

Class project: genetic drift simulation

What is genetic drift?

Why is it important?

How can we model it?

Population genetics

- The study of changes in gene frequencies within a population, or between populations, is the field of population genetics
- It is an important topic for:
 - Conservation biology
 - Studies of evolution
 - Plant and animal breeding
 - Biotechnology

Gene basics

- Genes – sequences of DNA
 - Some code for proteins
 - Some are regulatory
 - Some (called “pseudogenes”) have no function
- Alleles – alternative forms of a gene
 - Genes will often have more than one allele, sometimes many different ones, in a population
 - An individual can only have at most two different alleles
- Homozygote – an individual in which both maternal and paternal copies of a gene are the same allele
- Heterozygote – an individual in which maternal and paternal copies of a gene are different alleles

Importance of genetic diversity

- Genetic diversity is an important part of conserving biological natural resources
 - Sources of genes for biotechnology
 - Sources of genes for agriculture
 - Sources of genes for disease resistance
- Genetic diversity is also important to the survival of species
 - Individual level = important for immune function (mammalian MHC), reproductive success
 - Population level = increases a population's ability to adapt to changing conditions (new diseases, climate change)

Measuring diversity: heterozygosity

- Any given gene can either be homozygous or heterozygous
- The proportion of genes that are heterozygous is the “heterozygosity” for ***an individual***

Gene	Zygoty
A	Homozygous
B	Heterozygous
C	Heterozygous
D	Homozygous

$$\text{Heterozygosity} = 2/4 = 0.5$$

Population level: gene frequencies, allelic diversity

- Gene frequencies = relative number of each allele in a population
- Allelic diversity = number of alleles in the population
- At a population level, a gene is more diverse if:
 - There are more alleles in the population (greater allelic diversity)
 - Allele frequencies are even (for a given allelic diversity)

Alleles for gene A	Copies of allele found in population	Allele frequency
A1	10	$10/70 = 0.14$
A2	20	$20/70 = 0.29$
A3	5	$5/70 = 0.07$
A4	35	$35/70 = 0.50$
Total	70	

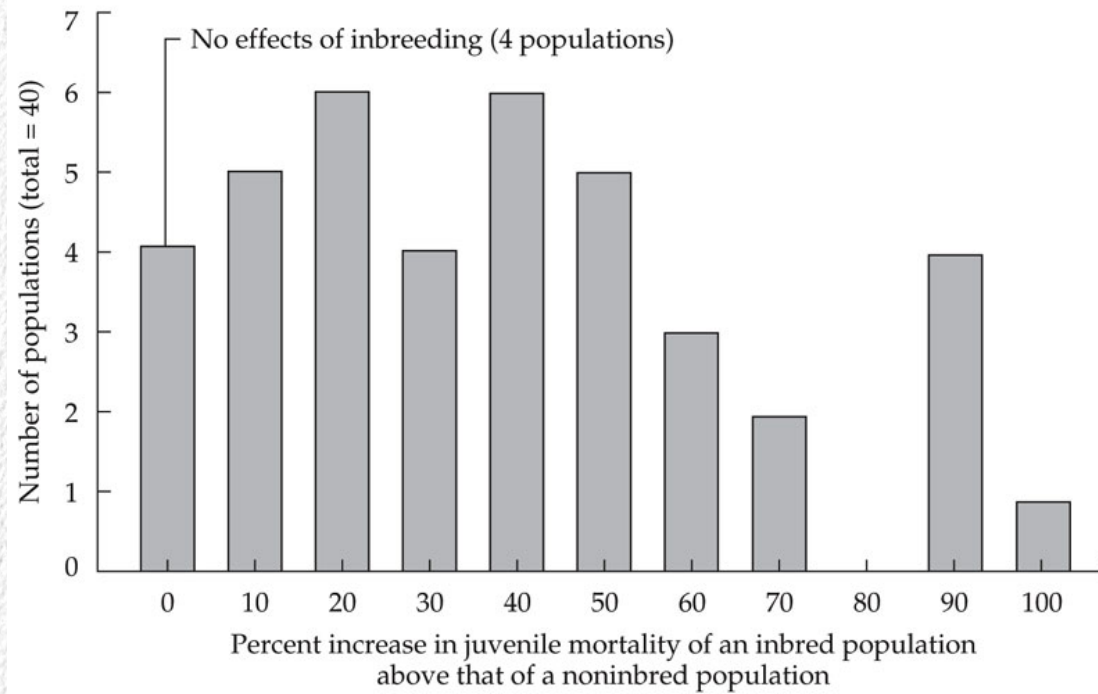
Heterozygosity at a population level

- The proportion of individuals that are heterozygous at a locus is the population heterozygosity
- Across multiple genes, average heterozygosity is often used

Gene	Heterozygous individuals	Heterozygosity
A	15	$15/100 = 0.15$
B	5	$5/100 = 0.05$
C	54	$54/100 = 0.54$
D	26	$26/100 = 0.26$
Total	100	Mean = 0.25

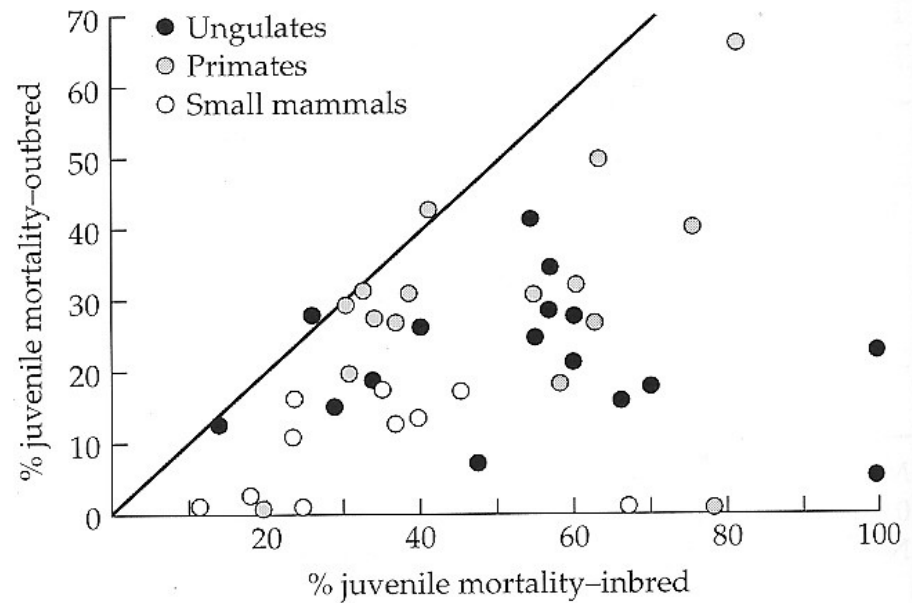
Consequences of reduced genetic diversity: inbreeding depression

- Caused by increase in homozygous recessive alleles (decreased heterozygosity)
- Effects include:
 - Decreased fertility
 - High juvenile mortality
 - Sperm, birth defects,
 - Weak, sterile offspring
 - Impaired immune function



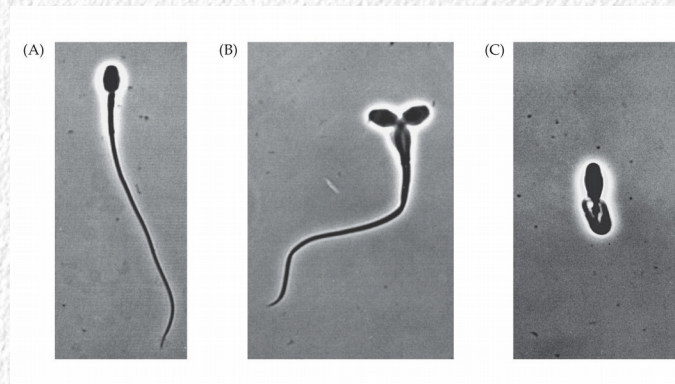
Mammals

ESSENTIALS OF CONSERVATION BIOLOGY, Fourth Edition, Figure 11.6 © 2006 Sinauer Associates, Inc.



Examples of Variability Loss

- Cheetah
 - None of 52 genes tested showed variability - expect 8 to 31%
 - Skin grafts accepted between individuals
 - Abnormal sperm (70%)
- Florida panther (30-50 animals)
 - Sperm defects (93.5), cryptorchid (testes that don't descend), congenital heart defects, decreased immunity
- Ngorongoro Lion – sperm defects



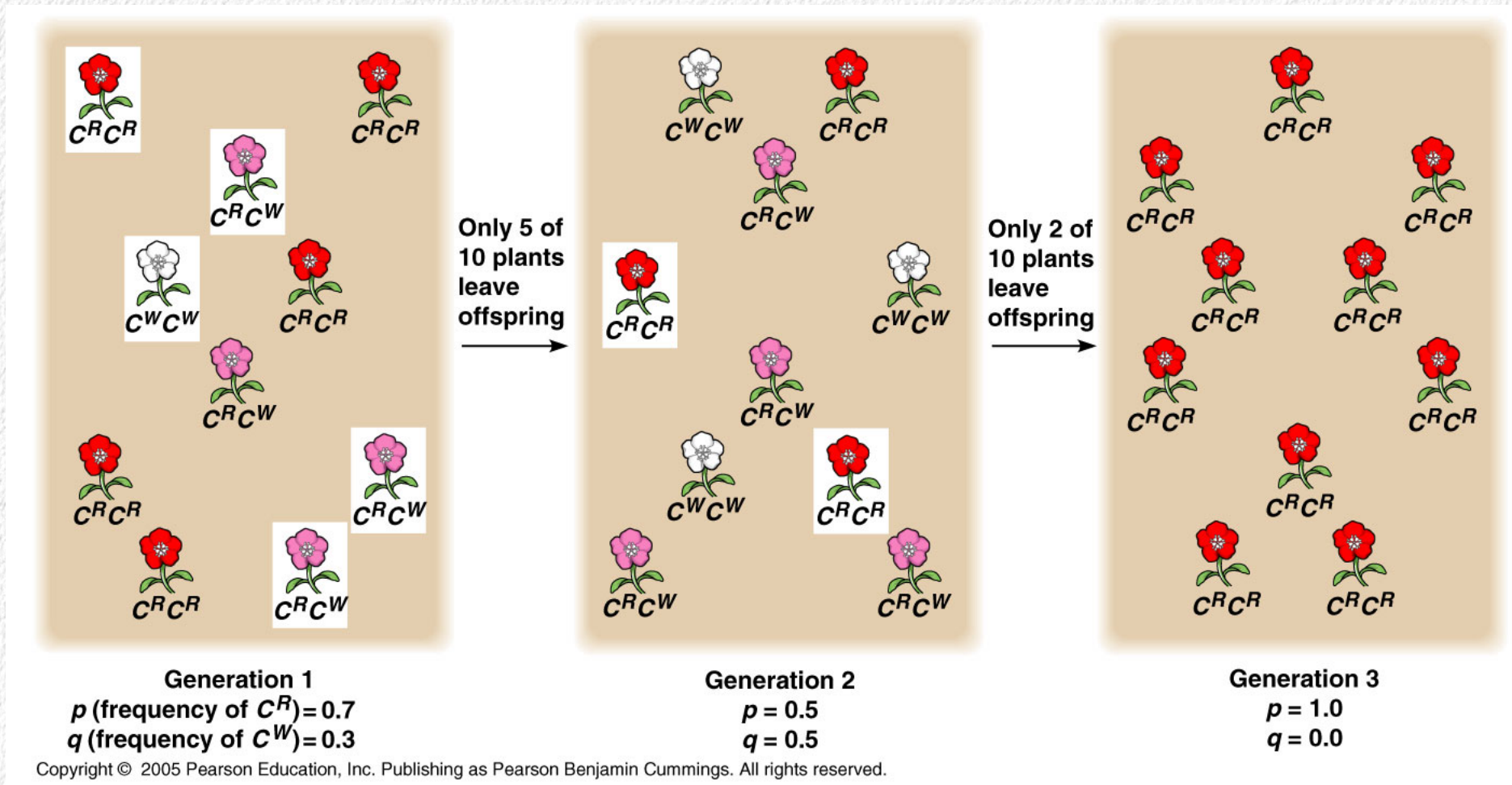
Changes in allele frequencies over time = evolution

- Population genetics definition of evolution = a change in allele frequencies in a population over time
- Evolution by natural selection changes gene frequencies in a predictable way
 - Beneficial mutations are retained, deleterious mutations are removed
- Evolution can also happen due to random changes in gene frequencies = **genetic drift**
- Genetic drift doesn't lead to adaptation, causes loss of diversity, loss of heterozygosity

Changes in gene frequencies

- Gene frequencies are a balance between loss and gain of alleles
- Gains in alleles come from:
 - Mutation
 - Immigration (gene flow)
- Losses of alleles come from:
 - Genetic drift (bottlenecks, founder effects)
 - Natural selection (under some circumstances)

Random changes in gene frequencies – genetic drift



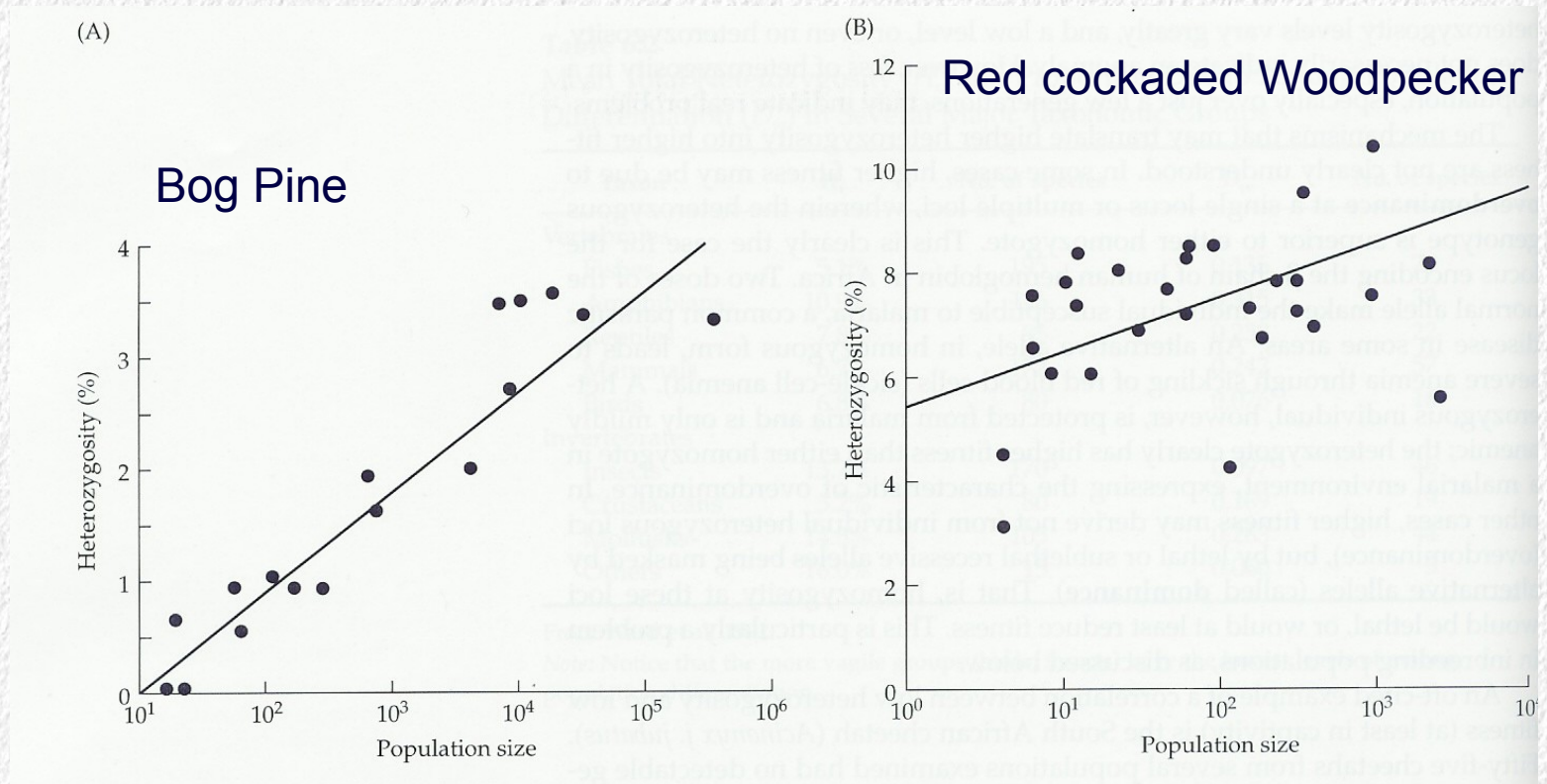
Effect of drift on genetic diversity

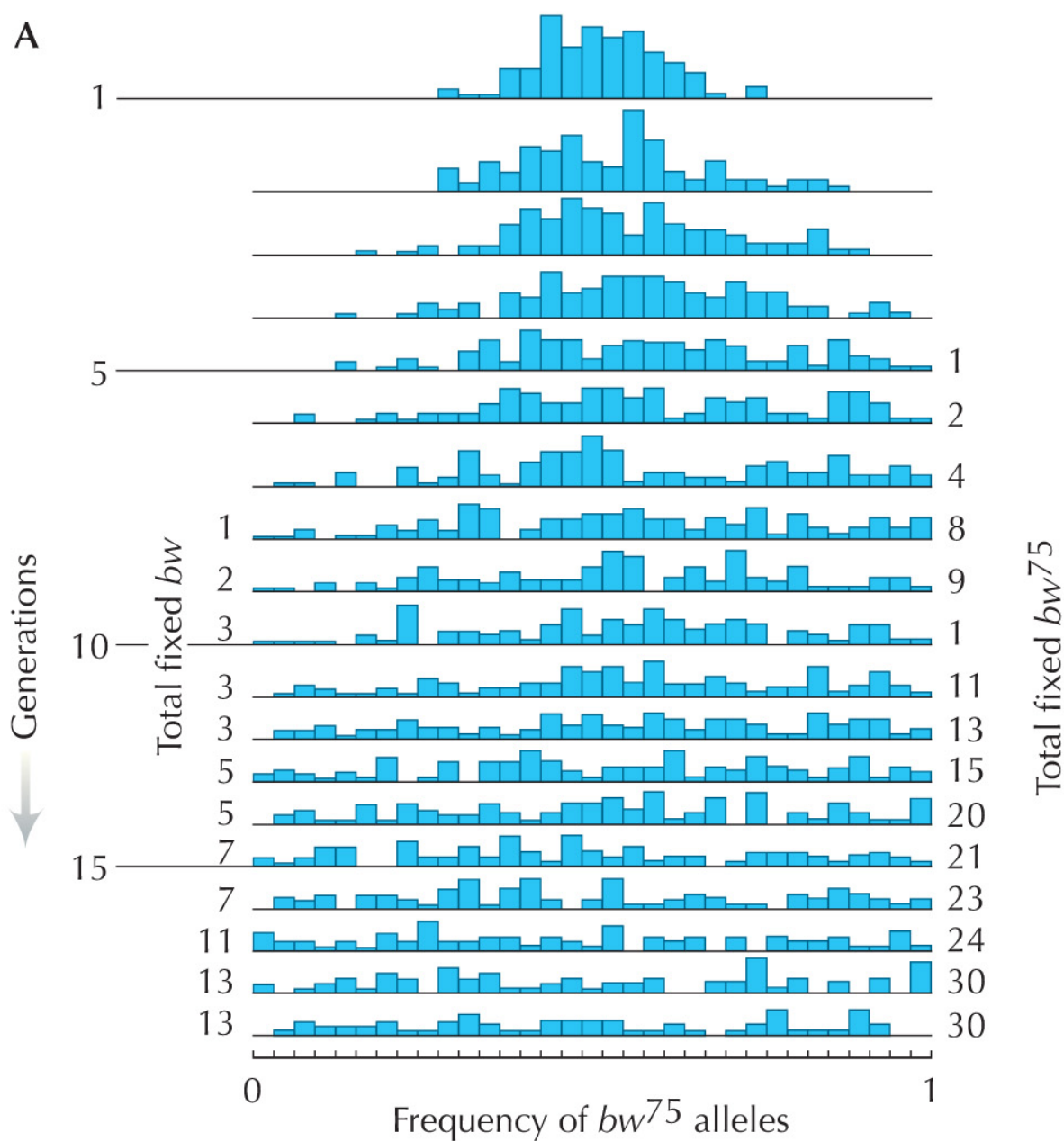
- Random mating → allele frequencies that “drift” from one generation to the next
 - When an allele frequency reaches 0 it is lost
 - When any one allele frequency reaches 1 all other alleles are lost (**fixation**)
- If no new alleles are added, populations will eventually lose genetic diversity due to drift

How do we study genetic drift?

- Field studies, laboratory studies
 - Look for genetic differences among populations in alleles that don't produce a phenotype (pseudogenes)
 - Or, genes that do produce a phenotype but that don't affect the fitness of the individual
- Mathematical models, computer simulations
 - Study the expected behavior of a gene pool subject to random mating
 - Study effects of various factors on drift (mating systems, population size, fluctuating environments, etc.)

Field studies





Experiment on lab populations of *Drosophila*, 8 mating pairs each

Histograms are the gene frequencies for the bw^{75} allele across many populations

All started at a frequency of 0.5

An increasing number of populations became fixed at a frequency of 1 for bw^{75} or 1 for bw , causing loss of the other allele over time

FIGURE 15.3. Random drift in experimental populations of *Drosophila melanogaster*, propagated with eight males and eight females. (A) The distribution of allele frequencies across replicate populations, all started at $p = 0.5$. Populations that had fixed one or the other allele are shown at *left* and *right*.

Mathematical models

- Models are useful for telling us how to expect gene pools to change due to drift
- Allow us to isolate factors
 - Effects of drift at a population size of 100
 - Effects of drift at a population size of 200
 - Effects of drift in the presence of migration
- Simple models can tell us general things, but won't be tailored to a particular species or situation
- Some types of models:
 - Deterministic vs. stochastic
 - Analytical vs. simulation

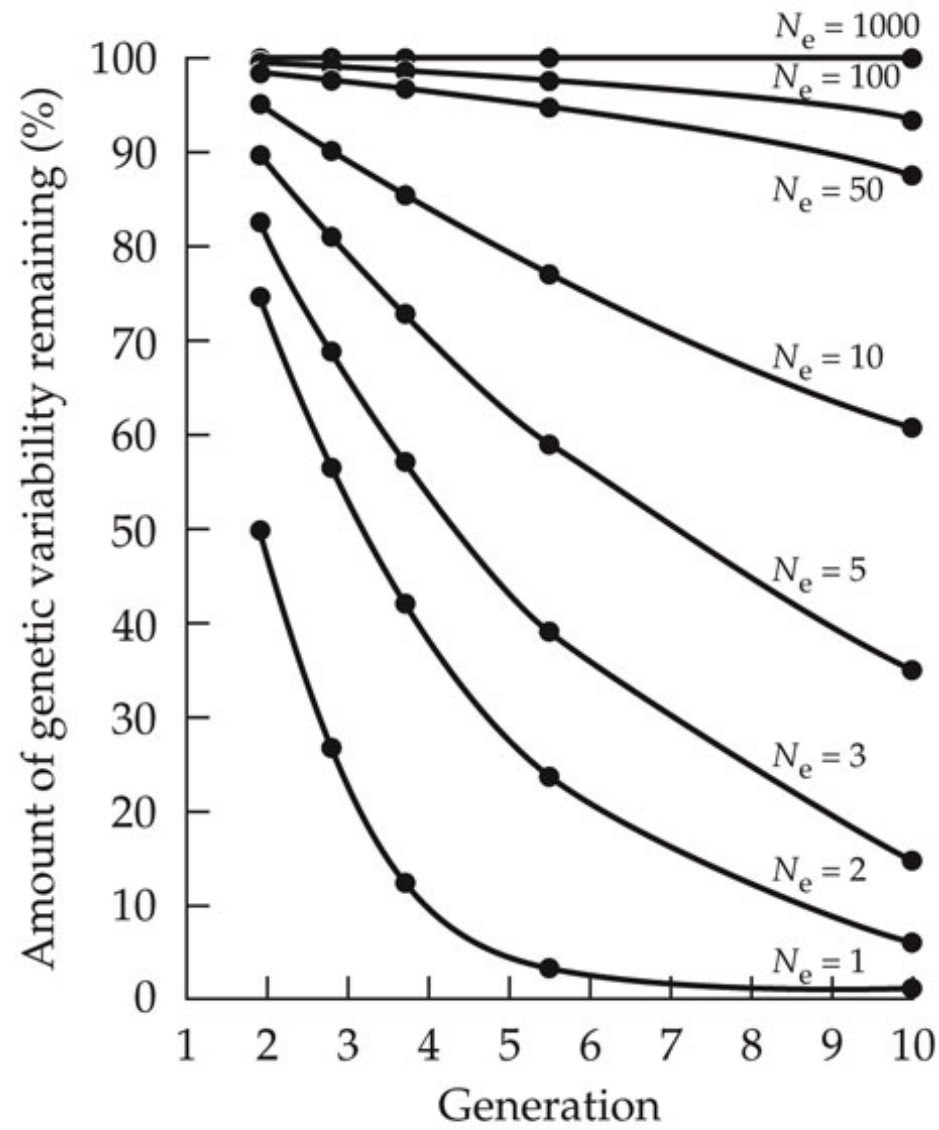
Deterministic model: change in heterozygosity due to drift

- Heterozygosity measures diversity at the individual scale
- Equation: loss of original heterozygosity after each generation due to drift

$$\Delta H = 1 - \frac{1}{2N_e}$$

- Thus, loss in heterozygosity is proportional to $2N_e$
- Deterministic model of a stochastic process – prediction of an average, but there will be variation around it

Proportion HZ remaining



Analytical vs. simulation

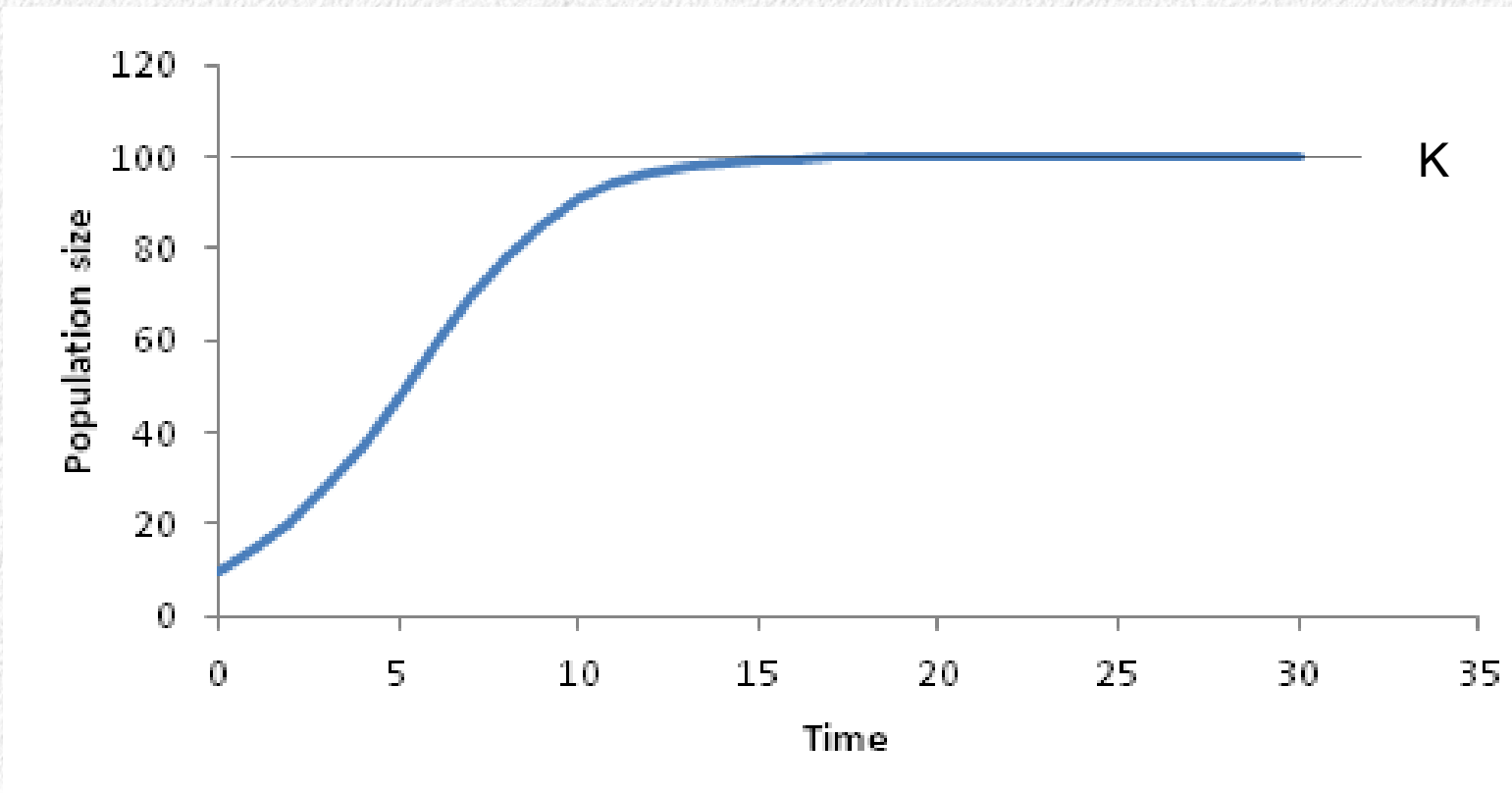
- Analytical models can be solved for parameters of interest
 - Can calculate equilibrium values
- Simulation models can't – need to be run, output studied

Example: deterministic model of logistic growth in discrete time

- Populations grow exponentially when they are small
- As the population size increases, competition intensifies
 - Birth rates decline
 - Death rates increase
- When the population size is enough that births = deaths, the population stops growing
- This is the “carrying capacity” - the number of individuals that can be supported indefinitely
- Discrete time = time passes in chunks, such as years
- Model:

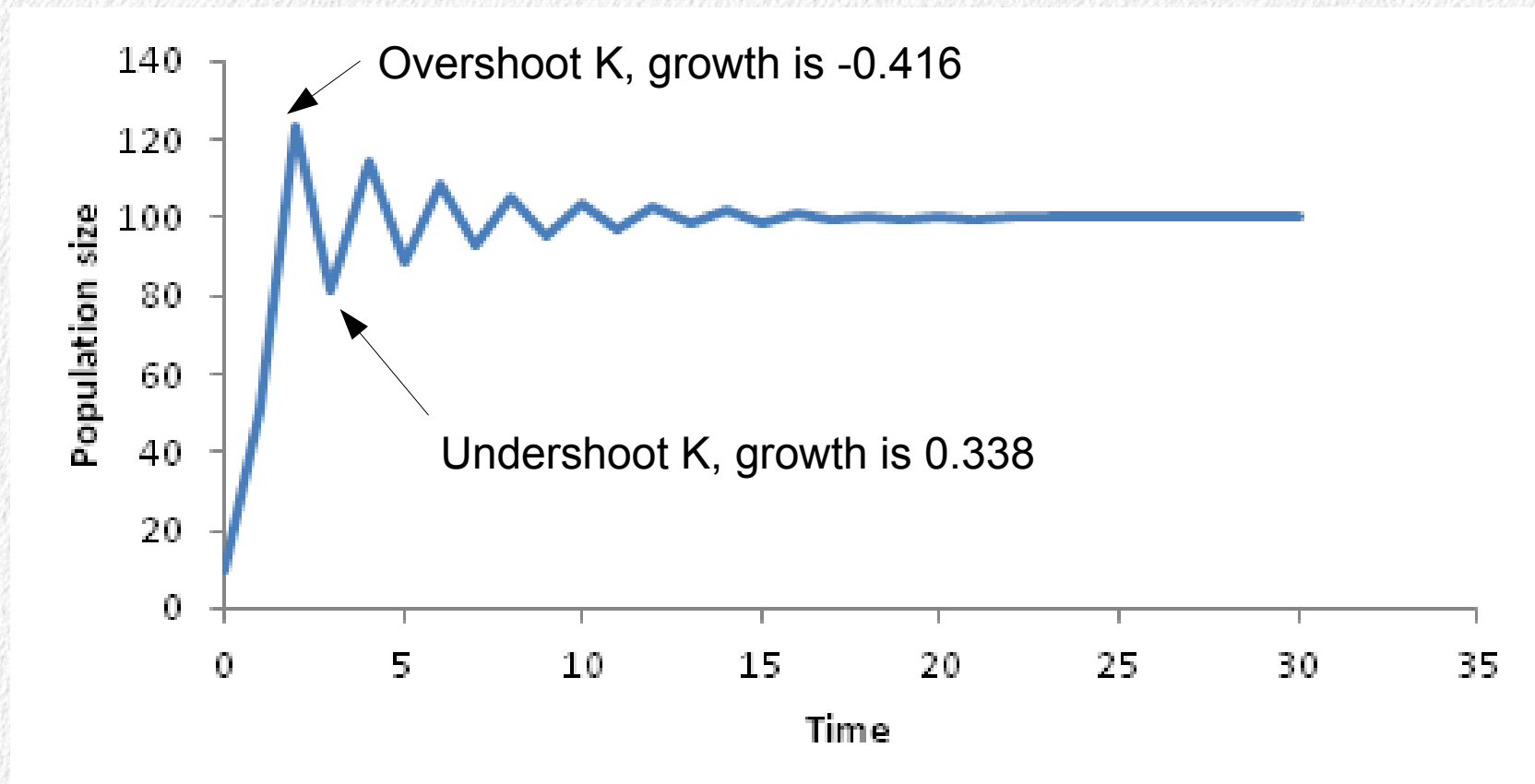
$$N_{t+1} = N_t \times e^{r(1 - N_t/K)}$$

$$r = \text{birth rate} - \text{death rate} = 0.4$$
$$\lambda = 1.5$$



Smooth curve, gradually approaches carrying capacity, reaches a stable equilibrium

$$r = 1.8, \lambda = 6.0$$

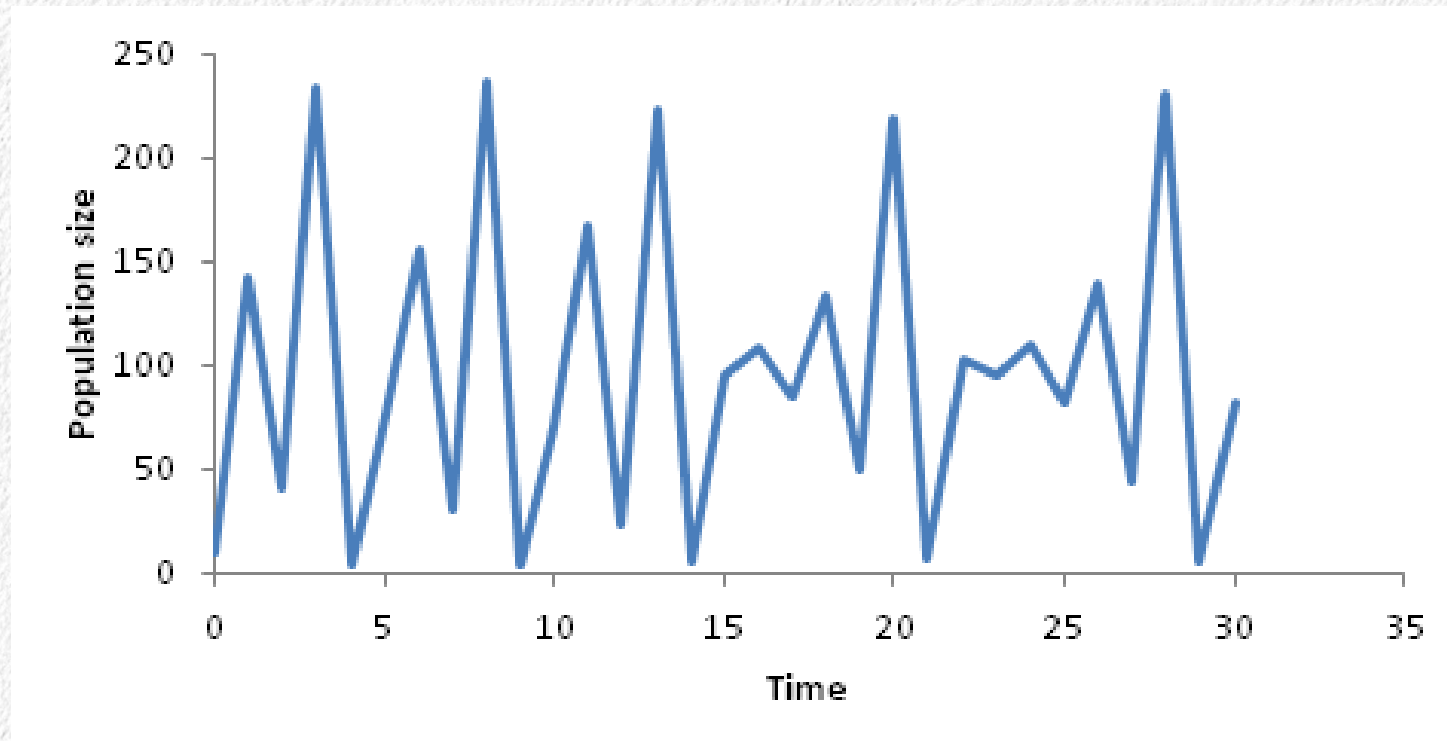


Overshoots carrying capacity, damped oscillations before reaching a stable equilibrium

This is a deterministic model, but not an analytical one

- You can't predict the population size at $t+2$ from the conditions at time t , you have to first calculate $t+1$
- “Damped” oscillations
 - First an overshoot leads to negative growth
 - Then, an undershoot leads to positive growth
 - The over/undershoots decrease over time until carrying capacity is reached

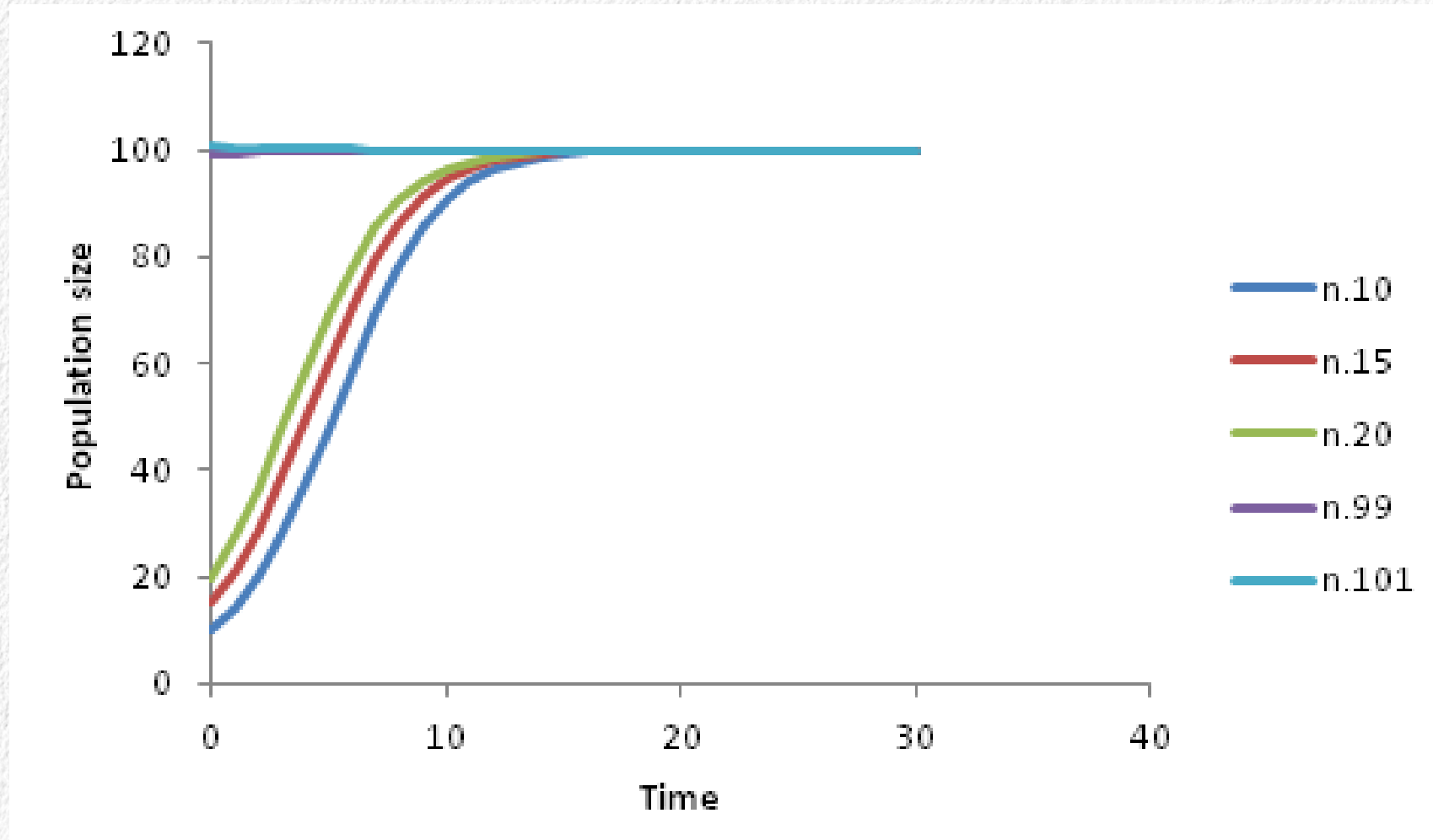
$$r = 2.95, \lambda = 19.1$$



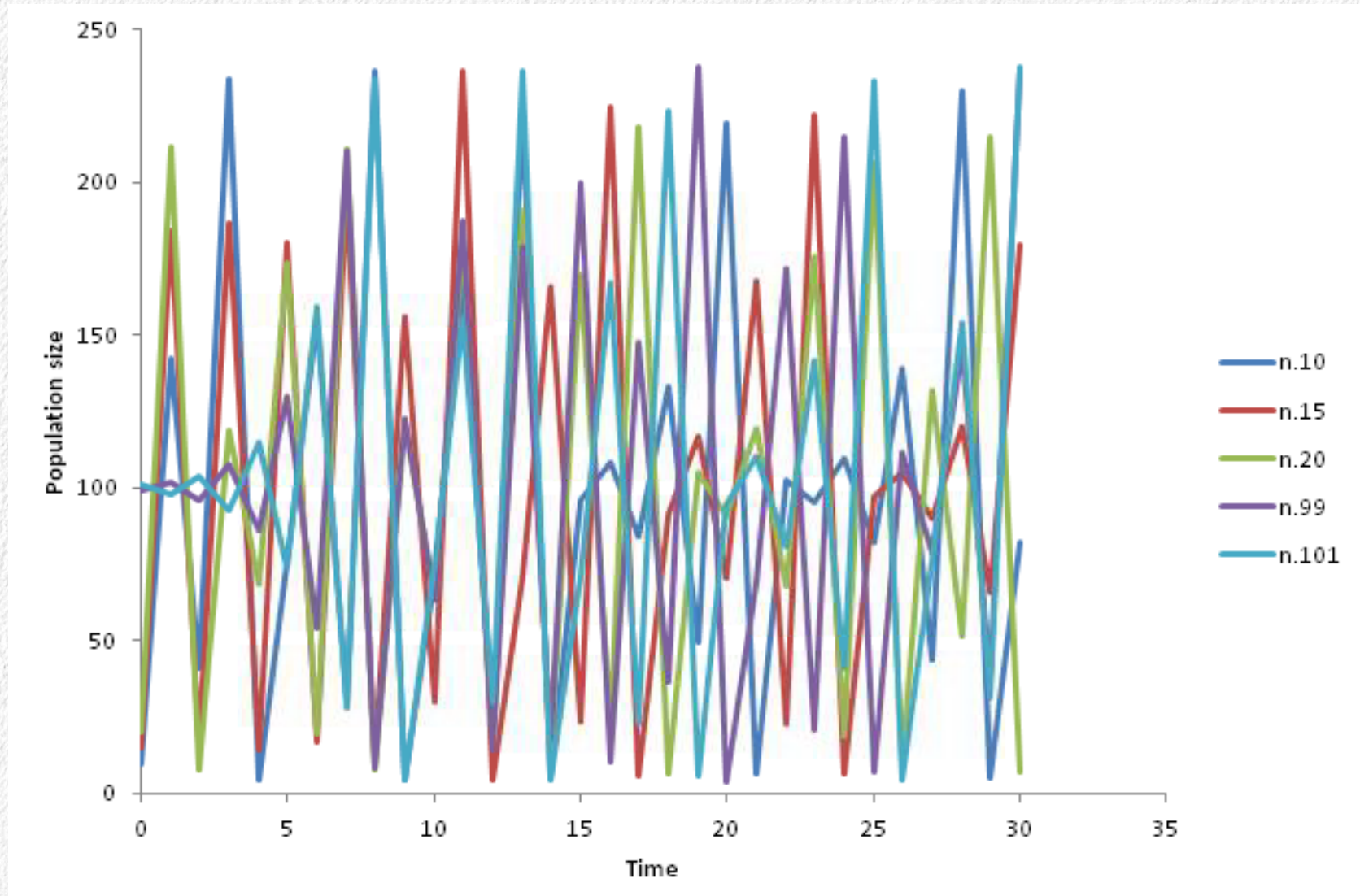
Deterministic chaos = unpredictable from initial conditions, no tendency towards reaching a stable equilibrium at K

But, the exact same trajectory will happen again if you use the same initial conditions

Different population sizes when $r = 0.40$



Same model when $r = 2.95$



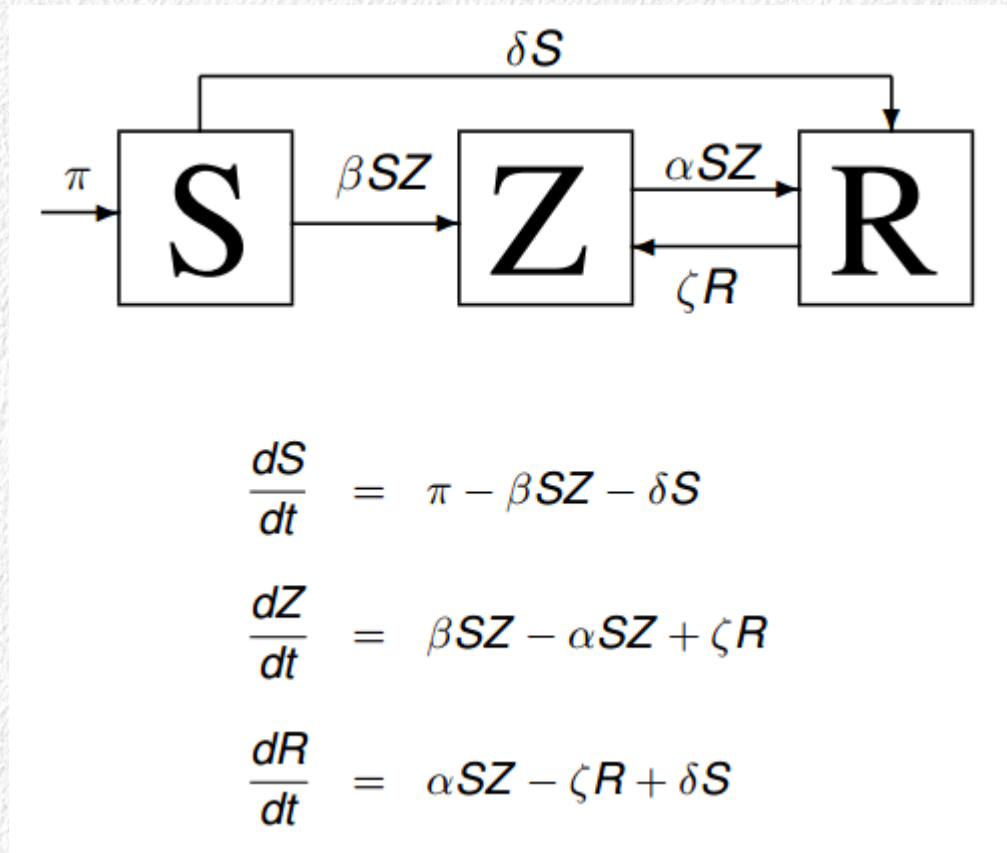
Model complexity, model realism

- Very simple models gloss over a lot of biological detail
 - Can't take into account exceptions
 - Any factors not included in the model aren't considered
- Complex models can more closely approximate reality
 - But, errors propagate
 - Need estimates of large numbers of parameters
- Want a “Goldilocks” model – one that is complex enough to meet your needs, but not too complex

Computer simulations

- Computer simulation models allow you to address complex questions without the complex mathematics that would be required from an analytical model
- Advantages
 - The complexity can be reduced by using algorithms in the place of equations
 - Different levels can be modeled (populations, individuals)
 - Randomness is easy to incorporate
- Disadvantages:
 - Approximate, numerical results rather than exact, analytical solutions
 - Results have to be studied to infer solutions, rather than simply calculating them

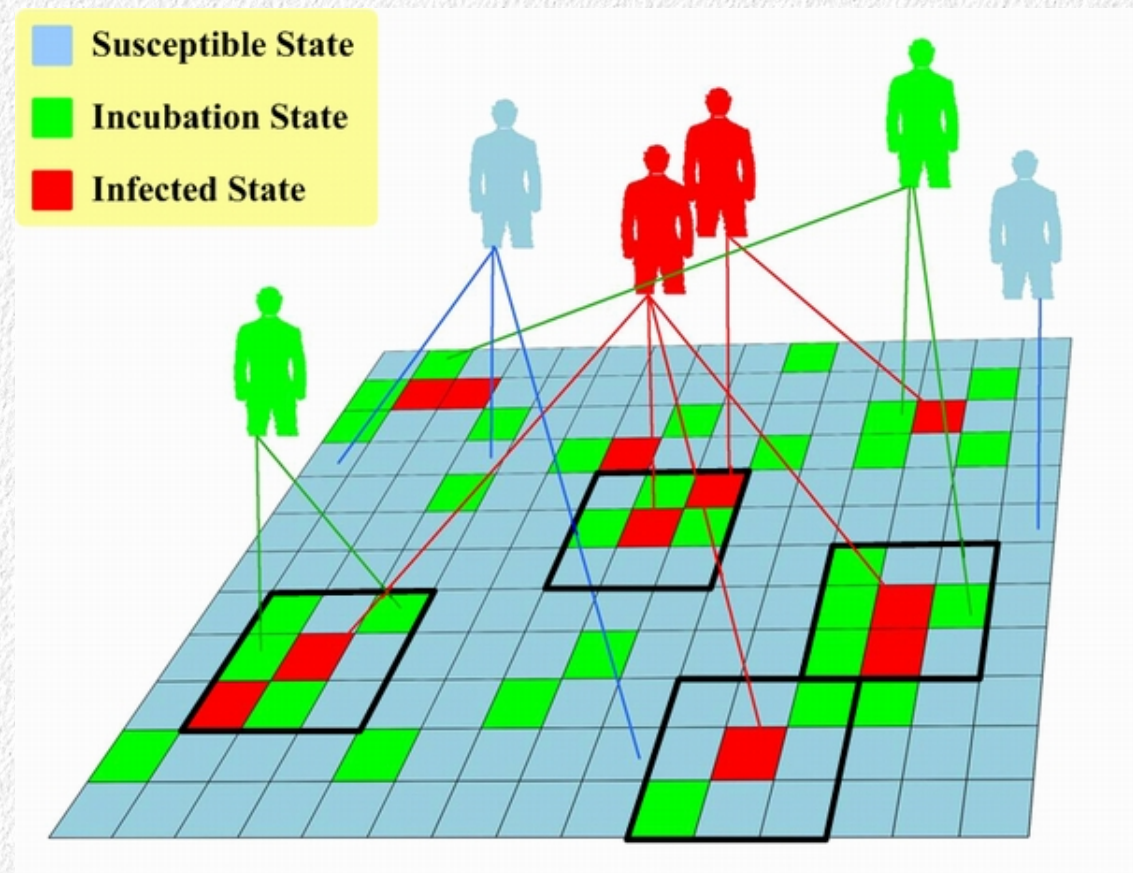
Spread of disease – systems of differential equations



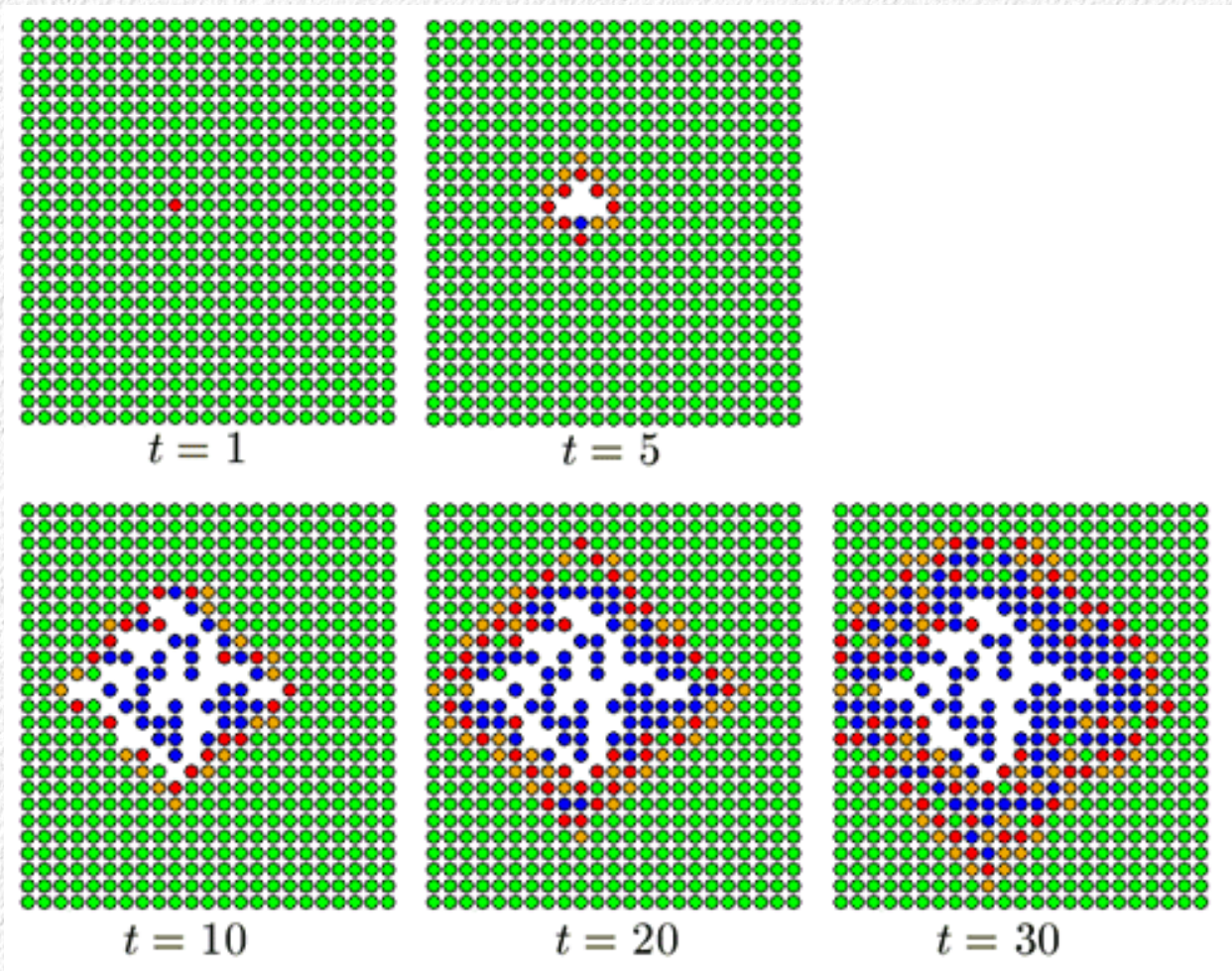
Actually a model of the Zombie apocalypse – hit them early, hit them hard!

Example: modeling disease spread

- If the spatial spread of the disease is to be modeled, then the math is even more complicated
- A simpler approach is to use a “cellular automata” model
- Each cell of a grid is established as susceptible (individuals, population)
- Infection is introduced in one or more cells
- Rules for spread: from infected to susceptible in adjacent cell be susceptible, must be in an adjacent cell



Running the model



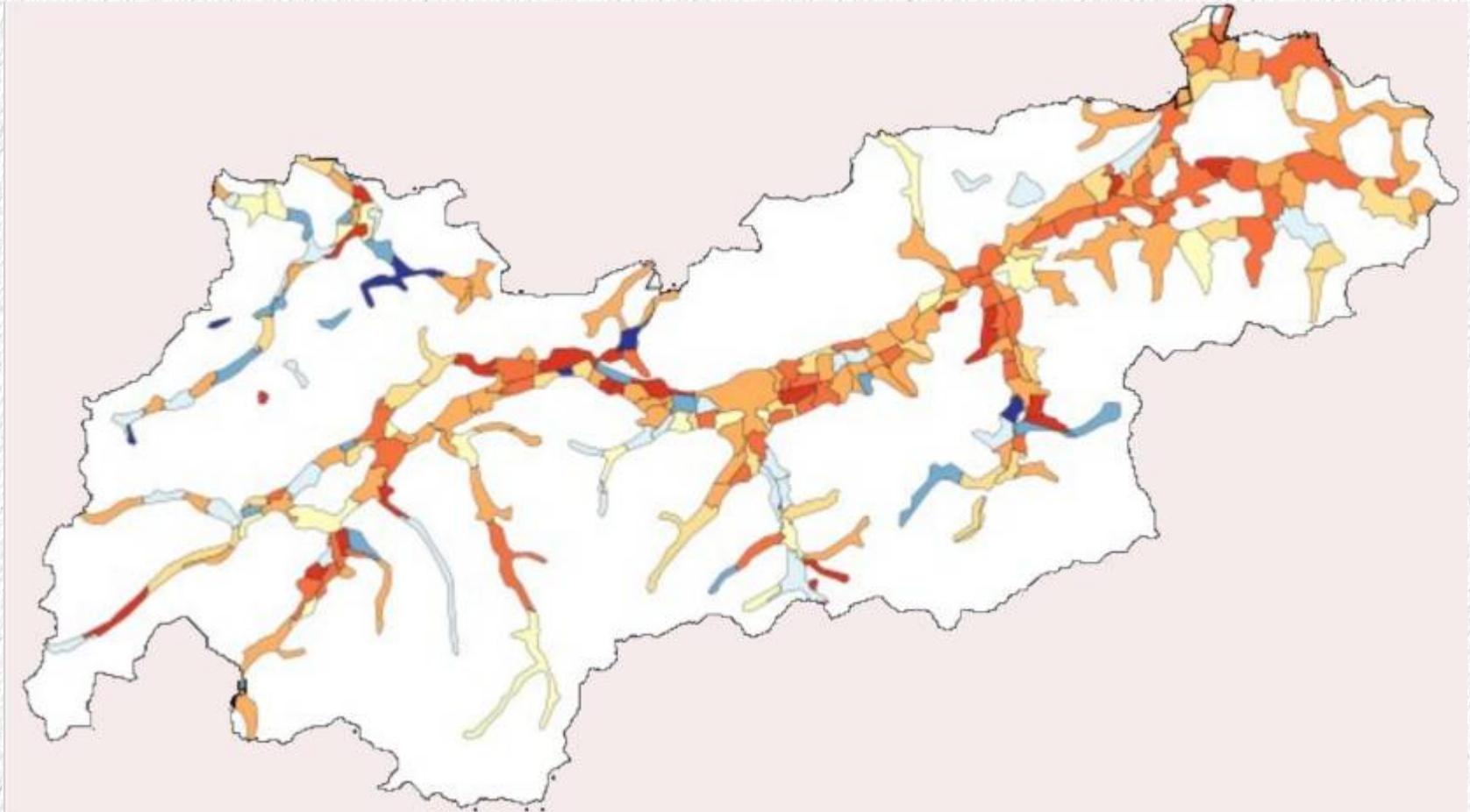
We can ask questions like:

How quickly will the disease spread?

What if we quarantine people with the disease?

If they are infectious during incubation phase, is it too late by the time they start showing symptoms?

More realistic layout



More realistic map – spread will be along travel corridors

More realistic models are generally more complicated

Our stochastic simulation model of drift

- Set up a population of individuals with genotypes set to match an initial, known frequency
 - The model is “individual-based”, or “agent-based”
- Simulate generations of random breeding over time
 - Not all individuals will successfully breed
 - If the breeders don't have an allele, it is lost from the population
- After 500 generations, assess how the gene pool has changed for the population
- Repeat 100 times, evaluate the patterns that emerge across multiple runs of the model