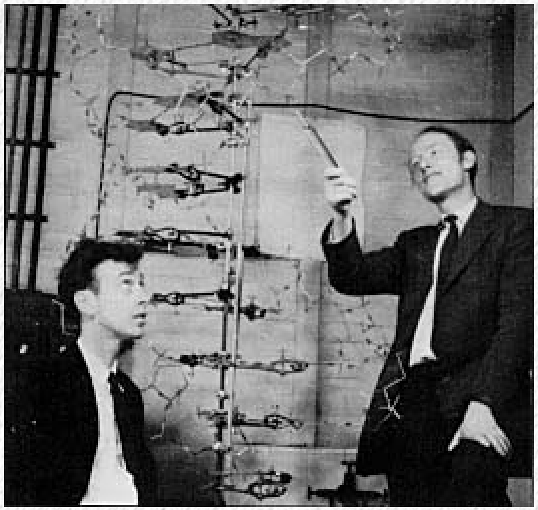
Simulation modeling

What are models good for?
When are simulations preferable to analytical models?
When are stochastic simulations better than deterministic ones?

What is a model?

- An abstract representation of a system or process
  - Physical models
  - Verbal models
  - Graphical models
  - Mathematical models

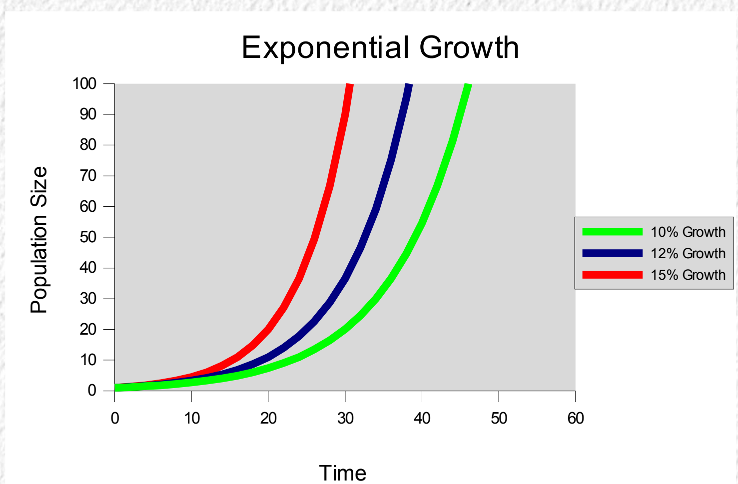
Quick – which is a model?



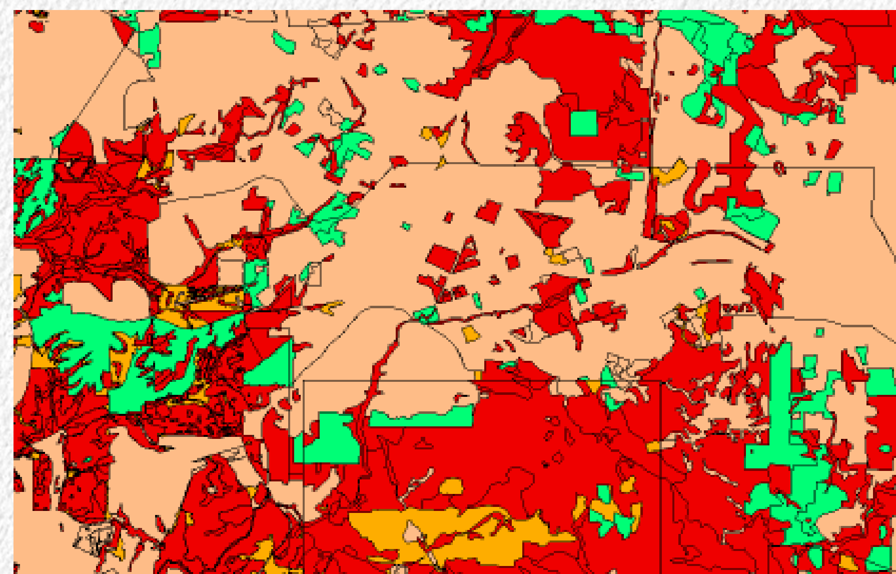
Physical representation

$$\frac{a}{\phi^2 e^{a/\phi}} = \frac{-\ln\left(e^{-a/\Theta} - p\right)\Theta}{\phi^2 e^{-\ln\left(e^{-a/\Theta} - p\right)\Theta/\phi}}$$

Equation



Plot



Land cover map

Why model?

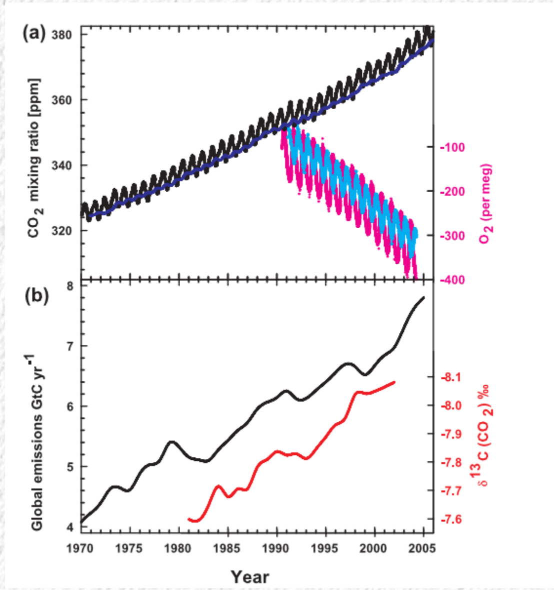
- Models are ways of converting assumptions into predictions so can test them against data
- Models allow a form of “experimentation” not possible otherwise
- Models allow us to “constrain complexity” in our thinking

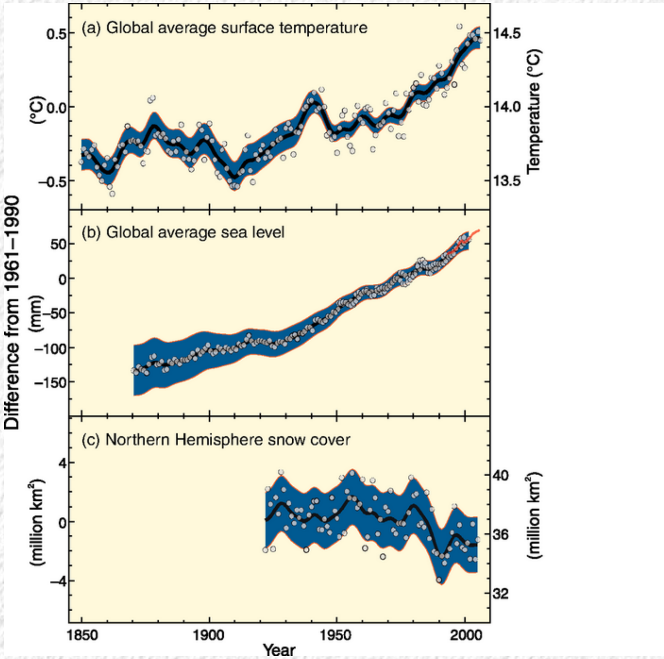
Constraining complexity

- The natural world is unimaginably complex
- As biologists, our job is to make sense of the living component of it
- Best we can do is to contemplate parts of it at any given time, so we need to simplify to understand
- The question is, how much can we simplify the world in our thinking before our conception of it becomes fundamentally inaccurate?
- In other words, we want our understanding of the world to be complex enough, but no more so

Example: global climate change

- We know that:
  - CO_2 is increasing in the atmosphere
  - Humans emit CO_2 when they burn fossil fuels
  - CO_2 is a greenhouse gas
  - Global average temperature has been increasing
- We want to know whether:
  - Anthropogenic CO_2 is the cause of the increase in temperature





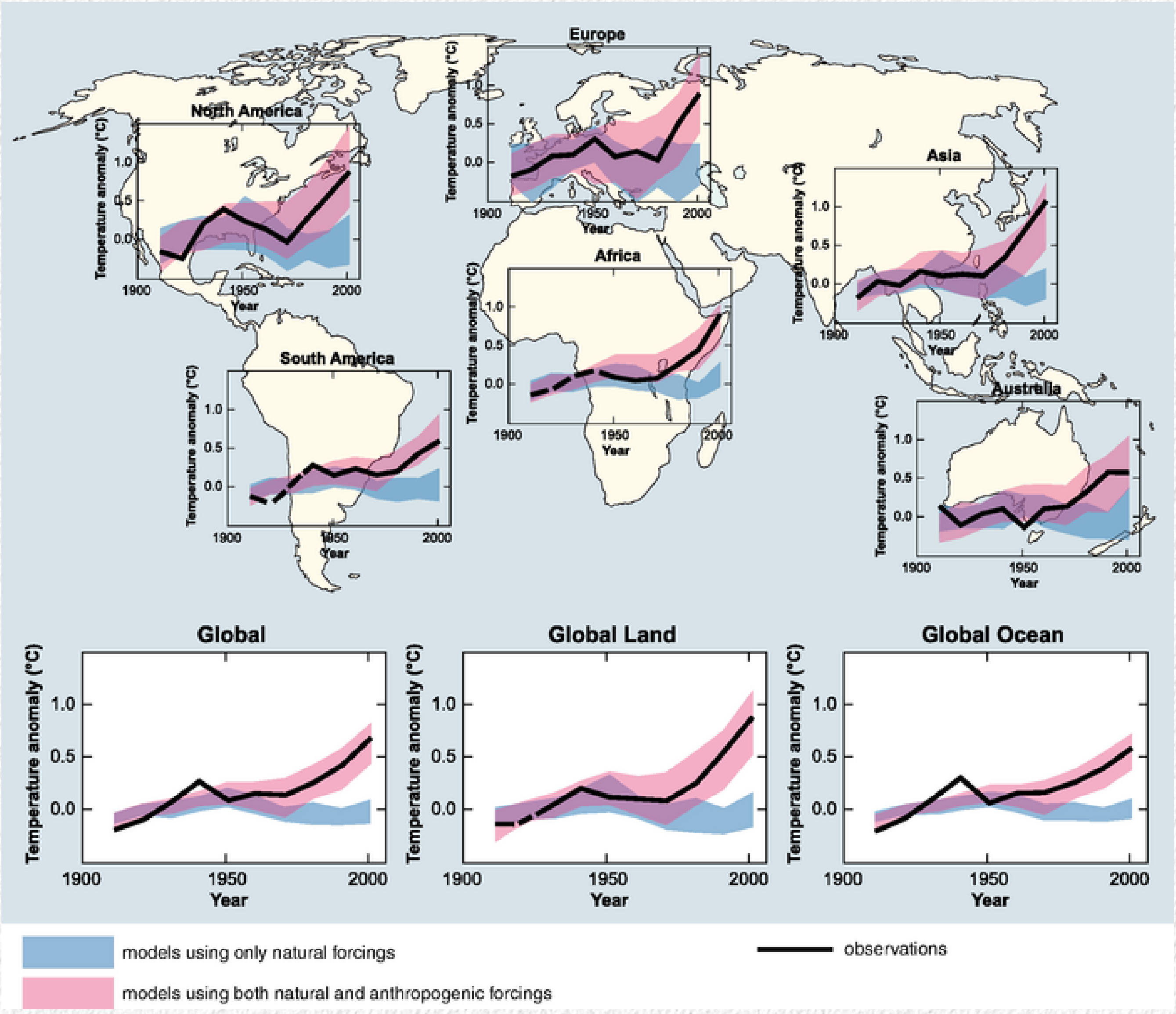
Just do the experiment

- All we have to do is to re-run the time since the industrial revolution, but without the CO_2 emissions
- If the planet still heats up due to all the other factors known to affect temperature, then CO_2 is not to blame
- If the planet doesn't heat up without the CO_2 emissions, then CO_2 is responsible
- Problem?

If you can't do the experiment in the real world, you can in a model

- Build a model of variation in climate that incorporates all the factors we know to be important in temperature regulation
- Run the model without an anthropogenic enhancement of CO_2
- Run the model with an anthropogenic enhancement of CO_2
- Compare the models to the temperature records we already have in hand
- What does this test?

Only predict the temperature increase correctly if we include anthropogenic effects



What do we learn from this model?

- We can reasonably accurately reproduce our existing temperature records
  - Thus, our understanding of global climate is not bad
- We can't reproduce the existing temperature records unless we include anthropogenic CO_2
  - Thus, the existing temperature records support an effect of anthropogenic CO_2 on global temperature

Some common dichotomies in math modeling

- There are many different kinds of models
- Some common characteristics that distinguish them:
  - Deterministic vs. stochastic
  - Analytical vs. simulation

Deterministic vs. stochastic

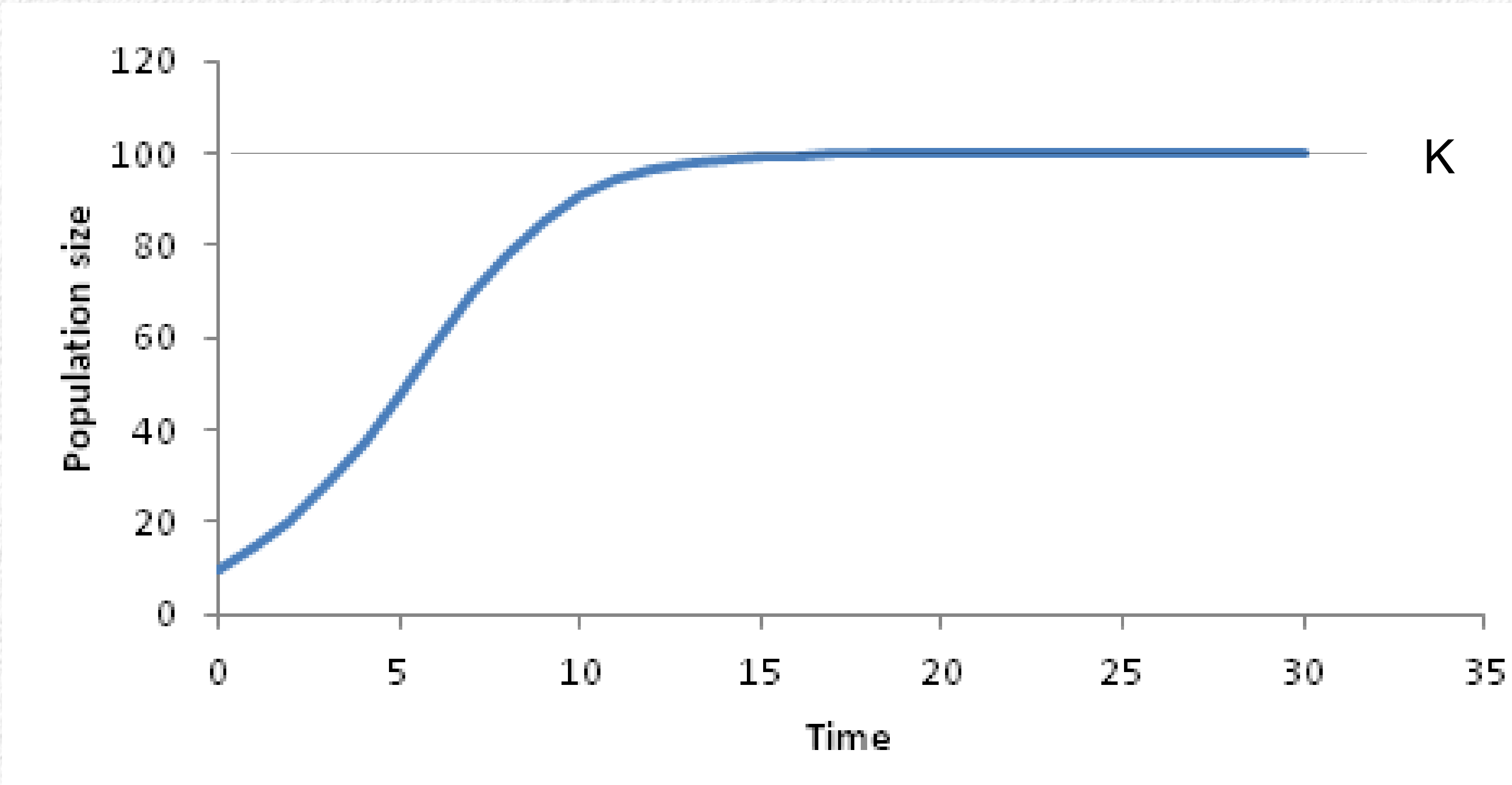
- Deterministic – Given the same inputs, the outputs are always the same
- Stochastic – The same inputs give different outputs
  - The outputs can be described probabilistically, but can't be predicted exactly

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:

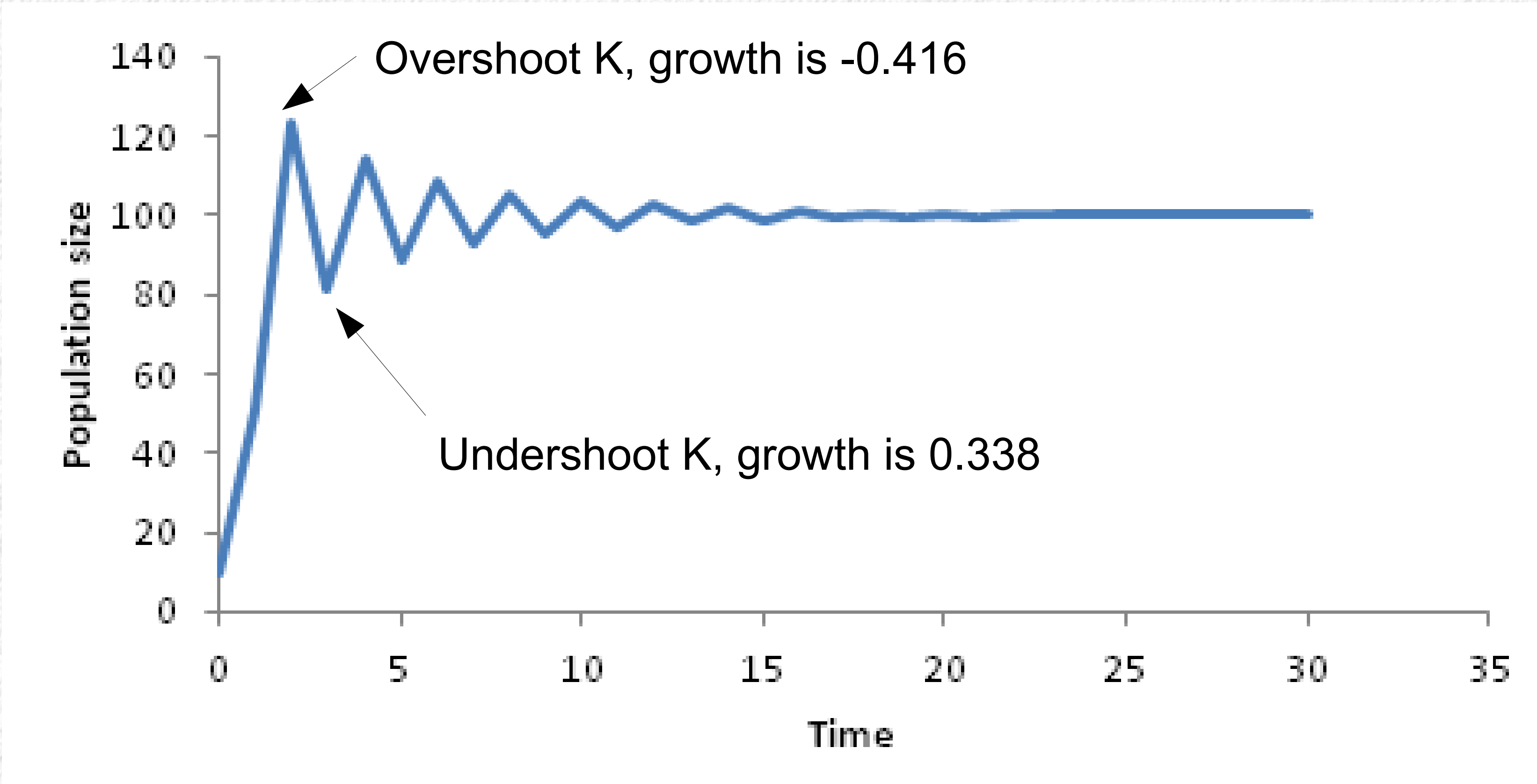
$$N_{t+1} = N_t \times e^{r\left(1 - n_t / K\right)}$$

r = birth rate – death rate = 0.4
λ = 1.5



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

$r = 1.8, \lambda = 6.0$

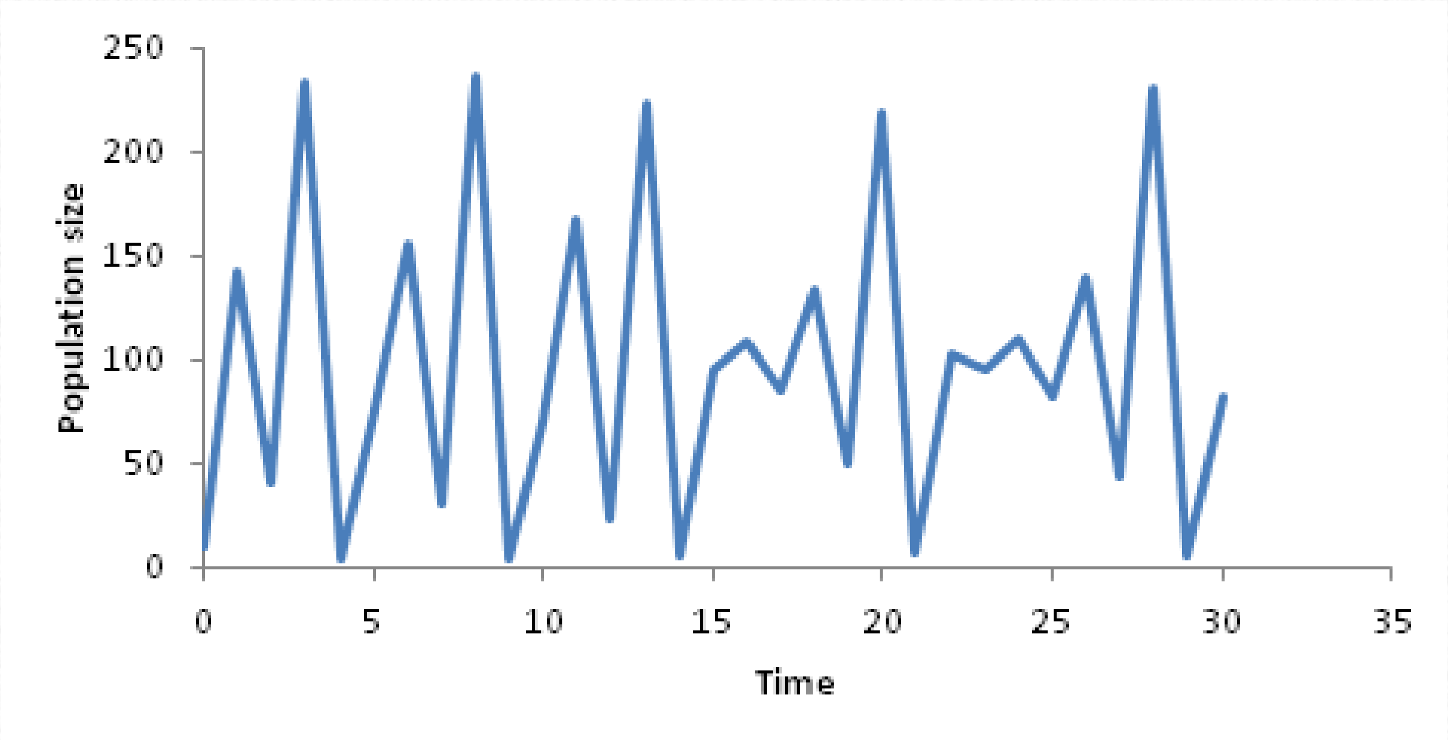


Overshoots carrying capacity, damped oscillations before reaching a stable equilibrium

This is a deterministic model, but...

- 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

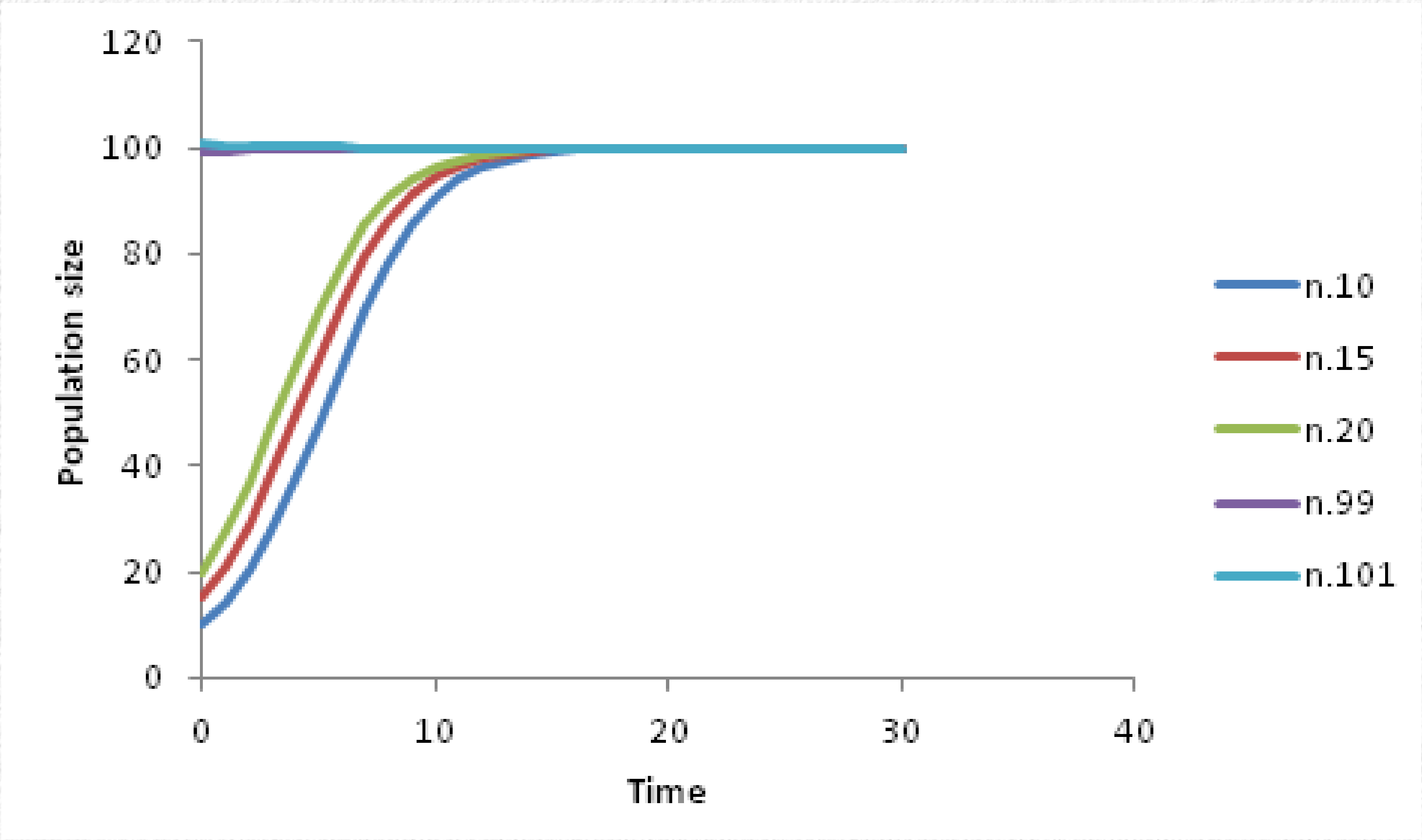
$r = 2.95, \lambda = 19.1$



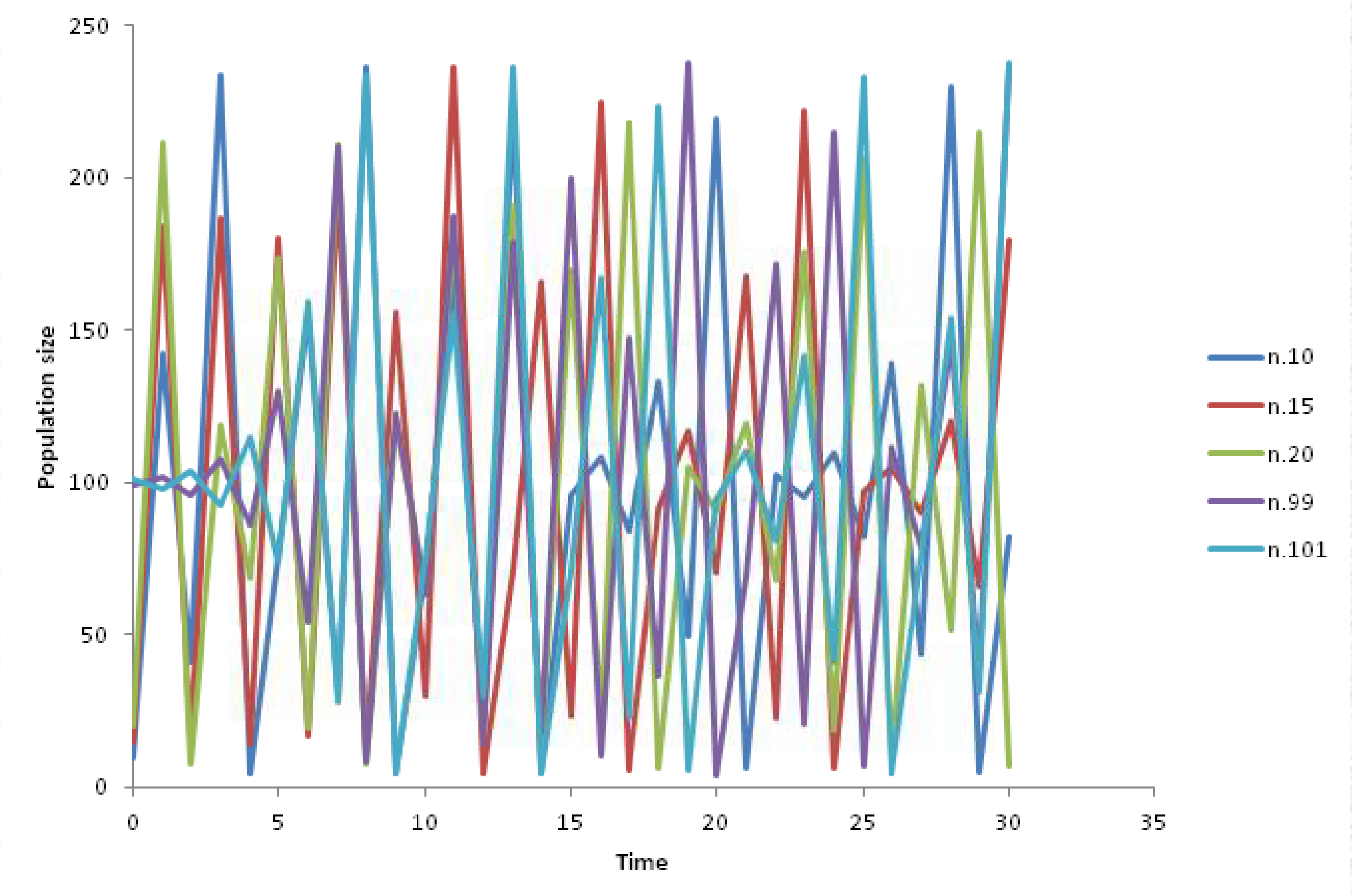
Deterministic chaos = unpredictable from initial conditions, no tendency towards reaching a stable equilibrium at K

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



Correlation between populations with different starting sizes

r = 0.40 (stable equilibrium)

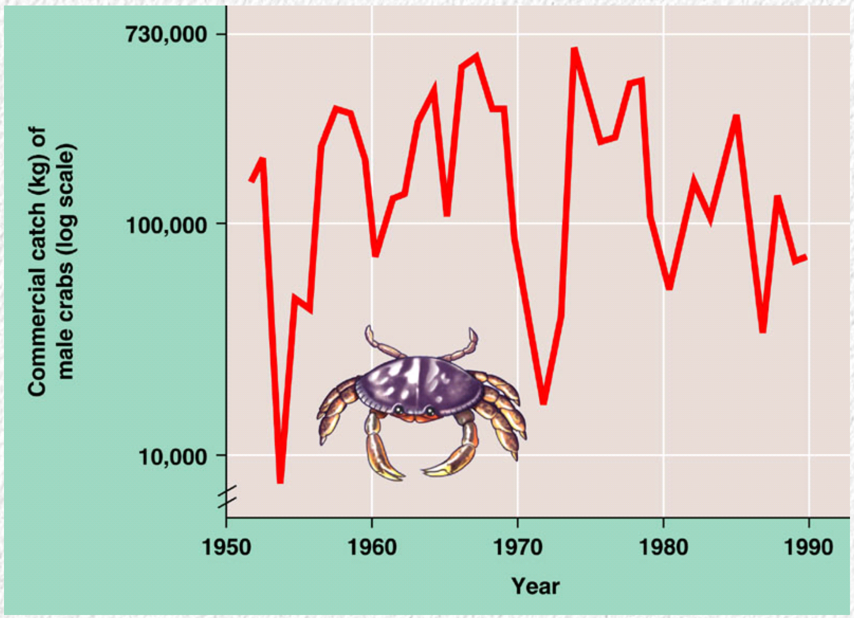
	n.10	*n.15*	*n.20*	*n.99*	*n.101*
n.10	1				
n.15	0.994714	1			
n.20	0.984298	0.997156	1		
n.99	0.817572	0.863003	0.894016	1	
n.101	-0.81423	-0.85988	-0.89113	-0.99997	1

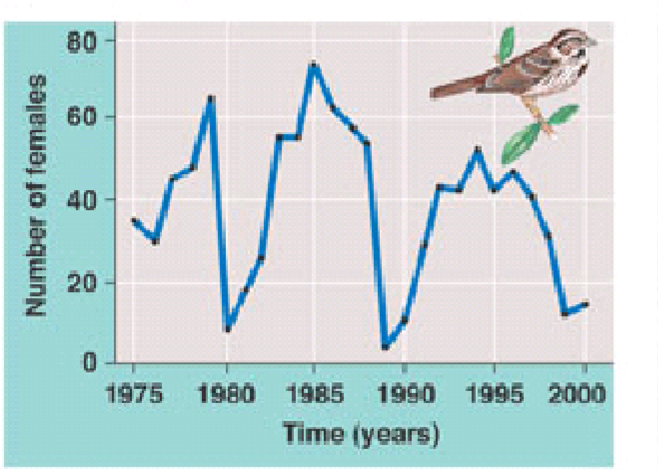
r = 2.95 (chaotic dynamics)

	n.10	*n.15*	*n.20*	*n.99*	*n.101*
n.10	1				
n.15	0.053815	1			
n.20	-0.03747	0.308216	1		
n.99	0.019601	0.072153	0.192389	1	
n.101	0.498576	0.080715	-0.10598	-0.16889	1

Constraining complexity – do we need to include randomness?

- The graphs look chaotic, do we need to explain the variations by something else (randomness)?
- Deterministic chaos looks superficially similar – does that tell us we need to look no further?



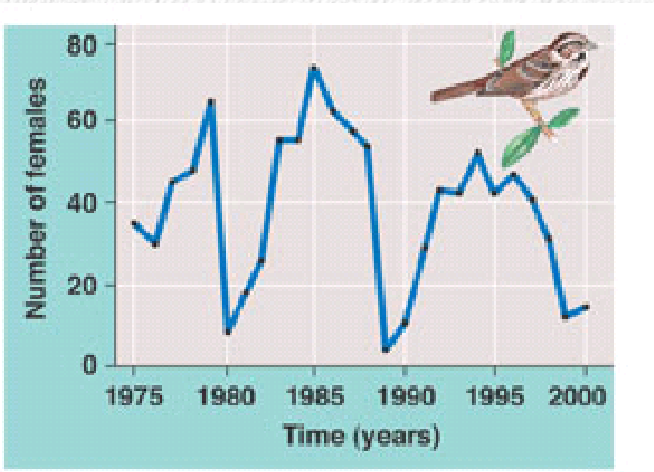


(c) A song sparrow population in its natural habitat

Why is randomness more complex than chaos?

- We can produce chaos with a deterministic model
- Randomness in biological systems comes from the unpredictable action of multiple factors acting at once

Birds can't increase by 20x in a year



(c) A song sparrow population in its natural habitat

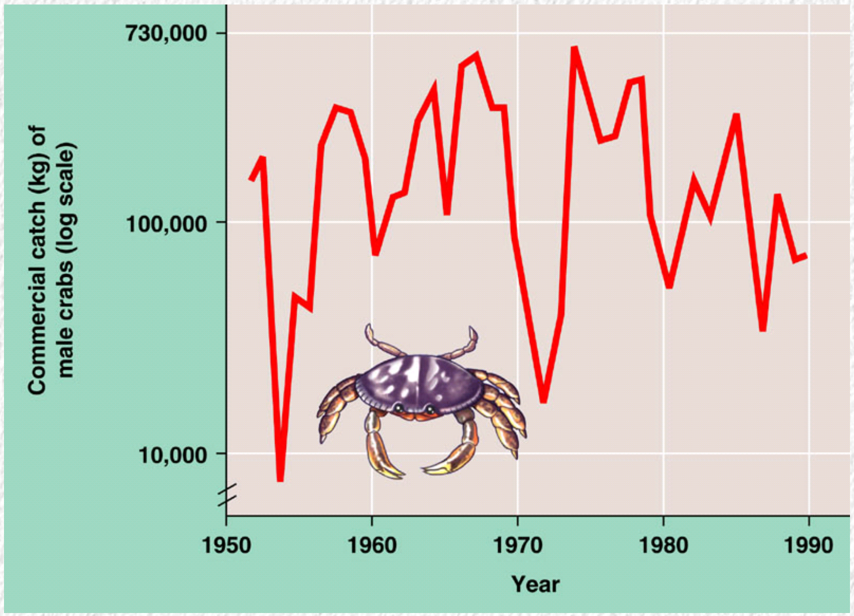
Song sparrows would be lucky to fledge half that many each year

Many would not survive to the next year

Big dropoffs are possible, since mortality can be up to 100%

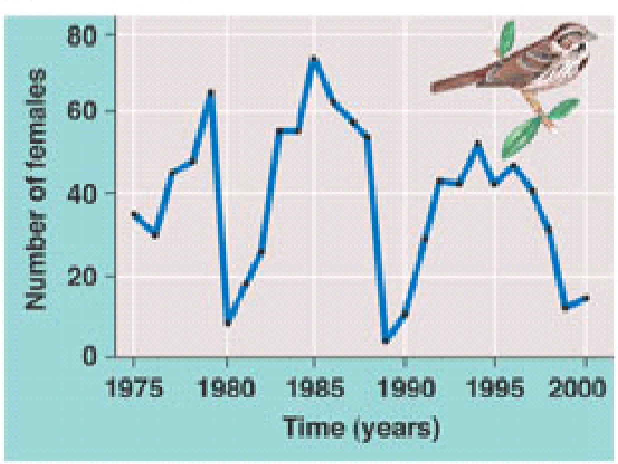
But, the biggest increases in a year are 4x – not big enough to produce deterministic chaos

We know there are environmental influences on r, K, and N_t



Crabs have the capacity to increase their population 20x in a year – massive reproductive rates

But, they are sensitive to the environment, which varies randomly from year to year



(c) A song sparrow population in its natural habitat

Birds are also sensitive to changes in their environments from year to year

Many biological systems are subject to random variation

- Some processes may be truly random
  - Random assortment of alleles during meiosis
  - Random sampling of alleles in a population over time
- May be so complex that it isn't possible to know all the factors that affect the system – model the system as stochastic

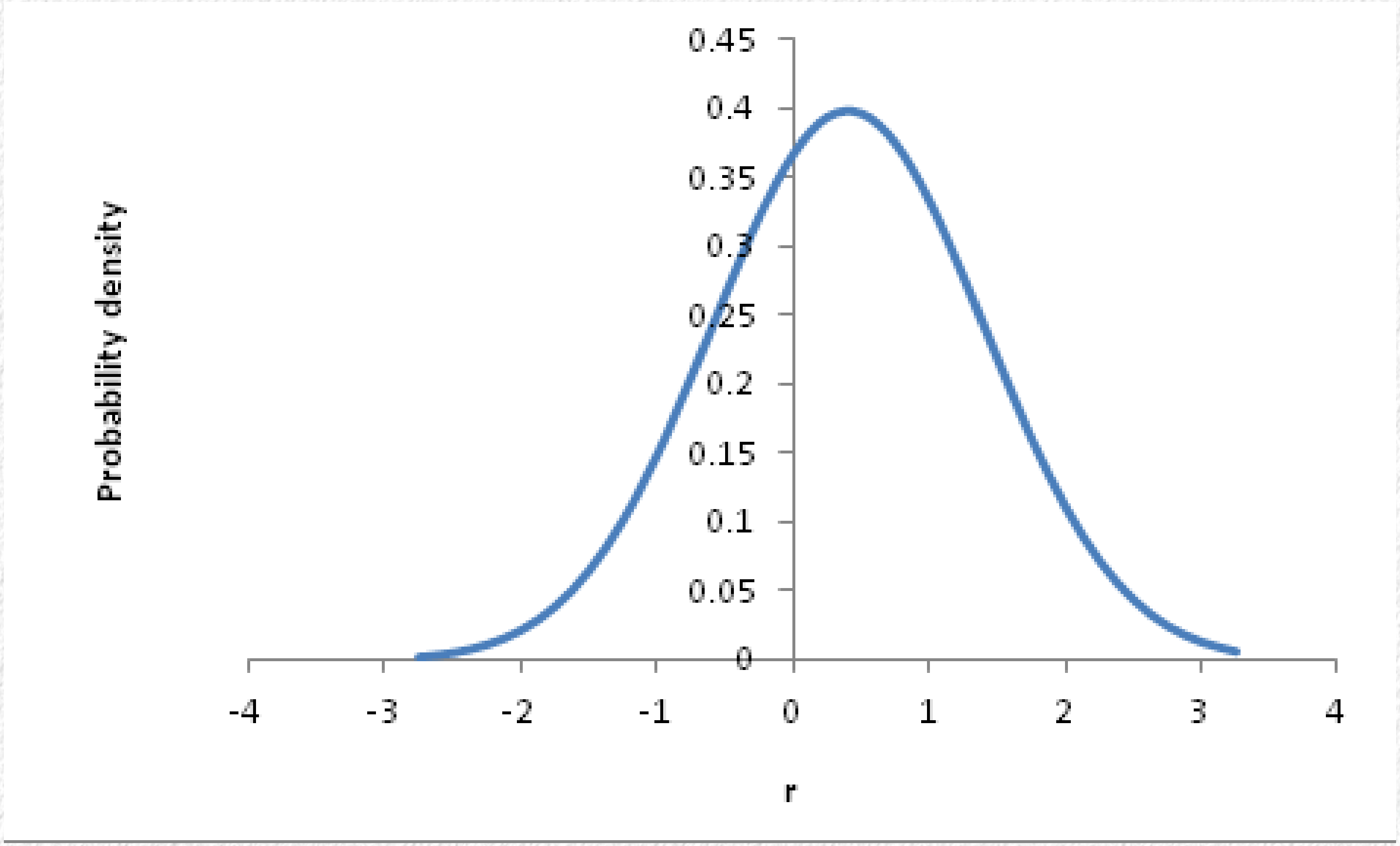
Making a model complex enough

- If we're trying to understand why song sparrow populations vary like they do, we will need to make the models complex enough to include all the most important factors
- We know from field data that the parameters vary over time
- Building a model in which the parameters ***don't*** vary over time will be too simple to reproduce the dynamics we see

Stochastic version of the model

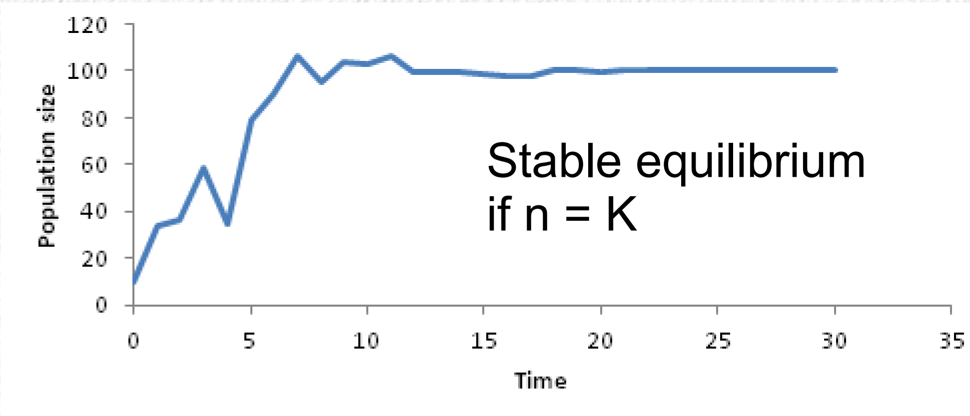
- A stochastic model will not be the same any two times through, even if you start with the same conditions
- We'll specify an average growth rate, but each time step it will be a random draw from a distribution of possible values
- Even when the growth rate is low enough that the deterministic model smoothly approaches K, the stochastic version will have random variation

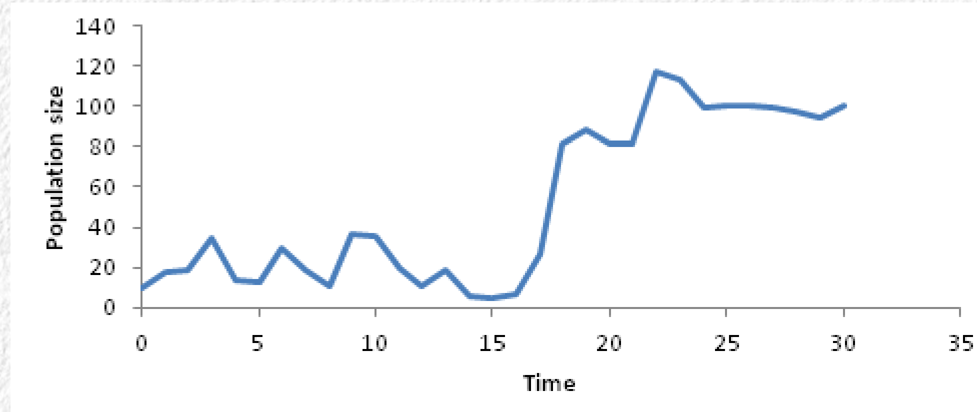
r as a distribution

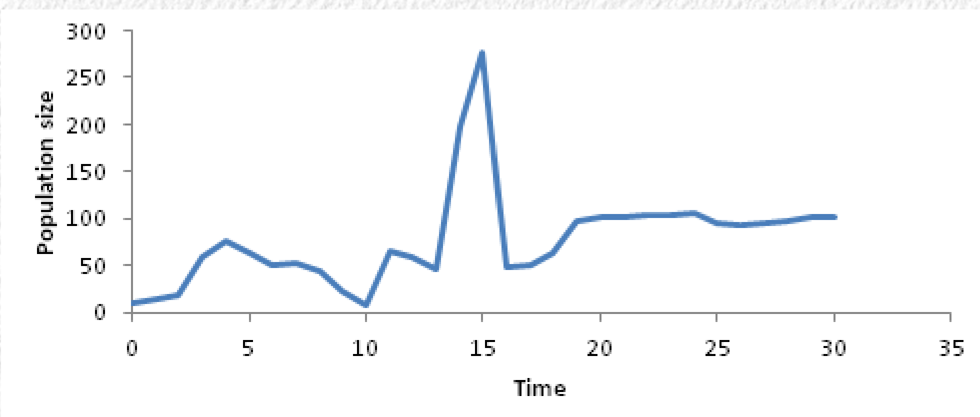


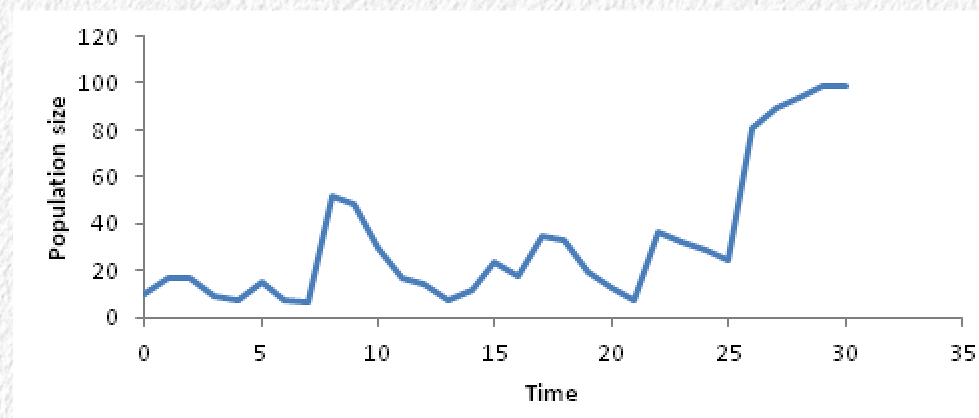
Mean r is 0.4
Standard deviation is 1
95% of r's will fall between -1.6 and 2.4
Actual r for a year will be a draw from this distribution

A few runs through – all starting at n = 10, all with K = 100









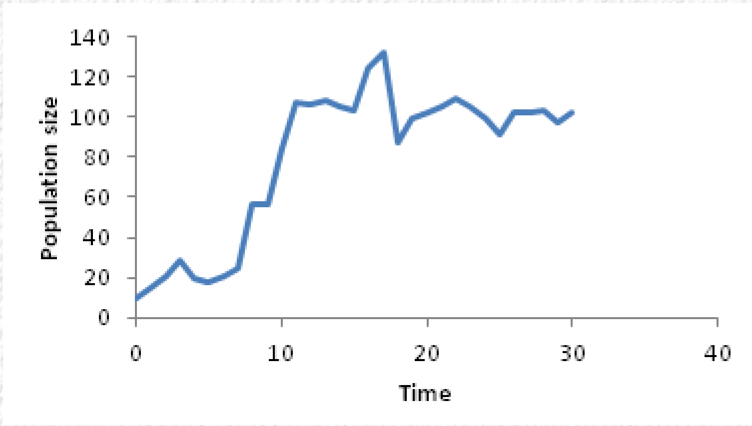
$$N_{t+1} = N_t \times e^{r(1-n_t/K)}$$

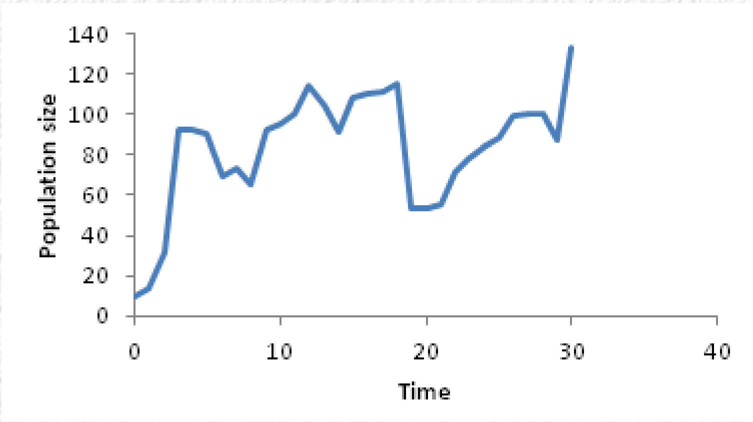
Other sources of randomness

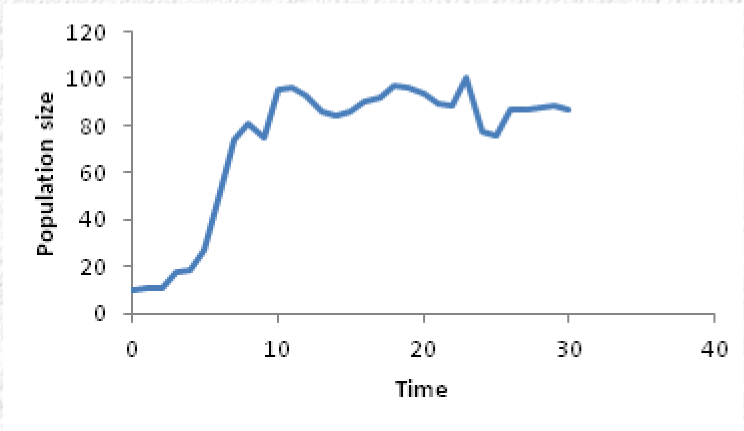
- This model only treated r as being subject to randomness
- What if K also varies at random?
- We expect this – K is a function of e.g. food levels, which will vary randomly over time
- Will we ever reach a stable equilibrium if K varies over time?

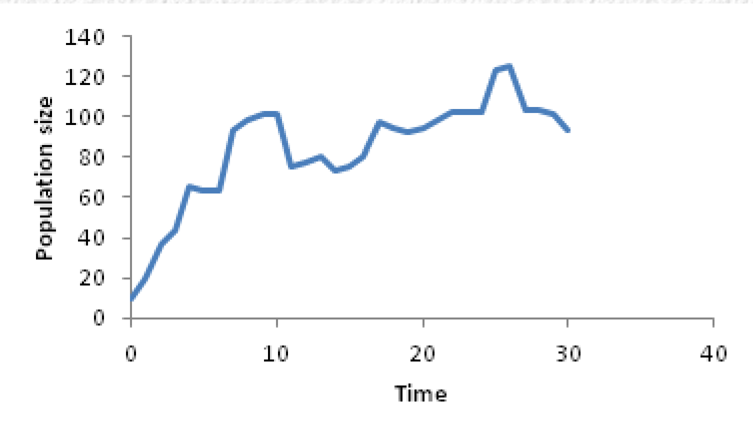
$$N_{t+1} = N_t \times e^{r(1 - n_t/K)}$$

Both r (s = 0.5) and K (s = 10) are stochastic







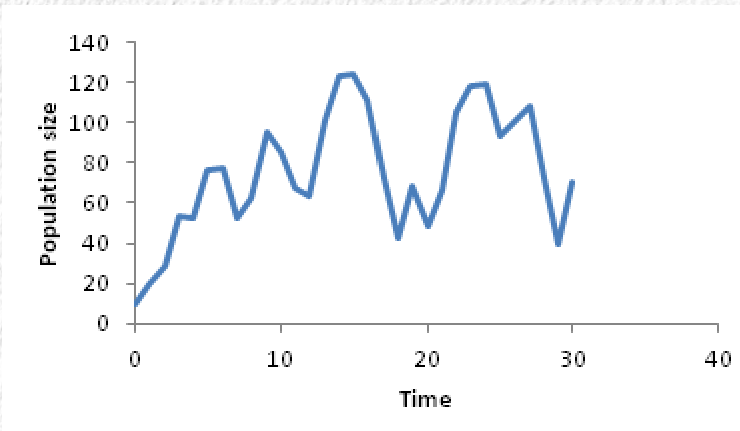


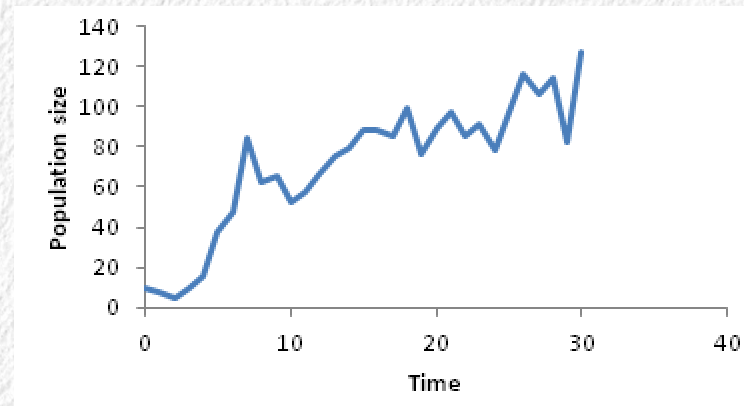
$$N_{t+1} = N_t \times e^{r(1 - n_t / K)}$$

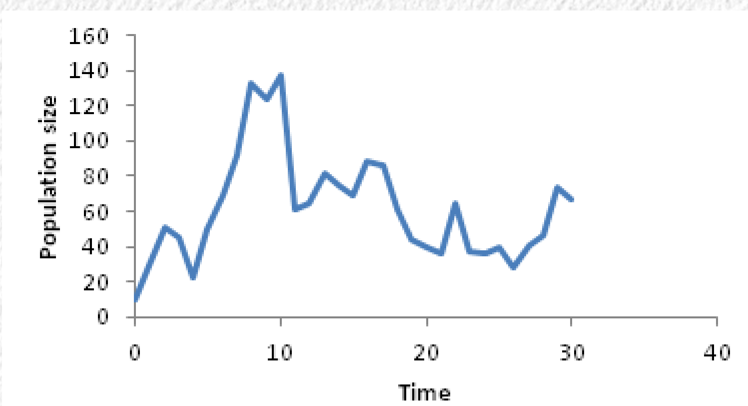
Random mortality

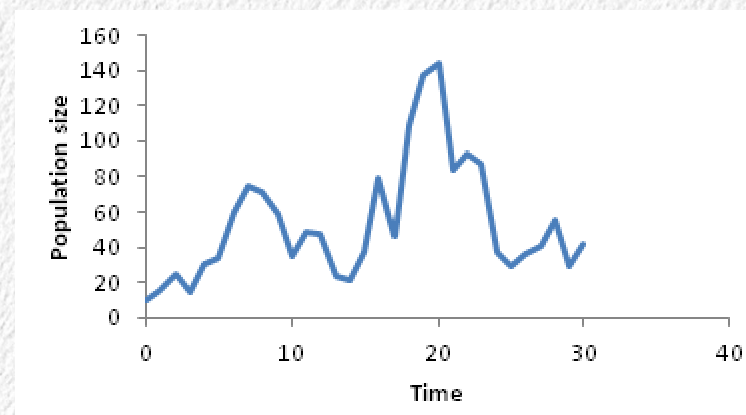
- Some sources of mortality occur at random, and take out some proportion of the population
  - "Density independent" = the amount of effect doesn't depend on the number of organisms in the population
  - Weather events are often like this
- We can simulate this by randomly selecting some proportion of the population to be taken out each year
  - i.e. if we choose 5% density independent mortality, then we'll lose 5% regardless of how many are there
  - "Density dependent" only if we lose a larger proportion when the population is larger
- Set a mean level, a standard deviation, then randomly select from the distribution each year
- Will we reach a stable equilibrium? Will carrying capacity be evident?

r and K are stochastic, random mortality (mean = 5%, s = 2%)







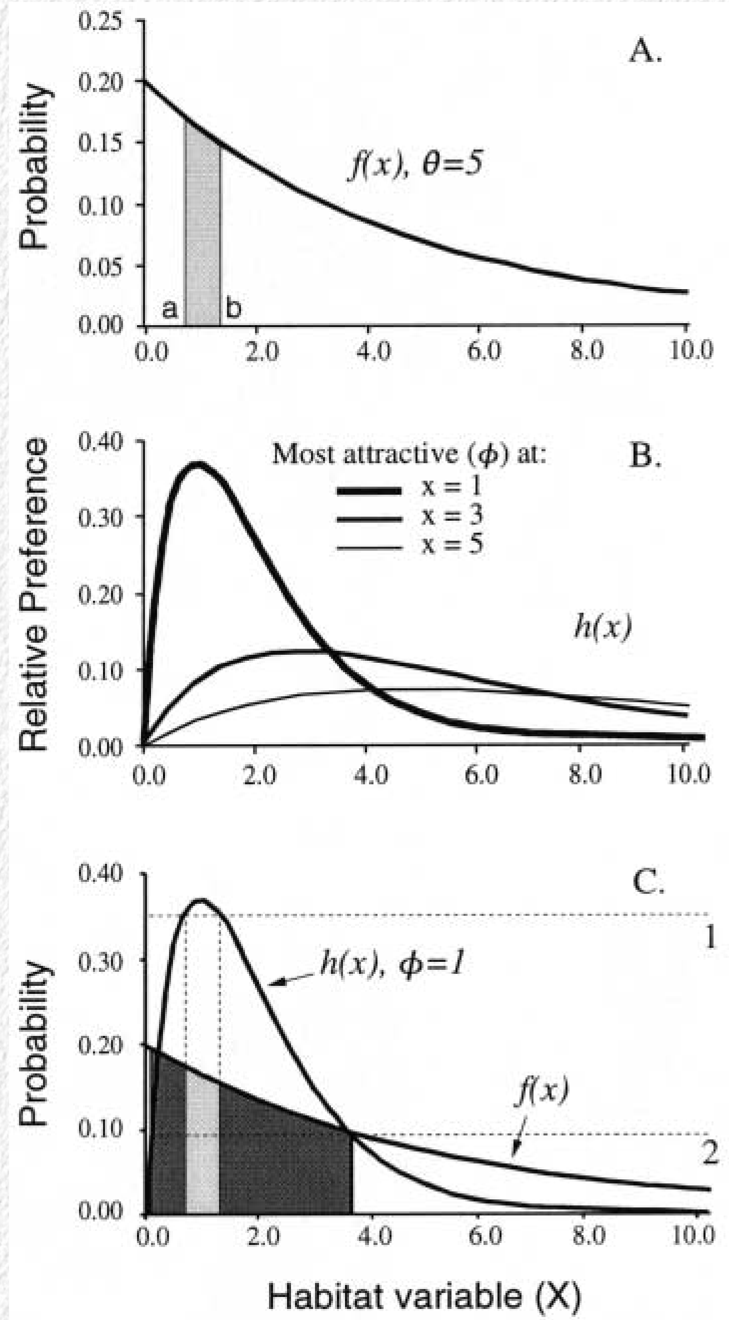


Analytical vs. simulation models

- Analytical – Solutions are mathematically “well formed,” and can be expressed as equations
  - e.g. if b > d, r > 0
  (if births > deaths, population growth will be positive)
- Simulation – Model structures are too complicated to be expressed as simple equations
  - Generally, simulations are done with a computer
  - Often stochasticity is included
  - Results of simulations are collected and studied, much like we study the real world
  - We will generally need to run the model many times and see what the average result is, as well as variability in results among runs

Example analytical model: sinks and traps

- From Kristan 2003
- Model of habitat selection by animals
- Assume that habitat is limited in availability
- Assume they when habitat is not completely filled, they will take the most attractive habitat available



Working with the model

- Knowing the equations for each of the curves, it's possible to solve for various quantities of interest
- But, the math gets a little hairy
- For example, the equation for the mean attractiveness of occupied habitat is:

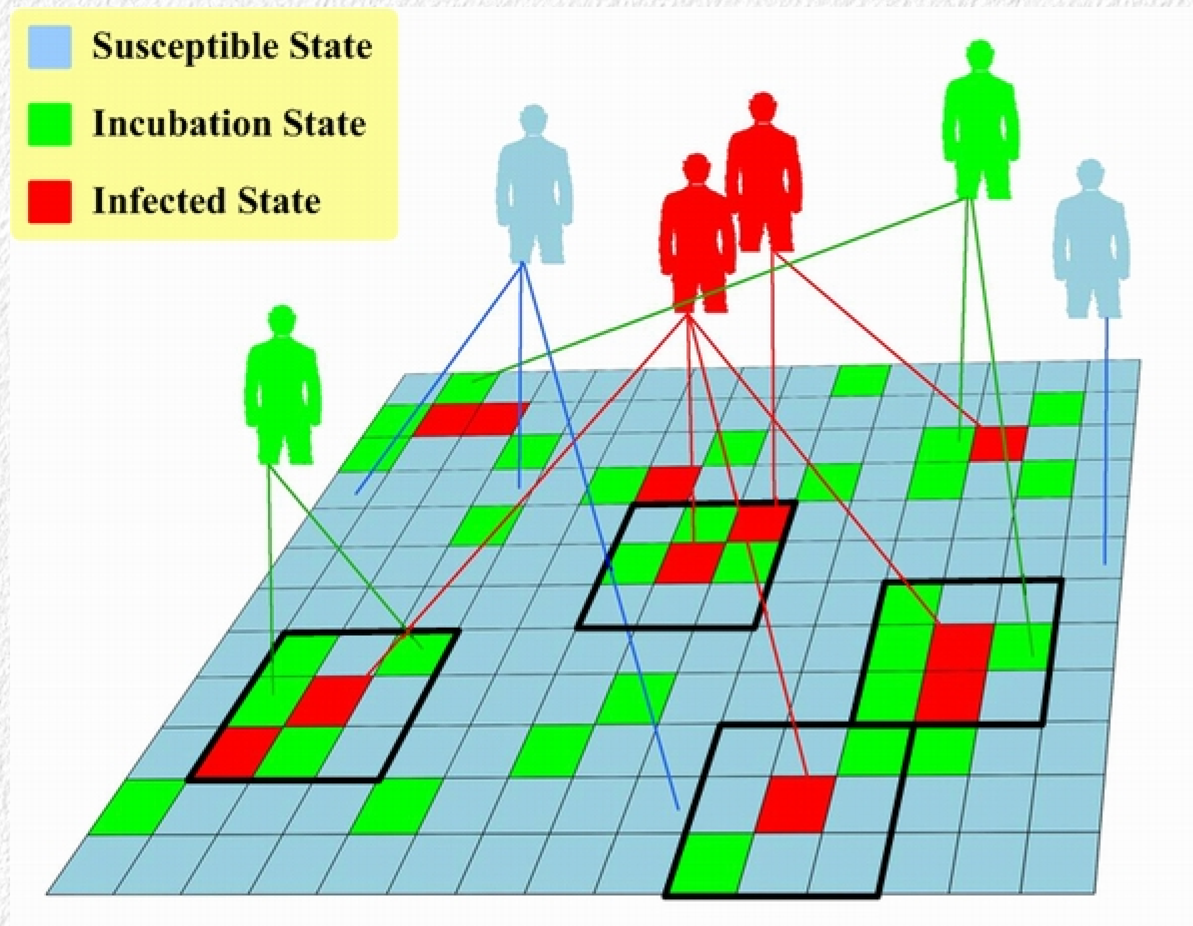
$$\bar{h} = \int_a^b \left(\frac{x}{\phi^2 e^{\frac{x}{\phi}}} \right) \left[\frac{\left(\frac{1}{\theta} e^{\frac{-x}{\theta}} \right)}{e^{\frac{-a}{\theta}} - e^{\frac{-b}{\theta}}} \right] dx$$

Computer simulations

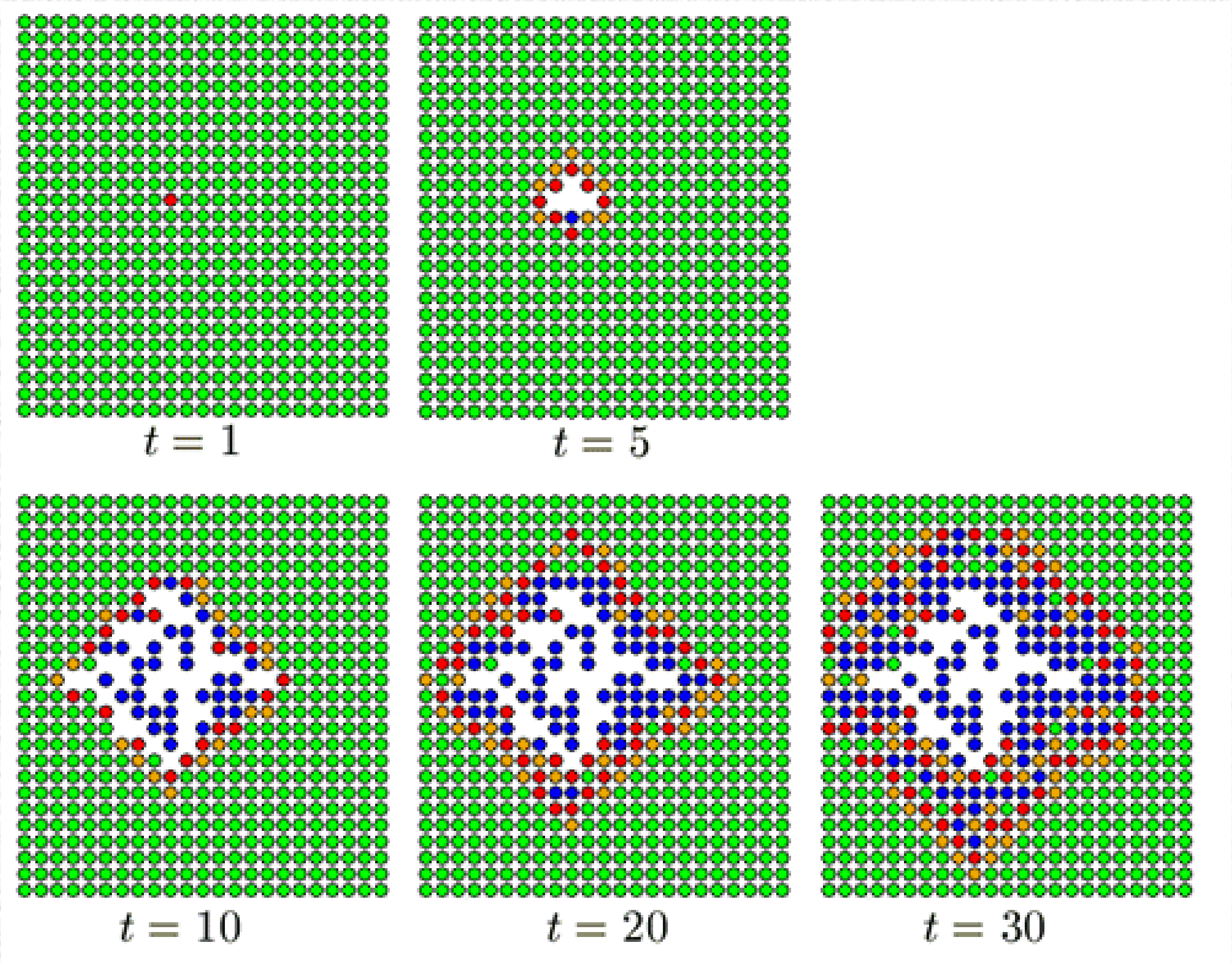
- Computer simulation models allow you to address complex questions without the complex mathematics that would be required from an analytical model
- Advantages
  - The complexity can be reduced by using algorithms in the place of equations
  - Different levels can be modeled (populations, individuals)
  - 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

Example: modeling disease spread

- Analytical models of spread of epidemics are very complex
- 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: must be susceptible, must be in an adjacent cell



Running the model

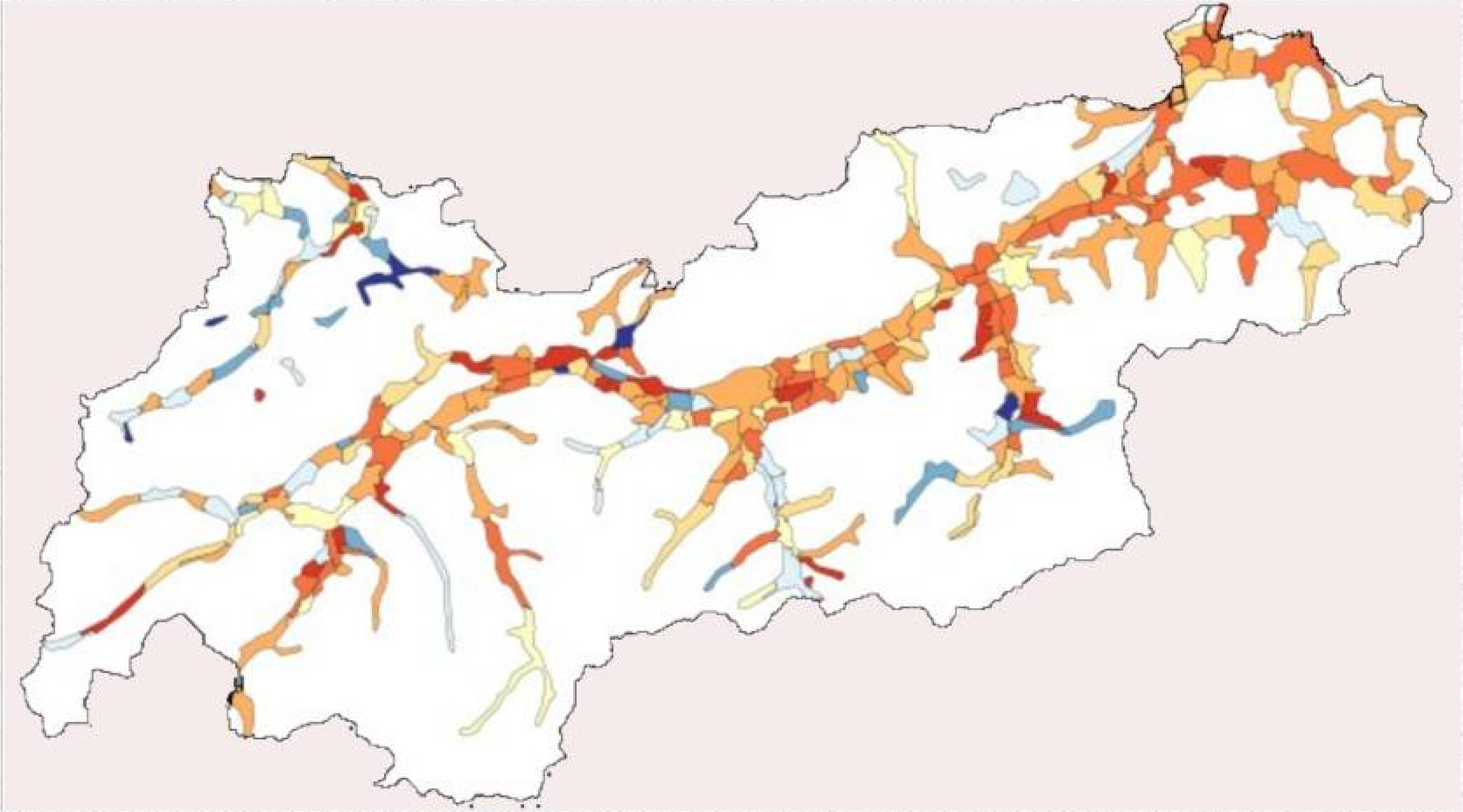


We can ask questions like:

How quickly will it spread?

What if we quarantine people with the disease? Is it too late by the time they start showing symptoms?

More realistic layout



More realistic – spread will be along the colored shapes, where the travel corridors are

Adding randomness to a computer simulation

- Simple matter of specifying a distribution for any parameter subject to randomness, drawing from it each step of the simulation
- With a stochastic simulation, results will never be identical twice
- Must be run repeatedly (hundreds, thousands, tens of thousands of times), and the outputs studied

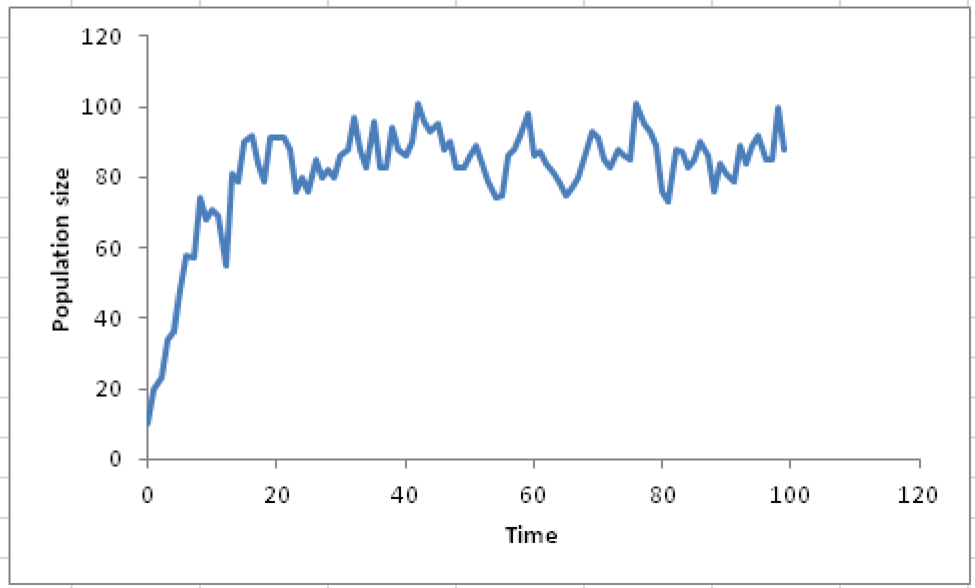
Simulating our stochastic population

- Run 1000 times for 100 years each
- Population sizes at each year recorded each time
- Mean size and range of 95% of runs identified, plotted
- N_0 = 10 for all
- If $N_t < 0.5$, then the population is extinct – set population to 0

Setup in Excel – first run

B3 =ROUND(NORMINV(RAND(), (1-E$4), F$4) * B2 * EXP(NORMINV(RAND(),E$2,F$2)*(1-B2/NORMINV(RAND(),E$3,F$3))),0)

	A	B	C	D	E	F
1	Time	Population size		Parameter	Mean	s
2	0	10		r	0.4	0.4
3	1	20		K	100	10
4	2	23		Mortality	0.05	0.01
5	3	34				
6	4	36				
7	5	48				
8	6	58				
9	7	57				
10	8	74				
11	9	68				
12	10	71				
13	11	69				
14	12	55				
15	13	81				
16	14	79				
17	15	90				
18	16	92				
19	17	84				
20	18	79				
21	19	91				
22	20	91				
23	21	91				
24	22	88				
25	23	76				
26	24	80				
27	25	76				
28	26	85				
98	96	85				
99	97	85				
100	98	100				
101	99	88				
102	100	88				
103						



Macro recorder

	A	B	C
1	Time	Population size	
2	0	10	
3	1	9	
4	2	11	
5	3	25	
6	4	24	
7	5	31	
8	6	38	
9	7	51	
10	8	55	
11	9	61	
12	10	63	
13	11	70	
14	12	68	
15	13	66	
16	14	77	
17	15	80	
18	16	80	
19	17	73	
20	18	68	
21	19	76	
22	20	75	
23	21	85	
24	22	82	
25	23	80	
26	24	97	
27	25	97	
28	26	99	
98	96	76	
99	97	74	
100	98	62	
101	99	59	
102	100	59	
103			

1. Copy the population sizes

2. Switch to results sheet, paste-special the values transposed into the first row

	A	B	C	D	E	F	G	H	I
1	Time.0	Time.1	Time.2	Time.3	Time.4	Time.5	Time.6	Time.7	Time.8
2	10	20	23	34	36	48	58	57	74
3									

3. Turn off the macro recorder

Add a loop

```
Sub PopSim()
'
' PopSim Macro
' Simulate stochastic population growth.
'
' Keyboard Shortcut: Ctrl+Shift+P
'
For i = 1 To 1000
    Sheets("CurrentSimulation").Select
    Range("B2:B102").Select
    Selection.Copy
    Sheets("Results").Select
    Range("A" & i + 1).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=True
Next i
End Sub
```

Select the right sheet

Select the pop sizes

Switch to results sheet

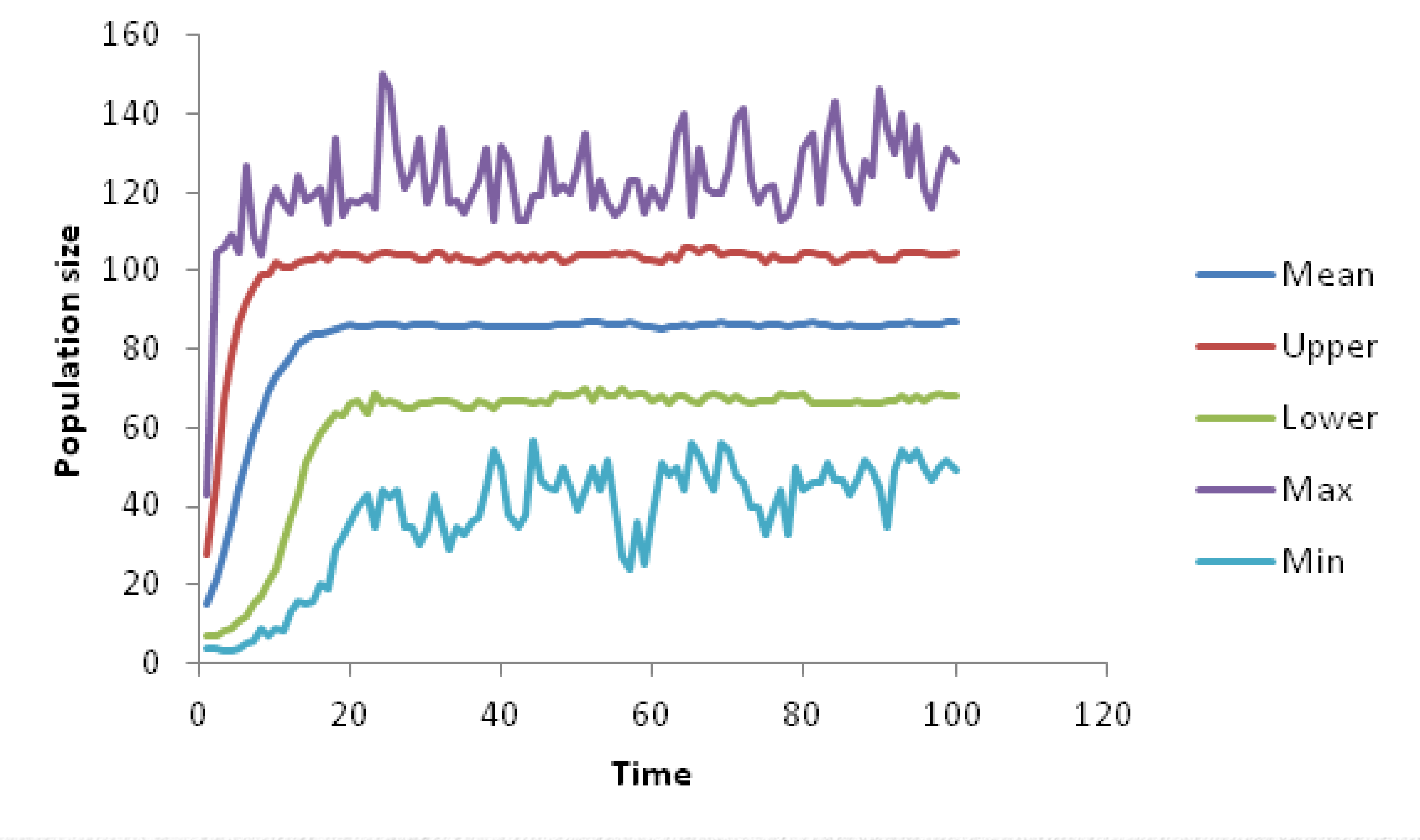
Select the first output cell

Paste special transposed the values

Run it, analyze the results

	A	B	C	D	E	F	G	H	I	J	K	L
1	Time.0	Time.1	Time.2	Time.3	Time.4	Time.5	Time.6	Time.7	Time.8	Time.9	Time.10	Time.
2	10	12	12	9	12	10	9	18	11	12	15	
3	10	8	5	9	9	16	15	35	44	46	84	
4	10	7	11	11	22	33	38	52	46	64	44	
5	10	16	31	34	31	43	49	61	69	83	80	
6	10	23	26	23	48	47	54	54	62	66	87	
7	10	11	13	12	11	9	14	16	27	31	75	
8	10	14	11	7	9	9	8	29	25	43	41	
9	10	9	16	16	47	68	73	72	102	99	88	
10	10	15	18	20	23	27	71	79	76	80	93	
11	10	12	19	19	28	35	56	62	70	65	92	
12	10	14	20	23	29	32	43	59	61	61	69	
13	10	13	13	19	26	26	31	47	53	53	58	
14	10	13	20	18	21	36	38	51	50	51	51	
15	10	14	43	82	86	91	85	87	78	89	80	
16	10	9	10	17	21	27	55	74	52	46	53	
996	10	10	21	37	38	45	60	71	74	80	87	
997	10	15	18	31	32	42	41	42	63	75	71	
998	10	14	25	47	49	79	89	86	103	98	95	
999	10	9	10	37	31	47	46	82	78	77	75	
1000	10	18	16	27	43	74	71	79	78	77	79	
1001	10	10	18	24	44	45	51	47	58	71	79	
1002												
1003												

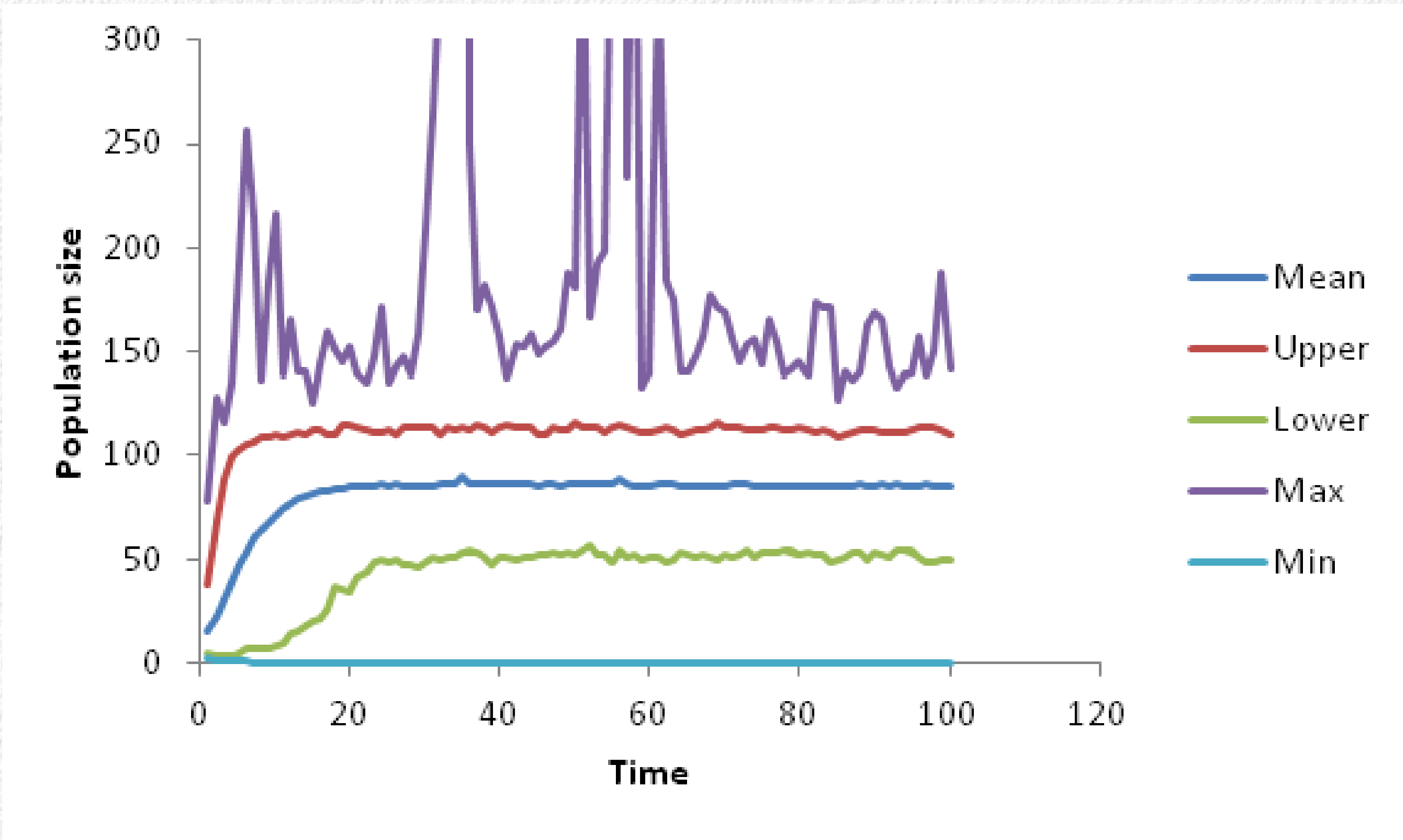
Mean, 97.5th and 2.5th percentiles, max, and min for each time



What do we learn?

- On average, pop size is fairly stable
- Population didn't go extinct in any run
- The mean was lower than K of 100
- The maximum and minimum values could be quite different from the mean, could even be different from the 2.5 and 97.5 percentile – occasional very high or very low population sizes can happen by chance
- What about more variation? Increase variation in r to 0.6

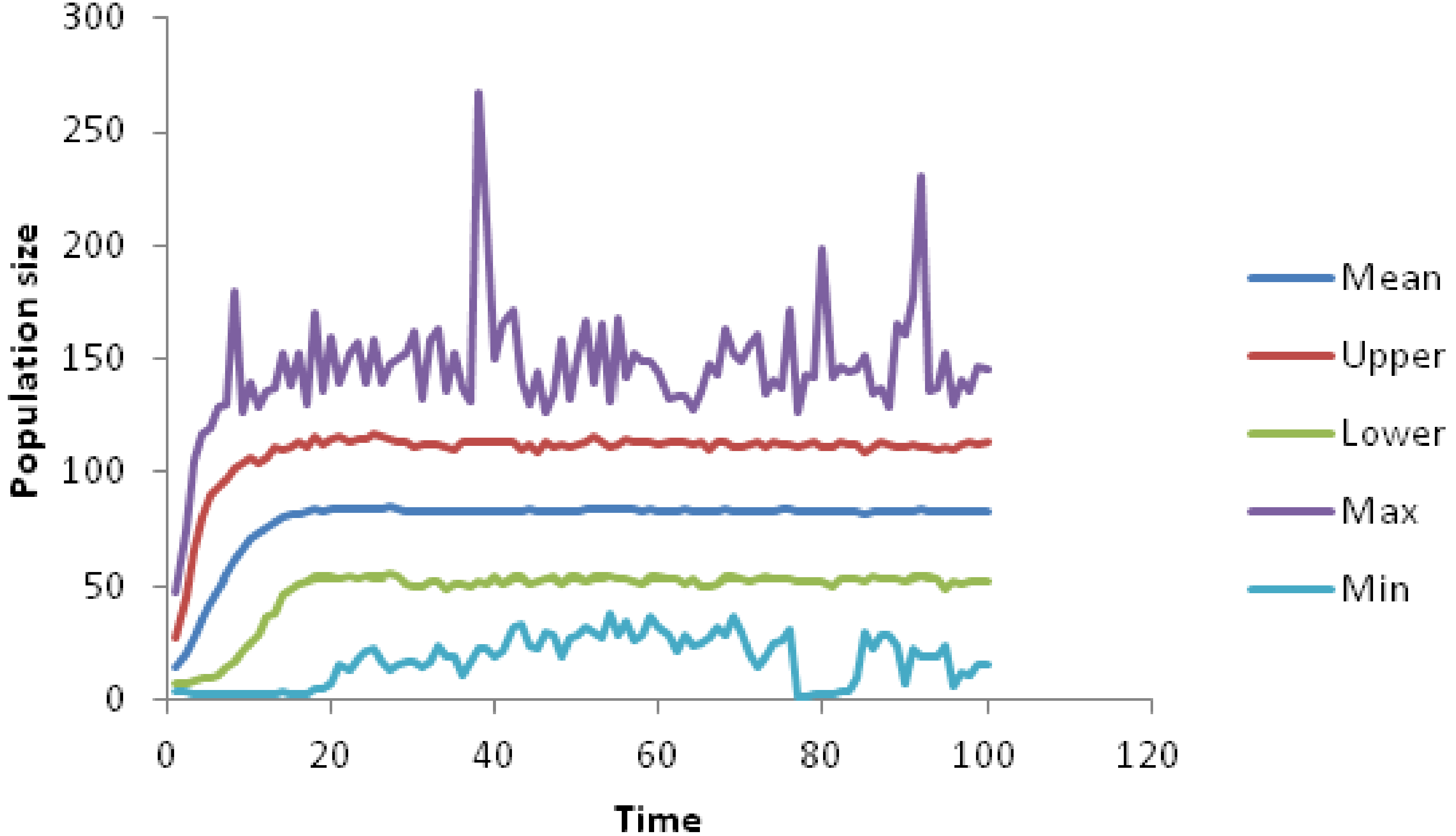
Mean, 97.5th and 2.5th percentiles, max, and min for each time



What did we learn?

- Still get an average that's fairly stable
- But, now some very large numbers and some extinctions (0.1 to 0.3% of runs)
- What if we put variation in r back to 0.4, and increase variation in carrying capacity to 20?

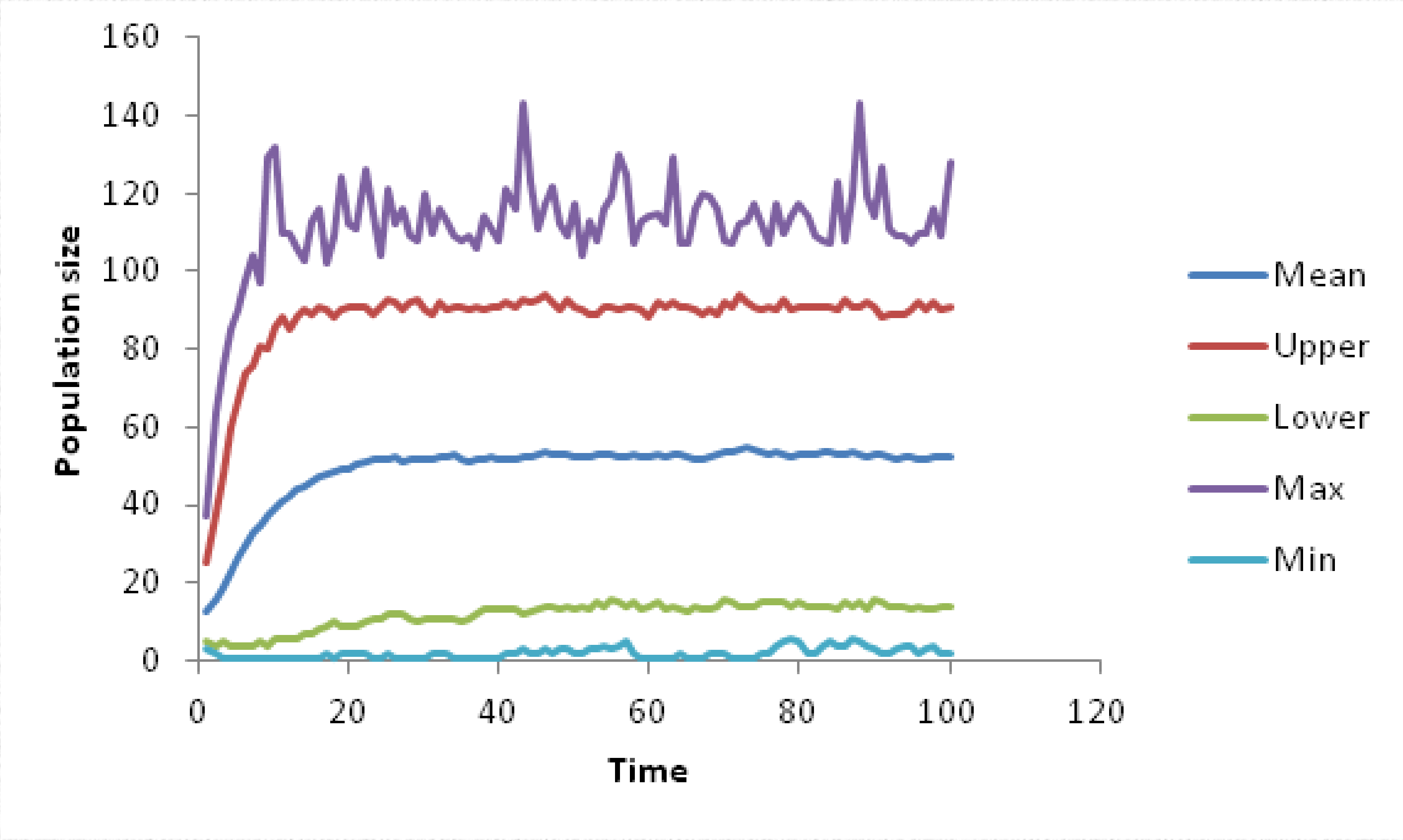
Mean, 97.5th and 2.5th percentiles, max, and min for each time



Variation in random mortality

- Set variation in K back to 10
- This time, make the average random mortality equal to 0, but set standard deviation to 20%
- Thus, small reductions will be most common, but large reductions will happen occasionally
  - Decreases of 40% will happen 5% of the time, and decreases of 60% will happen 1% of the time

Mean, 97.5th and 2.5th percentiles, max, and min for each time



Take-home re: computer simulation

- Computer simulations let us tackle complicated problems relatively simply
- The solutions will be numerical, rather than analytical
- We can incorporate randomness easily
- Once the simulation is set up, we can experiment with it in ways we can't in the real world
- But beware: computer simulations are abstractions of reality – when real experiments are possible, prefer them