Simulations: Experiments in a virtual world

Analytical models versus simulations

- Most models in this course will be analytical: Both the model and the results are described by closed mathematical expressions.
- For example, if the three genotypes at *AA*, *Aa* and *aa* have fitness w_{AA} , w_{Aa} , and w_{aa} , respectively, then the frequeny of the *A* allele changes according to

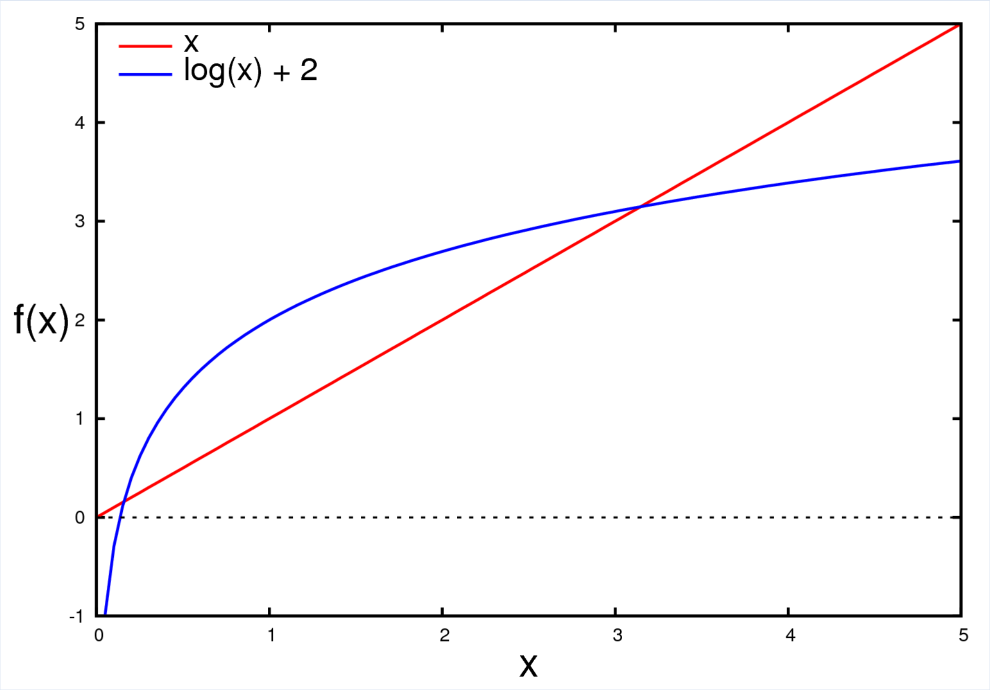
$$p' = \frac{p^2 w_{AA} + p(1-p) w_{Aa}}{p^2 w_{AA} + 2p(1-p) w_{Aa} + (1-p)^2 w_{aa}}$$

- With heterozygote advantage ($w_{Aa} > w_{AA}$ and $w_{Aa} > w_{aa}$) both alleles can be maintained, with the equilibrium frequency of *A* being

$$p_{eq} = \frac{w_{AA} - w_{Aa}}{(w_{AA} - w_{Aa}) + (w_{aa} - w_{Aa})}$$

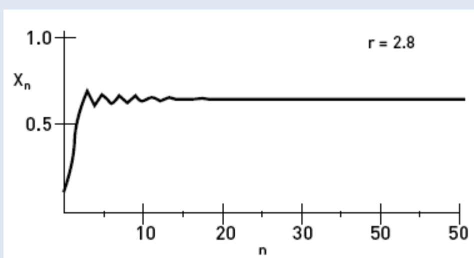
Numerical solutions of analytical expressions

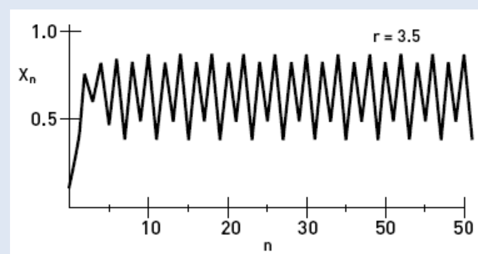
- Sometimes, a closed analytical solution does not exist.
- For example, if for some reason, we need to solve the equation, $x = \ln(x) + a$ for x, this is not possible, but for $a = 2$, we can numerically find two solutions, $x_1 \approx 0.159$ *and* $x_2 \approx 3.146$.

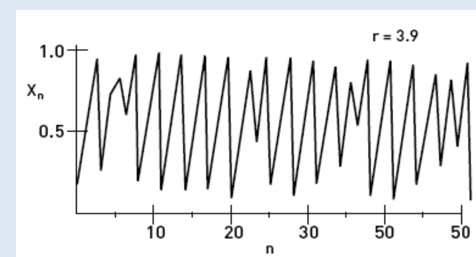


Simulations

- In contrast, in simulations, dynamical rules are set up, and these rules are then repeatedly applied for a given set of parameter values and initial conditions.
- For example, iterating the discrete logistic equation *n*' = *r n* (1-*n*) leads to the plots seen in a previous lecture:







- These solutions are not general, but only hold for specific values of *n*(0) and *r*.
- Of course, the rules for the simulation can be much more complex and include, for example, stochastic elements (i.e. chance events).

Example: Evolution of dispersal

- In fragmented landscapes, local populations are often to small for long-term survival.
- However, a meta-population can survive if local extinctions are balanced by recolonization.
- Classical model by Levins: $$\frac{dp}{dt} = cp(1-p) - ep$$
- Here, *p* is the proportion of occupied patches, *c* is colonization rate, and *e* is extinction rate.
- Simple analytical result: The equilibrium value of *p* is 1 – *e* / *c*. Thus, survival requires *c* > *e*.

Example: Evolution of dispersal

- But: Colonization rate depends on dispersal, which is an individual property.
- Individuals do not disperse in order to ensure population survival, but to maximize their own fitness.
- What happens if habitats are destroyed and migration becomes more dangerous?
- If individuals evolve reduced migration rate, this may increase extinction risk and lead to evolutionary suicide.
- In contrast, if individuals evolve increased migration rate, there may be evolutionary rescue.

Rescue or suicide?

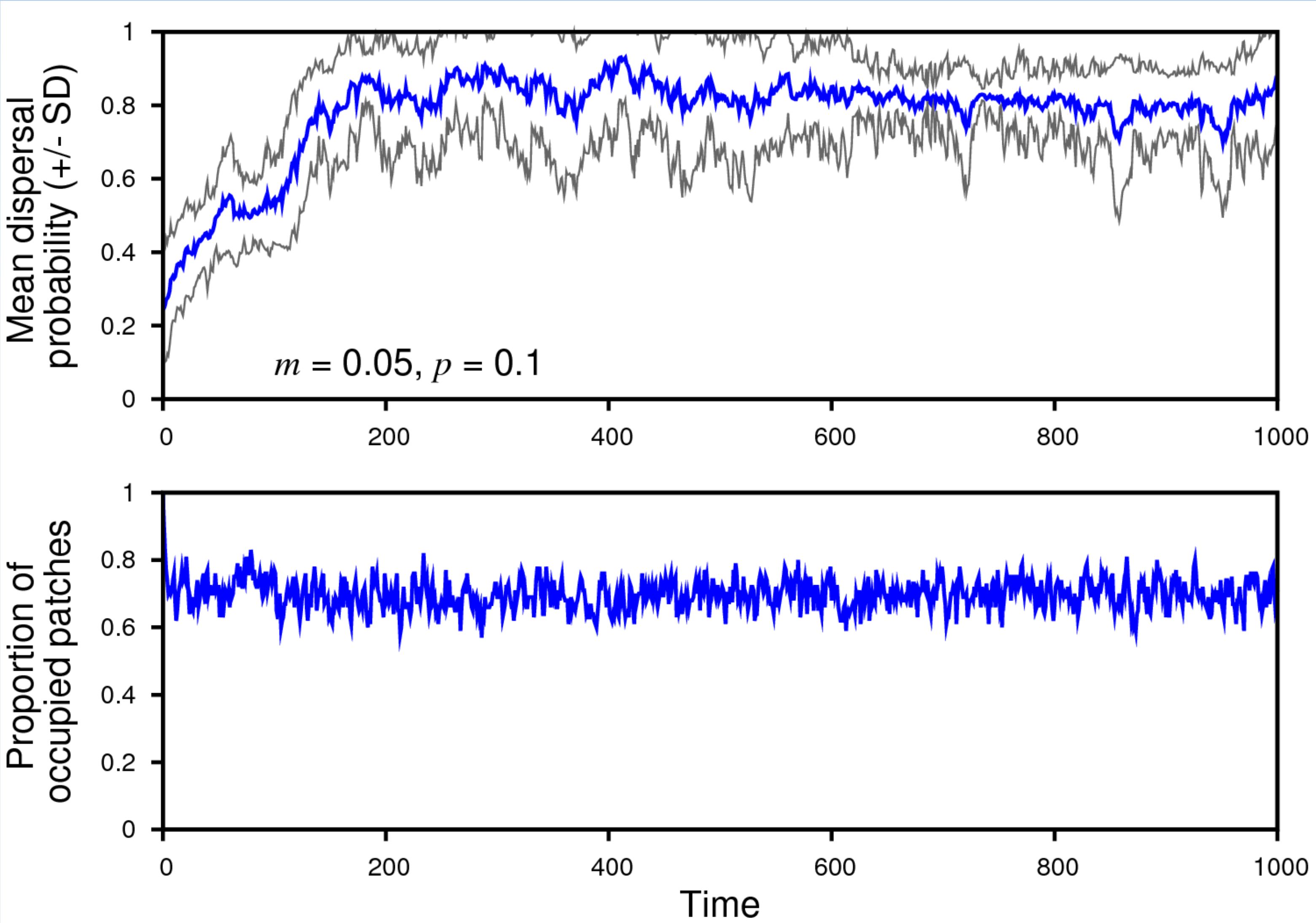
Whether dispersal rate will increase or decrease is hard to predict, because their are two opposing factors at work:

- As the chance of finding a new habitat decreases, individuals might be better off staying where they are.
- On the other hand, if few individuals colonize new habitats, the potential reward in being one of those is high.

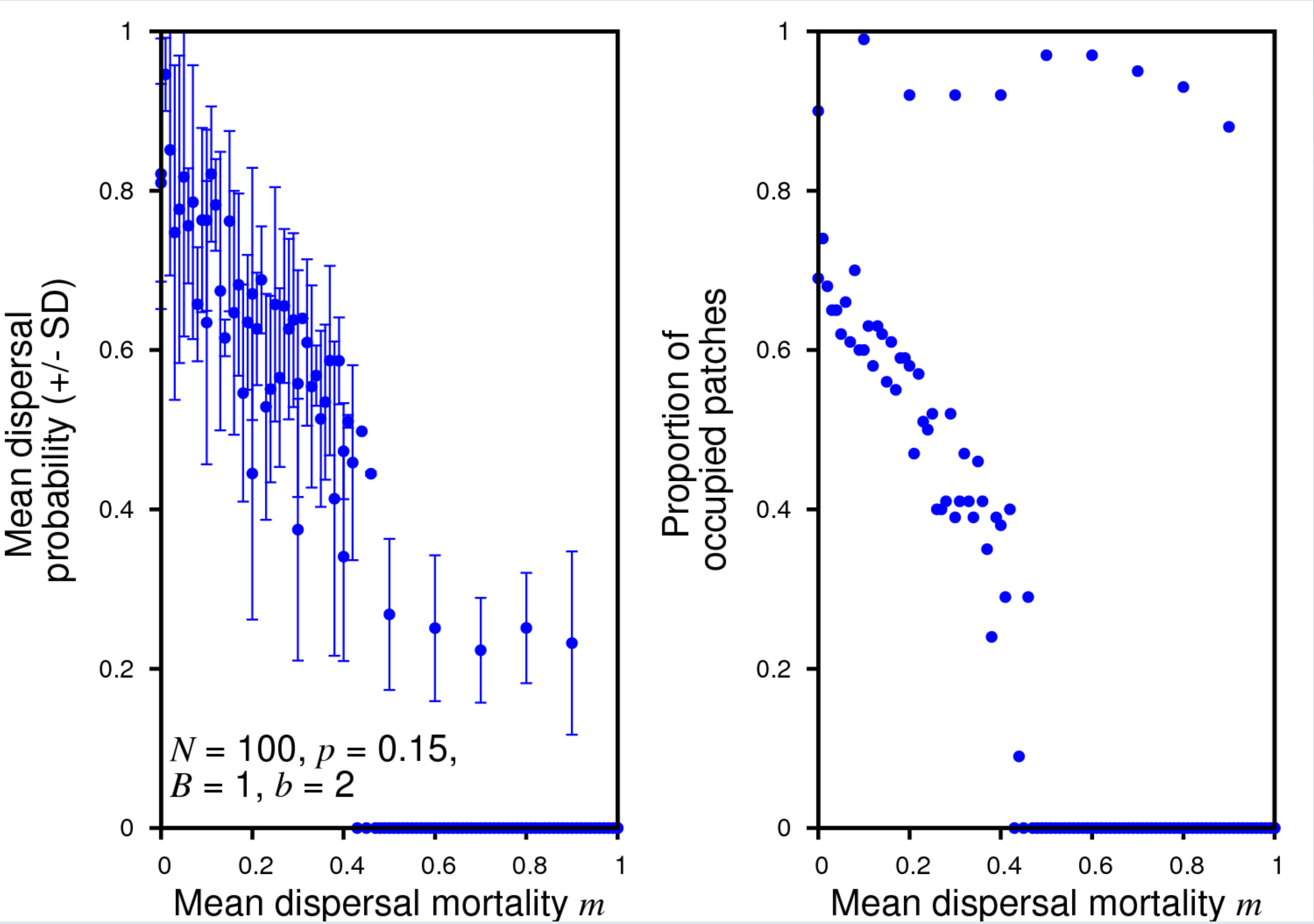
A simulation model

- *N* patches, initially inhabited by one individual each. The patches are identical in size and quality, and every patch can be reached from every other patch with equal probability.
- Individuals are asexual haploids, whose dispersal probability *x* is determined by a single genetic locus. The initial values of *x* are drawn from a uniform distribution between 0 and 1.
- Each generation, a fraction *p* of the patches are temporarily destroyed (i.e. all indivuals are killed).
- In the remaining patches, a maximum of *B* individuals are chosen for reproduction and produce *b* offspring each. All adult individuals die.
- The offspring are subject to mutation with probability *q*. The mutant phenotype is independent of the parent genotype.
- Individuals migrate to random new patches with probabilities determined by their genotype *x*. A fraction *m* of dispersers die.
- This procedure is repeated for 1000 generations.

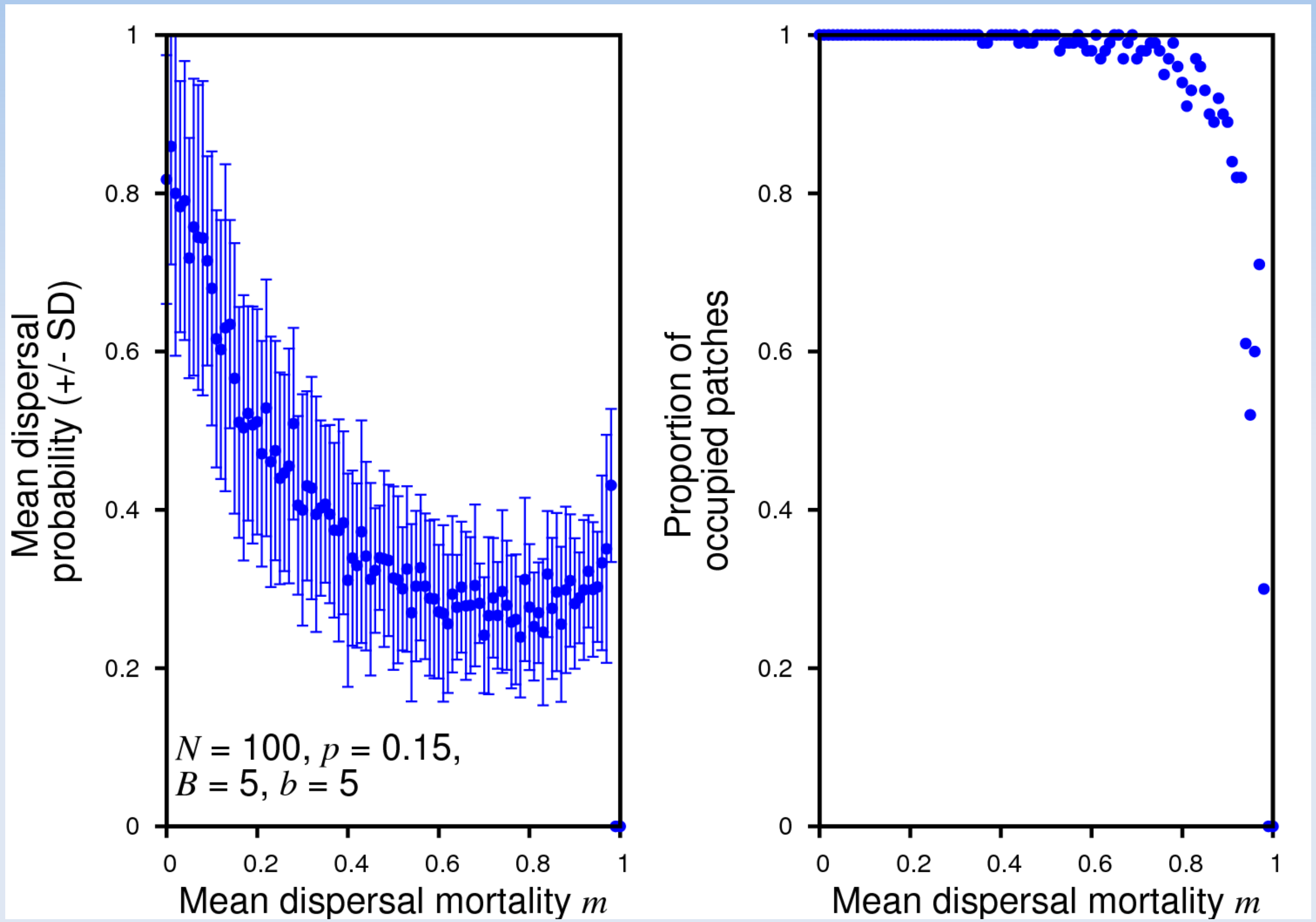
Simulation output



No evolutionary rescue (typical)



Evolutionary rescue



Evolutionary rescue is possible if both the carrying capacity B and the number of offspring *b* are large. Then, the proportion of occupied patches declines sharply for high dispersal mortality *m*, which selects for increased dispersal probabilities.

Types of simulations

- Deterministic versus stochastic
- Individual-based
- Agent-based
- Complex adaptive systems
- Artificial life

Analytical models versus simulations

Advantages of analytical models

- Exact
- General
- One equation says more than 1000 pictures.

Advantages of simulations:

- Scenarios can be much more complex.
- An investigator creates a virtual world, over which he or she has complete control and where experiments can be done at will.
- Simulations are straightforward (in principle) and are guaranteed to yield lots of results.

Disadvantages of simulations

- All results are special cases.
- Parameter space cannot be explored exhaustively.
- Often, one gets a result, but cannot explain it (because one doesn't understand the inner workings of the simulation).

For these reasons, analytical models tend to be more prestigeous than simulation studies.

Nevertheless, many problems can be studied only by simulations.

Futhermore, analytical models and simulations can often be combined.

Uses of simulations

- Learn something about systems that are analytically intractable.
- Validate analytical results (error checking).
- Combine with analytical results. Often analytical results are only possible
  - with additional simplifying assumptions
  - employing approximations
  - for limiting or special cases
- Simulations can be used to test the robustness of the approximations and relax the assumptions of the analytical model.

Examples for analytical approximations

- Weak selection
- Linkage equilibrium
- Infinite population
- Equivalent loci
- Symmetric migration
- Constant genetic variance
- Some factors much stronger / weaker than others (e.g., strong selection, weak mutation)
- Some processes much faster than others (separation of timescales)

Advice for studying simulation models

- Validate the code by checking special cases with known results.
- Start by playing around with the model.
- But then proceed to analyze it systematically (in an automated way).
- Find compact ways of representing your results (squeeze lots of information into a single plot, lots of plots onto a single page etc.).
- Explore as many parameter combinations (and initial conditions) as possible. Don't be afraid of having your computer run for a week.
- Try to get some analytical result, which can be used as a landmark when exploring the simulations.

Simulation software

- Simple iterations of discrete dynamical systems can be done in a spread sheet.
- There is specialized simulations software (like Stella, Simulink) that allows you to specify the model by setting up a flow diagram.
- It is advisable to learn a proper programming language. Programming skills sufficient to write biologically interesting simulations are not too difficult to learn.
- You can use software packages like Mathematica, R, or Matlab (octave, scilab), which have many built-in functions and graphics facilities but are not very fast.
- If simulations need to be fast, you might have to use a low-level language like C or Fortran.