Inferring moth emergence from abundance data: A novel mathematical approach using birth-death contingency tables

Dan Sheldon, Evan Goldman, Erin Childs, Olivia Poblacion, Jeffrey C. Miller, Julia A. Jones and **Thomas G. Dietterich**
Oregon State University

LTER

HJ Andrews Experimental Forest
Long Term Ecological Research

OSU
Oregon State
UNIVERSITY

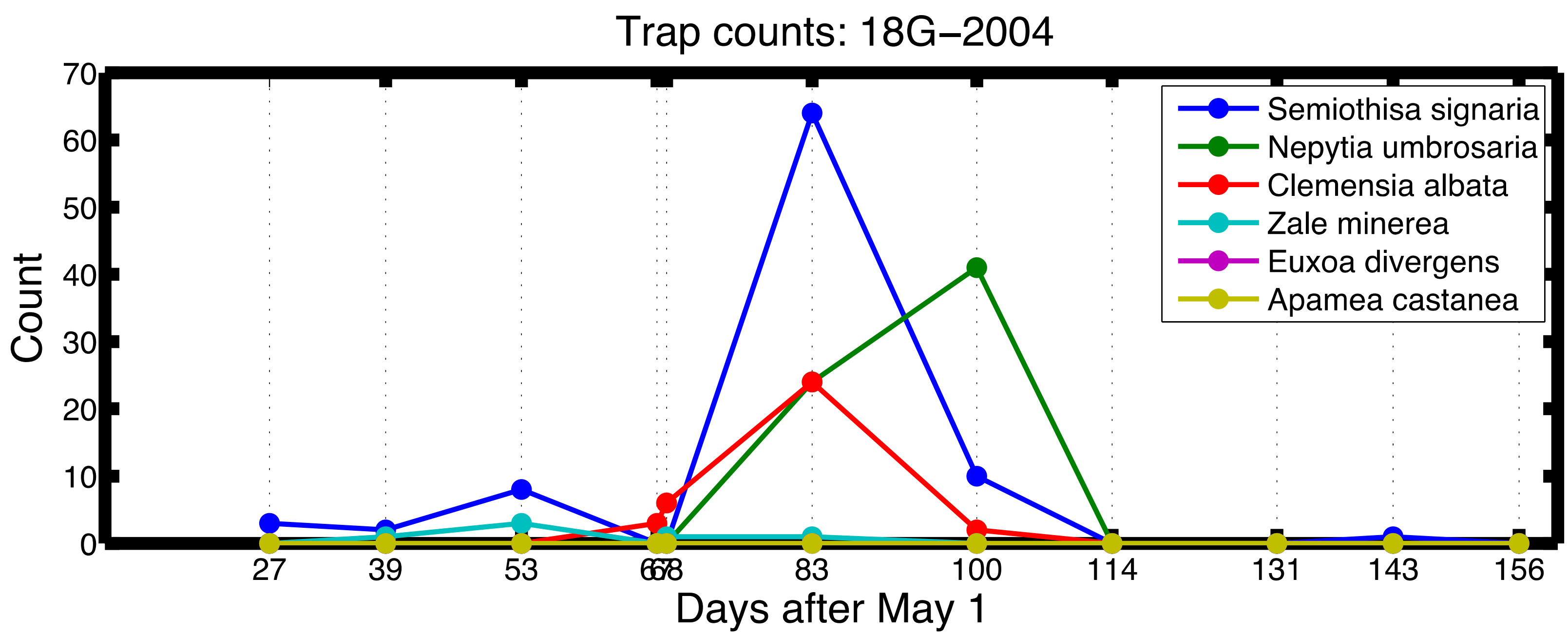
A Common Problem in Ecology

- *What we have*: periodic observations of organism "activity"
  - Moth trap counts
  - Bird surveys

- *What we want*: timing of life history events
  - When did adult moths emerge from cocoons?
  - When did migrating birds arrive?

- How to bridge the gap?

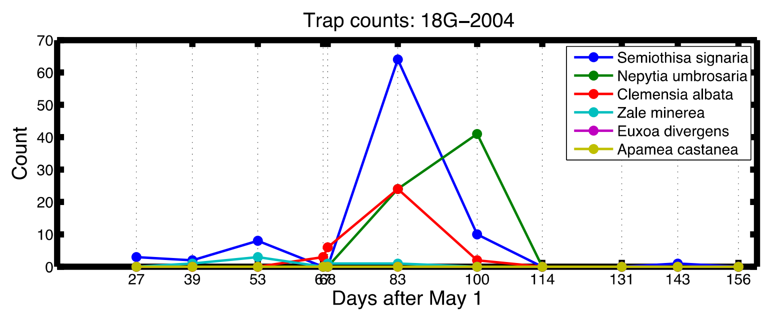
Example: Moth Trap Counts



What was the flight period of Nepytia umbrosaria in 2004?

Challenges

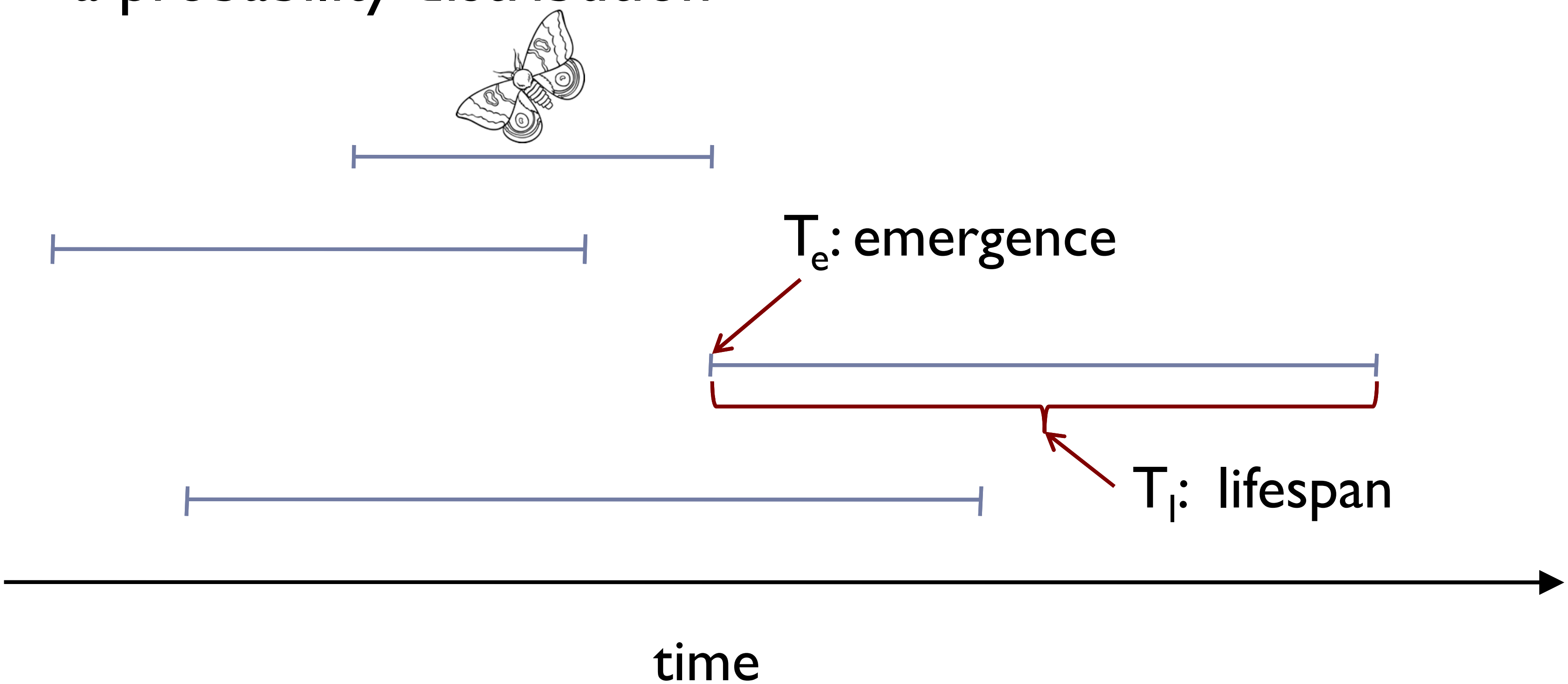
- We do not *directly* observe the events we are interested in
  - Moth emergence
  - Bird arrival
- Surveys are infrequent
  - May miss "peak" activity



- Naïve approaches don't use all of the data
  - Date of first moth
  - Date of maximum abundance

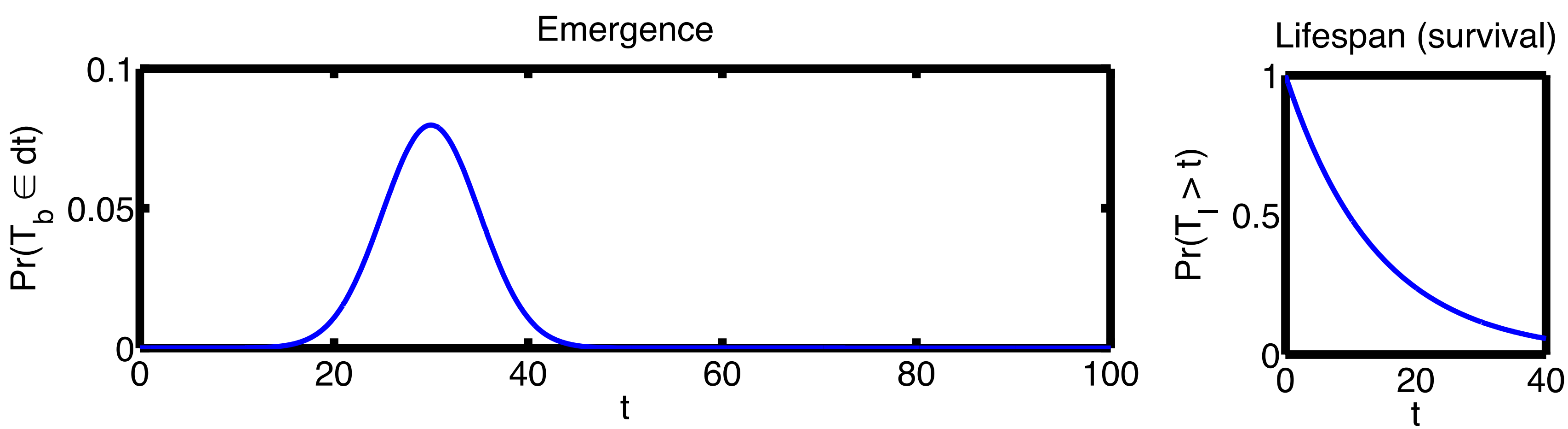
Probabilistic Modeling Approach

- Due to Zonneveld (1991), Manley (1974)
- Assume moths are independent and identical draws from a probability distribution



Choose Lifetime Distributions

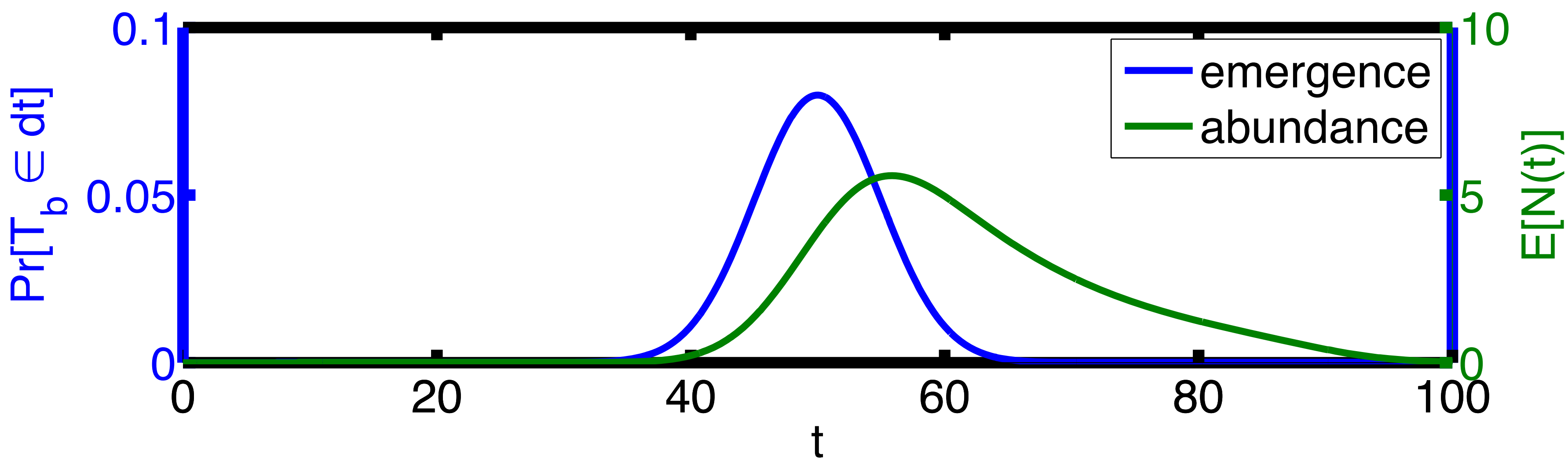
- Emergence date is $\mathrm{Normal}(\mu, \sigma^2)$
- Lifespan is $\mathrm{Exponential}(\lambda)$



Note: *any* other parametric models can work
(not restricted to Zonneveld's Logistic distribution)

Abundance

- Emergence and lifespan induce a model of abundance



- Goal: fit parameters (μ, σ^2, λ)

Zonneveld (1991)

- Compute expected number of moths $X(t)$ flying at time t by solving

$$\frac{dx(t)}{dt} = N^* f(t;\mu,\sigma^2) - \lambda x(t)$$

- Assume Poisson distributed observation counts

$$y(t) \sim \text{Poisson}(x(t))$$

- Implemented in INCA (www.urbanwildlands.org/INCA/)
  - Bruggeman, Longcore & Zonneveld

Differences between Transect Counts and Moth Trapping

- Traps are lethal
  - requires change in the likelihood and survival functions
  - eliminates issues of double counting (either within a single night or across multiple nights)
- Traps are very effective
  - Less problem with detection rates
  - Nonetheless, we still include detection rate in our model

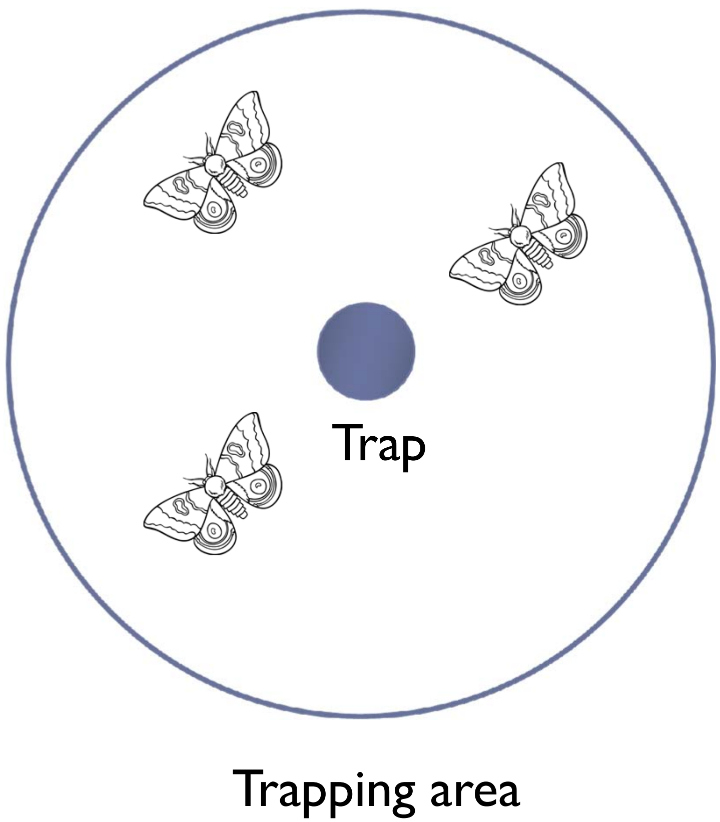
Limitations of the Zonneveld Method

- Differential equation approach makes it difficult to change the emergence distribution
- Likelihood requires an approximation (i.e., Poisson)

Our Contributions

- General approach that can work with any emergence and mortality distributions
  - Only requirement is the ability to evaluate a double integral over the convolution of the two distributions
    - Survival functions for emergence and mortality with weather covariates
    - Overlapping generations
  - Can model lethal and non-lethal trapping
- Exact likelihood rather than an approximation
  - Our likelihood converges to the Zonneveld likelihood in the limit where trapping probability goes to zero

Trapping model



- Closed trapping area
- N moths (unknown) emerge during season
- Trap dates $t_1, \ldots, t_k$
- Trapping probability α for each moth alive on trap date

Data Likelihood

- Can categorize each moth by its possible fates

Outcome	Probability	Trap Count
Trapped t_1	q_1	f_1
Trapped t_2	q_2	f_2
…	…	…
Trapped t_k	q_k	f_k
Not trapped	$r = 1 - (q_1 + \cdots + q_k)$	$U = N - (f_1 + \cdots + f_k)$

- Thus, likelihood of observed trap counts $f_1, \ldots, f_k$ is Multinomial:

$$(f_1, \ldots, f_k, U) \sim \text{Multinomial}(N; q_1, \ldots, q_k, r)$$

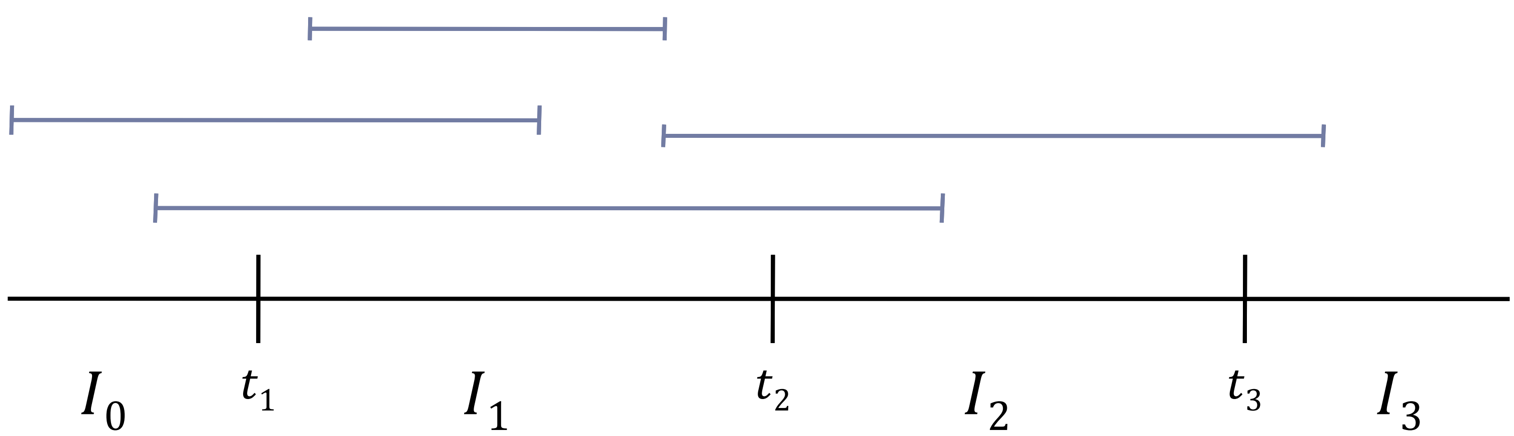
Likelihood Computation and Model Fitting

- To compute likelihood:
  - Compute birth-death table $P(i, j)$ from emergence and lifespan distributions (numerical integration)
  - Compute birth-trap table $Q(i, j)$
  - Compute trapping probabilities $q_1, \ldots, q_k$

- Find (μ, σ^2, λ) to maximize likelihood with numerical optimizer

Intervals

- Trapping times $t_1, \dots , t_k$
- Intervals $I_0, \dots , I_k$
- Trap counts $f_1, \dots , f_k$



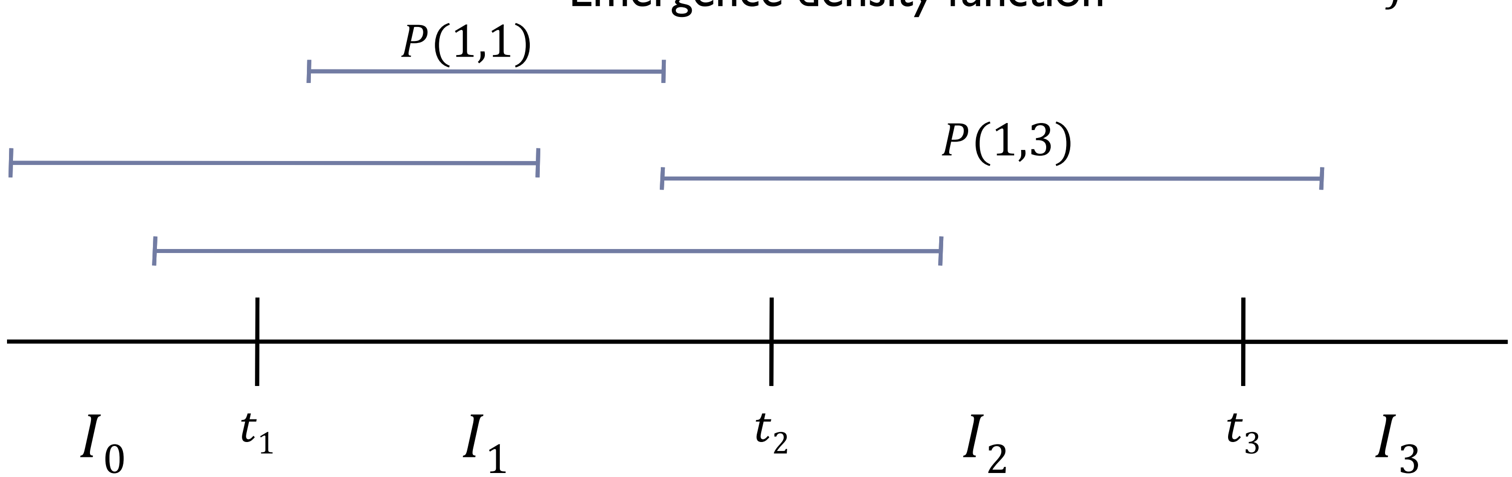
Birth-Death Table

- Let $P(i,j)$ be probability a moth is born in I_i and dies in I_j

$$P(i,j) = \int_{t_i}^{t_{i+1}} f(s;\mu,\sigma^2)\Pr[t_j \leq s + T_l \leq t_{j+1}]\, ds$$

Emergence density function

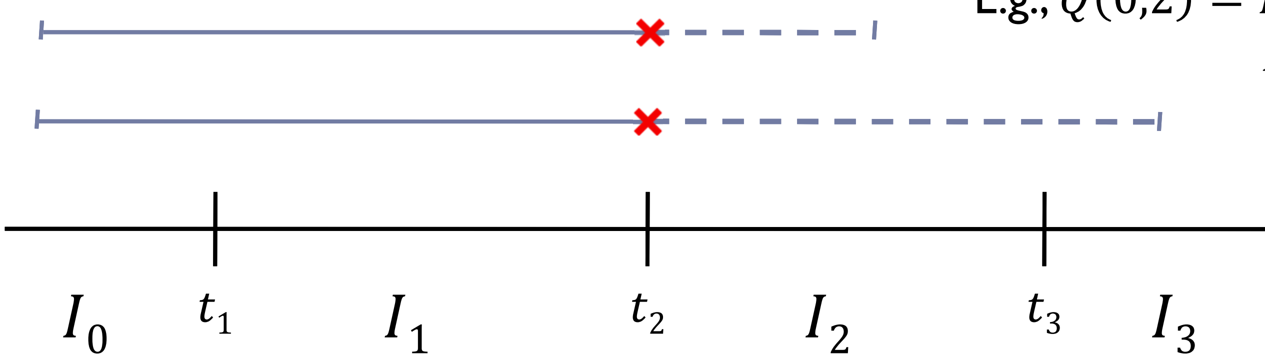
Dies in j



Birth-Trap Table

- Let $Q(i,j)$ be probability a moth is born in I_i and trapped at t_j

$$Q(i,j) = \sum_{k \geq j} P(i,k)(1-\alpha)^{j-i-1}\alpha$$



E.g., $Q(0,2) = P(0,2)(1-\alpha)\alpha + P(0,3)(1-\alpha)\alpha$

Trap Probabilities

- Overall probability of being trapped at t_j

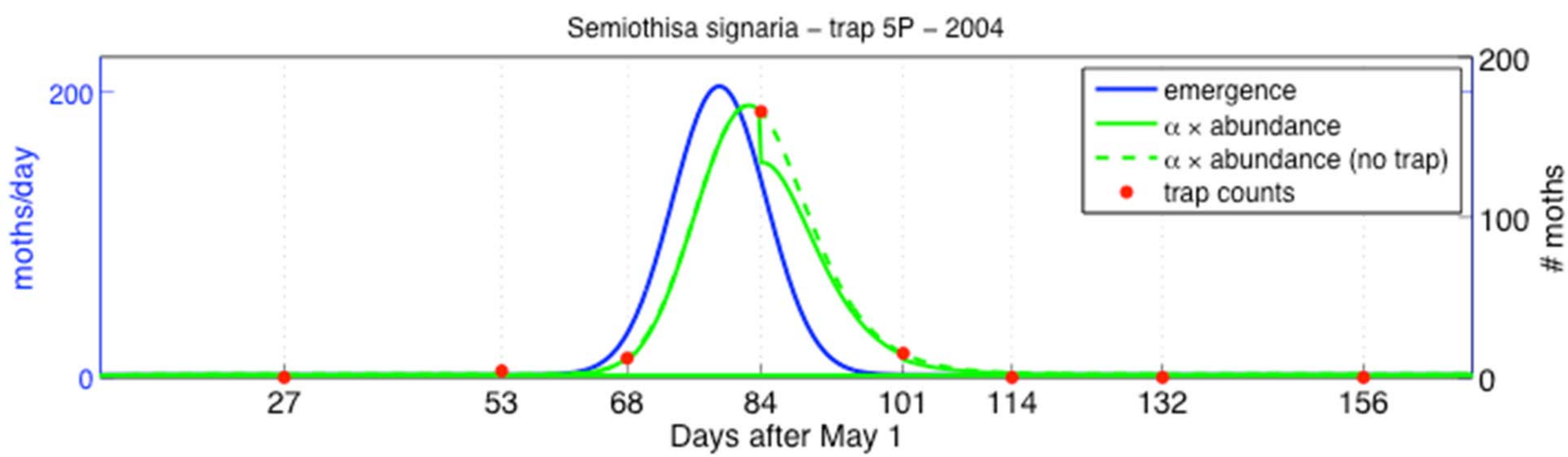
$$q_j = \sum_{i<j} Q(i,j)$$

- Probability not trapped at all

$$r = 1 - \sum_j q_j$$

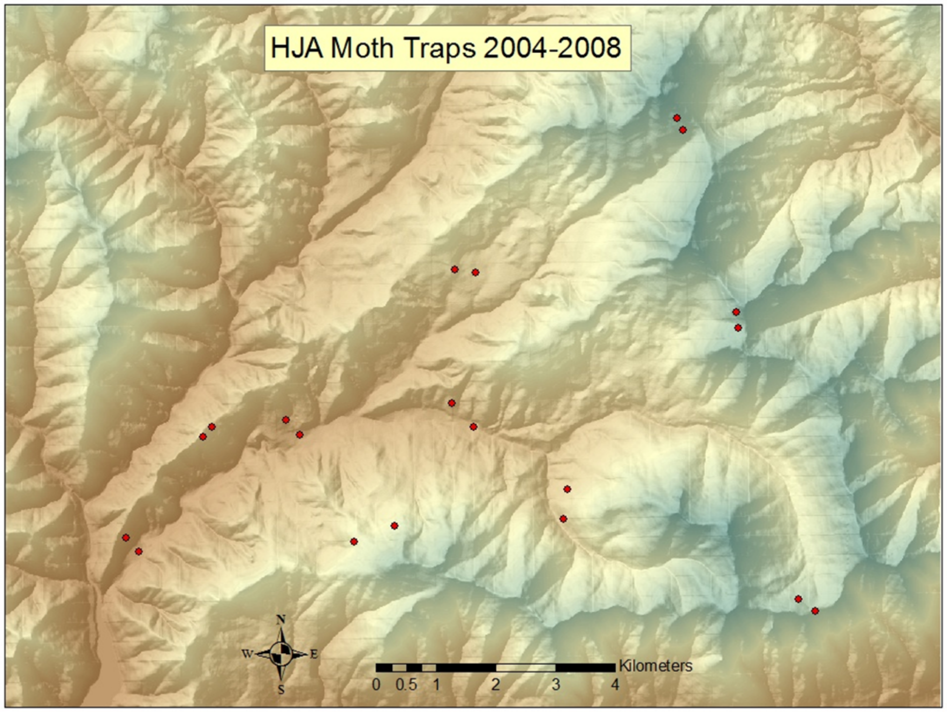
Results

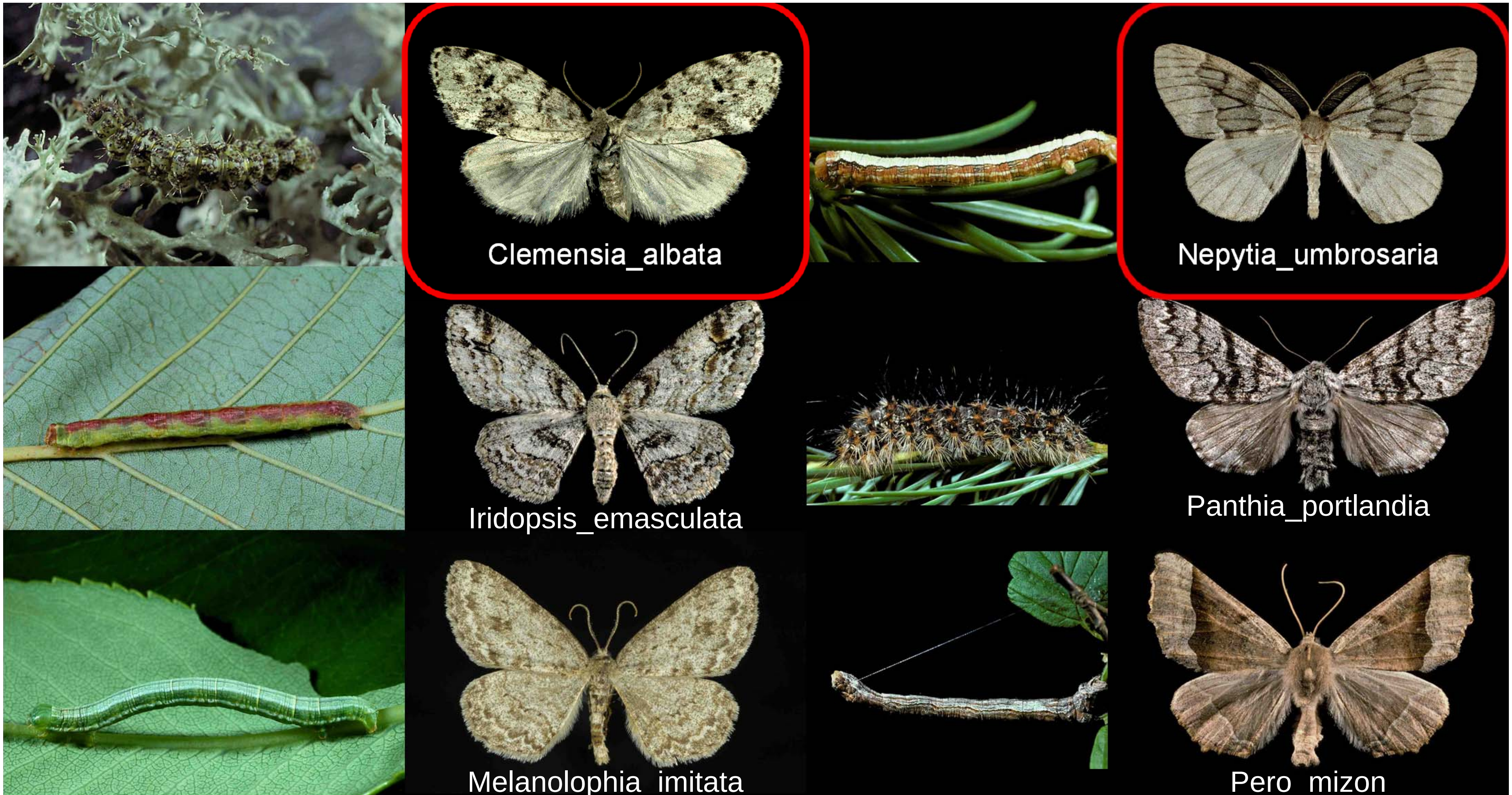
- Example of fitted model



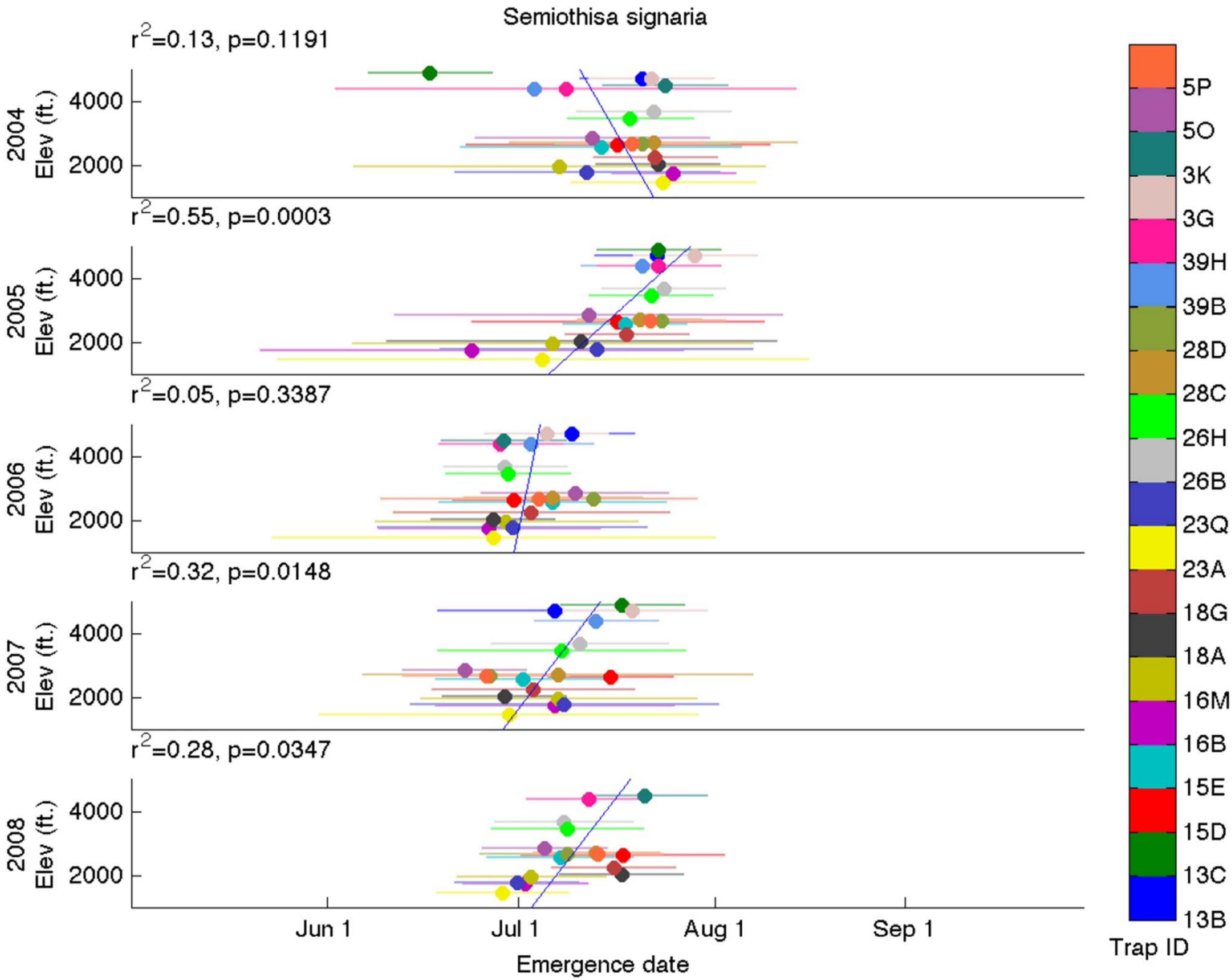
Data

- Moth trapping at HJA, 2004-2008
- 20 paired sites
- 10 trapping dates/year
- Approx. 2 week intervals
- > 500 species

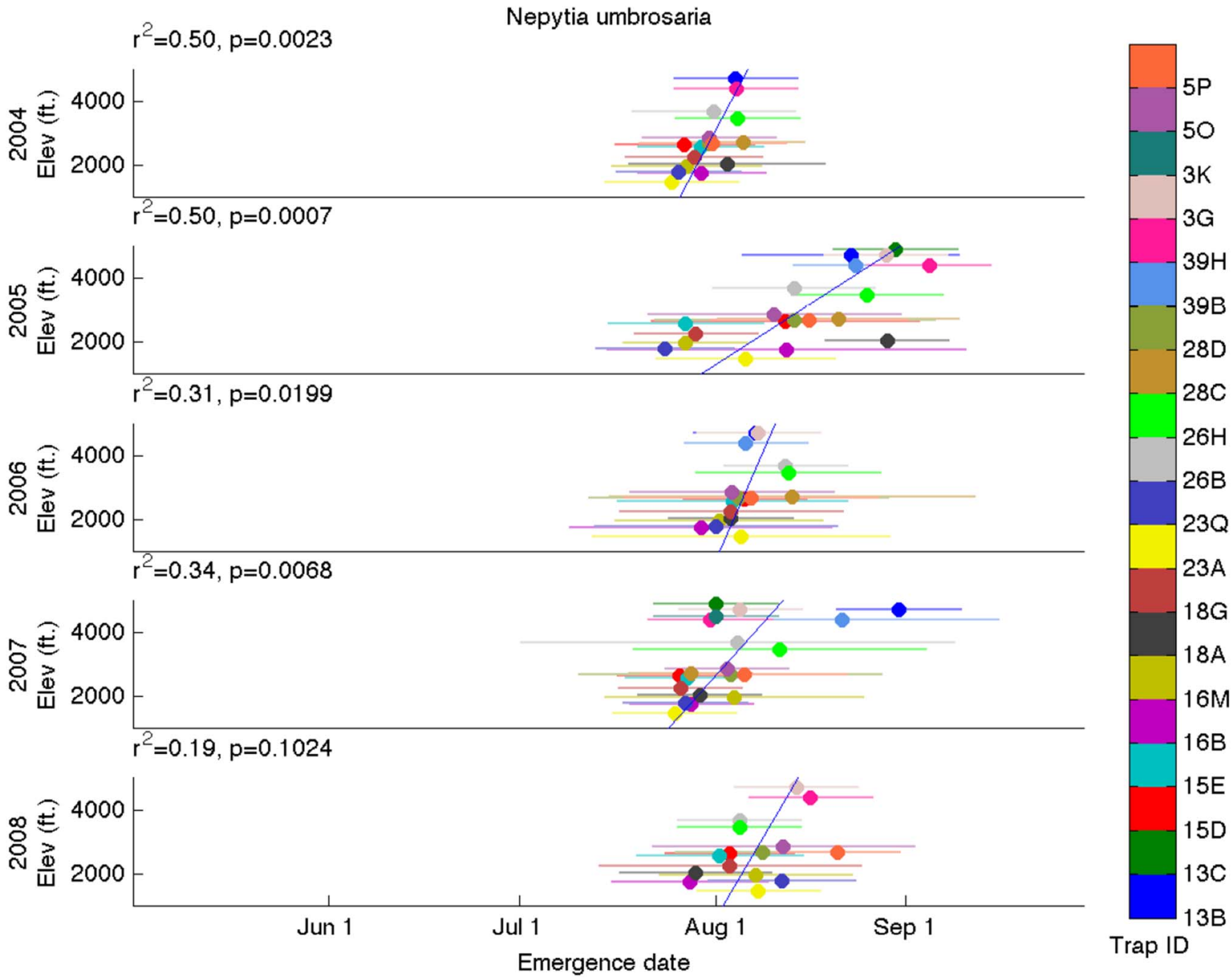


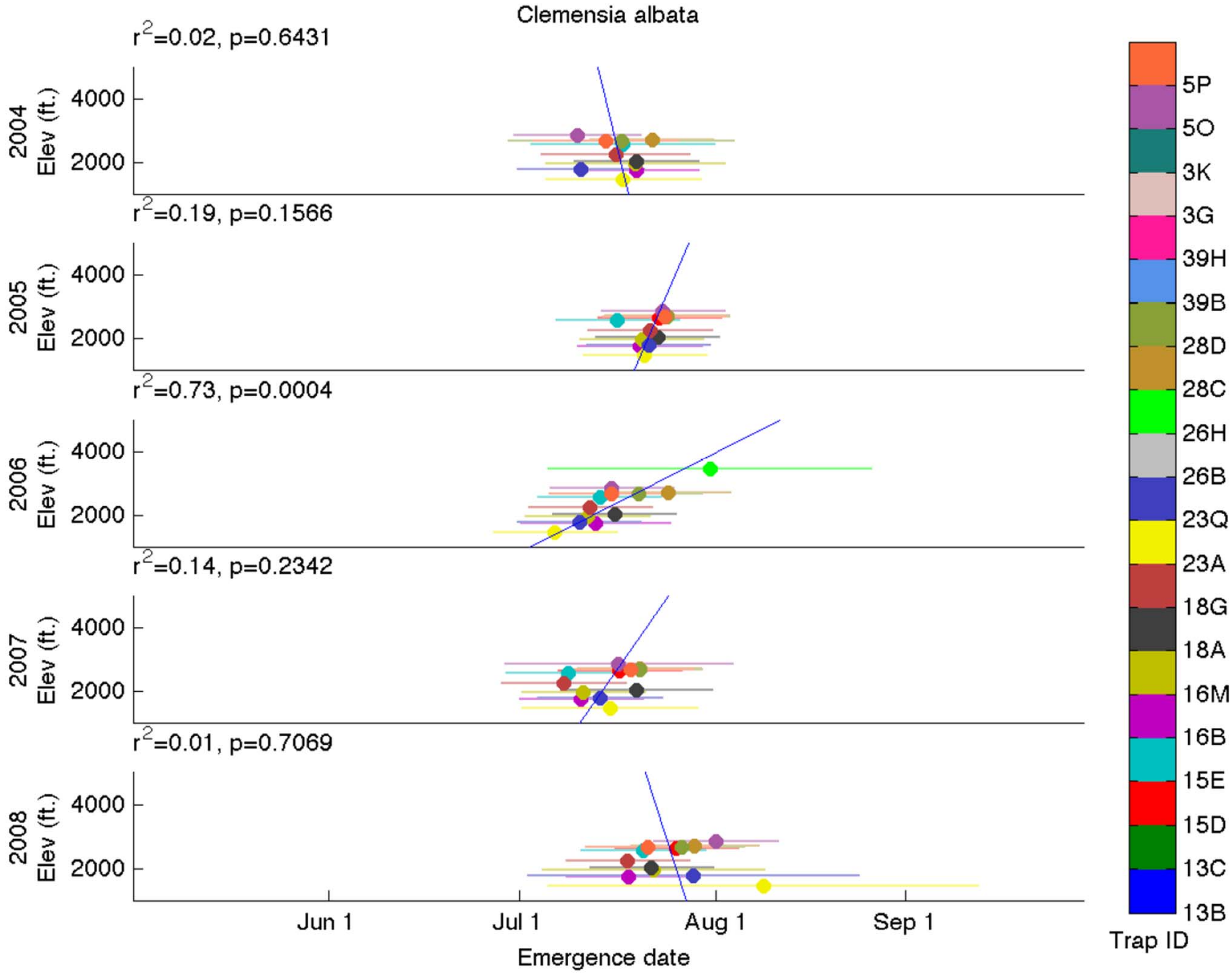


Seven of the eight most common moths in the HJ Andrews (photos by Jeff Miller)



Semiothisa signaria
r2=0.13, p=0.1191
r2=0.55, p=0.0003
r2=0.05, p=0.3387
r2=0.32, p=0.0148
r2=0.28, p=0.0347
2004
2005
2006
2007
2008
Elev (ft.)
4000
2000
Jun 1
Jul 1
Aug 1
Sep 1
Emergence date
5P
5O
3K
3G
39H
39B
28D
28C
26H
26B
23Q
23A
18G
18A
16M
16B
15E
15D
13C
13B
Trap ID


Nepytia umbrosaria
r2=0.50, p=0.0023
r2=0.50, p=0.0007
r2=0.31, p=0.0199
r2=0.34, p=0.0068
r2=0.19, p=0.1024
2004 Elev (ft.)
2005 Elev (ft.)
2006 Elev (ft.)
2007 Elev (ft.)
2008 Elev (ft.)
4000
2000
Jun 1
Jul 1
Aug 1
Sep 1
Emergence date
5P
5O
3K
3G
39H
39B
28D
28C
26H
26B
23Q
23A
18G
18A
16M
16B
15E
15D
13C
13B
Trap ID



Clemensia albata
r2=0.02, p=0.6431
r2=0.19, p=0.1566
r2=0.73, p=0.0004
r2=0.14, p=0.2342
r2=0.01, p=0.7069
2004
2005
2006
2007
2008
Elev (ft.)
4000
2000
Jun 1
Jul 1
Aug 1
Sep 1
Emergence date
5P
5O
3K
3G
39H
39B
28D
28C
26H
26B
23Q
23A
18G
18A
16M
16B
15E
15D
13C
13B
Trap ID

Summary

- Generalization of the Zonneveld model
  - Works with any parametric birth (arrival) and death (departure) processes
  - Provides exact likelihood

- At HJA, often significant correlation between elevation and moth emergence, but also unexplained variability in this pattern

- Future work:
  - Incorporate environmental covariates into model (e.g., degree days)
  - Explore model limitations via simulation study
  - Obtain confidence intervals on parameters
  - Extend to other phenology questions: Bird Migration

Thank-you to Nick Haddad for introducing us to the Zonneveld Model

NSF LTER; NSF Ecoinformatics IGERT (Jones); NSF Computational Sustainability Expedition (Dietterich); NSF Ecoinformatics Summer Institute; NSF BIO Postdoc (Sheldon)