Cell displacement; MC optimisation; varying motility

Steven Court

August 24, 2016

1 Distance travelled by cells

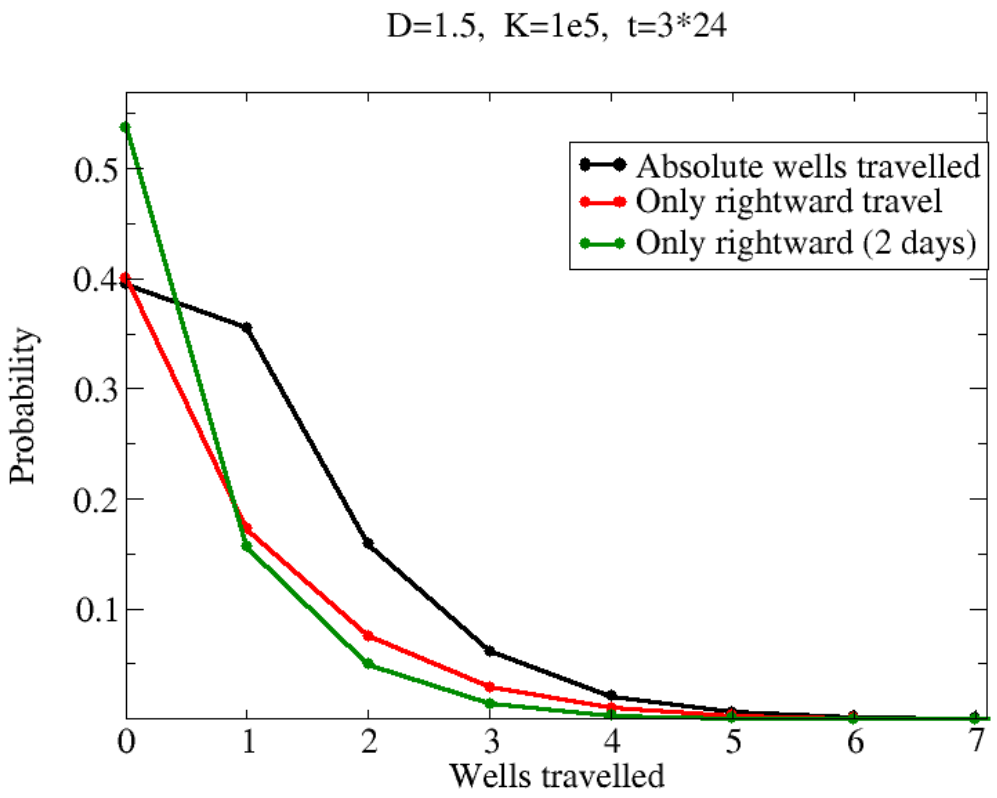

Figure 1: The probability that a given cell is found a certain number of wells away from it's well of birth (with bacterial diffusion constant $D = 1.5$ mm^2/h and carrying capacity $K = 10^5$). The black curve shows the absolute distance (cells can migrate left and right) travelled while the red curve ignores any cells that have migrated to a previous well, both after 3 days. The green curve shows the probability for only rightward travelling cells after 2 days. We see that after 3 days the minority, 40% of cells, are found in the well in which they were born. 17% of cells have advanced 1 well to the right after 3 days, about 8% have advanced 2 wells and 3% 3 wells. Might be interesting to produce these plots for different ciprofloxacin profiles.

2 Monte Carlo parameter optimisation

I re-wrote the program as a Tau-leaping algorithm in the hope that it would be efficient enough to use as part of an optimisation to determine unknown parameters. The first thing we tried was to produce a time-series of population density for each well using the deterministic ODEs for a discrete Fisher equation. We used this as our experimental data and taking the diffusion rate through each channel as our parameter set, tried to find the (aleady known) true values. To compare the stochastic simulations with the data set, I measure the distance between the 2 at a specific point in time, t, as:

$$d_t = \sum_w \left(n_w^{\text{sim}}(t) - n_w^{\text{data}}(t) \right)^2 \tag{1}$$

and sum this value for about 360 points throughout the time series to get the overall distance between a given stochastic run and the data as:

$$d = \sum_t d_t. \tag{2}$$

Attempting an optimisation revealed that the problem is quite tricky and that I need to think more carefully about our distance measure. The first issue is that for a fixed set of parameters (i.e. the correct set!) there can be quite a large stochastic spread in distance measures due to the random nature of the simulations. Figure 2 shows this and Figure 3 shows 2 overlays of the "experimental" and simulation time series illustrating how this can be the case.

(Aside: the initial implementation I did assigned diffusion parameters to each *well* rather than *channel*. Hence the left and right migration rates from a well were the same, rather than the left and right rates across a given channel being the same. This leads to odd dynamics in which certain cells can act as sinks, gathering populations much higher than their carrying capacity and huge distance measures. Fixing the parameter across each channel means that each well has the same steady state density and the optimisation works much better.)

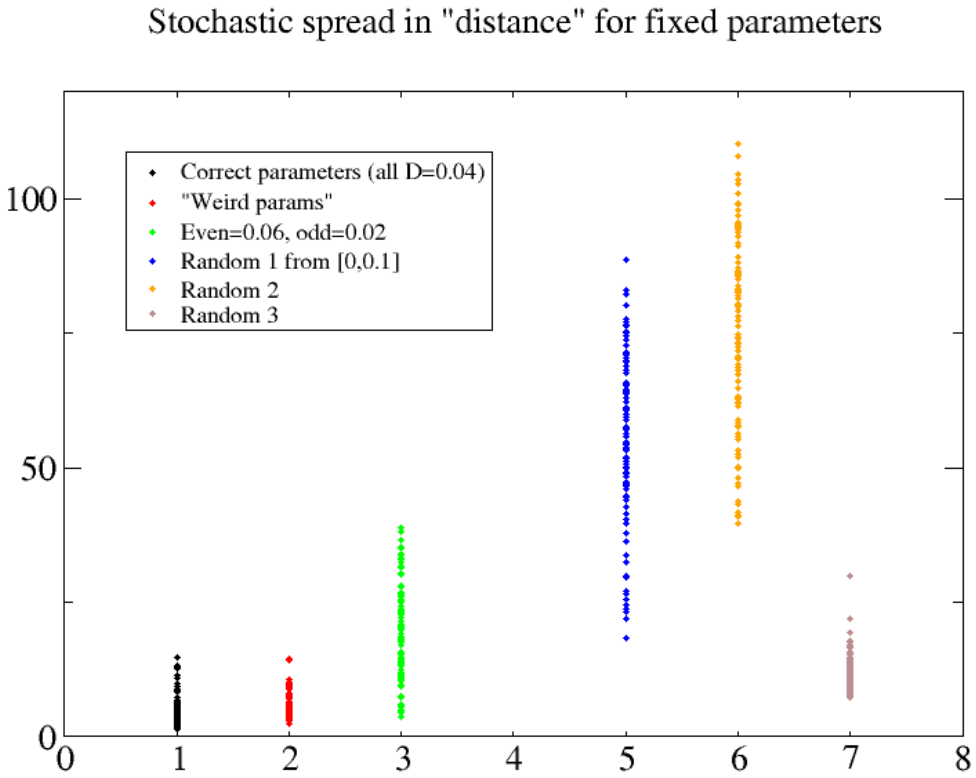

Figure 2: The stochastic spread in the distances measured for 100 simulations using the same parameter sets. Black points use known correct values (all migration rates 0.04) and give distance measurements from around 1.5 to 15. The red points show the results of a very different parameter set – $[0.025, 0.024, 0.038, 0.090, 0.034, 0.018, 0.137, 0.024, 0.080, 0.047, 0.024, 0.047, 0.114, 0.023, 0.043, 0.021, 0.036, 0.032, 0.029, 0.124, 0.100, 0.014, 0.016, 0.0]$ – but which has a similar spread in distances as the true set. This parameter set was found in my optimisation procedure. Green points correspond to the case in which even-numbered channels have a migration rate of 0.06 while odd-numbered channels have a value of 0.02. The other 3 data sets are 3 randomly chosen sets, in which all values are picked uniformly from $[0, 0.1]$. This illustrates that the range of distances and their mean value can vary hugely; there is a lot of potential for random and very different parameter sets to produce distances as good as the true solution.

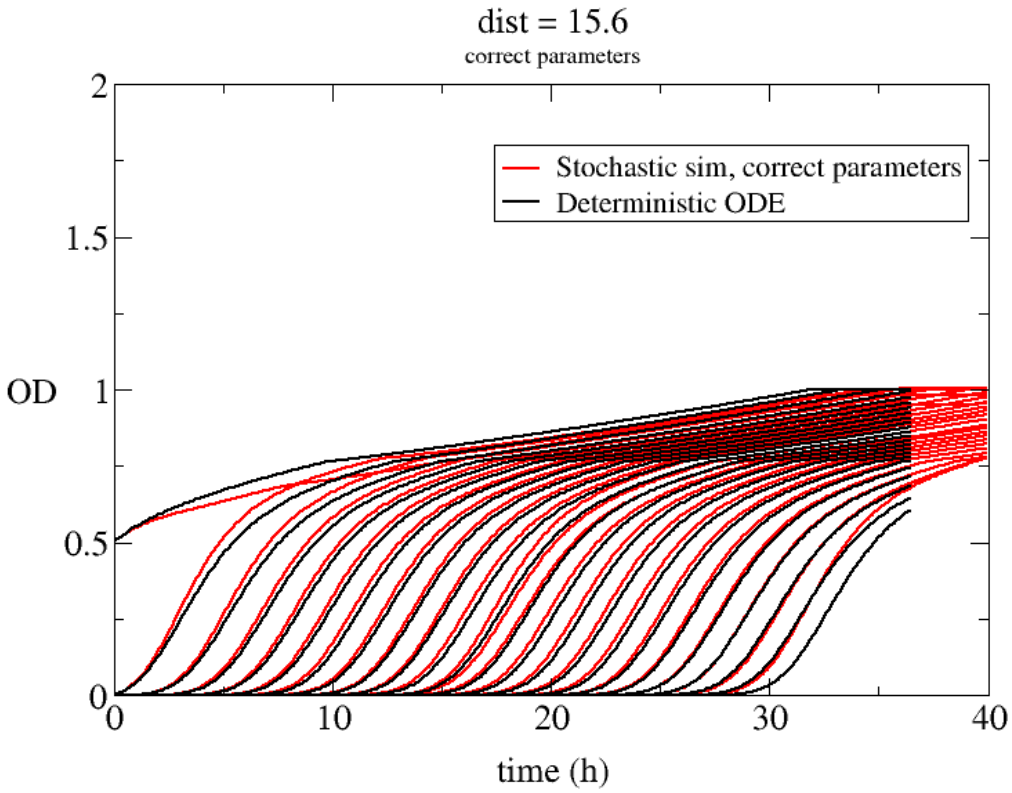

Figure 3: The time-series produced from the deterministic ODEs (black) and that of a single stochastic simulation which produced a high distance measure of 15.6. We can see that somewhere around the middle of the experiment, one well is colonised much earlier than expected and this leads to essentially all later growth curves being shifted by a certain amount, leading to a very high distance measure. For this sort of parameter-determining approach to be useful I think we need to know exactly what we are interested in. What we want to fit the simulation to could be very important in choosing an appropriate measure of distance.

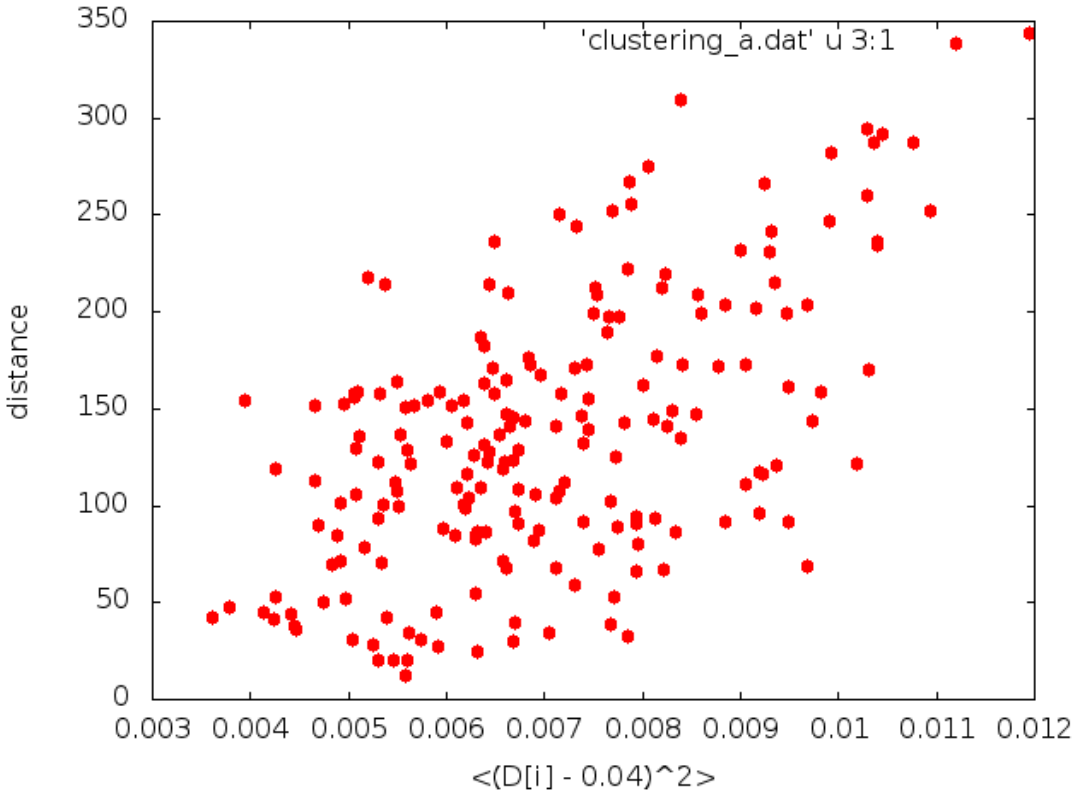

Figure 4: Shows that the measure of distance in general increases as the mean square displacement of the parameters from their true values, $\langle (D[i] - 0.04)^2 \rangle$, increases. For each data point, every parameter was chosen uniformly from $[0, 0.2]$.

3 Varying motility

We discussed two ways to vary motility with ciprofloxacin concentration.

3.1 1: motility depends on MIC of each strain

Here, I impose a hard cutoff: if the concentration of ciprofloxacin is higher than some fraction of a strains MIC then this strain is immotile. I performed 200 simulations for 3 different values: 0.2, 0.5 and 1.0 times the MIC and for two different maximum concentrations, 1000 and 20000 ng/ml. (The 1.0 values should be identical to the case in which there is no cutoff). The results can certainly different when there is a ciprofloxacin-dependent motility, but the results are maybe hard to interpret!

For the case of max cipro = 20000 ng/ml I find very little difference in the weighted mutational graphs and there is no drastic change in the likelihoods of finding the certain genotypes in the final well (Figure **??**). The biggest difference is the the probability of actually reaching the final well (simulations were run for 3 days). With no motility decrease, 144/200 simulations reached the final well. When the cutoff above which cells cannot move was set to 0.5MIC, 126/200 reached the final well and for a cutoff of 0.2MIC this was further reduced to 99/200. So it seems that a ciprofloxacin-dependence on motility will reduce the chance of resistance emerging, but that if it does emerge, for this extremely high concentration there is a relatively small number of ways that this is achieved. In each, 6-7 trajectories account for 80% of all trajectories. From the top 5 followed trajectories in the case with no change in motility, 4 of these are in the top 5 of the 0.5MIC case and 3 are in the top 5 of the 0.2MIC case.

The case of max cipro = 1000 ng/ml was more complicated. It appears that reducing motility increases the diversity of trajectories and noticeably alters the mutational graph of trajectories. In terms of reaching the final well though, no change in motility and the 0.5MIC cutoff both see 200/200 simulations reaching the final well while the 0.2MIC cutoff sees 198/200 – i.e. it thus seems less likely that this change in motility accounts for all of the differences between Bartek's experiments and the simulations.

3.2 2: motility depends on specific genes

The other way to implement a varying MIC we discussed was to attach a cutoff to each gene. For example, a mutated marR or acrR gene prevents motility over a certain concentration of ciprofloxacin but if a strain has a mutated gyrA1/A2 gene, it is always motile, i.e. compensating for the mutated marR/acrR. Setting this limiting concentration to 100, 500 and 1000 ng/ml and again performing 200 simulations for max cipro = 20,000 ng/ml for each, shows that there is very little change in the mutational graphs. There was also little change in the probability of reaching the final well: 138/200, 147/200 134/200 for cutoffs of 100, 500 and 1000 ng/ml (all reasonably close to the case in which there is no cipro-dependent motility, 144/200).

Another way to code this would be to say that a mutated gyrA1/A2 gene confers no change in motility but if the strain has a mutated marR or acrR then the cell becomes immotile past the chosen cutoff (i.e. gyrA mutations do not compensate the mutated marR or acrR genes). In this case the results are very different. For the same values used above and a max cipro of 20,000 ng/ml the mutational graph plots show that we are effectively disallowing marR and acrR mutations and thus severely restricting the number of possible trajectories (to about 3) and only 1 genotype (11100; no marR or acrR) is ever observed in the final well. The probability of reaching the final well for cutoffs of 100, 500 and 1000 are respectively 133/200, 60/200 and 46/200. The most probable pathway by far in each case is the same: 00000-100000-10100-11100.

3.3 Motility summary

The general result is that, perhaps not surprisingly, allowing for a ciprofloxacin-dependent motility can strongly affect the diversity of evolutionary trajectories observed an experiment, and in a non-obvious way. I have implemented this dependence in 3 different ways above and each has produced different results which is interesting: since the effect on motility due to ciprofloxacin is unknown, perhaps there is a way we could compare the simulations to Bartek's experiments to see which is the most likely mode of dependence (though maybe a greater knowledge of filamentation would be useful). I don't know what the best things to compare would be though: likelihood of detecting a given gene in a well, the actual gene abundances and the probability of reaching the final well are all measureable by Bartek. Does Bartek have enough data to say "30% of my 1000 ng/ml experiments reached the final well"? If so, I can choose an implementation and see what range of cutoff values would give rise to a similar proportion. I should also check how much I would need to reduce the mutation rate (base rate or function for cipro-dependent increase) to achieve the same thing. Perhaps it's not a huge change?

One problem is that we are still unsure what other features (and their effects) are missing from the model, filamentation being an obvious one.