

# Multivariate Analysis I

HES 505 Fall 2022: Session 19

Matt Williamson

# Objectives

By the end of today you should be able to:

- Recognize the link between regression analysis and overlay analysis
- Generate spatial predictions based on regression analysis
- Extend logistic regression to presence-only data models

# Estimating favorability

$$F(\mathbf{s}) = \prod_{M=1}^m X_m(\mathbf{s})$$

- Treat as binary
- Then if all inputs () are suitable
- Then if not

# Estimating favorability

- does not have to be binary (could be ordinal or continuous)
- could also be extended beyond simply 'suitable / not suitable'
- Adding weights allows incorporation of relative importance
- Other functions for combining inputs ()

# Weighted Linear Combinations

- is now an index based on the values of
- can incorporate weights of evidence, uncertainty, or different participant preferences
- Dividing by normalizes by the sum of weights

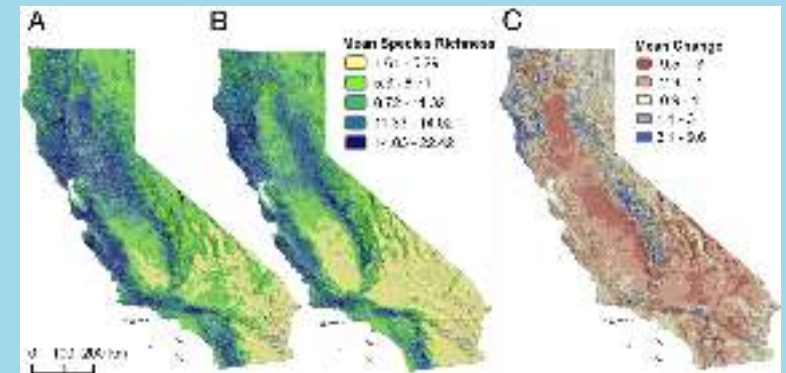
# Model-driven overlay

- If we estimate using data, we specify as the outcome of regression
- When is binary  $\rightarrow$  logistic regression
- When is continuous  $\rightarrow$  linear (gamma) regression
- When is discrete  $\rightarrow$  Poisson regression
- Assumptions about matter!!

# Logistic Regression and Distribution Models

# Why do we create distribution models?

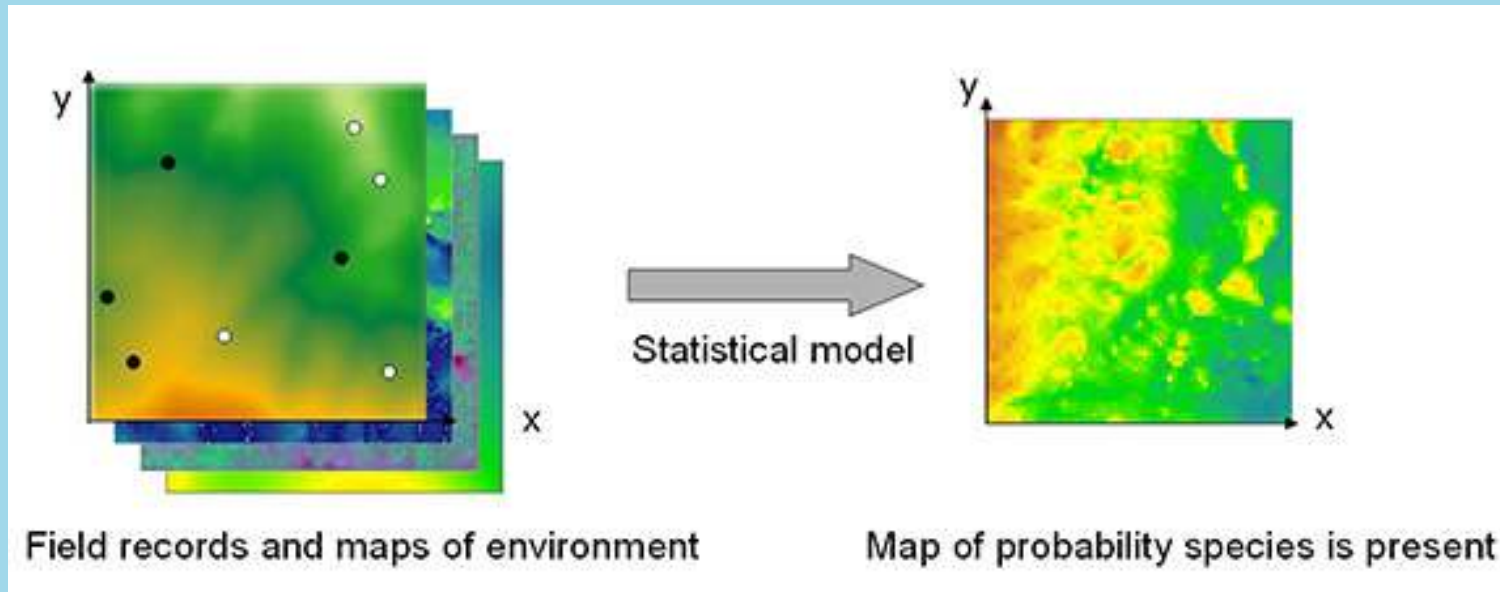
- To identify important correlations between predictors and the occurrence of an event
- Generate maps of the 'range' or 'niche' of events
- Understand spatial patterns of event co-occurrence
- Forecast changes in event distributions



From Wiens et al. 2009



# General analysis situation



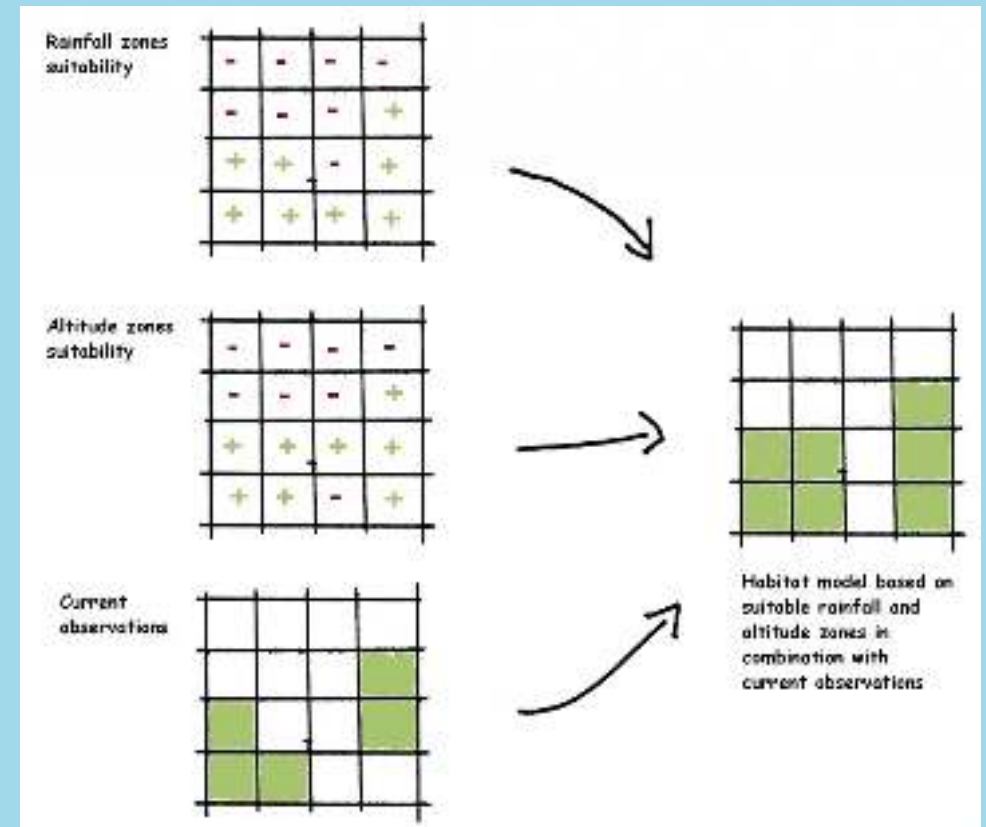
From Long

- Spatially referenced locations of events sampled from the study extent
- A matrix of predictors that can be assigned to each event based on spatial location

**Goal:** Estimate the probability of occurrence of events across unsampled regions of the study area based on correlations with predictors

# Modeling Presence-Absence Data

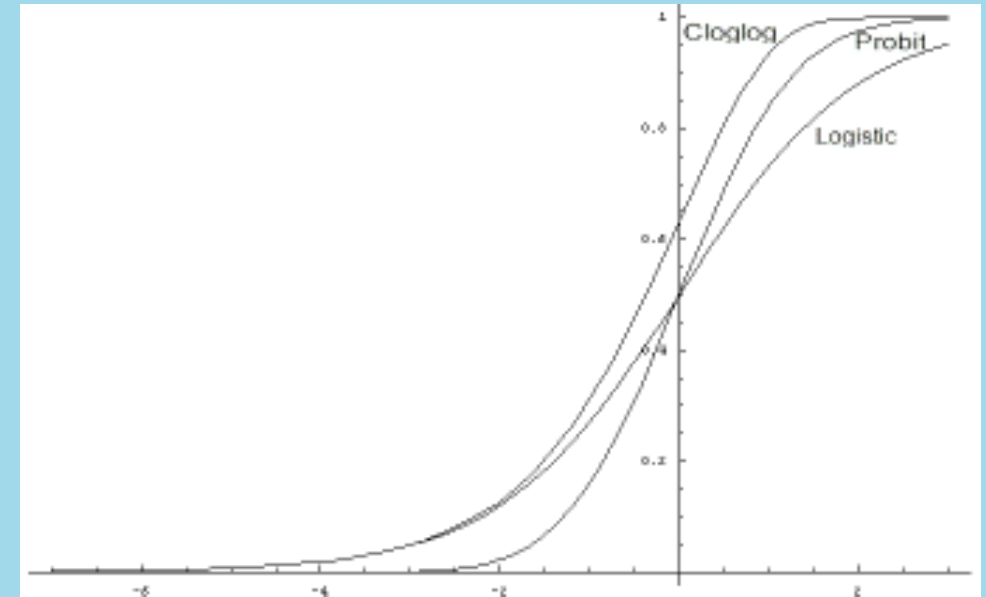
- Random or systematic sample of the study region
- The presence (or absence) of the event is recorded for each point
- Hypothesized predictors of occurrence are measured (or extracted) at each point



From By Ragnvald - Own work, CC BY-SA 3.0

# Logistic regression

- We can model favorability as the **probability** of occurrence using a logistic regression
- A *link* function maps the linear predictor onto the support (0-1) for probabilities
- Estimates of can then be used to generate ‘wall-to-wall’ spatial predictions



From Mendoza

# An Example

Inputs from the **dismo** package

# An Example

The sample data

```
1 head(pres.abs)
```

# An Example

## Building our dataframe

```
1 pts.df <- terra::extract(pred.stack, vect(pres.abs), df=TRUE)  
2 head(pts.df)
```

# An Example

## Building our dataframe

```
1 pts.df[,2:7] <- scale(pts.df[,2:7])  
2 summary(pts.df)
```

# An Example

Looking at correlations

```
1 pairs(pts.df[,2:7])
```



# An Example

## Looking at correlations

```
1 corrplot(cor(pts.df[,2:7]), method = "number")
```

# An Example

## Fitting some models

```
1 pts.df <- cbind(pts.df, pres.abs$y)
2 colnames(pts.df)[8] <- "y"
3 logistic.global <- glm(y~., family=binomial(link="logit"), data=pts.df[,2:8])
4 logistic.simple <- glm(y ~ MeanAnnTemp + TotalPrecip, family=binomial(link="logit"), data=pts.df[,2:8])
5 logistic.rich <- glm(y ~ MeanAnnTemp + PrecipWetQuarter + PrecipDryQuarter, data=pts.df[,2:8])
```

# An Example

## Checking out the results

```
1 summary(logistic.global)
```

# An Example

## Checking out the results

```
1 summary(logistic.simple)
```

# An Example

## Checking out the results

```
1 summary(logistic.rich)
```

# An Example

## Comparing models

```
1 AIC(logistic.global, logistic.simple, logistic.rich)
```

# An Example

## Generating predictions

```
1 preds <- predict(object=pred.stack, model=logistic.simple)
2 plot(preds)
3 plot(pres.pts$geometry, add=TRUE, pch=3, col="blue")
4 plot(abs.pts$geometry, add=TRUE, pch = "-", col="red")
```

# An Example

## Generating predictions

```
1 preds <- predict(object=pred.stack, model=logistic.simple, type="response")
2 plot(preds)
3 plot(pres.pts$geometry, add=TRUE, pch=3, col="blue")
4 plot(abs.pts$geometry, add=TRUE, pch = "-", col="red")
```



# An Example

## Generating predictions

```
1 preds <- predict(object=pred.stack, model=logistic.global, type="response")
2 plot(preds)
3 plot(pres.pts$geometry, add=TRUE, pch=3, col="blue")
4 plot(abs.pts$geometry, add=TRUE, pch = "-", col="red")
```

# An Example

## Generating predictions

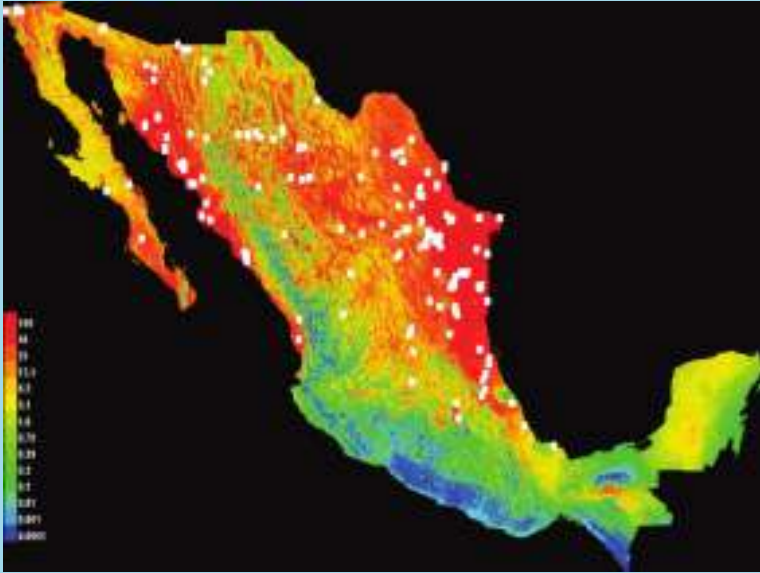
```
1 preds <- predict(object=pred.stack, model=logistic.rich, type="response")
2 plot(preds)
3 plot(pres.pts$geometry, add=TRUE, pch=3, col="blue")
4 plot(abs.pts$geometry, add=TRUE, pch = "-", col="red")
```

# Key assumptions of logistic regression

- Dependent variable must be binary
- Observations must be independent (important for spatial analyses)
- Predictors should not be collinear
- Predictors should be linearly related to the log-odds
- **Sample Size**

# Modelling Presence- Background Data

# The sampling situation



From Lentz et al. 2008

- Opportunistic collection of presences only
- Hypothesized predictors of occurrence are measured (or extracted) at each presence
- Background points (or pseudoabsences) generated for comparison

# The Challenge with Background Points

- What constitutes background?
- Not measuring *probability*, but relative likelihood of occurrence
- Sampling bias affects estimation
- The intercept

# Point Process Models

- Poisson Point Process Models model location (not )
- Number of points expected is given by a rate
- Model using Poisson regression

