

# Stochastic models in biology and their deterministic analogues

Alan McKane

Theory Group, School of Physics and Astronomy, University of Manchester

Newton Institute, May 2, 2006

- Introduction
- The master equation, with examples
- Hubbell's neutral theory
- Stochastic amplification
- Conclusions

with

Tim Newman, John Nagy, Marianne Stefanini (Arizona),  
David Alonso, Mercedes Pascual (Michigan)

# 1. Introduction

One of the most well known aspects of quantitative ecology is the study of deterministic differential equations for the change in time of populations of individuals — different species, for example

An example is the Lotka-Volterra equations describing competition between species A, having  $N_A$  individuals in the population, and species B, having  $N_B$  individuals in the population:

$$\frac{dN_A}{dt} = r_A N_A (1 - N_A/K_A) - b_A N_A N_B$$

$$\frac{dN_B}{dt} = r_B N_B (1 - N_B/K_B) - b_B N_A N_B$$

In reality, population dynamics is stochastic — populations fluctuate due to random births and deaths, for instance

The deterministic description is exact only when there are an infinite number of individuals — for realistic population sizes stochastic effects may be important

For the example of two species in competition — and more generally — would like to:

- Formulate an individual based model (IBM) which is stochastic
- Find the deterministic equation that it approaches for large population sizes (expect that population-level descriptions, such as the Lotka-Volterra equation, should emerge by taking the  $N \rightarrow \infty$  limit of the IBM)
- Carry out the same procedure for spatially explicit models
- Investigate the nature of the stochasticity for large, but finite,  $N$

## 2. The master equation, with examples

We begin with a stochastic system which consists of a set of  $N$  states labelled  $n = 1, 2, \dots, N$

For example,  $n$  could be the number of individuals in a population or the number of occupied patches in a metapopulation (to describe the number of individuals in two populations we need to generalise to  $\mathbf{n} = (n, m)$ )

There will be a transition rate from state  $n$  to state  $n'$  caused by births, deaths, competition, predation, colonisation, extinction,... This rate will be denoted by  $T(n'|n)$ .

Then the probability of finding the system in state  $n$  at time  $t$  changes according to the master equation:

$$\frac{dP(n, t)}{dt} = \sum_{n' \neq n} T(n|n') P(n', t) - \sum_{n' \neq n} T(n'|n) P(n, t).$$

For a one-step process, transition rates are only non-zero when  $n' = n + 1$  and  $n' = n - 1$ :

Then the master equation takes the simpler form

$$\begin{aligned} \frac{dP(n, t)}{dt} &= T(n|n+1)P(n+1, t) + T(n|n-1)P(n-1, t) \\ &\quad - \{T(n-1|n) + T(n+1|n)\} P(n, t) \end{aligned}$$

Given the  $T(n|n')$ , the master equation is a differential-difference equation for the probabilities  $P(n, t)$

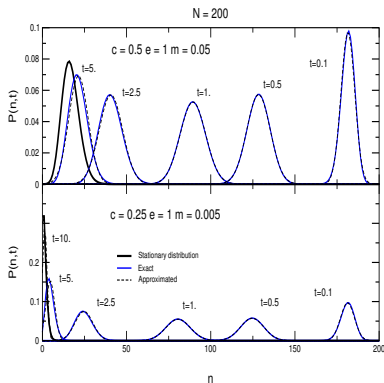
A simple example is the birth and death process:

$$T(n+1|n) = \lambda n \quad (\text{birth}) \quad \text{and} \quad T(n-1|n) = \mu n \quad (\text{death})$$

where  $\lambda$  and  $\mu$  are constants [D. G. Kendall, *Biometrika*, **35**, 6 (1948)]

What methods are available to analyse master equations in general?

For large  $N$  can use van Kampen's approximation scheme — writing  $n = N\phi(t) + N^{1/2}x$  and expanding master equation in powers of  $1/\sqrt{N}$ .

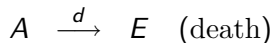
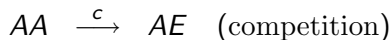


To leading order ( $N \rightarrow \infty$ ) get equation for  $\phi(t)$  (macroscopic equation).  
 Next order get the Gaussian broadening of  $P(n, t)$ . Next order after this  
 non-Gaussian corrections

## Examples

### 1. The simplest case of competition within one species

We suppose that the population dynamics of the system can be essentially described by three processes:



Here  $E$  is a null or space (in the spatial context) — it allows for the population of species  $A$ ,  $n$ , to grow or decline, since the population of the nulls is  $N - n$

What is the time evolution of the system?



- Sample the patch — for  $\mu$  of the time randomly choose two individuals and allow them to interact. For  $(1 - \mu)$  of the time choose only one individual randomly
- Simple combinatorics gives:

$$\text{Probability of picking } AA \text{ is } \mu \frac{n}{N} \frac{(n-1)}{(N-1)}$$

$$\text{Probability of picking } AE \text{ is } 2\mu \frac{n}{N} \frac{(N-n)}{(N-1)}$$

$$\text{Probability of picking } A \text{ is } (1 - \mu) \frac{n}{N}$$

This gives the following transition rates:

$$T(n+1|n) = 2\mu b \frac{n}{N} \frac{(N-n)}{N-1}$$

$$T(n-1|n) = \mu c \frac{n}{N} \frac{(n-1)}{N-1} + (1 - \mu) d \frac{n}{N},$$

Writing  $n/N = \phi(t) + N^{-1/2}x$  in the master equation and equating terms of order  $N^0, N^{-1/2}, N^{-1}, \dots$  gives at leading order the following deterministic equation

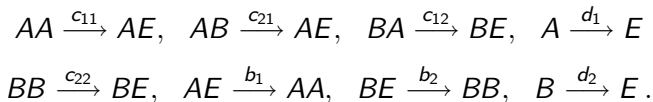
$$\frac{dN_A}{dt} = r_A N_A (1 - N_A/K_A),$$

where  $r_A$  and  $K_A$  are given in terms of  $b, c$  and  $d$

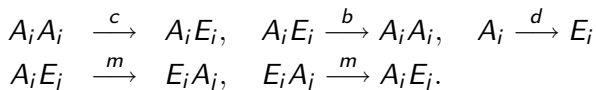
[Phys. Rev. E**70**, 041902 (2004)]

## 2. Spatial model of two species $A$ and $B$ in competition

First, suppose we have two species:



Now let us add space. For the case of one species we have:



We can now find the transition rates and set-up the master equation as before

The population-level model (PLM) corresponding to this model can also be derived:

$$\begin{aligned}\frac{\partial N_A}{\partial t} &= D_A \nabla^2 N_A \\ &+ D_1 (N_A \nabla^2 N_B - N_B \nabla^2 N_A) \\ &+ r_A N_A (1 - N_A/K_A) - b_A N_A N_B\end{aligned}$$

$$\begin{aligned}\frac{\partial N_B}{\partial t} &= D_B \nabla^2 N_B \\ &+ D_2 (N_B \nabla^2 N_A - N_A \nabla^2 N_B) \\ &+ r_B N_B (1 - N_B/K_B) - b_B N_A N_B\end{aligned}$$

These equations have the extra term  $(N_A \nabla^2 N_B - N_B \nabla^2 N_A)$  which does not appear in phenomenological models — which simply add diffusion terms to the non-spatial model

### 3. Metapopulations

A metapopulation is a population of populations — a group of local populations in patches which are either occupied or unoccupied.

Population size of each patch unimportant — focus only on persistence

Patches can become occupied by colonisation from other patches, or may become unoccupied due to extinction of the local population. Suppose that there are  $N$  patches, of which  $n$  are occupied.

If we also allow for the possibility of a mainland “raining down” migrants on to the islands, then there are 3 types of process: extinction, colonisation and migration.

This gives the following transition probabilities:

$$T(n+1|n) = \left(c\frac{n}{N} + m\right) \left(1 - \frac{n}{N}\right)$$
$$T(n-1|n) = e^{-\frac{n}{N}}$$

The  $N \rightarrow \infty$  limit of the model is the Levins model:

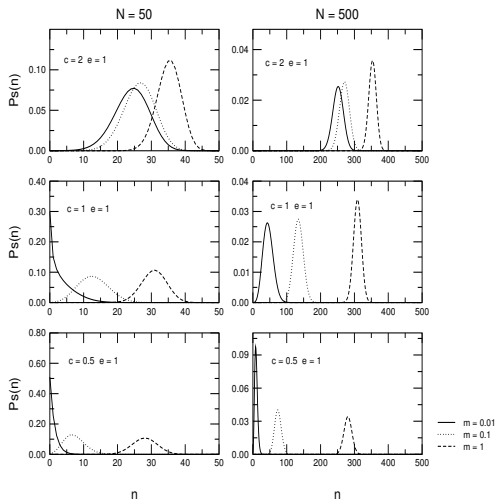
$$\frac{df}{dt} = (cf + m)(1 - f) - ef,$$

where  $f(t)$  is the fraction of occupied sites at time  $t$

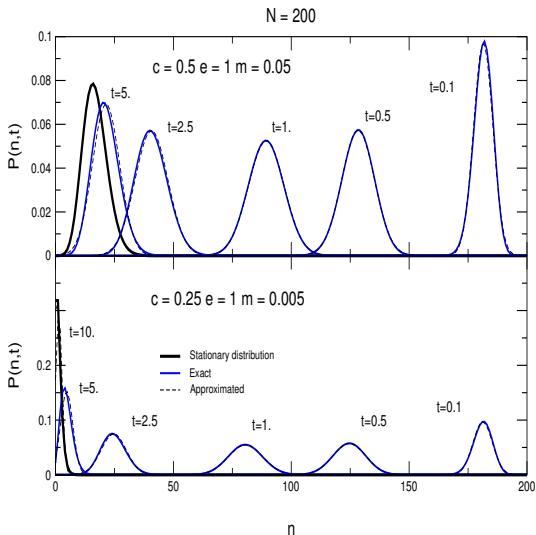
It is also possible to:

- (i) Obtain the stationary probability distribution analytically
- (ii) Carry out a large  $N$  expansion and solve the Fokker-Planck equation obtained at next-to-leading order
- (iii) Obtain the mean time to extinction of the metapopulation through a linear approximation
- (iv) Investigate the spatial version of this model through both a lattice version and a spatially realistic version

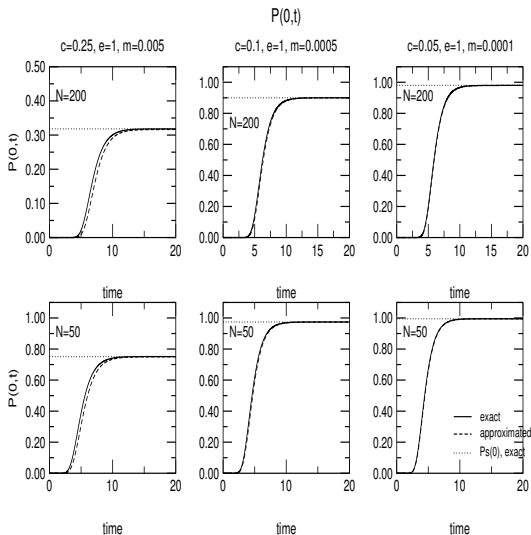
[Bull. Math. Biol. **64**, 913 (2002)]



Stationary probability distribution  $P_s(n)$



“Exact” = Numerical integration of equation; “Approximated” = Large  $N$  approximation



Probability of metapopulation extinction at time  $t$ ; “approximated” refers to a linear approximation used to solve the equations



### 3. Hubbell's neutral theory

[ S. Hubbell, "The unified neutral theory of biodiversity and biogeography" (Princeton, 2001) ]

Imagine an island (the *local community*) consisting of  $J$  individuals of  $S$  possible species (typical values used in simulations might be  $J$  equal to tens of thousands and  $S$  in the hundreds). The number of individuals is fixed, and not all possible species may be represented on the island

Immigration occurs from a *metacommunity*, but in such a way that any immigrants simply displace individuals in the local community, so that  $J$  still remains fixed

The metacommunity may be thought of as a large area (a continent, for instance) and the local community a small area of the metacommunity, or an island off the metacommunity mainland, or perhaps more accurately as a *sample* from the metacommunity

For the moment, let's forget about the metacommunity and focus on the local community

Hubbell's theory for the local community is based on two principles: (i) zero-sum dynamics (fixed number of individuals in the community), (ii) ecological equivalence of all individuals

We have mentioned (i) already. Principle (ii) is the controversial assumption of neutrality: neglect all differences in birth, growth, death rates, habitat, dispersal,... between trophically similar species. Might be a zeroth-order approximation?

The local community consists of  $N_i$  individuals of species  $i$  with the total number of individuals being  $J$ . The dynamics consists of picking one individual at random from the local community, removing it, and then replacing it with an individual also from the local community

This models birth, death, competition and predation

How can we specify the dynamics of this process?

Focus on one species (it doesn't matter which one — they're all the same!) Call it species  $j$ . At a given time step there is a probability of

$$\frac{N_j}{J}$$

that the individual you select is from species  $j$ . If you now select an individual to replace it which is *not* from species  $j$ , then the number of individuals of species  $j$  decreases by 1. The probability of this happening is

$$\frac{(J-1) - (N_j - 1)}{(J-1)} = \frac{J - N_j}{(J-1)}$$

$$\Rightarrow T(N_j - 1 | N_j) = \frac{N_j}{J} \frac{(J - N_j)}{(J - 1)}$$

Similarly,

$$T(N_j + 1 | N_j) = \frac{(J - N_j)}{J} \frac{N_j}{(J - 1)}$$

Once an initial condition has been specified, this is a well-defined stochastic dynamics which can be analysed using the master equation

This is a Polya process: start off with  $J$  balls of  $S$  possible colours in a bag (colour  $i$  has  $N_i$  balls at a given time) and replace them according to the above rules

It is fairly clear, and can be easily proved from the above, that after a sufficiently long time, all remaining balls will be of the same colour. Not a very biodiverse community!

But the whole point of the model is to couple local communities and regional metacommunities

So implement above dynamics only  $(1 - m)$  of the time. For the rest of the time choose to replace the individual with one from the metacommunity, with a probability  $P_i$ . Here  $P_i$  is the relative abundance of species  $i$  in the metacommunity

So the transition probabilities are modified as follows:

$$T(N_j - 1|N_j) = (1 - m) \frac{N_j}{J} \frac{(J - N_j)}{(J - 1)} + m \frac{N_j}{J} (1 - P_j)$$

$$T(N_j + 1|N_j) = (1 - m) \frac{N_j}{J} \frac{(J - N_j)}{(J - 1)} + m \frac{(J - N_j)}{J} P_j$$

This dynamics does lead to a biodiverse community — in fact the stationary distribution can be found in closed form

$$P_S(N_j) = \binom{J}{N_j} \frac{B(N_j + P_j^*, N_j^* - N_j)}{B(P_j^*, N_j^* - J)}$$

Here  $B(\cdot, \cdot)$  is the beta-function and

$$P_j^* = m \left( \frac{J - 1}{1 - m} \right) P_j, \quad N_j^* = \left( \frac{J - m}{1 - m} \right) - P_j^*$$

[Theor. Popul. Biol. **65**, 67–73 (2004)]

A quantity that is frequently measured in the field is the species abundance distribution,  $S(n)$ . This is the number of species in the community having  $n$  individuals. It is proportional to the stationary probability distribution,  $P_S(n)$ , in the model

To calculate this, one has first to create the metacommunity through birth, death and *speciation* events. Once this is done a master equation for the metacommunity can also be written down and analysed. Again it can be solved exactly

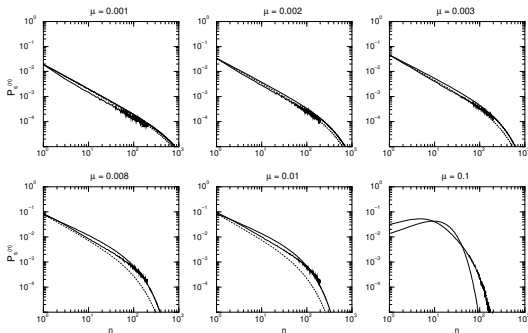
It turns out that Hubbell's neutral theory is able to describe the complexity of natural communities in a very concise way — only two numbers are needed to characterise a natural community in a given locality

These are:  $\theta$  (the fundamental biodiversity number — the potential species richness of the community) and  $m$  (the immigration parameter — its degree of isolation)

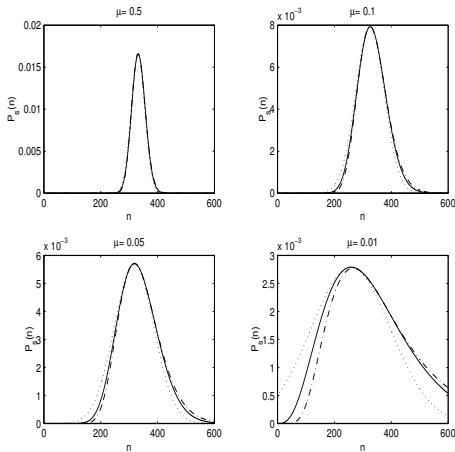
The species abundance distribution in Hubbell's model is generally in good agreement with data

[Ecol. Lett. **7**, 901 (2004), Trends. Ecol. Evol. (in press)]

$N = 5000, S = 300, C = 0.5$



Stationary probability distribution  $P_S(n)$  (solid line) from a model related to Hubbell's [Phys. Rev. E**62**, 8466 (2000)]; log-series approximation (dotted line); simulation of IBM (noisy solid line)  $N = 5000, S = 300, C = 0.5$



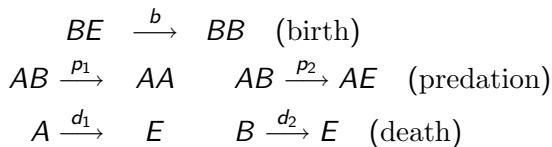
Stationary probability distribution  $P_S(n)$  (solid line) from a model related to Hubbell's [Phys. Rev. E**62**, 8466 (2000)]; log-normal approximation (long-dashed line); normal approximation (dotted line)  
 $N = 100000$ ,  $S = 300$ ,  $C = 0.5$



## 4. Stochastic amplification

The IBM for the simplest type of predator-prey model consists of  $n$  individuals of species  $A$  (the predators) and  $m$  individuals of species  $B$  (the prey)

The processes are taken to be



Defining new rates

$$\tilde{d}_i = (1 - \mu) \frac{d_i}{N} \quad ; \quad \tilde{p}_i = \mu \frac{p_i}{(N-1)} \quad ; \quad \tilde{b} = \mu \frac{b}{(N-1)}$$

gives the following transition rates:

$$\begin{aligned}
 T(n-1, m|n, m) &= \tilde{d}_1 n \\
 T(n, m+1|n, m) &= 2\tilde{b} \frac{m}{N} (N-n-m) \\
 T(n, m-1|n, m) &= 2\tilde{p}_1 \frac{n}{N} m + \tilde{d}_2 m \\
 T(n+1, m-1|n, m) &= 2\tilde{p}_2 \frac{n}{N} m
 \end{aligned}$$

Analyse with the system-size expansion:

$$\frac{n}{N} = f_1(t) + \frac{x}{\sqrt{N}} \quad ; \quad \frac{m}{N} = f_2(t) + \frac{y}{\sqrt{N}}$$

and expand systematically in powers of  $1/\sqrt{N}$ , equating terms of order  $N^0, N^{-1/2}, N^{-1}, \dots$

The explicit form of the population level model is found to be

$$\begin{aligned}
 \frac{df_1}{dt} &= n(f_2)f_1 - df_1 \\
 \frac{df_2}{dt} &= r f_2 \left(1 - \frac{f_2}{K}\right) - g(f_2)f_1
 \end{aligned}$$

The constants, numerical response and functional response are given by

$$d = d_1, \quad r = 2b - d_2, \quad K = 1 - (d_2/2b)$$

$$n(f_2) = 2p_1 f_2 ; \quad g(f_2) = 2(p_1 + p_2 + b)f_2$$

Note that  $n(f_2) = \lambda g(f_2)$

[Phys. Rev. Lett, **94**, 218102 (2005)]

So the simplest individual based model gives the Volterra model, as expected. The question is now: what is the stochastic behaviour of the model for relatively large values of  $N$ ?

The  $1/\sqrt{N}$  corrections to the population level model (PLM) are described by a linear Fokker-Planck equation, or alternatively by a set of linear Langevin equations

$$\begin{aligned}\dot{x} &= a_{11}x + a_{12}y + \eta_1(t), \\ \dot{y} &= a_{21}x + a_{22}y + \eta_2(t),\end{aligned}$$

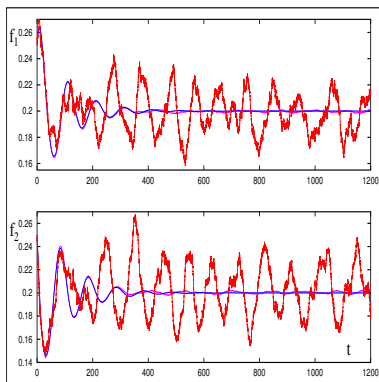
where  $\eta_i(t)$  is a Gaussian white noise with zero mean and

$$\langle \eta_i(t)\eta_j(t') \rangle = b_{ij}\delta(t - t')$$

The  $a_{ij}$  and  $b_{ij}$  are calculated from the  $1/\sqrt{N}$  expansion in terms of the original parameters of the stochastic model

At the fixed point of the PLM (after the transients have died away) these are constants and so since the Langevin equations are linear they can be solved exactly

As expected, the ensemble averaged population density of the IBM, determined from numerical simulation (purple line), agrees perfectly with the solution of the PLM (blue line) showing a decaying oscillatory transient followed by a constant steady-state density



In marked contrast, individual realisations of the IBM show large persistent cycles (red lines). The amplitude of these cycles is much larger than would be naively expected

To search for oscillations in noisy data, one of the most useful diagnostic tools is the power spectrum  $P(\omega) = \langle |\tilde{x}(\omega)|^2 \rangle$ , where  $\tilde{x}(\omega)$  is the Fourier transform of  $x(t)$ . Taking the Fourier transform of the Langevin equations we find

$$P(\omega) = \frac{\alpha + \beta\omega^2}{[(\omega^2 - \Omega_0^2)^2 + \Gamma^2\omega^2]}$$

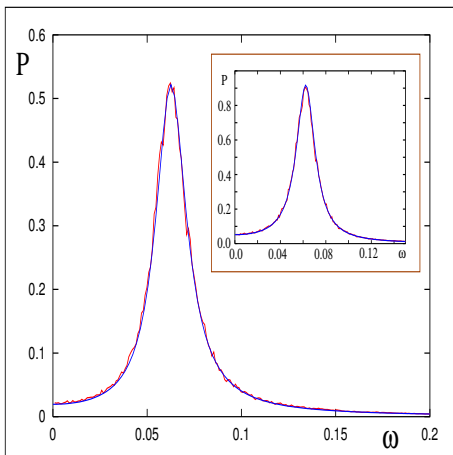
where  $\alpha$  and  $\beta$  are functions of the IBM rates:

$\alpha = b_{11}a_{22}^2 + 2b_{12}a_{12}|a_{22}| + b_{22}a_{12}^2$  and  $\beta = b_{11}$ . The constants in the denominator have the especially simple forms:  $\Omega_0^2 = a_{12}|a_{21}|$  and  $\Gamma = |a_{22}|$ .

The spectrum given above is reminiscent of that for a simple mechanical system — namely a linear damped harmonic oscillator, with natural frequency  $\Omega_0$  and driven at frequency  $\omega$ . Unlike the case of a mechanical oscillator the driving frequency is not tuned to achieve resonance

The noise which drives the system is not external, but arises from the demographic stochasticity contained in the individual processes which define the model; there is no environmental stochasticity in this model

The spectrum predicted by the above equation gives the solid line shown below. The agreement with the spectrum obtained from simulation of the IBM (noisy line) is very good



The naive  $O(1/\sqrt{N})$  estimate of the size of stochastic fluctuations is comparable to the zero frequency value of  $P(\omega)$ . This clearly illustrates the very large amplification of these fluctuations due to the resonance effect

The amplitude of the cycles will be roughly given by

$$\text{Amplitude} \sim \frac{R}{\sqrt{N}}$$

where  $R$  is related to the height of the resonant peak

So for instance, if  $R = 30$ , then for  $N \sim 1000$  the amplitude of the cycles are of order unity, whereas without the resonance they would be only 0.03.

Nevertheless, in the deterministic limit,  $N \rightarrow \infty$ , the cycles disappear



## Other Applications

### 1. Cycles in Epidemics

[D. Alonso, A. J. McKane, M. Pascual]

A stochastic version of the SIR (Susceptible-Infected-Recovered) model can be constructed by introducing the following four processes:

(i) Infection. An infected individual may come into contact with an susceptible individual giving rise to two infected individuals. This is assumed to take place with a rate  $\beta$ . A susceptible may also be infected by an external agent (that is, from the environment). This is assumed to happen with a rate  $\eta$ . These two mechanisms may be expressed as

$S I \xrightarrow{\beta} II$  and  $S \xrightarrow{\eta} I$ .

(ii) Death. This is a demographic effect rather than an epidemiological one, and consequently all three types of individuals are assumed to have the same death rate. These are represented by  $S \xrightarrow{\delta} E$ ,  $I \xrightarrow{\delta} E$  and  $R \xrightarrow{\delta} E$ .

(iii) Birth. This is again a demographic effect, but with all newly-born individuals being susceptible:  $E \xrightarrow{b} S$ .

(iv) Recovery. The rate of recovery of individuals from the infected class is taken to be  $\gamma$ :  $I \xrightarrow{\gamma} R$ .

Analysis of the master equation using the system-size expansion reveals:

- Cycles are observed for parameter values where damped oscillations were found in the deterministic model. The period of the cycles does not necessarily correspond to that of the decaying deterministic oscillations and so cannot generally be obtained from the deterministic system
- The cycles are robust and exist for a large range of parameter values. For example, they do not change appreciably when seasonality in transmission is assumed (as a result of school terms, for instance)

## 2. Cycles in biochemical reactions in cells

[q-bio.CB/0604001]

In cells the number of constituents in biochemical reactions are sufficiently small that the amplification phenomenon may be quite pronounced

As an example, one of the simplest models of glycolysis is due to Sel'kov. A key parameter in this model is  $\gamma$  (the number of ADP molecules required to activate a key enzyme)

Within the deterministic modelling framework it is found that  $\gamma > 1$  is required for cyclic behaviour to emerge from the model (even though the biochemistry of the the enzyme demands that  $\gamma = 1$ ), and that even then, the cycling exists only for a very narrow range of parameter values

When Sel'kov's scheme is reformulated using the stochastic framework, we find stochastically induced oscillations for a wide-range of parameters when  $\gamma = 1$

We also considered a simple model which has been used to explain circadian rhythms, and again found oscillations in parameter ranges where they were absent in the deterministic equations

## 5. Conclusions

Stochastic effects occur in modelling many different biological systems. Master equations — and the associated formalism — are a frequently a useful way of analysing these systems

The  $1/\sqrt{N}$  expansion provides a systematic method for investigating stochastic models of these types. It can uncover effects such as cycles in models of predator-prey interactions, epidemiology and biochemical reactions in the cell, which are significant for moderate values of  $N$ , but which vanish when  $N \rightarrow \infty$

Neutral theory has great merit in being relatively easy to formulate and in giving concrete predictions for interesting questions. Neutrality seems to be an important effect; at the very least, neutral theories serve as null models

In the field of ecological modelling the challenge is to go beyond neutral theory, and to construct models which are as clearly formulated, as amenable to analysis, and with as much predictive power as neutral theory