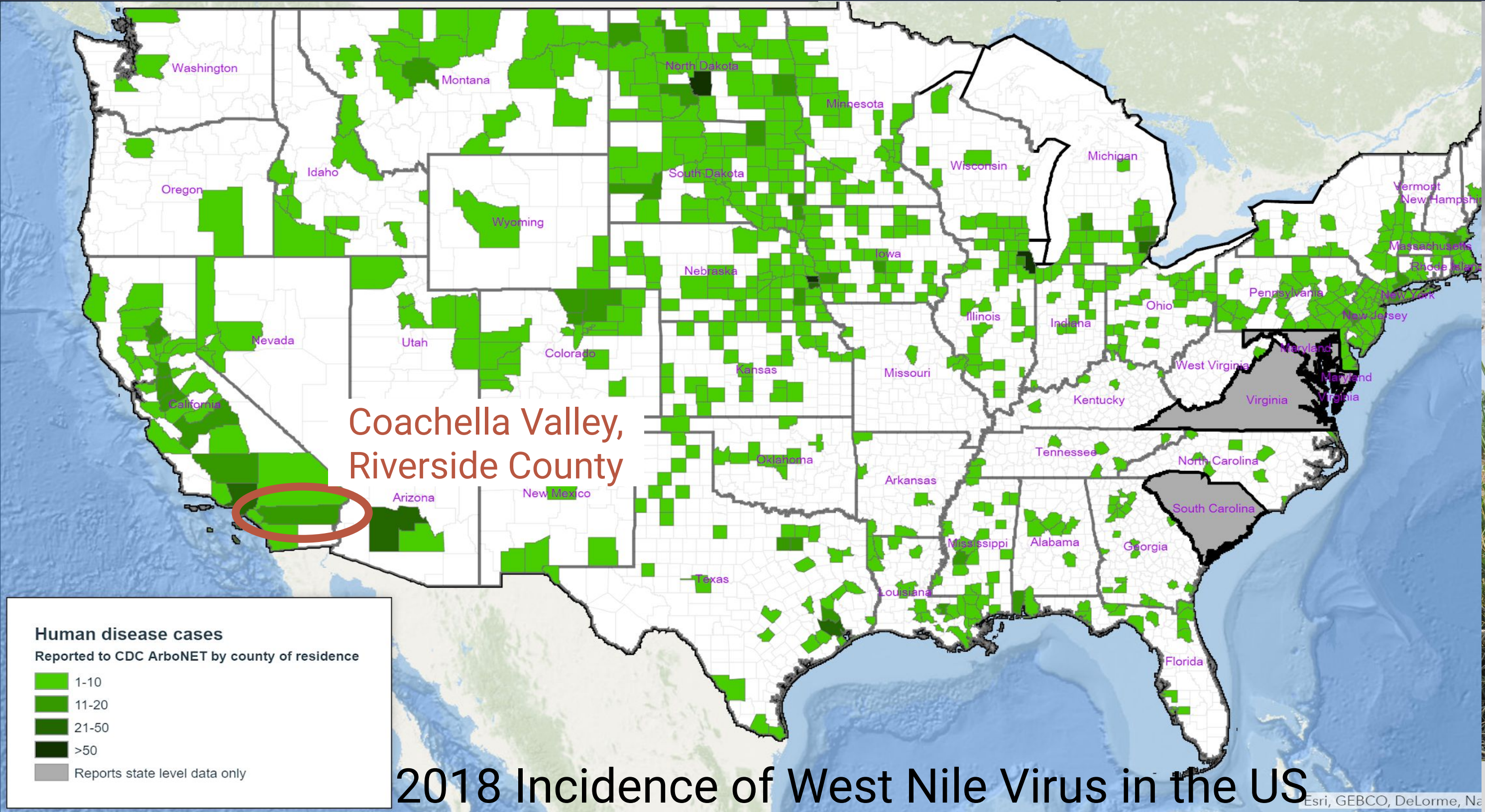
Mosquito Abundance Estimation in the Coachella Valley

May 29, 2019

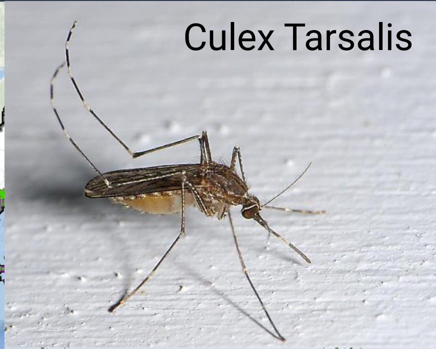
Jacob Shultz
Joseph Cole

Problem and Objectives



2018 Incidence of West Nile Virus in the US

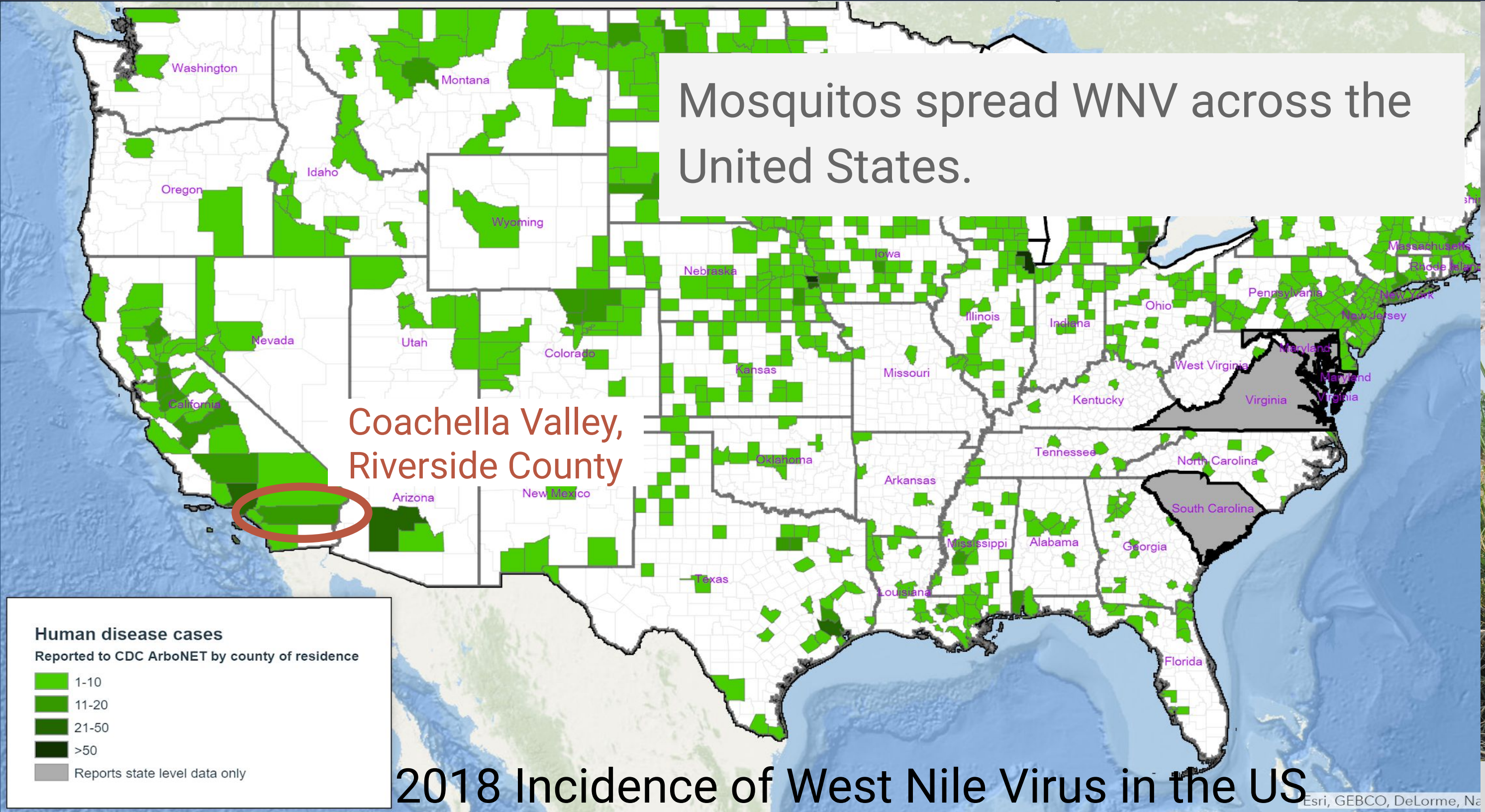
Culex Tarsalis



Pelecanus Occidentalis

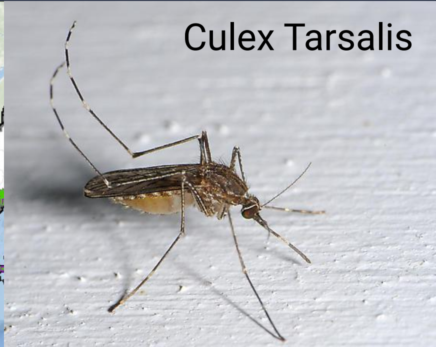
Problem and Objectives

Mosquitos spread WNV across the United States.



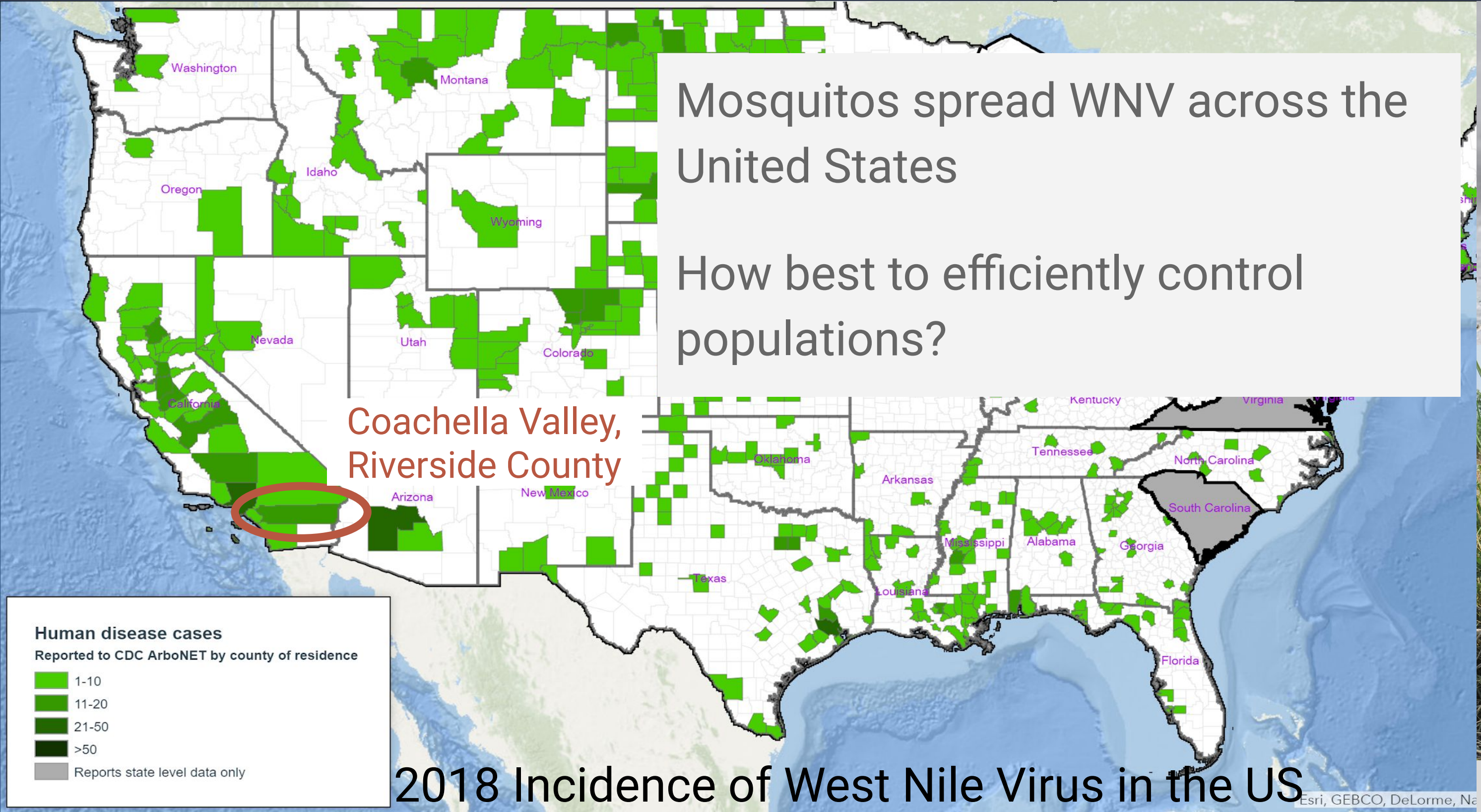
2018 Incidence of West Nile Virus in the US

Culex Tarsalis



Pelecanus Occidentalis

Problem and Objectives

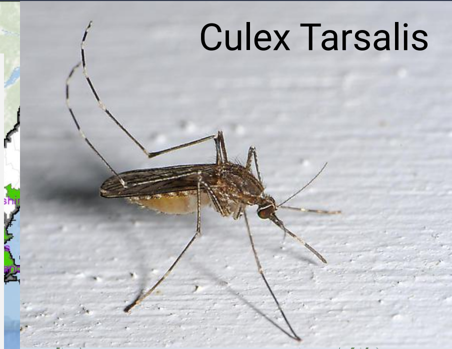


Mosquitos spread WNV across the United States

How best to efficiently control populations?

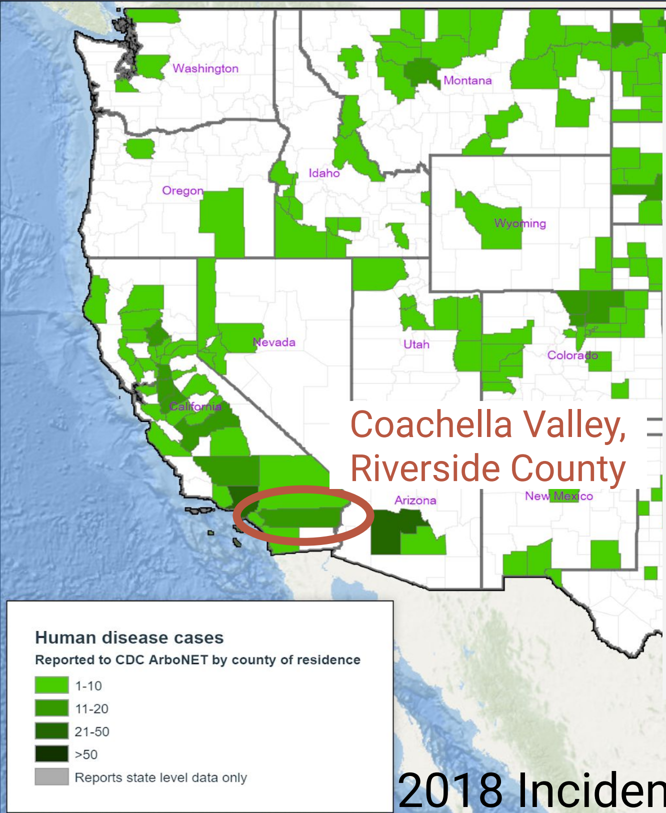
2018 Incidence of West Nile Virus in the US

Culex Tarsalis



Pelecanus Occidentalis

Problem and Objectives

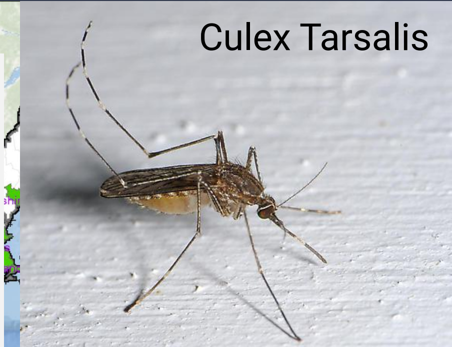


2018 Incidence of West Nile Virus in the US

Mosquitos spread WNV across the United States

How best to efficiently control populations?

There is a need to introduce modern statistical techniques throughout the mosquito control community.



Culex Tarsalis

Pelecanus Occidentalis

Data Description

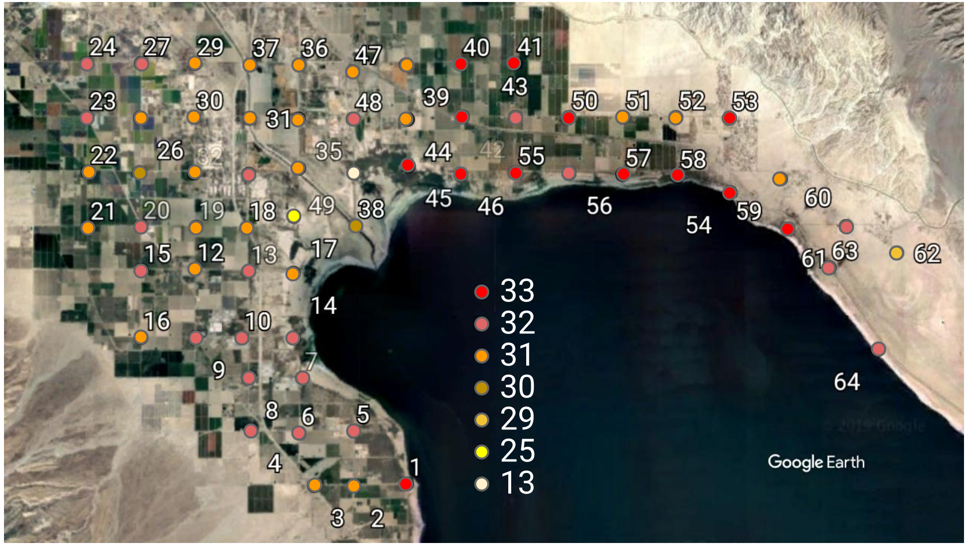
Data Summary

Primary Model Inputs
Trap Number: *i* - Integer in [1, 64] describing which trap the data was collected from - Trap 28 removed due to zero samples taken
Time: *t* - Integer in [1, 42] describing when the sample was taken - An increase in 1 corresponds to a change in two weeks -Ranges from Apr 1994 to Nov 1995 (Gap from Oct 1994 to Feb 1995)
Count: $n_{i\,t}$ - Integer in [0, 7936] describing the number of *Culex tarsalis* captured at trap *i* during time period *t* - Treated as a continuous response

Predictor	Site/Obs Level	Data Type	Description
Latitude	Site	Continuous [33.44, 33.55]	Latitude of trap *i*
Longitude	Site	Continuous [-116.15, -115.89]	Longitude of trap *i*
Dist to sea	Site	Continuous [0.20, 9.78]	Distance from trap *i* to the Salton Sea
Max Temp (tenths of degrees C)	Obs	Continuous [261.9, 453.6]	Average of all daily highs within time period *t*
Observed Temp (tenths of degrees C)	Obs	Continuous [233.7, 433.9]	Average of all temps taken at 17:00 within time period *t*
Biomes (9 types)	Site	Continuous [0, 1]	Percent biome surrounding a trap

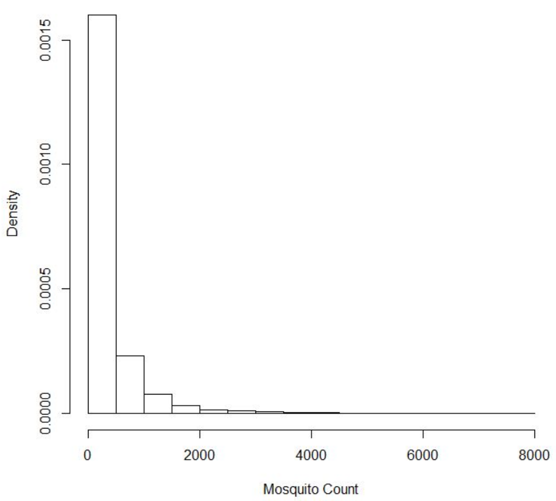
Exploratory Data Analysis

Number of Time Samples from Each Trap



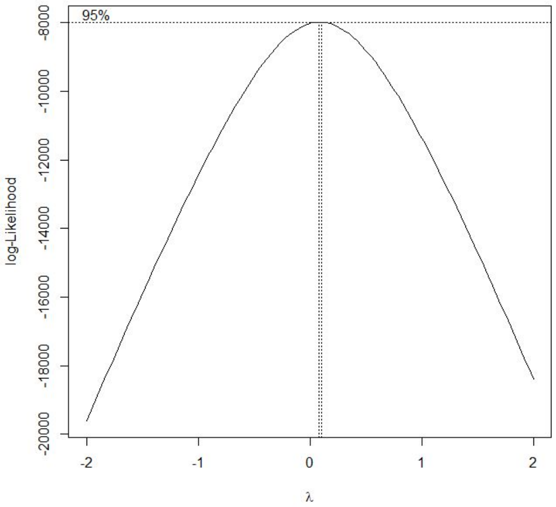
Exploratory Data Analysis

Distribution of Raw Count Data
(All traps, all time)



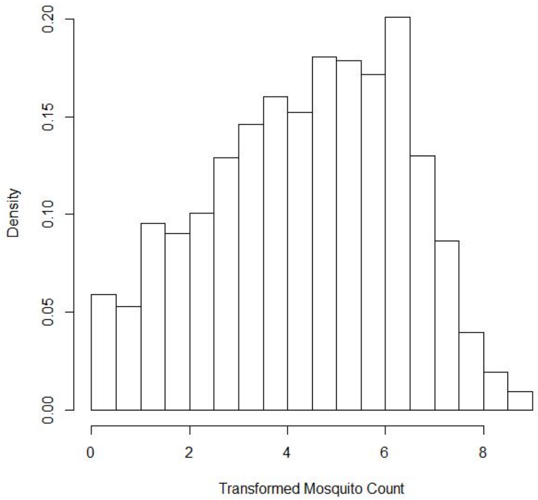
Skew: 5
Kurtosis: 38

Box-Cox Likelihood of Transformation Power



Optimal data transform (λ=0.1) is close to that used by Reisen (λ=0).

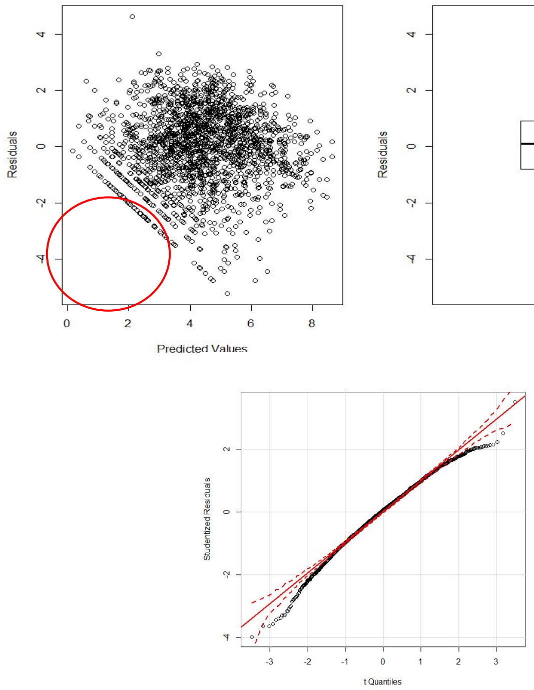
Distribution of Transformed Count Data



Skew: -0.3
Kurtosis: 2.3

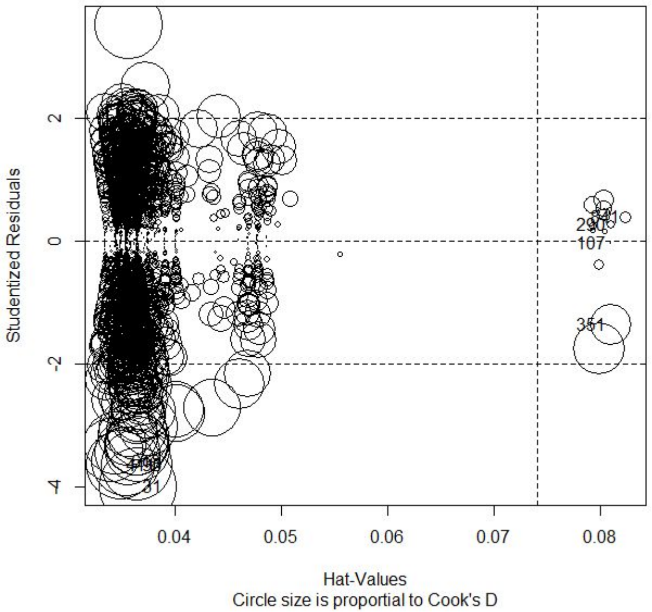
Exploratory Data Analysis

Linear Model Residual Diagnostics



Residuals form a left skewed distribution that has a heavier than expected tail on the left and a lighter than expected tail on the right. This is consistent with the failure of the response transformation to fully normalize the data.

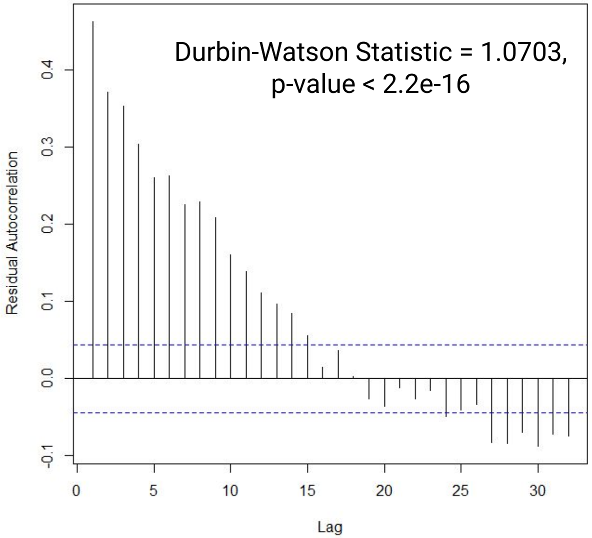
Identification of High Influence Data



Exploratory Data Analysis

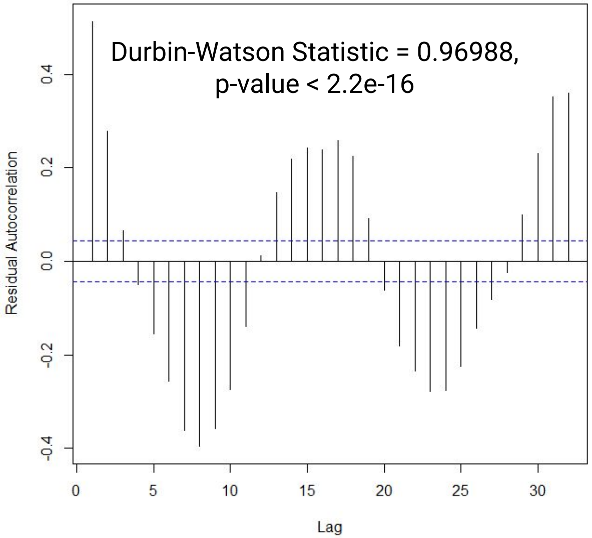
Evidence for the Need to Model Population Dynamics

Data ordered:
First by time period
Then by trap number



Correlation between nearby traps could indicate a need to model spatial dynamics.

Data ordered:
First by trap number
Then by time period



Correlation through time could indicate inadequate modelling of seasonal fluctuations.

Exploratory Data Analysis

Considerations for Model Selection

Covariate alternatives explored:

1) Latitude/Longitude vs. Distance to Salton Sea
2) Trap Number (as a factor) vs. Habitat ratios and Lat/Lon
3) Year and Month (as factors) vs. Time ID
4) Month (as a factor) vs. Temperature

(selected options)

- Variance Inflation Factors used to find incompatible groups of covariates
- R^2_{adj} used as a rough model selection metric

Methodology: N-Mixture Model for Closed Populations

Goal: Estimate these primary model parameters

p = P(trapping an individual | individual in the sphere of influence of trap)
λ = Abundance at a single site

Both parameters can be estimated with a simple intercept model or using vectors of covariates $\boldsymbol{\beta}_p$ and $\boldsymbol{\beta}_\lambda$.

Open/Closed Populations:

- Closed populations have a constant site-level populations over time
- Open populations can have "additions or deletions" in site-level populations (Dail and Madson, 2011)
    - Require modeling of population dynamics...

Methodology: N-Mixture Model for Open Populations

Goal: Estimate all of these (Note abundance is now *initial* abundance)

Primary model parameters:

p = P(trapping an individual | individual in the sphere of influence of trap)

λ = Abundance at a single site at the first time step

Population dynamics:

γ = Arrival rate

ω = Survival percentage between time steps

Again, parameters can be estimated with a simple intercept model or using vectors of covariates $\boldsymbol{\beta}_p$, $\boldsymbol{\beta}_\lambda$, $\boldsymbol{\beta}_\gamma$, and $\boldsymbol{\beta}_\omega$.

Fits a markov chain describing abundance through time

Testing For Closure

Key Concept:

Setting $\{\gamma = 0$ and $\omega = 1\}$ in the open model implies a closed model assumption. Therefore, these models are nested and we can use LRT to test for closure

$$LR = -2\ln\left(\frac{\sup(L\ under\ closed\ assuption)}{\sup(L\ under\ open\ assumption)}\right)$$

$$= -2\ln\left(\frac{\sup\big(\mathrm{L}(p, \lambda, |\gamma = 0, \omega = 1, \{n_{it}\})\big)}{\sup\big(\mathrm{L}(p, \lambda, \gamma, \omega|\{n_{it}\})\big)}\right)$$

LR is distributed as a mixture of $\chi^2_{(0)}$, $\chi^2_{(1)}$, and $\chi^2_{(2)}$ since γ and ω are on the edges of Θ.

Results:

For intercept-only models $LR = 2765.518$, providing strong evidence against closure

Intercept Model Parameter Estimates

All estimates were fit under intercept models (no covariates) using the `unmarked` package

	AIC	$\hat{\lambda}$	$\hat{p}$	$\hat{\gamma}$	$\widehat{\omega}$
Open Population	12974.97	952.541	0.6439376	2.0719353	0.6509723
Closed Population	15736.49	7069.499	0.05684929	N/A	N/A

Open population assumption seems more reasonable

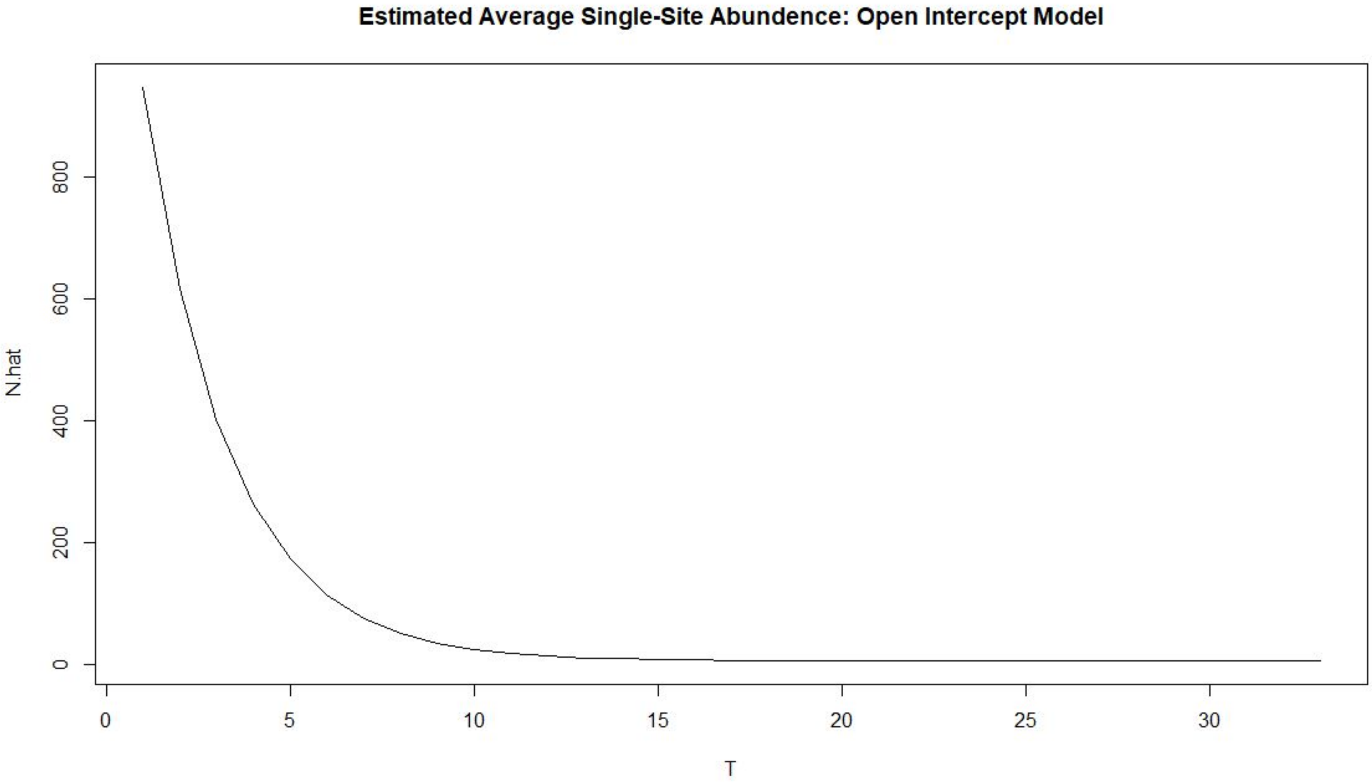
- Fitted abundance is more consistent with count summary statistics
- Smaller AIC in open population model

Intercept Model Parameter Estimates

We can step through the fitted Markov chain to estimate abundance over time

Intercept model is cleary underspecified

- No temporal effects
- No spatial effects



Computational Challenges

1) R package `unmarked` doesn’t allow for temporal modeling of population dynamics (only detection probability)

- Big concern: We observe counts cycling through time → dynamics are time-dependent
- Any model fit using `unmarked` will be underspecified

Computational Challenges

1) R package `unmarked` doesn't allow for temporal modeling of population dynamics (only detection probability)

- Big concern: We observe counts cycling through time → dynamics are time-dependent
- Any model fit using `unmarked` will be underspecified

2) MLE's take a long time to converge

- Closed intercept: < 30 Seconds
- Open intercept: ~10-30 Minutes
- Best model so far (next slide): ~12.5 Hours
  - Not an especially complex model

Computational Challenges

1) R package `unmarked` doesn't allow for temporal modeling of population dynamics (only detection probability)

- Big concern: We observe counts cycling through time → dynamics are time-dependent
- Any model fit using `unmarked` will be underspecified

2) MLE's take a long time to converge

- Closed intercept: < 30 Seconds
- Open intercept: ~10-30 Minutes
- Best model so far (next slide): ~12.5 Hours
  - Not an especially complex model

3) Some combinations of predictors lead to identifiability issues (e.g. predicting detection probability using temperature)

Best Model So Far by AIC

Takeaways:

- Very high significance for all covariates
- Improved AIC compared to intercept models
- Model likely underspecified → Coefficients could be unreliable

```
Abundance (log-scale):
            Estimate     SE     z  P(>|z|)
(Intercept)    3.370 0.0885  38.1 0.00e+00
d.to.sea      -0.396 0.0311 -12.8 2.69e-37

Recruitment (log-scale):
            Estimate     SE     z   P(>|z|)
(Intercept)    1.696 0.0711 23.87 6.91e-126
d.to.sea      -0.128 0.0225 -5.69  1.25e-08

Apparent Survival (logit-scale):
            Estimate     SE    z  P(>|z|)
(Intercept)   0.6647 0.1419 4.68 2.81e-06
d.to.sea      0.0371 0.0326 1.14 2.55e-01

Detection (logit-scale):
             Estimate     SE      z  P(>|z|)
(Intercept)    -1.065 0.1055 -10.10 5.70e-24
seasonspring    1.154 0.0690  16.72 8.81e-63
seasonsummer   -0.311 0.0526  -5.92 3.28e-09

AIC:  12048.1
```

Conclusions

Contrasting Methods

Reisen (1999)

- Best response transformation for a linear regression model (judging by R^2_{adj})
- Still violates residual independence, normality, and constant variance assumptions
- Can’t model population dynamics or imperfect detection

Conclusions

Contrasting Methods

Reisen (1999)

- Best response transformation and set of covariates for a linear regression model (judging by R^2_{adj})
- Still violates residual independence, normality, and constant variance assumptions
- Can't model population dynamics or imperfect detection

Dail, Madson (2011)

- Flexible models that account for imperfect detection and population dynamics
- Computation time and current R functions limit usability
- Lack of temporal modeling of population dynamics + high significance on included covariates → Current best model is almost certainly underspecified

Recommendations for Future Work

Two Paths

1) Likelihood methods
- Continue finding MLE's using Dail, Madson model
- Requires custom optimization code to work around limitations in `unmarked`

Recommendations for Future Work

Two Paths

1) Likelihood methods
- Continue finding MLE’s using Dail, Madson model
- Requires custom optimization code to work around limitations in `unmarked`

2) Bayesian Methods (Our Recommendation)
- Requires the likelihood + priors for all parameters
    - Likelihood given by Dail, Madson □
    - Priors can be intentionally vague or based on outside sources
        - e.g. Prior on *p* could be based on existing studies of CO_2 trap effectiveness
- Fit model by sampling from posterior (MCMC) → obtain posterior densities for all parameters

Questions?