

A adaptive feedback methodology for determining
information content in demographic data for
ecotoxicology studies

H. T. Banks

Center for Research in Scientific Computation
N. C. State University
Raleigh NC

NC STATE UNIVERSITY

*Center for Research
in Scientific Computation
North Carolina State University*

Acknowledgments

This represents joint work with

Rebecca A. Everett at North Carolina State University,
J. E. Banks at California State University, Monterey Bay, and
J. D. Stark at Washington State University, Puyallup

An adaptive feedback methodology for determining information content in population studies, (HTB, J.E. Banks, Rebecca Everett, and John Stark), CRSC-TR15-12, Center for Research in Scientific Computation, N. C. State University, Raleigh, NC, November, 2015; *Mathematical Biosciences and Engineering*, to appear.

Ecotoxicology and Ecological Risk Assessments

- Pesticides protect crops and improve their productivity by harming or repelling target species



ipm.ucdavis.edu

- Can also present a risk to non target species and ecosystems (e.g., spray-drift, leaching, and runoff)
- Ecological risk assessments (ERAs) are used to evaluate the effects of a chemical on an environment

- Ecotoxicology combines ecology and toxicology to study of the effects of toxic substances on biological organisms and then environment
- ERAs are often based on individual measurements, but protection goals are in terms of population levels
- Demographic analysis: predictions about population growth in response to disturbances based upon life tables
- Collection of demographic data can be time consuming and costly

Can partial demographic data replace full demographic data while still providing an accurate picture of the impact of a toxicant on a population? (Stark and J. Banks, 2015)

Simple Logistic Growth Model

We want to develop a predictive mathematical methodology for determining when enough data has been collected while still obtaining an accurate estimate of the endpoint of interest.

Consider the well-known logistic growth model

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K} \right),$$

where $x(t)$ represents the population size at time t , K represents the carrying capacity, and r represents the growth rate.

The exact solution of this model is given by

$$x(t) = \frac{K}{1 + \left(\frac{K}{x_0} - 1 \right) e^{-rt}}, \quad (1)$$

where x_0 represents the initial population size.

Simple Logistic Growth Model

We first created simulated data $\{y_j\}_{j=1}^n$ using the “true” or nominal parameter set $\theta_0 = [K, r] = [17.5, .7]$ and $x_0 = 0.1$ and added Gaussian noise. Thus we have the following realization of an absolute error statistical model given by

$$y_j = f(t_j, \theta_0) + \epsilon_j, j = 1, \dots, n,$$

where $f(t_j, \theta) = x(t_j)$.

We estimate the parameters $\theta = [K, r]$ by minimizing the cost functional

$$J_n(\theta) = \sum_{j=1}^n [y_j - f(t_j, \theta)]^2$$

with respect to θ . We provide an initial guess $\theta^0 = [8, 1]$ and obtain the minimized cost $J_n(\hat{\theta})$, where $\hat{\theta} = \hat{\theta}^n$ represents the optimized value of θ using n number of data points.

Simple Logistic Growth Model

In order to calculate the standard errors for each estimated parameter, we must first calculate the sensitivity matrix χ , where

$$\chi^T = \begin{pmatrix} \frac{\partial x(t_1)}{\partial K} & \cdots & \frac{\partial x(t_n)}{\partial K} \\ \frac{\partial x(t_1)}{\partial r} & \cdots & \frac{\partial x(t_n)}{\partial r} \end{pmatrix}.$$

The standard errors are then calculated using

$$SE_j = \sqrt{\hat{\sigma}^2 (\chi^T \chi)^{-1}_{jj}}, j = 1 \dots n,$$

where

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{j=1}^n [y_j - f(t_j; \hat{\theta}^n)]^2,$$

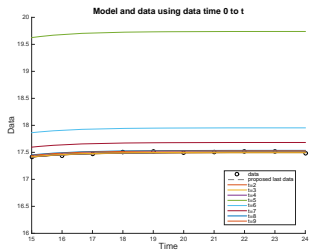
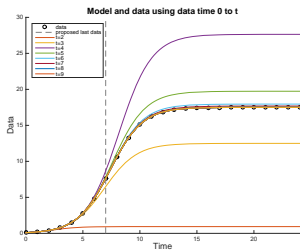
and p is the number of parameters being estimated.

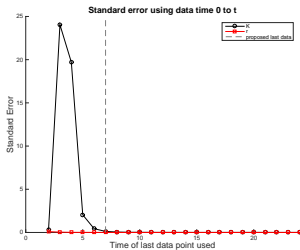
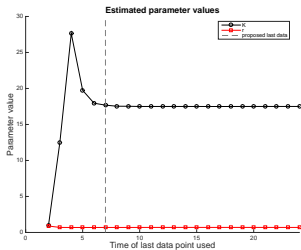
We repeated this process for each increasing number of data points.

Our *adaptive feedback methodology* is summarized in the following steps. *Each time data are collected*

- 1 Perform an inverse problem using all the available data to obtain the estimated parameter(s).
- 2 Calculate the *associated standard error(s)*.
- 3 Calculate the *percent change in parameter values*.
- 4 Calculate the *percent change in standard errors*.
- 5 Stop collecting data once the percent change in parameter values decreases below a threshold *and* the percent change in standard errors decreases below a threshold.

Simple Logistic Growth Model





Adaptive feedback methodology on the parameter λ

Based on the results from our simple example using the logistic population growth model, we investigate the minimum amount of daphnia population data needed by estimating λ , the parameter of interest, and by calculating the standard errors for an increasing number of data; once the estimated parameter values and standard errors level off, we will deem that sufficient data will have been collected. We can determine this by using the methodology described above. We first calculate the **percent change in parameter values** for increasing n using equation

$$\frac{P_n - P_{n-1}}{P_{n-1}} 100,$$

and determine when this reaches below a threshold. We then repeat this process for **percent change in standard error** using equation

$$\frac{|s_n - s_{n-1}|}{\max_{j=1, \dots, n} s_j} 100$$

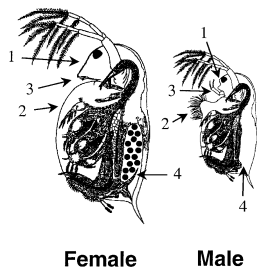
Why Do We Care About *Daphnia magna*?

- *Daphnia* is a water flea used as a model organism in many scientific investigations.
- *Daphnia* exhibit several complex behaviors, but are easy to study.
- Ease of study: transparent, and *Daphnia magna* are large enough to visually see with the human eye.
- **Ideal for generational toxicology studies:**
 - ▶ Reproduction starts around day 10.
 - ▶ Typical life span is 40-80 days.



(en.wikipedia.org/wiki/Daphnia_magna)

Why Do We Care About *Daphnia magna*?



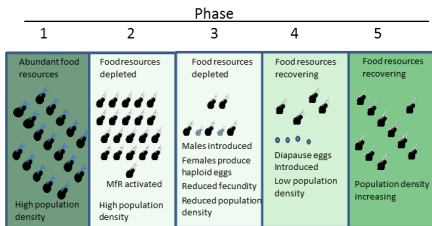
Olmstead 2000

- Differentiating males and females is straightforward.
- They are found in several ecosystems, are primary consumers (feeding on algae), and are food to secondary consumers (fish).
- The new "canary in the mineshaft".
- Ultimate goal: create a comprehensive baseline model for the population growth of *Daphnia magna*, which can then be perturbed in future work to include the effects of chemicals, toxins, temperature fluctuations, food disturbances, etc.
- Previous modeling efforts have missed important effects, or have failed to quantify uncertainty.

Difficulties in Modeling: Individual Variability

Fig. 9

- Cyclic parthenogenesis (LeBlanc 2013), as well as random male offspring (Olmstead 2001).



(LeBlanc 2013)

- High variability in reproduction. Mild
- individual variability in length.
- Individual variability in survivorship .

Difficulties in Modeling: Density-Dependent Effects

Density is known to affect:

- A delayed negative effect on fecundity (that is, fecundity goes down as density goes up, but this effect is delayed by an unknown and possibly variable number of days).
- An effect on the size of each daphnid. As the population density goes up, the maximum size of the daphnids in the population goes down. Affects classification of adults and juveniles, and other density effects.
- An effect on survivorship seen in juveniles (as population density goes up, survivorship goes down as daphnids compete for food).

(Frank1960, Pratt1943, Preuss2009 and the references therein)

Our Lab and Undergraduate Experimenters



Acknowledgments

Kevin Flores

Collaborator, Postdoc, now Asst Prof at CRSC/
NCSU

Kaska Adoteye, Chelsea Ross, Amanda Laubmeier

Collaborators, Graduate Students at CRSC/NCSU

Karissa Cross, Sarah Stokely, Michael Stemkovski, Timothy Nguyen, Robert Baraldi, and Emmaline Smith

Collaborators, Undergraduate Students at CRSC/NCSU

Gerald A. LeBlanc and Stephanie Eytcheson

Experimental Collaborators,
Department of Environmental & Molecular Toxicology, NCSU

RTG Fellowships, NSF, AFOSR, etc. grants

Adaptive feedback methodology on the parameter λ

Based on the results from our simple example using the logistic population growth model, we investigate the minimum amount of daphnia population data needed by estimating λ , the parameter of interest, and by calculating the standard errors for an increasing number of data; once the estimated parameter values and standard errors level off, we will deem that sufficient data will have been collected. We can determine this by using the methodology described above. We first calculate the percent change in parameter values for increasing n using equation

$$\frac{P_n - P_{n-1}}{P_{n-1}} 100,$$

and determine when this reaches below a threshold. We then repeat this process for percent change in standard error using equation

$$\frac{|s_n - s_{n-1}|}{\max_{j=1, \dots, n} s_j} 100$$

Our proposed *last data point collection* is then the maximum of these two times. However, there is a difference between our simulated data in the logistic model example and the data from Stark and Banks. In the logistic model example, the state variable represented total population size and the data was total population size at given times.

For the daphnia population model described above, the state variable also represents total population size, however the data are not total population size. Since the offspring are removed daily, the data consists of only two generations, not the total population. Since the life tables were built on the data (see Appendix A in [BBES1TechRep] for how to develop the life table), we used the calculated λ from each week of the life table to determine the total population for each week using equation

$$x(t) = x_0 \lambda^t.$$

We repeated this process for each replicate of each species. Thus, the datasets $\{y_j\}_{j=1}^n$ now consist of the total population for each week.

We use the inverse problem methodology to carry out estimation procedures for $\theta = \lambda$ using the ordinary least squares method. Again, we did not try to estimate x_0 since we know $x_0 = 10$. Using the MATLAB function *fminsearch*, we minimized the cost functional

$$J_n(\theta) = \sum_{j=1}^n [y_j - f(t_j, \theta)]^2$$

with respect to θ , where $f(t_j, \theta) = x(t_j)$. We provided the MATLAB function with an initial guess which was equivalent to the λ calculated from the life table for that week, and obtained the minimized cost $J_n(\hat{\theta})$, where $\hat{\theta} = \hat{\theta}^n$ represents the optimized value of θ using n number of points. To compute the standard errors for λ , we used the partial derivative

$$\frac{\partial x}{\partial \lambda} = x_0 t \lambda^{t-1} = 10 t \lambda^{t-1}.$$

Following the procedure in the logistic model example, we performed **multiple inverse problems to estimate $\theta = \lambda$** for an increasing number of data points. We first used only two data points $\{y_1, y_2\}$ to estimate $\theta^{\wedge 2}$ and then repeated this process until all n data points had been used, resulting in the set $\{\theta^{\wedge 2}, \dots, \theta^{\wedge n}\}$. We then **plotted the total population data (calculated from the λ for each week from the life tables) along with the estimated curves** over the entire time period using each $\theta^{\wedge i}$, for $i = 1, \dots, n$. The estimated parameter values were then plotted along with the weekly λ values from the life table. As expected, the two sets of λ values were extremely similar. **The percent change in estimated parameter values, standard errors, and percent change in standard errors were also plotted.** This process was then repeated for each replicate and for each species. Appendix B in [BBES1TechRep] contains all of the graphs for each replicate of each species.

Apply this same methodology to the daphnia population data:

- Studied three species of Daphniids *Ceriodaphnia dubia*, *Daphnia magna*, and *D. pulex*
- For each species, 3 replicates of 10 individuals were observed for their entire lifespan
- Recorded number of births each day and survival of initial population
- Stark and J. Banks developed life tables weekly, focusing on the parameter λ , the multiplication rate of the population

$$x(t) = x_0 \lambda^t = x_0 e^{rt}$$

Assumptions:

- Only changes in population are from birth and death
- Population is growing at a constant rate (birth and death rates are constant)

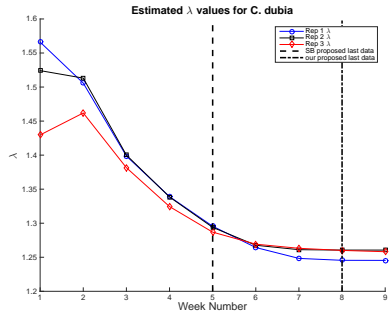
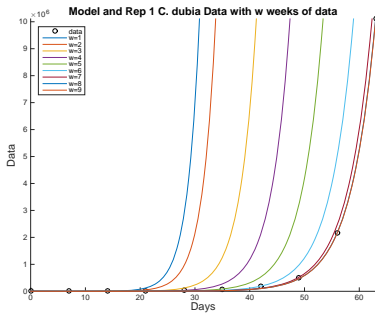
Predictive Methodology on Parameter λ

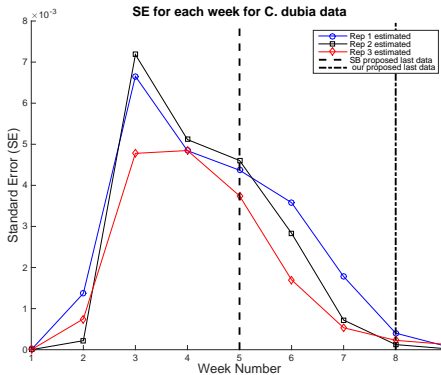
Using the λ from the life table, we calculated the total population and carried out the estimation procedure for λ

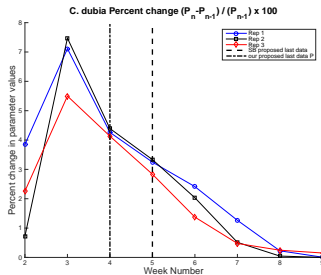
C. dubia number of offspring and survival data were used to build a life table as in [StarkBanks2015]. The calculated λ from the life table for each week was then used to calculate the total population each week. Using this total population data, parameter values $\hat{\theta}^n$ are estimated each week and the resulting curves are plotted along with the total population data. The estimated parameter values $\hat{\theta}^n$ are plotted as well as the standard errors for $\theta^{\hat{n}}$ for each replicate.

The dashed vertical line represents the proposed last data by Stark and Banks. The dash-dot vertical line represents our proposed last data.

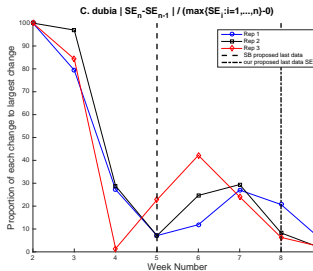
C. dubia





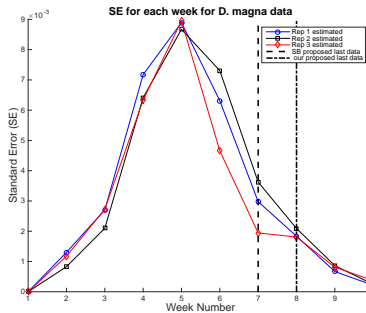
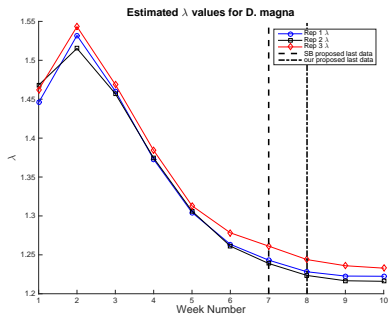


The percent change in parameter values calculated using n and $n - 1$ number of *C. dubia* data points. For each method (parameter or standard error), the dash-dot vertical line is chosen as the proposed time for each replicate that occurred the most number of times.

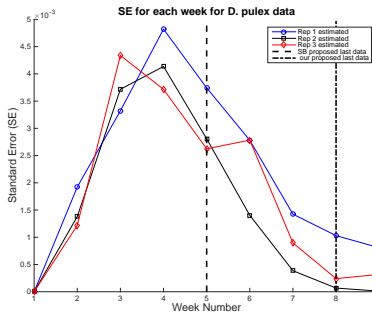
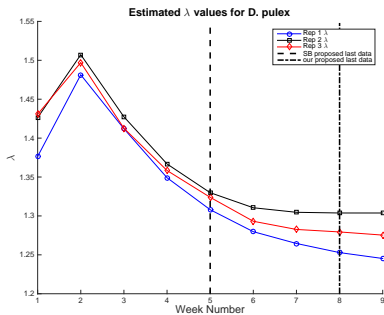


The percent change in standard error relative to the maximum standard error. The percent change in standard errors are calculated using n and $n - 1$ number of *C. dubia* data points. For each method (parameter or standard error), the dash-dot vertical line is chosen as the proposed time for each replicate that occurred the most number of times.

D. magna



D. pulex

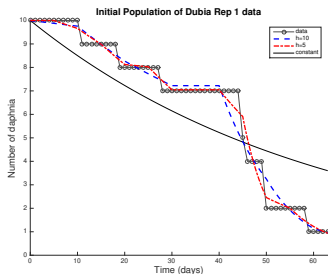


Conclusions and Future Work

- Developed predictive methodology for determining when enough data has been collected using the simple Logistic growth model
- Using constant growth, our results supported the results of Stark and J. Banks in that partial demographic data can replace full demographic data
- Time-varying rates fit the model statistically significantly better than constant rates
- We need to better understand the improved mathematical model, and we need to investigate a possibly more appropriate statistical model and include the adaptive features in optimal design methodology

Current and Future Work

In the formulation of the life tables, the birth and death rates are assumed to be constant, however these assumptions may not necessarily hold true



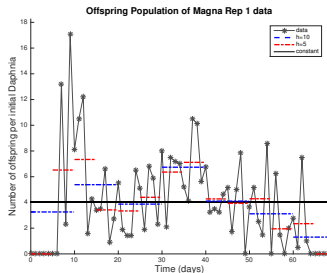
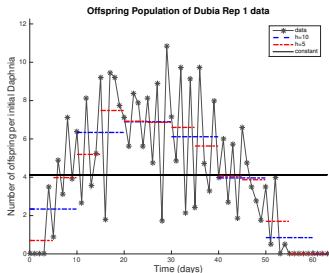
Compare models assuming constant and time-varying death rates:

$$\frac{dx_1}{dt} = -dx_1$$

$$\frac{dx_2}{dt} = -d(t)x_2$$

Current and Future Work

Fertility rate: average number of births per daphnia per day:



- Need to consider equations that are used to build a life table assuming non-constant rates
- Understand better how life table equations are used with cohort data