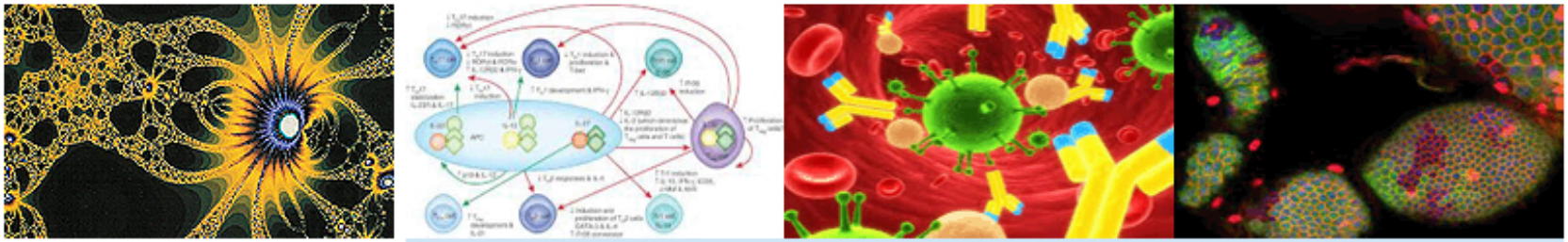


IBMC

CENTRO DE MATEMÁTICA UNIVERSIDADE DO PORTO

Building Bridges in Quantitative Biology
Joint seminar IBMC/CMUP

Mathematical models of the dynamical properties of biological systems aim to improve our understanding of the studied system with the ultimate goal of being able to predict system responses in the absence of experimentation.

Stochastic Simulation in Biology

A survey

João Nuno Tavares and Ricardo Cruz

CMUP

Jan 25, 2015

Contents

Abstract

Many processes in cell and molecular biology are comprised of **biochemical reaction networks**.
It has proven useful to study these networks using computer simulations because they allow us to quantitatively investigate hypotheses about the networks.
Deterministic simulations are sufficient to predict average behaviors at the population level, but they cannot address questions about noise, random switching between stable states of the system, or the behaviors of systems with very few molecules of key species.
These topics are investigated with stochastic simulations.
In this first **Bridges talk**, we will review, in an intuitive, and hopeffuly comprehensive manner, some key **Stochastic Simulation Algorithms** used in **Sistems Biology**. We will give several illustrations of these methods.

What are we talking about?

Mathematics: **Markov jump processes (MJP=CTMC)**

What are we talking about?

Mathematics: **Markov jump processes (MJP=CTMC)**

Biology. **Biochemical reaction networks (BRN)**.

MJP + BRN = SKN= Stochastic kinetic networks

What are we talking about?

Mathematics: **Markov jump processes (MJP=CTMC)**

Biology. **Biochemical reaction networks (BRN)**.

MJP + BRN = SKN= Stochastic kinetic networks

Applications ([9]) [Systems Biology, Computational Biology, Immunology, etc.]

What are we talking about?

Mathematics: **Markov jump processes (MJP=CTMC)**

Biology. **Biochemical reaction networks (BRN)**.

MJP + BRN = SKN= Stochastic kinetic networks

Applications ([9]) [Systems Biology, Computational Biology, Immunology, etc.]

- Molecule Synthesis and Degradation
- Enzymatic Reactions
- Receptor-Ligand Interaction
- Gene Expression and Regulation
- Ion Channel Dynamics and Ion Transport Across Membranes
- Immunological processes (inter and intra cellular)
- Epidemiological models

Deterministic/Stochastic

The cellular machinery differs a lot from the machines we see in our everyday life:

- Cells are microscopic reactors. On a microscopic level, individual molecules are constantly formed and destroyed by chemical reactions.

Deterministic/Stochastic

The cellular machinery differs a lot from the machines we see in our everyday life:

- Cells are microscopic reactors. On a microscopic level, individual molecules are constantly formed and destroyed by chemical reactions.
- Proteins and other molecules tumble back and forth, diffuse, change their conformations, assemble and disassemble in permanent thermal movement.

Deterministic/Stochastic

The cellular machinery differs a lot from the machines we see in our everyday life:

- Cells are microscopic reactors. On a microscopic level, individual molecules are constantly formed and destroyed by chemical reactions.
- Proteins and other molecules tumble back and forth, diffuse, change their conformations, assemble and disassemble in permanent thermal movement.
- Biochemical reactions are probabilistic collisions between randomly moving molecules, with each event resulting in the increment or decrement of molecular species by integer amounts. They can be described mathematically by random processes: reactions happen unpredictably, and each sequence of random events leads to a different history of the system.

Deterministic/Stochastic

- The amplified effect of fluctuations in a molecular reactant, or the compounded of fluctuations across many molecular reactants, referred to as molecular noise, often can accumulate as an observable phenotype.

Deterministic/Stochastic

- The amplified effect of fluctuations in a molecular reactant, or the compounded of fluctuations across many molecular reactants, referred to as molecular noise, often can accumulate as an observable phenotype.
- Stochastic models allow us to compute mean values, fluctuations, and temporal correlations of system states.

Deterministic/Stochastic

- The amplified effect of fluctuations in a molecular reactant, or the compounded of fluctuations across many molecular reactants, referred to as molecular noise, often can accumulate as an observable phenotype.
- Stochastic models allow us to compute mean values, fluctuations, and temporal correlations of system states.
- Furthermore, individual realizations of random processes can often be obtained by **stochastic simulation**.

Deterministic/Stochastic

1. On larger space and time scales, the microscopic processes translate into an effective macroscopic behavior, for instance, the dynamics of metabolic pathways governed by kinetic laws.

Deterministic/Stochastic

1. On larger space and time scales, the microscopic processes translate into an effective macroscopic behavior, for instance, the dynamics of metabolic pathways governed by kinetic laws.

2. Random models provide a more detailed description than the deterministic kinetic models. Whenever the random fluctuations remain small, deterministic models provide a good and numerically cheap approximation.

Deterministic/Stochastic

1. On larger space and time scales, the microscopic processes translate into an effective macroscopic behavior, for instance, the dynamics of metabolic pathways governed by kinetic laws.

2. Random models provide a more detailed description than the deterministic kinetic models. Whenever the random fluctuations remain small, deterministic models provide a good and numerically cheap approximation.

3. However, random fluctuations can become important if molecule numbers are low, which typically happens in gene expression, or in models with nonlinear and unstable dynamics.

General settings

- Consider a well-stirred system of "molecules" of s species $\{\mathscr{X}_1, \cdots, \mathscr{X}_s\}$, which interact through r reaction channels $\{\mathscr{R}_1, \cdots, \mathscr{R}_r\}$. We assume the system in thermal equilibrium and confined to a finite constant volume V.

General settings

- Consider a well-stirred system of "molecules" of s species $\{\mathscr{X}_1, \cdots, \mathscr{X}_s\}$, which interact through r reaction channels $\{\mathscr{R}_1, \cdots, \mathscr{R}_r\}$. We assume the system in thermal equilibrium and confined to a finite constant volume V.
- Consider the following reaction network

$$\mathscr{R}_\mu : \underbrace{\sum_{i=1}^{s} \alpha_{i\mu} \mathscr{X}_i}_{\text{reactants}} \xrightarrow{\kappa_\mu} \underbrace{\sum_{i=1}^{s} \beta_{i\mu} \mathscr{X}_i}_{\text{products}}, \qquad \mu = 1, 2, \cdots, r$$

General settings

- Consider a well-stirred system of "molecules" of s species $\{\mathscr{X}_1, \cdots, \mathscr{X}_s\}$, which interact through r reaction channels $\{\mathscr{R}_1, \cdots, \mathscr{R}_r\}$. We assume the system in thermal equilibrium and confined to a finite constant volume V.
- Consider the following reaction network

$$\mathscr{R}_\mu : \underbrace{\sum_{i=1}^{s} \alpha_{i\mu} \mathscr{X}_i}_{\text{reactants}} \xrightarrow{\kappa_\mu} \underbrace{\sum_{i=1}^{s} \beta_{i\mu} \mathscr{X}_i}_{\text{products}}, \qquad \mu = 1, 2, \cdots, r$$

- Define the **stoichiometric matrix** $\mathbf{S} = (S_{i\mu})$, where

$$S_{i\mu} = \beta_{i\mu} - \alpha_{i\mu}, \quad i = 1, \cdots, s; \mu = 1, \cdots, r$$

General settings

- **Let**:

$$N_i(t) \quad = \quad \text{number of molecules of species } \mathscr{X}_i \text{ present in instant } t \tag{1}$$

General settings

- **Let**:

$$N_i(t) \quad = \quad \text{number of molecules of species } \mathscr{X}_i \text{ present in instant } t \tag{1}$$

- **Aim**. Compute the state vector time evolution $\mathbf{N}(t) = (N_1(t), \cdots, N_s(t))$, assuming the initial state $\mathbf{N}(0) = \mathbf{n}_o$.

General settings

- **Let**:

$$N_i(t) \quad = \quad \text{number of molecules of species } \mathscr{X}_i \text{ present in instant } t \tag{1}$$

- **Aim**. Compute the state vector time evolution $\mathbf{N}(t) = (N_1(t), \cdots, N_s(t))$, assuming the initial state $\mathbf{N}(0) = \mathbf{n}_o$.
- We can also use other state variable descriptions

$$\begin{aligned} \mathbf{X}(t) &= (X_1(t), \cdots, X_s(t))^\top \quad \text{(real concentrations)} \\ \mathbf{Z}(t) &= (Z_1(t), \cdots, Z_r(t))^\top \quad \text{(event reaction counts on } [0, t]) \end{aligned}$$

General settings

- each time reaction $\mathscr{R}_\mu$ triggers, the $\mathscr{X}_i$ state changes according to

$$N_i \longrightarrow N_i + S_{i\mu}$$

General settings

- each time reaction $\mathscr{R}_\mu$ triggers, the $\mathscr{X}_i$ state changes according to

$$N_i \longrightarrow N_i + S_{i\mu}$$

- So, the system update is made through

$$\Delta \mathbf{N}(t) = \mathbf{S} \Delta \mathbf{Z}(t)$$

General settings

- each time reaction $\mathscr{R}_\mu$ triggers, the $\mathscr{X}_i$ state changes according to

$$N_i \longrightarrow N_i + S_{i\mu}$$

- So, the system update is made through

$$\Delta \mathbf{N}(t) = \mathbf{S}\Delta \mathbf{Z}(t)$$

- The stoichiometry matrix therefore encodes important structural information about the reaction network. In particular, vectors in the left null-space of $\mathbf{S}$ correspond to **conservation laws** in the network: any s-vector $\mathbf{a}$ that satifies $\mathbf{a}^\top \mathbf{S} = 0$ has the property that $\mathbf{N}(t)$ remains constant for all t.

General settings

- Assuming very big numbers, we can approximate **N**, **X** and **Z** by real (t-smooth dependent) numbers **n**, **x** and **z** and write ODE's

$$\dot{\mathbf{n}}(t) = \mathbf{S}\dot{\mathbf{z}}(t), \qquad \dot{\mathbf{x}}(t) = \mathbf{S}\frac{\dot{\mathbf{z}}(t)}{V},$$

General settings

- Assuming very big numbers, we can approximate **N**, **X** and **Z** by real (t-smooth dependent) numbers **n**, **x** and **z** and write ODE's

$$\dot{\mathbf{n}}(t) = \mathbf{S}\dot{\mathbf{z}}(t), \qquad \dot{\mathbf{x}}(t) = \mathbf{S}\frac{\dot{\mathbf{z}}(t)}{V},$$

- However, these ODE's are only useful if we can establish a relationship between the derivative $\dot{\mathbf{z}}(t)$ and the variables **n** or **x**.

$$\dot{\mathbf{z}}(t) = \widehat{\upsilon}(\mathbf{n}) = V\ \upsilon(\mathbf{x})$$

Mass-action law

- Usually we adopt the so called **mass-action law (MAL)**

$$\widehat{v}_\mu(\mathbf{n}) = \widehat{k}_\mu \prod_{i=1}^{s} n_i^{\alpha_{i\mu}}, \qquad v_\mu(\mathbf{x}) = k_\mu \prod_{i=1}^{s} x_i^{\alpha_{i\mu}}$$

which leads to **RRE (reaction rate equations)** according to MAL

$$\dot{\mathbf{n}}(t) = \mathbf{S}\ \widehat{v}(\mathbf{n}(t)), \qquad \dot{\mathbf{x}}(t) = \mathbf{S}\ v(\mathbf{x}(t)),$$

An example: Lotka-Volterra Predator-prey model

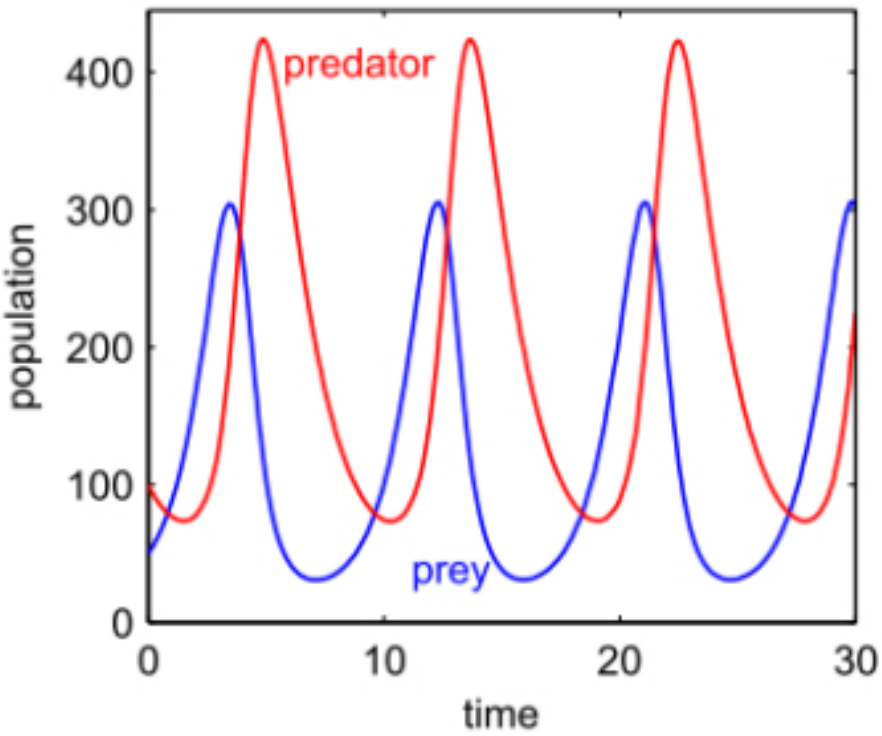
Species: X_1=prey; X_2=predator. $n_1(t), n_2(t)$ the corresponding copy numbers. The number n_A of food items is assumed unchanged.

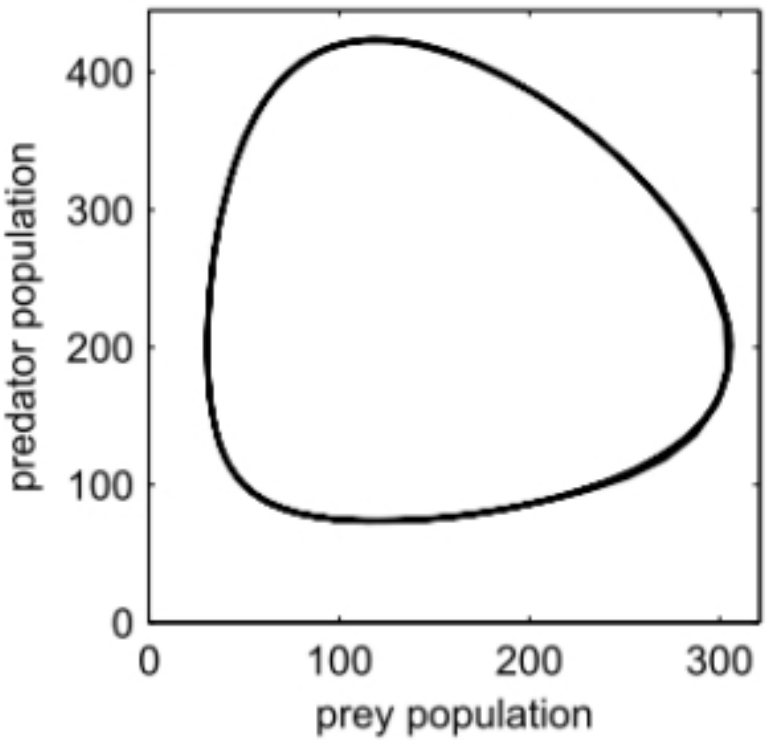
- The reaction network and action-mass conversion rates are

$$\begin{array}{rcl||rcl} X_1 + A & \xrightarrow{\widehat{k}_1} & 2X_1 & \widehat{v}_1 & = & \widehat{k}_1 n_1 n_A \\ X_1 + X_2 & \xrightarrow{\widehat{k}_2} & 2X_2 & \widehat{v}_2 & = & \widehat{k}_2 n_1 n_2 \\ X_2 & \xrightarrow{\widehat{k}_3} & \emptyset & \widehat{v}_3 & = & \widehat{k}_3 n_2 \end{array}$$

- We will see later the SSA.

An example: Lotka-Volterra Predator-prey model





An example: SIRS epidemic model

- The reaction network and action-mass conversion rates are

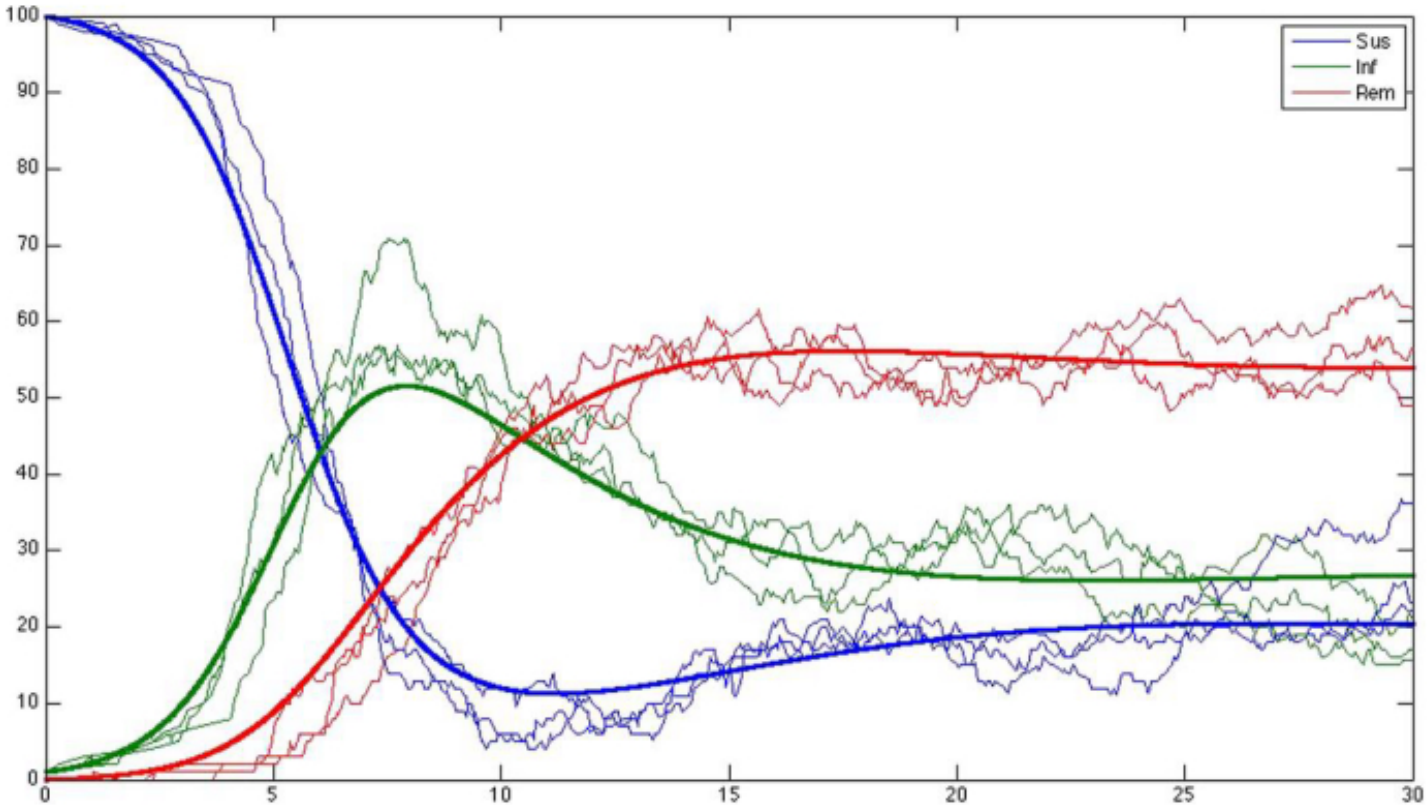
$$\begin{array}{rcl||rcl} S+I & \xrightarrow{\widehat{k}_1} & 2I & \widehat{v}_1 & = & \widehat{k}_1 n_S n_I \\ I & \xrightarrow{\widehat{k}_2} & R & \widehat{v}_2 & = & \widehat{k}_2 n_I \\ R & \xrightarrow{\widehat{k}_3} & S & \widehat{v}_3 & = & \widehat{k}_3 n_R \end{array}$$

- The RRE's are

$$\begin{cases} \dot{n}_S &= -\widehat{k}_1 n_S n_I + \widehat{k}_3 n_R \\ \dot{n}_I &= \widehat{k}_1 n_S n_I - \widehat{k}_2 n_I \\ \dot{n}_R &= \widehat{k}_2 n_I - \widehat{k}_3 n_R \end{cases}$$

We will see later the SSA.

An example: SIRS epidemic model



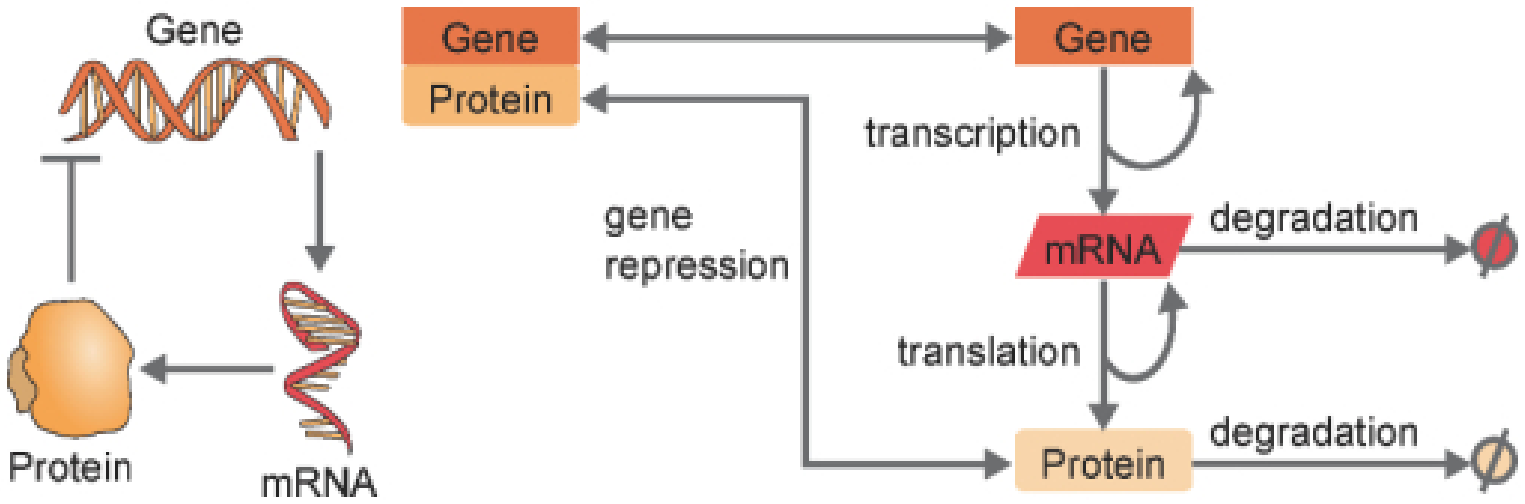
An example: Michaelis-Menten enzyme kinetics

The reaction network is

$$\begin{aligned} S + E &\rightleftharpoons C \\ C &\rightarrow P + E \end{aligned}$$

where S is the substance which is transformed by the reaction, E is the enzyme which facilitates the reaction, C is an intermediary species and P the final product.

An example: Gene regulation



where the gene G is transcribed to the mRNA M with rate constant k_m , the mRNA is translated to the protein P with rate constant k_p, and the protein binds to (and represses) the gene with rate constant k_b and unbinds back with rate constant k_u. The mRNA and protein are degraded with respective rate constants k_m^- and k_p^-.

An example: Gene regulation

The reaction network is

G	$\xrightarrow{k_m}$	$G + M$	υ_1	$=$	$k_m x_G$	transcription
M	$\xrightarrow{k_p}$	$M + P$	υ_2	$=$	$k_p x_M$	traslation
$G + P$	$\xrightarrow{k_b}$	$G \cdot P$	υ_3	$=$	$k_b x_G x_P$	binding
$G \cdot P$	$\xrightarrow{k_u}$	$G + P$	υ_4	$=$	$k_u(x_G^{\text{total}} - x_G)$	unbinding
M	$\xrightarrow{k_m^-}$	$\emptyset$	υ_5	$=$	$k_m^- x_M$	degradation
P	$\xrightarrow{k_p^-}$	$\emptyset$	υ_6	$=$	$k_p^- x_P$	degradation

An example: HIV infection

The reaction network is

$$\begin{aligned} \emptyset &\xrightarrow{s} T \\ T &\xrightarrow{\delta} \emptyset \\ T + V &\xrightarrow{\kappa} T^* \\ T^* &\xrightarrow{\eta} NV \\ V &\xrightarrow{\gamma} \emptyset \end{aligned}$$

where T is a CD4+ T-cell, V is virus and T^* is an infected active T-cell. Refering to concentrations $copies/\mu L$, we have the RRE's

$$\begin{cases} \dot{x}_T &= s - \delta x_T - \kappa x_V x_T \\ \dot{x}_{T^*} &= \kappa x_V x_T - \eta x_{T^*} \\ \dot{x}_V &= N\eta x_{T^*} - \gamma x_V \end{cases}$$

Deterministic/Stochastic

Even though the deterministic approach has proven very successful, it comes with some issues:

1. Small particle numbers in cellular subsystems (e.g. in signaling pathways) lead to random fluctuations which can change the dynamic behavior considerably.

Deterministic/Stochastic

Even though the deterministic approach has proven very successful, it comes with some issues:

1. Small particle numbers in cellular subsystems (e.g. in signaling pathways) lead to random fluctuations which can change the dynamic behavior considerably.
2. Bi- or multi-stable systems can not be described adequately.

Deterministic/Stochastic

Even though the deterministic approach has proven very successful, it comes with some issues:

1. Small particle numbers in cellular subsystems (e.g. in signaling pathways) lead to random fluctuations which can change the dynamic behavior considerably.
2. Bi- or multi-stable systems can not be described adequately.
3. Stochasticity itself can be an important property of the system, e.g. in evolution, noise-induced amplification of signals or noise-driven divergence of cell fates.

Deterministic/Stochastic

Even though the deterministic approach has proven very successful, it comes with some issues:

1. Small particle numbers in cellular subsystems (e.g. in signaling pathways) lead to random fluctuations which can change the dynamic behavior considerably.
2. Bi- or multi-stable systems can not be described adequately.
3. Stochasticity itself can be an important property of the system, e.g. in evolution, noise-induced amplification of signals or noise-driven divergence of cell fates.
4. For very small particle numbers (e.g. single genes) the concept of continuous concentrations is not appropriate.

Fundamental assumptions

- **Hypothesis 1.** The chemical system is under thermal equilibrium conditions.

Fundamental assumptions

- **Hypothesis 1.** The chemical system is under thermal equilibrium conditions.
- **Hypothesis 2.** The chemical system is such that, at any time t, the concentration of each species is homogeneous in the reaction vessel (i.e., does not depend on space).
  Homogeneity is in fact achieved if nonreactive collisions are much more frequent than reactive ones, which ensures diffusion processes proceed at much higher rate than any reaction in the system.

Fundamental assumptions

- **Hypothesis 1.** The chemical system is under thermal equilibrium conditions.
- **Hypothesis 2.** The chemical system is such that, at any time t, the concentration of each species is homogeneous in the reaction vessel (i.e., does not depend on space).
  Homogeneity is in fact achieved if nonreactive collisions are much more frequent than reactive ones, which ensures diffusion processes proceed at much higher rate than any reaction in the system.
- **Hypothesis 3.** In a bimolecular reaction, the time to the occurrence of the reaction is largely determined by the time to the reactive collision whereas the time necessary for the chemical transformation of the colliding species into the reaction products is negligible.

Discussion

To each reaction channel $\mathscr{R}_\mu$ we associate two quantities:

- the vector $\boldsymbol{\vartheta}_\mu = \mathbf{S}_{\cdot\mu}$ - stoichiometric matrix column μ. Thus, if the sytem is in state $\mathbf{n}$ and if reaction $\mathscr{R}_\mu$ triggers, the sytem changes instantly its state to $\mathbf{n} + \boldsymbol{\vartheta}_\mu$.

Discussion

To each reaction channel $\mathscr{R}_\mu$ we associate two quantities:

- the vector $\boldsymbol{\vartheta}_\mu = \mathbf{S}_{\cdot\mu}$ - stoichiometric matrix column μ. Thus, if the sytem is in state $\mathbf{n}$ and if reaction $\mathscr{R}_\mu$ triggers, the sytem changes instantly its state to $\mathbf{n} + \boldsymbol{\vartheta}_\mu$.
- the **propensity** or **hazard** $h_\mu(\mathbf{n}, c_\mu)$, defined by

$$h_\mu(\mathbf{n}, c_\mu)dt \quad = \quad \text{probability that, assuming that } \mathbf{N}(t) = \mathbf{n}, \text{ reaction } \mathscr{R}_\mu \text{ triggers in interval } [t, t+dt) \text{ somewhere in volume } V. \tag{2}$$

Here c_μ is the **stochastic rate constant**.

Examples of reactions

Reaction	Propensity $h_\mu(\mathbf{n}, c_\mu)$	Description
$\emptyset \longrightarrow \mathscr{X}_1$	c_1	Inflow
$\mathscr{X}_1 \longrightarrow \emptyset$	$c_2 n_1$	Degradation
$\mathscr{X}_1 + \mathscr{X}_2 \longrightarrow \mathscr{X}_3$	$c_3 n_1 n_2$	Catalysation
$2\mathscr{X}_1 \longrightarrow \mathscr{X}_2$	$c_4 n_1 (n_1 - 1)/2$	Dimerisation
$3\mathscr{X}_1 \longrightarrow \mathscr{X}_3$	$c_5 n_1 (n_1 - 1)(n_1 - 2)/6$	Trimerisation

Chemical Master Equation(CME)

The first thing we can do is to deduce a **master equation** (Kolmogoroff-Chapmann).

- Let

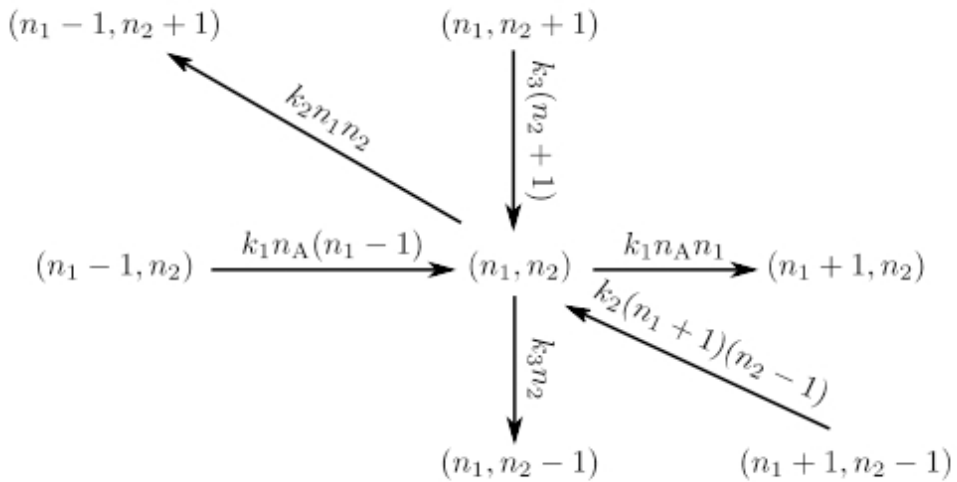
$$P(\mathbf{n}, t) = \text{Prob}\{\mathbf{N}(t) = \mathbf{n} | \mathbf{N}(0) = \mathbf{n}_o\}$$

- We write $P(\mathbf{n}, t + dt)$ as the sum of the probabilities of the number of ways in which the network can arrive in state $\mathbf{n}$ at time $t + dt$

$$\begin{aligned} P(\mathbf{n}, t + dt) &= o(dt) + \left(1 - \sum_{\mu=1}^{r} h_\mu(\mathbf{n}, c_\mu) dt\right) P(\mathbf{n}, t) + \\ &\quad + \sum_{\mu=1}^{r} h_\mu(\mathbf{n} - \boldsymbol{\vartheta}_\mu, c_\mu) dt \; P(\mathbf{n} - \boldsymbol{\vartheta}_\mu, t) \end{aligned}$$

Chemical Master Equation(CME)

- The first quantity in is the probability that the system undergoes no reactions in $[t, t + dt)$
- The term $h_\mu(\mathbf{n} - \boldsymbol{\vartheta}_\mu, c_\mu)P(\mathbf{n} - \boldsymbol{\vartheta}_\mu, t)\ dt$ is the probability that the system is one $\mathscr{R}_\mu$ reaction removed from state $\mathbf{n}$ at time t and then undergoes such a reaction in $[t, t + dt)$.



State transitions of the Lotka–Volterra model

Chemical Master Equation(CME)

The CME is then

$$\frac{dP(\mathbf{n},t)}{dt} = \sum_{\mu=1}^{r} [h_\mu(\mathbf{n}-\boldsymbol{\vartheta}_\mu, c_\mu)\ P(\mathbf{n}-\boldsymbol{\vartheta}_\mu, t) - h_\mu(\mathbf{n}, c_\mu)\ P(\mathbf{n}, t)]$$

Chemical Master Equation(CME)

The CME is then

$$\frac{dP(\mathbf{n},t)}{dt} = \sum_{\mu=1}^{r} [h_\mu(\mathbf{n}-\boldsymbol{\vartheta}_\mu, c_\mu)\ P(\mathbf{n}-\boldsymbol{\vartheta}_\mu, t) - h_\mu(\mathbf{n}, c_\mu)\ P(\mathbf{n}, t)]$$

- This a set of linear, autonomous ODEs. One ODE for each possible **n**-state of the system. Solution of the **n** th equation at time t gives the probability of system being in that particular state at time t.

Chemical Master Equation(CME)

The CME is then

$$\frac{dP(\mathbf{n},t)}{dt} = \sum_{\mu=1}^{r} [h_\mu(\mathbf{n}-\boldsymbol{\vartheta}_\mu, c_\mu)\ P(\mathbf{n}-\boldsymbol{\vartheta}_\mu, t) - h_\mu(\mathbf{n}, c_\mu)\ P(\mathbf{n}, t)]$$

- This a set of linear, autonomous ODEs. One ODE for each possible **n**-state of the system. Solution of the **n** th equation at time t gives the probability of system being in that particular state at time t.
- Unfortunately, the CME is only tractable for a handful of cases. Hence, for most systems of interest, an analysis via the CME will not be possible and then stochastic simulation techniques will present the only practical approach to gaining insight into a system's dynamics.

Gillespie Algorithm (SSA)

Typically, the CME is too high dimensional to deal with computationally. The SSA gets around this issue by computing single realizations of the state vector rather than an entire probability distribution.

Gillespie Algorithm (SSA)

Typically, the CME is too high dimensional to deal with computationally. The SSA gets around this issue by computing single realizations of the state vector rather than an entire probability distribution.

The key point to design SSA is to compute the joint density function for the two random variables μ and τ. We can prove that this joint density is the product of two individual density functions.

- **Next reaction index μ** - discrete pdf $\frac{h_\mu(\mathbf{n}, c_\mu)}{h_0(\mathbf{n}, c)}$, ie, choose one of the reactions with the rule that the chance of picking the μth reaction is proportional to its propensity $h_\mu(\mathbf{n}, c_\mu)$. Here $h_0(\mathbf{n}, c) = \sum_{\mu=1}^{r} h_\mu(\mathbf{n}, c_\mu)$.
- **Time τ until next reaction.** With exponential pdf $h_0(\mathbf{n}, c)e^{-h_0(\mathbf{n},c)\tau}$.

Gillespie Algorithm (SSA)

- **1.** Initialize in $t = 0$, with molecular numbers $\mathbf{n} = (n_1, \cdots, n_s)$ and stochastic constant rates $c = (c_1, \cdots, c_r)$.
- **2.** For each $\mu = 1, \cdots, r$, compute $h_\mu(\mathbf{n}, c_\mu)$, based in actual state $\mathbf{n}$, and $h_0(\mathbf{n}, c)$. STOP if $h_0 = 0$.
- **4.** Simulate time τ until next reaction $\tau \sim \text{Exp}(h_0(\mathbf{n}, c))$.
- **5.** Simulate the type μ of next reaction using $(h_\mu(\mathbf{n}, c_\mu)/h_0(\mathbf{n}, c))_{\mu=1,\cdots,r}$.
- **6.** Upadate time $t = t + \tau$.
- **7.** Update state $\mathbf{n}$ according to $\mathbf{n} = \mathbf{n} + \boldsymbol{\vartheta}_\mu$.
- **8.** If $t < t_{\text{max}}$, go to step 2.

τ-leaping

Assume that there exists $\tau > 0$ that satisfies the following **Leap condition** : "$\forall \mu, h_\mu(\mathbf{n}, c_\mu) \approx$ constant in interval $[t, t+\tau)$".

- We then can show that each reaction count number is Poisson with mean (and variance) $h_\mu(\mathbf{n}, c_\mu)\tau$.
- So, we can leap the system by τ time units

$$\mathbf{N}(t+\tau) = \mathbf{n} + \sum_{\mu=1}^{r} \mathcal{P}_\mu(h_\mu(\mathbf{n}, c_\mu)\tau)\ \boldsymbol{\vartheta}_\mu$$

with $\mathbf{N}(t) = \mathbf{n}$ and $\mathcal{P}_\mu(m_\mu)$ independent Poisson RV with mean (and variance) $m_\mu = h_\mu(\mathbf{n}, c_\mu)\tau$.

τ-leaping algorithm

- **1.** Initialize in $t = 0$, with molecular numbers $\mathbf{n} = (n_1, \cdots, n_s)$ and stochastic constant rates $c = (c_1, \cdots, c_r)$.
- **2.** For each $\mu = 1, \cdots, r$, compute $h_\mu(\mathbf{n}, c_\mu)$, based in actual state $\mathbf{n}$, and $h_0(\mathbf{n}, c)$. STOP if $h_0 = 0$.
- **3.** Simulate rv's $\{p_\mu\}_{\mu=1}^r$ from RVMs $\{\mathcal{P}_\mu(h_\mu(\mathbf{n}, c_\mu)\tau\}_{\mu=1}^r$.
- **4.** Update

$$\mathbf{N}(t+\tau) = \mathbf{n} + \sum_{\mu=1}^{r} p_\mu \, \boldsymbol{\vartheta}_\mu$$

  and $t = t + \tau$.
- **5.** Go to step 1.

Langevin equation (CLE)

- Assume that we choose τ so that

$$h_\mu(\mathbf{n}, c_\mu)\tau >> 1, \quad \forall \mu$$

- Then

$$\mathcal{P}_\mu(h_\mu(\mathbf{n}, c_\mu)\tau) \approx h_\mu(\mathbf{n}), c_\mu)\tau + \sqrt{h_\mu(\mathbf{n}, c_\mu)\tau}\ \mathcal{N}_\mu(0,1)$$

- and substituting, we get we get the update (Euler-Maruyama)

$$\begin{aligned}\mathbf{Y}(t+\tau) &= \mathbf{Y}(t) + \tau \sum_{\mu=1}^{r} \boldsymbol{\vartheta}_\mu\ h_\mu(\mathbf{Y}(t), c_\mu) + \\ &\sqrt{\tau} \sum_{\mu=1}^{r} \boldsymbol{\vartheta}_\mu \sqrt{h_\mu(\mathbf{Y}(t), c_\mu)}\ \mathcal{N}_\mu(0,1)\end{aligned}$$

Why we call this "diffusion approximation"?

- **Drift**. When the average value of a stochastic process changes with time, we say that the process has a **drift**.

Why we call this "diffusion approximation"?

- **Drift**. When the average value of a stochastic process changes with time, we say that the process has a **drift**.
- **Diffusion.** When the (co)variance, a measure of spread of the distribution, changes with time, we say that the process has **diffusion**.

Why we call this "diffusion approximation"?

- **Drift**. When the average value of a stochastic process changes with time, we say that the process has a **drift**.
- **Diffusion.** When the (co)variance, a measure of spread of the distribution, changes with time, we say that the process has **diffusion**.
- Let $\Delta Z_\mu(t) = Z_\mu(t + \Delta t) - Z_\mu(t$ be the **short time $\mathscr{R}_\mu$-reaction count increment**. Then

$$(\Delta Z_\mu)_{\mathbf{n}} \approx h_\mu(\mathbf{n}, c_\mu)\Delta t + \sqrt{h_\mu(\mathbf{n}, c_\mu)\Delta t}\, \mathcal{N}_\mu(t)$$

Why we call this "diffusion approximation"?

- **Drift**. When the average value of a stochastic process changes with time, we say that the process has a **drift**.
- **Diffusion.** When the (co)variance, a measure of spread of the distribution, changes with time, we say that the process has **diffusion**.
- Let $\Delta Z_\mu(t) = Z_\mu(t + \Delta t) - Z_\mu(t$ be the **short time $\mathscr{R}_\mu$-reaction count increment**. Then

$$(\Delta Z_\mu)_{\mathbf{n}} \approx h_\mu(\mathbf{n}, c_\mu)\Delta t + \sqrt{h_\mu(\mathbf{n}, c_\mu)\Delta t}\, \mathcal{N}_\mu(t)$$

and since $(\Delta N_i)_{\mathbf{n}} = \mathbf{S}\ (\Delta Z_\mu)_{\mathbf{n}}$, we obtain

$$(\Delta N_i)_{\mathbf{n}} \approx \sum_{\mu=1}^{r} S_{i\mu} h_\mu(\mathbf{n}, c_\mu)\Delta t + \sum_{\mu=1}^{r} S_{i\mu} \sqrt{h_\mu(\mathbf{n}, c_\mu)\Delta t}\, \mathcal{N}_\mu(t)$$

CLE. Langevin equation

- The increment vector then takes the form

$$(\Delta \mathbf{N})_{\mathbf{n}} \approx A(\mathbf{n})\Delta t + D(\mathbf{n})\sqrt{\Delta t}\,\mathcal{N}(t)$$

- The factor $\sqrt{\Delta t}\,\mathcal{N}(t)$ in the second summation on the right can be recognized as the Wiener increment,

$$\Delta \mathcal{W} = \mathcal{W}(t+\Delta t) - \mathcal{W}(t) = \sqrt{\Delta t}\,\mathcal{N}(t)$$

of an r vector $\mathcal{W}(t)$ of independent standard Brownian motions, or standard Wiener processes.

- Finally we obtain the CLE

$$d\mathbf{Y}(t) = \underbrace{\mathbf{A}(\mathbf{Y})(t))}_{\text{drift-rate function}}\, dt + \underbrace{\mathbf{D}(\mathbf{Y})(t))}_{\text{diffusion-rate function}}\, d\mathcal{W}(t)$$

CLE. Langevin equation

- This SDE (CLE equation) represents the diffusion process that most closely matches the dynamics of the associated Markov Jump Process (MJP), and can be shown to approximate the Stochastic Kinetic Network (SKM) increasingly well in high concentration scenarios.

CLE. Langevin equation

- This SDE (CLE equation) represents the diffusion process that most closely matches the dynamics of the associated Markov Jump Process (MJP), and can be shown to approximate the Stochastic Kinetic Network (SKM) increasingly well in high concentration scenarios.
- However, it should be noted that the approximation breaks down in low-concentration scenarios, and therefore should not be expected to work well for models involving species with very low copy-number. This is quite typical for many SKN's; yet the approximation often turns out to be adequate for inferential purposes in practice.

Summary

We summarize the high-level differences between the CME, CLE, and RRE philosophies.

Summary

We summarize the high-level differences between the CME, CLE, and RRE philosophies.

- **Chemical Master Equation.** A set of linear, autonomous ODEs. One ODE for each possible state of the system. Solution of the **n**th equation at time t gives the probability of system being in that particular state at time t.

Summary

We summarize the high-level differences between the CME, CLE, and RRE philosophies.

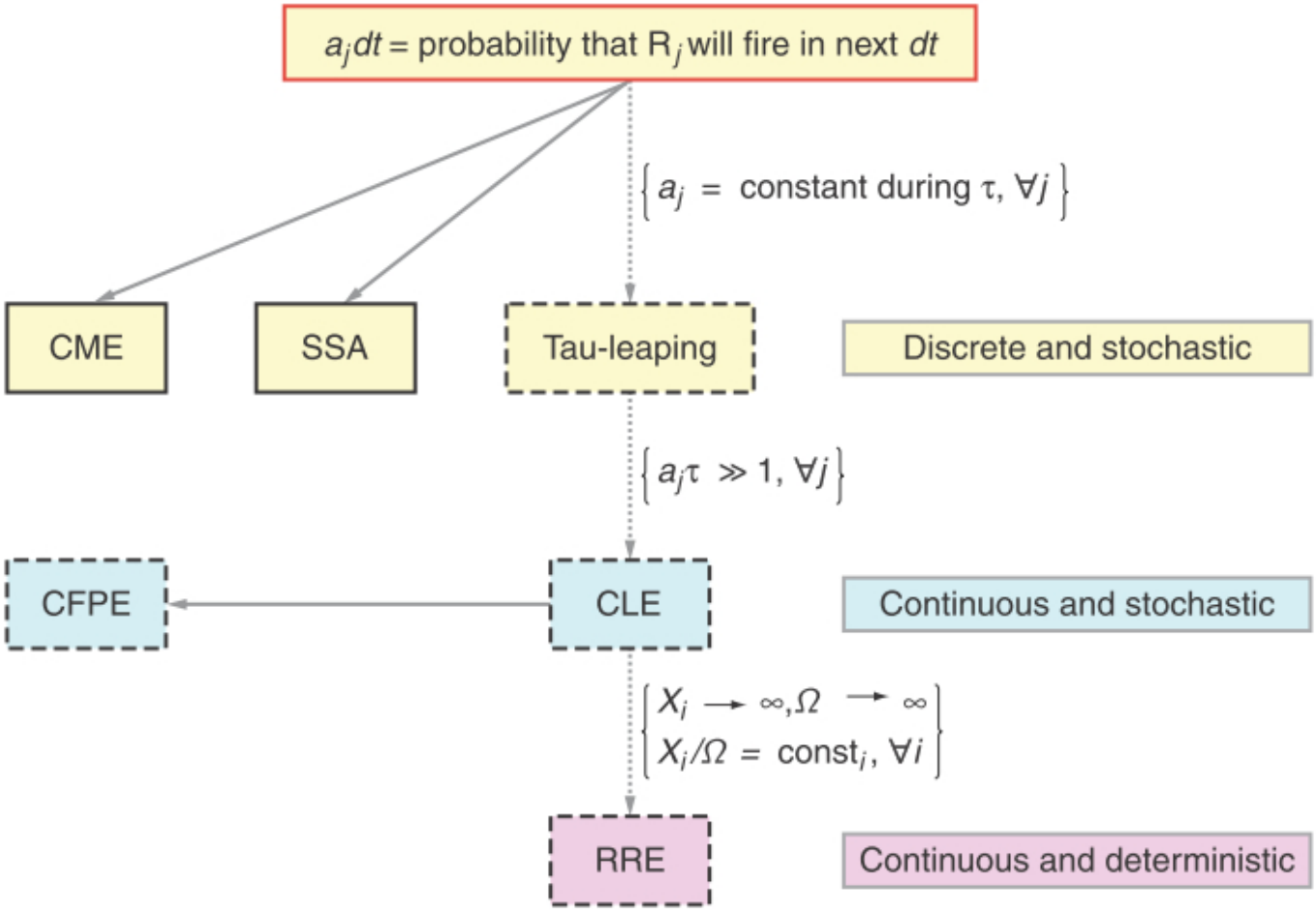
- **Chemical Master Equation.** A set of linear, autonomous ODEs. One ODE for each possible state of the system. Solution of the **n**th equation at time t gives the probability of system being in that particular state at time t.
- **Chemical Langevin Equation.** A set of nonlinear, autonomous SDEs. One SDE for each chemical species. Solution of the ith equation at time t is a real-valued random variable representing the amount of species i at time t.

Summary

We summarize the high-level differences between the CME, CLE, and RRE philosophies.

- **Chemical Master Equation.** A set of linear, autonomous ODEs. One ODE for each possible state of the system. Solution of the **n**th equation at time t gives the probability of system being in that particular state at time t.
- **Chemical Langevin Equation.** A set of nonlinear, autonomous SDEs. One SDE for each chemical species. Solution of the ith equation at time t is a real-valued random variable representing the amount of species i at time t.
- **Reaction Rate Equations.** A set of nonlinear, autonomous ODEs. One ODE for each chemical species. Solution of the ith equation at time t is a real number representing the concentration of species i at time t.

Summary (from [5])



Logical structure of stochastic chemical kinetics. Everything follows from the fundamental premise at the top via the laws of probability theory.

Computational Experiment. Michaelis-Menten system

- **Reaction network**

$$\begin{aligned} \mathscr{R}_1: \quad S + E &\xrightarrow{c_1} C \\ \mathscr{R}_2: \quad C &\xrightarrow{c_2} S + E \\ \mathscr{R}_3: \quad C &\xrightarrow{c_3} P + E \end{aligned}$$

- **Stoichiometry vectors** Species ordering (S, E, C, P):

$$\boldsymbol{\vartheta}_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\vartheta}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\vartheta}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix},$$

- **Propensities**

$\mathscr{R}_1 : h_1(\mathbf{n}, c) = c_1 n_S(t) n_E(t); \mathscr{R}_2 : h_2(\mathbf{n}, c) = c_2 n_C(t); \mathscr{R}_3 : h_3(\mathbf{n}, c) = c_3 n_C(t)$

Computational Experiment. Michaelis-Menten system

- **Initial data and constant rates (deterministic)**

$$\begin{pmatrix} x_S(0) \\ x_E(0) \\ x_C(0) \\ x_P(0) \end{pmatrix} = \begin{pmatrix} 5 \times 10^{-7} M \\ 2 \times 10^{-7} M \\ 0M \\ 0M \end{pmatrix}, \quad k_1 = 10^6, k_2 = 10^4, k_3 = 10^{-1}$$

- In a volume $V = 10^{-15} L$, this correspondes to molecular data

$$\begin{pmatrix} n_S(0) \\ n_E(0) \\ n_C(0) \\ n_P(0) \end{pmatrix} = \begin{pmatrix} \lfloor 5 \times 10^{-7} \mathcal{N}_A V \rfloor \\ \lfloor 2 \times 10^{-7} \mathcal{N}_A V \rfloor \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 312 \\ 125 \\ 0 \\ 0 \end{pmatrix}$$

$$c_1 = \frac{10^6}{\mathcal{N}_A V}, c_2 = 10^{-4}, c_3 = 10^{-1}$$

where $\mathcal{N}_A \approx 6.023 \times 10^{23}$ is **Avogadro number**.

Open Challenges

Stochastic modeling of biological dynamics, especially at the cellular level, is increasingly making its way to the mainstream of quantitative biology investigation. The CME and its accompanying SSA have proven to be invaluable computational tools for such studies. There are, however, many challenges that need to be addressed in order to make stochastic modeling a widely applicable tool for realistic biological problems.

Open Challenges

Stochastic modeling of biological dynamics, especially at the cellular level, is increasingly making its way to the mainstream of quantitative biology investigation. The CME and its accompanying SSA have proven to be invaluable computational tools for such studies. There are, however, many challenges that need to be addressed in order to make stochastic modeling a widely applicable tool for realistic biological problems.

- **Efficient Stochastic Simulation and Analysis for Systems Evolving at Disparate Temporal and Spatial Scales**.

Open Challenges

Stochastic modeling of biological dynamics, especially at the cellular level, is increasingly making its way to the mainstream of quantitative biology investigation. The CME and its accompanying SSA have proven to be invaluable computational tools for such studies. There are, however, many challenges that need to be addressed in order to make stochastic modeling a widely applicable tool for realistic biological problems.

- **Efficient Stochastic Simulation and Analysis for Systems Evolving at Disparate Temporal and Spatial Scales**.
- **Efficient Spatiotemporal Simulations.** Biological networks in practice consist of components that interact in a three-dimensional space and are not necessarily distributed homogeneously as they diffuse between different cellular compartments.

Open Challenges

- **Holistic understanding of biological systems.**
  Understanding of biological systems often involves the probing of cellular biochemical networks in the context of the cell, of cells in the context of a tissue, and of a tissue in the context of the organism. How to account for and move between these spatial scales remains an open problem for stochastic modeling.

Open Challenges

- **Holistic understanding of biological systems.** Understanding of biological systems often involves the probing of cellular biochemical networks in the context of the cell, of cells in the context of a tissue, and of a tissue in the context of the organism. How to account for and move between these spatial scales remains an open problem for stochastic modeling.
- **Parametrization and Sensitivity Analysis of Stochastic Models.** Stochastic models of biological systems typically depend on a set of kinetic parameters whose values are often unknown or fluctuate due to an uncertain environment. These parameters determine the dynamic behavior of the model, and changes in them may alter the system's output in nonintuitive ways.

Next steps

Space as the final frontier in stochastic simulations of biological systems.

First Reaction Method

In the his 1976 paper [4], Gillespie writes two algorithms: • **Direct Method (DM)** that we have explored before and which requires 2 random numbers per iteration. • **First Reaction Method (FRM)** which will be exemplified therefore:

$\mathscr{R}_1$	$\mathscr{R}_2$	$\mathscr{R}_3$	$t = 0$	
0.75	0.20	1.20	$t = 0.20$	, $\mathscr{R}_2$

First Reaction Method

In the his 1976 paper [4], Gillespie writes two algorithms: • **Direct Method (DM)** that we have explored before and which requires 2 random numbers per iteration. • **First Reaction Method (FRM)** which will be exemplified therefore:

$\mathscr{R}_1$	$\mathscr{R}_2$	$\mathscr{R}_3$	$t = 0$	
0.75	0.20	1.20	$t = 0.20$	, $\mathscr{R}_2$
0.60	0.63	1.30	$t = 0.20+0.60$	, $\mathscr{R}_1$

First Reaction Method

In the his 1976 paper [4], Gillespie writes two algorithms: • **Direct Method (DM)** that we have explored before and which requires 2 random numbers per iteration. • **First Reaction Method (FRM)** which will be exemplified therefore:

$\mathscr{R}_1$	$\mathscr{R}_2$	$\mathscr{R}_3$	$t = 0$	
0.75	0.20	1.20	$t = 0.20$	, $\mathscr{R}_2$
0.60	0.63	1.30	$t = 0.20+0.60$	, $\mathscr{R}_1$

- algorithm: generate t_i for each reaction i, based on its propensities $\rightarrow$ choose smallest t_i
- requires ν random numbers (number of reactions)
- Gillespie proved it to be compatible (yet slower) than DM

First Reaction Method

- **1.** Initialization: initial conditions.
- **2.** Calculate reaction propensities $h_i(\mathbf{x}, c_i)$, for $i = 1, 2, \cdots, \nu$.
- **3.** Simulate a time to the next reaction i, $t_i \sim Exp(h_i(\mathbf{x}, c_i))$, for $i = 1, 2, \cdots, \nu$.
- **4.** Let j be the index of the smallest time t_i.
- **5.** Set $t := t + t_j$.
- **6.** Update the state $\mathbf{x}$ according to the reaction with index j. That is, set $\mathbf{x} := \mathbf{x} + S^{(j)}$.
- **7.** If $t < T_{\text{max}}$, return to step **2.**

Next Reaction Method

The former algorithm was improved by Gibson and Bruck [3] in what they called **Next Reaction Method (NRM)**:

R1	R2	R3		R2	R1	R3	
0.60	0.20	1.30	$\Rightarrow$	0.20	0.60	1.30	R2 , $t := 0.20$

R2	R1	R3		R1	R2	R3	
0.63	0.60	1.30	$\Rightarrow$	0.60	0.63	1.30	R1 , $t := t + 0.60$

$= t + 0.43$

- Generate times only for reactions whose propensities have been affected by the current one.
- Always keep t_i in absolute time by adding current time: $t_i := t + Exp(h_i(\mathbf{x}, c_i))$.

Next Reaction Method

- **1.** Initialization: initial conditions.
- **2.** Let j be the index of the smallest t_i.
- **3.** Set $t := t_j$.
- **4.** Update $\mathbf{x}$ according to reaction with index j.
- **5.** Update $h_j(\mathbf{x}, c_j)$ according to the new state $\mathbf{x}$ and simulate a new time $t_j := t + Exp(h_j(\mathbf{x}, c_j))$.
- **6.** For each reaction $i(\neq j)$ whose hazard is changed by reaction j:
  - **a)** Update $h_i' = h_i(\mathbf{x}, c_i)$ (but temporarily keep the old h_i).
  - **b)** Set $t_i := t + (h_i/h_i')(t_i - t)$.
  - **c)** Forget the old h_i.
- **7.** If $t < T_{\max}$, return to step **2.**

Next Reaction Method

- Requires extra structures for:
  - **a)** graph identifying reactions whose propensities may have been affected by current reaction.
  - **b)** event queue structure to keep t_i sorted, avoiding performing a full sort.
- It may actually be faster than the **Direct Method** (DM), depending on details of the model (DM always requires 2 random numbers; the model may require only 1 or more under NRM) and implementation (structures used).
- Iterations also have the potential to be run in parallel. (Wilkinson, 2005)

Next Subvolume Method

- For spatial modeling, one algorithm is **Next Subvolume Method (NSM)** by Elf and Ehrenberg [2].

Next Subvolume Method

- For spatial modeling, one algorithm is **Next Subvolume Method (NSM)** by Elf and Ehrenberg [2].
- This method modifies **Next Reaction Method**, so that it uses "subvolumes" instead.
- Each subvolume is a homogeneous mixture under the **Direct Method**.
- We may in fact look at these subvolumes as reactions, where each molecule is subject to diffusion to another subvolume, according to a diffusion rate (its h_i):

$$\mathscr{S}_1 : A_{(1,1)} \xrightarrow{D_1} A_{(2,1)}$$

Next Subvolume Method

- For spatial modeling, one algorithm is **Next Subvolume Method (NSM)** by Elf and Ehrenberg [2].
- This method modifies **Next Reaction Method**, so that it uses "subvolumes" instead.
- Each subvolume is a homogeneous mixture under the **Direct Method**.
- We may in fact look at these subvolumes as reactions, where each molecule is subject to diffusion to another subvolume, according to a diffusion rate (its h_i):

$$\mathscr{S}_1 : A_{(1,1)} \xrightarrow{D_1} A_{(2,1)}$$

- Each spatial diffusion affects only 2 subvolumes, so only 2 times need to be regenerated (like in **DM**)

Next Subvolume Method

- For spatial modeling, one algorithm is **Next Subvolume Method (NSM)** by Elf and Ehrenberg [2].
- This method modifies **Next Reaction Method**, so that it uses "subvolumes" instead.
- Each subvolume is a homogeneous mixture under the **Direct Method**.
- We may in fact look at these subvolumes as reactions, where each molecule is subject to diffusion to another subvolume, according to a diffusion rate (its h_i):

$$\mathscr{S}_1 : A_{(1,1)} \xrightarrow{D_1} A_{(2,1)}$$

- Each spatial diffusion affects only 2 subvolumes, so only 2 times need to be regenerated (like in **DM**)
- Can potentially represent any spatial geometry (2D, 3D, graph)

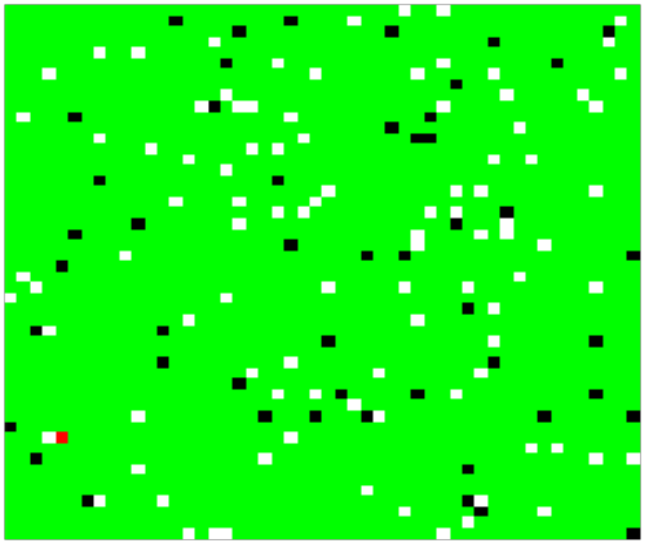
Next Subvolume Method

- **1.** For each subvolume k, calculate rates r_k and s_k, and generate $t_k \sim Exp(D_k(r_k + s_k))$. Sort vector $\mathbf{t}$.
- **2.** For the smallest time τ of subvolume λ, choose either a *reaction event* if rand $< r_\lambda/(r_\lambda + s_\lambda)$, or a *diffusion event*
- **3.** If reaction:
  - **a)** Reuse the previous random number to determine which reaction occurred, as in the DM.
  - **b)** Recalculate r_λ, s_λ and t_λ for this subvolume.
- **4.** If diffusion:
  - **a)** Reuse the previous random number to determine which molecule diffused away, given their respective diffusion rates.
  - **b)** Another random number chooses a neighbor γ based on the connectivity matrix.
  - **c)** Recalculate r, s and t for both subvolumes λ and γ.
- **5.** Go back to step **2.**

Lotka-Volterra simulation

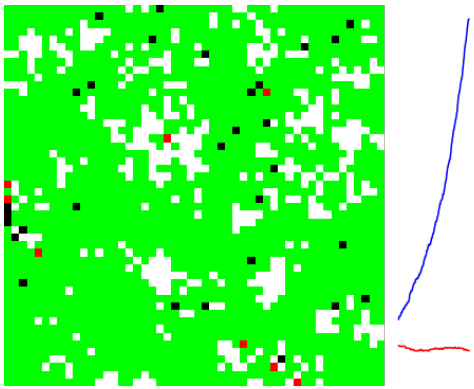
Using this method, we simulate Lotka-Volterra predator-prey model, in a 50x50 grid, by initially spreading around 100 X_1 (sheep) and 50 X_2 (wolves), using the reaction rates $\mathbf{C} = (0.2, 0.5, 0.1)$ and diffusion rates $\mathbf{D} = (1, 5)/3^2$, and the reaction network as:

$$\begin{aligned} X_1 &\xrightarrow{c_1} 2X_1 \\ X_1 + X_2 &\xrightarrow{c_2} 2X_2 \\ X_2 &\xrightarrow{c_3} \emptyset \end{aligned}$$

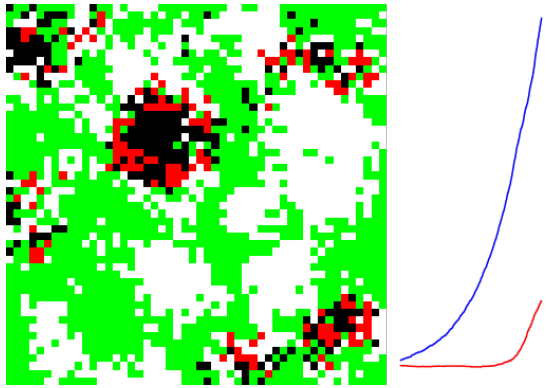


Lotka-Volterra simulation

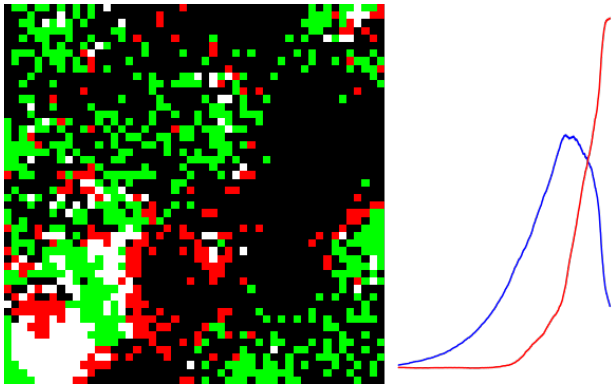
t = 10 (2.624s)



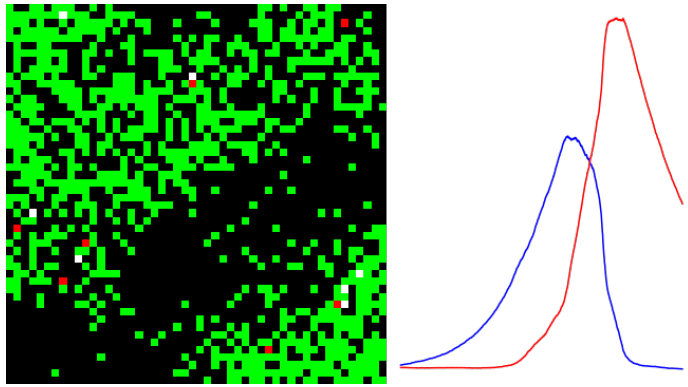
t = 20 (18.213s)



t = 30 (535.229s)



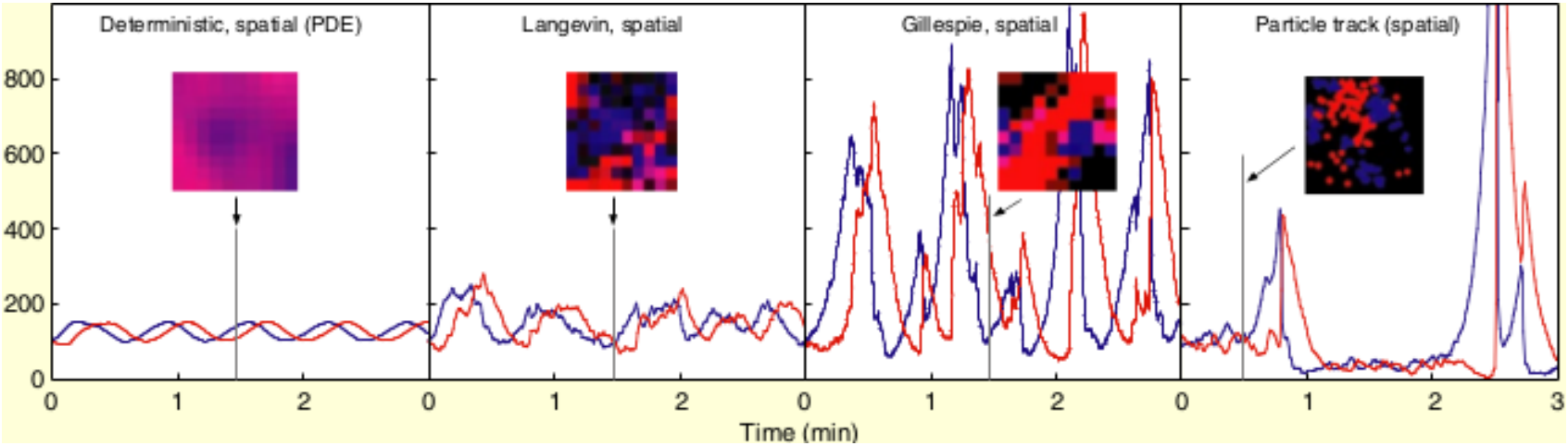
t = 40 (1013.944s)



Legend: Green=$(0X_1, 0X_2)$, White=$(> 0X_1, 0X_2)$, Black=$(0X_1, > 0X_2)$, Red=$(> 0X_1, > 0X_2)$

Several spatial methods

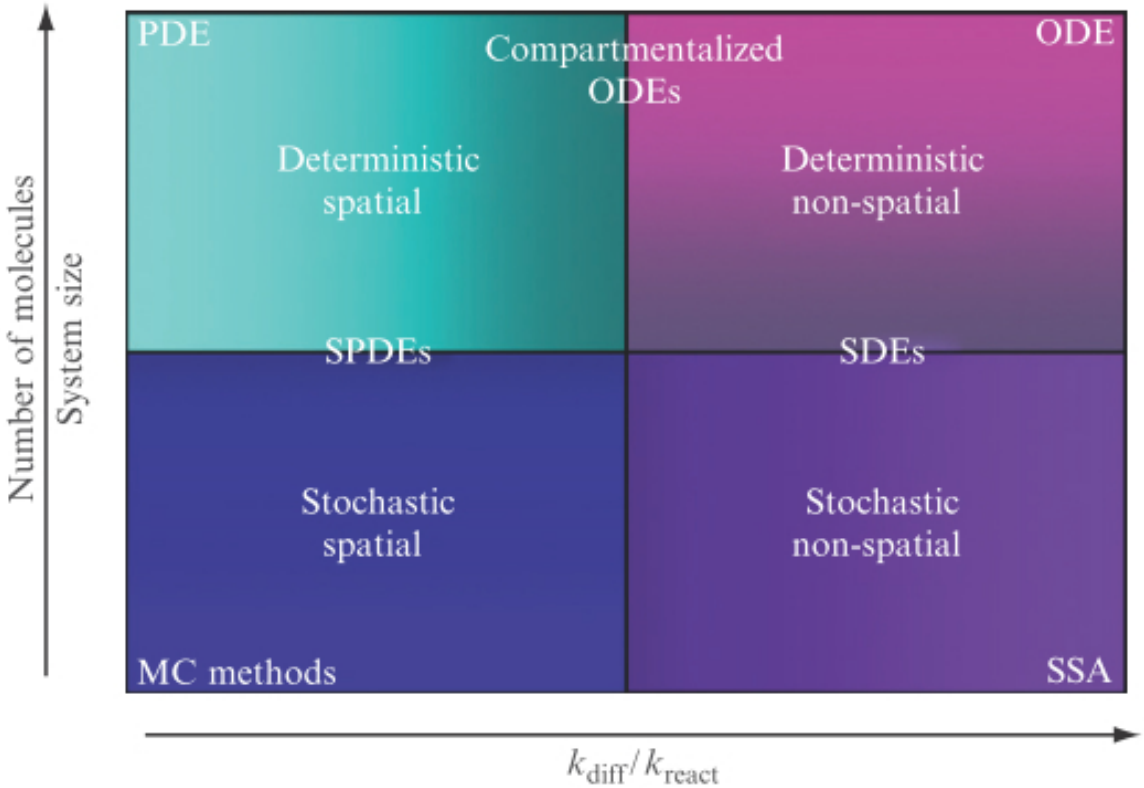
A review of several spatial methods by Andrews and Arkin [1]:



The authors found that:

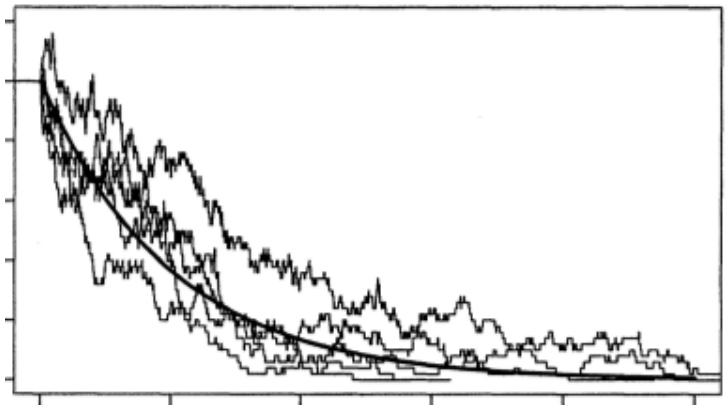
- the Gillespie simulation has larger peaks than the Langevin one
- the particle tracking simulation shows larger and fewer bursts than does the Gillespie simulation because it accurately treats diffusion at all length scales (this difference can be reduced by using smaller subvolumes)

Summary of Spatial Biological System Simulations



Mathematical methods often employed in kinetic studies: ODEs, PDEs, SPDEs, SDEs ; MC, Monte Carlo; SSA, Gillespie-type stochastic simulation approach; k_{react}= rate of biochemical processes; k_{diff} = rate of diffusion processes.

Summary of Biological System Simulations



- The world is governed by deterministic laws
- We need exogenous stochasticity to keep our models simple [9, pp.2]
- Using stochastic methods matter more when working with fewer molecules/agents

References I

Steven S Andrews and Adam P Arkin. "Simulating cell biology". In: *Current biology* 16.14 (2006), R523–R527.

Johan Elf and Måns Ehrenberg. "Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases". In: *Systems biology* 1.2 (2004), pp. 230–236.

Michael A Gibson and Jehoshua Bruck. "Efficient exact stochastic simulation of chemical systems with many species and many channels". In: *The journal of physical chemistry A* 104.9 (2000), pp. 1876–1889.

Daniel T Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of computational physics* 22.4 (1976), pp. 403–434.

Daniel T Gillespie. "Stochastic simulation of chemical kinetics". In: *Annu. Rev. Phys. Chem.* 58 (2007), pp. 35–55.

References II


Desmond J Higham. "Modeling and simulating chemical reactions". In: *SIAM review* 50.2 (2008), pp. 347–368.


Thomas E Turner, Santiago Schnell, and Kevin Burrage. "Stochastic approaches for modelling in vivo reactions". In: *Computational biology and chemistry* 28.3 (2004), pp. 165–178.


Mukhtar Ullah and Olaf Wolkenhauer. "Stochastic approaches in systems biology". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 2.4 (2010), pp. 385–397.


Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2011.