PNAS
www.pnas.org

Supplementary Information for
Ctenophores are direct developers that reproduce continuously beginning very early after hatching

Allison Edgar, José Miguel Ponciano, Mark Q. Martindale

Mark Q. Martindale
Email: mqmartin@whitney.ufl.edu

This PDF file includes:

Supplementary text: Statistical approach ("Hierarchical stochastic time series model of early reproduction in ctenophores"; includes Figure S1)
Tables S1 to S3
Legends for Datasets S1 to S6
SI References

Other supplementary materials for this manuscript include the following:

Data S1: BIC summary table for all analyses (BICsummary.xls)
Data S2: Raw data for parental size experiments.
Data S3: Raw data for temperature experiments.
Data S4: Raw data for density experiments.
Data S5: Raw data for nutritional experiments.
Data S6: Raw data for selected individuals that grow from ≤2 to ≥4 mm.

Supplementary Information Text:

Hierarchical stochastic time series model of early reproduction in ctenophores

Model derivation

Survival, extinction, and growth of populations are inherently stochastic (1). At its core, the problem of modeling the growth of the accumulation of embryos from ctenophores is a stochastic population dynamics problem. Existing time series methods to fit biologically sound stochastic growth models allow the estimation of quantities related to growth rates, the average time until any given event (first passage times, for instance), and the chances of different events of interest. Of particular interest in this case is a process-based estimation of the time-varying probability that within a given window of time, the number of embryos will grow by any given amount under different experimental settings.

Conservation biology (2), wildlife management (3, 4) and evolutionary microbial population dynamics (5) are only a few of the research areas in ecology and evolution that have long benefited from the application of stochastic population growth models. In conservation biology for instance, the so-called Population Viability Analyses (PVA) were changed dramatically by the contribution of papers like Dennis et al.'s 1991 (1) estimation of growth and extinction parameters. Their key contribution was illustrating how fundamental results and concepts from stochastic processes could be adequately used to better understand and predict population dynamics processes.

To model the early reproduction process in ctenophores, we originally used a "stochastic birth process" (6), which is a well-known continuous time and discrete states Markov Process (7) used in various population dynamics contexts (7–10), notably to model pathogens' reproduction (11). One key property of such processes is that the total (cumulative) number of births can only increase or remain the same over time, but not decrease. The stochasticity in this general model is called "demographic stochasticity" (1, 12) and represents chance variation due to individual heterogeneities in birth rates. This type of process variability is particularly relevant at low numbers (1).

Let $N(t)$ be the number of accumulated embryos at time t. We denote the initial number of embryos $N(0)$, as m. Heuristically, the birth process can be described as a process where embryos are being born one at a time, at random time points since the beginning of each trial. Births occur at a rate $\lambda_n(t)$ and at random time intervals. One of the key characteristics of this stochastic process that makes it suitable to model the observed accumulation of births is that it is not necessary to record the exact moment at which every single event occurs. Inference can still be done when observations are made at unequal sampling intervals. The only data needed to fit different models for the birth rate are the number of embryos alive observed at different time points. We note in passing that here we use the convention that random variables are written with capital letters and realized values as lowercase letters. Hence, $n(t)$ denotes the observed number of embryos at time t and thus, one realization of the random variable $N(t)$.

Let $p_n(t)$ denote the probability of observing $n(t)$ births at time t or $Pr\big(N(t) = n(t)\big)$. Using standard stochastic process results one may arrive at an analytical expression for $p_n(t)$ as a function of the birth rate $\lambda_n(t)$ (7). The utility of our modeling approach lies in the fact that we were able to translate different hypotheses/models regarding the factors modulating the birth rate into specific mathematical functions for $\lambda_n(t)$. Thus, we explicitly stated how $\lambda_n(t)$ is modulated by the categorical and quantitative variables of interest. The resulting expression for $p_n(t)$ under each biological hypothesis serves as the direct link that connected the observations with the proposed probabilistic model via the likelihood function.

The birth process for the problem at hand was indeed formulated so that it embodies a suite of plausible biological hypotheses regarding the unfolding of the reproduction process, and in particular, the dependency of its intensity on a handful of biological factors of interest. These factors included the density of coetaneous individuals in the experimental container, the temperature, and the type of food given to the ctenophores. Details about this formulation are given in the "Parameter Estimation" section below.

The general birth process formulation applied to our case assumes that there exists an arbitrarily small amount of time Δt during which at most one new birth occurs with probability $\lambda_n(\Delta t)$ and that the probability that no birth occurs is $1 - \lambda_n(t)$. This model also assumes that the probability of any other event is negligible, *i.e.*, that only birth of cases can be observed. Accordingly, the probability that starting at time t, after waiting a small amount of time Δt the observed number of births is n is:

$$p_n(t + \Delta t) = (\Delta t)\lambda_{n-1}p_{n-1}(t) + (1 - (\Delta t)\lambda_n)p_n(t). \text{(Eq. 1)}$$

The first term in the right-hand side (RHS) of Eq. 1 is the probability that the cumulative number of births at time t was $n - 1$ and a birth occurred with probability $\lambda_n(\Delta t)$. The second term is the probability that at time t the number of cumulative births was already n and that within that small time window Δt no birth occurred.

As $\Delta t \rightarrow 0$, Equation 1 tends to the following system of Ordinary Differential Equations (ODEs):

$$\frac{dp_n(t)}{dt} = \lambda_{n-1}p_{n-1}(t) - \lambda_n p_t(t),$$

Where $n = m, m + 1, m + 2, \ldots$ Letting $\lambda_n = \lambda$, a positive constant, the above system of equations can be readily solved (7) to yield the probability of observing $n(t)$ births at time t $p_n(t)$ in the form of a Poisson random variable, *i.e.*:

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^{n(t)}}{n(t)!}$$

Thus, under this stochastic process, the total number of births occurring during a time period τ is Poisson distributed with parameters $\lambda\tau$ and the waiting time until the first event or between any two events follows the exponential distribution with parameter λ.

Therefore, the Poisson distribution and the exponential distribution are tightly linked under this model: assuming one implies assuming the other. However, the initial motivation for this analysis using stochastic processes was the observation that counts of new births appear clumped over time. As well, the counts of new births over time are characterized by an excess of 0's. Hence, neither the assumption of Poisson distributed counts nor of exponentially distributed inter-event times seemed appropriate.

In survival analyses applied to econometrics and engineering, when exponentially distributed inter-event times are not appropriate due to temporal clumping of events, Pareto distributions are used instead. As it turns out, a special case of this distribution, known as the Lomax distribution (13) arises naturally in our context by incorporating the birth rate as a random effect and formulating our stochastic model as a hierarchical time-series model. Accordingly, if we model individual variation in birth rates using a Gamma distribution with parameters α and k and integrate the Poisson probability of observing $n(t)$ births at time t $p_n(t)$ over this probability model we get that

$$p_n(t) = P(N(\tau) = n) = \binom{n+k-1}{n} \left(\frac{\alpha}{\tau+\alpha}\right)^k \left(\frac{\tau}{\tau+\alpha}\right)^n. \quad \text{(Eq. 2)}$$

Hence, the number of new births in a time interval of size τ is Negative Binomially distributed with parameters $P = \frac{\alpha}{\tau+\alpha}$, $Q = \frac{\tau}{\alpha+\tau}$ and overdispersion parameter k. Under this model, the waiting time T

between birth events is no longer exponentially distributed. Noting that, starting at time 0, the event "no birth was recorded during the time interval $\tau = \tau - 0$" is equivalent to the event "the waiting time T until the next birth is greater than τ ", the cumulative distribution function (cdf) of T is readily found to be

$$F(\tau) = P(T \leq \tau) = 1 - P(N(\tau) = 0) = 1 - \left(\frac{\alpha}{\tau+\alpha}\right)^k. \quad \text{(Eq. 3)}$$

Differentiating the cdf in Eq. 3 with respect to τ we immediately obtain the probability density function (pdf) for the waiting time between births in our hierarchical model:

$$f_T(\tau) = \frac{dF(\tau)}{d\tau} = \left(\frac{\alpha}{\tau+\alpha}\right)^k \left(\frac{k}{\tau+\alpha}\right). \quad \text{(Eq. 4)}$$

This equation is recognized as the pdf of the Lomax distribution (13), which is a special case of a Pareto distribution.

The salient feature of any stochastic birth model is that one or multiple forms of biological variability or heterogeneity can be acknowledged. Consequently, the population projections of these models will deviate at random (but according to the type of biological heterogeneity considered) from the usual deterministic and smooth projections obtained from differential or difference equation models (See Fig. S1 A). A hypothetical single growth trajectory of the cumulative number of offspring produced in given experiment is shown in Fig. S1 A. In stochastic processes terminology, this sample trajectory is usually called a "single realization" of the stochastic growth process. Visualizing multiple independent realizations of the stochastic model of population growth starting from the same initial conditions (See Fig. S1 B.) is important because it helps convey the fundamental property of these models: the population size at any point in time obeys a probability law shaped by the birth rate $\lambda_n(t)$. Put in another way, the population size at any time can be modeled with a time-dependent random variable. Hence, unlike deterministic models which give an exact point projection of what the population size of interest will be at any point in time, these models specify the probabilities that the population size at time t will be of any given size. This probability law is at the center of our Parameter Estimation approach as it provides the link between the data and the models tested.

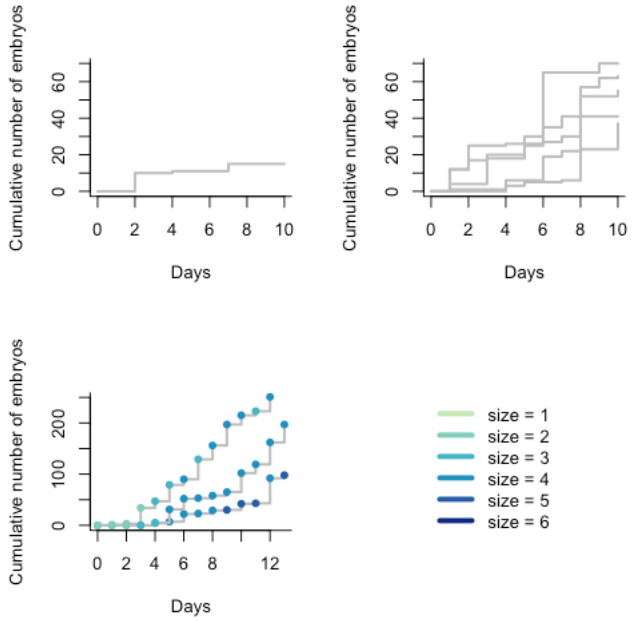

Figure S1. A. A single projection/trajectory of a stochastic birth model showing the cumulative number of births is usually referred to as a "single realization" of the stochastic process. Multiple realizations of the process, depicted in B, lead to a general description of the process in terms of means and variances of the cumulative number of births over time. With our experiments, we aimed at inferring the properties of the mean of the process over time and under different experimental conditions. C. Visualization of real individuals' embryo production over time, overlaid with measured parental body size data at each step. Additional data contained in Supplemental Data File S6 can be used with the provided R code to visualize other individuals' trajectories.

Parameter estimation

With the modification incorporating individual heterogeneity, our stochastic birth process is strictly no longer a Markov process. Rather, this type of modification belongs to a class of models known as a "Semi-Markov Processes" (SMP's, (14)). Although the waiting time is no longer exponentially distributed, the embedded discrete-time chain counting the cumulative number of births retains the Markov property. This fact is crucial for writing the likelihood function which connects the different hypotheses we tested with the observations. We call this general hierarchical stochastic model the semi-Markov Birth Process (SMBP).

Critically for our model-fitting approach, our hierarchical model formulation admits an explicit expression for the mean number of random birth events occurring in any given time interval τ:

$$\mathrm{E}[N(\tau)] = \frac{kQ}{P} = k\tau\frac{1}{\alpha}. \quad \text{(Eq. 5)}$$

Hypothesized drivers of mean number of births in a time interval were subsequently modeled using a traditional regression format using this expression. Our approach, akin to Generalized Linear Models (GLMs, (15)), allowed us to write the mean of the stochastic process (Eq. 5) as a function of continuous covariates (like temperature) or other experimental treatments or factors (See "Parameter Estimation" section below). As stated in the main text, traditional GLM approaches should not be used here because they erroneously ignore the biological time dependency in the counts, and therefore may result in excessive Type I errors in hypothesis testing as well as severe model choice errors using information criteria (16).

Accordingly, each experimental time series for a single culture was modeled as a single realization of our stochastic process model and treated as an independent replicate. Therefore, an entire time series (usually of about 7 days) was treated as a single experimental unit observation. The mean for each replicate is given by Eq. 5 but modified to account for the particular experimental conditions of each replicated time series. To model the mean of the stochastic process (Eq. 5) as function of hypothesized drivers/covariates we proceeded as follows:

The basic data unit consists of a recorded pair of numbers for every day for one experimental setting. These two numbers are the day, t_i and the new number of embryos n_i produced between the previous day (t_{i-1}) and the current day (t_i). Starting with $n_0 = 0$ and $t_0 = 0$ the data for a basic experimental time series unit is denoted as:

$$(n_0, t_0), (n_1, t_1), (n_2, t_2), \ldots, \left(n_q, t_q\right),$$

where q is the total number of counts done after day 0. Because the SMBP evolves by increments (jumps) during each time interval $\tau_i = t_i - t_{i-1}$, and since the probability distribution of the jump is negative binomial, the likelihood function for a single time series is equal to a product of q negative binomial expressions evaluated at the observations. Care must be taken, however, in the parameterization of that probability distribution for every jump, as it depends on the size of the time increment τ and on the value of the parameter α. Critically, it is through the parameter α that we phrase different hypotheses regarding the drivers of the mean size of the increments. We phrase these hypotheses as follows: First, define $\theta = \frac{1}{\alpha}$. Then, we arranged the observed values of the covariates in

matrix format by computing the design matrix (17) according to the particular model tested using R's function "model.matrix()". This function is at the core of the off-the-shelf GLM programs in R. For example, if the model to be fitted states that the jump increments every day only depend on the number of parental types (a continuous variable) using standard "formula" notation in R code, the design matrix would be given by

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{q1} \end{bmatrix},$$

where $x_{11}, \ldots, x_{q1}$ are the values of the number of parental types present every day from 1 to q. The second sub-index (always a 1 in this example) is used to enumerate/name the covariate of interest. In this case, we are simply naming the number of parental types as the first (and only) covariates. For this example, we would connect the jump model parameters τ and θ with the hypothesis that the mean jump size depends on the number of parental types through the log-link regression function

$$\log \boldsymbol{\theta} = \boldsymbol{X\beta}, \quad \text{(E. 6)}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is a vector containing the unknown parameters to be estimated (here the bold-face notation is used to denote vectors). The hypothesis of no effect of the number of parental types would then dictate that the slope parameter $\beta_1$would be equal to 0. Effectively then, we make the parameter θ a function of the covariates, whatever these may be. Categorical variables can be readily accommodated using standard dummy variable notation for design matrices, all while making sure the model is not overparameterized (17). What changes from model to model is the form and dimension of the design matrix and the number of parameters to be estimated but the general log-link function (Eq. 6) remains unchanged. We designed and wrote our own code for Maximum Likelihood (ML) parameter estimation, keeping standard R GLM notation for ease of use, and uploaded the code to GitHub (https://github.com/jmponciano/ctenophores) that can be used for any SMBP. An example of the data analysis is provided in the program "ExampleCtenophores.R". The output of the function gives the model parameter estimates, the optimized negative log likelihood value and the Bayesian Information Criterion (BIC) model score. Confidence intervals for the mean growth in the number of embryos as a function of time were computed using parametric bootstrap.

Written as such, this new hierarchical model is one of the few hierarchical models that does not necessitate an MCMC numerical integration machinery as an analytical solution is readily obtained for the likelihood function. Then, a simple numerical optimization routine is enough to maximize this likelihood. Since hierarchical models first were used in Ecology, Bayesian methods were increasingly being applied for statistical inference for these models as these methods bypass difficult integration problems to obtain the likelihood function (18). Soon, hierarchical models were taken as synonyms of Bayesian statistics, a point of view that has long shown to be erroneous (e.g., (19)). Lele et al (2007) (20) first presented a technique that uses the numerical integration benefits of Bayesian statistics to compute Maximum Likelihood estimates for hierarchical models. However, the model presented here is like one of the most popular hierarchical (also called state-space models in ecology) used in population dynamics, the Gompertz State Space model (4) in that no numerical integration is needed to compute the likelihood function, as it can be derived analytically. Therefore, there is no need of MCMC approaches. Besides the GLM model inadequacy (due to temporal dependencies) mentioned in the main text, we add that we did not use a Bayesian mixed-model GLM approach because statistical research spanning more than a decade and a half and recently reviewed (21, 22) has shown that Bayesian inference for stochastic, hierarchical models (or state-space models) can be unreliable due to parameter identifiability issues and non-transformation invariance of (uninformative) priors among other things (see (4, 18, 19, 23–29) for original papers) .

Nonparametric Bootstrap and Model Selection

The problem of how to measure support for the different tested models of the drivers of embryos production is at the core of our statistical methodology relying on Evidential Statistics (30–32). The main goal of the evidential statistics paradigm (16, 25, 28, 32–34) is to quantify the strength of the evidence in the data for a reference model relative to another model. This quantification is done through an evidence function, which is a statistic for comparing two models (35). The salient property of evidence functions is that their associated probabilities of making a wrong model choice approach 0 as sample size increases. These probabilities, which are analogous to Type I and II errors in the classical Neyman-Pearson Hypothesis Testing (NPHT) framework are pre-data error rates and measure the probability of obtaining "weak misleading evidence" and the probability of obtaining strong misleading evidence. Here, making the wrong model choice refers to deeming as best a model that is not the closest to the true generating process model and "misleading evidence" corresponds to obtaining observations that either weakly or strongly support a wrong model, *i.e.*, not the model that is closest to the data-generating process.

The main reason for choosing this statistical paradigm is that, unlike the classic NPHT and Bayesian approaches it provides solid guidelines to assess inferential errors when none of the statistical models at hand are a perfect representation of the data-generating process (26), that is, when all the models tested are imperfect mathematical "misspecifications" of the data-generating process as it is the case here. Indeed, the NP framework for instance depends critically on either the Null or the Alternative hypothesis being a statistical model that perfectly represents the data generating mechanism, fixes one of the error probabilities (alpha) and thus problematically assesses the evidence *against* the null hypothesis with alpha constant regardless of sample size and remains silent with respect to the evidence *for* the null hypothesis. The asymmetry of the error structure has often led to difficulties in interpretation of hypotheses tests. On one side, the decision to pick the alternative model over the null hypothesis is not controversial as it has some intuitively desirable statistical properties: for example, the probability to reject the null hypothesis given that the alternative is true converges to 0 as sample size increases (28). On the other side however, the probability of wrongly choosing the alternative when the null is true remains stuck at the chosen level α regardless of how large a sample size is collected (16, 28). Matters get only more complicated when it is considered that the original Neyman-Pearson theorem (36) assumes that the data was generated under one of the two models but provides no guidance whatsoever in the event of model misspecification, a scenario commonly encountered in science (16). The fact that in scientific practice model comparison rarely stops at two models complicates even further the interpretation of experimental results using the NPHT. An overconfidence in model selection procedures also results in Bayesian Statistics when the model misspecification is ignored (37).

The evidential approach, on the other hand, proposes instead to fix cutoff values for the evidence statistic, not the error probabilities. Under this concept of evidence, the value of a statistic like the likelihood ratio is evidence, not an error rate that is pre-set. Then, the evidential error probabilities mentioned above can be made to converge to 0 as sample size grows large. Finally, under this evidential statistics approach, the conclusion structure of say, a comparison between two models H_1 and H_2 has a trichotomy of outcomes: *i)* strong evidence for H_1, *ii)* weak or inconclusive evidence and *iii)* strong evidence for H_2.

Some, but not all, information criteria commonly used for model selection are evidence functions. That is, not all information criteria have a vanishingly small probability of making the wrong model choice as sample size gets large. An Information Criterion is an estimator of the sample size scaled difference of divergences between the generating mechanism and the competing models. One such estimator is the ΔBIC, or difference in the Bayesian Information Criterion (38). As a note, we mention that this criterion is one of a series of criteria that all have frequentist derivations (39) so to avoid Bayesian implications, in other papers we refer to it as the "Schwartz Information Criterion", or SIC. In keeping with the most common naming convention for this statistic, we call it here "the BIC". Suppose we are comparing two models, 1 and 2. Then, if BIC_1 is the BIC for model 1 and BIC_2 is the BIC score for model two and $\Delta BIC_{21} = BIC_2 - BIC_1 > 0$ then that means that model 1 has the lowest score and hence, it is the model that is closest to the unknown, true generating mechanism. Repeating the same calculation between all possible pairs of models yields the best model as that one with the overall lowest BIC score. Because these statistics depend on sample space probabilities and hence on the sample at hand, there exists a non-negligible probability that this comparison procedure leads to the wrong model choice, i.e., not the

model that is closest to the generating mechanism. For the BIC, the probability of such mistake becomes, however, smaller as sample size gets larger (16).

Information criteria are all functions of the log-likelihood maximized under the model being fitted plus a penalty term. For example, Akaike's Information Criterion (AIC, (40)) is computed as -2 times the log likelihood plus twice the number of parameters in the model being tested and is an estimate of the expected, scaled divergence between the generating process and the approximating model at hand. Unlike the AIC, besides considering the number of estimated parameters, the BIC is scaled by the sample size. As a result, as sample size increases, the error in deeming a model as "best" using the BIC statistics becomes vanishingly small. This desirable property, called "Information consistency" (16, 25, 28) is lacking in the AIC (16). Inconsistent criteria, such as the AIC, tend to overfit at all sample sizes. Hence, the AIC as a model selection tool is insufficient because it is not information consistent. Furthermore, the mistakes done in model selection using AIC get worse specially when, as it is the case here, none of the models tested is a perfect specification of the true, underlying generating process (that is, all models are "mis-specified") (16). To be fair, the same is true for model selection under mis-specification using Bayesian Statistics (37) as we mention above.

Although all paradigms of statistical science (NPHT, Bayesian statistics, Evidential Statistics) have flaws (reviewed in (27)), the Evidential Statistics paradigm (16, 25, 27, 28, 30, 35, 41) possesses better, desirable characteristics for the quantification of uncertainty (see (27)) and ultimately, for the design of inferential statements about the models' proximity to the true, generating process.

Here, we used the approach proposed and exemplified by Taper et al (2021) (28) to attach an uncertainty measure to the model selection results via non-parametric bootstrap. As Cox (1958) (42) put it, any statement about a model form or about a model parameter, that is, any inferential statement becomes a statistical inferential statement only when an uncertainty measure is attached to it (27, 28). Taper et al (2021) (28) develop a non-parametric bootstrap approach to evaluate the support of a model that relies in the following idea extensively used also in other areas of statistical inference in biology like phylogenetics: if we were to obtain datasets after datasets of the same experiments, how often would one obtain evidence for one or the other inferential statement? Applied to our case, the question becomes: if we were to sample with replacement via non-parametric bootstrap the observed data sets for every set of experiments, how often would the model originally chosen as best would be deemed to be best? And what is, on average, the strength of the evidence of that model over the second-best model and so on?

The first question was answered by doing a non-parametric bootstrap resampling of each experimental data set to obtain a bootstrap replicate of the same structure as the original data set. Then, for each bootstrap replicate we fitted the best five models (6 in some cases) according to the original data set model fitting and inference and kept track of the BIC-based rank of all the models. After repeating this process 100 times we could compute the percentage of time that the original best model was again chosen as the best model. That percentage granted our model choice a degree of support akin to nodal support measures in phylogenetic inference (24), which we present in with each figure in the main text, and list comprehensively in Supplemental File Data S1. The second question above regarding the average strength of evidence relates to the following problem: when one does model selection and one model is deemed best because its information criterion score is smaller, how small is small enough before it is decided that one model is effectively better than the other? That question, partially answered in (25) and nicely illustrated in (34), Box 3 can be briefly summarized as follows:

Applied scientists have provided guidelines to evaluate the strength of evidence using differences in information criteria. Burnham and Anderson (2002) (43) for instance, famously suggested using the cutoffs $0 > \Delta AIC_{ij} = AIC_i - AIC_j > 2$, $4 < \Delta AIC_{ij} < 7$ and $\Delta AIC_{ij} > 10$ to deem the support for model j over model i as "essentially none", "considerably less" and "substantial". Different discretizations of the intervals of support have been suggested (44). These discretizations are to certain extent arbitrary and vague and thus result in the same types of confusions stemming from arbitrary critical alpha levels under NPHT. Following Royall (2000) (31) who proposed to fix cutoff values for the evidence statistic, not the error probabilities, Jerde et al. (28) design a system where every cutoff point in the scale of the strength of evidence corresponds to the likelihood of one model being better than the other by twice the amount,

four times the amount, eight times, sixteen times, thirty-two times and so forth. These cutoffs correspond respectively to approximately 1.4 points of difference in the BIC, 3 points, 4 points, 5.5 points, 7 points. Then, if a model is 32 times more likely than the other, the evidence for that model is deemed strong, 16 times moderate, 8 times marginal, 4 times weak and two times very weak.

Therefore, by computing the average strength of the support for the first model (i.e., average size of the difference between the best model and the next best model) in our bootstrap replicates, we provided an extra level of nuance to the uncertainty measure (percent support for the best model) that we devised. Thus, to each model deemed as best in our results we attached three measures: the observed ΔBIC between the best model and the next best model, its percent support via non-parametric bootstrap and the average ΔBIC between that model and the next best model and between the second-best model and the third-best model over all non-parametric bootstrap runs.

Supplemental display items

Table S1.

species	small-size spawning (mm)	large-size spawning (mm)	mean embryos per day	max clutch size	individuals examined (N)	reference (main text citation #)
M. ovum	0.75-1.6 (0.6-1.3)	ND	0.5	2	23	Jaspers *et al.* 2012 (45)
M. leidyi	1.5-2.8 (1.2-2.2)	~35	2.9	21	26 (7 whole)	Martindale, 1987 (46)
P. bachei	1.1-2.7 (0.9-2.2)	9.1	2.9	40	6-7 each	Hirota, 1972; Baker and Reeve, 1987 (47, 48)
P. globosa	1-2.6 (0.8-2.1)	8	9*	42	28	Wang *et al.* 2020 (49)

Size at onset and fecundity of “larval” spawning previously reported in ctenophores. The fourth column lists mean fecundity, the fifth column lists maximum clutch size during the period of small-size spawning, and the sixth column lists the number of spawning individuals that were observed. Parenthetical numbers for small-size spawning uses 0.8 conversion factor from body length, used in the cited work, to body width used in this manuscript, rounded to the nearest 0.1 mm.

Table S2.

	ISO-1800 (high DHA)	NANNO-3600 (high EPA)	RGcomplete (high DHA and EPA)
protein	44%	58.6%	55.7%
carbohydrate	25%	20%	19%
lipid	19%	14.5%	19.5%
DHA (mg/g dry wt)	11.2	0.3	41.9
EPA (mg/g dry wt)	1.8	28.7	18.1
ARA (mg/g dry wt)	0.4	2.7	2.1

Approximate nutritional content of each feed used in the prey rotifer diet experiments, as reported by supplier labeling.

Table S3. Confidence intervals for best-supported models for each data set (column 1).

experiment	*variable(s)*	*model*	*beta_k (SE)*	*beta_0 (SE)*	*beta_1 (SE)*	*beta_2 (SE)*	*beta_3 (SE)*	*beta_4 (SE)*
embryos: size	day	model5	0.2030078 (0.2703520)	1.0568879 (0.6208427)	0.3221804 (0.1203708)			
cydippids: size	day	model5	0.1494056 (0.4130690)	-0.4035272 (0.8919687)	0.4096800 (0.1499842)			
embryos: temperature	number + temperature	model1.2	1.066666 (0.36136462)	-2.16306112 (1.79162612)	0.08242668 (0.04780375)	0.17022280 (0.06267349)		
cydippids: temperature	number + temperature	model1.2	0.9443697 (0.37823491)	-3.60815982 (1.95915579)	0.07441569 (0.05069297)	0.21651703 (0.06820714)		
embryos: density1	bio rep + number	model1.3	0.5240834 (0.3224816)	-1.2850684 (0.8146496)	3.1660914 (0.7616562)	1.3318275 (0.8138532)	0.5708982 (0.8583342)	0.1568974 (0.1117714)
cydippids: density1	bio rep + number	model5	0.2775632 (0.4960618)	0.7331545 (0.9043939)	-3.6813343 (1.3136280)	-3.4322243 (1.2102000)	0.3990200 (0.2199577)	
embryos: feed2	bio rep + number + DHA	model12	0.6779661 (0.40666905)	3.43807047 (3.65792882)	-1.60657827 (0.79180671)	-0.01970894 (0.22105712)	0.04672436 (0.01968718)	
cydippids: feed2	bio rep + DHA	model12	0.5356867 (0.50657108)	1.65550682 (0.76961512)	-1.70752863 (0.77664501)	0.07196409 (0.02466868)		
embryos: density2	bio rep + number	model5	0.5732126 (0.24281054)	-1.25969425 (0.67444368)	1.33588121 (1.09882301)	2.65789800 (0.60752062)	0.29732809 (1.12146461)	0.88846527 (0.62702242)
(continued)			*beta_5 (SE)*	*beta_6 (SE)*	*beta_7 (SE)*	*beta_8 (SE)*	*beta_9 (SE)*	
			0.85068654 (1.10604344)	0.85068654 (0.68217098)	-0.06230228 (0.68217098)	3.32315493 (1.03586591)	0.27702925 (0.04982021)	
cydippids: density2	bio rep + number	model5	0.5497029 (0.3355522)	-2.1554548 (2.2402377)	0.5052674 (1.3848481)	2.3552265 (1.9090575)	-0.6675430 (1.3916894)	-1.3250314 (2.0772470)
(continued)			*beta_5 (SE)*	*beta_6 (SE)*	*beta_7 (SE)*	*beta_8 (SE)*	*beta_9 (SE)*	
			0.1326285 (1.3583131)	-0.8455202 (1.7846930)	3.5486164 (1.8186160)	0.3416078 (0.1131649)		

Data S1. (separate file)

BIC summary table for all analyses “BICsummary.xls”.

Data S2. (separate file)

Raw data for size experiment “dissogeny_size_singletons.csv”. Individual cydippids were measured and cultured individually in 50 ml cultures as described in Methods. Measurements were repeated daily.

Data S3. (separate file)

Raw data for temperature experiment “temp_fecundity_revised.csv”. Groups of cydippids were cultured at two temperatures as described in Methods.

Data S4. (separate file)

Raw data for density experiment “dissogeny_density_1.csv”. Groups of 1, 2, 4, or 8 parental cydippids were kept in 50 ml cultures as described in Methods.

Data S5. (separate file)

Raw data for nutritional experiment “feed2.csv”. Groups of cydippids were fed one of four diets as described in Methods.

Data S6. (separate file)

Raw data for selected individuals that grow from ≤2 to ≥4 mm “single-Ml-spawn-growth_v3.csv”. This data set duplicates some data entries from data set S2 (individual and replicate identifiers are preserved in the data) as well as some additional individuals fed more heavily to promote rapid growth. Three of the latter individuals are shown in Supplemental Figure S1C, and our R code can be used to visualize any of these individuals’ growth vs. reproduction.

SI References

1. B. Dennis, P. L. Munholland, J. M. Scott, Estimation of Growth and Extinction Parameters for Endangered Species. *Ecological Monographs* **61**, 115–143 (2011).
2. D. F. Staples, M. L. Taper, B. B. Shepard, Risk-based viable population monitoring. *Conservation Biology* **19**, 1908–1916 (2005).
3. J. M. Ferguson, J. M. Ponciano, Predicting the process of extinction in experimental microcosms and accounting for interspecific interactions in single-species time series. *Ecology Letters* **17**, 251–259 (2014).
4. B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, D. F. Staples, Estimating density dependence, process noise, and observation error. *Ecological Monographs* **76**, 323–341 (2006).
5. J. M. Ponciano, L. de Gelder, E. M. Top, P. Joyce, The population biology of bacterial plasmids: A hidden Markov model approach. *Genetics* **176**, 957–968 (2007).
6. S. Karlin, *A first course in stochastic processes* (Academic Press, 2014).
7. L. J. S. Allen, An introduction to stochastic processes with applications to biology, second edition. *An Introduction to Stochastic Processes with Applications to Biology, Second Edition*, 1–461 (2010).
8. B. Dennis, Allee Effects: Population Growth, Critical Density, and the Chance Of Extinction. *Natural Resource Modeling* **3**, 481–538 (1989).
9. L. J. S. Allen, G. E. Lahodny, Extinction thresholds in deterministic and stochastic epidemic models. *Journal of Biological Dynamics* **6**, 590–611 (2012).
10. L. J. S. Allen, A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling* **2**, 128–142 (2017).

11. L. J. S. Allen, R. K. McCormack, C. B. Jonsson, Mathematical models for hantavirus infection in rodents. *Bulletin of Mathematical Biology* **68**, 511–524 (2006).
12. J. M. Ferguson, J. M. Ponciano, Evidence and implications of higher-order scaling in the environmental variation of animal population growth. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2782–2787 (2015).
13. N. L. Johnson, S. Kotz, N. Balakrishnan, "20 Pareto distributions" in *Continuous Univariate Distributions, Volume 1*, 2nd Ed., (John Wiley & Sons, 1994), p. 573.
14. A. Asanjarani, B. Liquet, Y. Nazarathy, Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches. *International Journal of Biostatistics*, 1–20 (2021).
15. P. McCullagh, J. A. Nelder, *Generalized linear models* (Routledge, 2019).
16. B. Dennis, J. M. Ponciano, M. L. Taper, S. R. Lele, Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC. *Frontiers in Ecology and Evolution* **7** (2019).
17. W. N. Venables, B. D. Ripley, *Modern applied statistics with S-PLUS.* (Springer Science and Business Media LLC, 2002).
18. J. M. Ponciano, M. L. Taper, B. Dennis, S. R. Lele, Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* **90**, 356–362 (2009).
19. S. R. Lele, B. Dennis, Bayesian Methods for Hierarchical Models: Are Ecologists Making a Faustian Bargain? *Ecological Applications* **19**, 581–584 (2009).
20. S. R. Lele, B. Dennis, F. Lutscher, Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters* **10**, 551–563 (2007).
21. M. Auger-Méthé, *et al.*, State-space models' dirty little secrets: Even simple linear Gaussian models can have estimation problems. *Scientific Reports* **6**, 1–10 (2016).
22. J. W. Smith, L. R. Johnson, R. Q. Thomas, Assessing Ecosystem State Space Models: Identifiability and Estimation. 1–37 (2021).
23. S. R. Lele, Model complexity and information in the data: Could it be a house built on sand? *Ecology* **91**, 3493–3496 (2010).
24. J. M. Ponciano, J. G. Burleigh, E. L. Braun, M. L. Taper, Assessing Parameter Identifiability in Phylogenetic Models Using Data Cloning. *Systematic Biology* **61**, 955–972 (2012).
25. M. L. Taper, J. M. Ponciano, Evidential statistics as a statistical modern synthesis to support 21st century science. *Population Ecology* **58**, 9–29 (2016).
26. S. R. Lele, Consequences of Lack of Parameterization Invariance of Non-informative Bayesian Analysis for Wildlife Management: Survival of San Joaquin Kit Fox and Declines in Amphibian Populations. *Frontiers in Ecology and Evolution* **7** (2020).
27. S. R. Lele, How Should We Quantify Uncertainty in Statistical Inference? *Frontiers in Ecology and Evolution* **8** (2020).
28. M. L. Taper, S. R. Lele, J. M. Ponciano, B. Dennis, C. L. Jerde, Assessing the global and local uncertainty of scientific evidence in the presence of model misspecification. *Frontiers in Ecology and Evolution* **In press** (2021).
29. M. Auger-Méthé, *et al.*, A guide to state–space modeling of ecological time series. *Ecological Monographs* **91**, e01470, 1–38 (2021).
30. R. Royall, *Statistical evidence: a likelihood paradigm* (Chapman & Hall, 1997).
31. R. Royall, On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association* **95**, 760–768 (2000).
32. R. M. Royall, "The Likelihood Paradigm for Statistical Evidence" in *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*, M. L. Taper, S. R. Lele, Eds. (University of Chicago Press, 2004), pp. 119–152.
33. M. L. Taper, S. R. Lele, Eds., *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations* (The University of Chicago Press, 2004).
34. C. L. Jerde, *et al.*, Strong Evidence for an Intraspecific Metabolic Scaling Coefficient Near 0.89 in Fish. *Frontiers in Physiology* **10**, 1–17 (2019).

35. S. R. Lele, “Evidence functions and the optimality of the law of likelihood” in *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, M. L. Taper, S. R. Lele, Eds. (The University of Chicago Press, 2004), pp. 191–216.
36. J. Neyman, E. S. Pearson, On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A* **231**, 289–337 (1933).
37. Z. Yang, T. Zhu, Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 1854–1859 (2018).
38. G. Schwarz, Estimating the Dimension of a Model. *The Annals of Statistics* **6**, 461–464 (1978).
39. R. Nishii, Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivariate Analysis* **27**, 392–403 (1988).
40. H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle. 199–213 (1998).
41. M. L. Taper, “Model identification from many candidates” in *Model Identification from Many Candidates*, M. L. Taper, S. R. Lele, Eds. (The University of Chicago Press, 2004), pp. 448–524.
42. D. R. Cox, *Planning of experiments.* (Wiley, 1958).
43. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Second (Springer, 2003).
44. K. P. Burnham, D. R. Anderson, K. P. Huyvaert, AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65**, 23–35 (2011).
45. C. Jaspers, *et al.*, Ctenophore population recruits entirely through larval reproduction in the central Baltic Sea. *Biology Letters* **8**, 809–812 (2012).
46. M. Q. Martindale, Larval reproduction in the ctenophore *Mnemiopsis mccradyi*. *Marine Biology* **94**, 409–414 (1987).
47. L. D. Baker, M. R. Reeve, Laboratory culture of the lobate ctenophore *Mnemiopsis mccradyi* with notes on feeding and fecundity. *Marine Biology* **26**, 57–62 (1974).
48. J. Hirota, “Laboratory Culture and Metabolism of the Planktonic Ctenophore, Pleurobrachia bachei A. Agassiz” in *Biological Oceanography of the Northern North Pacific Ocean*, (1972), pp. 465–484.
49. S. Wang, G. Zhang, K. Zhou, S. Sun, Long-term population variability and reproductive strategy of a northward expanded ctenophore Pleurobrachia globosa Moser, 1903 in a temperate bay, China. *Journal of Experimental Marine Biology and Ecology* **533**, 151457 (2020).