Ryan Williams
Post(Ghost)-doctoral Researcher
Agricultural and Biosystems Engineering
Iowa State University

# Modeling and Simulating Data for a Better Understanding of Microbial Ecology

Ryan J. Williams

# What is a model?

- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

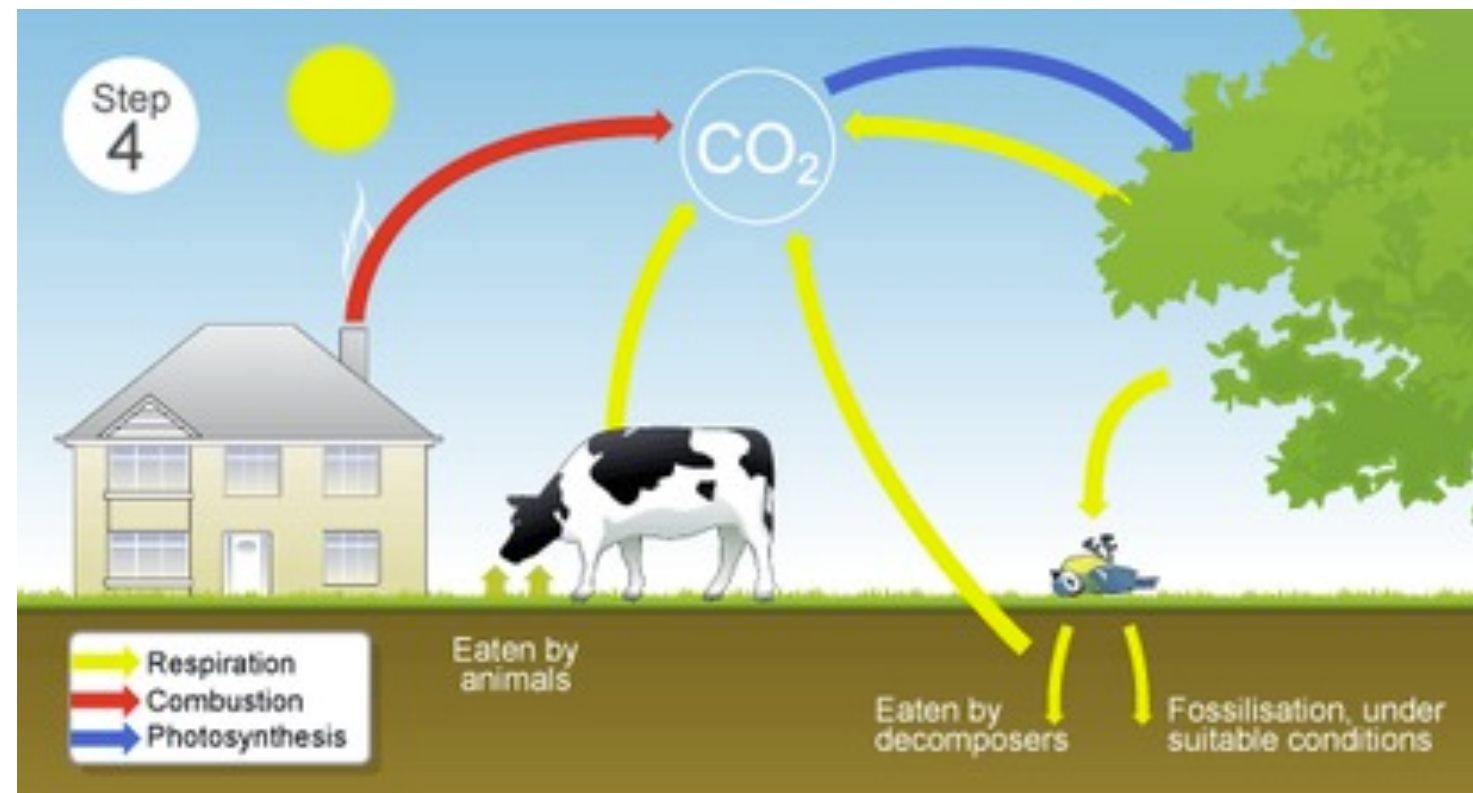- Can be mathematical

- Can be statistical

# What is a model?

- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical

# What is a model?

- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical

# What is a model?

- Conceptualization of how we understand phenomena in nature.

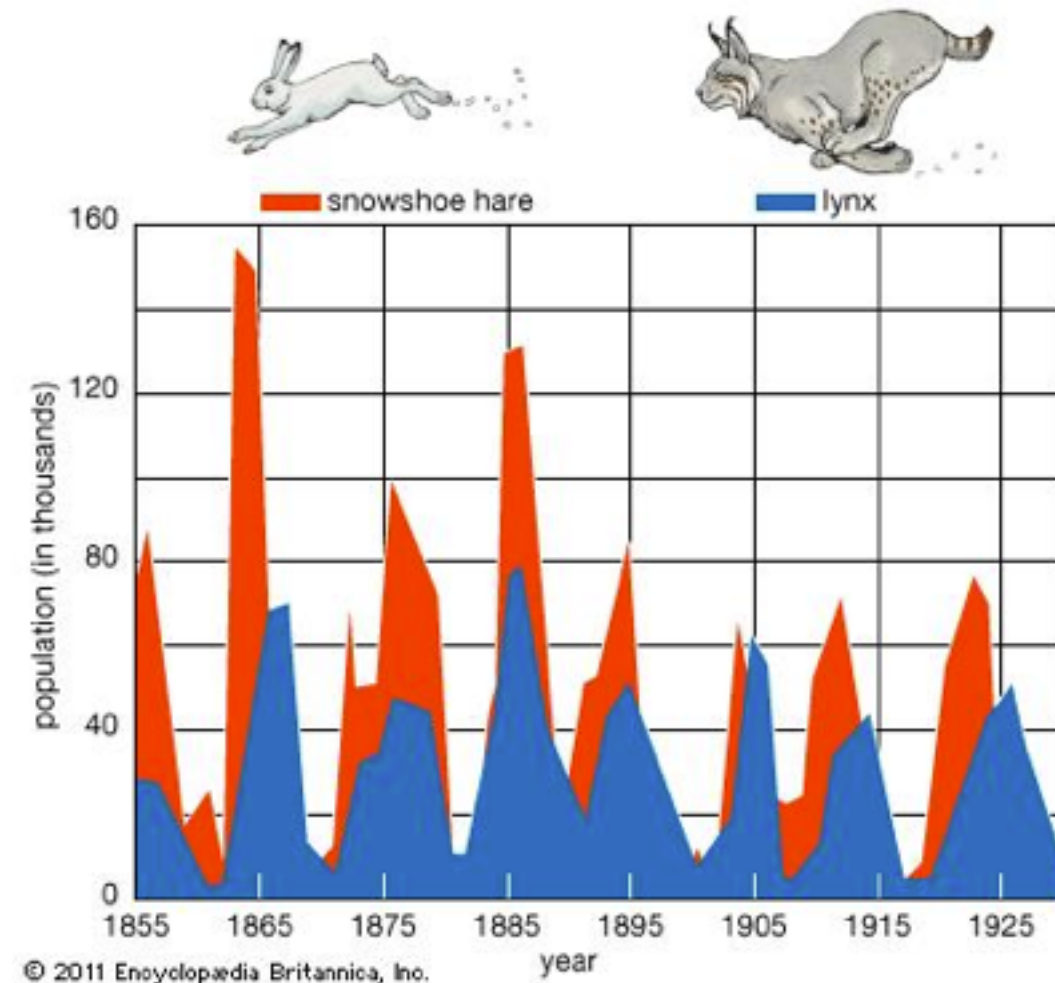- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical



snowshoe hare    lynx

160
120
80
40
0

population (in thousands)

1855  1865  1875  1885  1895  1905  1915  1925

year

© 2011 Encyclopaedia Britannica, Inc.

# What is a model?

- Conceptualization of how we understand phenomena in nature.

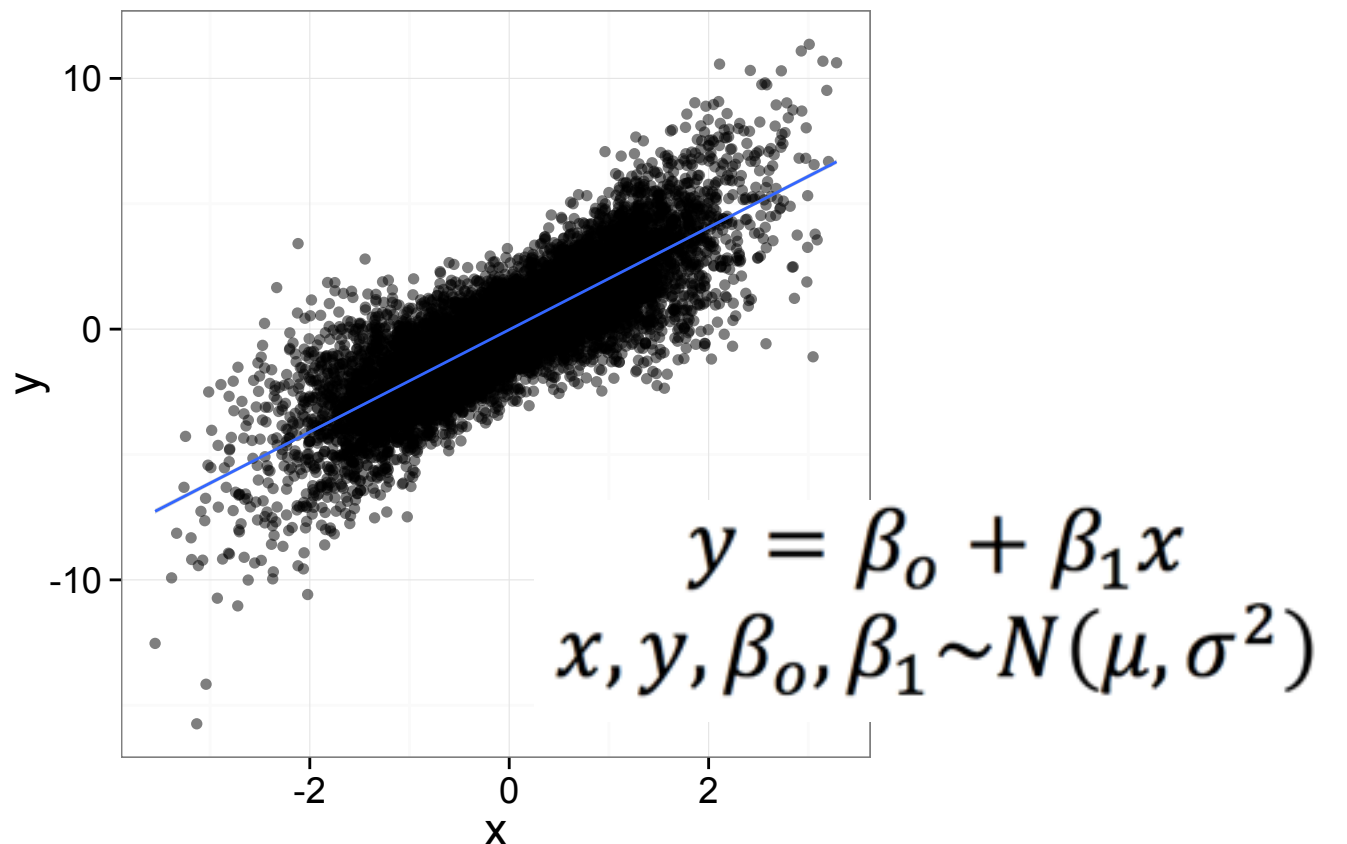- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical



$$y = \beta_o + \beta_1 x$$
$$x, y, \beta_o, \beta_1 \sim N(\mu, \sigma^2)$$
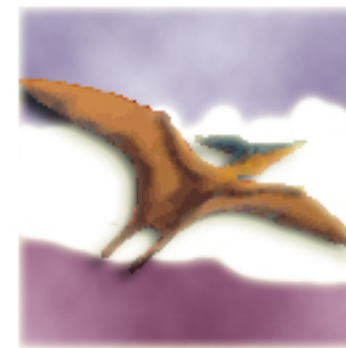
# What is simulation?

- Producing information to mimic something we're interested in

  - Simulating sequences

  - Simulating numerical data

# What is simulation?

- Producing information to mimic something we're interested in

  - Simulating sequences

  - Simulating numerical data

# What is simulation?

- Producing information to mimic something we're interested in

  - Simulating sequences

**MetaSim**

Daniel H. Huson and Felix Ott

with contributions from:
R. Schmid, A.F. Auch and D.C. Richter

www-ab.informatik.uni-tuebingen.de/software/metasim

OPEN ACCESS Freely available online

**PLOS** | ONE

## NeSSM: A Next-Generation Sequencing Simulator for Metagenomics

Ben Jia[1,9], Liming Xuan[3,4,9], Kaiye Cai[4,9], Zhiqiang Hu[2,4], Liangxiao Ma[4],

1 School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, 2 Department of Bioinformat
Biotechnology, Shanghai Jiao Tong University, Shanghai, China, 3 School of Bioengineering, East China University of Sci
Center for Bioinformation Technology, Shanghai, China

Johnson et al. BMC Bioinformatics 2014, 15(Suppl 9):S14
http://www.biomedcentral.com/1471-2105/15/S9/S14

**BMC Bioinformatics**

**PROCEEDINGS**

**Open Access**

## Grinder: a versatile amplicon and shotg sequence simulator

Florent E. Angly[1,*], Dana Willner[1,2], Forest Rohwer[3], Philip Hugenh
Gene W. Tyson[1,5]

## A better sequence-read simulator program for metagenomics

Stephen Johnson[1*], Brett Trost[1], Jeffrey R Long[1], Vanessa Pittet[2], Anthony Kusalik[1]

From RECOMB-Seq: Fourth Annual RECOMB Satellite Workshop on Massively Parallel Sequencing
Pittsburgh, PA, USA. 31 March - 05 April 2014

# What is simulation?

- Producing information to mimic something we're interested in

  - Simulating sequences

- Simulating numerical data

# What is simulation?

- Producing information to mimic something we're interested in

  - Simulating sequences

  - Simulating numerical data
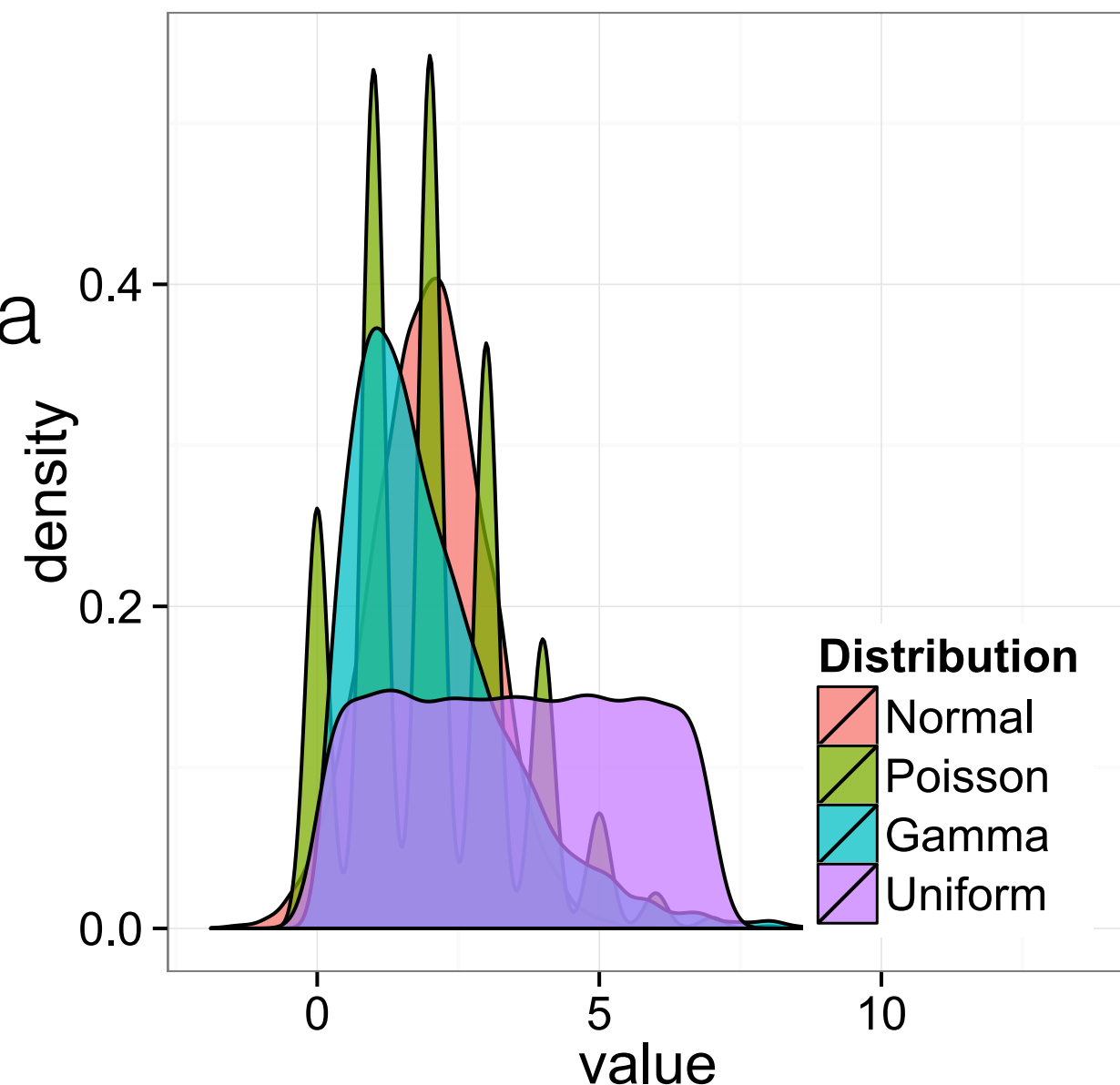
# What's the use?

- Microbes are hard to observe in nature

- Microbial data is expensive (relatively, it depends on who you talk to)

- Microbial data is highly multivariate

- Microbial ecologists love to use a variety of methods and pre-packaged tools that are not necessarily easy to understand

- Numerical simulation is universal; simulating sequences is applied (both can be very slick but simple tools!)