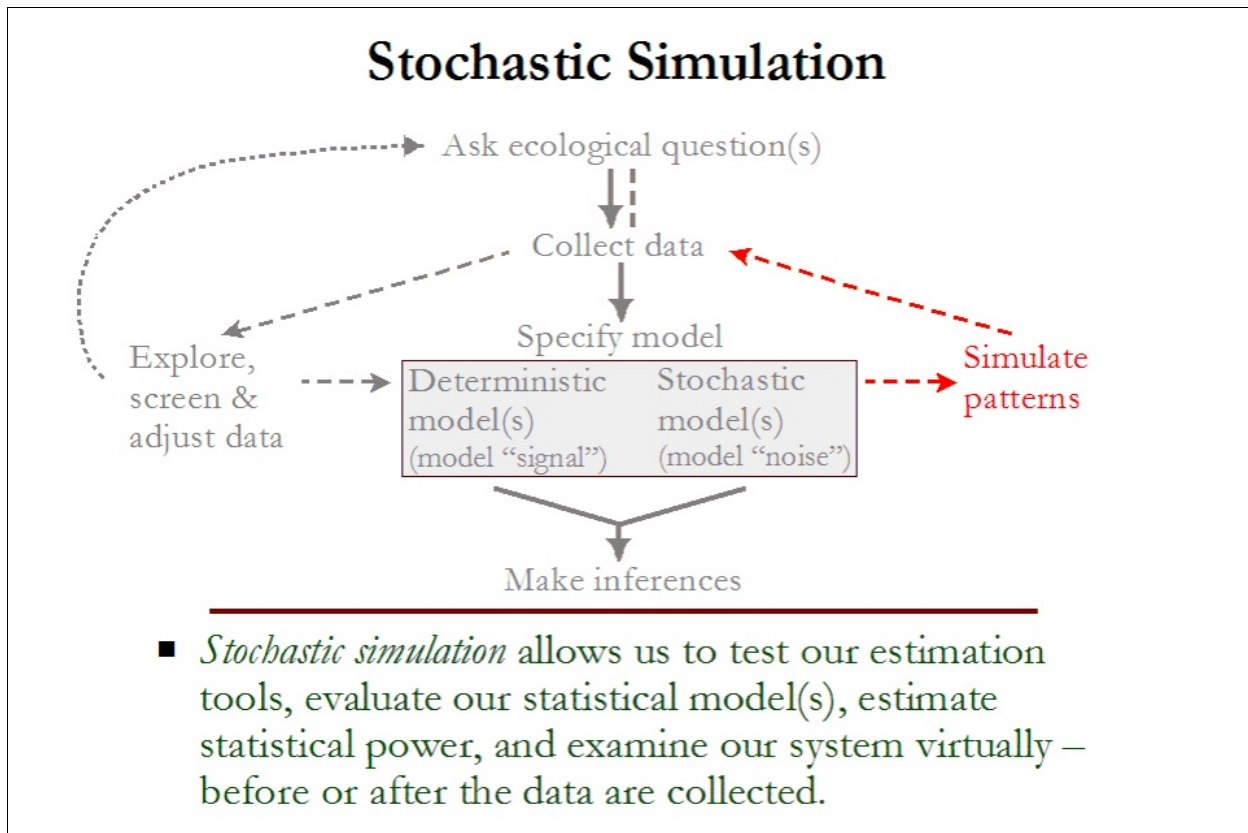


Analysis of Environmental Data

Chapter 10: Conceptual Foundations:

Stochastic Simulation

1. What is stochastic simulation?	2
2. Simulating static environmental processes.	3
3. Simulating dynamic environmental systems.	8
4. Applications of stochastic simulation.	12
4.1 Test estimation procedures.	13
4.2 Evaluate statistical models.	15
4.3 Conduct power analysis.	18
4.4 Evaluate model sensitivity.	22
4.5 Conduct virtual experiments.	24
4.6 Predict system behavior.	26



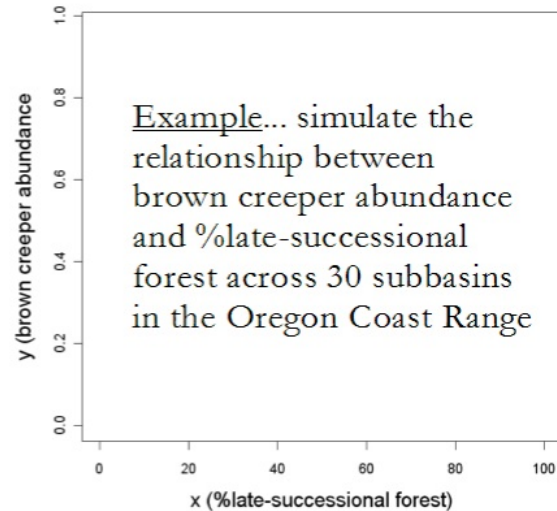
1. What is stochastic simulation?

Simulation is sometimes called *forward modeling*, to emphasize that we pick a model and parameters and work forward to predict patterns in the data. Simulation is the act of imitating a real system – in our case an environmental system. The “stochastic” part of the name simply means that the simulation model has a stochastic component so that each simulated pattern is a random outcome of the model. Typically, stochastic simulation involves repeatedly generating hundreds or even thousands of simulated patterns and then summarizing the range of variation in the patterns. Simulation is used by environmental scientists to gain insight into the functioning of environmental systems (more on this below).

Stochastic Simulation... static process

What is a static simulation?

- Environmental scientists use simulation to imitate real environmental systems
- In a static simulation, the data represent a *snapshot* of the environmental system
- We can use a single function to simulate the deterministic process and then add heterogeneity or stochasticity



2. Simulating static environmental processes

Static environmental processes, where the data represent a snapshot of some environmental system, are relatively easy to simulate. For static data, we can use a single function to simulate the deterministic process and then add heterogeneity or stochasticity. Often, however, we will chain together several different functions and probability distributions representing different stages in an environmental process to produce surprisingly complex and rich descriptions of environmental systems.

Example: We illustrate the process of simulating a static environmental process using a simple ecological model based on the now familiar Oregon birds data set. For this example, let's examine the relationship between brown creeper abundance and the extent of late-successional forest across 30 subbasins in the central Oregon Coast Range. For now, we will ignore the real data and simply simulate some data based on a particular statistical model.

Stochastic Simulation... static process

Example: *Brown creeper abundance*

Statistical model:

$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

Read as Y is a random variable drawn from a normal distribution with a mean $a + b \cdot x$ and standard deviation σ



Decomposing the model:

- What is the *data*?
- What are the *parameters*?
- What is the *deterministic* component?
- What is the *stochastic* component?

Let's assume a linear model (deterministic component) with normal errors (stochastic component), which we can write as:

$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

which specifies that Y (brown creeper abundance) is a random variable drawn from a normal distribution with a mean $a + b \cdot X$ and standard deviation σ . In this notation, a and b are parameters of the deterministic linear model (intercept and slope, respectively), X is data (values of the explanatory variable, %late-successional forest), and σ is a parameter of the stochastic component of the model (the standard deviation of the normally distributed errors). This means that the i^{th} value of Y , y_i , is equal to $a + b \cdot x_i$ plus a normally distributed error with mean equal to the linear model and standard deviation σ .

Stochastic Simulation... static process

Example: Brown creeper abundance

Statistical model:

$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

$$y_i = a + b \cdot x_i + \varepsilon_i$$

First random observation = y_1

$$y_1 = \{0.099 + 0.006 \cdot 0.71\} + \{-0.188\}$$

$$y_1 = \{0.103\} + \{-0.188\}$$

$$y_1 = -0.085$$

Let's assume:

- $a = 0.099$
- $b = 0.006$
- $\sigma = 0.14$
- $x = (0.71, 80.15, \dots, 59.25)$

Next, we create a simulation for our linear model that draws a vector y from the specified model given values of x and the model parameters: a , b and σ . Note, each draw results in a different vector y because of the stochastic part of the model – a random error is added to each fitted value from the deterministic model. Thus, each simulated data set represents a single stochastic outcome of the model.

There are two ways to think about this. First, because for the normal distribution the mean is independent of the variance (error), we can compute the mean from the deterministic portion of the model and then literally add an error drawn from a normal distribution with a mean=0 and sd= σ , as shown in the top figure. This is because the deterministic model gets us to the mean and the error model gives us a distribution about the mean; in this way we are

Stochastic Simulation... static process

Example: Brown creeper abundance

Alternatively

Statistical model:

$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

$$y_i = a + b \cdot x_i + \varepsilon_i$$

First random observation = y_1

$$y_1 = \text{Norm}(\text{mean} = 0.099 + 0.006 \cdot 0.71, \text{sd} = 0.14)$$

$$y_1 = -0.085$$

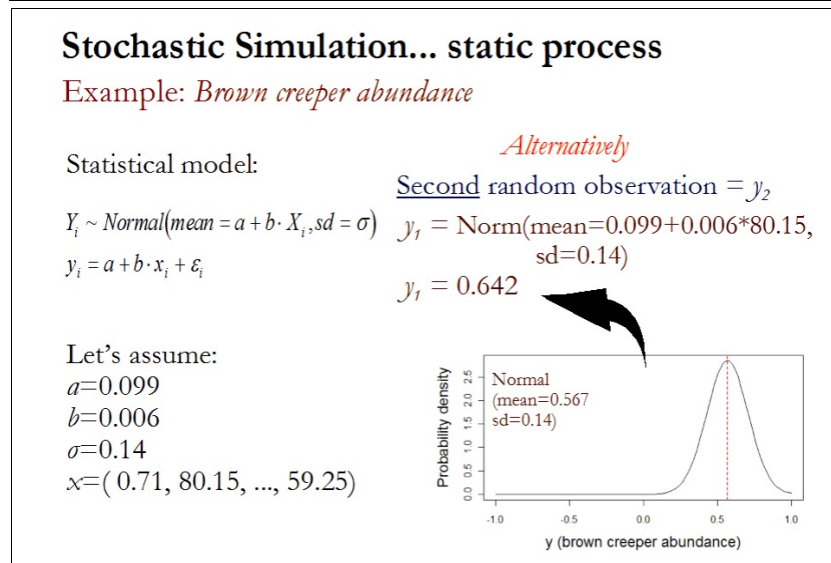
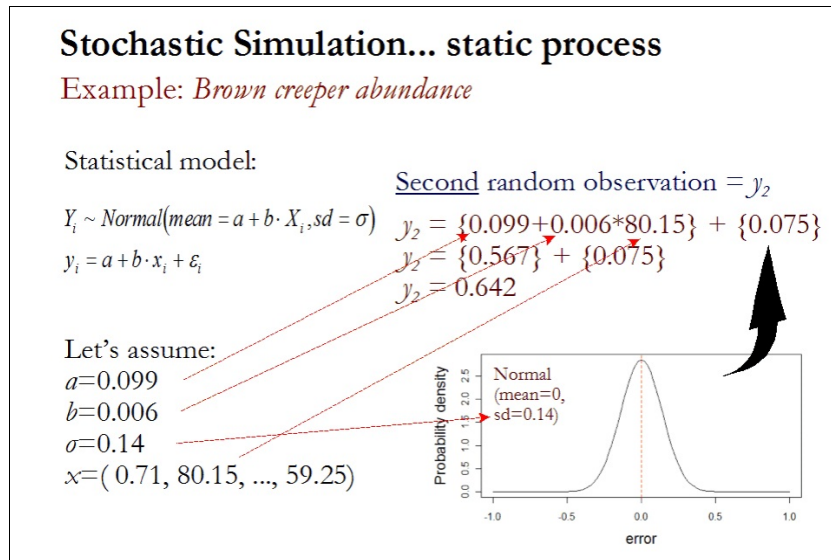
Let's assume:

- $a = 0.099$
- $b = 0.006$
- $\sigma = 0.14$
- $x = (0.71, 80.15, \dots, 59.25)$

centering the normal on the mean, which is the same as specifying a mean of zero. In the example shown here, for the first random point taken at $x=0.71$, we compute the mean from the linear model to be equal to 0.103 and then add a randomly drawn error from a normal distribution with mean=0 and sd=0.14, which in this example was an error of -0.188, giving a value of -0.085 for the first random point.

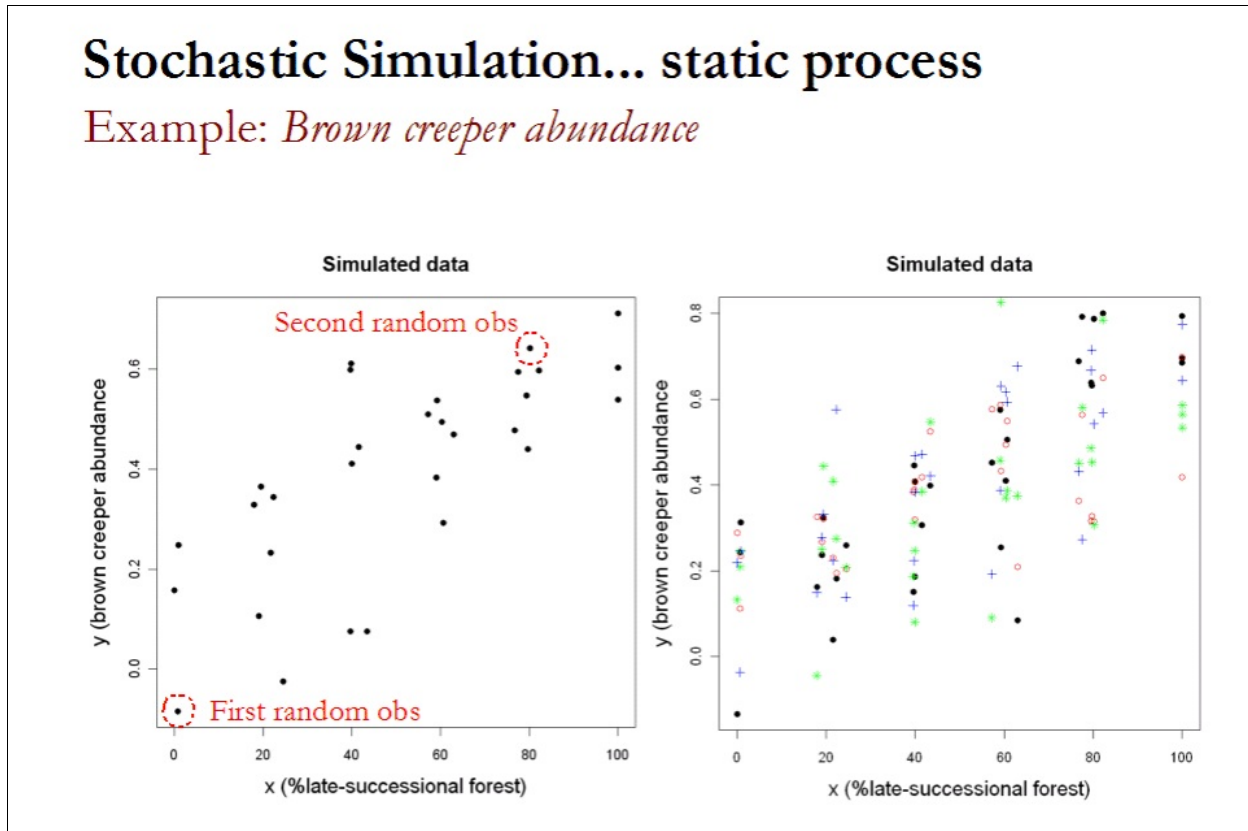
Alternatively, we can compute the random value of y directly from the normal distribution with a mean=0.103 (from the linear model) and sd=0.14, which in this case produced a value of -0.085. This is the same as saying, if my expected value is 0.103 and my spread is 0.14, then pick a value of y at random from a normal distribution defined by those parameter values. This is shown in the second figure.

We repeat this process for as many points as we desire. The figures below illustrate the process for the second random point.



Stochastic Simulation... static process

Example: *Brown creeper abundance*

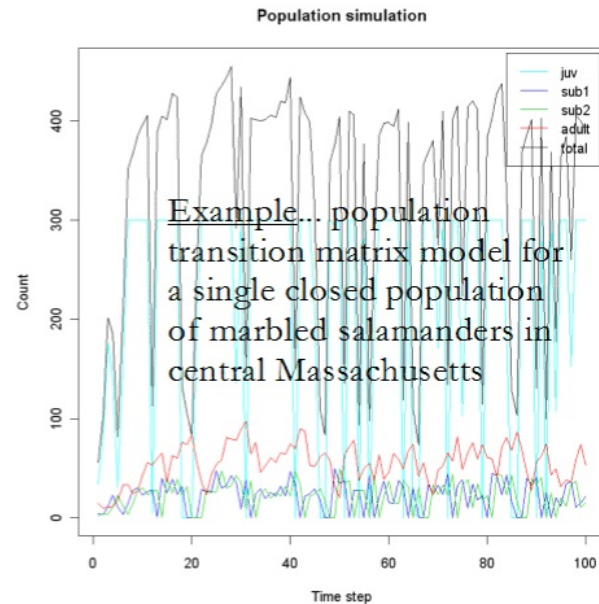


Next, we plot the simulated data to see what we got. Remember, each simulated data set represents a new set of points randomly drawn from the same model. If we run the simulation many times, we get lots of new data sets, each one representing a single stochastic outcome of the model. There are lots of things we can do with a simulation like this – which we will explore later. For now, it is sufficient to understand how we simulated a static environmental process.

Stochastic Simulation... dynamic process

What is a dynamic simulation?

- Environmental systems are *dynamic* – they change over time
- In a dynamic simulation, we imitate how environmental processes drive environmental systems to change (e.g., how populations change over time)



3. Simulating dynamic environmental systems

The previous example dealt with a *static* environmental process; the data represented a snapshot of the environmental system. However, environmental systems are *dynamic* – they change over time. Therefore, many problems require a dynamic modeling approach. Dynamic models are models that describe how environmental processes drive environmental systems to change over time. Dynamic models are a vast and increasingly complex subject and therefore one that we cannot hope to cover in depth. The most common dynamic modeling application involves modeling population change over time. Here, we will simply introduce dynamic models by constructing a very simple population transition matrix model. Knowing how to simulate dynamic models is important because fitting dynamic models to data is so tricky that it is essential to model simulated data to confirm that our inference methods work.

Stochastic Simulation... dynamic process

Example: *Population matrix model*

Population structure:

juvenile (0)
subadult1 (1)
subadult2 (2)
adult (3+)



Leslie transition matrix:

0.00	0.00	0.00	22.32
0.14	0.00	0.00	0.00
0.00	1.00	0.00	0.00
0.00	0.00	1.00	0.66

← fecundity = number young produced per female per timestep

survival = probability of surviving to the next age class


Let's build a dynamic population model for a single closed population of a local population of marbled salamanders without immigration or emigration. We will represent four classes: juvenile, subadult1, subadult2, and adult, corresponding to year 0, 1, 2, and 3+. To control the key population vital rates, fecundity (number of young produced per breeding female per timestep) and survival (probability of surviving between timesteps), we will employ a *Leslie transition matrix*, which records the fecundity and survival rates for each age class in the population.

Briefly, the transition matrix defines how the population will change at each time step based on fecundity (i.e., addition of new individuals through reproduction) and mortality (i.e., loss of individuals through death). Specifically, the transition matrix contains the age-class specific fecundities in the first row. In the example shown here, only the adults produce offspring, and they do so at the average rate of 22.32 young per breeding female per year. Survival rates for each age class are given in the remaining rows. For example, in this example, the juvenile age class individuals survive to the subadult1 age class at the average rate of 0.14. Thus, for every 100 young that are produced, only 14 will survive on average to the next year. Similarly, adult survival rate is given as 0.66, so that on average 66 out of 100 adults will survive to the next year. The subadult1 and subadult2 age class individuals each survive at a rate of 1.0, which means that we pass individuals through the subadult stages without any mortality; this parameterization is necessary to account for the fact that our field estimates of juvenile survival represent survival from juvenile to adult stages over a three year period and that we want to use a one year timestep in the model (which corresponds to the annual breeding cycle) rather than a three year timestep. For our purposes, it is not critical to understand this model parameterization further.

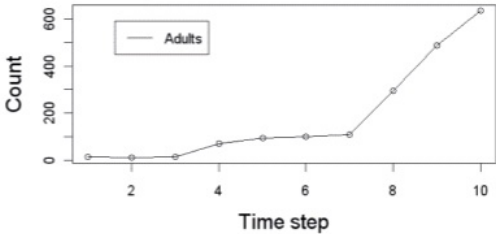
Stochastic Simulation... dynamic process

Example: Population matrix model

<u>Age class</u>	Population P_0	Transition matrix (T)	P_1	P_2																								
juvenile	40	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>0.00</td><td>0.00</td><td>0.00</td><td>22.32</td></tr> <tr><td>0.14</td><td>0.00</td><td>0.00</td><td>0.00</td></tr> <tr><td>0.00</td><td>1.00</td><td>0.00</td><td>0.00</td></tr> <tr><td>0.00</td><td>0.00</td><td>1.00</td><td>0.66</td></tr> </table>	0.00	0.00	0.00	22.32	0.14	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.66	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>446</td></tr> <tr><td>6</td></tr> <tr><td>2</td></tr> <tr><td>15</td></tr> </table>	446	6	2	15	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>335</td></tr> <tr><td>62</td></tr> <tr><td>6</td></tr> <tr><td>12</td></tr> </table>	335	62	6	12
0.00	0.00	0.00	22.32																									
0.14	0.00	0.00	0.00																									
0.00	1.00	0.00	0.00																									
0.00	0.00	1.00	0.66																									
446																												
6																												
2																												
15																												
335																												
62																												
6																												
12																												
\times		$=$		$\times T =$																								
\times		$\times T$		$\times T \dots$																								



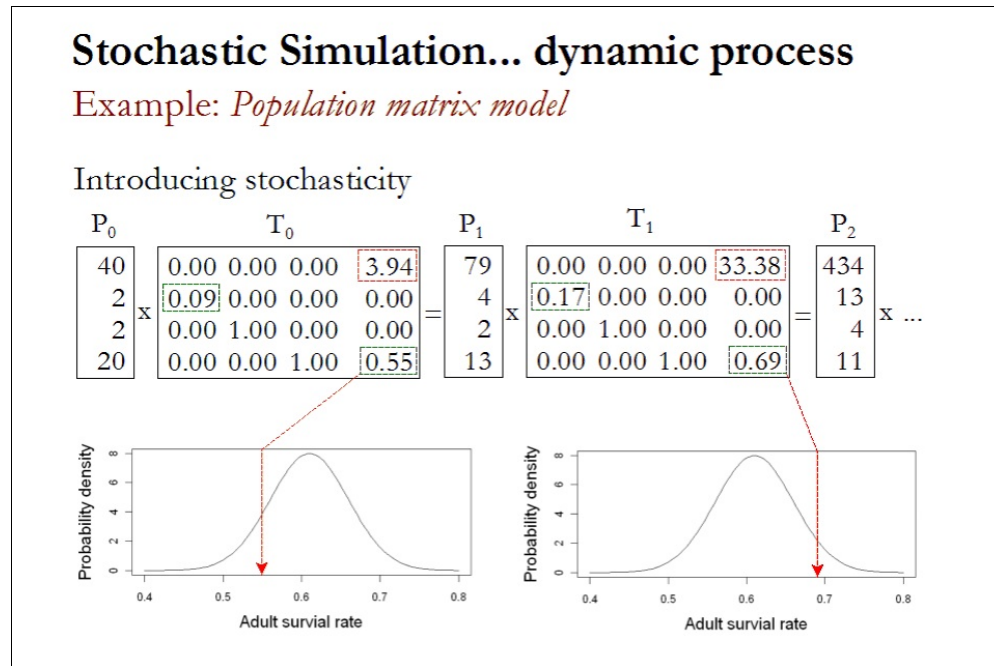
Population simulation



The population model works as follows. First, we define an initial population vector which gives the number of individuals in each age class at the start of the simulation (P_0), so in our case it is a vector of length four. Then we multiply this population vector by the transition matrix (T), which accounts for age-specific fecundity and survival, to determine the population vector for the next timestep (P_1). Lastly, we repeat this process for as many timesteps as we desire. The result is a population trajectory over time, as shown here in the plot of adult population size (y-axis) over time (x-axis).

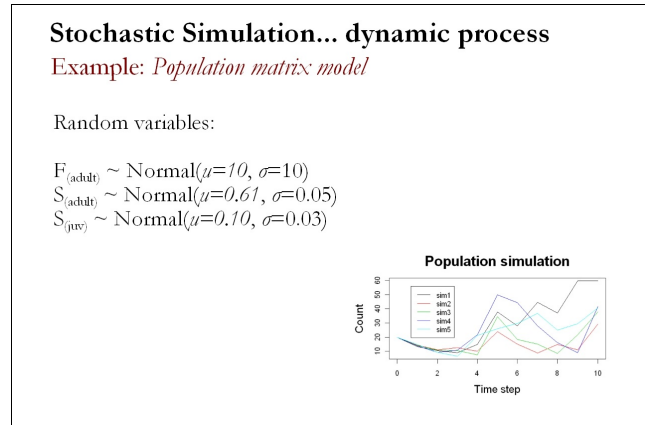
If we simulate the population in this manner, is our dynamic model stochastic?

No, we have not introduced any source of random variability to the model. The population trajectory is absolutely predetermined by the starting population vector and the values in the transition matrix.



How might we make this model stochastic to better reflect the true nature of population dynamics? First, we have to recognize that our estimates of fecundity and survival are only just that – estimates. We do not know what the true rates are, we only have our uncertain estimates. Moreover, the population vital rates are not likely to remain constant year after year, since there are numerous factors affecting these vital population processes every year. For example, a drought in one year may cause fecundity rates to drop near zero; a disease outbreak may cause survival rates to drop temporarily; favorable weather conditions in the fall and winter might leave more juveniles alive for the following year than usual, and so on.

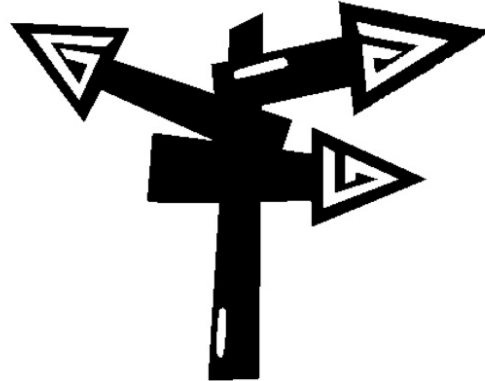
The easiest way to make this dynamic simulation model stochastic (and thus more realistic) is to incorporate random variation into our vital rates. We do this by drawing our vital rates from a random distribution (e.g., a random normal distribution) at each timestep. Consequently, in the first timestep the adult fecundity rate might be 35.2, while in the second timestep it might be 0.5. Let's rerun the model, only this time let's draw our fecundity and survival estimates from random normal distributions with pre-specified means and standard deviations, and see what happens. Now the population trajectory is no longer predetermined. Each time we run the simulation we get a different population trajectory. We now have a truly stochastic and dynamic simulation model. There are lots of things we can do with a simulation model like this – which we will explore later. For now, it is sufficient to understand how we simulated a dynamic ecological process.



Stochastic Simulation... applications

How can we use stochastic simulation?

- Test estimation procedures
- Evaluate statistic models
- Conduct power analysis
- Evaluate model sensitivity
- Conduct virtual experiments
- Predict system behavior



4. Applications of stochastic simulation

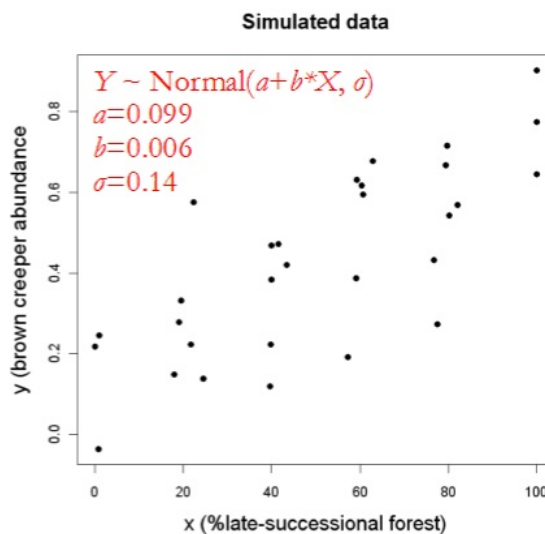
Environmental scientists use stochastic simulation in all sorts of ways and, thanks to increasing computing power and the accessibility of programming languages like R, the use of stochastic simulation in environmental modeling is increasing rapidly. Some of the most common uses are to:

- Test estimation procedures
- Evaluate statistic models
- Conduct power analysis
- Evaluate model sensitivity
- Conduct virtual experiments
- Predict system behavior

Stochastic Simulation... applications

Test estimation procedures

- We never know the true answer to an environmental question – we only have imperfect measurements with which we're trying to get as close to the answer as possible
- Simulation is the only way to test whether we can correctly estimate the parameters of an environmental system



4.1 Test estimation procedures

We can use stochastic simulation to test our estimation procedures. Since we never know the true answer to an environmental question – we only have imperfect measurements with which we're trying to get as close to the answer as possible – simulation is the only way to test whether we can correctly estimate the parameters of an environmental system. It's always a good idea to test such a best-case scenario, where we know that the functions and distributions we're using are correct, before we proceed to real data.

Example.—We continue with our previous example of modeling brown creeper abundance in relation to the extent of late-successional forest across 30 subbasins in the central Oregon Coast Range. Let's see how well our estimation procedure estimates the true population parameters. Let's assume the same values as before for each of the parameters of our linear model: $a=0.099$, $b=0.006$ and $\sigma=0.14$. Next, let's simulate some data from this model. Remember, each simulation produces a different outcome because it is a stochastic model. Now, let's use an estimation procedure to estimate the population parameters based on the simulated ample data set. For this purpose, we will use the method of *ordinary least squares* which, under the assumption of independent and normally distributed errors, gives us the maximum likelihood estimates, or the most likely parameter values; i.e. the ones that make our sample data the most likely outcome under hypothetical repeated sampling. Here, it is not important to understand the particular estimation method, only that we have chosen one and wish to see whether it can estimate the known population parameters.

Stochastic Simulation... applications

Test estimation procedures

$$Y \sim \text{Normal}(a+b*X, \sigma)$$

True values:

$$a=0.099$$

$$b=0.006$$

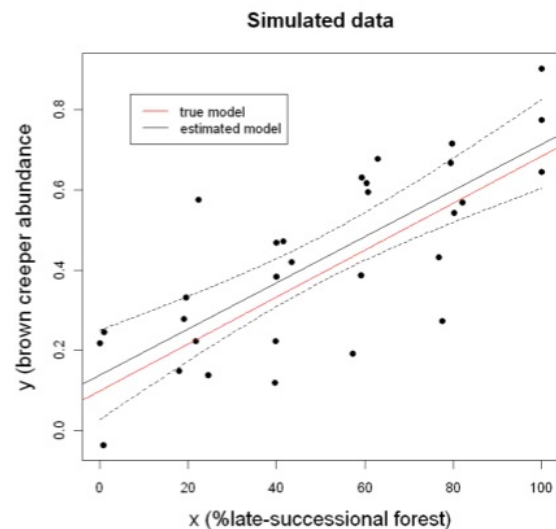
$$\sigma=0.140$$

Estimated values (95% ci):

$$a=0.139 \text{ (0.029-0.251)}$$

$$b=0.006 \text{ (0.004-0.008)}$$

$$\sigma=0.151$$

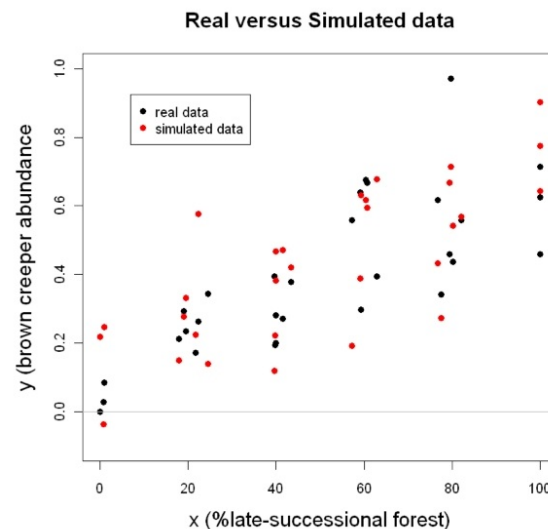


The results of our estimation procedure are shown here. The true intercept (a) is likely to be between 0.029 and 0.251 with a 95% level of confidence; that is to say, there is a 95% chance that the true intercept falls between these values. Similarly, the true slope (b) is likely to be between 0.004 and 0.008 with a 95% level of confidence. Lastly, the standard deviation (σ) of the errors – the deviations of the observed values from the fitted or predicted values, was 0.151, which is pretty close to the true value. Thus, we can conclude that our estimation procedure did a pretty good job of estimating the population parameters. Granted, with a simple linear model, it is not too surprising that our estimation method is effective. However, with more complex models and more sophisticated estimation procedures, the end result is not always going to be so obvious.

Stochastic Simulation... applications

Evaluate statistical models

- After we collect data, we can use stochastic simulation to explore the functions and distributions we chose for our statistical model
- If we can choose parameters that make the simulated output look like our real data, we can confirm that the model is reasonable



4.2 Evaluate statistical models

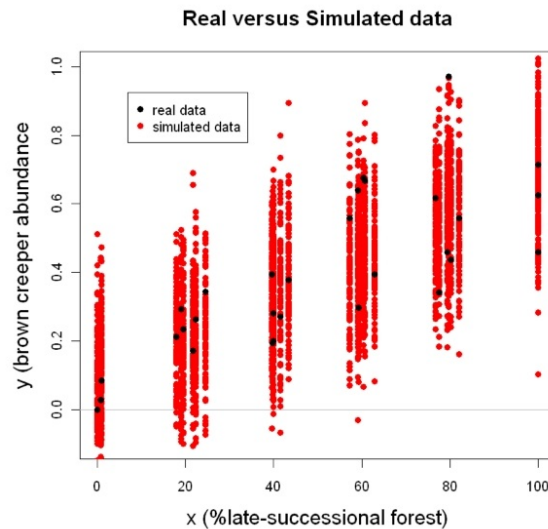
After we collect data, we can use stochastic simulation to explore the functions and distributions we chose for our statistical model. If we can choose parameters that make the simulated output from those functions and distributions look like our real data, we can confirm that the model is reasonable – and simultaneously find a rough estimate of the parameters (which we may need in the estimation procedure).

Example.—We continue with our previous example of modeling brown creeper abundance in relation to the extent of late-successional forest across 30 subbasins in the central Oregon Coast Range. Let's see how well our statistical model approximates the patterns observed in the real data. First we simulate a data set using our linear model with normal errors and plot the simulated data with the real data. Is the simulated pattern of points the same as the original pattern? Are there any notable discrepancies?

Stochastic Simulation... applications

Evaluate statistical models

- Is the simulated pattern of points the same as the original pattern?
- Are there any problems with the model?



We might want to run the model several more times to see how variable the results are – it is a stochastic model after all. After running the model several times, are there any notable problems with the model? In other words, does the model reproduce the patterns in the original data perfectly or are there issues with the spread of values or with the generation of illogical values?

One problem with the use of the *normal* distribution is that it is unbounded on the lower limit. Thus, negative values are possible. In this case, because the y -intercept is close to 0, the simulation is likely to produce negative values occasionally when $x \rightarrow 0$. Since brown creeper abundance cannot be negative, this is an undesirable behavior of the model.

Another apparent problem with the use of the normal distribution is that the mean and variance are independent, which translates into a constant variance across the full range of x . In other words, as the mean of y changes as x increases, the variance is assumed to stay the same. This does not appear to be the case with the brown creeper data, as the variance appears to increase with the mean.

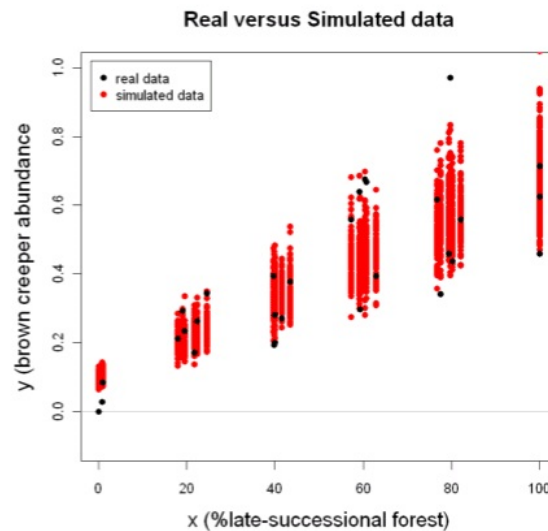
Stochastic Simulation... applications

Evaluate statistical models

Let's try a gamma error model:

$$Y \sim \text{Gamma}(\text{shape}, \text{scale} = (a + b \cdot X) / \text{shape})$$

- Are there any problems with this model?





One way to fix this problem is to use the *gamma* distribution, which allows only positive value and has a variance that is dependent on the mean, which allows the variance to increase with the mean. Note, the gamma distribution has two parameters called shape and scale, instead of the mean and variance parameters of the normal distribution. Scale is equal to the mean divided by shape, where the mean is the deterministic part of the model or the expected value for any given value of x . See the lecture on probability distributions for more detail on the gamma distribution. Now, let's simulate data using our linear model but this time with gamma errors instead of normal errors, and like before, let's replicate the simulation several times and plot the simulated data with the real data. How does the simulated pattern compare to the real data using the new model? Are there still problems with the model?

One issue that arises with the gamma distribution is that $y_i=0$ is not allowed, so we never predict zeros for the response, even though zeros are legitimate values of the response variable, brown creeper abundance, in this example. This appears to be the only notable problem with this model. Even in this simple example, stochastic simulation has helped us choose an appropriate model. In more complex situations, the use of stochastic simulation to evaluate and choose a statistical model can be vitally important.

Stochastic Simulation... applications

Conduct power analysis

- Power = the probably of correctly rejecting the null hypothesis when it is false  How large a sample size we would need to achieve a certain power?
- Power analysis involves asking a more general question: How do the quality and quantity of our data and the true properties (parameters) of our environmental system affect the quality and quantity of the answers to our questions about environmental systems?  How large a sample size we would need to get a reasonably precise estimate of our parameters?
Which study design would lead to the most precise estimates of our parameters?

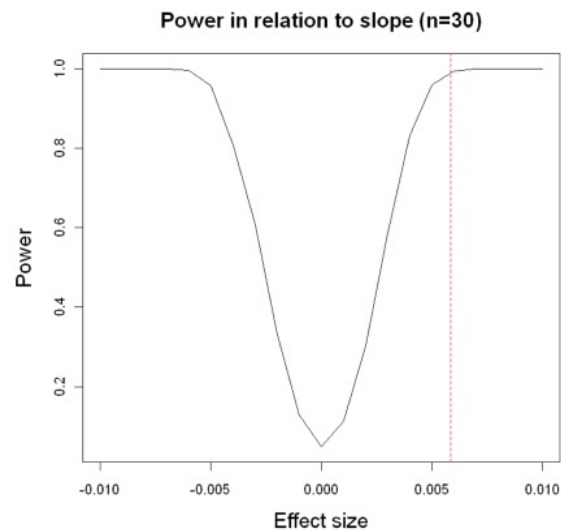
4.3 Conduct power analysis

Power analysis in the narrowest sense means figuring out the (frequentist) statistical power – the probably of correctly rejecting the null hypothesis when it is false. While we are generally less concerned with power analysis in the conventional sense of hypothesis testing, we are very interested in the role of power analysis in addressing a much broader question: How do the quality and quantity of our data and the true properties (parameters) of our environmental system affect the quality and quantity of the answers to our questions about environmental systems? For any real experiment or observation situation, we don't know what is really going on (the “true” model or parameters), so we don't have the information required to answer these questions from the data alone. But we can approach them by analysis or simulation. Historically, questions about statistical power could only be answered by sophisticated analyses, and only for standard statistical models and experimental designs such as one-way analysis of variance and linear regression. However, increases in computing power have extended power analysis to many new areas. Stochastic simulation is a powerful tool for conducting power analysis. It is a specific kind of simulation testing where we explore how large a sample size we would need to get a reasonably precise estimate of our parameters. The more precisely we can estimate our parameters, the more likely is it that we will be able to reject our null hypothesis (if we have one) in favor of our alternative given that the null hypothesis is actually wrong – which is the right thing to do. We can also use power analysis to explore how variations in study design would change our ability to answer questions; e.g., which of several alternative study designs would lead to the most precise estimates of our parameters?

Stochastic Simulation... applications

Conduct power analysis

1. Specify model:
 - a. $Y \sim \text{Normal}(0.099+0.006X, \sigma=0.14)$
2. Simulate data set ($n=30$)
3. Estimate parameters
4. Compute p -value and decide whether to accept/reject null
5. Repeat process many times
6. Compute proportion of times null rejected = power



Example.—We illustrate the use of stochastic simulation for power analysis using the linear model above. Let's start by finding out whether we can reject the null hypothesis in a single experiment. To do this, we simulate a data set with a known parameters: intercept ($a=0.099$), slope ($b=0.006$) and error ($\sigma=0.14$), and number of data points ($n=30$); then we estimate the parameters of the model from the data (using an estimation procedure); then we compute the probability of observing the data if in fact the null hypothesis (of no relationship between x and y) were true (we call this the p -value); and then we see whether it is less than our specified alpha (α) criterion (usually 0.05). In this example, we used ordinary least squares (see above) on the simulated data set and estimated the slope parameter to be 0.007, and the probability is less than 0.001 that we would have observed a slope this large if in fact the true underlying population slope was 0 (i.e., $p < 0.001$). Pretty strong evidence that the null hypothesis of no slope is wrong, so we reject the null hypothesis. This is the correct thing to do because we know that the true slope from which this sample was drawn is 0.006, not 0. However, to estimate the probability of successfully rejecting the null hypothesis when it is false (the power), we have to repeat this procedure many times and calculate the proportion of the time that we correctly reject the null hypothesis. We do so and find out that we correctly reject the null hypothesis 1,000/1,000 times. So what is our power in this case?

But this is the power to detect a slope of roughly $b=0.006$ with a sample size of $n=30$, given our specified statistical model (i.e., linear model with independent normal errors). Usually we don't just want to know the power for a single experimental design. Rather, we want to know how the power changes as we change some aspect of the design, such as the sample size or the effect size (slope, in this case). Thus, we have to repeat the entire procedure multiple times, each time changing some parameter of the simulation such as the slope or the sample size.

Shown in the figure above is how power changes as a function of the slope (effect size). The vertical line shows the power for our initial slope estimate of 0.006.

What is the power for a slope of say .002?

We can do the same thing for a gradient in sample sizes. Shown below is how power changes as a function of the sample size, while holding slope constant at 0.006. The vertical line shows the power for our original sample size of 30.

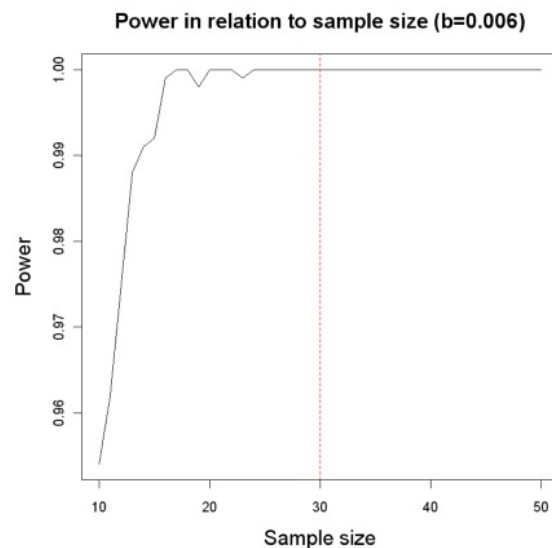
How much power is lost if we reduce the sample size from 30 to 20, or to 10?

What would happen if our model error doubled; i.e., if the residual variation (stochastic component) about the expected value (deterministic component) doubled?

Stochastic Simulation... applications

Conduct power analysis

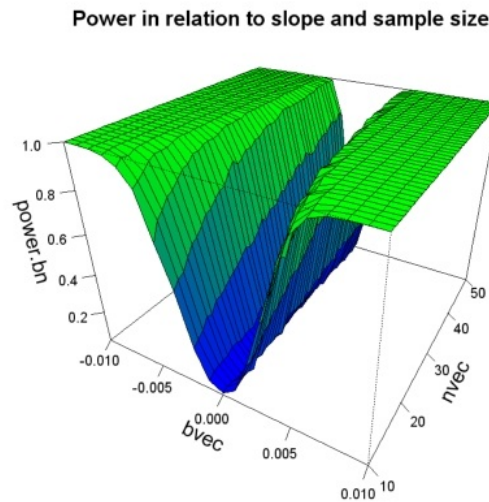
1. Specify model:
 - a. $Y \sim \text{Normal}(0.099+0.006X, \sigma=0.14)$
2. Simulate data set ($n=30$)
3. Estimate parameters
4. Compute p -value and decide whether to accept/reject null
5. Repeat process many times
6. Compute proportion of times null rejected = power



Stochastic Simulation... applications

Conduct power analysis

1. Specify model:
 - a. $Y \sim \text{Normal}(0.099 + 0.006X, \sigma=0.14)$
2. Simulate data set ($n=30$)
3. Estimate parameters
4. Compute p -value and decide whether to accept/reject null
5. Repeat process many times
6. Compute proportion of times null rejected = power



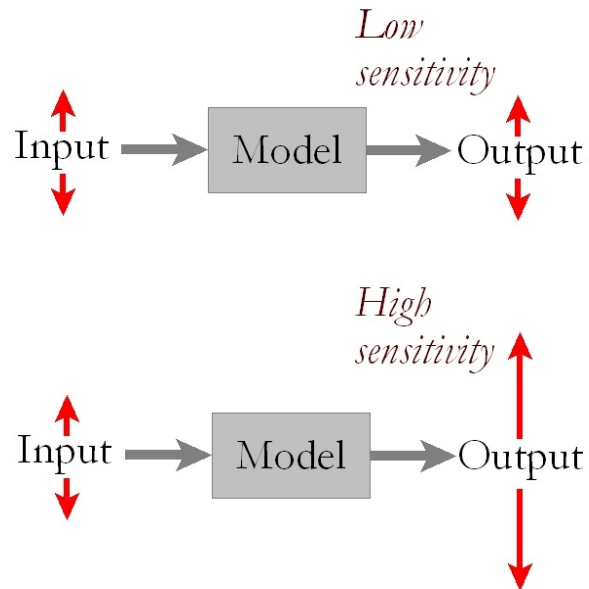
We could repeat this process for other parameters such as the error component of the model, but you get the idea. While we can do these power analysis simulations for one parameter at a time, it might be more interesting to vary combinations of parameters, say of slope and sample size. Shown here is how power changes as a function of the slope (effect size) and sample size, let both parameters vary over reasonable ranges.

What does the power surface reveal about the relationship between slope and sample size? If we wanted say a power of >0.8 to detect a slope of $b=.002$, how large would our sample size need to be?

Stochastic Simulation... applications

Evaluate model sensitivity

- Sensitivity = how much do the result changes per unit of change in each of the input parameters?
- The parameters with the greatest sensitivity are the ones we need to worry about the most
- Sensitive parameters have the biggest impact on results



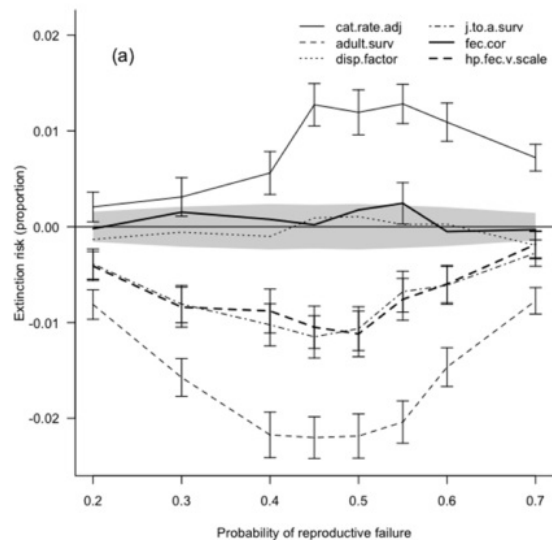
4.4 Evaluate model sensitivity

We can use stochastic simulation to evaluate model sensitivity. Sensitivity refers to how much the result changes per unit of change in each of the input parameters. The parameters with the greatest sensitivity are the ones we need to worry about the most, since even little changes in their values can have a relatively big impact on the results. Sensitivity analysis can be quite useful as a guide for future studies, since it can direct future research towards the most sensitive parameters – the ones that matter most. This can be especially important in complex multi-parameter models, where complex interactions among parameters make it difficult to understand and predict the sensitivity of the model to each parameter. Fortunately, sensitivity analysis is exceedingly simple with stochastic simulation. We simply simulate the environmental system according to our model and vary each of the input parameters over a range of values and compare the results. There are lots of options for conducting the sensitivity analysis. For example, we might vary input values for each parameter by $\pm 10\%$ of our original estimates, and might do so one parameter at a time holding all others constant at their nominal values or vary them all simultaneously by randomly sampling values from distributions. Alternatively, we might vary each parameter over the range of uncertainty in our original estimate to reflect that fact that the precision of our estimates might vary dramatically among parameters.

Stochastic Simulation... applications

Evaluate model sensitivity

1. Specify model
2. Simulate data set many times; each time vary input parameters (e.g., $cv=10\%$)
3. Quantify model outcome
4. Compute change in model outcome per unit change in each input parameter



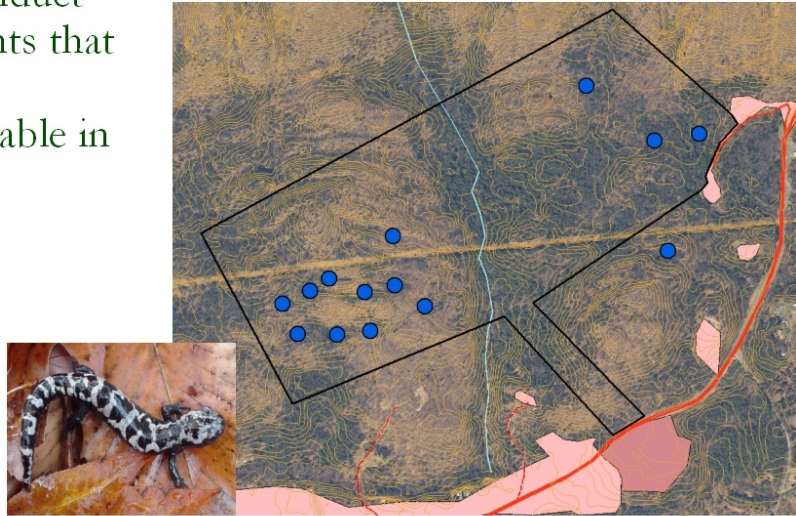
Example.—This example involves a spatially realistic population viability model for a metapopulation of marbled salamanders in western Massachusetts. Here, we varied each of the major model parameters by a fixed amount by sampling from a normal distributions with a mean set to our nominal estimate and a coefficient of variation (spread) equal to 10%. We ran 1,000 simulations, each time randomly generating parameter estimates from the above distributions, and computed the risk of metapopulation extinction (by running 100 simulations and determining how many times the metapopulation went extinct) under each set of parameter values. Lastly, using regression procedures (not to worry about the details here) we calculated the expected absolute change in the response variable per percentage change in the parameter. As shown in the figure here, our analysis reveals that extinction risk is most sensitive to adult survival (*adult.surv*), less but roughly equally sensitive to the breeding failure rate (*cat.rate.adj*), fecundity (*hp.fecundity.v.scale*), and juvenile survival (*j.to.a.surv*), and insensitive to juvenile dispersal rates (*disp.factor*) and correlation among sub-populations in fecundity (*fec.cor*). Moreover, these trends hold true across a broad range of nominal breeding failure rates (x-axis). The y-axis represents the expected change in the response per percent change in each parameter. Error bars indicate 95 % confidence intervals and are not shown within the grey region that represents points whose confidence intervals span zero (insensitive). These results suggest that if we wish to estimate extinction risk well, it is more important that we derive accurate and precise estimates of adult survival rate than juvenile dispersal rates.

Stochastic Simulation... applications

Conduct virtual experiments

- We can use stochastic simulation to conduct virtual experiments that are not practical, possible, or desirable in the real world
- What are some examples?

Marbled salamander metapopulation



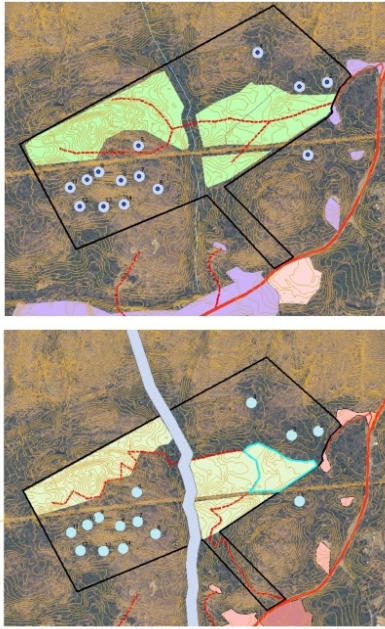
4.5 Conduct virtual experiments

We can use stochastic simulation to conduct virtual experiments that are not practical, possible, or desirable in the real world. For example, it would not be practical to conduct a field experiment to determine how forest management practices affect the wildfire disturbance regime of a forested landscape over an environmentally meaningful time scale, since the dynamic occurs over centuries not years. In addition, it would not be ethical to conduct a field experiment to determine the level of anthropogenic stress necessary to cause a species to go extinct. Simulation models allow us to conduct such studies in the virtual world instead of the real world. Can you think of other examples of where computer simulation is the highly preferred or even the only way to conduct the desired experiment?

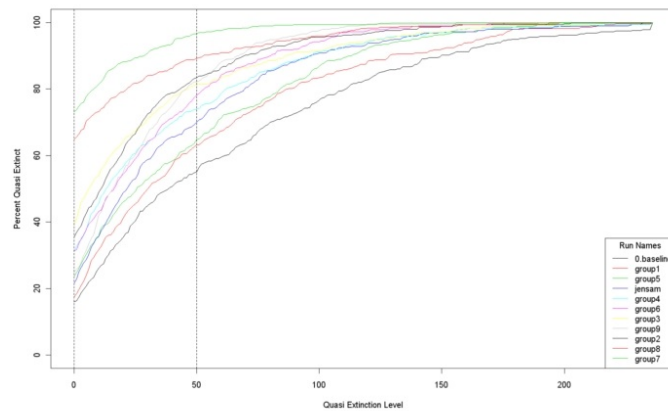
Example.—We continue with our spatially realistic metapopulation viability model. Here, we use our stochastic simulation model to evaluate the impacts of alternative forest management scenarios on metapopulation extinction risk – the probability that the metapopulation will go extinct over a 100-year period. Briefly, as part of a former class exercise, each group of forest managers was asked to devise a forest management plan to harvest timber while minimizing adverse impacts to the marbled salamander metapopulation, subject to certain objectives (e.g., extract a certain amount of timber) and constraints (e.g., minimize stream crossings).

Stochastic Simulation... applications

Conduct virtual experiments



Effects of forest harvesting scenarios on metapopulation quasi extinction



Some of the results are shown here. The figures on the left depict two different harvesting scenarios in which stands were partially cut for timber and roads were constructed for access. The altered land cover was input to the model and affected population vital rates such as adult upland survival, juvenile dispersal between ponds, etc. The figure on the right depicts the probability of quasi extinction under the various implemented scenarios. Quasi extinction is the probability that the metapopulation will fall below a threshold level of abundance sometime during the simulation period (100 years in this case). Thus, draw a vertical line up from any particular quasi extinction level on the x axis, say 50, until you intercept the curve. From that point draw a horizontal line to the y axis. The value of the y axis gives the probability that the metapopulation will fall below that quasi-extinction level during the next 100 years. The best scenario is one that minimizes quasi extinction risk.

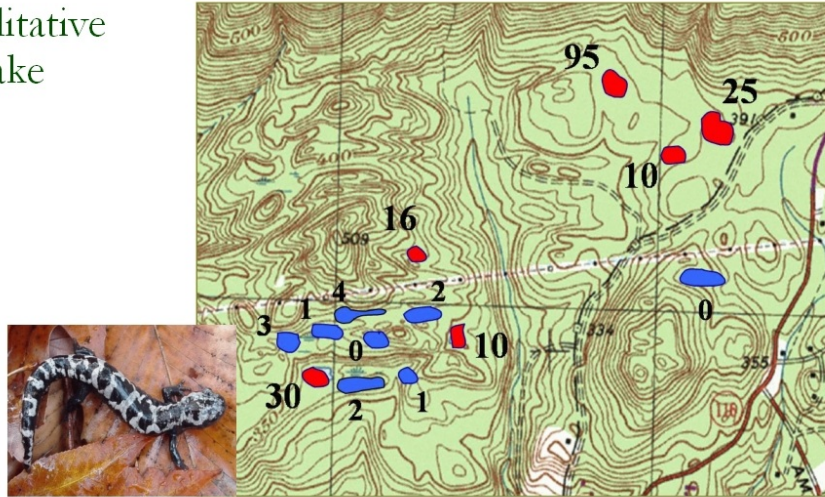
Clearly, there is no easy way to conduct an experiment like this in the real world, since we can only implement one scenario on the study site, and we would have to wait 100 years for the result. Stochastic simulation provides the opportunity to explore the possible outcomes of hundreds of scenarios without ever cutting one tree or building a single foot of new road. Of course, for the results to be trustworthy, we have to trust the model and its parameterization.

Stochastic Simulation... applications

Predict system behavior

- We can use stochastic simulation to understand qualitative patterns and make predictions
- What are some examples?

Marbled salamander metapopulation



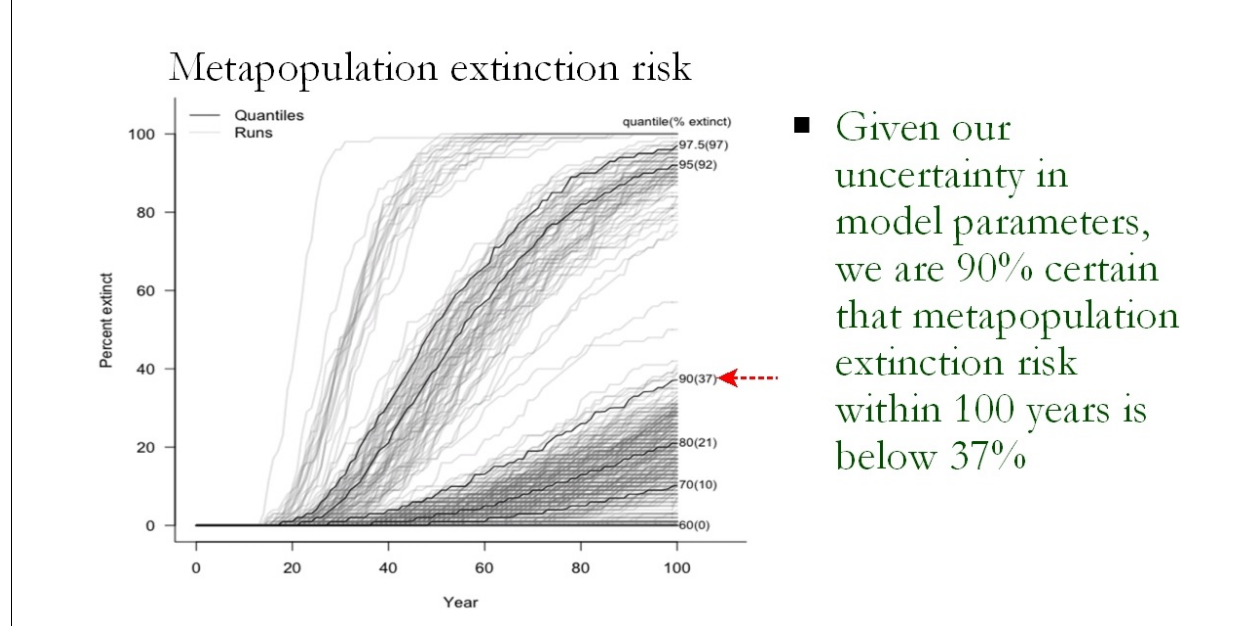
4.6 Predict system behavior

Lastly, we can use stochastic simulation to examine the behavior of our environmental system in order to understand qualitative patterns and/or make predictions. In some ways, this purpose subsumes all of the previous applications, since they too help us understand our model and our system and help us make predictions, e.g., that a sample size of n will give us the power we want. However, in some cases a stochastic simulation model is constructed for the primary purpose of making predictions of system behavior and quantifying the uncertainty in those predictions. In such cases, the uncertainty in our parameter estimates is used in the randomization process to produce uncertainty in our estimates of the system behavior or model output.

Example.—We continue with our spatially realistic metapopulation viability model. Here, we use our stochastic simulation model to predict the future fate of the metapopulation and to quantify the uncertainty in our prediction. Note, this is very similar to the previous example, where we evaluated alternative scenarios or model parameterizations, except here we are interested in a single scenario and the focus is on quantifying the model outcome in terms of uncertainty.

Stochastic Simulation... applications

Predict system behavior



In the example here, the data from the marbled salamander field study were combined with published studies to parameterize the spatially realistic metapopulation viability model. The figure shown here depicts the risk of metapopulation extinction under 1,000 different model parameterizations in which two key model parameters, reproductive failure rate and adult survival rate, were allowed to vary within the bounds of our uncertainty in these parameters. Based on these simulation results, we can be 90% certain that metapopulation extinction risk within 100 years is below 37%. If we find this level of uncertainty unacceptable, we can direct additional research to improve our estimates of reproductive failure rate and adult survival rate.