

An effort to improve nonlinear modeling practice

John C. Nash
Telfer School of Management
University of Ottawa (retired)
with the
NCEAS Nonlinear Modeling Working Group



What?

- Funded by the National Center for Ecological Analysis and Synthesis (NCEAS)
- Nonlinear Modeling Working Group
 - assembly of approximately 20 international researchers
 - 2 formal meetings of 1 week each in Santa Barbara (UCSB)
 - January 2011 & July 2011.
- Use only open source s/w (and hopefully improve s/w or practice)
 - R
 - ADMB (Automatic Differentiation Model Builder)
 - Open versions of BUGS (OpenBUGS / JAGS)
- Suite of test problems of an ecological nature requiring nonlinear modeling
- Write-ups of problems, **Web-site**, conferences and papers



Who?

Casper Willestofte Berg	Technical University of Denmark
Ben Bolker	McMaster University
Mollie Brooks	University of Florida
Liza S. Comita	University of California, Santa Barbara
Elizabeth E. Crone	Harvard University
Sarah Cubaynes	Centre d'Ecologie Fonctionnelle et Evolutive-Centre national de la recherche scientifique CEFE-CNRS
Trevor Davies	Dalhousie University, Halifax
Perry de Valpine	University of California, Berkeley
Jessica Ford	University of Tasmania / CSIRO
Beth Gardner	North Carolina State University
Olivier Gimenez	Centre d'Écologie Fonctionnelle et Évolutive
Marc Kéry	Swiss Ornithological Institute
Eunjung Kim	University of Hawaii, Mānoa
Cleridy Lennert-Cody	InterAmerican Tropical Tuna Commission
Arni Magnusson	Marine Research Institute
Steve J. Martell	University of British Columbia
Mark N. Maunder	InterAmerican Tropical Tuna Commission
Mihoko Minami	Keio University
John C. Nash	University of Ottawa
Anders Nielsen	Technical University of Denmark
Jim Regetz	University of California, Santa Barbara
J. Andy Royle	USGS Patuxent Wildlife Research Center
Hans J. Skaug	University of Bergen
Elise Zipkin	US Geological Survey (USGS)



Main products

- Collection of worked examples
 - Different computational packages
 - Worked example “user” code for the different tools
 - Parameter estimates, uncertainties and times
 - Simulated data runs
 - Write-up
 - Advice and recommendations on software for problem types

Main products (2)

- Accelerated tool development
 - R packages R2admb, kktc
 - Influences on R package optimx, some ADMB
- Subversion and wiki collaborative environment
 - Automated parallel simulation using *make*
- Public wiki: currently just a stub
 - Planned launch at end of November

<https://groups.nceas.ucsb.edu/non-linear-modeling>



Problem span

- Linear regression (but may need nonlinear computations for dispersion estimates)
- Nonlinear regression & MLE
- Zero Inflated Poisson models
- Linear mixed models
- Nonlinear mixed models
- State-space population dynamics
- Mark-Recapture

Example Problems

- Cod
 - Zero-inflated regression model (fixed-effect) for cod parasite data
- Min
 - Mineralization of terbuthylazine linear ODE solution by mle.
- Nmix
 - N-mixture models with random observer effects
- Occupancy
 - Dynamic occupancy model analysis for the Swiss finch (crossbill)
- OrangeTree
 - Pinheiro & Bates (2000, Ch.8.2) logistic growth curve with random effects

Example Problems (2)

- Owls
 - Zero-inflated regression model (mixed-effect) for owlet begging behaviour
- Skate
 - Bayesian State-space length disaggregated population model: Estimating total mortality by decade for skates
- Spatialcar
 - Spatial Poisson CAR model (for Swiss bird counts??)
- Spatialmr
 - Hair-snare mark recapture to estimate bear population on Fort Drum

Example Problems (3)

- Stage
 - Stage structured population dynamics model
- Tadpole
 - Predation of tadpoles modelled by size (nonlinear fit).
- Theta
 - Parameter and state estimation using hidden Markov models using theta logistic population model
- Weeds
 - Sigmoid growth curve; serious numerical issues.
- Wildflower
 - Binomial GLMM with the interesting features of multiple (3) random effects, including two that are crossed.

How we worked?

- Initial meeting – list of problems
- Teams to work on problems in 3 s/w systems
 - R
 - ADMB
 - OpenBUGS / JAGS (mostly settled on the latter)
- Once reference problem “solved”, add simulated data
- Eventually set up a directory template and makefile in Subversion repository

Tadpole example problem

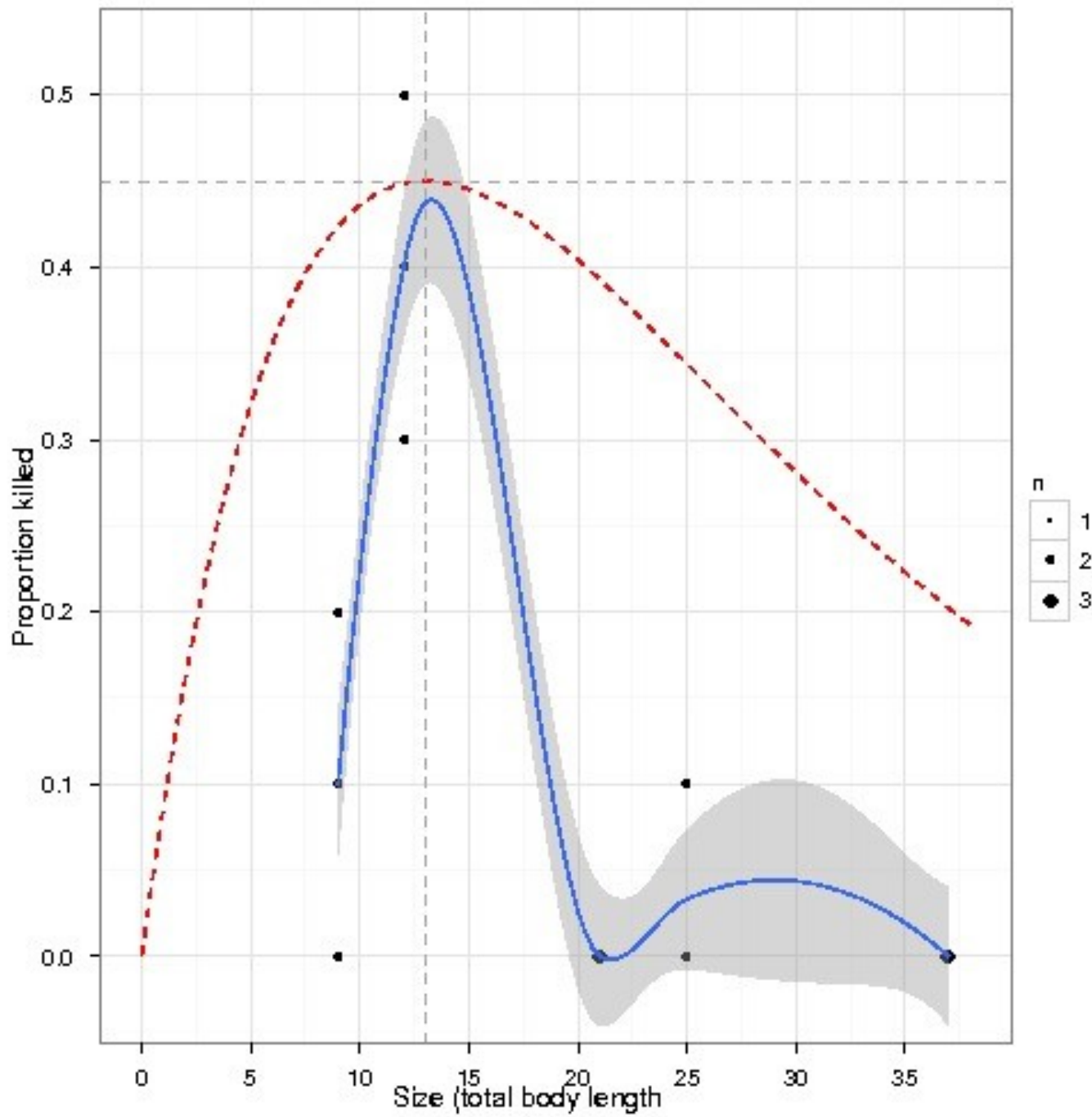
- Model of predation

$$P(\text{kill}) = c((S/d) \exp(1 - (S/d)))^g$$

- Parameters are c , d , g ; S is size
- Tadpoles can't be eaten if 0 length, and get too big at some point, so need a function that has a particular mound-shape

Data

	TBL	Kill	Exposed
1	9	0	10
2	9	2	10
3	9	1	10
4	12	3	10
5	12	4	10
6	12	5	10
7	21	0	10
8	21	0	10
9	21	0	10
10	25	0	10
11	25	1	10
12	25	0	10
13	37	0	10
14	37	0	10
15	37	0	10



Project directory – tadpole

```
|-- ADMB
|  |-- tadpole0.tpl
|  |-- tadpole_ADMB_fits.RData
|  |-- tadpole_ADMB_funs.R
|  |-- tadpole_ADMB_run.R
|  |-- tadpole_ADMB_runx.R
|  `-- tadpole.tpl
|-- BUGS
|  |-- model.txt
|  |-- rout
|  |-- tadpole_BUGS_funs.R
|  |-- tadpole_BUGS_run.R
|  `-- tadpole_bugs.txt
|-- DATA
|  `-- tadpole.dat
|-- Makefile

|-- R
|  |-- fit_fits.RData
|  |-- fitx.RData
|  |-- jngraph.R
|  |-- Rplots.pdf
|  |-- tadpole_bugs.txt
|  |-- tadpole_R_funs.R
|  |-- tadpole_R_run.R
|  |-- tadpole_R_runxj.R
|  |-- tadpole_R_runx.R
|  `-- tadpole_R_sum.R

|-- SIMS
|  |-- allsim.RData
|  |-- Makefile
|  |-- sim.R
|  |-- tadpole1
|  |-- tadpole1.tpl
|  |-- tadpole_gen
|  |-- tadpole_gen.tpl
|  |-- tadpole_sim_funs.R
|  |-- tadpole_sims.RData
|  |-- tadpole_sims_run.R
|  |-- tadpole_sims_sum.R
|  |-- tadpole.tpl
|  `-- true.dat
|-- WRITEUP
|  |-- tadpole.pdf
|  `-- tadpole.Rnw
```

Automation

- Makefile (per project)
 - R, ADMB, BUGS
 - Not yet WRITEUP
 - SIMS by 'make sim' (but tadpole lacks this)
- Global “make” ?
 - Likely need compute-cluster to run this, as some problems are compute intensive

Lessons for WG project

- Context and experience are IMPORTANT
 - What you regularly use is always easier to use
- The problems and tools are both complicated.
 - Control parameters (many) and messages not obvious.
- Even experienced users are sometimes surprised at the (initial) failure rate of optimization tools
 - Tuning / scaling / options sometimes critical to success
 - Automatic choices good, but problems are challenging

Lessons for WG project (2)

- Problem setup – devils & details
 - Parametrizations (e.g. log) & scaling affect outcomes and convergence tests
 - Importance of tests such as KKT optimality check
 - Separability, list vs dataframe, vectorization, use of sparsity, and other specification issues
 - Selection of sub-tools (optimizers, samplers)
- Tools are critical to managing overall work
 - Subversion, make, LaTeX, Sweave

Lessons to share (1)

- Importance of analytic derivatives
 - Getting “the” answer
 - Efficiency and correctness of KKT tests etc.
 - Memory space / time trade-off
- Platform issues for software use
 - This is changing over time, but “glitches” are common with all platforms and s/w
 - Need a lot of tools – may result in inconvenient system (computer and OS issues)

Lessons to share (2)

- Caution! Check “answers” to see if OK
- WG method may be useful s/w development model
- R is a VERY useful wrapper
 - ADMB: glimmadmb, R2admb, graphing output,
 - OpenBUGS/JAGS: R2jags, rjags, graphing output, and some other packages
 - FLR: flr-project.org (Fisheries Library in R)

“Products”

- Already resulted in new R packages (kktc, R2admb) and mods / feature requests for others
- Likely will influence ADMB and possibly JAGS or OpenBUGS
- Write-ups (vignettes)
- Talks (like this!)
- New variation on a s/w development process

Software and IT issues

- Setting up and maintaining a system to run all problems is challenging
 - ADMB, R + packages, JAGS/BUGS, LaTeX, make
 - Each platform (Linux, Win, Mac) a bit different, but less so than a few years ago
 - JAGS 2.2 → 3.0 caused a version hiccup with R2jags and rjags
- “Clean” setup as bootable system
 - Debian squeeze, Ubuntu Lucid on USB HDD, #! Linux (based on squeeze) on USB Flash 8GB.

Dissemination

- Different audiences
 - Tool builders – R / ADMB / BUGS – JAGS
 - Subject specialists
 - Ecology
 - Other bio-science, agriculture
 - Non-bio: business, physical science, economics
 - Statistical scientists
 - Computer scientists / software developers
 - Policy folk – lesson of open source and collaboration

Acknowledgements

- The Working Group activity was conducted as a part of the Nonlinear Modeling Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #EF-0553768), the University of California, Santa Barbara, and the State of California.
- The presenter's participation in UseR!2011 is supported by private funds made available by Nash Information Services Inc.

