for simulating the dynamics of entire cell populations. CPB models are among the most comprehensive and accurate approaches for this task. However, they can quickly become intractable if multiple species participate in several reactions that need to be explicitly taken into account.

In this work, we demonstrated how one can use reduction techniques to efficiently simulate cell population heterogeneity. Thus, we showed how singular perturbation can be used to reduce the dimensionality of a system of equations pertaining to multiple species, thereby deriving a single cell model that faithfully captures the original dynamics. We contrasted this detailed approach with an *ad hoc* technique which entails setting the temporal derivatives of the fast species to zero. This latter approach was shown to incur large errors in the prediction of the dynamics of the system.

We further showed how a hybrid model can be used to simulate the dynamics of a cell population. This class of models was introduced in previous work (Stamatakis, 2010), in which it was shown that under certain biologically plausible assumptions, the behavior of the hybrid models are identical to those of the full CPB. This theoretical prediction was confirmed computationally for the single gene with feedback, for which the temporal behavior of the hybrid model was found to be identical to that of the CPB. In addition, by analyzing the error and computational times of the two models, the high accuracy and efficiency of the hybrid model was demonstrated.

Even in the case that the assumptions discussed in (Stamatakis, 2010) do not hold true, hybrid models can still be used to obtain approximate solutions of the full CPB, rather than the exact solution thereof. Such investigations open opportunities for further research on improving the accuracy and expanding the applicability of hybrid models in simulating cell population heterogeneity.

Acknowledgements

MS would like to thank Prof. Kyriacos Zygourakis for providing access to the computational facilities at Rice University. This work was supported in part by the Shared University Grid at Rice funded by NSF under Grant EIA-0216467, and a partnership between Rice University, Sun Microsystems, and Sigma Solutions, Inc.

Appendix

Equivalence of Reduced Intensive and Extensive Single Cell Models

To validate the reductions we have to check whether the models are identical under the transformation $y^{(0)} = V\cdot[Y]^{(0)}$. To this end, we will subtract equation (46) from (32)/V. First, note that:

$$y^{(0)} = [Y]^{(0)} \cdot V \tag{63}$$

and thus:

$$\Phi\left(y^{(0)}\right) = \Psi\left([Y]^{0}\right) \tag{64}$$

Furthermore:

$$\Phi\left(y^{(0)}\right)\cdot\frac{\mathrm{d}\,y^{(0)}}{\mathrm{dt}} + \Theta\left(y^{(0)}\right) = \Phi\left(y^{(0)}\right)\cdot\left[\frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}}\cdot V + g\cdot V\cdot[Y]^{(0)}\right] + \Theta\left(y^{(0)}\right) \tag{65}$$

Thus, equation (46) subtracted from (32)/V and with the use of equations (64, 65) becomes:

$$\begin{aligned}
&\Psi\left([Y]^{(0)}\right)\cdot\frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} - \Phi\left(y^{(0)}\right)\cdot\left[\frac{\mathrm{d}[Y]^{(0)}}{\mathrm{dt}} + g\cdot[Y]^{(0)}\right] - \frac{1}{V}\cdot\Theta\left(y^{(0)}\right) = \\
&\quad (k_0 - k_1)\cdot\frac{\alpha\cdot\beta\cdot[O]_T}{\left([Y]^{(0)}\right)^2 + \alpha\cdot\beta} + k_1\cdot[O]_T - (\lambda + g)\cdot[Y]^{(0)} - \frac{2\cdot g}{\beta}\cdot\left([Y]^{(0)}\right)^2 + \\
&\quad -(k_0 - k_1 - 2\cdot g)\cdot\frac{\alpha\cdot\beta\cdot[O]_T\cdot V^2}{\left(y^{(0)}\right)^2 + \alpha\cdot\beta\cdot V^2} - k_1\cdot[O]_T + \frac{\lambda}{V}\cdot y^{(0)}
\end{aligned} \tag{66}$$

The above equation simplifies to:

$$-\Phi\left(y^{(0)}\right)\cdot g\cdot[Y]^{(0)} - \frac{1}{V}\cdot\Theta\left(y^{(0)}\right) = 2\cdot g\cdot\frac{\alpha\cdot\beta\cdot[O]_T}{\left([Y]^{(0)}\right)^2 + \alpha\cdot\beta} - g\cdot[Y]^{(0)} - \frac{2\cdot g}{\beta}\cdot\left([Y]^{(0)}\right)^2 \tag{67}$$

From the definitions of Θ and Ψ (equations 30, 31) it is easy to show that the above equation is an identity. Thus, the intensive and extensive reduced models are equivalent.

References

Anderson, E. C., G. I. Bell, D. F. Petersen and R. A. Tobey. Cell Growth and Division: IV. Determination of Volume Growth Rate and Division Probability. Biophysical Journal 1969; 9(2): 246-263.

Ataai, M. M. and M. L. Shuler. Simulation of CFSTR through development of a mathematical model for anaerobic growth of *Escherichia coli* cell population. Biotechnology and Bioengineering 1985; 27(7): 1051-1055.

Avery, S. V. Microbial cell individuality and the underlying sources of heterogeneity. Nature Reviews Microbiology 2006; 4: 577-587.

Bell, G. I. Cell Growth and Division: III. Conditions for Balanced Exponential Growth in a Mathematical Model. Biophysical Journal 1968; 8(4): 431-444.

Bell, G. I. and E. C. Anderson. Cell Growth and Division: I. A Mathematical Model with Applications to Cell Volume Distributions in Mammalian Suspension Cultures. Biophysical Journal 1967; 7(4): 329-351.

Davidson, C. J. and M. G. Surette. Individuality in Bacteria. Annual Review of Genetics 2008; 42: 253-268.

Fadda, S., A. Cincotti and G. Cao. A novel population balance model to investigate the kinetics of in vitro cell proliferation: Part I. model development. Biotechnology and Bioengineering 2012a; 109(3): 772-781.

Fadda, S., A. Cincotti and G. Cao. A novel population balance model to investigate the kinetics of in vitro cell proliferation: Part II. numerical solution, parameters' determination, and model outcomes. Biotechnology and Bioengineering 2012b; 109(3): 782-796.

Faraday, D. B. F. and N. F. Kirkby. The Application of Population Balance Techniques to the Modeling of the Cell-Cycle. Chemical Engineering Research & Design 1992; 70(2): 174-176.

Fernandes, R. L., M. Nierychlo, L. Lundin, A. E. Pedersen, P. E. P. Tellez, A. Dutta, M. Carlquist, A. Bolic, D. Schapper, A. C. Brunetti, S. Helmark, A. L. Heins, A. D. Jensen, I. Nopens, K. Rottwitt, N. Szita, J. D. van Elsas, P. H. Nielsen, J. Martinussen, S. J. Sorensen, A. E. Lantz and K. V. Gernaey. Experimental methods and modeling techniques for description of cell population heterogeneity. Biotechnology Advances 2011; 29(6): 575-599.

Fredrickson, A. G. Population balance equations for cell and microbial cultures revisited. AIChE Journal 2003; 49(4): 1050-1059.

Fredrickson, A. G. and N. V. Mantzaris. A new set of population balance equations for microbial and cell cultures. Chemical Engineering Science 2002; 57(12): 2265-2278.

Fredrickson, A. G., D. Ramkrishna and H. M. Tsuchyia. Statistics and dynamics of prokaryotic cell populations. Mathematical Biosciences 1967; 1(3): 327-374.

Gerdtzen, Z. P., P. Daoutidis and W. S. Hu. Non-linear reduction for kinetic models of metabolic reaction networks. Metabolic Engineering 2004; 6(2): 140-154.

Godin, F. B., D. G. Cooper and A. D. Rey. Development and solution of a cell mass population balance model applied to the SCF process. Chemical Engineering Science 1999; 54(5): 565-578.

Henson, M. A., D. Müller and M. Reuss. Cell population modelling of yeast glycolytic oscillations. Biochemical Journal 2002; 368(2): 433-446.

Immanuel, C. D. and F. J. Doyle III. Computationally efficient solution of population balance

models incorporating nucleation, growth and coagulation: application to emulsion polymerization. Chemical Engineering Science 2003; 58(16): 3681-3698.

Kavousanakis, M. E., N. V. Mantzaris and A. G. Boudouvis. A novel free boundary algorithm for the solution of cell population balance models. Chemical Engineering Science 2009; 64(20): 4247-4261.

Kolewe, M. E., S. C. Roberts and M. A. Henson. A population balance equation model of aggregation dynamics in Taxus suspension cell cultures. Biotechnology and Bioengineering 2012; 109(2): 472-482.

Liu, Y. H., J. X. Bi, A. P. Zeng and J. Q. Yuan. A population balance model describing the cell cycle dynamics of myeloma cell cultivation. Biotechnology Progress 2007; 23(5): 1198-1209.

Mantzaris, N. V. A cell population balance model describing positive feedback loop expression dynamics. Computers & Chemical Engineering 2005a; 29(4): 897-909.

Mantzaris, N. V. Single-cell gene-switching networks and heterogeneous cell population phenotypes. Computers & Chemical Engineering 2005b; 29(3): 631-643.

Mantzaris, N. V. Stochastic and deterministic simulations of heterogeneous cell population dynamics. Journal of Theoretical Biology 2006; 241(3): 690-706.

Mantzaris, N. V. From Single-Cell Genetic Architecture to Cell Population Dynamics: Quantitatively Decomposing the Effects of Different Population Heterogeneity Sources for a Genetic Network with Positive Feedback Architecture. Biophysical Journal 2007; 92: 4271-4288.

Mantzaris, N. V. and P. Daoutidis. Cell population balance modeling and control in continuous bioreactors. Journal of Process Control 2004; 14(7): 775-784.

Mantzaris, N. V., P. Daoutidis and F. Srienc. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. Computers & Chemical Engineering 2001a; 25: 1411-1440.

Mantzaris, N. V., P. Daoutidis and F. Srienc. Numerical solution of multi-variable cell population balance models: II. Spectral methods. Computers & Chemical Engineering 2001b; 25: 1441-1462.

Mantzaris, N. V., P. Daoutidis and F. Srienc. Numerical solution of multi-variable cell population balance models: III. Finite element methods. Computers & Chemical Engineering 2001c; 25: 1463-1481.

Mantzaris, N. V., F. Srienc and P. Daoutidis. Control of cell mass distribution in continuous bioreactors using population balance models. Dynamics & Control of Process Systems 1998, Volumes 1 and 2 1999: 571-576.

Pinto, M. A., C. D. Immanuel and F. J. Doyle. A feasible solution technique for higher-dimensional population balance models. Computers & Chemical Engineering 2007; 31(10): 1242-1256.

Schauer, M. and R. Heinrich. Quasi-Steady-State Approximation in the Mathematical-Modeling of Biochemical Reaction Networks. Mathematical Biosciences 1983; 65(2): 155-170.

Sharifian, M. and M. A. Fanaei. Dynamic Simulation and Control of a Continuous Bioreactor Based on Cell Population Balance Model. Iranian Journal of Chemistry & Chemical Engineering-International English Edition 2009; 28(2): 15-27.

Sinko, J. W. and W. Streifer. A New Model for Age-Size Structure of a Population. Ecology 1967; 48(6): 910-918.

Sinko, J. W. and W. Streifer. Applying Models Incorporating Age-Size Structure of a Population

to Daphnia. Ecology 1969; 50(4): 608-615.
Sinko, J. W. and W. Streifer. A Model for Populations Reproducing by Fission. Ecology 1971; 52(2): 330-335.
Smith, M. and T. Matsoukas. Constant-number Monte Carlo simulation of population balances. Chemical Engineering Science 1998; 53(9): 1777-1786.
Stamatakis, M. Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches. Chemical Engineering Science 2010; 65(2): 1008-1015.
Stamatakis, M. and K. Zygourakis. A Mathematical and Computational Approach for Integrating the Major Sources of Cell Population Heterogeneity. Journal of Theoretical Biology 2010; 266(1): 41-61.
Stamatakis, M. and K. Zygourakis. Deterministic and Stochastic Population Level Simulations of an Artificial *lac* Operon Genetic Network. BMC Bioinformatics 2011; 12: 301.
Subramanian, G. and D. Ramkrishna. On the solution of statistical models of cell populations. Mathematical Biosciences 1971; 10(1-2): 1-23
Tsuchiya, H. M., A. G. Fredrickson and R. Aris. Dynamics of microbial cell populations. Advances in Chemical Engineering 1966; 6: 125-206.
Zhang, Y. C., M. A. Henson and Y. G. Kevrekidis. Nonlinear model reduction for dynamic analysis of cell population models. Chemical Engineering Science 2003; 58(2): 429-445.
Zhu, G. Y., A. Zamamiri, M. A. Henson and M. A. Hjortso. Model predictive control of continuous yeast bioreactors using cell population balance models. Chemical Engineering Science 2000; 55(24): 6155-6167.

Tables

Table 1: Parameter values for the reduced single cell models and the cell population models

Parameter	Value	Units	Description
α	200	nM	Dimer-operator dissociation constant
β	0.8	–	Dimerization constant
k_0	0.1	min^{-1}	Protein expression kinetic constant from bare operator
k_1	16	min^{-1}	Protein expression kinetic constant from bound operator
λ	0.5	min^{-1}	Protein degradation kinetic constant
$[O]_T$	2	nM	Total operator concentration
g	0.0231	min^{-1}	Cell growth rate constant
V_{div}	$1.1 \cdot 10^{-15}$	L	Parameter quantifying critical cell size for division
n_{div}	25	–	Parameter quantifying sharpness of division
q	80	–	Parameter quantifying symmetry of cell fission

Figure Captions

Figure 1: Panel (a): comparison of transients obtained from the reduced intensive models (equations 46 and (49) with that of the full model (11-13) using the parameters of Table 1 and $\varepsilon = 1$. Panel (b): difference between the results of each reduced model and that of the full model as a function of ε. For equation (46) derived through SPA, the error drops linearly with ε, whereas for equation (49) the error remains high.

Figure 2: Bifurcation diagram for the reduced intensive model in the case of positive feedback (panel a: $k_0 = 0.1$ min^{-1} and $k_1 = 16$ min^{-1}) and negative feedback (panel b: $k_0 = 16$ min^{-1} and $k_1 = 0.1$ min^{-1}). Solid and dashed lines correspond to stable and unstable steady states, respectively. Bistability is observed when positive feedback exists. Parameters not mentioned have the values of Table 1.

Figure 3: Panel (a): cell contents for all cells in the population as a function of time. Panel (b): cell concentrations for all cells in the population. Parameter values as in Table 1.

Figure 4: Temporal evolution of the normalized number density of a cell population. Panels (a, c, e, g) correspond to a hybrid ensemble/volume CPB simulation, whereas panels (b, d, f, h) correspond to the full cell population Monte Carlo simulation. Parameter values as in Table 1 and for the IC: $\mu_V = 0.6$ fL (where 1 fL = 10^{-15} L), $\sigma_V = 0.05 \cdot \mu_V$, $\mu_C = 3.763$ nM, $\sigma_C = 0.05 \cdot \mu_C$. For an animation of these results see Supplementary Material.

Figure 5: Panel (a): error of the full CPB model as a function of the population size in the constant-number KMC scheme. Population sizes: 6250, $1.25 \cdot 10^4$, $2.5 \cdot 10^4$, $5 \cdot 10^4$. The error was calculated from equation (62) with $\Delta t = 55$ min, and $N_t = 5$. For the discretization of the number density functions: $\Delta V = 0.01$ fL (where 1 fL = 10^{-15} L), $\Delta y = 0.05$ nM. The most accurate solution was the one obtained by solving each component of the hybrid model for $8 \cdot 10^5$ cells. The dashed line shows the inverse square root convergence rate ($294.6/\sqrt{N_{population}}$). Panel (b): Ratio of computational times for the full CPB model and the hybrid model. The solid horizontal line marks a value of this ratio equal to 1. The dashed line is a power law fit of the data for the computational times ($4.97 \cdot 10^{-9} \cdot N_{population}^{1.75}$).

Figures

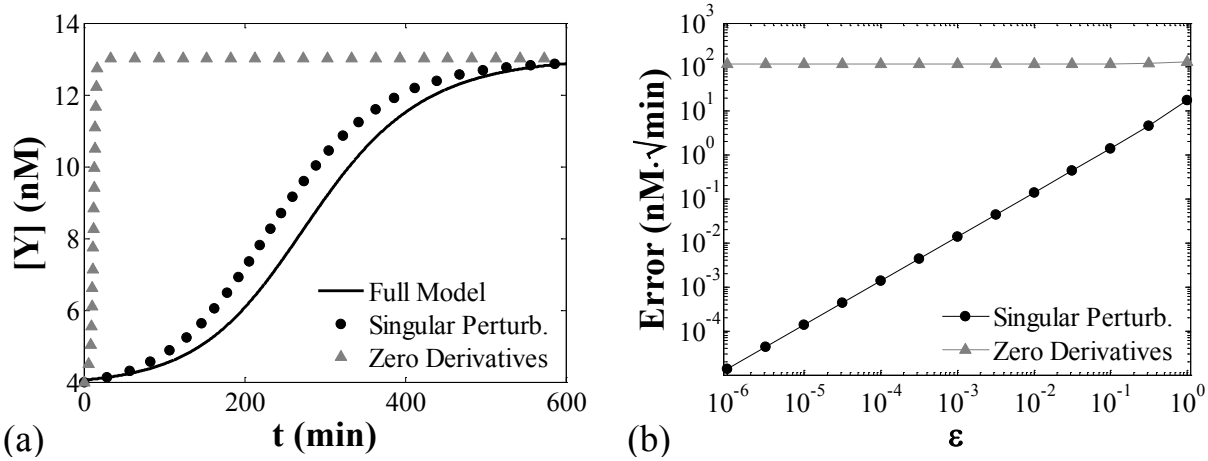

Figure 1

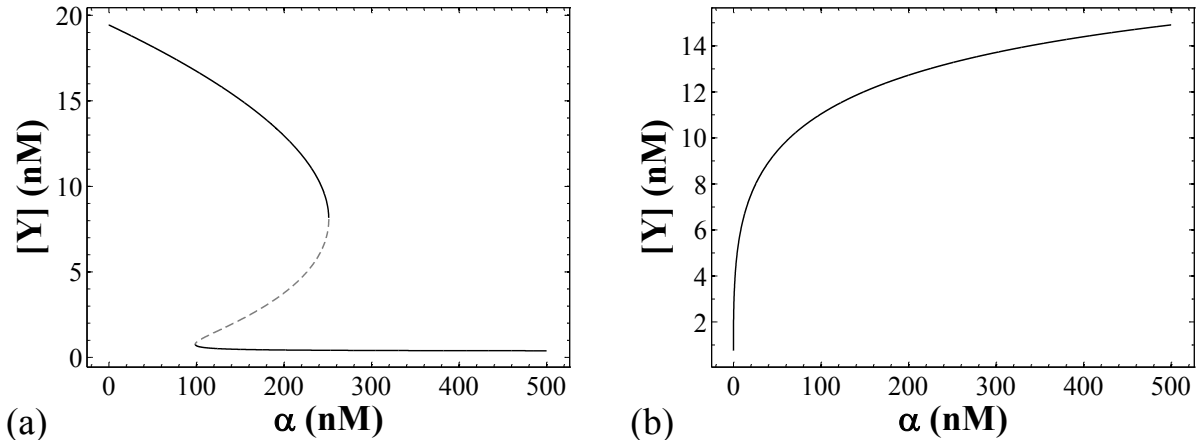
20
15
10
5
0
[Y] (nM)
0
100
200
300
400
500
(a)
α (nM)
14
12
10
8
6
4
2
[Y] (nM)
0
100
200
300
400
500
(b)
α (nM)

Figure 2

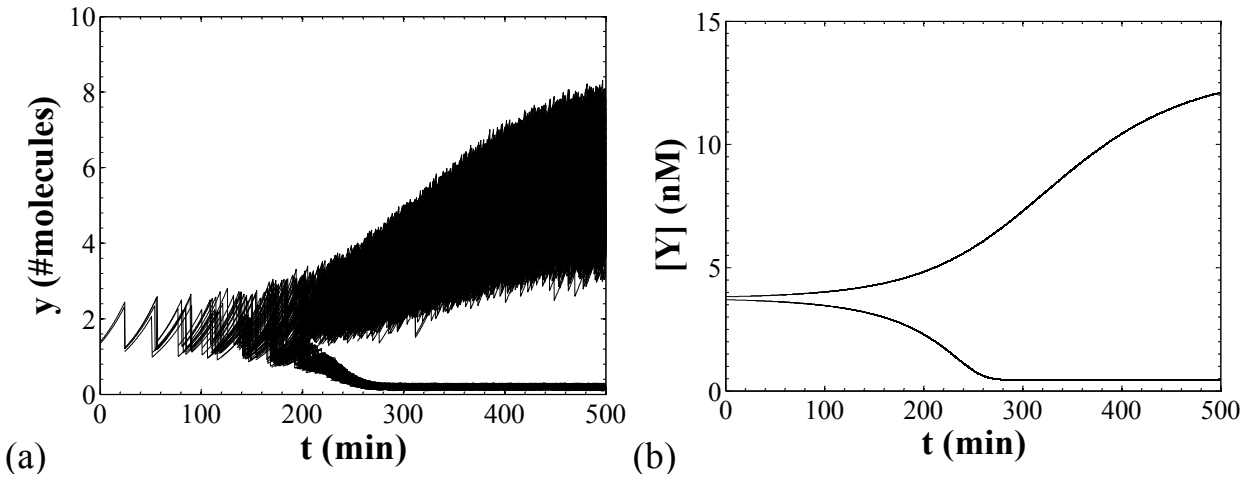

Figure 3

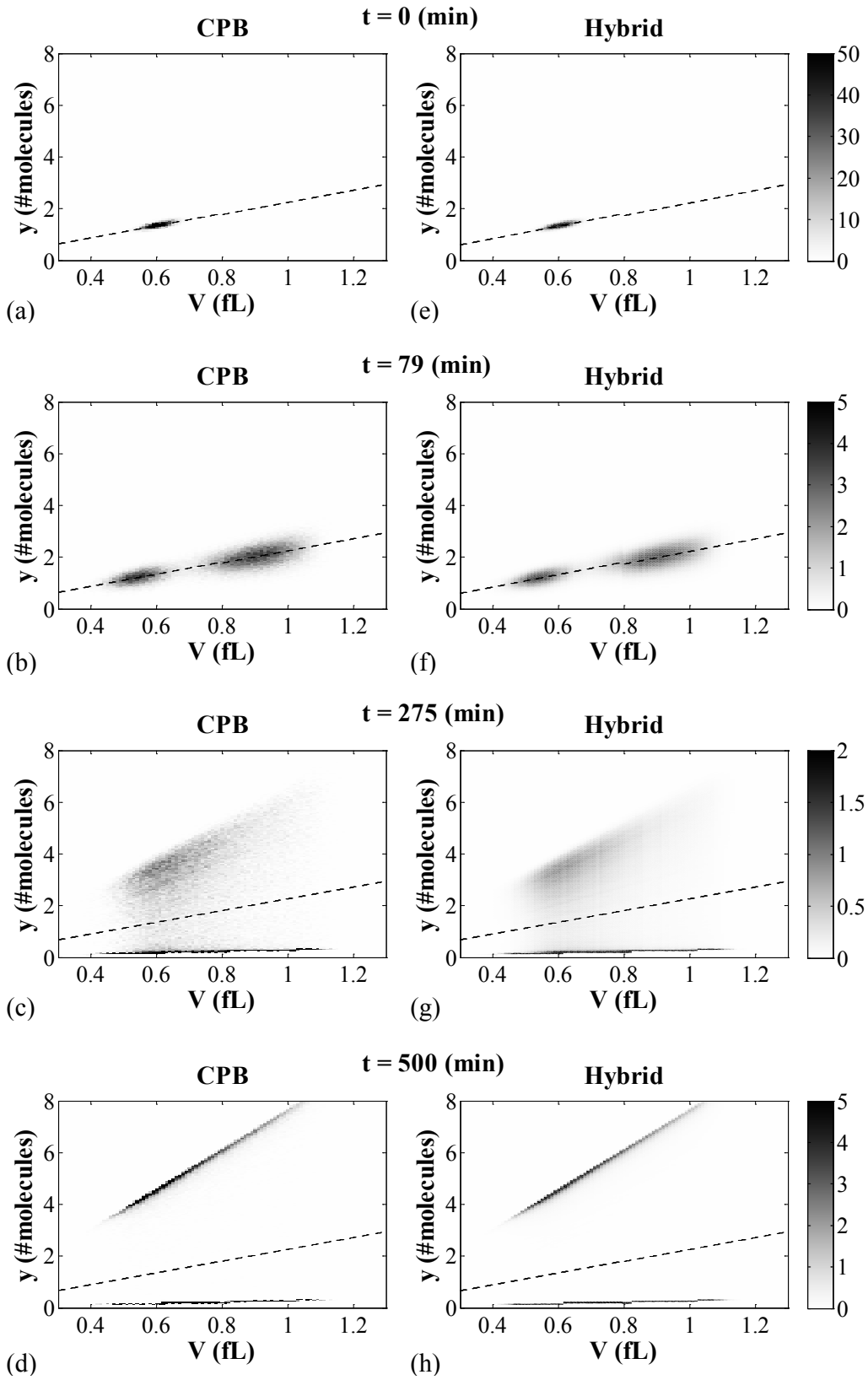
t = 0 (min)
CPB
Hybrid
t = 79 (min)
CPB
Hybrid
t = 275 (min)
CPB
Hybrid
t = 500 (min)
CPB
Hybrid
y (#molecules)
V (fL)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

Figure 4

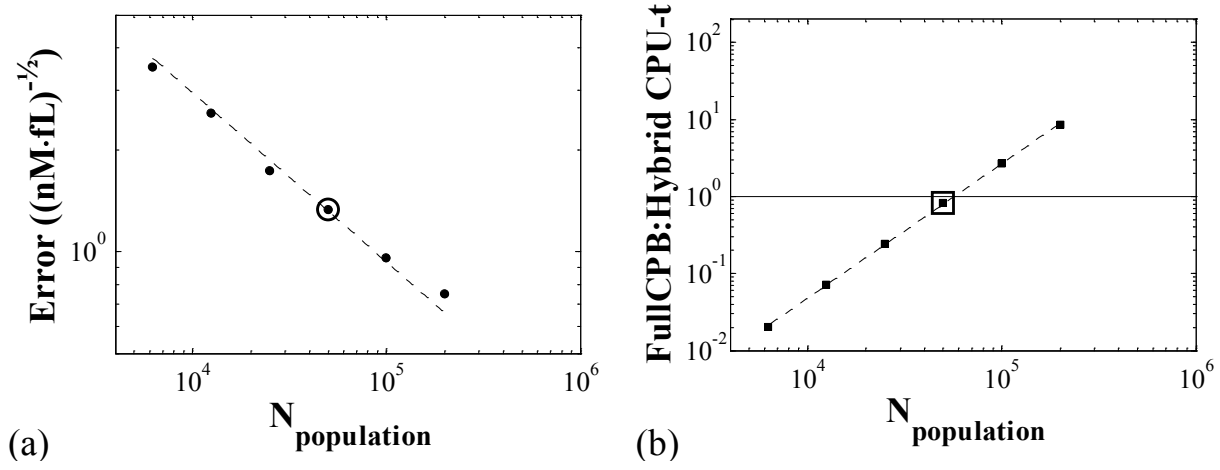
Error ((nM·fL)^-½)
10^0
10^4
10^5
10^6
N population
(a)
FullCPB:Hybrid CPU-t
10^2
10^1
10^0
10^-1
10^-2
10^4
10^5
10^6
N population
(b)

Figure 5