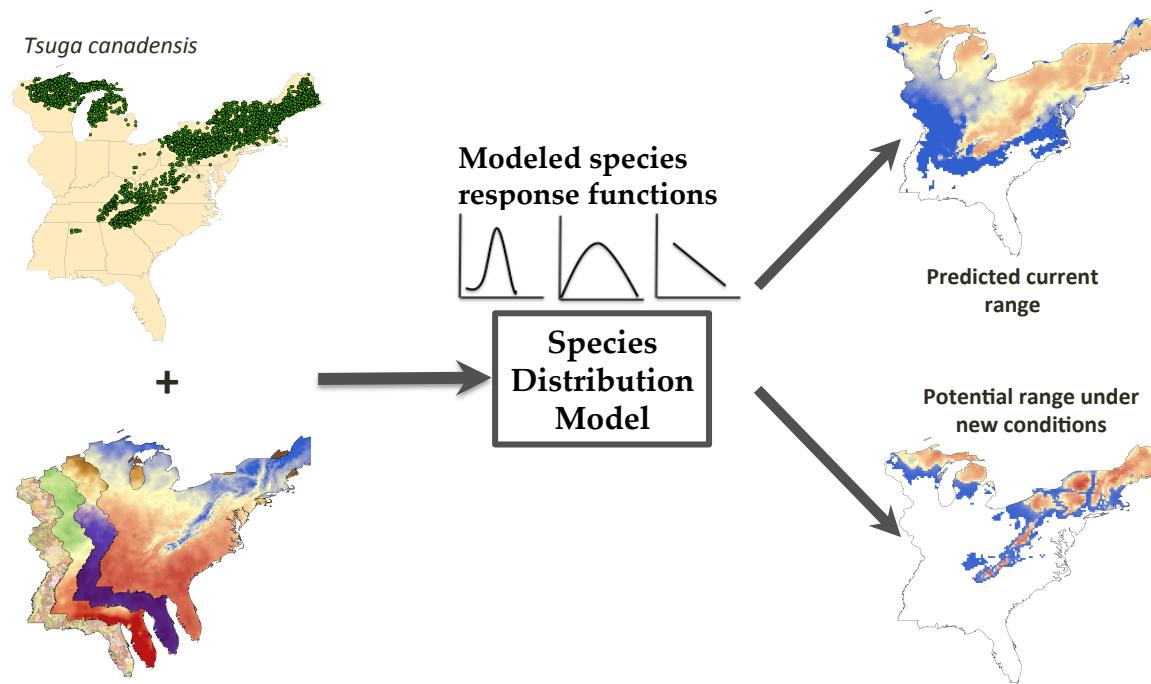
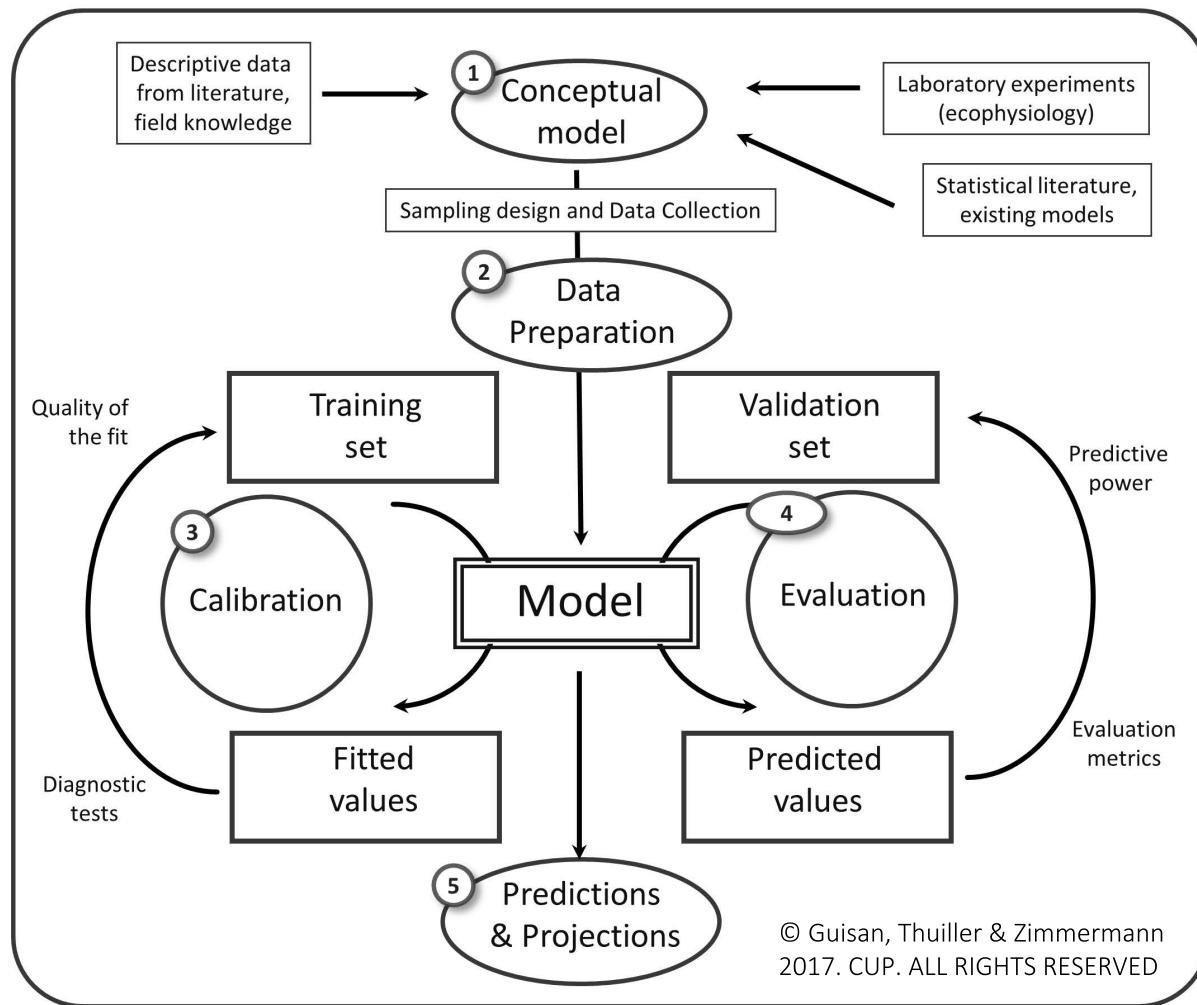


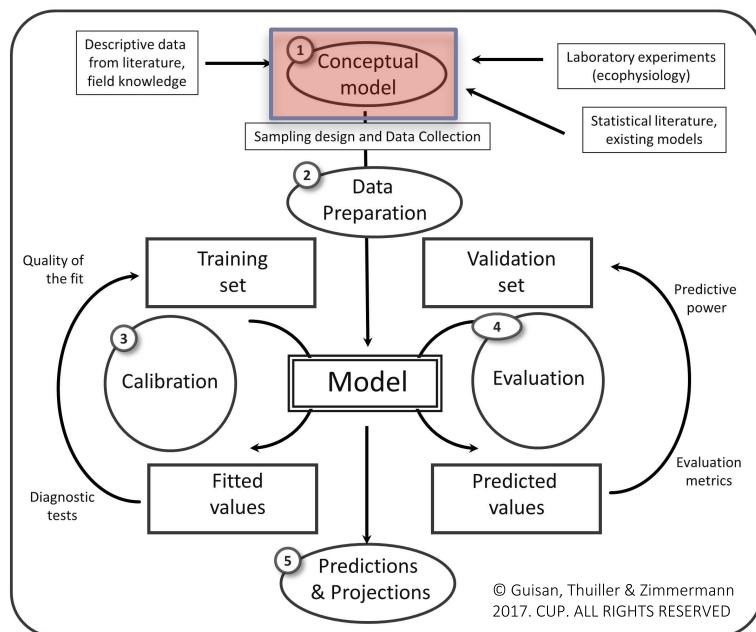
# The five steps of Species Distribution Modeling



# Five modeling steps

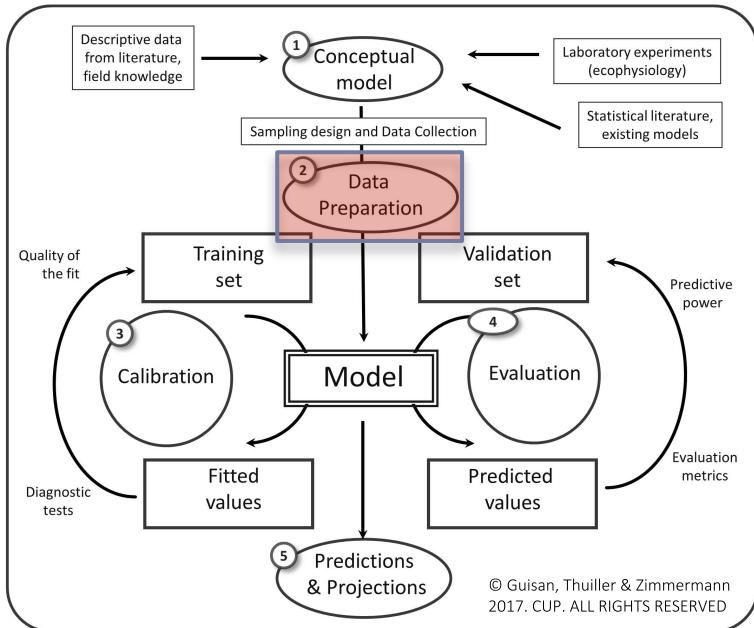


# Step 1: Conceptualization



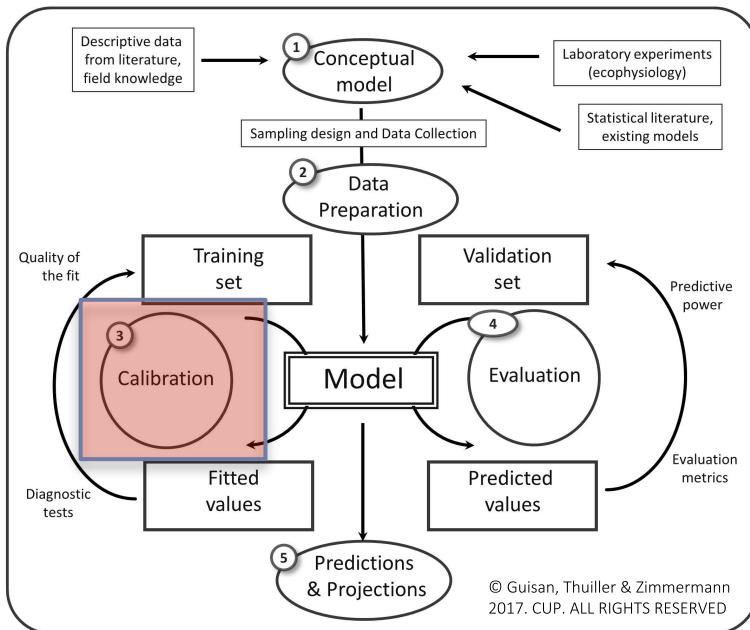
- Formulate main research objectives / study design
- What data are available?
- Check theory & important assumptions
- Things to consider:
  - Type of organism
  - Types of predictors
  - Model(s)

# Step 2: Data preparation



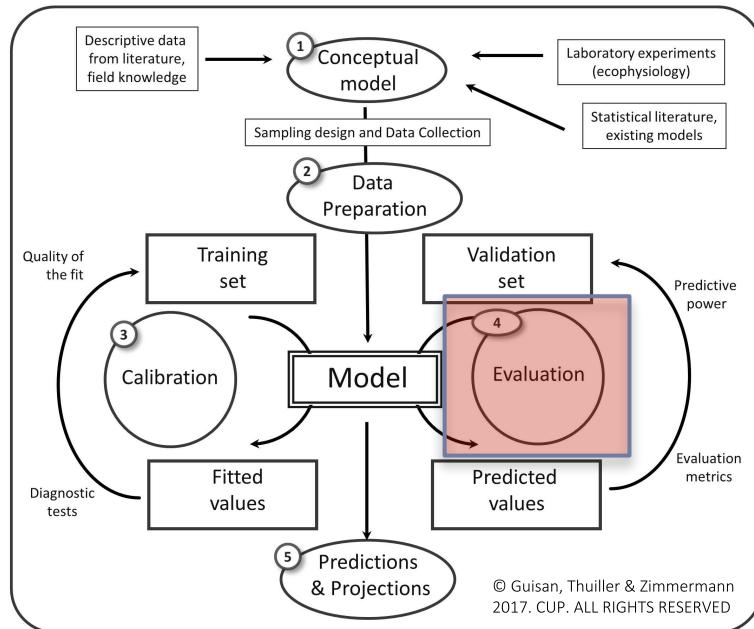
- Often takes longer than all other steps combined
- Gather and process biological and environmental data
  - p/a, p-o, current, future
- Check for scaling / temporal mismatches, etc
- Manipulate data as needed to ensure they are “stackable”

# Step 3: Model fitting



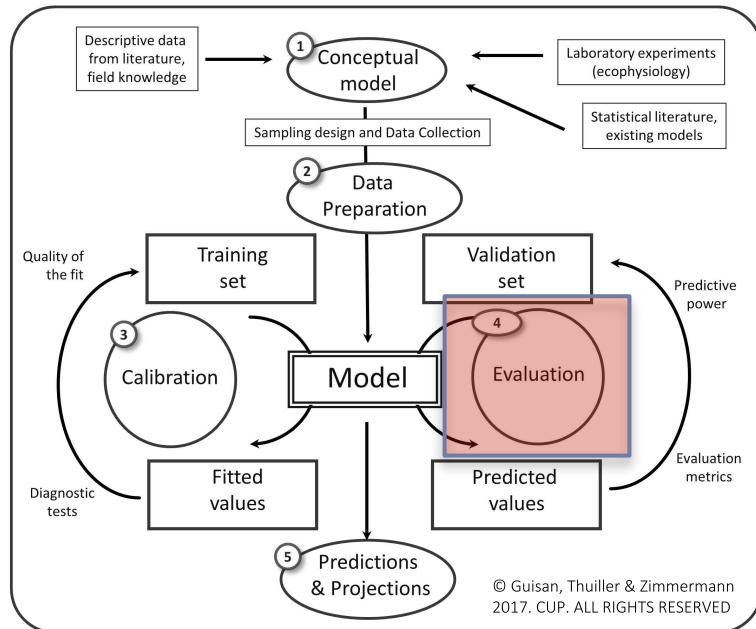
- Which algorithm(s) to use?
- How to deal with collinearity?
- Variable selection - how many and which ones?
- Which model settings?
- Test / correct for non-independence (spatial autocorrelation)
- Select final model (AIC, etc)

# Step 4: Model evaluation



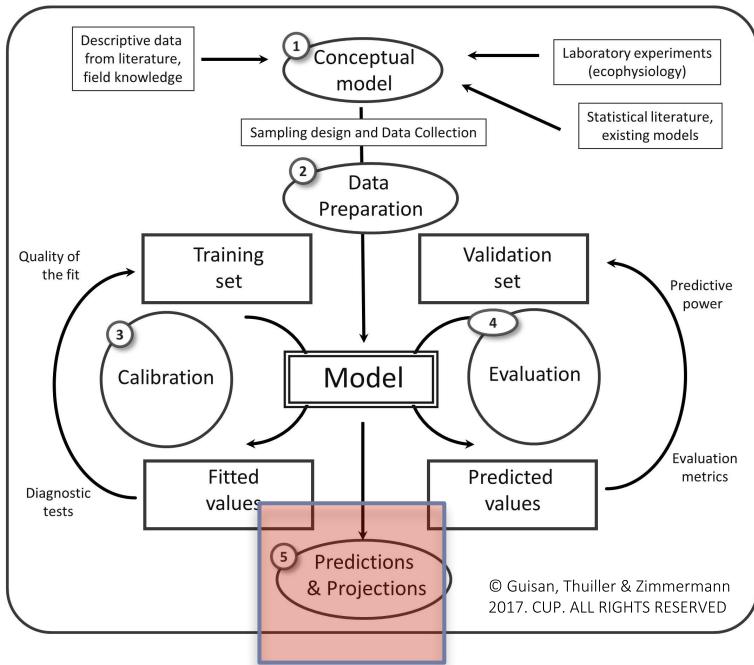
- Assess plausibility of:
  - fitted species-env relationships
  - Model coefficients
  - Variable importance

# Step 4: Model evaluation



- Assess predictive performance
  - Cross-validation, leave-one-out
  - Costs of over-prediction vs. under-prediction
  - Presence-only vs. p/a

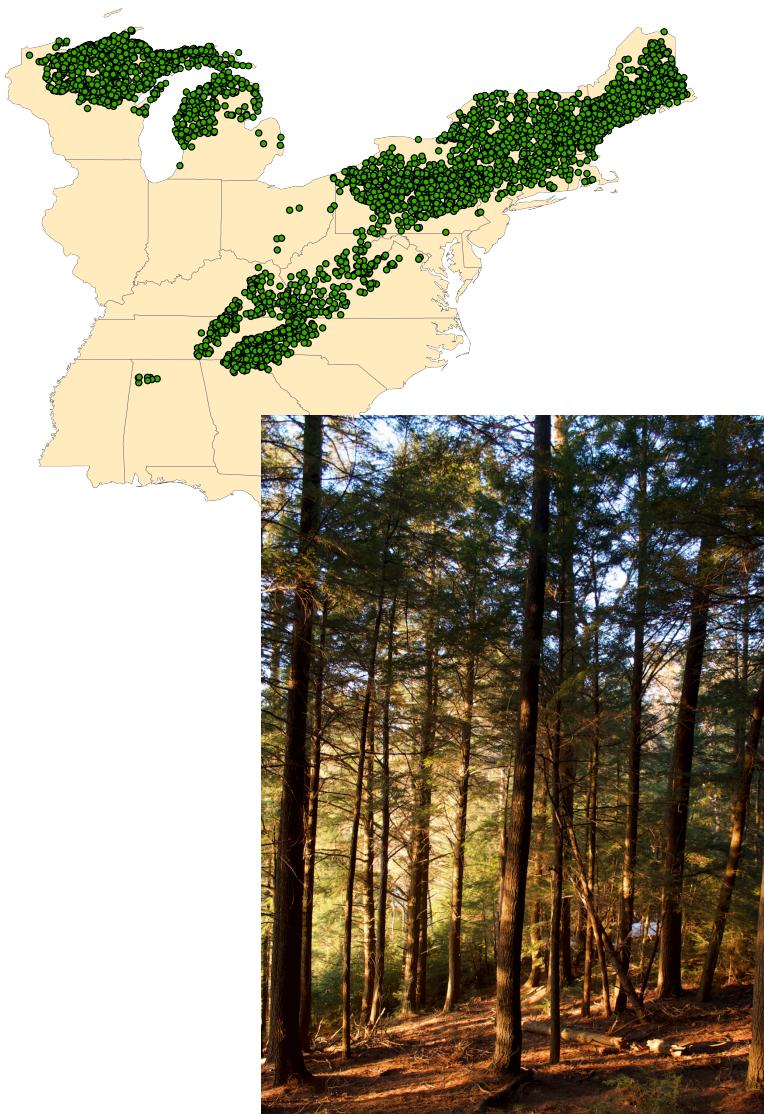
# Step 5: Model prediction / projection



- Quantification of uncertainty
  - Train / validation splits
  - Different algorithms
  - Model complexity
  - Different future scenarios
  - Novel environments

# SDM: Step 1

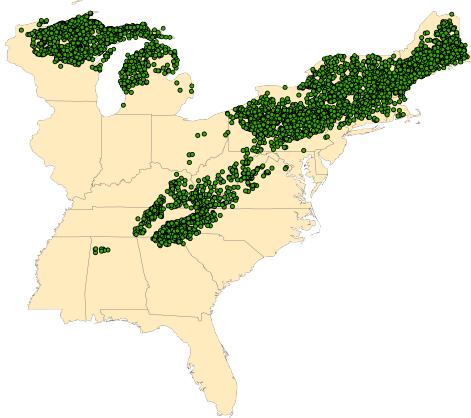
*Tsuga canadensis*



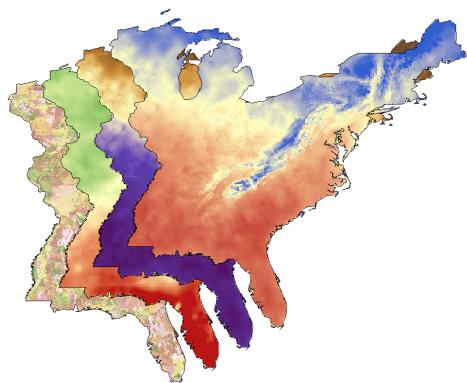
- How will climate change impact eastern hemlock?
- Late successional conifer
- Easy to identify
- Cool, damp environments
- Climate, soils, topography
- Presence-absence, but only in US

# SDM: Step 2

*Tsuga canadensis*

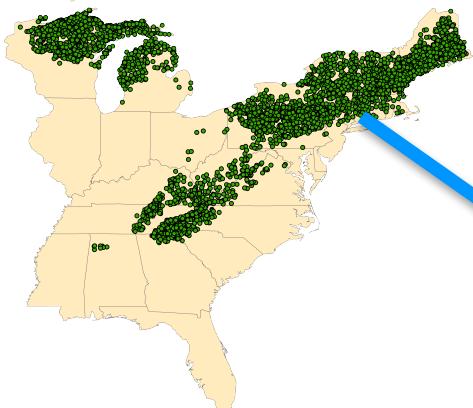


+

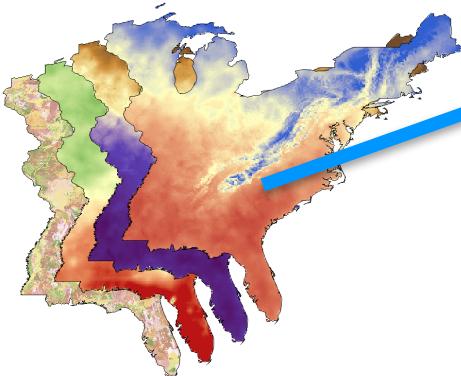


# SDM: Step 2

*Tsuga canadensis*



+



Biological survey data

	$Spp_1$
$Site_1$	0
$Site_2$	1
$Site_3$	1
...	...
$Site_j$	1

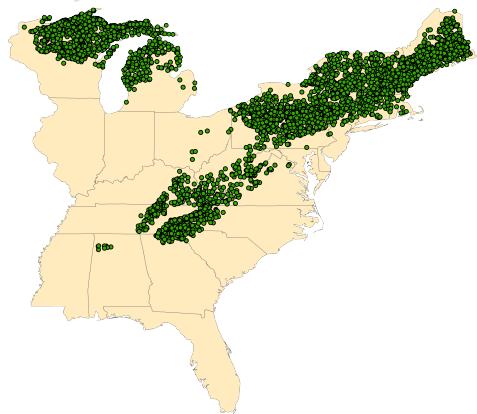
$$= f \{$$

Environmental covariates

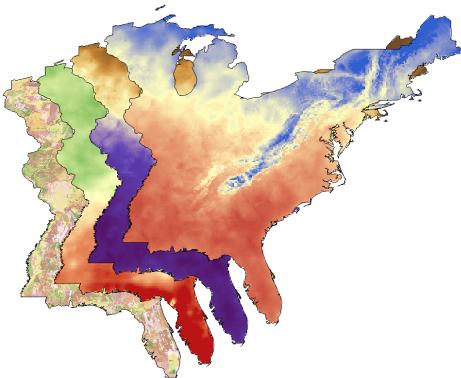
	$Env_1$	$Env_2$	$Env_3$	...	$Env_k$
$Site_1$	23.4	545.5	0.64	...	4.1
$Site_2$	22.1	89.0	0.22	...	8.0
$Site_3$	24.9	439.5	0.61	...	3.4
...	...	...	...	...	...
$Site_j$	25.3	321.7	0.88	...	3.9

# SDM: Step 2

*Tsuga canadensis*



+

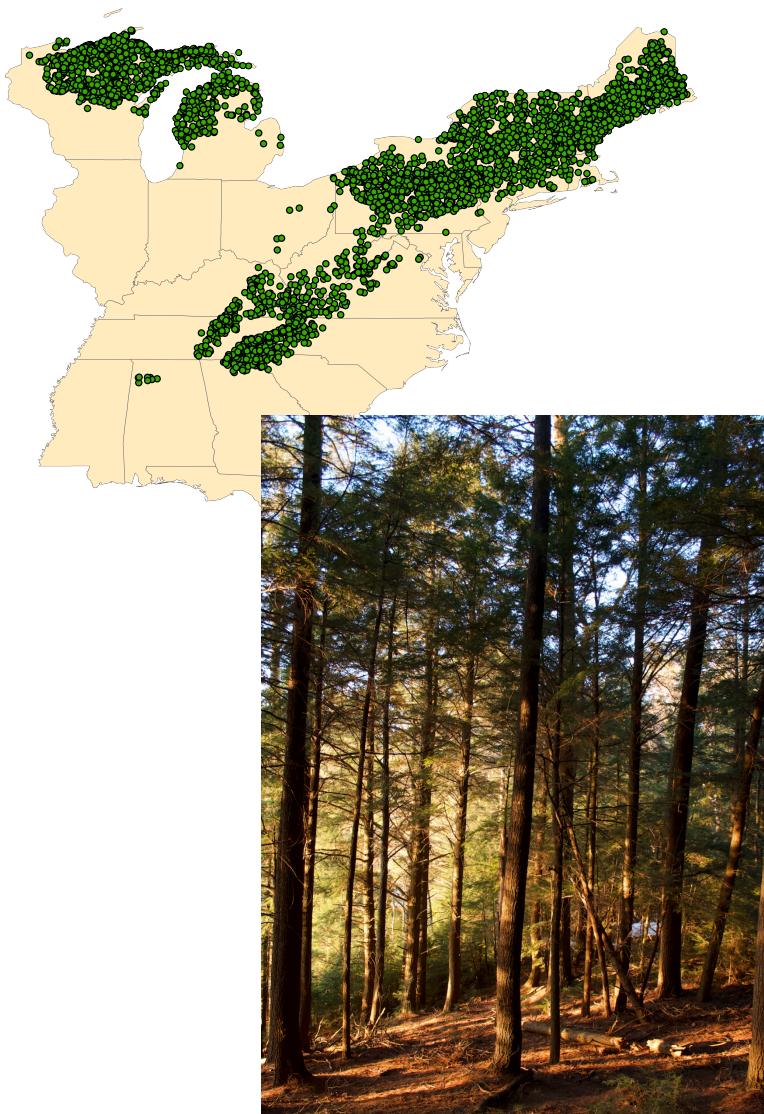


Response &  
predictors  
table

ID	Sp1	Tave	Prcp	Geol	...
1	0	3.25	1390	X	...
2	1	5.57	1840	b	...
3	1	7.21	2130	a	...
4	0	2.94	1420	Y	...
5	1	8.43	1789	a	...
...	...	...	...	...	...

# SDM: Step 3

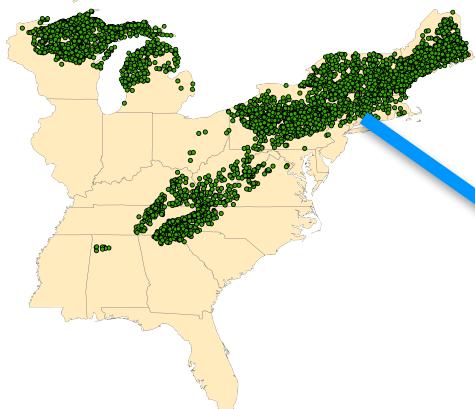
*Tsuga canadensis*



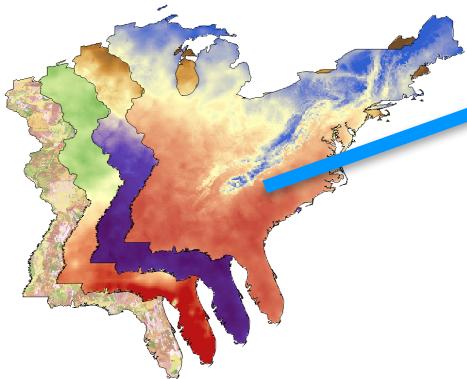
- How will climate change impact eastern hemlock?
- Late successional conifer
- Easy to identify
- Cool, damp environments
- Climate, soils, topography
- Presence-absence, but only in US

# Species distribution modeling

*Tsuga canadensis*



+



Biological  
survey data

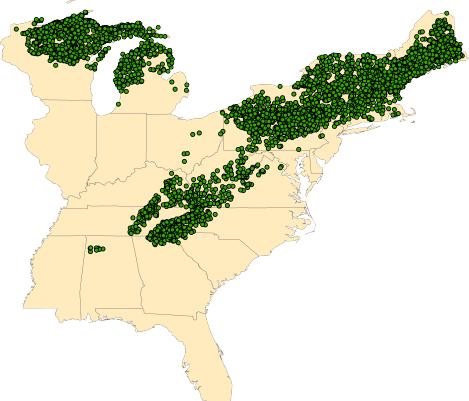
	$Spp_1$
$Site_1$	0
$Site_2$	1
$Site_3$	1
...	...
$Site_j$	1

$$\left[ \begin{array}{c} Spp_1 \\ Site_1 \\ Site_2 \\ Site_3 \\ \dots \\ Site_j \end{array} \right] = f \left\{ \left[ \begin{array}{cccccc} Env_1 & Env_2 & Env_3 & \dots & Env_k \\ Site_1 & 23.4 & 545.5 & 0.64 & \dots & 4.1 \\ Site_2 & 22.1 & 89.0 & 0.22 & \dots & 8.0 \\ Site_3 & 24.9 & 439.5 & 0.61 & \dots & 3.4 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 25.3 & 321.7 & 0.88 & \dots & 3.9 \end{array} \right] \right\}$$

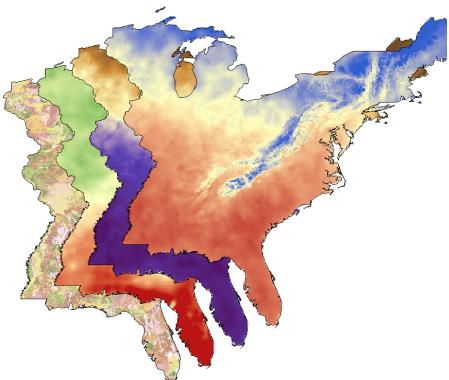
Environmental covariates

# SDM: Step 3

*Tsuga canadensis*



+



Modeled species  
response functions

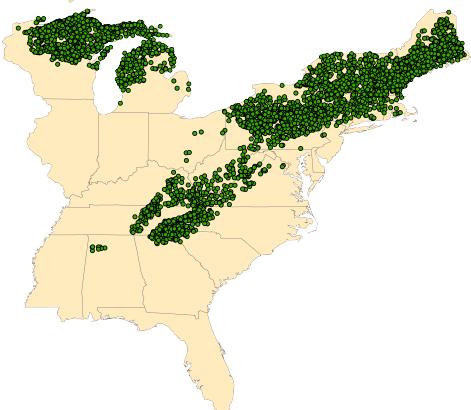


Species  
Distribution  
Model

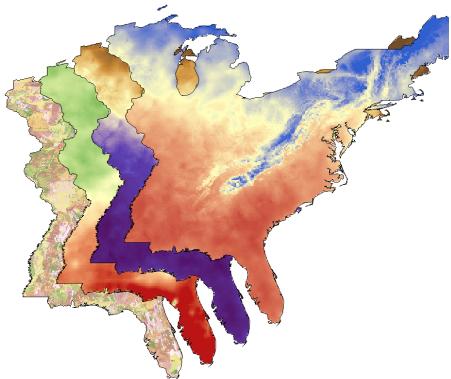
- Use p-a method
- Iterative process of checking collinearity and variable importance
- Lots of data, but want to avoid overfitting

# SDM: Step 4

*Tsuga canadensis*



+



Modeled species  
response functions



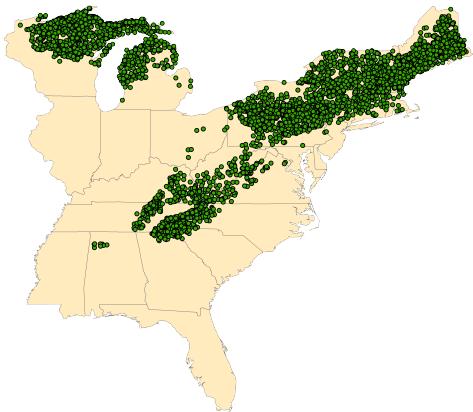
Species  
Distribution  
Model

Model evaluation

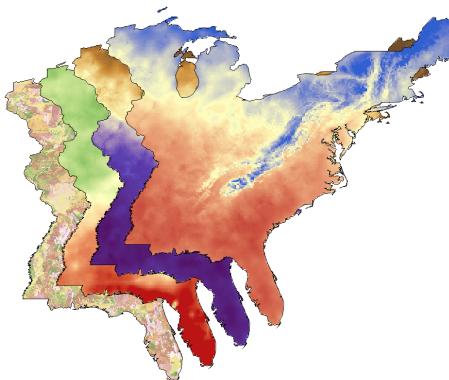
		Observed	
		p	a
Predicted	p		
	a		

# SDM: Step 5

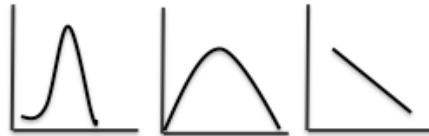
*Tsuga canadensis*



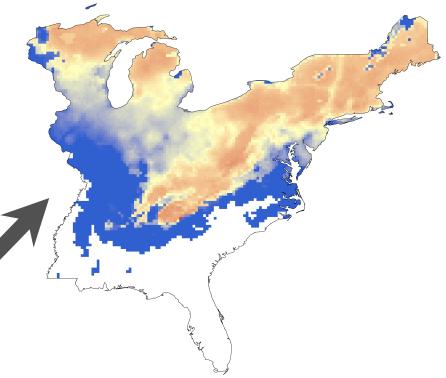
+



Modeled species  
response functions



Species  
Distribution  
Model



Predicted current  
range

# SDM: Step 5

