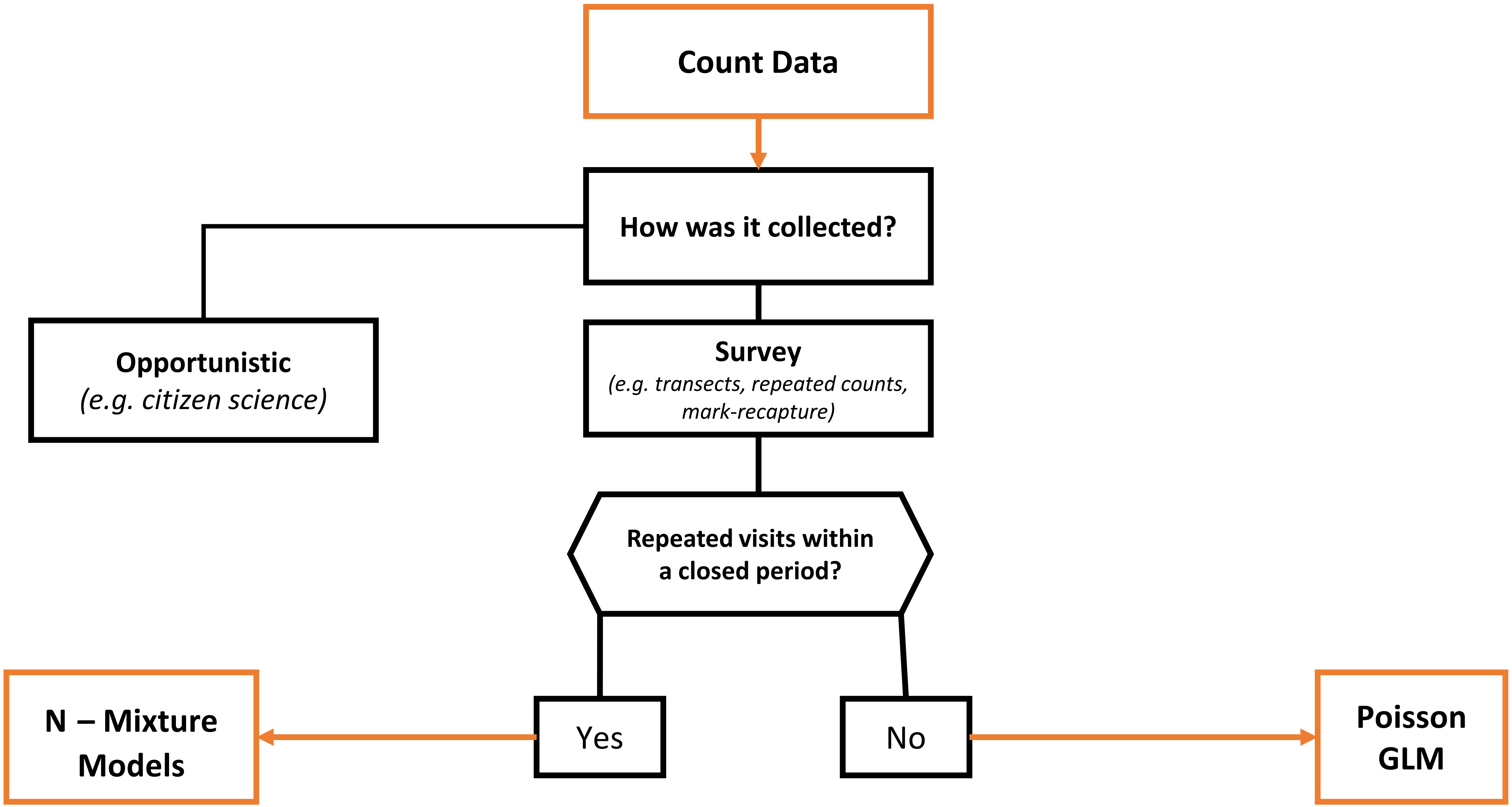
Count data

SSA 200

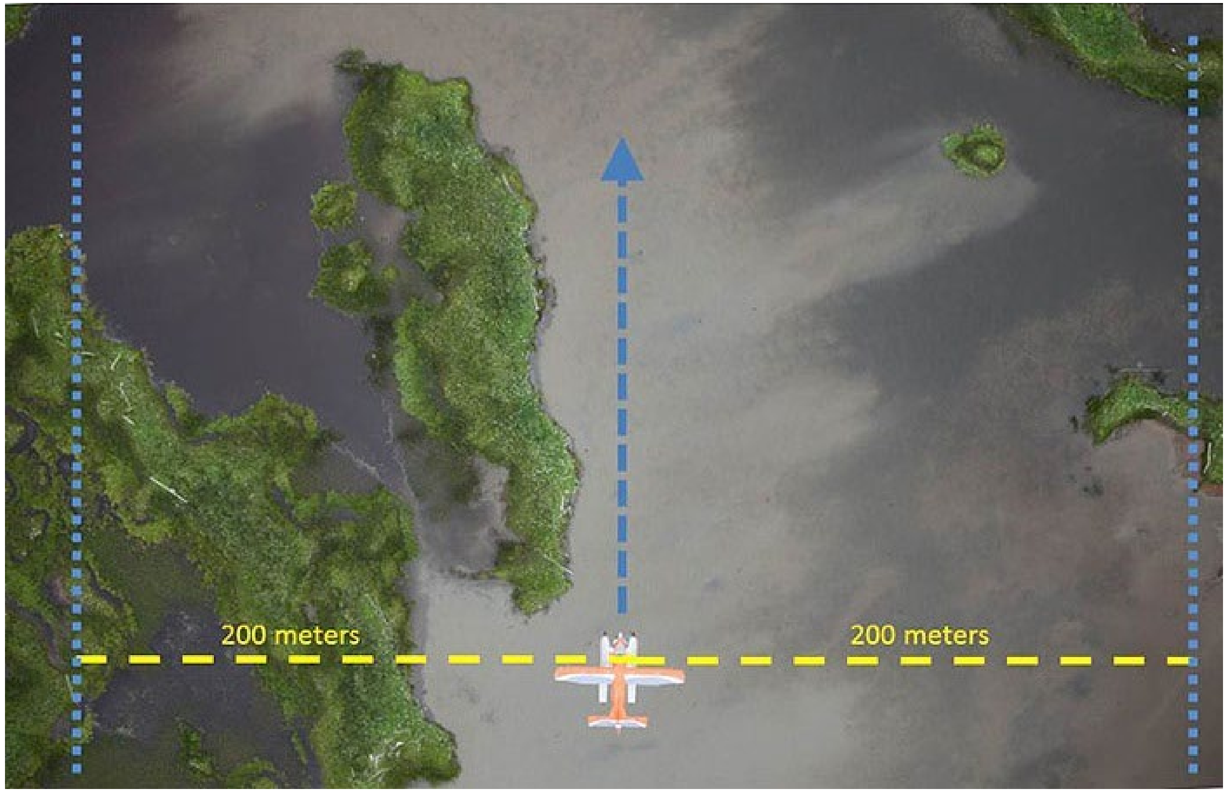


alabama cooperative fish and wildlife research unit

U.S. FISH & WILDLIFE SERVICE

Count data

- Typically collected annually or seasonally
- Data collection methods
  - Camera surveys
  - Aerial surveys
  - Point counts
  - Transects



Problems with count data

- Sampling and observation errors
  - Target population not fully sampled
  - Individuals present but not detected
  - Double counting
  - Misidentified individuals

Problems with count data

- Sampling and observation errors
  - Target population not fully sampled
  - Individuals present but not detected
  - Double counting
  - Misidentified individuals

Correcting count data

- We can correct count data for detection if data was specifically collected to estimate detection
  - Study design was set to capture information on detection and counts
  - A separate study was designed to estimate detection probability
- Detection data can be collected by
  - Distance sampling
  - Double observer study
  - Repeated counts at several sites within a close period
  - Etc.

What if we can't correct for detection?

- We can move forward but we need to understand the limitations of our data
  - What can we estimate?
    - **Apparent** abundance or an abundance **index**
    - **Trend** in apparent abundance
    - Relationships with ecological covariates
  - Assumptions
    - Detection is constant over time
  - How can we do it?
    - Generalized linear models (GLM)
    - State-space models (SSM)

Generalized linear models

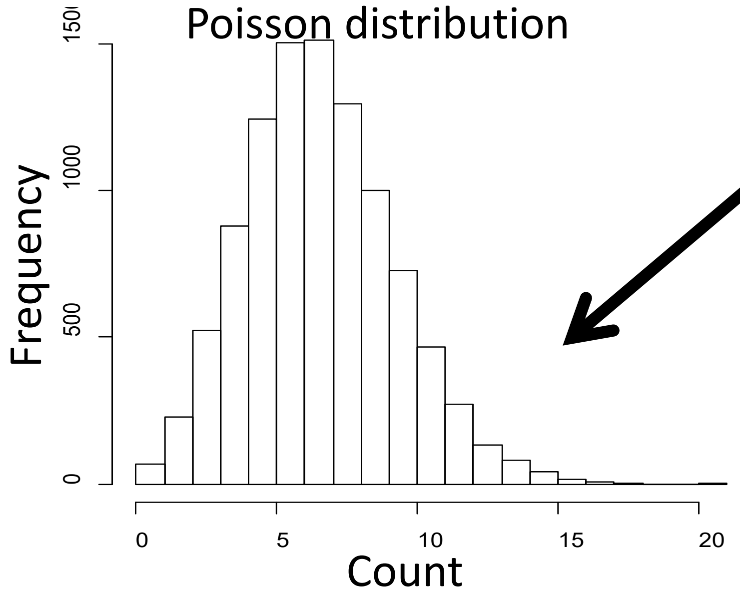
- GLM's are based on an assumed relationship called a link function between a linear predictor of the explanatory variables **(ecological covariates)** and the response variable **(count)**
- GLM's are an extension of the general linear model
  - Used when error is non-normally distributed
    - Most ecological data is non-normal!

Generalized linear models

- Counts typically modeled using a Poisson distribution
  - If data is over dispersed or there are a lot of 0's
    - Negative Binomial
    - Zero-inflated Poisson
- Count data must reasonably follow the chosen distribution
- Model selection
  - May test a number of ecological covariates
  - Use AIC to compare candidate models

Generalized linear models

- Poisson Generalized Linear Model
  - Discrete, positive integers (0, 1, 2, ..)
  - One parameter guides mean and variance



$$C = \alpha + \beta x$$

Environmental/habitat covariates

State-space models

- Time series models
  - Model the **true state** of the system (abundance) as an unobserved process
  - Observed data (**counts**) are modeled conditional on the true state (abundance) and the observation error
- Partitions variance in counts
  - Process error – Biological or process variation (e.g. demographic stochasticity)
  - Observation error – sampling variation

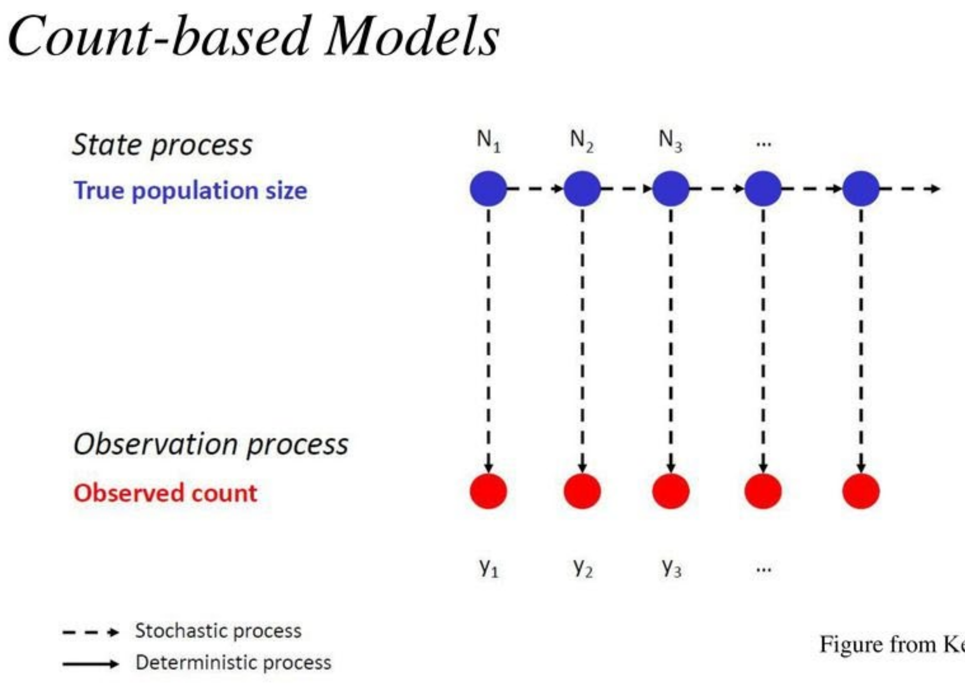


Figure from Kéry & Schaub 2012

State-space models

- Provides estimates of population growth rate
- Accounts for sampling variation (observation error) and process error (variation in abundance)
- Drawbacks
    - Cannot correct for bias in counts relative to true abundance
    - Can be relatively complex
    - Simple models suffer from estimation problems
    - Model fit and selection are difficult

What if we can correct for detection?

- If we have an estimate(s) of detection probability we can correct the counts and estimate **abundance**
  - $\hat{N} = C/\hat{p}$
    - $\hat{N}$ is estimated abundance
    - C is the count
    - $\hat{p}$ is detection probability
- If we have repeated counts at several sites in a closed period
  - We can use N-Mixture Models!

N-mixture models

- Use *repeated counts* at *several sites* to estimate detection probability directly
- Can include covariates associated with either abundance or detection
    - Explicitly model spatial and temporal variation
- Called “mixture” because it combines two GLMs
    - Poisson GLM – abundance
    - Binomial GLM (Logistic regression) – detection

N-mixture models

- Model detection as a function of covariates
  - Survey timing, observer
  - Habitat or weather
- Model abundance as a function of covariates
  - Habitat type
  - Presence/absence of predators
- Include these covariates as predictors of species abundance

N-mixture models

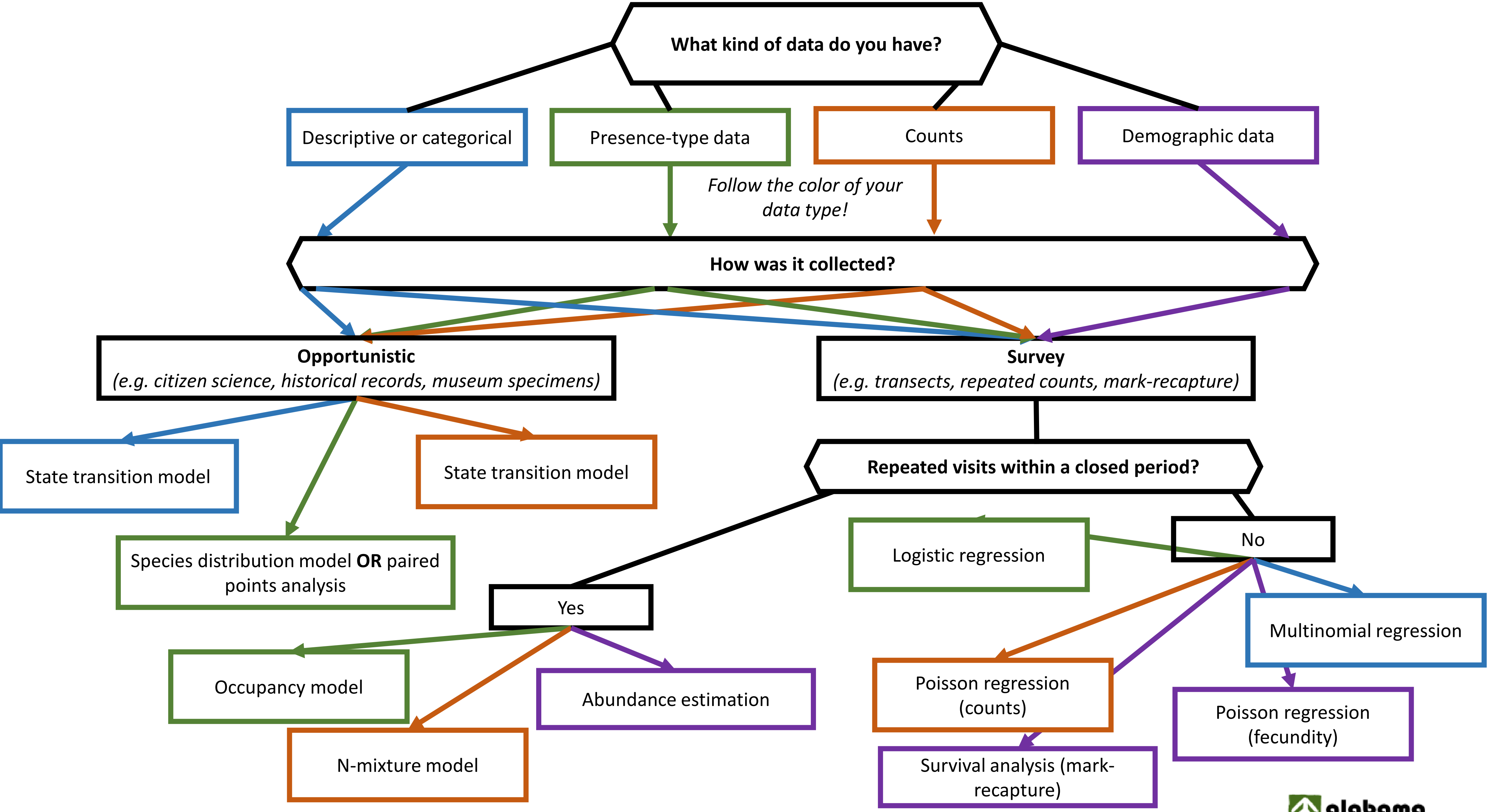
- Assumptions
  - Sites closed to immigration/emigration between surveys
  - Detection process is independent at each site but can vary among sites
  - No double counting
  - Equal detection probability for all individuals within a sample
- Model selection
  - Typically use AIC for both detection and abundance models
  - Assess relative fit of model sets to the data

Review

- What types of questions should you ask before you choose an analysis for your count data?

- What types of models are available for count data that are not corrected for detection?

- Can you estimate abundance or only relative abundance with repeated counts at several sites within a closed period?



This roadmap is to serve as a general guide and is not an exhaustive list of all analysis options. Also, ***always check the specific assumptions of your planned modeling approach!***

Questions?