Systems Biology 2

Stochastic Modelling

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

- Any questions from the lab?

Today's lecture

- Deterministic modeling review.
- Stochastic simulation as an alternative.
- Stochastic Simulation - Gillespie algorithm.
- Gillespie - weaknesses and extensions.

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

University of Glasgow

Recall the simple model:

$$
\begin{aligned}
A + B &\xrightarrow{k_1} 2B + A \\
B + C &\xrightarrow{k_2} 2C \\
C &\xrightarrow{k_3} \emptyset
\end{aligned}
$$

Resulting in the following set of ODEs:

$$
\begin{aligned}
[\dot{A}] &= 0 \\
[\dot{B}] &= k_1 \cdot [A] \cdot [B] - k_2 \cdot [B] \cdot [C] \\
[\dot{C}] &= k_2 \cdot [B] \cdot [C] - k_3 \cdot [C]
\end{aligned}
$$

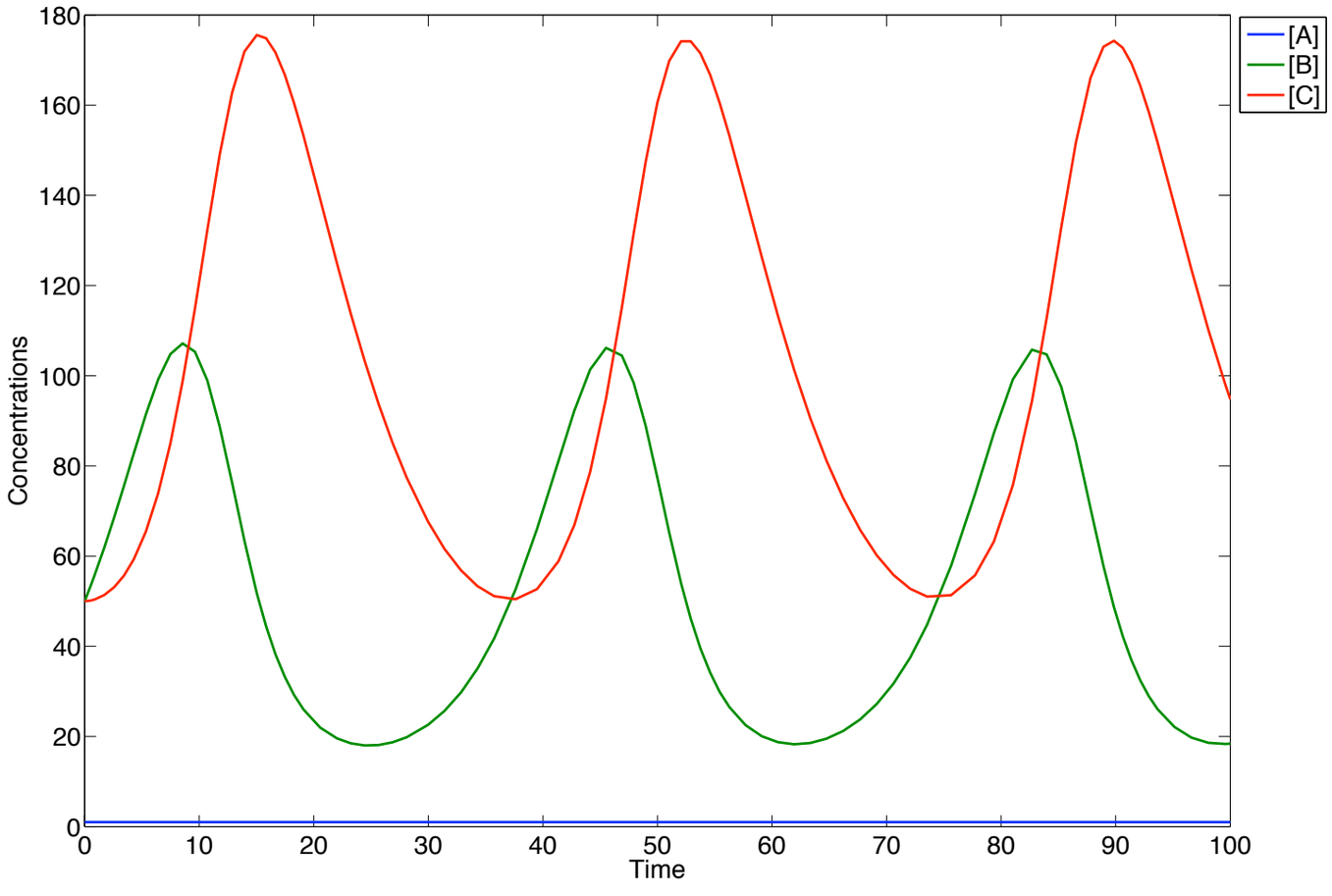
Simulation

Given initial concentrations and constants:

$$A|_{t=0} = 1 \qquad B|_{t=0} = 50 \qquad C|_{t=0} = 50$$

$$k_1 = 0.25 \qquad k_2 = 0.0025 \qquad k_3 = 0.125$$

We can simulate concentrations over time:



simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

All assumptions are wrong, some are useful...

- In order to construct these ODEs what assumptions have we made?
    - Deterministic
    - Mass action kinetics
    - Continuous values
    - Closed system
    - Well mixed

We have assumed that we can accurately model based on some notion of average behavior.

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Exact simulation

- An alternative to ODE approaches is exact stochastic simulation.
- Exact?
  - *We deal with populations rather than concentrations.*
  - *We explicitly model each reaction.*
  - *Not: "This is exactly what will happen"!*
- *Stochastic?*
  - *Model the inherent uncertainty of the system.*
  - *Particularly important for species with small populations.*

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Exact stochastic simulation

- The most famous stochastic simulation algorithm is the Gillespie algorithm:

Exact Stochastic Simulation of Coupled Chemical Reactions

Daniel T. Gillespie*

Research Department, Naval Weapons Center, China Lake, California 93555 *(Received May 12, 1977)*

Publication costs assisted by the Naval Weapons Center

The Journal of Physical Chemistry, Vol. 81, No. 25, 1977

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

What does Gillespie do?

- Deals with integer populations of molecules
- Assumes the model is inherently stochastic (random)

- E.g. protein decay:

$$C \xrightarrow{k} \emptyset$$

$$C|_{t=0} = 100$$

$$k = 0.1$$

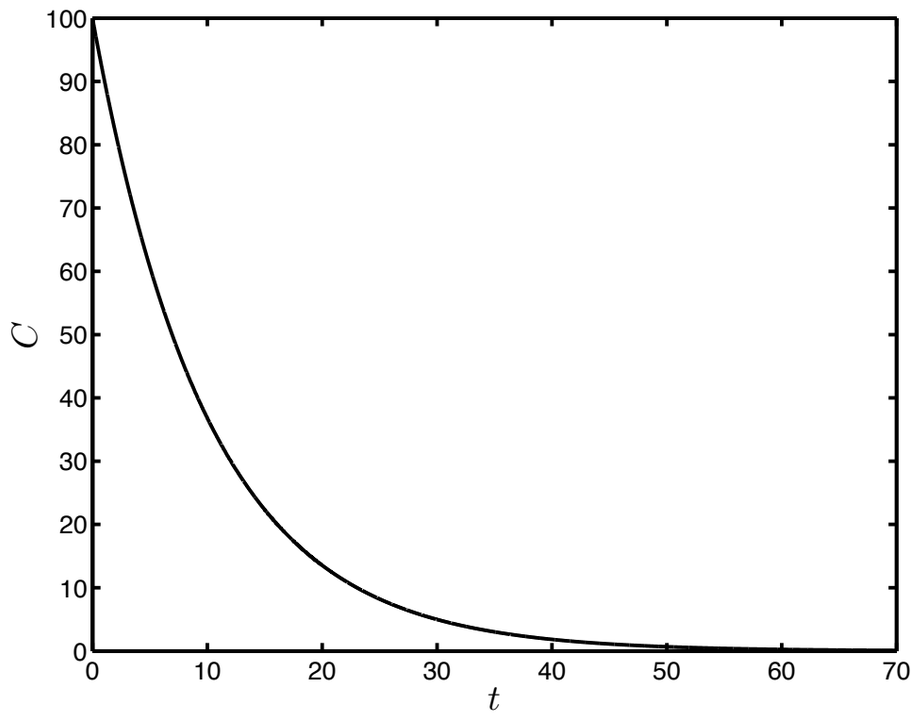
- Deterministic model (ODE) can be solved analytically

$$\dot{C} = -kC$$

$$C(t) = C_0 \exp\{-kt\}$$

- And tells us that:

$$\text{at } t = 20, \;\; C(20) = 13.5335$$



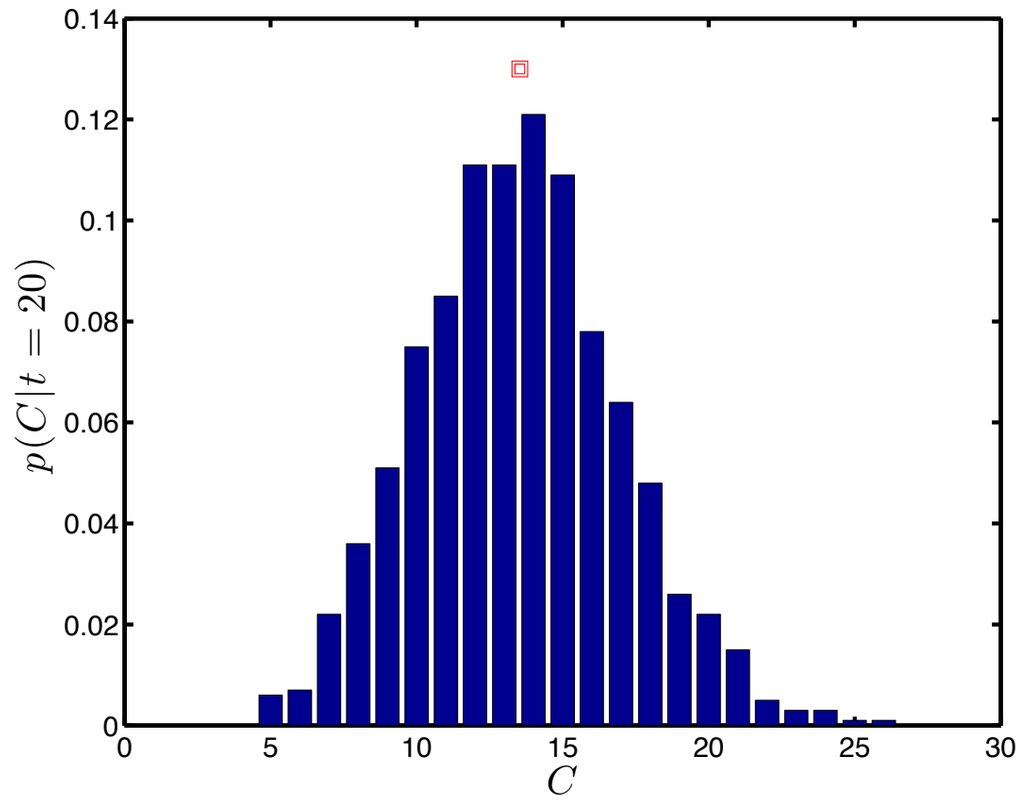
simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

What does Gillespie do? II

In stochastic simulation, we are interested in distributions

$$p(C|t=20) \text{ or } p(C=13|20)$$

Here is the distribution for t = 20 created using Gillespie - we can see that the ODE (red square) in this case is in rough agreement with the most likely number of C molecules.

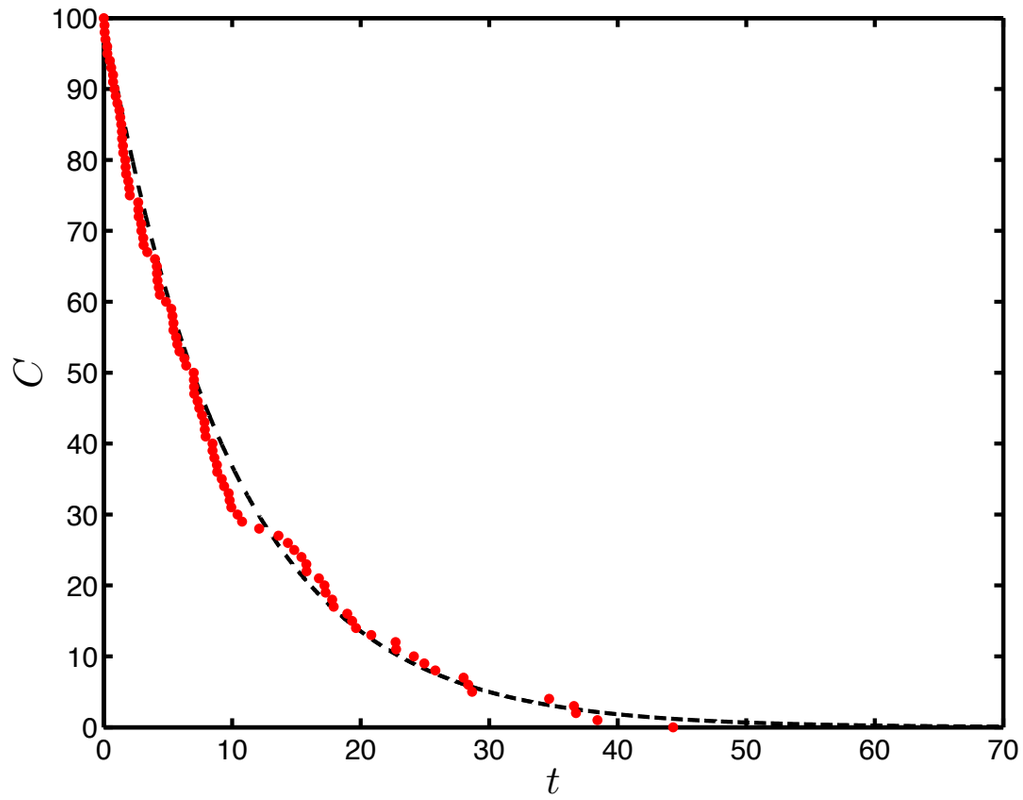


Unfortunately, for any remotely interesting model, analytically computing these probabilities is impossible. But, we can simulate...

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

What does Gillespie do? III

The Gillespie algorithm allows us to generate samples from the stochastic model

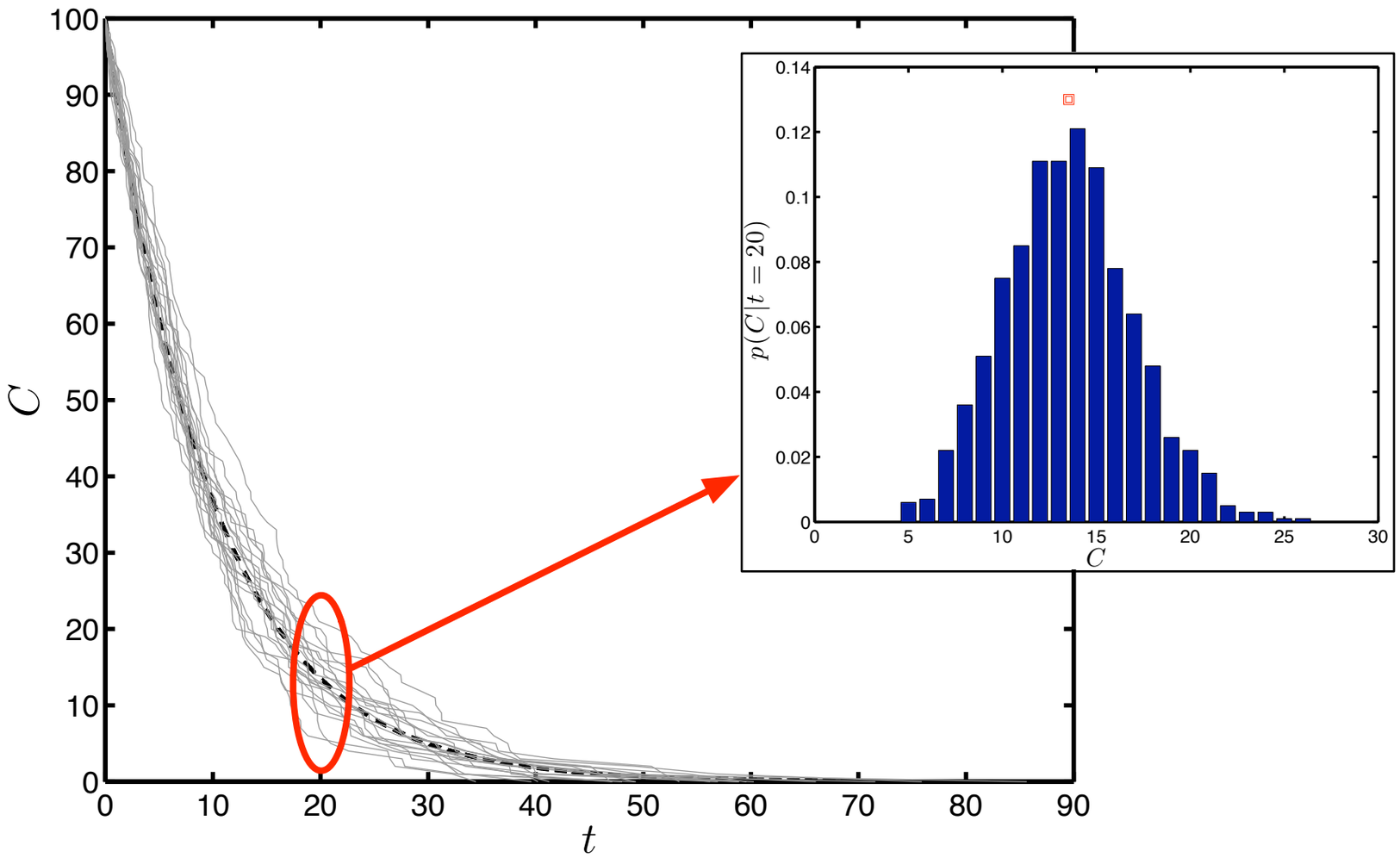


Each sample is a trajectory of the species' populations through time

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

What does Gillespie do? IV

We may be interested in individual samples or computing empirical distributions from sets of samples...



simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

The Gillespie algorithm

N molecular species, M reactions

At time t, population sizes (state) given by: $X_1, X_2, \ldots . X_N$

(Integers, >0)

We need to generate two things:

1. The time until the next reaction occurs
2. The type of reaction that occurs

Gillespie shows that the two can be de-coupled (we can sample the time and then sample which reaction takes place) resulting in a very simple procedure.

For more details and derivations, refer to Gillespie's paper.

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

The Gillespie algorithm

At each iteration....

For each of the M reactions compute: $a_m = c_m h_m$

Reaction constant (analogous to k)

$$X_1 + X_2 \xrightarrow{k} X_3$$

Number of reactant combinations

Reaction	Number of combination, h
$X_1 + X_2 \rightarrow X_3$	$X_1 X_2$
$X_1 + X_1 \rightarrow X_3$	$\frac{1}{2} X_1 (X_1 - 1)$

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

The Gillespie algorithm II

University of Glasgow

Then....

Let: $$a_0 = \sum_{m=1}^{M} a_m$$

Generate: $$r_1, r_2 \sim U(0,1)$$

Compute time to reaction: $$\tau = (1/a_0)\ln(1/r_1)$$

Choose reaction v for which: $$\sum_{m=1}^{v-1} a_m/a_0 < r_2 \leq \sum_{m=1}^{v} a_m/a_0$$

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Gillespie - Example

Two species, two reactions:

$$\begin{aligned} A &\xrightarrow{c_1} B \\ B &\xrightarrow{c_2} \emptyset \end{aligned}$$

$$c_1 = c_2 = 0.1$$

$$\text{at } t = 0, \;\; A = 100, B = 0$$

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

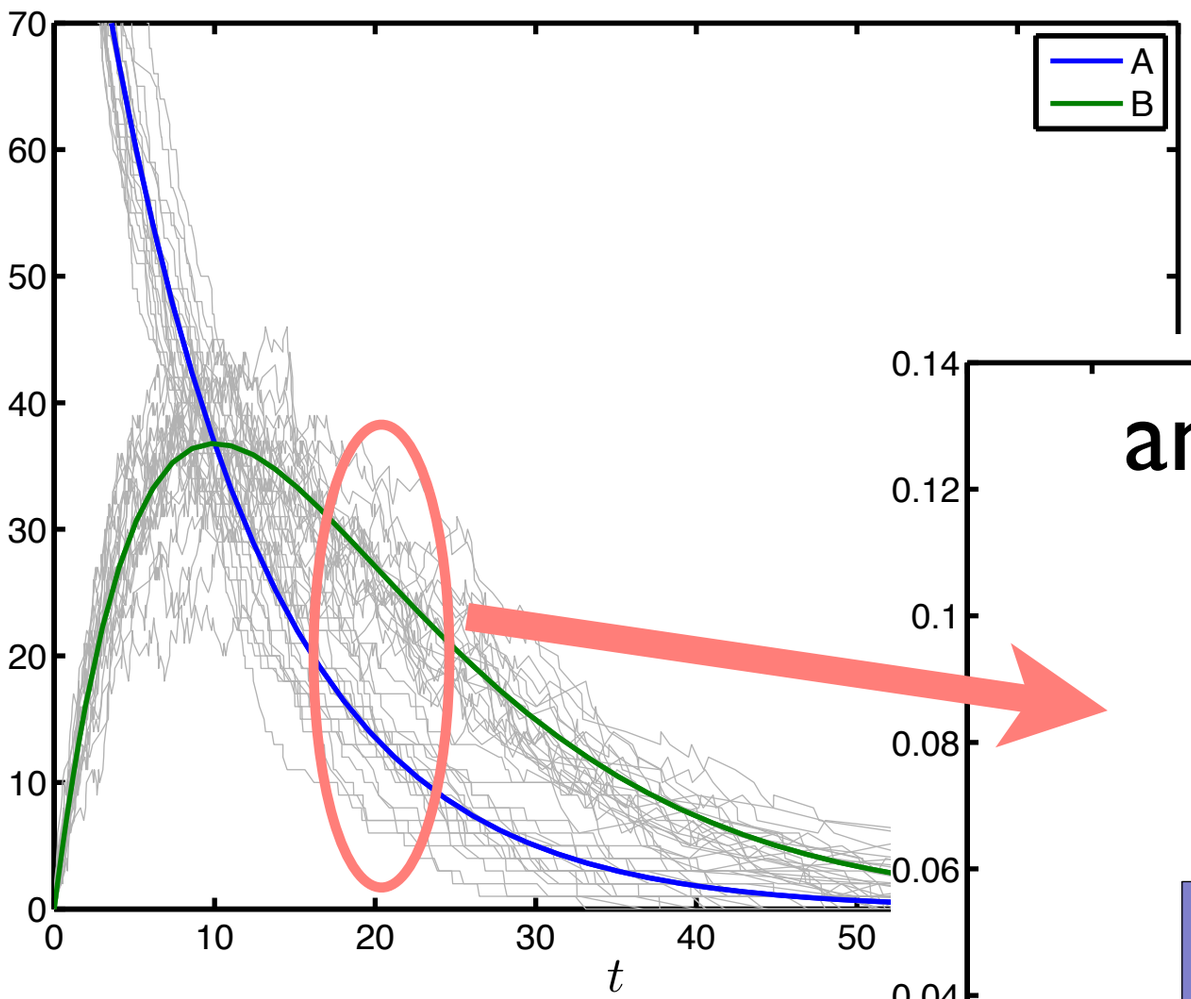
Gillespie Example

lt	t	A	B	h	a/a0	Sum
1	0	100	0	[100,0]	[1,0]	[1,1]
2	0.4315	99	1	[99,1]	[0.99,0.01]	[0.99,1]
3	0.5528	98	2	[98,2]	[0.98,0.02]	[0.98,1]
...	...	...	...	...	...	...
32	2.8841	73	24	[73,24]	[0.75,0.25]	[0.75,1]

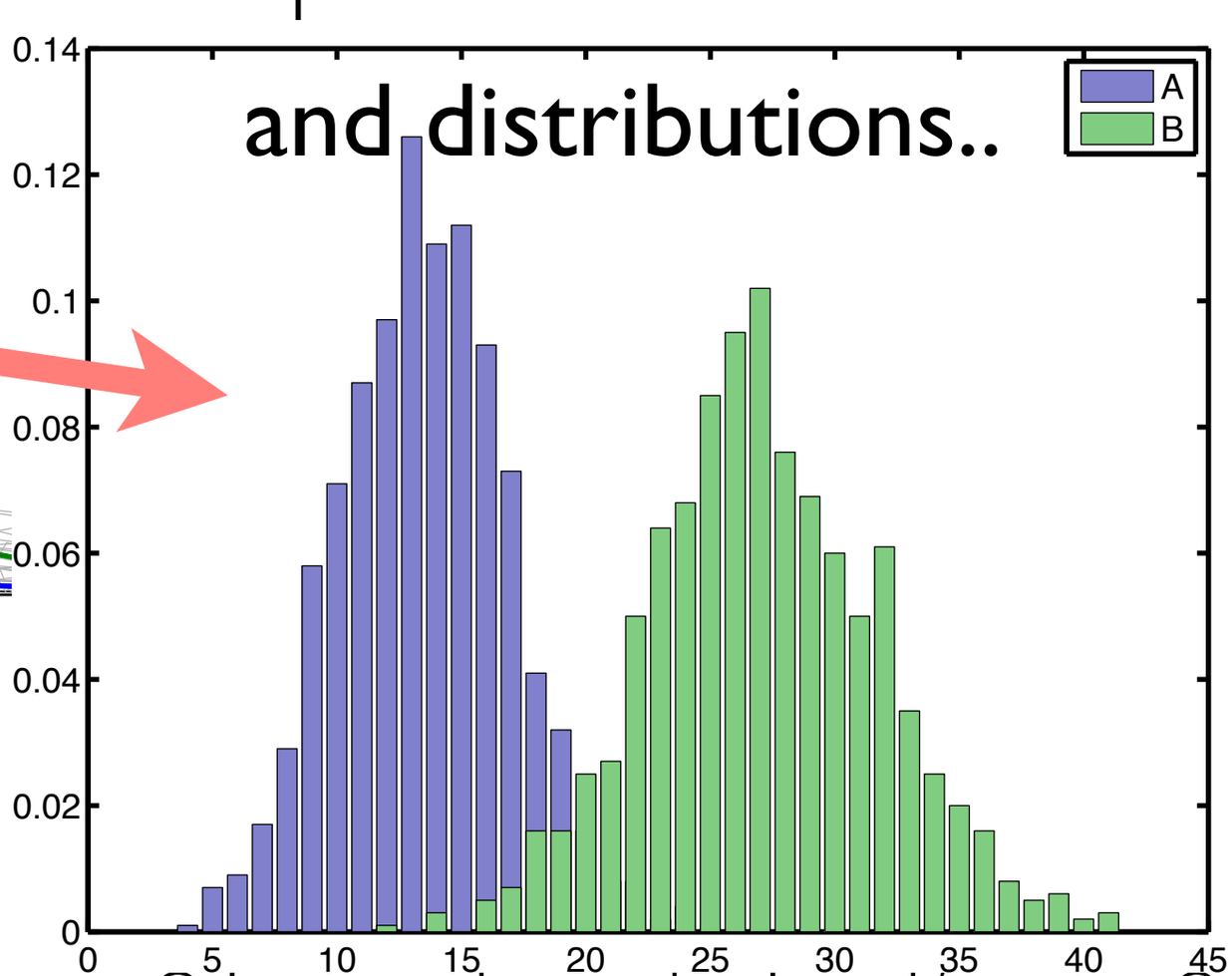
simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Gillespie example

Lots of simulations...



and distributions..



simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Testing our assumptions

- As Gillespie is exact we can use it to test the assumptions we use in ODE models.
- As an example, lets test the assumptions used in the 3 species model already discussed.

University of Glasgow

Recall the simple

$$\begin{aligned} A + B &\xrightarrow{k_1} 2B + A \\ B + C &\xrightarrow{k_2} 2C \\ C &\xrightarrow{k_3} \emptyset \end{aligned}$$

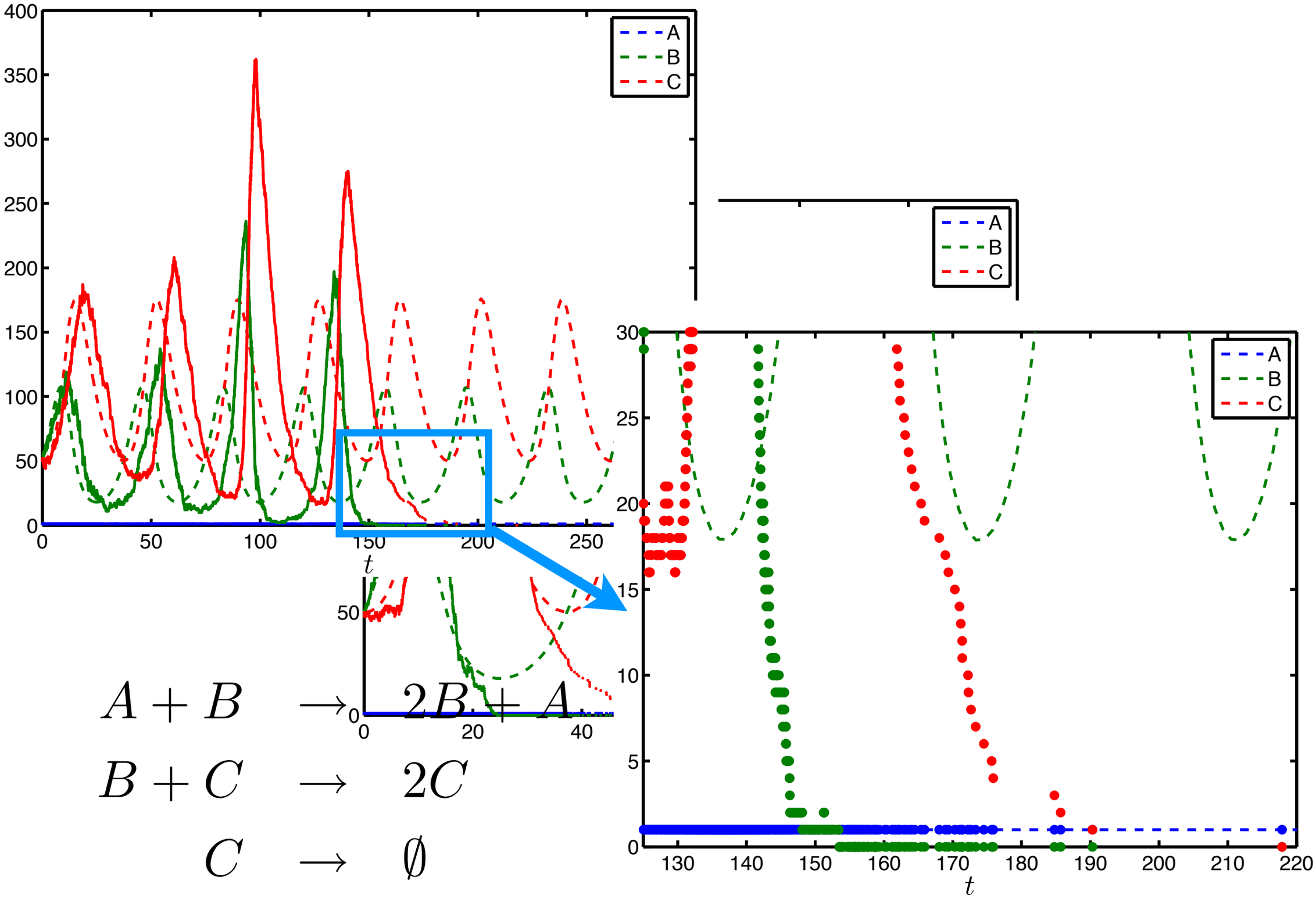
Resulting in the following set of ODEs:

$$\begin{aligned} [\dot{A}] &= 0 \\ [\dot{B}] &= k_1 \cdot [A] \cdot [B] - k_2 \cdot [B] \cdot [C] \\ [\dot{C}] &= k_2 \cdot [B] \cdot [C] - k_3 \cdot [C] \end{aligned}$$

Lotka-Volterra

- Run the Gillespie simulator using the lotka project with T=100

Lotka-Volterra



simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Lotka-Volterra

- ODE model predicts sustained oscillations (B>0).
- Stochastic model stops after a number of cycles (B=0) and never recovers.
- Very different interpretations!

Limitations of Gillespie

- For large populations and fast reactions, Gillespie becomes computationally impractical.
- For example, dimerisation:

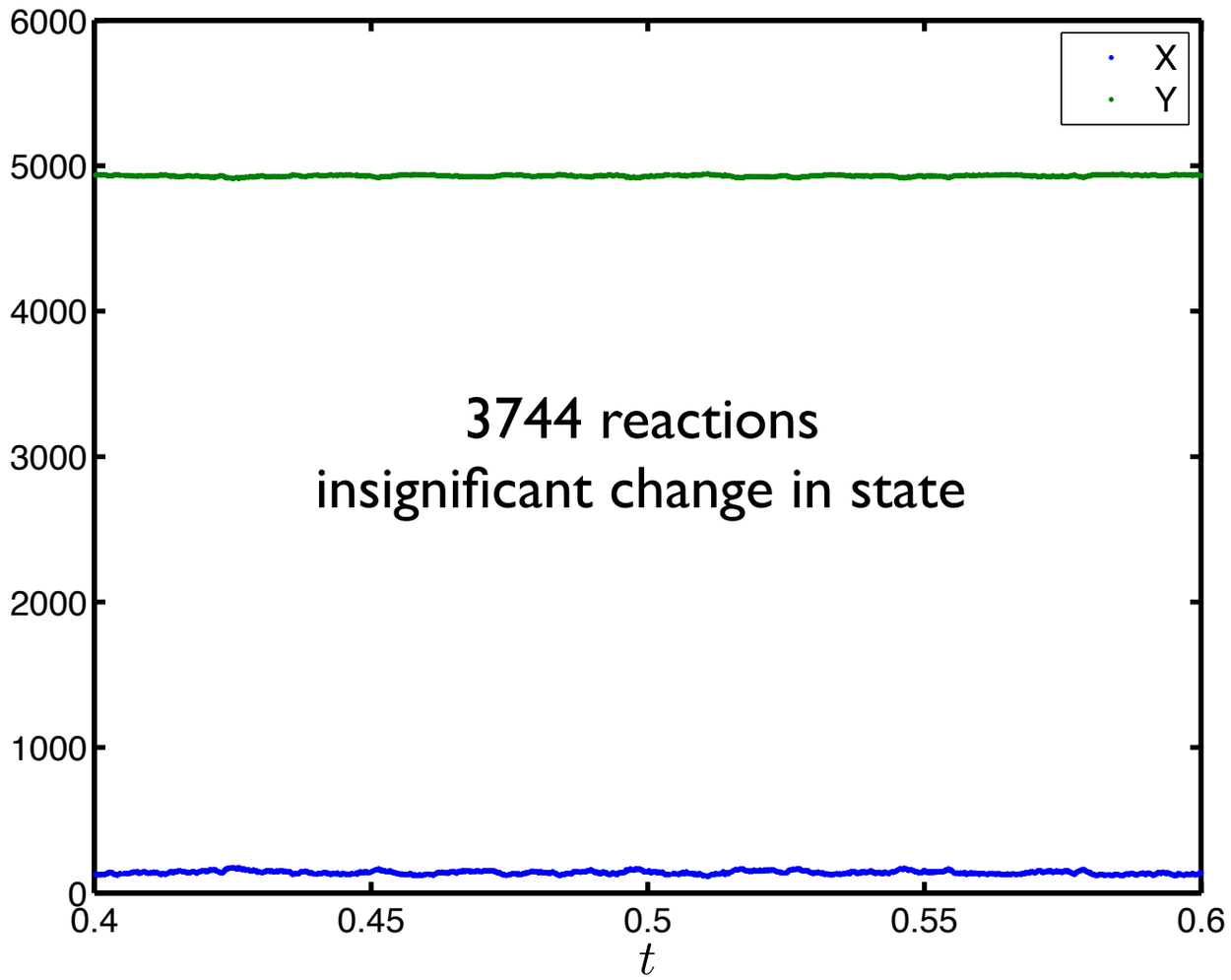
$$\begin{aligned} 2X &\xrightarrow{1} Y \\ Y &\xrightarrow{2} 2X \end{aligned} \qquad X_0 = 100, \quad Y_0 = 1$$

- Run this simulation (project: dimer) for T=100. How many reactions are simulated? Plot the results.

Extensions to Gillespie

- Extensions have been proposed to speed things up:
- Exact methods:
    - Tricks to make the sampling faster.
- Approximations:
    - e.g. Tau-leaping:
        - Choose a time, tau, to 'leap'. Compute how many reactions one would expect in that time jump.
        - Key Assumption: State does not change much in tau.

Extensions to Gillespie II



In the dimerisation example, tau-leaping may be appropriate.

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Hybrid models

- There is no reason why our model need be exclusively stochastic or deterministic.
- Mix and match:
    - Use exact for slow, important reactions
    - Use ODEs for fast, less critical reactions
- Example - Lotka-Volterra 2!

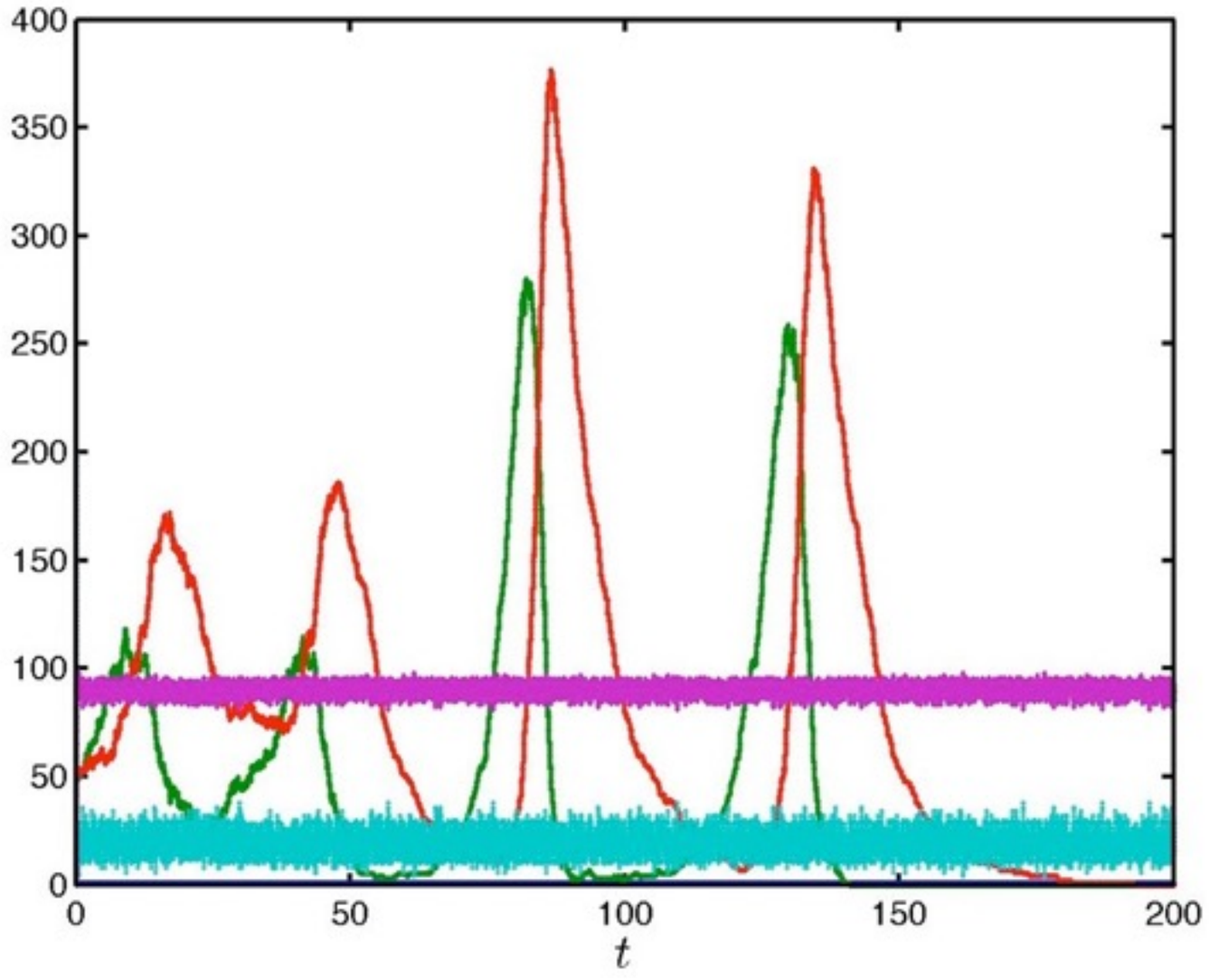
Lotka-Volterra hybrid

$$
\begin{aligned}
A + B + Y &\xrightarrow{0.0025} 2B + A + Y \\
B + C &\xrightarrow{0.0025} 2C \\
C &\xrightarrow{0.125} \emptyset \\
2X &\xrightarrow{1} Y \\
Y &\xrightarrow{2} X
\end{aligned}
$$

$$A_0 = 1, \;\; B_0 = 50, \;\; C_0 = 50, \;\; X_0 = 200, \;\; Y_0 = 0$$

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Lotka-Volterra hybrid

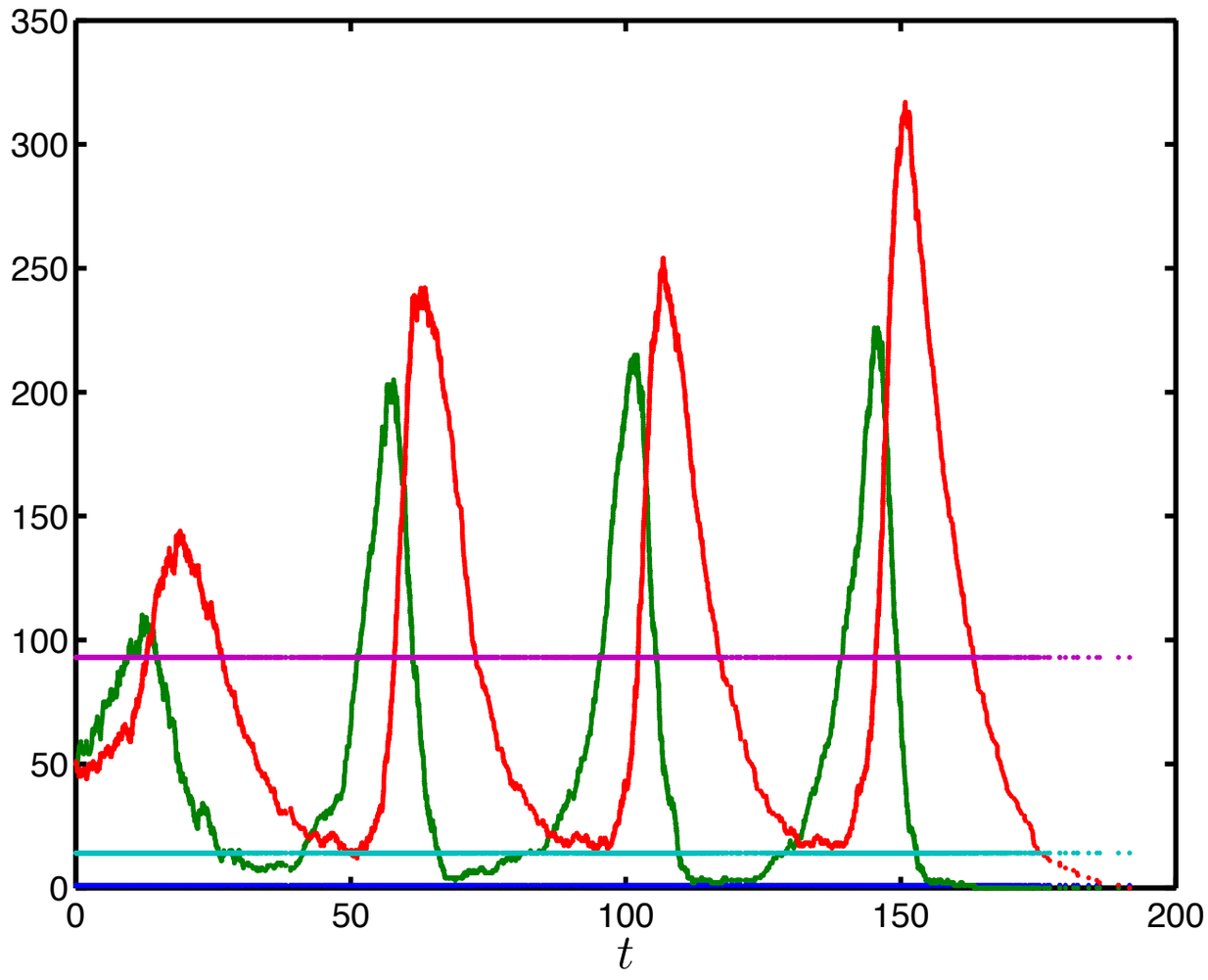


Standard Gillespie - 78,258 reactions

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Lotka-Volterra hybrid

As X and Y only change in dimerisation reactions, remove these and substitute X and Y populations by steady state from ODEs (X=14 , Y=93).



Same behavior - 5773 reactions (93% reduction!)

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers

Parameter estimation

- On Wednesday, we briefly mentioned parameter estimation for ODE models
- Can the same be done with stochastic simulation?
  - Q: what kind of data?
- **Very** computationally expensive (why?)

Summary

- Exact stochastic simulation an an alternative to ODEs.
- More realistic.
- More computation!
- Can show us where ODEs breakdown.
- In large systems, stochastic simulation is not feasible. Hybrid models show great potential.

simon.rogers@glasgow.ac.uk www.dcs.gla.ac.uk/~srogers @sdrogers