

RESEARCH

Open Access



Analysis and remedy of negativity problem in hybrid stochastic simulation algorithm and its application

Minghan Chen and Yang Cao*

From International Workshop on Computational Network Biology: Modeling, Analysis and Control
Washington, D.C., USA. 29 August 2018

Abstract

Background: The hybrid stochastic simulation algorithm, proposed by Haseltine and Rawlings (HR), is a combination of differential equations for traditional deterministic models and Gillespie's algorithm (SSA) for stochastic models. The HR hybrid method can significantly improve the efficiency of stochastic simulations for multiscale biochemical networks. Previous studies on the accuracy analysis for a linear chain reaction system showed that the HR hybrid method is accurate if the scale difference between fast and slow reactions is above a certain threshold, regardless of population scales. However, the population of some reactant species might be driven negative if they are involved in both deterministic and stochastic systems.

Results: This work investigates the negativity problem of the HR hybrid method, analyzes and tests it with several models including a linear chain system, a nonlinear reaction system, and a realistic biological cell cycle system. As a benchmark, the second slow reaction firing time is used to measure the effect of negative populations on the accuracy of the HR hybrid method. Our analysis demonstrates that usually the error caused by negative populations is negligible compared with approximation errors of the HR hybrid method itself, and sometimes negativity phenomena may even improve the accuracy. But for systems where negative species are involved in nonlinear reactions or some species are highly sensitive to negative species, the system stability will be influenced and may lead to system failure when using the HR hybrid method. In those circumstances, three remedies are studied for the negativity problem.

Conclusion: The results of different models and examples suggest that the Zero-Reaction rule is a good remedy for nonlinear and sensitive systems considering its efficiency and simplicity.

Keywords: Hybrid stochastic algorithm, Negative population, Second slow reaction firing time

Background

The stochastic simulation algorithm (SSA), also often called Gillespie's algorithm [1, 2], is a major stochastic simulation method for simulating stochastic effects in biochemical networks. Although the SSA is quite reliable in numerous applications in computational biology, the algorithm is computationally intensive and inefficient for systems with fast reactions or large populations. Though many optimizations have been proposed to improve the

efficiency of the algorithm [3–8], the essential idea of simulating each reacting event in a dynamical system makes it unpromising for large and complex biochemical systems compared to traditional deterministic methods.

To avoid the expensive computational cost of the SSA, researchers began studying approximation strategies. One well-known approximation is the τ -leap method [9], which approximates many reaction events in an interval of τ instead of simulating each reaction. As biological networks at single cell levels usually have large scale discrepancies in populations of species such as mRNAs and proteins, as well as rate constants among different reactions, research is increasingly focused on hybrid

*Correspondence: ycao@cs.vt.edu
Department of Computer Science, Virginia Tech, Blacksburg 24061, VA, USA



methods targeting multiscale systems that contain species populations or reaction rates with widely varying scales [10–13]. One branch of the hybrid method is the piecewise deterministic Markov process [10, 14, 15], which mixes the deterministic evolution with random jumps. Under the SSA branch, one hybrid method is to combine the τ -leap algorithm and the SSA for multiscale features among species populations [13]. Species and corresponding reactions are partitioned into two sets based on their populations, one simulated by the SSA and the other simulated by the τ -leap method. In a multiscale system, fast reactions can reach partial equilibrium or quasi-steady-state under certain conditions. Hybrid methods, like the slow-scale SSA method (ssSSA) [16, 17] and the stochastic quasi-steady-state SSA method (SQSSA) [12], were proposed based on this property. The ssSSA partitions the system into fast reaction and slow reaction sets, assuming partial equilibrium for the fast reactions, while simulating the slow reactions with the SSA. Similarly, the SQSSA first separates intermediate species and their corresponding reactions from the system, then assumes that the separated subsystem is at a steady state and simulates the rest of the system with the SSA. But both methods have limitations on parameter space to ensure the system validity [18, 19].

For general cases where fast reactions do not always reach a steady state or partial equilibrium, Haseltine and Rawlings [11] proposed a hybrid method (hereafter referred to as the HR hybrid method), which modeled part of the system by continuous dynamics (ordinary differential equations (ODEs) or Langevin equations), while keeping the rest discrete. The idea of the HR hybrid method was further explored, improved, and extended to several hybrid methods [20–23]. In Salis et al.'s work [20], they partitioned the system into fast and slow reaction groups, and modeled the fast group by Langevin equations and the slow group by the SSA. Later, Liu et al. [21] improved the efficiency of the HR hybrid method by a different partitioning strategy: reactions that have both low-density reactants and small reaction rates were put into the slow reaction subsystem and all the other reactions into the fast reaction subsystem. Wang et al. [24] optimized the implementation efficiency for the HR hybrid method and compared the efficiencies of the hybrid method coupled with three traditional ODE solvers RADAU5, DASSL, and DLSODAR. Lecca et al. [22] further divided the system into three sets: fast reactions, moderate reactions, and slow reactions, where the simulation of moderate reactions can be switched between stochastic and deterministic processes based on the reaction firing time during the system evolution. For spatial models or domains, hybrid methods were introduced under reaction-diffusion systems, where diffusion was approximated by differential equations to improve simulation efficiency [23, 25, 26].

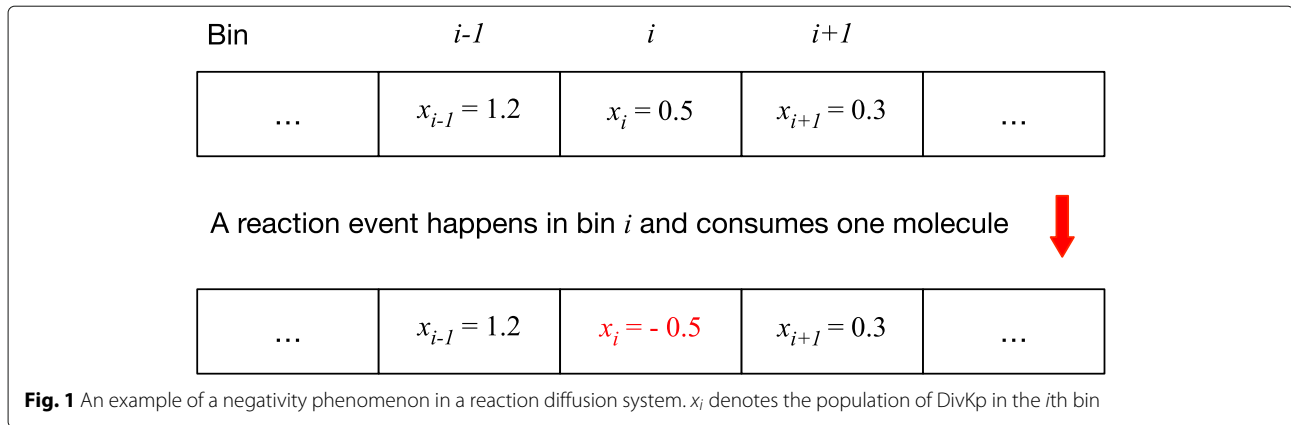
Simulation tools, e.g. Hy3S [27] and MoBioS [22], and software like COPASI [28] were investigated based on the HR hybrid method and provided users with different simulation choices and implementation rules. As to the application on complex biochemical models, Wang et al. [29] used the HR hybrid method to model a budding yeast cell cycle. The method largely reduced the simulation time and the results matched well with experimental data on cell cycle properties and prototypes of most mutant cells. To mathematically analyze the accuracy of the HR hybrid method, Chen et al. [30] used the next reaction time of the slow reaction event as the accuracy benchmark and showed that the HR hybrid method is accurate in linear chain systems under certain conditions (either large populations of reactants in the fast subsystem or large scale differences of reaction rates between fast reactions and slow reactions). It also demonstrated that the HR hybrid method is valid for a much greater region in system parameter space than those for the ssSSA and the SQSSA methods.

However, in the HR hybrid method framework, populations of some reactant species may become negative if they are involved in both deterministic and stochastic systems. Take system (1) as an example. If reaction rate constants satisfy $f_1 \gg k_c$ and $b_1 \gg k_c$, the system can be divided into two groups: the fast reaction group and the slow reaction group, containing the reversible and irreversible reactions, respectively.



Assume that this system has two S_1 molecules at the beginning, and the system parameters are $f_1 = 1, b_1 = 9, k_c = 0.01$. Then, compared with the slow system, the fast system can be considered at equilibrium, which gives $x_1 = 1.8$ and $x_2 = 0.2$, where x_i denotes the mean population of species S_i . Thus, when a slow reaction fires, x_2 is reduced to -0.8 . Only after a certain period of time, it may become nonnegative again through the reaction $S_1 \rightarrow S_2$.

Negative populations may also appear in stochastic simulations of reaction-diffusion systems, especially when low-density species are distributed in a well-meshed space. For example, in a one-dimensional spatial model of the *Caulobacter* cell cycle [31], the spatial domain is divided into 50 equally spaced bins. Since diffusion happens much faster than chemical reactions in the cell, diffusion events are modeled as continuous deterministic equations whereas chemical reactions are modeled by the SSA. In the initial stage of the cell cycle, protein DivKp has a low population (< 50). Thus the average population of DivKp inside each bin is < 1 . Note that the mean population of a species is a real number if it is involved in the fast subsystem. As illustrated in Fig. 1,



if there is a degradation reaction of DivKp firing in the i th bin, its population would become negative. Therefore, any consumption of those low population species in the spatial stochastic domain may lead to a negative population. The phenomenon of species' populations becoming negative, as shown in the above two examples, is called the negativity problem for the HR hybrid method.

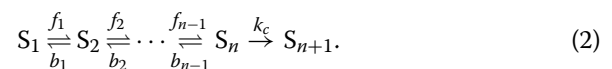
This paper is organized as follows. In the Methods section, we present the theoretical derivation of the second exit time of the chemical master equation (CME), the HR hybrid method, and three proposed remedies for the negativity problem. The Results and Discussion section analyzes the potential negativity effects on the accuracy of linear chain systems for a simple case ($n = 2$) and a complex case ($n = 10$). We test three remedies on three examples: a closed linear chain system, a nonlinear system, and a realistic biological system. Summary and conclusions are given last.

Methods

Second slow reaction firing time

Our prior work [30] analyzed the accuracy of the HR hybrid method by studying the next slow reaction firing time (NSRFT, also called the first exit time). Since the negative population problem mostly emerges after a slow reaction, it is not enough to just study the first exit time. So, we further extend that work and study the second slow reaction firing time (SSRFT, which can also be referred to as the second exit time). The SSRFT reflects the influence of a (possible) negative population on the firing of slow reactions. With this analysis, we hope to gain certain insight on the impact of the negativity problem on algorithm accuracy. In the HR hybrid method we assume reactants become negative only after a slow reaction happens, which is after the first exit time. Negative populations may also arise, with a much smaller probability, in the numerical integration of ODEs. That is not our focus here.

We use the same linear chain reaction network in Refs. [30, 32] as a study example, shown below.



A particle can exit the reversible chain system through S_n with reaction rate k_c . In most cases, k_c is comparably less than reaction rates f_i and b_i for $1 \leq i \leq n - 1$. In many applications, the reversible chain reactions can be considered as a fast subsystem and the irreversible reaction (exit to S_{n+1}) as a slow subsystem. With this partitioning strategy, if $x_n < 1$, then S_n will become negative whenever a slow reaction fires.

SSRFT for the CME

While the first exit time (NSRFT) denotes the time when the next slow reaction fires in the linear chain system (2), the second slow reaction firing time (SSRFT) is the time period from the system start to the second time the slow reaction fires.

Recapping the derivation of NSRFT in Ref. [30], the SSRFT can be considered as the probability that two independent events (NSRFT) happen in a time interval $[0, t]$. In system (2), $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_{n+1}(t))^T$ represents the system state at time t . If there is only one particle in the system, we denote the probability of $x_i = 1$ as

$$p_i(t) = P[x_i(t) = 1], \quad \text{for } i = 1, \dots, n + 1.$$

and the probability vector for species S_1, S_2, \dots, S_n as

$$\mathbb{P}(t) = [p_1(t), \dots, p_n(t)]^T$$

In the chemical master equation system, we have

$$\frac{d\mathbb{P}}{dt} = -A\mathbb{P}, \quad (3)$$

where A is stoichiometric matrix,

$$A = \begin{bmatrix} f_1 & -b_1 & 0 & 0 & \cdots & 0 \\ -f_1 & b_1 + f_2 & -b_2 & 0 & \cdots & 0 \\ 0 & -f_2 & b_2 + f_3 & -b_3 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -f_{n-2} & b_{n-2} + f_{n-1} & -b_{n-1} \\ 0 & \cdots & 0 & 0 & -f_{n-1} & b_{n-1} + k_c \end{bmatrix}.$$

As there is only one particle all the time in the system, we have $\sum_{i=1}^{n+1} p_i(t) = 1$. And $x_{n+1}(t) = 1$ if and only if the first exit time $T_1 \leq t$. Thus

$$P[T_1 \leq t] = p_{n+1}(t) = 1 - \sum_{i=1}^n p_i(t) = 1 - \vec{e}^T \mathbb{P}(t),$$

where $\vec{e} = [1, \dots, 1]^T$. Given an initial condition \vec{e}_j (a vector with the j th element equal to 1 and all other elements equal to 0), the NSRFT is (see Ref. [30])

$$\begin{aligned} q_j(T_1) &= P[T_1 > t] = \vec{e}^T e^{-At} \vec{e}_j \\ &= 1 - p_{n+1}(t) = 1 \int_0^t k_c p_n(s) ds. \end{aligned}$$

In a general case where there are m particles ($\vec{x}_0 = [m_1, m_2, \dots, m_n]^T, m = \sum_{i=1}^n m_i$) in this linear system, as particles are independent of each other, the NSRFT is

$$q(T_1) = \prod_{j=1}^n q_j^{m_j}(T_1).$$

Based on similar analysis to NSRFT, the second slow reaction firing time can be written as

$$\begin{aligned} P[T_2 \leq t] &= 1 - \prod_{j=1}^n q_j^{m_j} - \sum_{i=1, m_i \neq 0}^n C_{m_i}^1 (1 - q_i) q_i^{m_i - 1} \prod_{j=1, j \neq i}^n q_j^{m_j}. \end{aligned} \tag{4}$$

For a simple case where $n = 2$ and the initial condition: $m_1 = 2$ and $m_2 = 0$, we have

$$P[T_2 \leq t] = (1 - q_1(t))^2. \tag{5}$$

SSRFT for the HR hybrid method

In the HR hybrid method, define the state vector in the fast subsystem as $\vec{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$. The fast subsystem is modeled as a linear ODE system, denoted as

$$\frac{d\vec{x}(t)}{dt} = -\tilde{A}\vec{x}(t). \tag{6}$$

\tilde{A} is a $n \times n$ matrix given by

$$\tilde{A} = A - k_c \vec{e}_n \vec{e}_n^T,$$

where only the last elements of matrices \tilde{A} and A are different, $\tilde{A}(n, n) = b_{n-1}, A(n, n) = b_{n-1} + k_c$.

Denote T_1 as the NSRFT, we have

$$\int_0^{T_1} k_c x_n(t) dt = \int_0^{T_1} \vec{e}_n^T e^{-\tilde{A}t} \vec{x}_0 dt = r,$$

where r is a unit exponential random number. In order to compare it with the CME result, we have to change the exponential random number to a unit uniform random number u by the relation $u = 1 - e^{-r}$. The above equation can be written as

$$P[T_1 \leq t] = 1 - e^{-\int_0^t k_c \vec{e}_n^T e^{-\tilde{A}s} \vec{x}_0 ds} = u.$$

And the density function of the NSRFT is

$$p(T_1) = k_c \vec{e}_n^T e^{-\tilde{A}T_1} \vec{x}_0 e^{-\int_0^{T_1} k_c \vec{e}_n^T e^{-\tilde{A}s} \vec{x}_0 ds}. \tag{7}$$

The SSRFT for the HR hybrid method can be considered as the next slow reaction firing time with a different initial condition \vec{x}_{T_1} (the system state after the first exit time T_1),

$$\int_{T_1}^{T_2} k_c x_n(t) dt = \int_{T_1}^{T_2} \vec{e}_n^T e^{-\tilde{A}(t-T_1)} \vec{x}_{T_1} dt = r,$$

where $\vec{x}_{T_1} = e^{-\tilde{A}T_1} \vec{x}_0 - \vec{e}_n$. Thus

$$\begin{aligned} p(T_2|T_1) &= \left(k_c \vec{e}_n^T e^{-\tilde{A}(T_2-T_1)} \vec{x}_{T_1} - k_c \vec{e}_n^T \vec{x}_{T_1} \right) \\ &\quad e^{-\int_{T_1}^{T_2} k_c \vec{e}_n^T e^{-\tilde{A}(t-T_1)} \vec{x}_{T_1} dt}. \end{aligned} \tag{8}$$

Therefore, the SSRFT is

$$p(T_2) = \int_0^\infty p(T_1) p(T_2|T_1) dT_1. \tag{9}$$

Below we present three strategies to handle the negativity problem and compare the corresponding impact on the SSRFT.

SSRFT for remedy I: zero-population

For the negativity problem in the HR hybrid method, one simple treatment is to immediately change any negative value to zero. So we name this strategy the **Zero-Population** remedy: after a slow reaction happens in the stochastic subsystem, detect whether any corresponding reactants become negative, if so set them as zero and continue the simulation, otherwise continue without modification.

To check the effect of the Zero-Population remedy, we study the second slow reaction firing time of the system under this rule. Since the rule takes effect after the first slow reaction, only the conditional density function $p(T_2|T_1)$ is different from the HR hybrid method when $x_n < 0$. In this scenario, the initial condition of the second slow reaction firing time in linear chain system (2) is

$$\vec{x}_{T_1} = e^{-\tilde{A}T_1} \vec{x}_0 - \vec{e}_n, \quad \vec{x}_{T_1}(n) = \max(0, \vec{x}_{T_1}(n)). \tag{10}$$

The conditional density function of the second exit time in the Zero-Population remedy is similar to Eq. (8), just replace the initial condition with $\vec{x}_{T_1}(n)$ in Eq. (10).

SSRFT for remedy II: zero-reaction

While the Zero-Population remedy avoids the negative effect, it changes the conservation law in the system. For example, the total amount of all species in the closed system (2) should always be m , but simply changing the negative population ($-m_\delta$) to zero causes the total population to increase to $m + m_\delta$. In order to follow the law of conservation, which is important in many practical applications, one idea is to scale down a reaction when the slow reaction happens with a reactant population less than one. Take the reaction $X \rightarrow Y$ as an example. Suppose the population of reactant X is less than one ($0 < m_X < 1$), then instead of consuming one molecule of X , scale the reaction by a ratio of m_X , which produces m_X molecule of Y . However, this method breaks the natural discrete feature of slow reaction firing and can cause significant errors for stochastic models.

Alternatively, one can simply set all related reaction propensities as zero when a negative population appears. So we have the second remedy named the **Zero-Reaction** rule: set all reaction propensities involving negative species as zero in corresponding subsystems until the negative species become nonnegative.

After the first slow reaction, the system status is $\vec{x}_{T_1} = e^{-\tilde{A}T_1}\vec{x}_0 - \vec{e}_n$. For $x_n < 0$, reaction rates for all reactions that S_n participated in the fast subsystem are treated as zero, and the corresponding coefficient matrix for the ODEs is denoted as \hat{A} . The propensity of the slow reaction, which S_n participated in, is also set to zero. Since system (2) has only one slow reaction, no reaction in the slow subsystem will fire for negative x_n . The fast subsystem keeps running until $x_n = 0$. We denote τ as the time period of x_n evolving from negative to zero in the ODE system:

$$\vec{e}_n^T e^{-\hat{A}\tau} \vec{x}_{T_1} = 0, \tag{11}$$

where the above equation has a unique solution τ . So after time $T_1 + \tau$, the system is back to normal, with the system state $\vec{x}_\tau = e^{-\hat{A}\tau} \vec{x}_{T_1}$.

Thus, the conditional density function of the SSNFT under the Zero-Reaction remedy is

$$p(T_2|T_1) = \left(k_c \vec{e}_n^T e^{-\tilde{A}(T_2-T_1-\tau)} \vec{x}_\tau - k_c \vec{e}_n^T \vec{x}_\tau \right) e^{-\int_{T_1+\tau}^{T_2} k_c \vec{e}_n^T e^{-\tilde{A}(t-T_1-\tau)} \vec{x}_\tau dt}. \tag{12}$$

SSRFT for remedy III: zero-time

Another remedy for the negativity problem is called the **Zero-Time** rule: whenever a species' population become negative, pause the system and run a separate virtual ODE system G'_f that only contains reactions related to the negative species. When the species in the virtual system recovers to a nonnegative state, restart the original

hybrid system with the updated system state. Because the system recovers under existing reaction systems, the conservation law is obeyed.

In system (2), the system state is still $\vec{x}_{T_1} = e^{-\tilde{A}T_1}\vec{x}_0 - \vec{e}_n$ after the first slow reaction. For $x_n < 0$, build a separate ODE system that only contains the last reversible reaction,

$$G'_f : S_{n-1} \xrightleftharpoons[b_{n-1}]{f_{n-1}} S_n. \tag{13}$$

Run the G'_f system until $x_n = 0$, which costs time ρ from the below equation.

$$\vec{e}_2^T e^{-B\rho} \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix} = 0, \quad B = \begin{bmatrix} f_{n-1} & -b_{n-1} \\ -f_{n-1} & b_{n-1} \end{bmatrix} \tag{14}$$

Since the recovery time ρ is not counted in system evolution, we have the conditional density function of the second exit time under the Zero-Time remedy as

$$p(T_2|T_1) = (k_c \vec{e}_n^T e^{-\tilde{A}(T_2-T_1)} \vec{x}_\rho - k_c \vec{e}_n^T \vec{x}_\rho) e^{-\int_{T_1}^{T_2} k_c \vec{e}_n^T e^{-\tilde{A}(t-T_1)} \vec{x}_\rho dt}, \tag{15}$$

where \vec{x}_ρ is the system state after running the G'_f system for ρ .

Results and discussion

Theoretical analysis of SSRFT

A simple case ($n = 2$)

This subsection examines the SSRFT of the linear system (2) with $n = 2$, as shown in (1).

We first check conditions that may cause the negativity problem, such as parameters and initial conditions. In system (1), after the first slow reaction, we have

$$\vec{x}_{T_1} = e^{-\tilde{A}T_1}\vec{x}_0 - \vec{e}_2, \quad \tilde{A} = \begin{bmatrix} f_1 & -b_1 \\ -f_1 & b_1 \end{bmatrix},$$

thus the population of S_2 at time T_1 is

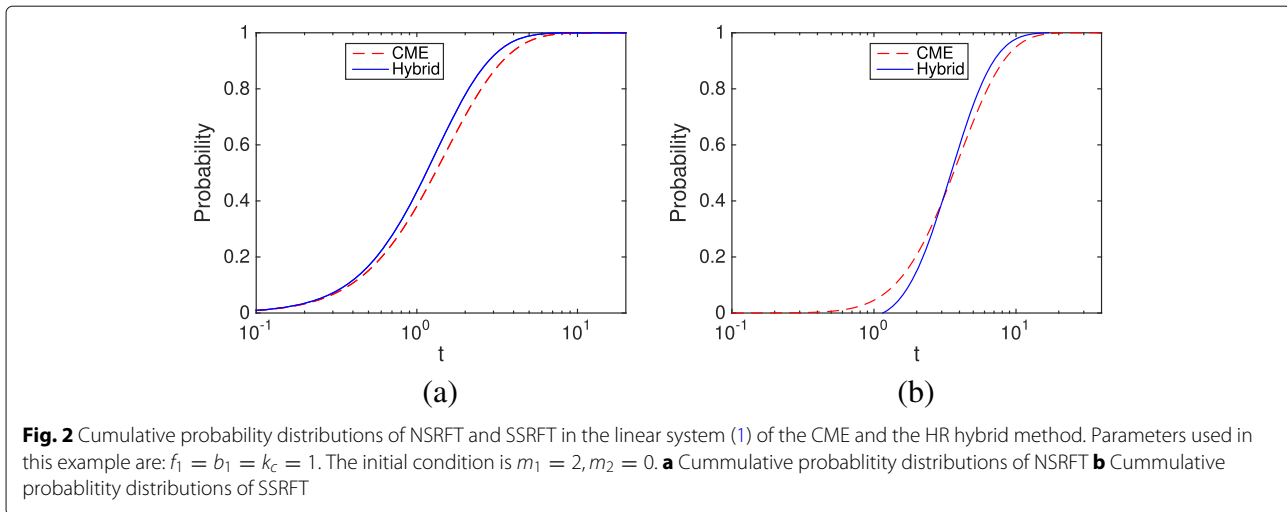
$$\vec{x}_{T_1}(2) = \frac{(m_2 b_1 - m_1 f_1) e^{-(b_1+f_1)T_1} + (m_1 + m_2) f_1}{b_1 + f_1} - 1.$$

The first slow reaction firing may happen from time 0 to ∞ . To ensure that the population of S_2 is nonnegative for all possible T_1 , we should have $\vec{x}_{T_1}(2) \geq 0$ for all T_1 . We solve this inequality and get

$$m_1 + m_2 \geq \frac{f_1 + b_1}{f_1} \quad \text{and} \quad m_2 \geq 1. \tag{16}$$

For the parameter set $f_1 = b_1 = 1$, the population of S_2 may become negative when the initial condition satisfies $m_2 = 0$ (Assume $m_1 + m_2 \geq 2$, so a second slow reaction is possible). For the parameter set $f_1 = 1, b_1 = 10$, the population of S_2 may become negative when the initial condition satisfies $m_2 = 0$ or $m_1 + m_2 < 11$.

Figure 2 presents the cumulative distribution functions (CDFs) of NSRFT and SSRFT from both the CME and the HR hybrid method, respectively. The model parameters



and initial conditions are chosen so that the negativity problem may arise. For the NSRFT, when t is small, the two methods are close enough. When $t \in [1, 10]$, the HR hybrid method has the first slow reaction firing earlier than the CME. In this time interval, the mismatch between the two methods comes from the error of the hybrid method, rather than the negativity problem. For the SSRFT, the CDFs of two methods have an intersection at around $t^* = 2$, where before this point the HR hybrid method tends to fire the second slow reaction later than the CME. This difference shows the impact of the negativity problem. After the first slow reaction, x_2 becomes negative, and the system needs extra time to recover, which causes a delay for the SSRFT in the HR hybrid method.

For comparison, we calculate the relative error

$$e_r = \frac{|T_c - T_h|}{T_c}, \quad (17)$$

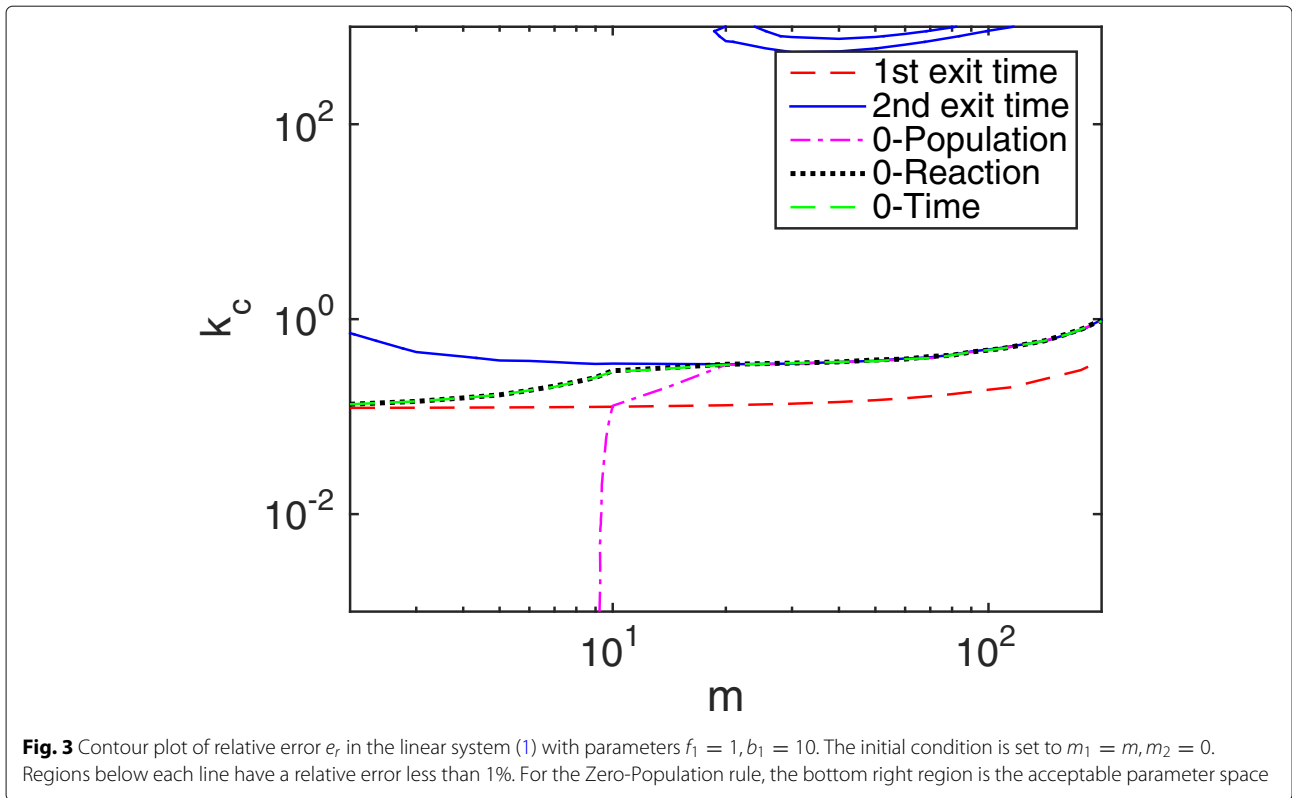
where T_c and T_h are the second slow reaction firing times of the CME and the HR hybrid method, respectively. Here we sample the mean SSRFT as T_c and T_h , i.e. $P(T_2 < T_c) = 0.5, P(T_2 < T_h) = 0.5$.

When we increase b_1 from 1 to 10 and set the initial condition as $m_1 = m, m_2 = 0$, then S_2 has a low population and a great chance to become negative when the slow reaction fires. In Fig. 3, the acceptable region for $e_r < 0.01$ of the SSRFT is surprisingly larger than that of the NSRFT. But it can be well explained. First, it is known that the NSRFT of the HR hybrid method is faster than that of the CME from the previous work [30] and the example in Fig. 2a. If there is no negativity problem, then the SSRFT of the hybrid method should be even faster than that of the CME as it accumulates error from two slow reactions. However, when there is a negative species, e.g. S_2 in this

case, the system slows down to recover from its negative state. The negativity recovery time, to some extent, reduces the numerical error of the hybrid method in the linear chain system. This runs like a way of coordination between two subsystems: when a species becomes negative caused by a slow reaction, i.e. the slow subsystem runs faster than expected, then the slow subsystem has to wait longer for the next slow reaction to fire again.

For the Zero-Reaction and Zero-Time remedies, though they have a smaller acceptable parameter space compared with the HR hybrid method, they still have a larger parameter region than the NSRFT. Note that in the linear chain system, the Zero-Time rule is similar to the Zero-Reaction rule as G'_f is the fast subsystem and time τ, ρ are comparably small, but they are different for other cases, such as when G'_f is different from the fast subsystem or when there are more slow reactions in the slow subsystems. For the Zero-Population remedy, the acceptable parameter space is only half of the other methods. All three remedies converge to the contour line of the HR hybrid method when $m \approx 11$, satisfying one of the nonnegative requirements (16). On the other hand, when the total population is large (or $m \geq \frac{f_1 + b_1}{f_1}$), a negative population appears in a shorter time window, whereas Ref. [30] has demonstrated that the error of the HR hybrid method becomes smaller with larger populations. This leads us to a conjecture regarding the HR hybrid method for the linear chain reaction system: the error caused by negative populations is negligible compared to the original error of the hybrid method. In other words, the negativity effect is substantial only when the method is already problematic in accuracy. Sometimes it acts as a positive sign to reduce the numerical error of the HR hybrid method.

For general cases that include both negative and non-negative situations, we calculate the mean relative error based on all possible initial conditions \vec{x}_0 , which follows

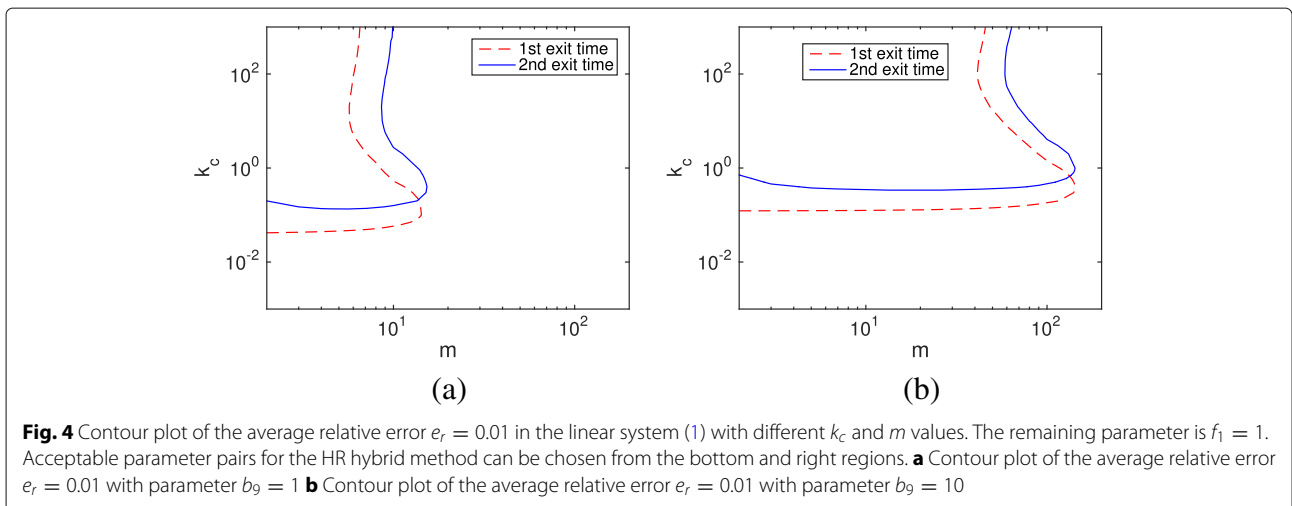


the steady state distribution of the fast subsystem subject to f_1 and b_1 . Figure 4 illustrates the acceptable system parameter region ($e_r < 0.01$) for $b_1 = 1$ and $b_1 = 10$. In both cases, the SSRFT keeps the same pattern but with improved accuracy horizontally resulting from the negativity phenomenon and decreased accuracy vertically due to the accumulative method error. Since nonnegative situations occur much more frequently than negative situations (where the initial condition must be either $m_2 = 0$ or $m < 11$), the three proposed remedies do not make a difference in the acceptable region of the HR hybrid method.

A larger system ($n = 10$)

If we increase the length of the linear chain system to $n = 10$, it is hard to calculate the derived formulas (4, 9) for the second exit time. Instead, we run simulations of each method and collect samples for the NSRFT and the SSRFT. For each pair of (m, k_c) , the reported NSRFT and SSRFT are the mean values from one million simulation results. Consider that all the contour plots in this subsection are processed to make the outline smooth.

First look at one negative case where the initial condition is $m_1 = m, m_i = 0 (i = 2, 3, \dots, 10)$. In Fig. 5,



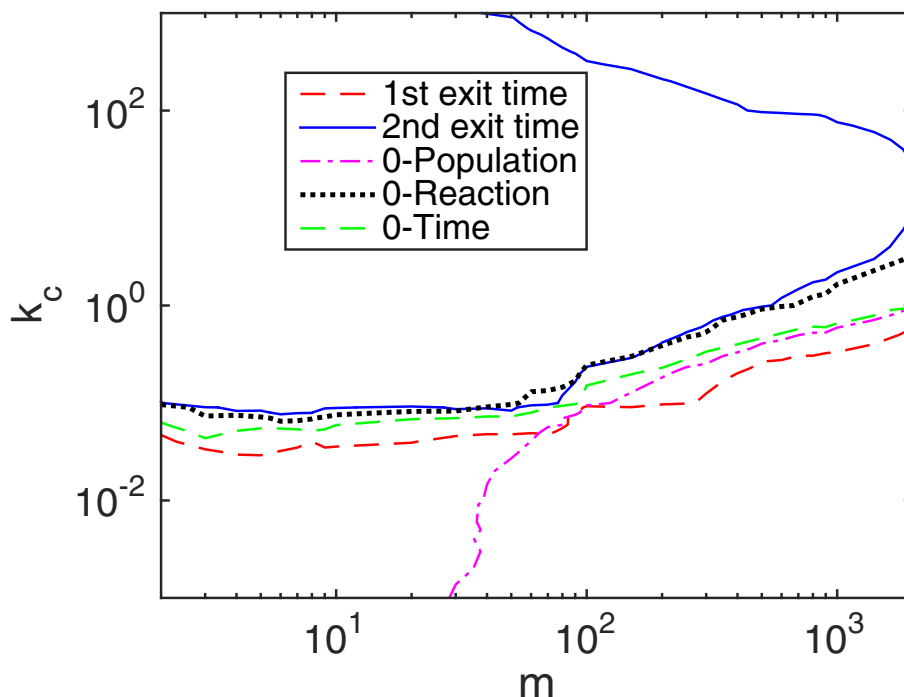


Fig. 5 Contour plot of relative error of e_r in the linear chain system (2) with parameters $f_i = b_i = k_c = 1, b_9 = 10$. The chain length is $n = 10$ and the initial condition is set to $m_1 = m, m_i = 0 (i = 2, 3, \dots, 10)$. Regions below each line have a relative error less than 1%. Note that the HR hybrid method has an extra top right region of acceptable parameter pairs

the shape of the acceptable region ($e_r < 0.01$) is similar to the case when $n = 2$. For the Zero-Reaction rule, it has the same accuracy as the original HR hybrid method, except for the top right region; while the Zero-Population rule is only accurate for large m and small k_c . So the observation that negativity does not influence the accuracy still holds for large linear chain systems, if the Zero-Reaction rule is applied. We do not consider the Zero-Time rule as it is much less

efficient than the other two remedies in this larger system.

For general cases, we randomly sample initial conditions for each (m, k_c) pair. And the probability distribution of \vec{x}_0 satisfies the steady state of the reversible reactions controlled by f_i and b_i . Figure 6 exhibits the same pattern as Fig. 4 for both cases $b_9 = 1$ and $b_9 = 10$. The acceptable parameter space for the second exit time is smaller when $n = 10$, which is similar to the first exit time

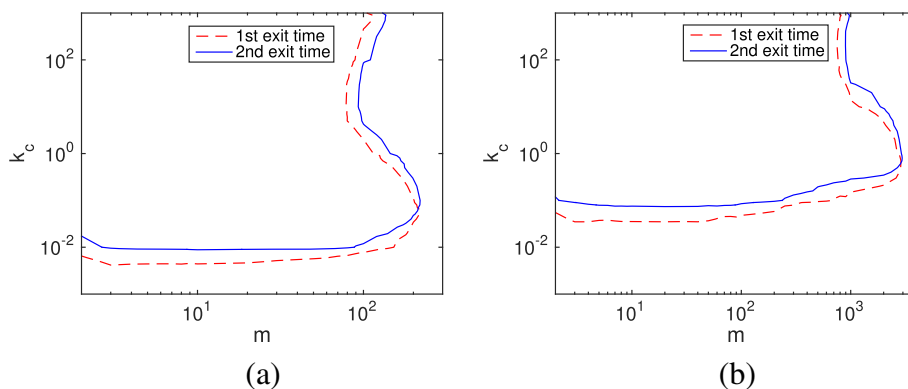


Fig. 6 Contour plot of the average relative error $e_r = 0.01$ in the linear chain system (2) with different k_c and m values. The chain length is $n = 10$. Acceptable parameter pairs for the HR hybrid method can be chosen from the bottom and right parts. **a** Contour plot of the average relative error $e_r = 0.01$ with parameter $b_9 = 1$ **b** Contour plot of the average relative error $e_r = 0.01$ with parameter $b_9 = 10$

discussed in Ref. [30]. Thus, for large linear chain systems, the negativity problem is insignificant for the hybrid method.

Numerical experiments

The previous subsection studied the accuracy based on the first and the second slow reaction times. In this section, we apply the HR hybrid method and three remedies to different systems and compare statistics.

A closed linear chain system

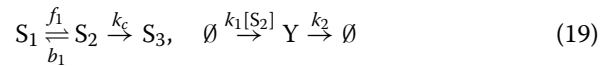
The first example shown below is a similar linear chain system with an extra reaction $S_3 \xrightarrow{k_s} S_1$ that forms a closed system. We divide the system into two parts, the ODE system contains the reversible reaction, while the SSA system takes the remaining reactions involving S_3 .



We choose model parameters and initial conditions so that negative populations occur frequently in the system. Figure 7 shows the average evolution of species S_2 and S_3 from different simulation methods and rules over 10,000 simulations. It is obvious that under the Zero-Population rule, the populations of S_2 and S_3 keep increasing with time while the other methods reach a steady state. The evolution from the Zero-Reaction rule is slightly closer to the results of SSA than the other methods. We then look at the final distributions of species S_2 and S_3 based on 10,000 simulations, shown in Fig. 8. In system (18), x_2 is negative for 30% of the time if using the original HR hybrid method. With the Zero-Time rule, x_2 is always nonnegative, while the Zero-Reaction rule does not change the distribution of S_2 much. For the final distribution of S_2 , though the differences between methods are fairly close (< 10%), the Zero-Reaction rule also works better than others. The results from the Zero-Population rule are much different

from the SSA results, and the final distribution will shift further to the right if we run the system longer.

The above example demonstrates that the system can finally recover from negative populations without using any remedies. To test an extreme case, suppose there is another species Y highly sensitive to S_2 , shown below



The first part is a simple linear chain system with $n = 2$. A particle can exit the reversible chain system through S_2 with reaction rate k_c . The other part is a birth and death process of species Y. S_2 acts as the enzyme activating the synthesis of Y. Following the same partition strategy used above, we isolate the irreversible reaction in a slow subsystem as both the rate constant k_c and quantity of S_2 are small. All the remaining reactions are put into a fast subsystem.

By setting $f_1/b_1 = 0.1$, S_2 maintains a low population and has a high chance to become negative when the slow reaction fires. Once $x_2 < 0$, the slow reaction propensity is negative, which ensures that the slow reaction will not fire until x_2 goes back to a positive value in the fast subsystem. So the negative population of S_2 has no effect on S_3 in this case. But for Y, it can provoke large fluctuations especially when $k_1[S_2] \gg k_2$. Figure 9 shows one simulation result of S_2 and Y. When S_2 first drops to -0.7 , Y is also driven negative and then becomes positive after one time unit, while under the Zero-Reaction rule, Y is always nonnegative. During the recovery period of Y, if there is another species Z dependent on Y and a species A dependent on Z, then Y may incur a cascade of negativity which significantly increases the simulation error. For situations where negative reactants heavily affect other species, a remedy is definitely needed to prevent a simulation failure.

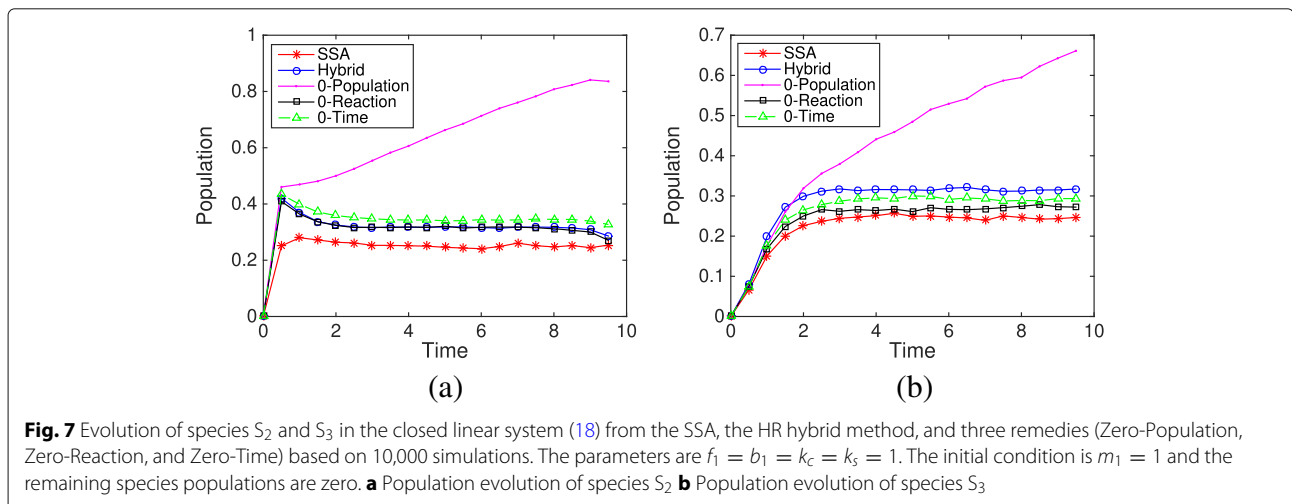
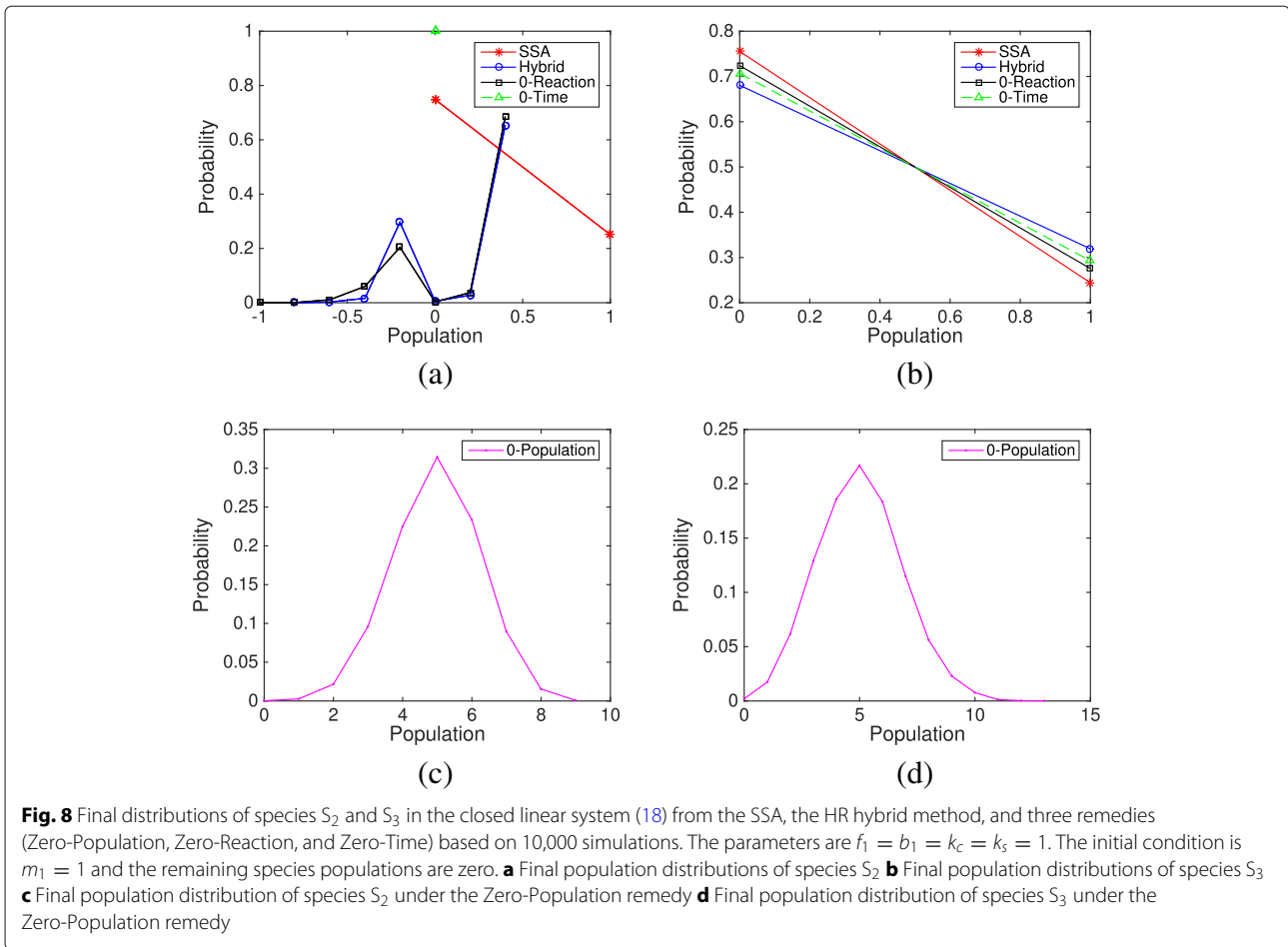
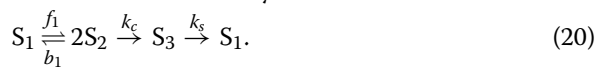


Fig. 7 Evolution of species S_2 and S_3 in the closed linear system (18) from the SSA, the HR hybrid method, and three remedies (Zero-Population, Zero-Reaction, and Zero-Time) based on 10,000 simulations. The parameters are $f_1 = b_1 = k_c = k_s = 1$. The initial condition is $m_1 = 1$ and the remaining species populations are zero. **a** Population evolution of species S_2 **b** Population evolution of species S_3



A closed nonlinear system

From the previous studies, we found that the HR hybrid method works fine for linear systems even with a high frequency of negative populations. Here we want to examine the effect of negative values on nonlinear systems, e.g. bimolecular reactions. Slightly modifying reactions involving S_2 into bimolecular reactions in system (18), we generate a new nonlinear system shown below.



Similarly, we partition the system into groups: the fast group containing the reversible reactions and the slow group containing the remaining reactions. For the bimolecular reaction $S_2 + S_2 \xrightarrow{k_c} S_3$ in the slow subsystem, the propensity is $a_{bi} = k_c x_2(x_2 - 1)$. When $x_2 < 0$, a_{bi} is positive which can potentially cause the reaction to fire and further decrease the value of x_2 . Thus, the SSA system becomes unstable when $x_2 < 0$.

For the fast ODE system, we have

$$\begin{aligned} \frac{dx_1}{dt} &= -f_1 x_1 + b_1 x_2^2 \\ \frac{dx_2}{dt} &= 2f_1 x_1 - 2b_1 x_2^2 \end{aligned}$$

The Jacobian matrix is

$$J = \begin{bmatrix} -f_1 & 2b_1 x_2 \\ 2f_1 & -4b_1 x_2 \end{bmatrix}$$

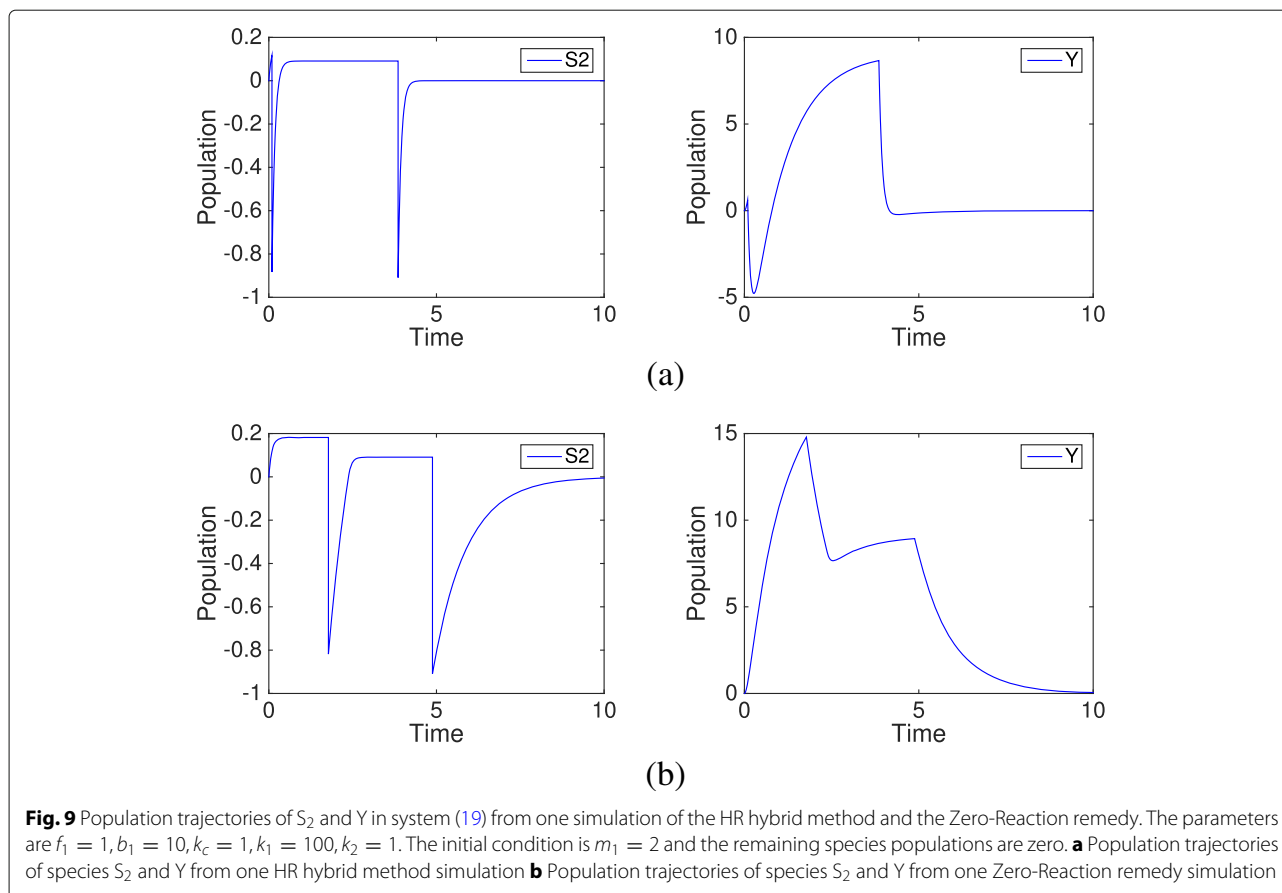
The determinant is

$$\begin{aligned} |\lambda I - J| &= \begin{vmatrix} \lambda + f_1 & -2b_1 x_2 \\ -2f_1 & \lambda + 4b_1 x_2 \end{vmatrix} \\ &= \lambda(\lambda + f_1 + 4b_1 x_2) \end{aligned}$$

The two eigenvalues of the ODE system are $\lambda_1 = 0$, $\lambda_2 = -f_1 - 4b_1 x_2$. For $x_2 > 0$, $\lambda_2 < 0$, so the fast system is stable. But in the HR hybrid method, x_2 could be negative under certain conditions. Let $\frac{dx_2}{dt} = 2f_1 x_1 - 2b_1 x_2^2 = 0$, species S_2 has the two equilibrium points:

$$x_2^* = -a \pm \sqrt{a^2 + am}, \quad a = \frac{f_1}{4b_1} \tag{21}$$

where m is the initial total population. One equilibrium point is positive, thus it is a stable point in the system. But for the other point, since $x_2^* = -a - \sqrt{a^2 + am} < -2a < -\frac{f_1}{4b_1}$, $\lambda_2 > 0$, it is an unstable point. So the HR hybrid method fails in this nonlinear system when the population of S_2 gets smaller than $-\frac{f_1}{4b_1}$ by chance.

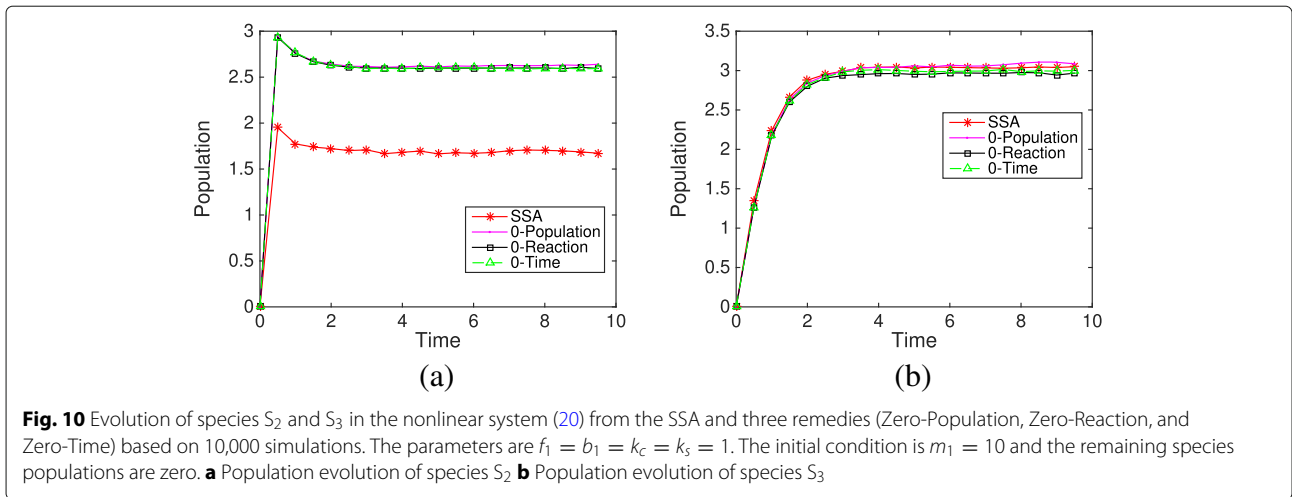


The above analysis is consistent with our simulation results. In our experiments, a simulation is considered a failure when the ODE solver is unable to meet integration tolerances with the smallest step length, or when one of the species' population reaches an extremely abnormal value, for example if a species' population becomes abnormally large (e.g. 1000) or below the negative value of the total population ($-m$). In Fig. 10, with an initial condition $m_1 = 10$ and $m_2 = 0$, it is found that even if x_2 has a probability less than 1% to be negative (see Fig. 11a), the system still suffers a significant error and breaks down after certain simulation time when using the original HR hybrid method. Particularly, 191 simulations of the original HR hybrid method failed among the total 10,000 trials (each trial runs from time $t = 0$ to $t = 10$). In Fig. 10, while the evolution of S_3 from the three rules is pretty close to that of the SSA, there is an approximate one molecule difference in the S_2 population between the remedy rules and the SSA, which mainly comes from the method error rather than the influence of negative value of x_2 . The final distributions of species S_3 from the Zero-Reaction and Zero-Time rules are close to the bell shape of the SSA results. Although the Zero-Population remedy did not fail in the simulation, the results are quite erroneous. Note

that in Fig. 10, the population of S_2 and S_3 from the Zero-Population rule does increase slowly with time. If we run the system to a much larger time (e.g. $t = 1000$), S_3 can reach 30, as shown in the final distribution of Fig. 11d.

When decreasing the total population to $m = 3$, only the Zero-Reaction rule still works and generates stable results similar to the SSA except the approximate one molecule difference in S_2 population, see Figs. 12 and 13. The Zero-Time rule failed because the separate G'_f system (which is the fast subsystem in this case) is unstable. Among the 10,000 simulations where the system was simulated from time $t = 0$ to $t = 10$, the original HR hybrid method failed in 2468 trials, while the Zero-Time rule failed in 1302 trials.

In general, the system stability will be affected if negative species are involved in nonlinear reactions where either the corresponding reacting terms in the ODE system or the corresponding propensities in the SSA system are still positive. Through the above comparison, the Zero-Reaction rule shows its ability to avoid the instability of nonlinear systems caused by negative values and at the same time keeps the accuracy of HR hybrid method. In application, it is easy and efficient to implement.

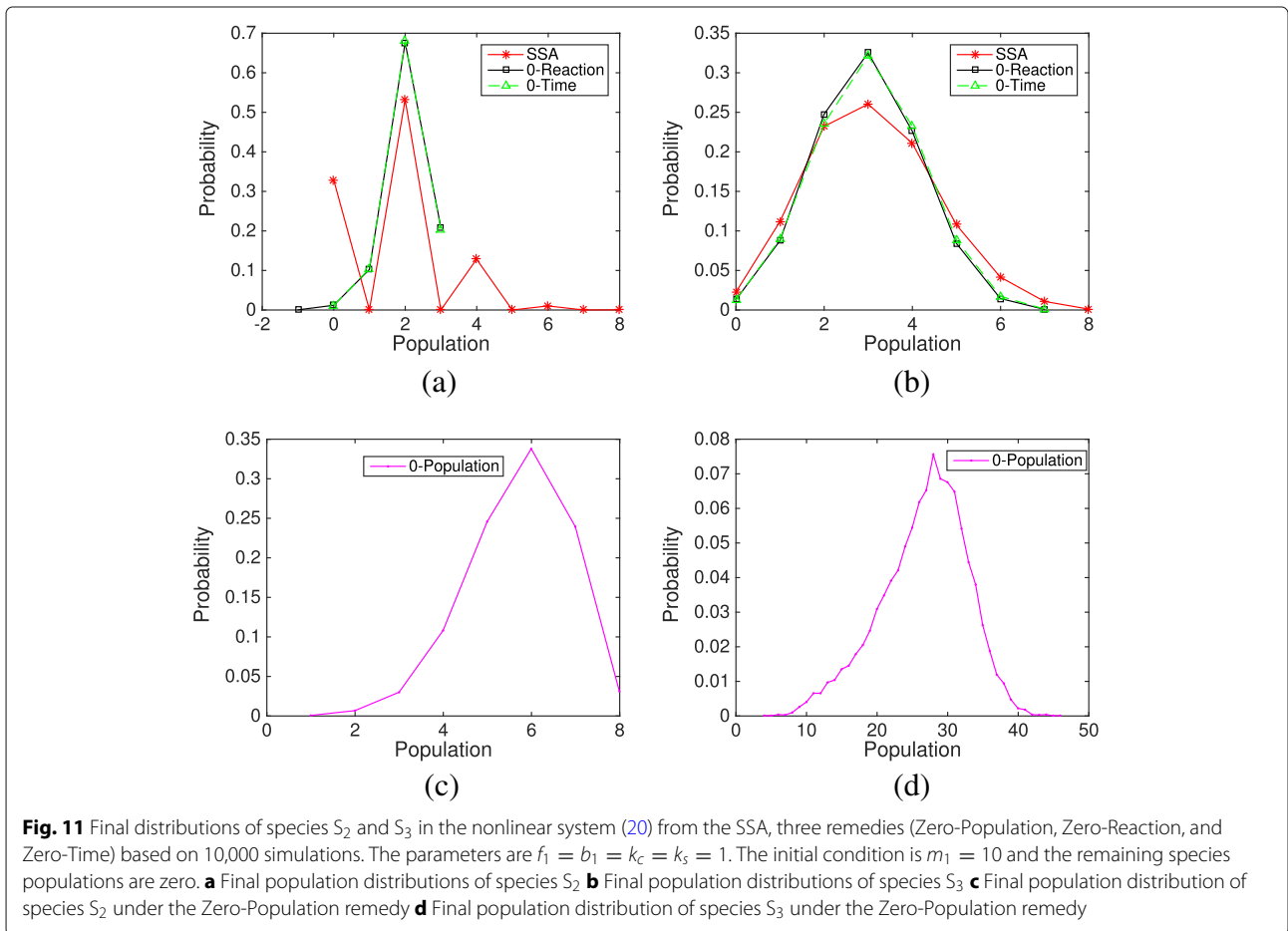


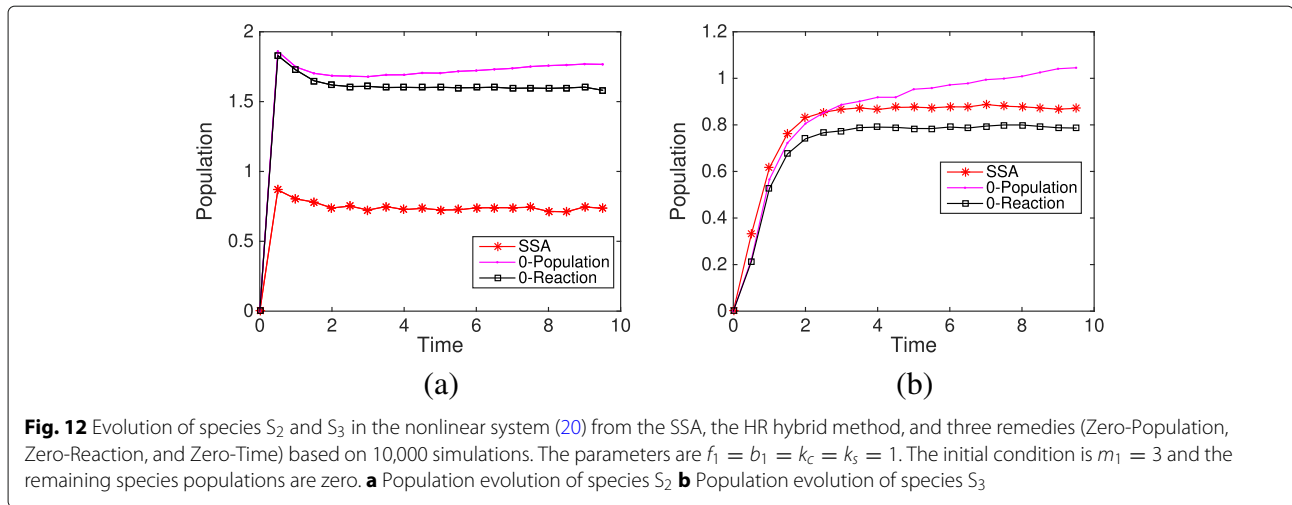
***Caulobacter crescentus* cell cycle model**

Caulobacter crescentus is a bacteria that lives in freshwater like streams and lakes. It has an asymmetrical division that produces two morphologically different daughter cells, which makes it an important study organism for cell cycle modeling. Li et al. [31] studied the stochastic spatiotemporal model of a response-regulator network in the

cell cycle. The stochastic model focused on the bistable switch of PleC that functioned as both kinase and phosphatase and successfully captured the viability of mutant cells. But it took the stochastic simulation three days for a single run.

To improve the efficiency, we applied the HR hybrid method and compared different partitioning strategies. In





the stochastic model, there are 141 reactions involving 45 species, in which eight proteins and their corresponding mRNAs are diffusive. The rod cell shape was modeled as $50 \times 1 \times 1$ cubics. In a test run, protein diffusion took 99.826% of the total number of reactions, putting it in the ODE system would greatly decrease the time cost. The catalytic reactions $\text{CtrA} \rightleftharpoons \text{CtrAp}$ took 0.148%. Compared to the diffusion of proteins, the diffusion of mRNAs only occupied 0.024%, while the remaining 140 reactions took 0.002%.

The computational cost of the hybrid method is proportional to the number of slow reaction firings, but is also affected by the size of the ODE subsystem. Based on the firing number of different reactions in the SSA simulation, we investigated three partitioning strategies as shown in Table 1. In Strategy I, only diffusion events of eight proteins are simulated by the ODE subsystem, the remaining events are put into the SSA subsystem. While Strategy II further partitions catalytic reactions of CtrA into the ODE subsystem, the size of the ODE subsystem does not

change (reactants and products of the catalytic reaction are diffusive and already included in the ODE subsystem). But the average slow reaction firing time is one order of magnitude less than Strategy I, which decreases the time cost by approximately a factor of ten. Strategy III partitions the system by species type: mRNA reactions are all in the SSA subsystem and protein reactions are in the ODE subsystem. This strategy greatly reduces the probability of negativity problems. On the other hand, although Strategy III has the least interruption by slow reactions (every $6e^{-5}$ min), its size for the ODE subsystem increases to 50 (bin number) \times 37 (types of species). This is quite a large ODE system, which imposes a high computational burden on the ODE solver. Overall, Strategy II is the most efficient partitioning strategy for the HR hybrid method in this *Caulobacter* cell cycle model, the time cost for a single cell cycle simulation is significantly reduced to one hour from three days!

However, as mentioned in the second example in the introduction, the negativity problem appears when

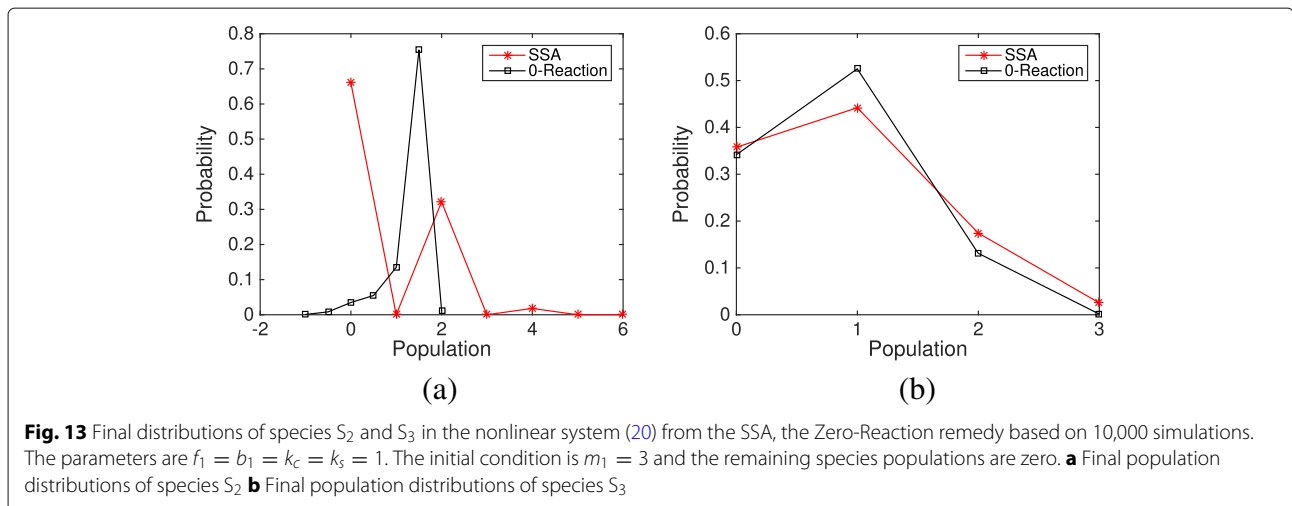


Table 1 Comparison of different partition strategies and complexities

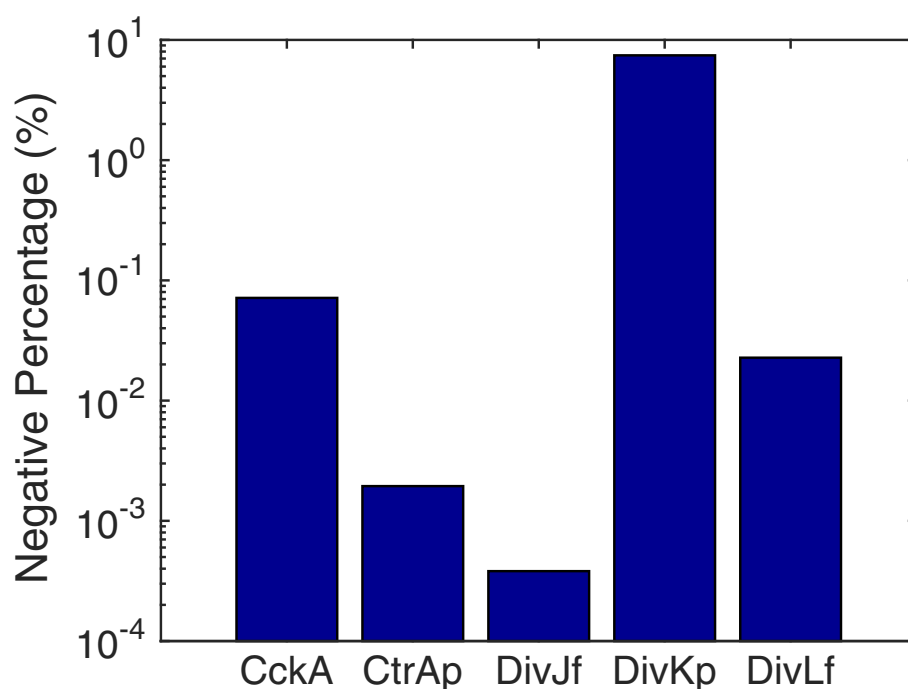
	Strategy I	Strategy II	Strategy III
ODE Reactions	Protein diffusion	Protein diffusion, catalytic reaction	Diffusion, reactions involving proteins
System Size	50× 8 equations	50× 8 equations	50× 37 equations
SSA Reactions	mRNA diffusion, all 141 reactions	mRNA diffusion, rest 140 reactions	mRNA diffusion, synthesis, degradation
Firing Interval	$1e^{-6}$	$2e^{-5}$	$6e^{-5}$
Running Time of HR hybrid Method	9.5h	1h	4h

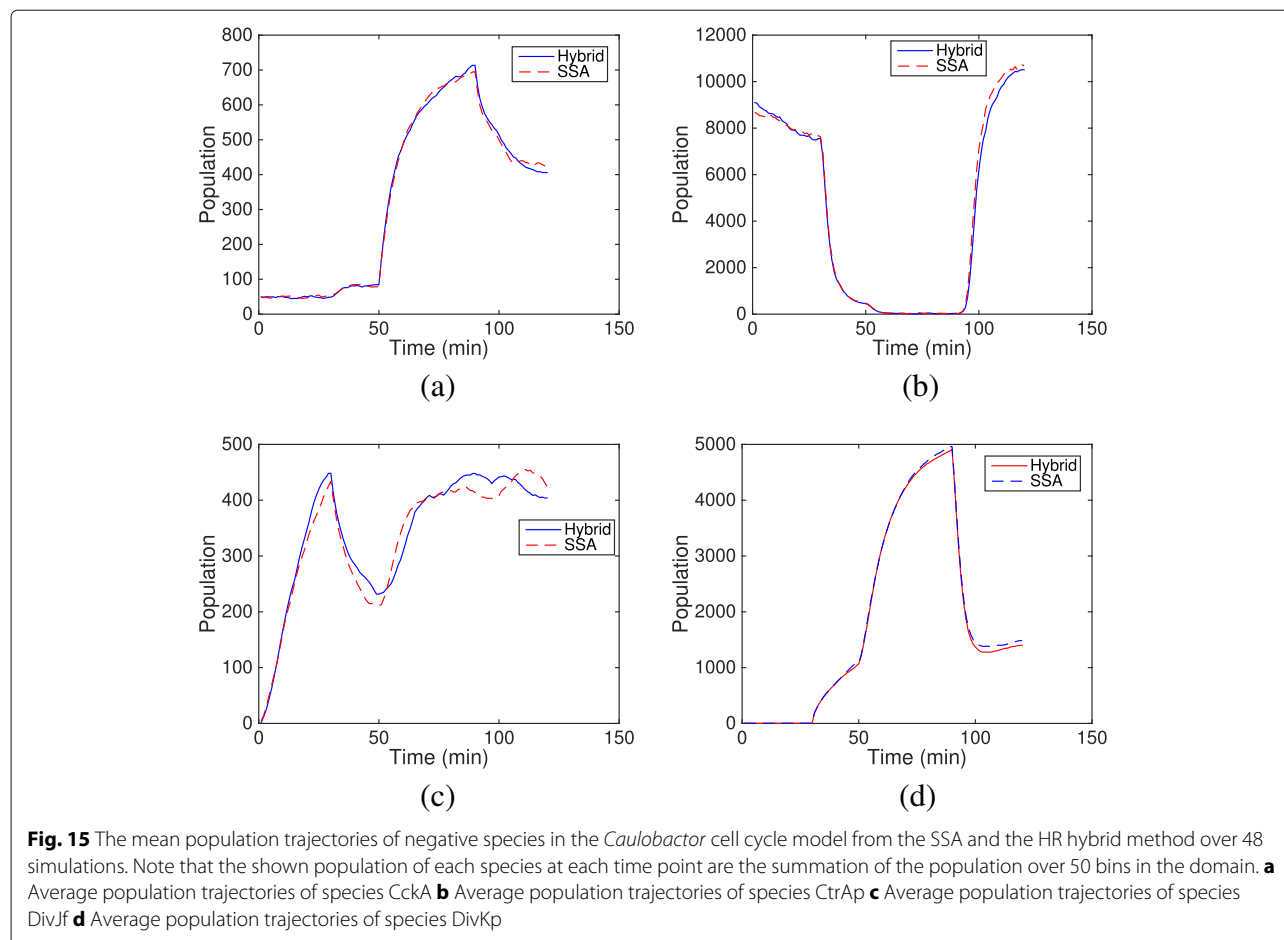
species density is low. Figure 14 summarizes the total time of negativity state for each species during one cell cycle (~ 120 min). It is observed that protein DivKp has a negative value for almost 10% of a cycle period, followed by proteins CckA, DivL(free), CtrAp, and DivJ(free), which are negative for less than 0.1% of the total time. Figure 15 shows the average population trajectories of four negative species from the SSA and the HR hybrid method using Strategy II. We can see that the hybrid method matches well with the SSA except for a slight difference in DivJ(free). It is also found that all species with negative values have a period of a low population during the cell cycle. The scarce density of DivKp (nearly zero for the initial 30 min) results in a high occurrence of negative value. Yet the negativity problem in this model has no significant impact on simulation accuracy because the diffusion of proteins happens much faster than chemical reactions (at least one order of magnitude faster). Whenever a bin has a negative population resulting from a slow reaction firing,

proteins in neighboring bins (with positive populations) quickly diffuse to the negative bin in the ODE system and make it positive before any chemical reaction happens. The HR hybrid method does not even need a remedy rule for the negativity problem in this case. But in general, the Zero-Reaction rule is recommended since the added computational cost is minimal but the potential impacts of the negativity problem can be avoided.

Conclusion

This paper presents an analysis on the negativity problem of the HR hybrid stochastic simulation algorithm. Based on the second slow reaction firing time, the error caused by negative populations is shown to be negligible compared to the approximation error of the method itself. In the linear chain system, the negativity phenomenon actually helps to increase the method's accuracy. But for nonlinear systems, negative values may lead to system failure. Three remedies for negativity are proposed and

**Fig. 14** The percentage of the cell cycle time where species have negative populations



studied in the context of SSRFT where Zero-Time and Zero-Reaction rules have acceptable accuracy. Particularly, the Zero-Reaction remedy can handle both extreme negative cases and nonlinear systems whereas the other two methods may fail. Without any remedy for negative populations, the HR hybrid method may still be successfully applied to a real biological network and significantly improves the efficiency via an optimized partition strategy. Overall, we conclude that the negativity phenomenon does not influence the biochemical network unless the negative species are involved in nonlinear reactions that generate positive reacting terms or propensities. In general, the Zero-Reaction remedy is recommended due to easy implementation and minimal additional computational cost.

Acknowledgements

Not applicable.

Funding

This work was partially supported by the National Science Foundation under awards CCF-0953590, CCF-1526666, and MCB-1613741.

Availability of data and materials

The accuracy and negativity analysis of different systems and examples are detailed in the main text.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 12, 2019*: Selected original research articles from the Fifth International Workshop on Computational Network Biology: Modeling, Analysis and Control (CNB-MAC 2018): Bioinformatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-12>.

Authors' contributions

YC conceived the idea and supervised the study. MC designed the study models, implemented the simulations, and analyzed the results. MC drafted the manuscript, and YC gave critical revisions on the writing. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 20 June 2019

References

- Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys.* 1976;22(4): 403–34.
- Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.* 1977;81(25):2340–61.
- Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Chem Phys.* 2000;104: 1876–89.
- Cao Y, Li H, Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys.* 2004;121(9): 4059–67.
- McCollum JM, Peterson GD, Cox CD, Simpson ML, Samatova NF. The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *Comput Biol Chem.* 2006;30: 39–49.
- Li H, Petzold L. Logarithmic Direct Method for discrete stochastic simulation of chemically reacting systems. Santa Barbara: University of California Santa Barbara; 2006.
- Slepoy A, Thompson AP, Plimpton SJ. A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *J Chem Phys.* 2008;128:205101.
- Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J Chem Phys.* 2007;127:214107.
- Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys.* 2001;115(4):1716–33.
- Davis MHA. Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models. *J R Stat Soc Ser B Methodol.* 1984;46(3): 353–88.
- Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys.* 2002;117(15):6959–69.
- Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J Chem Phys.* 2003;118(11):4999–5010.
- Cao Y, Gillespie DT, Petzold LR. Avoiding negative populations in explicit Poisson tau-leaping. *J Chem Phys.* 2005;123(5):054104.
- Franz U, Liebscher V, Zeiser S. Piecewise-Deterministic Markov Processes as limits of markov jump processes. *Adv Appl Probab.* 2012;44(3):729–48.
- Jahnke T, Kreim M. Error bound for piecewise deterministic processes modeling stochastic reaction systems. *Multiscale Model Simul.* 2012;10(4): 1119–47.
- Cao Y, Gillespie DT, Petzold LR. The slow-scale stochastic simulation algorithm. *J Chem Phys.* 2005;122(1):014116.
- Cao Y, Gillespie DT, Petzold LR. Multi-scale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J Chem Phys.* 2005;122(6):395–411.
- Sanft KR, LRP, Gillespie DT. Legitimacy of the stochastic Michaelis-Menten approximation. *IET Syst Biol.* 2011;5(1):58.
- Thomas P, Straube AV, Grima R. Communication: Limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks. *J Chem Phys.* 2011;135(18):181103.
- Salis H, Kaznessis Y. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J Chem Phys.* 2005;122(5): 054103.
- Liu Z, Pu Y, Li F, Shaffer CA, Hoops S, Tyson JJ, et al. Hybrid modeling and simulation of stochastic effects on progression through the eukaryotic cell cycle. *J Chem Phys.* 2012;136(3):034105.
- Lecca P, Bagagiolo F, Scarpa M. Hybrid deterministic/stochastic simulation of complex biochemical systems. *Mol BioSyst.* 2017;13: 2672–2686.
- Lo WC, Zheng L, Nie Q. A hybrid continuous-discrete method for stochastic reaction-diffusion processes. *R Soc Open Sci.* 2016;3(9):160485.
- Wang S, Chen M, Watson LT, Cao Y. Efficient implementation of the hybrid method for stochastic simulation of biochemical systems. *J Micromech Mol Phys.* 2017;02(02):1750006.
- Chiam KH, Tan CM, Bhargava V, Rajagopal G. Hybrid simulations of stochastic reaction-diffusion processes for modeling intracellular signaling pathways. *Phys Rev E.* 2006;74:051910.
- Rossinelli D, Bayati B, Koumoutsakos P. Accelerated stochastic and hybrid methods for spatial simulations of reaction-diffusion systems. *Chem Phys Lett.* 2008;451(1):136–40.
- Salis H, Sotiropoulos V, Kaznessis YN. Multiscale Hy3S: Hybrid stochastic simulation for supercomputers. *BMC Bioinformatics.* 2006 Feb;7(1):93.
- Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI—a complex pathway simulator. *Bioinformatics.* 2006;22:3067–74.
- Wang S, Ahmadian M, Chen M, Tyson JJ, Cao Y. A Hybrid Stochastic Model of the Budding Yeast Cell Cycle Control Mechanism. In: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. BCB '16. New York: ACM; 2016. p. 261–70.
- Chen M, Wang S, Cao Y. Accuracy analysis of hybrid stochastic simulation algorithm on linear chain reaction systems. *Bull Math Biol.* 2018. <https://doi.org/10.1007/s11538-018-0461-z>.
- Li F, Subramanian K, Chen M, Wang S, Cao Y. A stochastic spatiotemporal model of a response-regulator network in the *Caulobacter crescentus* cell cycle. *Phys Biol.* 2016;13(3):e0133295.
- Wang S, Cao Y. The abridgement and relaxation time for a linear multi-scale model based on multiple site phosphorylation. *PLoS ONE.* 2015;10(8):e0133295.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

