UNCERTAINTY AND STOCHASTICITY IN ECOLOGICAL MODELING

Certainty: An event is considered certain if it is 100% likely to happen.

Uncertainty: Anything that falls short of absolute certainty

COMMON FORMS OF UNCERTAINTY

Linguistic

Epistemic

Statistical uncertainty

Observational error

Structural uncertainty

Reducible

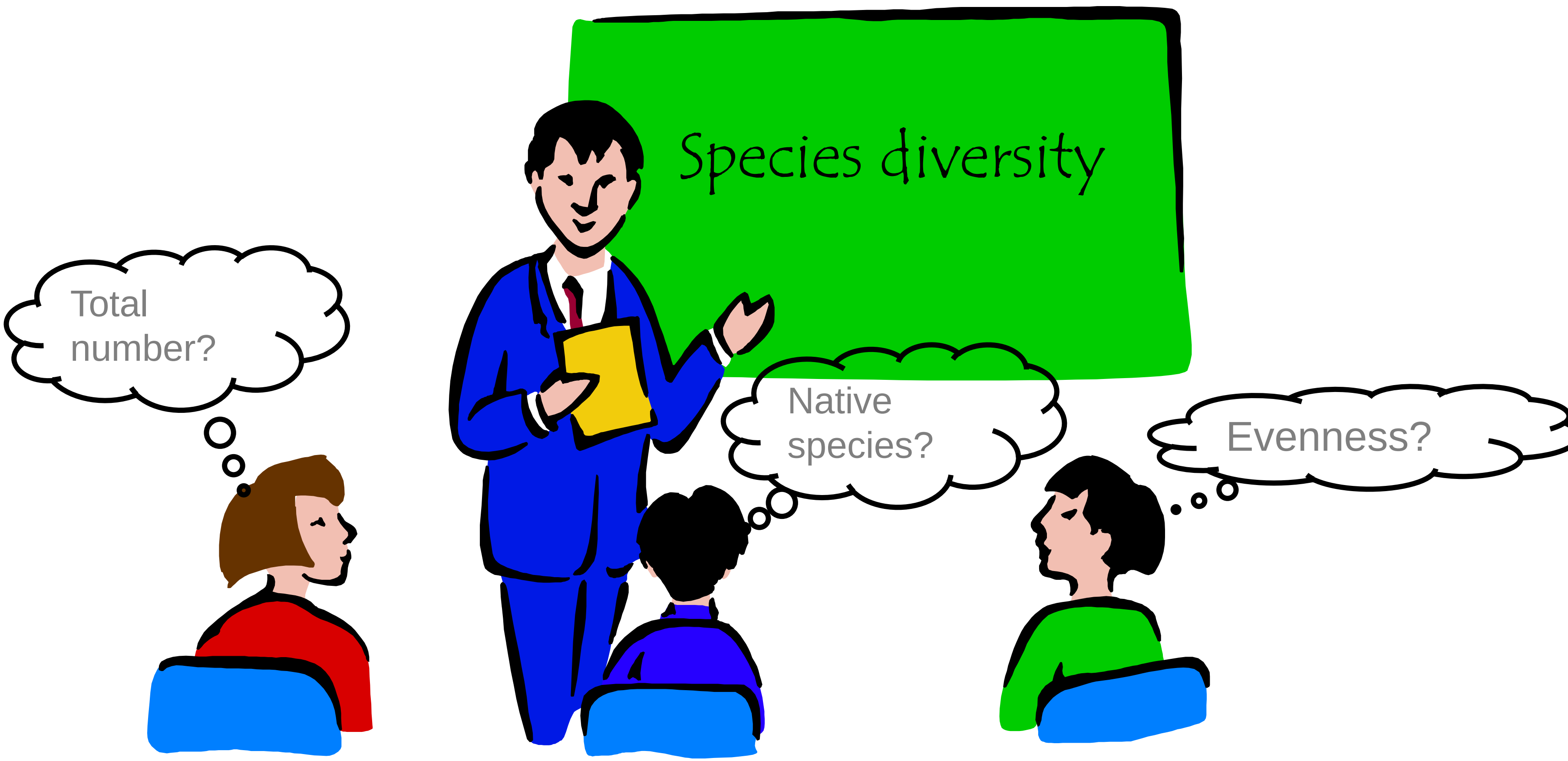
Aleatory

Environmental variability

Demographic variability

Irreducible

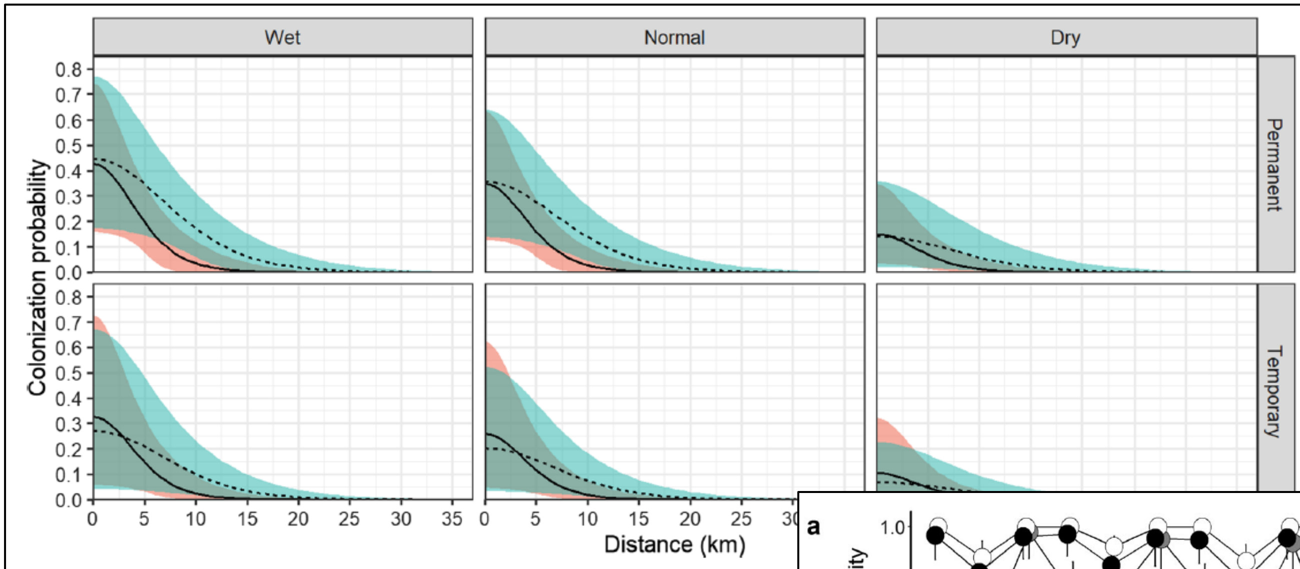
LINGUISTIC UNCERTAINTY

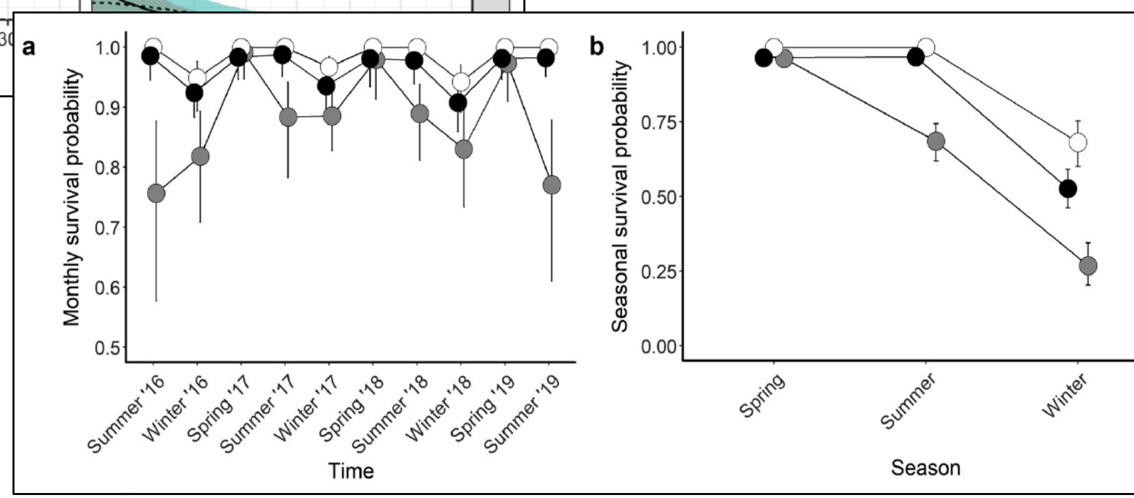


EPISTEMIC UNCERTAINTY

Statistical uncertainty

Due to the use of sample data to estimate parameters





EPISTEMIC UNCERTAINTY

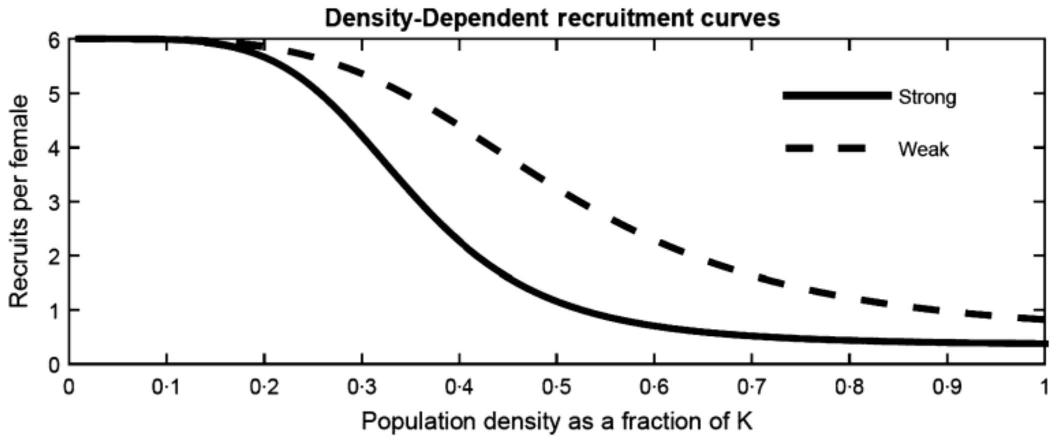
Observational uncertainty (Partial observablility)

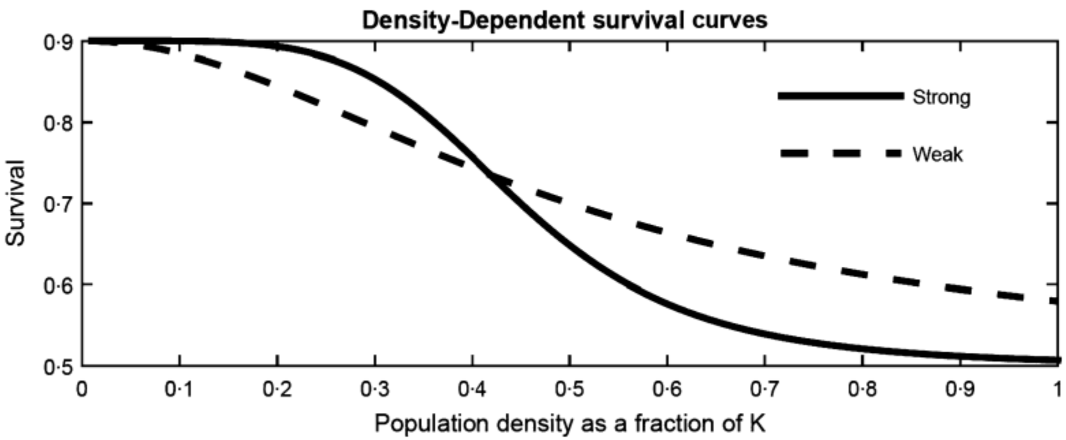
Inability to accurately assess the condition of the system

EPISTEMIC UNCERTAINTY

Structural (system) uncertainty

Incomplete understanding of system dynamics





ALEATORY UNCERTAINTY

Environmental/demographic stochasticity

ALEATORY UNCERTAINTY

Partial Controllability

The mismatch between the intention and the actual outcome

VARIABLES VS. RANDOM VARIABLES

Variable: A characteristic or feature that exhibits variability among units

Random variables: A variable in which the values come from a probability distribution

Stochasticity is incorporated into simulation models using random variables

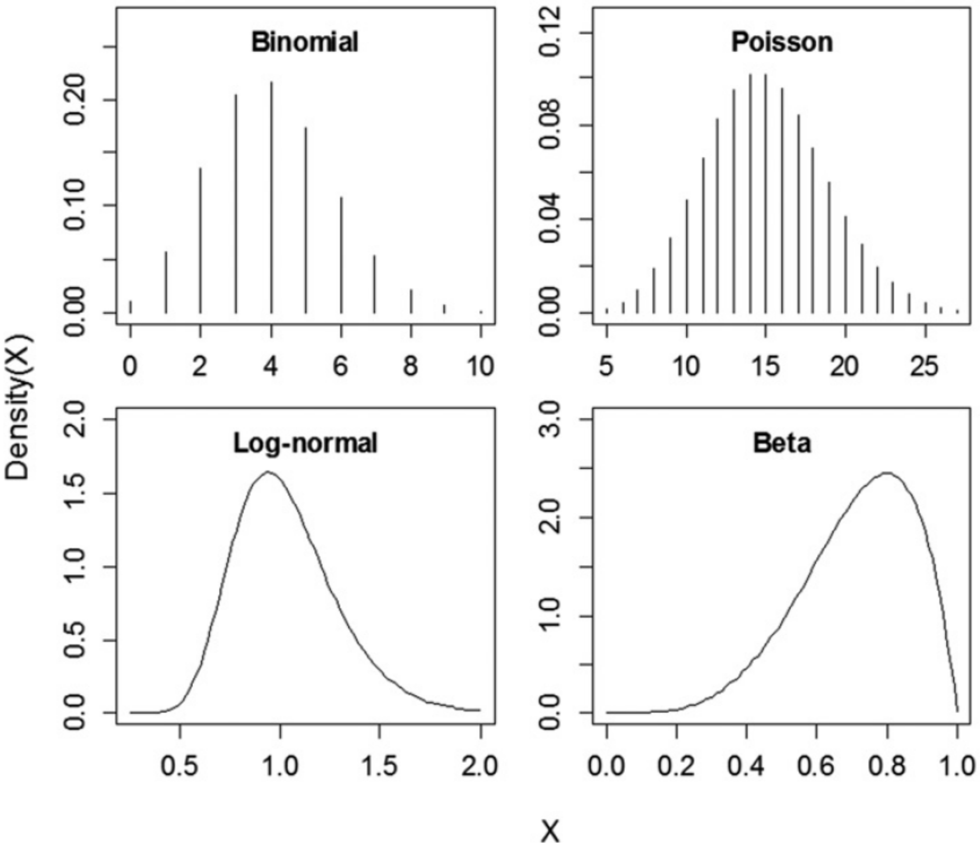
PROBABILITY DISTRIBUTION FUNCTIONS

Discrete

- The probability of result x from a random sample. For example, the probability that you obtained 5 heads (x=5) on 10 coin flips

Continuous

- The probability that the result x takes a value in a given interval. For example, the probability that the temperature next Wednesday will be between 5 and 10 ºC.



BASIC PROBABILITY – STATISTICAL INDEPENDENCE

If two events are independent:

Let the probability of catching a fish on a fishing trip = P(fish)

Let the probability of seeing another person fishing on a fishing trip = P(person)

Description	Probability statement
Probability of catching a fish and seeing another person fishing	P(fish)×P(person)
Probability of not catching a fish and seeing another person fishing	[1 – P(fish)]×P(person)
Probability of catching a fish and not seeing another person fishing	P(fish)×[1 – P(person)]
Probability of not catching a fish and not seeing another person fishing	[1 – P(fish)]×[1-P(person)]

BASIC PROBABILITY – TOTAL PROBABILITY

<u>Law of total probability</u> specifies that the overall probability of an event of interest occurring is the weighted sum of the conditional probabilities of the event, with the weights determined by the probabilities of the event that the event of interest is conditioned on

BASIC PROBABILITY – TOTAL PROBABILITY

Probability of catching a fish = P(fish) = 0.15

Probability of seeing another person fishing = P(person) = 0.90

Description	Probability statement	Probability
Probability of catching a fish and seeing another person fishing	P(fish)×P(person)	0.135
Probability of not catching a fish and seeing another person fishing	[1 – P(fish)]×P(person)	0.765
Probability of catching a fish and not seeing another person fishing	P(fish)×[1 – P(person)]	0.015
Probability of not catching a fish and not seeing another person fishing	[1 – P(fish)]×[1-P(person)]	0.085

Total sum = 1.000

Probability of at least one event occurring:

<u>P(fish)×P(person)</u> + <u>[1 – P(fish)]×P(person)</u> + <u>P(fish)×[1 – P(person)]</u>

= 1 – [1 – P(fish)]×[1-P(person)] = 0.915

BASIC PROBABILITY – CONDITIONAL DEPENDENCE

Description	Probability statement
Probability salmon are running	P(run)
Probability you catch a salmon given they are running	P(catch | run)
Probability salmon are running and you catch a salmon	P(run)×P(catch | run)

Description	Probability statement
Salmon are running and you missed	P(run)×[1 – P(catch | run)]
Salmon are not running (and you didn't catch one)	[1 – P(run)]

Total probability of the event not catching a salmon:

P(run)×[1 – P(catch | run)] + [1 – P(run)]

BRIEF REVIEW OF LINEAR MODELS

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_z x_{z,i} + \varepsilon$$

- Y is the response (a.k.a. dependent) variable
- x_1 and x_Z are the values of the explanatory (a.k.a. predictor or independent) variables for observation i
- β_0 is the intercept
- β_1 and β_z are the model coefficients (a.k.a. slope parameters)
- ε is the residual error

BRIEF REVIEW OF LINEAR MODELS

$$\text{Water Temp} = 0.5 + 0.02 \times \text{Air Temp} + 12.2 \times \text{Discharge} + 0.2 \times \text{Site}_{\text{Watershed\_B}}$$

Which probability distribution make sense?

BRIEF REVIEW OF LOGIT-LINEAR MODELS

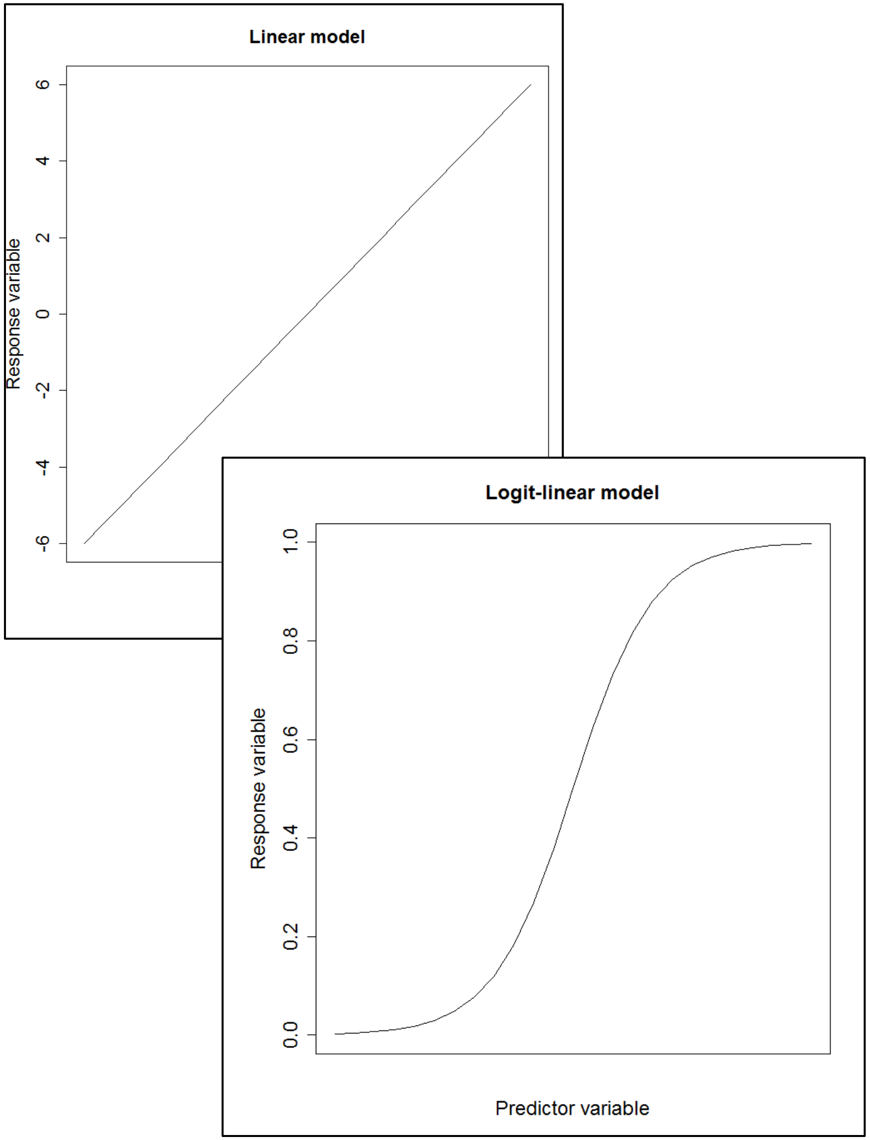
$$\eta = ln\left(\frac{p}{1-p}\right)$$ ← **Logit link**

- η is the log odds
- p is the probability of an event

Inverse logit

$$p = \frac{1}{1+\exp(-\eta)}$$

- η is the log odds
- p is the probability of an event



BRIEF REVIEW OF LOGIT-LINEAR MODELS

$$\eta_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_z x_{z,i}$$

- η is the log odds
- β_0 is the intercept on a logit scale
- β_1 and β_z are the model coefficients (a.k.a. slope parameters) on a logit scale

BRIEF REVIEW OF LOGIT-LINEAR MODELS

$$ln\left(\frac{survival}{1-survival}\right) = \eta = 0.12 + 0.025 \times \text{Body Size} - 1.2 \ \times \text{Sex}_{\text{Male}}$$

BRIEF REVIEW OF LOGIT-LINEAR MODELS

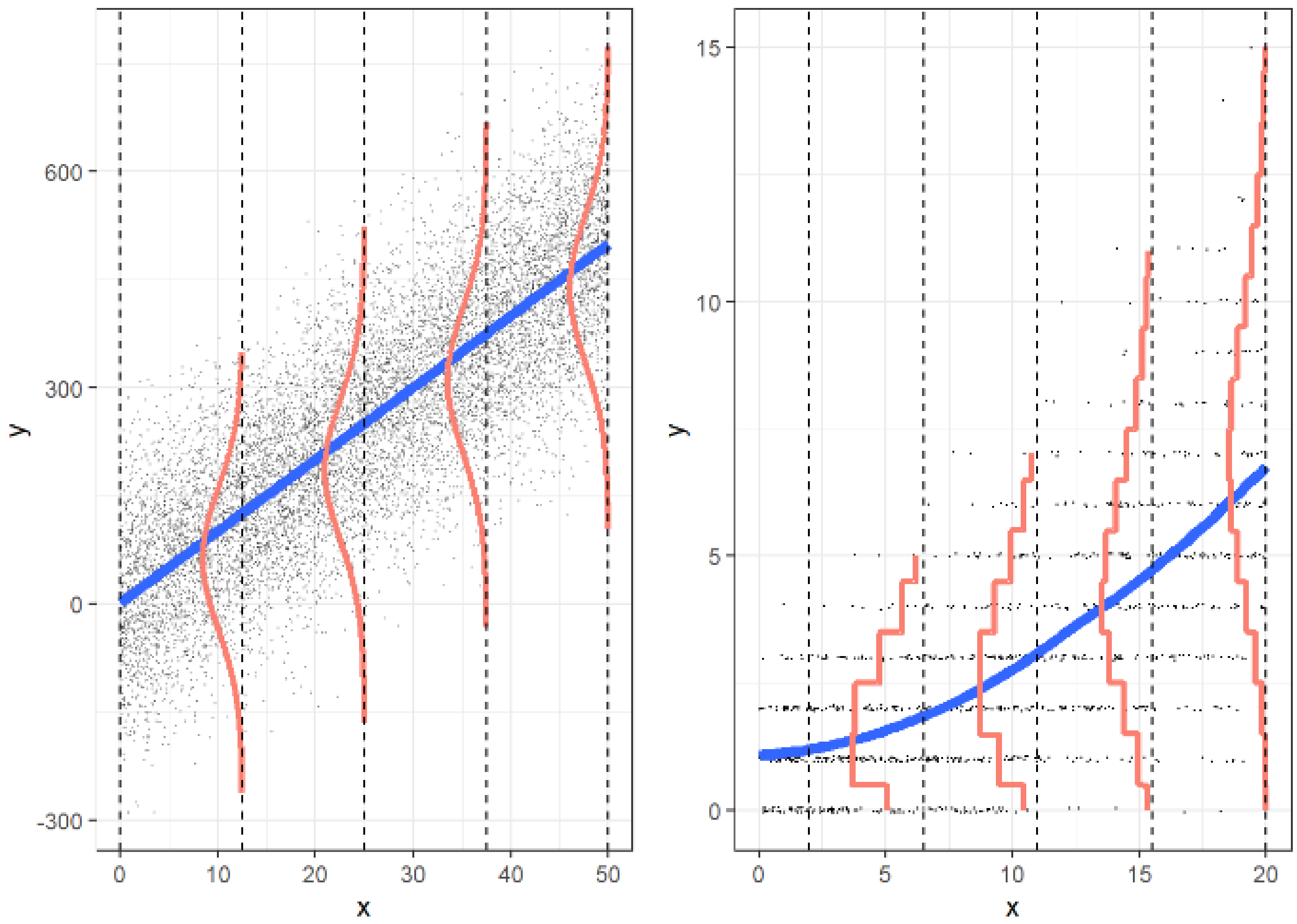
$$ln\left(\frac{survival}{1-survival}\right) = \eta = 0.12 + 0.025 \times \text{Body Size} - 1.2 \ \times \text{Sex}_{\text{Male}}$$

$$\text{surv}ival = \frac{1}{1 + \exp(-\eta)}$$

Body Size	Sex_{Male}	η	survival
40	1	-0.08	0.480
45	1	0.05	0.511
50	1	0.17	0.542
55	1	0.30	0.573
60	1	0.42	0.603
40	0	1.12	0.754
45	0	1.25	0.776
50	0	1.37	0.797
55	0	1.50	0.817
60	0	1.62	0.835

Which probability distribution make sense?

BRIEF REVIEW OF LOG-LINEAR MODELS



BRIEF REVIEW OF LOG-LINEAR MODELS

$$\ln(\lambda_i) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_z x_{z,i}$$

- $\ln(\lambda_i) =$ is the log of the rate parameter
- β_0 is the intercept on a log scale
- β_1 and β_z are the model coefficients (a.k.a. slope parameters) on a log scale

This is equivalent to:

$$\lambda_i = e^{\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_z x_{z,i}}$$

Which probability distribution make sense?

REPRESENTING EPISTEMIC UNCERTAINTY

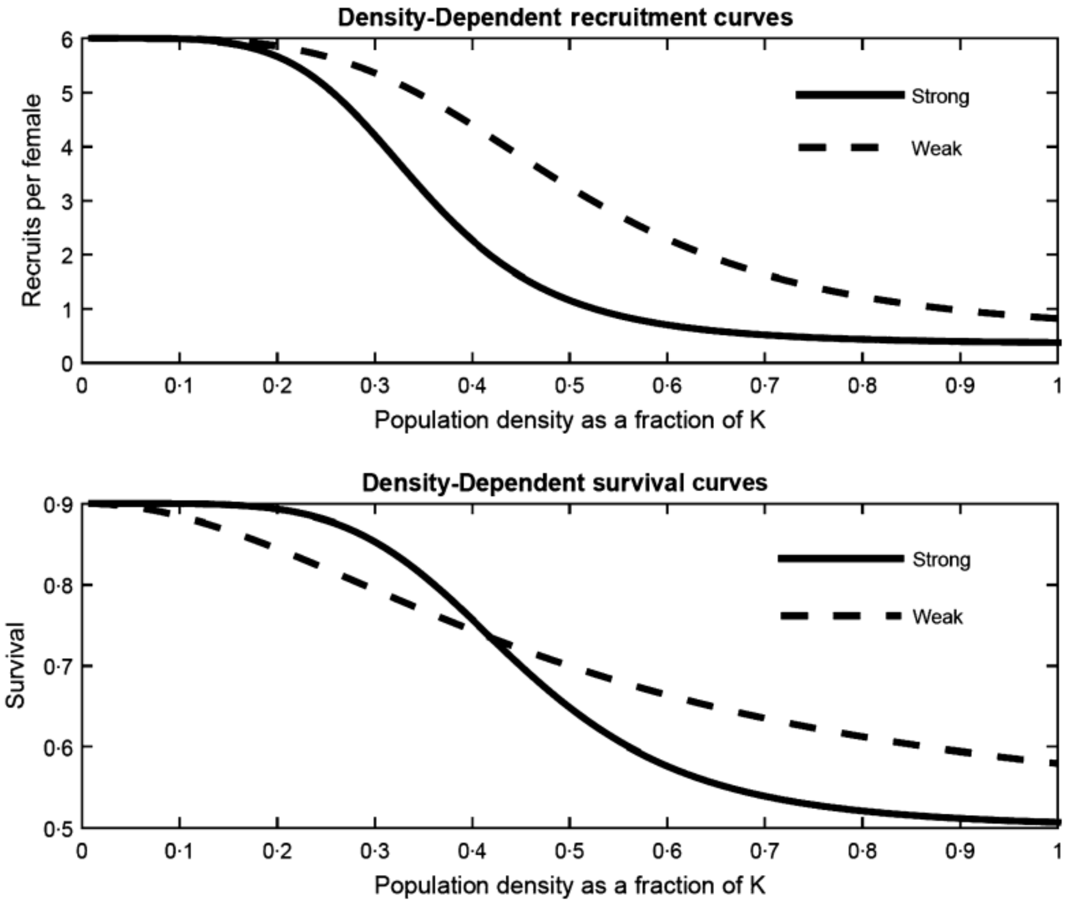
Statistical uncertainty

Table 2
Parameter estimates, SE, and upper and lower 95% confidence intervals from regression models that estimate downstream environmental conditions. CVP = Central Valley Project and SWP = State Water Project.

Estimates	Estimate	SE	Lower 95%	Upper 95%
Flow at Stockton (R^2 = 0.9132; residual error = 20.83)				
Intercept	806.54	15.46	776.23	836.86
log(Flow at Freeport)	-157.01	2.93	-162.76	-151.27
log(Flow at Vernalis)	-205.33	3.30	-211.79	-198.87
log(Minimum tide + 1)	131.36	17.42	97.20	165.52
sqrt(SWP exports)	4.79	0.46	3.90	5.69
CVP exports	-0.316	0.043	-0.40	-0.23
log(Flow at Vernalis) * log(Flow at Freeport)	40.66	0.60	39.49	41.83
log(Flow at Vernalis) * log (minimum tide + 1)	-32.10	4.39	-40.71	-23.50
log(Flow at Vernalis) * sqrt (SWP_exports)	-1.09	0.11	-1.31	-0.87
log(Flow at Vernalis) * log (CVP_exports)	0.091	0.011	0.068	0.11

REPRESENTING EPISTEMIC UNCERTAINTY

Structural (system) uncertainty



What should we do?

SCENARIO DEVELOPMENT

Table 1
Productivity (F), juvenile survival (ϕ_J), and adult survival (ϕ_A) estimates and their associated temporal process variances (σ^2) used to simulate golden-cheeked warbler (*Setophaga chrysoparia*) population dynamics under different scenarios.

	F		ϕ_J		ϕ_A	
Scenario	$\bar{X}$	σ^2	$\bar{X}$	σ^2	$\bar{X}$	σ^2
I	1.42	0.2415	0.28	0.0076	0.47	0.0120
II	2.52	0.2415	0.28	0.0076	0.52	0.0120
III	3.60	0.2415	0.28	0.0076	0.57	0.0120

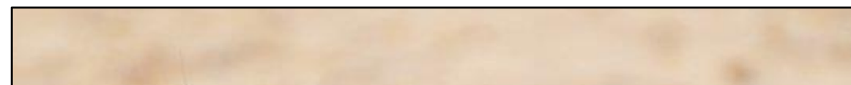
AKAIKE'S INFORMATION CRITERION (AIC)

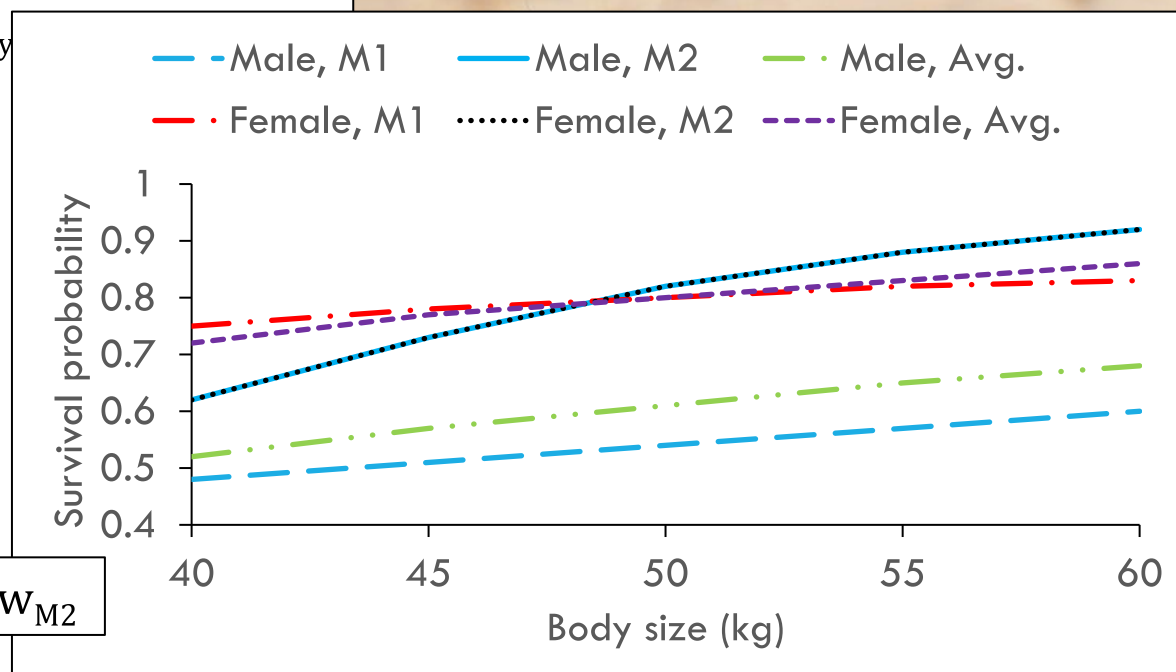
- Represents tradeoff between model fit and parsimony
- AIC by itself is relatively meaningless, buy we find the "best" model by
  - Delta AIC
  - AIC weight
    - Ranges 0–1, 1 = best model
    - Sums to 1
    - Interpreted as the relative likelihood that model is best, <u>given the data and the other models in the set</u>

AKAIKE'S INFORMATION CRITERION (AIC)

Model averaging:

- Model 1
  - logit(survival)=0.12 + 0.025 × Body Size − 1.2 × Sex_{Male}
  - AIC weight = 0.75
- Model2
  - logit(survival)= -3.5 + 0.1 × Body
  - AIC weight = 0.25



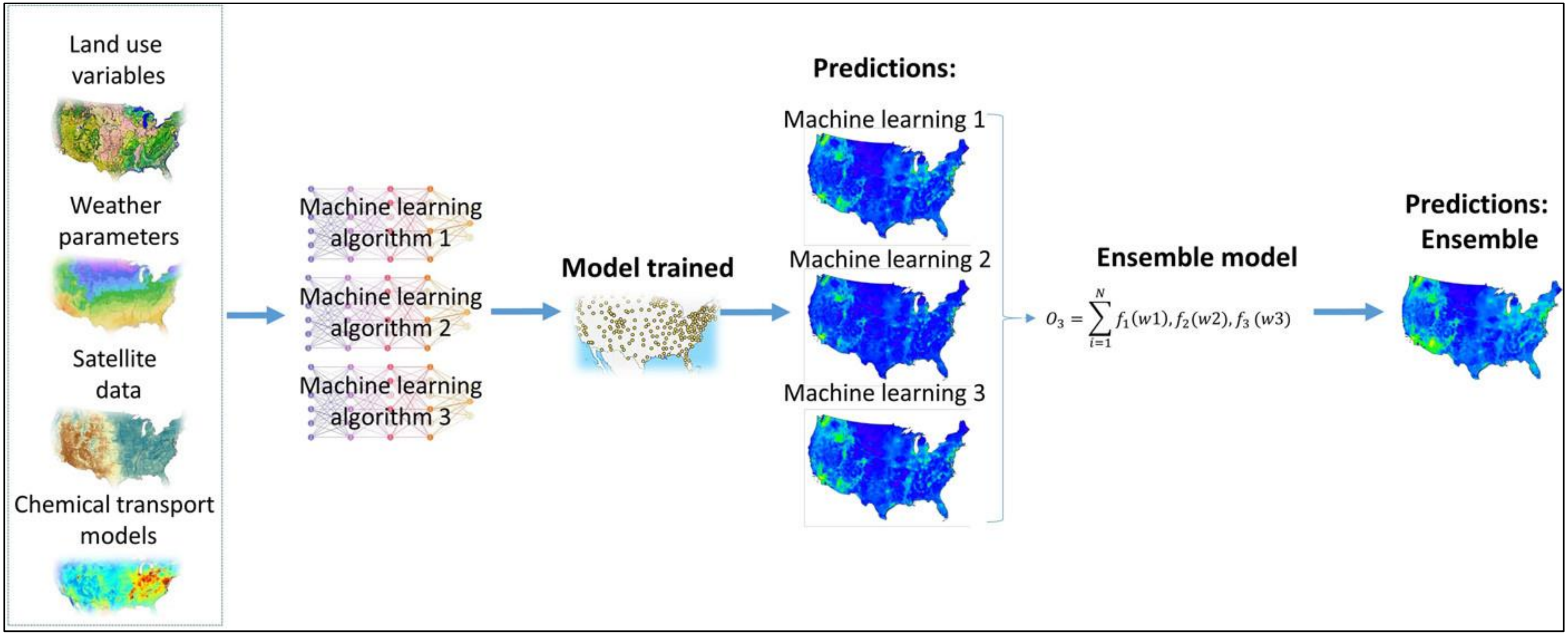


$$\text{Surv.}_{\text{Avg.}} = \text{Surv.}_{\text{M1}} \times w_{\text{M1}} + \text{Surv.}_{\text{M2}} \times w_{\text{M2}}$$

ENSEMBLE MODELS

Weighting

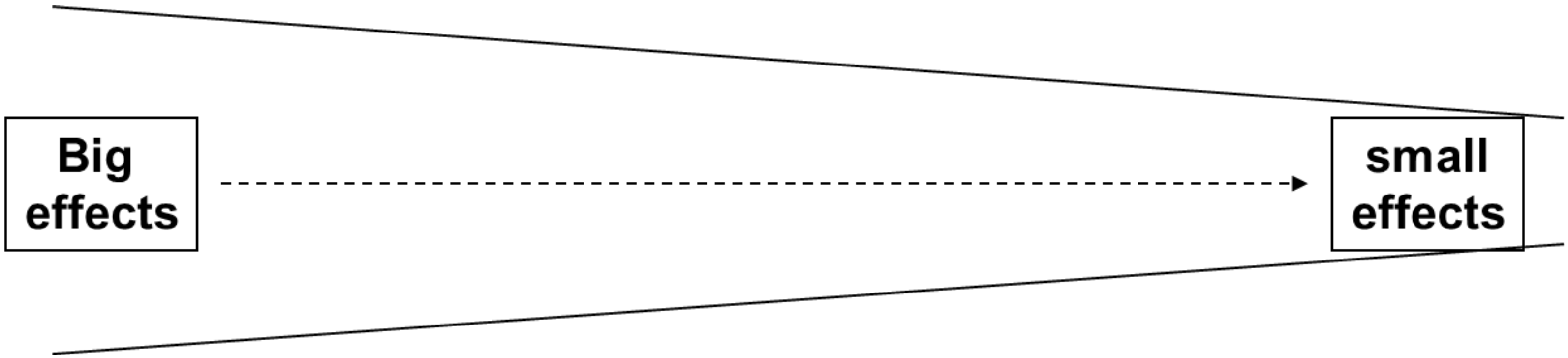
- Equally
- Relative belief
- Predictive accuracy



TAPERING EFFECT SIZES

These effects might be sequentially revealed as sample size increases because information content increases

- Rare events yet are more difficult to study (e.g., fire, flood, etc.)
- Fine scale events are also difficult to study (e.g., bioenergetics, individual choices, etc.)



WHAT TO DO WHEN DATA ARE LIMITED

1. Use estimate from the literature
2. Combine estimates across existing studies (meta-analysis)
3. Borrow information from other closely related system
4. Expert elicitation or hypothesized relationship
5. Model calibration

META-ANALYSIS

Combines results of previous research to provide estimates to parameterize models

Pros

- Uses existing information
- Provides estimates that are more defensible and empirically grounded than expert judgment

Cons

- No clear way of accounting for qualitative differences between/among studies
- “Comparable” studies are often hard to find
- Selection bias - “bad” outcomes, e.g., non significant results often do not get reported or published

EXPERT ELICITATION OR HYPOTHESIZED RELATIONSHIPS

Captures information within mental models of experts

Pros

- Fast
- Can capture hypothesized fine scale processes

Cons

- Not a substitute for empirical data… <u>but experts know their system and expert driven models are the baseline</u>

CALIBRATION

Makes model run match observed data

Pros

- Can be used to find missing parameters
- Stakeholder groups like things to be “right”

Cons

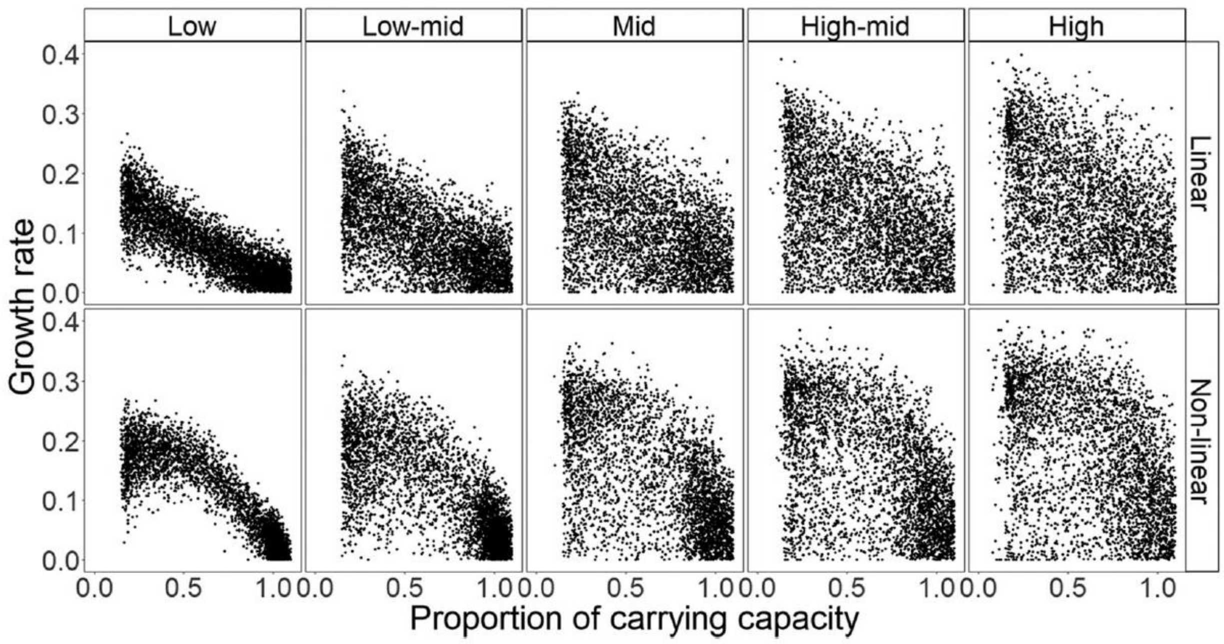
- Requires quality data to match
- Can lead to overfitting and missing ecological processes

SIMULATION MODELS

Remember:

- Keep it simple, stick to the important factors
- Think about the process, visualize it
- Draw a conceptual model (diagram)
- Consider the nature of the relationships
- Choose the appropriate functions and statistical distributions

"uncertainty begets uncertainty"



SENSITIVITY ANALYSIS

Definition: the change that occurs in model output, given a small change in some model parameter or other input

Evaluate relative influence of model parameters and inputs

Identify most influential

Guide model simplification

All models should be evaluated with sensitivity analysis

SENSITIVITY ANALYSIS

Basic idea: Vary the values of each parameter and examine the effect on the predicted values

Several types of sensitivity analysis:

1. One-way sensitivity analysis
2. Two-way sensitivity analysis
3. Response profile
4. Indifference curves
5. And many more…

There is no single best method of examining model sensitivity

AND

You can report on multiple types

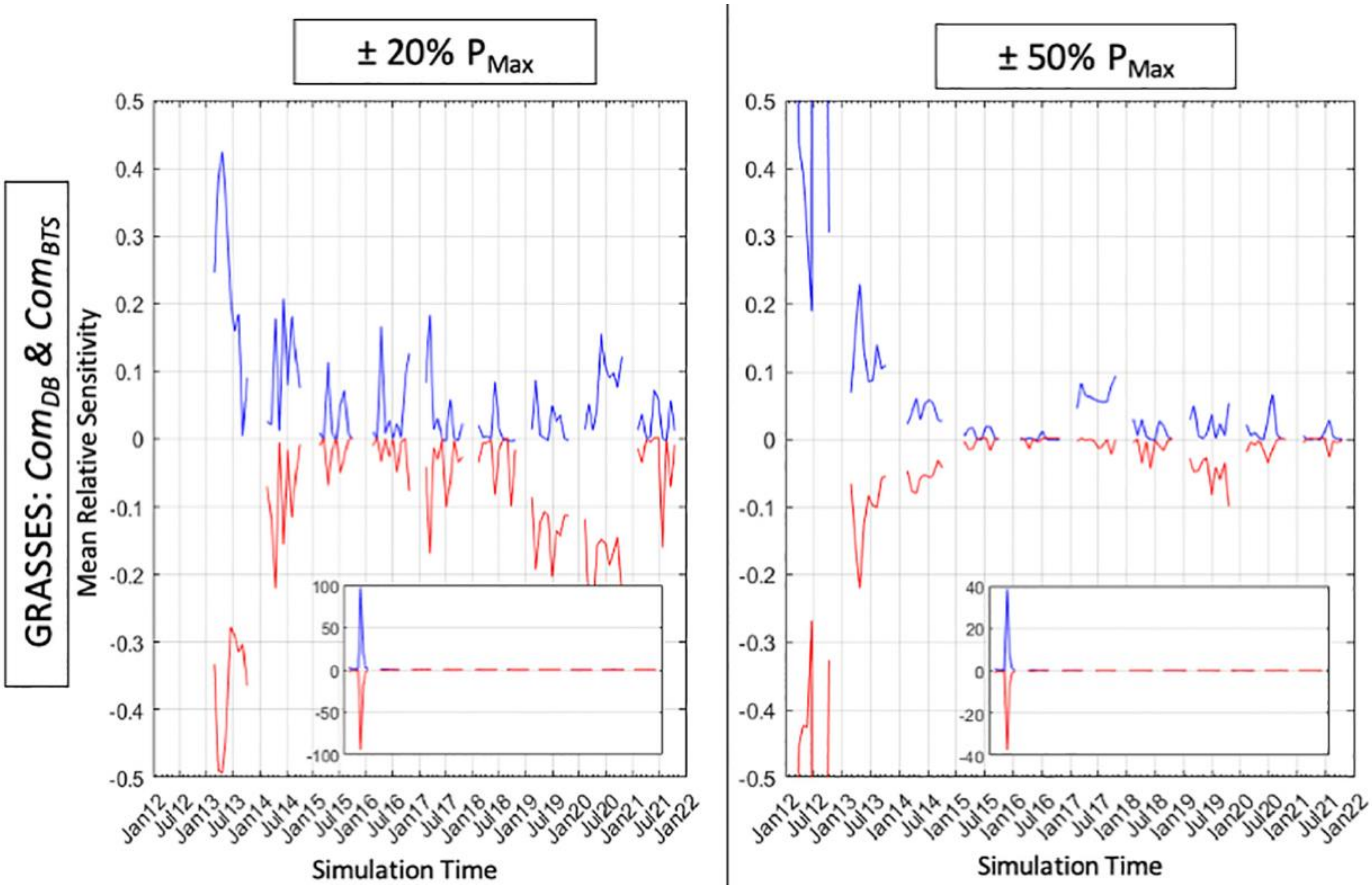
ONE-WAY SENSITIVITY ANALYSIS

For simulation models

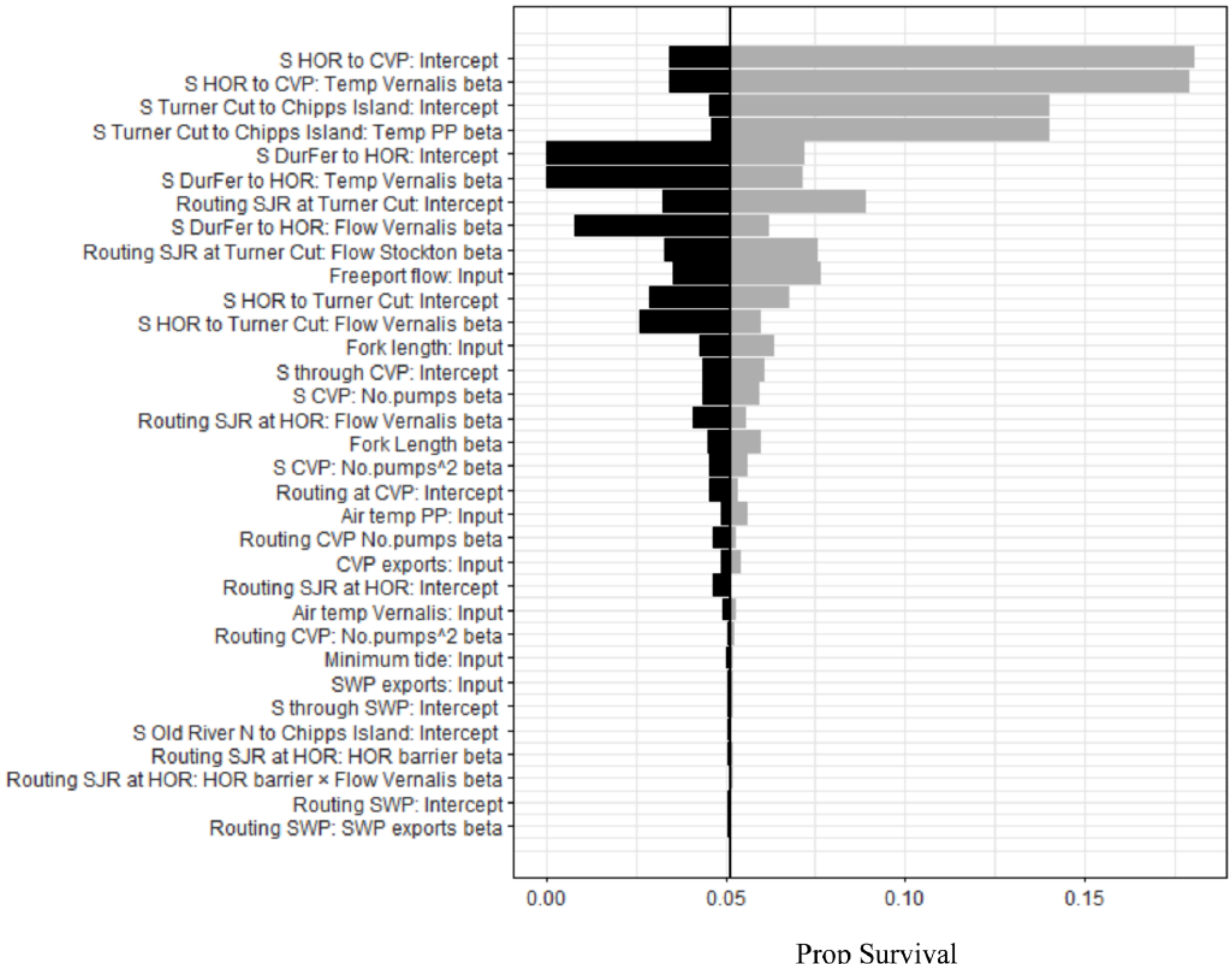
- Fix parameters at their means
- Vary 1 parameter across its range

Important point: keep unknown ranges relatively consistent across parameters and inputs

ONE-WAY SENSITIVITY ANALYSIS



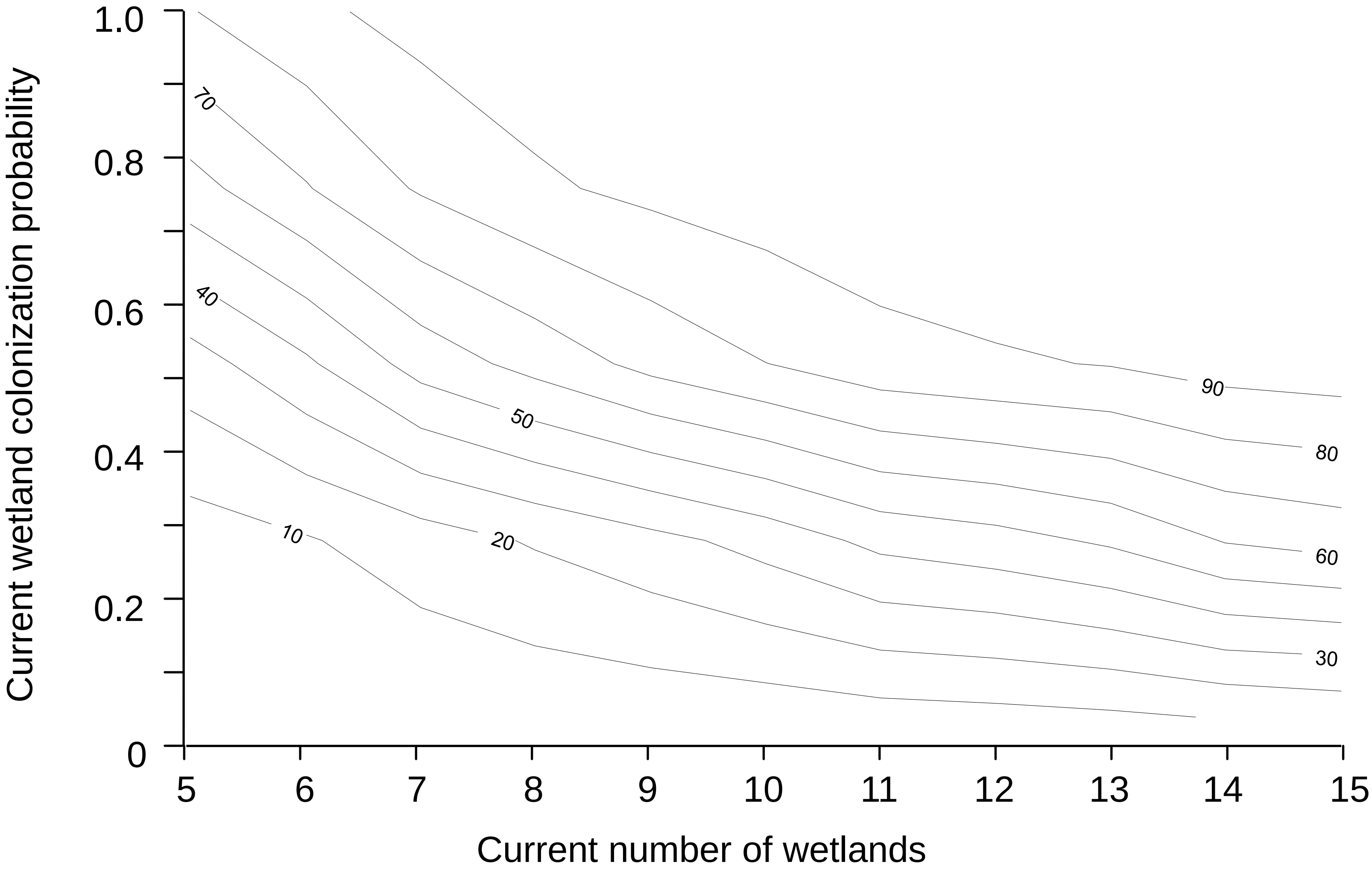
ONE-WAY SENSITIVITY ANALYSIS



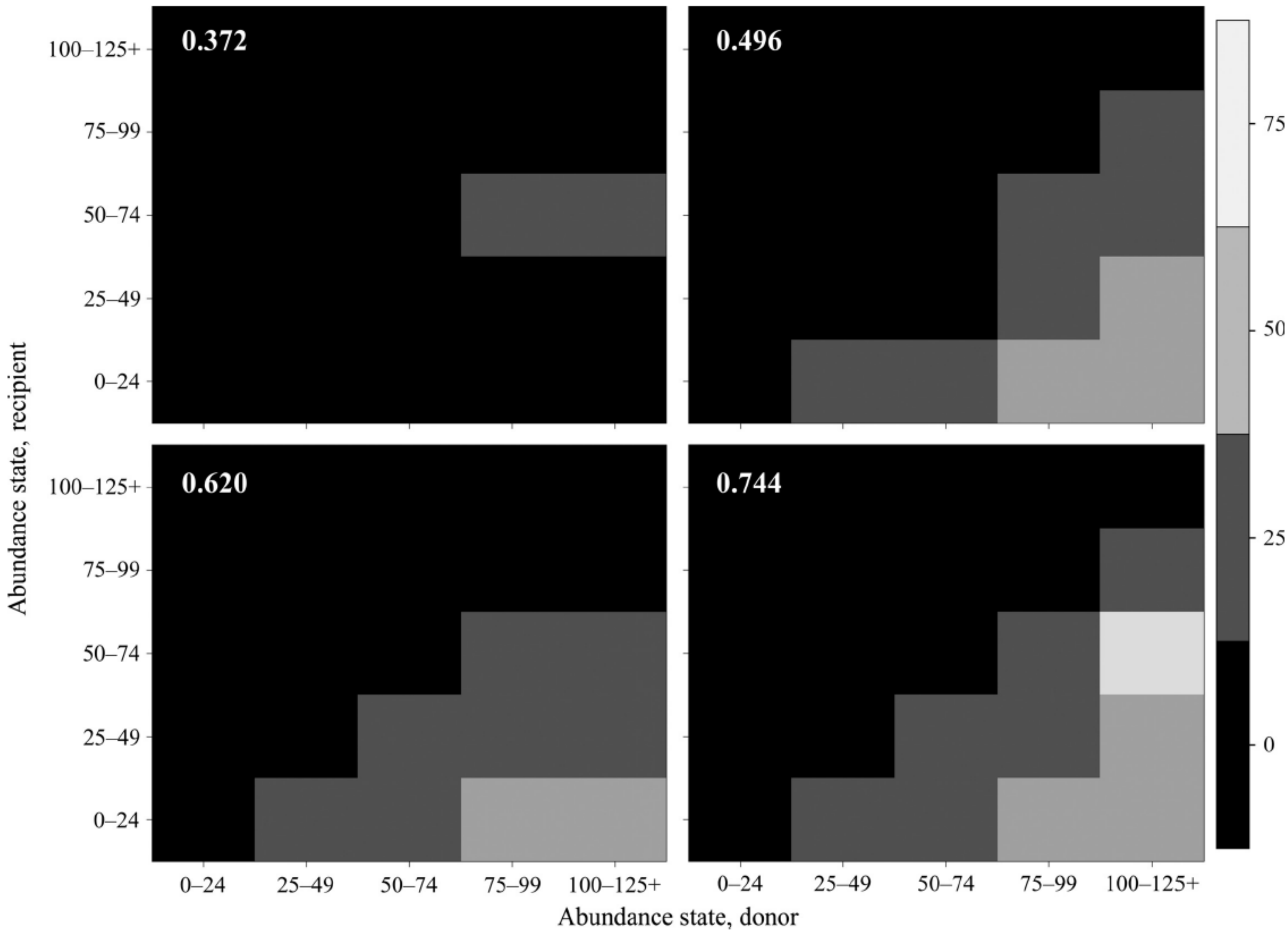
ONE-WAY SENSITIVITY ANALYSIS

	Changes in Top Decisions		
	Fall	*Winter*	*Spring*
Habitat			
Bypass habitat area	0	0	0
Adult holding habitat (spring-run only)			0
Spawning habitat	0	3	1
Perennially inundated juvenile habitat	3	5	3
Seasonally inundated juvenile habitat	3	5	1
Delta juvenile rearing habitat	0	0	0
Discharge and temperature			
Delta inflow	0	4	0
Median discharge	9	19	8
Proportion natal discharge	0	0	0
Proportion flow at Yolo and Sutter bypasses	0	0	0
Proportion pulse flow	0	0	0
Mean temperature	6	14	5
Operations			
Cross channel gate is open			
Exports from CVP and SWP	0	0	0
Total diversion	0	5	0
Miscellaneous			
Initial number of adults	0	5	1
Predator contact points	0	4	0
Hatchery smolts			
Angler (instream) and ocean harvest rates	0	0	0
Predator prevalence	0	0	0
Hatchery returning adults' stray rate	0	0	0

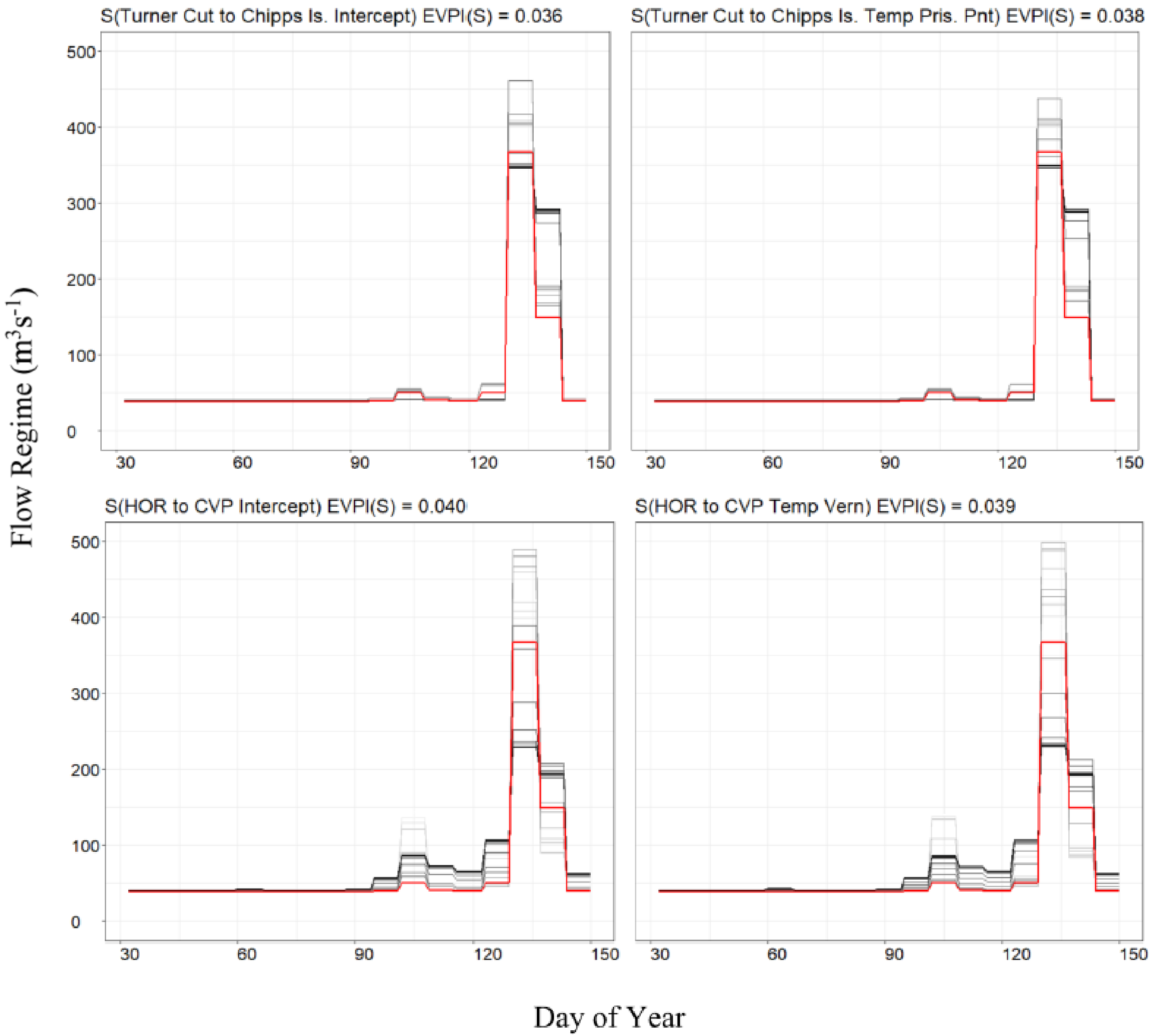
TWO-WAY SENSITIVITY ANALYSIS



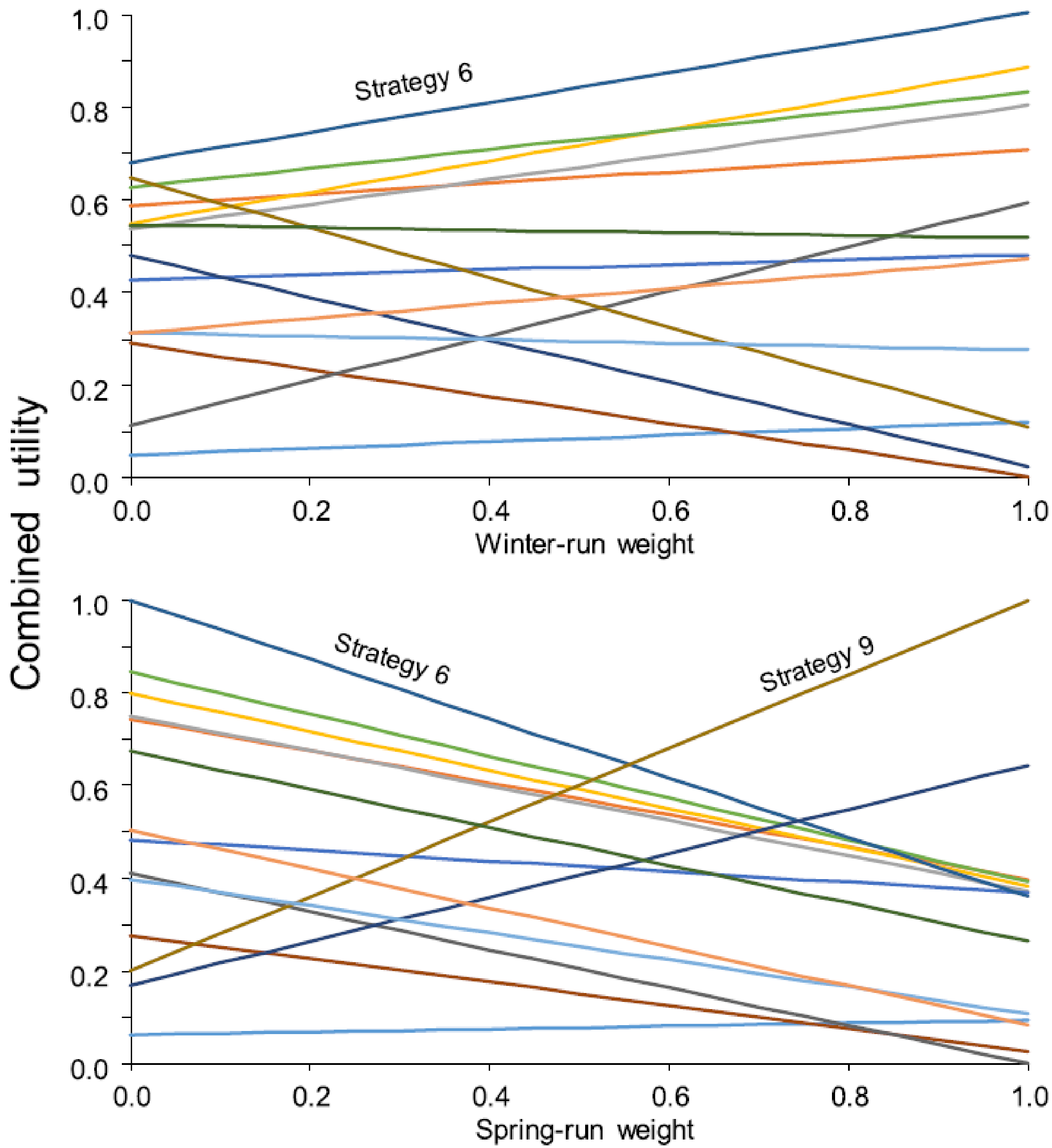
RESPONSE PROFILE



RESPONSE PROFILE



INDIFFERENCE CURVES



SUMMARY

Ecological systems are complex and uncertain

Stochasticity and muti-model inference are two common ways to incorporate uncertainty in ecological models

Control your world (model)

- Look toward simplification
- Control stochasticity
- Examine the influence of uncertainty
- Look toward uncertainty reduction

REDUCING UNCERTAINTY:
LEARNING HOW A SYSTEM WORKS

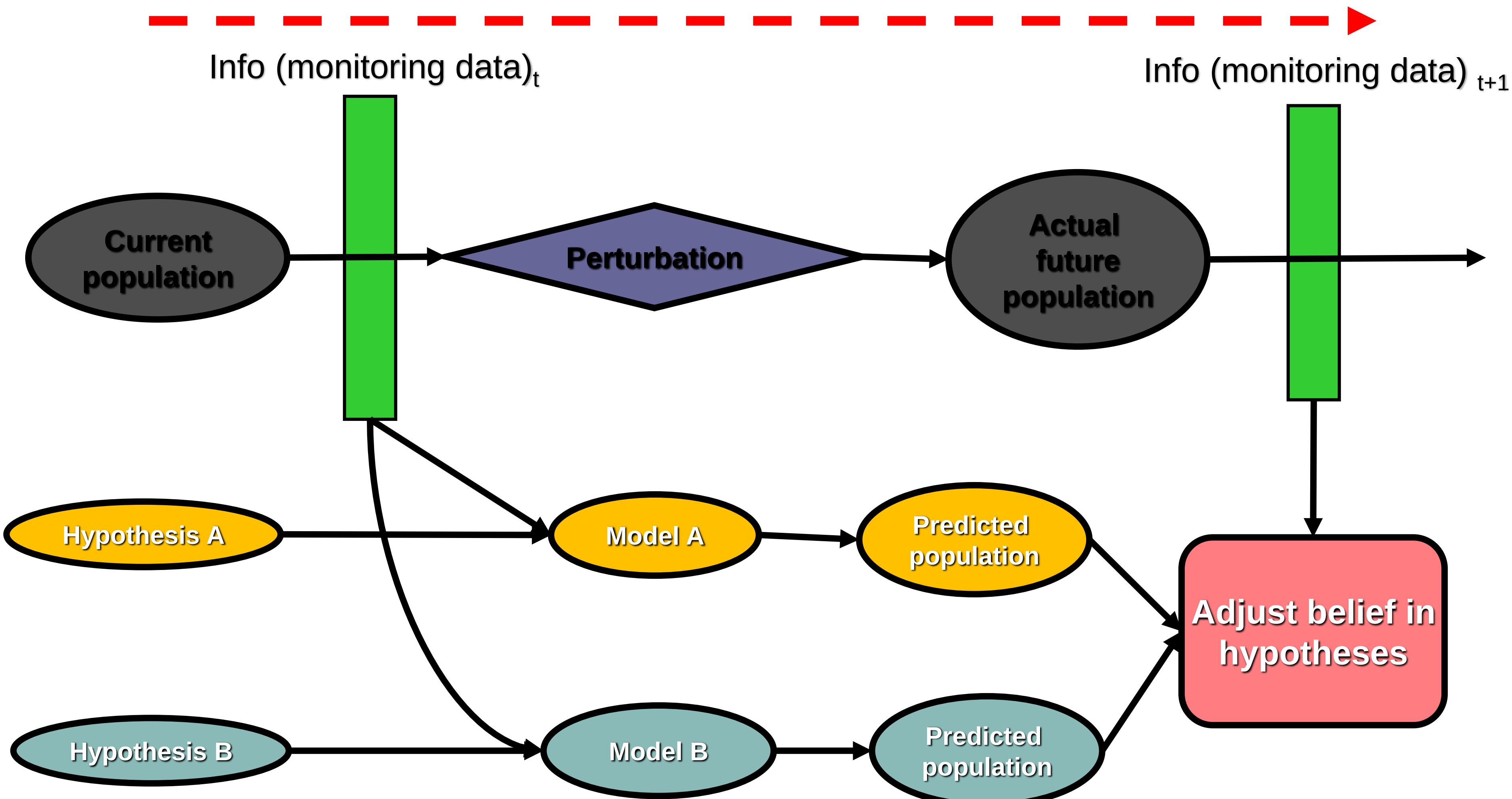
Retrospective study

- Analyze existing data
- Correlative, usually basis initial models

Experimentation

- Replication, randomization, treatments
- Feasibility
- Expensive

INTEGRATING MODELS AND MONITORING



INTEGRATING MODELS AND MONITORING