E²M²: Ecological and Epidemiological Modeling in Madagascar

Data and Models

Centre ValBio
Ranomafana National Park, Madagascar
December 2022
Thanks to our sponsors!

Lecture contributions from:
Tanjona Ramiadantsoa
Steve Bellan
Goals for this lecture

• To explain what we’re doing here
Goals for this lecture

• To explain what we’re doing here
• To define “science”
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
• To introduce the “E” in E²M²
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
• To introduce the “E” in E²M²
  • Ecology
Goals for this lecture

• To explain what we’re doing here
• To define ”science”
• To define ”data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
• To introduce the “E” in E²M²
  • Ecology
  • Epidemiology
All course materials are available at: E2M2.org
What is science?

the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts.

– Academic Press Dictionary of Science & Technology
Observations and Laws and Principles
Data and Models
Data and Models

• What is data?
Data and Models

• What is data?
  • Backbone of science
What is science?

the **systematic observation** of natural events and conditions in order to **discover** facts about them and to **formulate laws and principles**
Data and Models

• What are data?
  • Evidence to support a claim
Are these data?
Are these data?

What do we need to make these data?
Are these data?
Are these data?
Are these data?

Number of views for Justin Bieber’s “Baby” video on Youtube

[Graph showing the number of views over time from 1970 to 2010]
Are these data?

Total population size of China

Million

Year

Are these data?

Total population size of Madagascar

- Year
- Million
Are these data?

What are data?

• A relationship between at least two variables
  • x: explanatory, control, driver, independent variable(s)
  • y: response, dependent variable(s)
What are data?

• A relationship between at least two variables
  • x: explanatory, control, driver, independent variable(s)
  • y: response, dependent variable(s)
• x and y should be clearly defined
  • with respect to the question!
What are data?

• A relationship between at least two variables
  • x: explanatory, control, driver, independent variable(s)
  • y: response, dependent variable(s)
• x and y should be clearly defined
  • with respect to the **question**!
• Backbone of science
• **Evidence** to support a **claim**
Data provide evidence to support a claim.

**CLAIM:** The population size of Madagascar has increased throughout the past 50 years.
**Data: Sources of x and y**

- **Observational**
  - Just measure x and y

- **Experimental**
  - Interfere with x or the relationship between x and y
  - Create a relationship between x and y

- **Simulated**
  - [Diagram of simulated processes]

---

**Empirical data**

[Graph showing data analysis over time]
Data: Types

Numerical

Categorical
Data: Types

**Numerical**
- A variable is numerical when you can transform it with mathematical operation
- Examples?

**Categorical**

Data
Data: Types

**Numerical**
- A variable is numerical when you can transform it with mathematical operation
- Examples:
  - Integer, real number, multi-dimensional number

**Categorical**
Data: Types

Numerical
• A variable is numerical when you can transform it with mathematical operation
• Examples:
  • Integer, real number, multi-dimensional number

Categorical
• A variable is categorical when it is not numerical but a categorical can be numerical?
• Examples?
Data: Types

Numerical
• A variable is numerical when you can transform it with mathematical operation
• Examples:
  • Integer, real number, multi-dimensional number

Categorical
• A variable is categorical when it is not numerical but a categorical can be numerical?
• Examples:
  • Colors, (blood) types, species name
Data: Things to consider
Data: Things to consider

• Data acquisition
  • Impossible, example?
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?

• Data quality and quantity
  • In practice there is always a trade-off
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?

• Data quality and quantity
  • In practice there is always a trade-off
    • Example: monetary cost, human effort -> power analysis, sampling design etc.
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?

• Data quality and quantity
  • In practice there is always a trade-off
    • Example: monetary cost, human effort -> power analysis, sampling design etc.

• Reproducibility
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?

• Data quality and quantity
  • In practice there is always a trade-off
    • Example: monetary cost, human effort -> power analysis, sampling design etc.

• Reproducibility

• Measurement errors
Data: Things to consider

• Data acquisition
  • Impossible, example?
  • Theoretically possible but practically unfeasible, examples?

• Data quality and quantity
  • In practice there is always a trade-off
    • Example: monetary cost, human effort -> power analysis, sampling design etc.

• Reproducibility

• Measurement errors
  • Examples?
Data and Models
Data and Models

• What is a model?
What is science?

the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles.
Laws and Principles

• A theory = a declaration to explain a phenomenon
  • Logical and falsifiable
• A model = an abstract representation of a phenomenon
• A hypothesis = a testable declaration that is derived from a theory
Theory, Models, Hypotheses

- Theory
- Model
- Hypothesis

- General
- Specific
Models

- Human Ecology
- Evolution
- Car Ecosystem
- Climate Economy
- Economic systems
- Resource allocation
- Income - wages
- Financial sector
- Government sector
- Exports
- Income - exports
• When you make a model, you include the elements that you feel are most important to explain a phenomenon.
• When you make a **model**, you include the **elements that you feel are most important** to explain a phenomenon.

• Generally, we try to make **models** that can reproduce real-world **data**.
• When you make a model, you include the elements that you feel are most important to explain a phenomenon.

• Generally, we try to make models that can reproduce real-world data.

• In $E^2M^2$, we distinguish between statistical and mechanistic models.
Statistical vs. Mathematical Model

The choice depends on the research question!
Statistical Models

• Goal: To rigorously assess the strength of relationship between x and y
  • Find a significant relationship using a p-value as a measure of relationship strength
  • **Statistical models can demonstrate correlations.**
Statistical Models

• Goal: To rigorously **assess** the strength of relationship between x and y (describe patterns)
  • Find a significant relationship using a p-value as a measure of relationship strength
  • **Statistical models can demonstrate correlations.**

• Steps:
  1. Formulate a research question
  2. Formulate a hypothesis
  3. Develop a model to demonstrate your hypothesis.
  4. Collect **data (required!!!)**
  5. Evaluate hypothesis with appropriate statistical tools
     • t-test, Chi-square, ANOVA
     • Ordination (PCA)
     • Regression (LM, GLM, GLMM, GAM)
1. Example Question: **What** is the trajectory Malagasy population size through time?

Source: World Bank
1. Example Question: **What** is the trajectory of Malagasy population size through time?

2. Hypothesis: Malagasy population size increases with time
1. Example Question: **What** is the trajectory of Malagasy population size through time?

2. Hypothesis: Malagasy population size increases with time

3. Statistical Model: \( y = mx + b \)
   
   Linear Regression

Source: World Bank
1. Example Question: **What** is the trajectory of Malagasy population size through time?

2. Hypothesis: Malagasy population size increases with time

3. Statistical Model: 
   \[ y = mx + b \]
   Linear Regression

4. Data:

Source: World Bank
1. Example Question: **What** is the trajectory of Malagasy population size through time?

2. Hypothesis: Malagasy population size increases with time

3. Statistical Model:
   \[ y = mx + b \]

4. Data:
   \[ m = .372 \text{ million} \]
   \[ p = .003 \]

5. Evaluation

What can we conclude from this fitted model?
1. Example Question: **What** is the trajectory of Malagasy population size through time?

7. Adapt your model and re-evaluate:

\[ y = e^{mx+b} \]

Exponential Regression

\[ m = 0.029 \text{ mil.} \]

\[ p < .001 \]

What can we conclude from this fitted model?

Source: World Bank
Statistical Models: Beware!

• Statistical models and tests are based on specific assumptions
  • data normally distributed
  • y and y independent
  • etc.
Statistical Models: Beware!

• Statistical models and tests are based on specific assumptions
  • data normally distributed
  • y and y independent
  • etc.

• Assessing a model means you need to make sure the assumptions are not violated.
Statistical Models: Beware!

• Statistical models and tests are based on specific assumptions
  • data normally distributed
  • y and y independent
  • etc.

• Assessing a model means you need to make sure the assumptions are not violated.

• There are so many statistical models...
Statistical vs. Mathematical Model

The choice depends on the research question!
Mechanistic Models

• Goal: To **demonstrate the processes** that underlie a relationship between x and y
  • Find a significant relationship using a p-value as a measure of relationship strength
  • **Mechanistic models can demonstrate causation.**

• Steps:
  1. Formulate a research question
  2. Formulate a hypothesis
  3. Develop a model to demonstrate your hypothesis.
  4. Collect **data** (for certain questions)
  5. Evaluate the extent to which your model-simulated data matches that from the real world.
1. Example Question: **How** does Malagasy population size change with time?

2. Hypothesis: Malagasy population size increases because people are having children.

*Can you think of an alternative hypothesis?*

4. Data:

Source: World Bank
1. Example Question: How does Malagasy population size change with time?

2. Hypothesis: Malagasy population size increases because people are having children.

3. Mechanistic Model:

\[ P_{t+1} = P_t + b \times P_t - d \times P_t \]

\[ P_{t+1} = P_t + r \times P_t \]

Source: World Bank
1. Example Question: **How** does Malagasy population size change with time?

2. Hypothesis: Malagasy population size increases because people are having children.

3. Mechanistic Model:

```
| birth | Population | death |
```

4. Data:

```
<table>
<thead>
<tr>
<th>Year</th>
<th>Population (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>5</td>
</tr>
<tr>
<td>1980</td>
<td>10</td>
</tr>
<tr>
<td>1990</td>
<td>15</td>
</tr>
<tr>
<td>2000</td>
<td>20</td>
</tr>
<tr>
<td>2010</td>
<td>25</td>
</tr>
</tbody>
</table>
```

5. Evaluation:

\[ r = \frac{349}{person/yr} \]

**What can we conclude from this fitted model?**
Mechanistic Models: Beware!

- Parameters used in the mechanistic models sometimes are not measurable!
Mechanistic Models: Beware!

- Parameters used in the mechanistic models sometimes are not measurable!
- Simulations can be computationally intensive
Mechanistic Models: Beware!

• Parameters used in the mechanistic models sometimes are not measurable!
• Simulations can be computationally intensive
• Advances in computational power often inspire development of more complex models which are not necessarily better
Mechanistic Models: Beware!

• Parameters used in the mechanistic models sometimes are not measurable!
• Simulations can be computationally intensive
• Advances in computational power often inspire development of more complex models which are not necessarily better

“All models are wrong but some are useful…”

-George Box
Mechanistic Models: Beware!

• Parameters used in the mechanistic models sometimes are not measurable!
• Simulations can be computationally intensive
• Advances in computational power often inspire development of more complex models which are not necessarily better

“All models are wrong but some are useful...”
-George Box

We use models to both predict and explain.
It is ideal when statistical and mechanistic models meet:
A Tool for E²M²

- Computer power keeps increasing
- Language/software
  - Fortran, C, C++
  - Julia, Java, Python
  - Matlab, Maple, Mathematica,
    - SAS, SPSS, Stata
- Specific programs
  - Vortex, RAMAS, NetLogo for IBM
  - NicheMapper for physiology, iLand for forest dynamics
  - MaxEnt for species distribution modeling
  - Zonation for reserve selection etc...
- The compromise: R---very powerful for
  - Visualization
  - Data formatting and sorting
  - Statistical analyses
  - Simulation (mechanistic model)
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
• To introduce the “E” in E²M²
  • Ecology
  • Epidemiology
Goals for this lecture

• To explain what we’re doing here
• To define “science”
• To define “data”
• To define “models”
• To introduce many different types of models
  • Statistical
  • Mathematical
• To introduce the “E” in E²M²
  • Ecology
  • Epidemiology
What is Epidemiology?

• “the study of what is on the people”
  – coined in 1802 to describe diseases in the Spanish population
• Emphasis on the study and analysis of the distribution and determinants of health and disease (“risk factors”)
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)

2. Miasmatic theory of disease (1500s)
   • Sickness results from emanations of ‘bad air’
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
2. Miasmatic theory of disease (1500s)
   • Sickness results from emanations of ‘bad air’
3. Germ theory of disease
   • Leeuwenhoek’s microscope (1675)
   • Koch’s postulates (1890)
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)

2. Miasmatic theory of disease (1500s)
   • Sickness results from emanations of ‘bad air’

3. Germ theory of disease
   • Leeuwenhoek’s microscope (1675)
   • Koch’s postulates (1890)

4. Classical epidemiology
   • John Snow and London cholera (1854)
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)

2. Miasmatic theory of disease (1500s)
   - Sickness results from emanations of ‘bad air’

3. Germ theory of disease
   - Leeuwenhoek’s microscope (1675)
   - Koch’s postulates (1890)

4. Classical epidemiology
   - John Snow and London cholera (1854)

5. Mathematical epidemiology
   - Kermack and McKendrik (1927)
What is Ecology?

• The study of the interactions of organisms and their environment
  - Coined in 1866 by German scientist Ernst Haeckel
  - Nile crocodiles opening mouths for sandpipers (Herodotus)

• Emphasis on explaining dynamical processes in nature
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)

2. Food Chains
   - Charles Elton (1920s)
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)

2. Food Chains
   - Charles Elton (1920s)

3. Predator-Prey Dynamics
   - Lotka-Volterra (1920s)
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)

2. Food Chains
   - Charles Elton (1920s)

3. Predator-Prey Dynamics
   - Lotka-Volterra (1920s)

4. Mathematical Ecology
   - Robert MacArthur (1950s)
   - Island biogeography
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)

2. Food Chains
   - Charles Elton (1920s)

3. Predator-Prey Dynamics
   - Lotka-Volterra (1920s)

4. Mathematical Ecology
   - Robert MacArthur (1950s)
   - Island biogeography

5. Disease Ecology
   - Anderson and May (1980s)
   - Island biogeography

\[ \text{Susceptible} \xrightarrow{\beta I/N} \text{Infectious} \xrightarrow{\gamma} \text{Recovered} \]
Misaotra!
International Clinics on Infectious Disease, Dynamics, & Data

MMED: *Clinic on the Meaningful Modeling of Epidemiological Data*

May-June 2023, Cape Town, South Africa

DAIDD: *Clinic on Dynamical Approaches to Infectious Disease Data*

December 2023, Virtual

South African Center for Epidemiological Modeling and Analysis (SACEMA), Director
Dr. Juliet Pulliam
*University of Stellenbosch*

www.ici3d.org