

## A Bayesian approach to estimating copepod development times from stage frequency data

Wim Kimmerer\* and Alison Gould

Romberg Tiburon Center, 3152 Paradise Drive, Tiburon, CA 94920

### Abstract

We developed Bayesian hierarchical models to estimate life stage durations of copepods from data on life stage frequencies over time in laboratory cohorts. This approach can determine stage duration or development rate, as well as other parameters of the development process, with probability distributions for each parameter. Uncertainty arising from sources such as experimental replication and the variability inherent in count data can easily be incorporated. Prior probability distributions can be uninformative, or they can apply constraints (e.g., stage durations > 0), general knowledge of development, or results of previous experiments. The approach is flexible, with the capability to model any number of life stages from experiments using replicated or unreplicated designs. We verified the model by accurately recovering the life stage distributions used to produce data in a simulation. We then applied the method to laboratory data on the development of two calanoid copepods and one cyclopoid copepod from the San Francisco Estuary. With replication (3 or 4 replicates), the method can determine stage durations with ~30 copepods per sample, although the uncertainty around estimates of stage duration increases as the number of copepods per sample decreases or the sampling interval increases.

Copepods develop through a series of fixed life stages from egg to adult. The duration of a life stage varies among stages and species and depends on temperature and the degree of food limitation (McLaren 1965; Vidal 1980). Estimates of stage duration are required for estimating growth by the molt rate method (Hirst et al. 2005) and mortality in field populations of copepods (Aksnes and Ohman 1996). Accurate estimates of stage duration and their errors are essential for population dynamic studies of copepods.

Our focus is on laboratory estimates of development time through all life stages or a series of stages (e.g., all nauplii or all copepodite stages). Two general experimental methods have been applied. In one method, individual copepods are observed over time, and times of molting are noted (Twombly

and Burns 1996) so that development times, with their errors, can be estimated directly.

In the second, more common method, copepods are reared in bulk cultures from eggs collected over a short time period, samples are taken over time, and the number in each life stage at each time is determined (e.g., Landry 1983). An approach is then needed to calculate stage durations from these data. In most papers, researchers have simply fit straight lines or logistic curves to the data on cumulative fraction at or beyond each stage and estimated cumulative development time by regression of cumulative fraction of each stage versus time (sometimes with transformation, Landry 1983). Then cumulative development for each stage is determined by interpolating to the median development time, or the time when half of the copepods have entered a given stage. This procedure is simple but can be subjective, provides no clear way of estimating error, and does not use all the available information (Klein Breteler et al. 1994). Because durations are usually determined by stage rather than jointly, these estimates do not take advantage of the similarity of cumulative development curves for successive stages. Furthermore, because estimates are based on fractions rather than count data, information is lost about uncertainty. Finally, it is not clear how to combine results of replicate experiments given that most of the information about errors within each replicate is lost. Thus, the statistical properties of the resulting estimates of stage duration are unknown (Klein Breteler et al. 1994).

\*Corresponding author: E-mail: kimmerer@sfsu.edu,  
telephone: 415-338-3515, fax: 415-435-7120

### Acknowledgments

We thank R. Mac Nally for convincing us to go Bayesian and J. Thomson for help in coding the models, and both for comments on an earlier draft. M.A. McCarthy's very readable book, *Bayesian methods for ecology*, was an enormous help. J. Durand, M.H. Nicolini, C. Peñalva, N. Ferm, T. Ignoffo, A. Slaughter, R. Paul, L. Sullivan, V. Greene, and L. Kara assisted with culturing phytoplankton and in conducting the development experiments. We thank M. Weaver for helpful comments on the manuscript. Three anonymous reviewers helped us to refine and improve this paper. Funding was provided by CALFED grant SCI-05-C107 and the California Interagency Ecological Program.

DOI 10.4319/lom.2010.8.118

Increasingly rigorous methods have been applied to fit curves to the cumulative development data. Klein Breteler et al. (1994) fit cumulative gamma functions to data from development experiments, using a binomial error distribution to allow for the increase in reliability of estimates with increasing numbers of copepods counted. This was a marked improvement over previous methods in that it allowed for error estimates in replicated or unreplicated measurements and solved some of the problems with methods applied previously. Furthermore, the gamma functions allowed for skewness in the frequency distributions of stage durations. However, Klein Breteler et al. (1994) could not resolve the joint distributions of individual stage durations by this method.

In this article, we present a Bayesian approach (Ellison 2004; Gelman et al. 2004) for estimating the durations of life stages in copepods using simulated data and experimental results from our laboratory. Bayesian methods have become common in fields such as evolutionary biology (Huelsenbeck et al. 2001) and cosmology (Trotta 2008), but remain somewhat less common in ecological applications (Ellison 2004; although see Clark 2005; Van den Meersche et al. 2008).

The chief advantages of the Bayesian approach for estimating stage durations are that it provides estimates of model parameters with their probability distributions and that it can use information available before the experiment begins. This information can be as rudimentary as the fact that stage durations must be  $> 0$ , or it can come from the results of previous studies conducted using similar or different methods. In addition, error from any source of data with any statistical distribution can be propagated into the results, and any parameter of the underlying model can be calculated as output with its own probability distribution.

## Materials and procedures

The approach uses a hierarchical model to allow individual stage durations and their probability distribution functions to be determined jointly and directly from the data, regardless whether the experiments were replicated. In this approach, the reported uncertainty is that of each estimate of stage duration and there is no need to subtract within-replicate variance to get the variance among replicates (Klein Breteler et al. 1994). Whereas Klein Breteler et al. (1994) used a gamma function to fit the cumulative fraction molted, we applied a skewed version of the logistic function because of its ease of implementation.

**Model**—Data (below) consisted of counts by life stage in successive samples of copepods from cultures begun with eggs collected over a 12- or 24-h period. These could include all stages or a subset of stages (e.g., nauplii), and stages or sexes could be pooled or kept separate. Samples were generally taken at fixed intervals (e.g., 12 or 24 h), and the experiments were replicated either simultaneously or sequentially; the intervals need not be identical for all replicates.

The key parameter of interest in the model was the vector of mean stage durations  $\bar{D}_s$ , one for each stage  $s$  except the terminal (i.e., adult) stage. To fit the model to the entire data set, the data were combined into a single block of counts with each row representing a single sample of all life stages, identified by index  $i$ , which links each sample to time  $t_i$  and experimental replicate  $r_i$ .

We initially fit the data using logistic curves but the asymmetry around the median cumulative development times introduced some bias into the results. Klein Breteler et al. (1994) fit a cumulative gamma distribution to stage distribution data to account for this asymmetry, and this practice has been followed by several other researchers (e.g., Souissi and Ban 2001). However, Klein Breteler et al. (1994) chose a fixed value of 3 for the shape parameter of the gamma distribution. It is more appropriate to leave that parameter to be fitted, but this makes the cumulative gamma distribution difficult to determine analytically. We therefore applied a skewed version of the logistic function to the fraction of the population at or above each successive stage in each time period, linking this fraction to the cumulative development time:

$$F_{s,i} = 1 - \left( \frac{1}{1 + (2^{1/b} - 1)e^{-a_s(DC_{s,i} - t_i)}} \right)^b \quad (1)$$

where  $F_{s,i}$  is the estimated proportion of copepods of stage  $\geq s$  in sample  $i$ ,  $a_s$  is a parameter representing the spread in time of the logistic curve for stage  $s$ ,  $DC_{s,i}$  is the cumulative development time to the beginning of stage  $s$  estimated in sample  $i$ , and  $t_i$  is time (days) from the beginning of the experiment (ideally the midpoint of the period of egg release or hatching) to the time of sample  $i$ . The parameter  $b$  ( $0 < b \leq 1$ ) determines the degree of asymmetry in the logistic curve, and the first term in parentheses in the denominator has been inserted to force the function to have a value of 0.5 when  $DC_{s,i} = t_i$  as in a logistic function. The function in Eq. 1 has a very similar shape to a cumulative gamma function, the principal difference being a smoother initial rise from zero of Eq. 1 compared to the gamma function, which increases rather abruptly. When Eq. 1 was fit to data generated by cumulative gamma functions with randomly selected parameters the two functions were always correlated at  $r > 0.999$ .

Stage duration  $D_{s,r}$  for each replicate  $r$  is estimated as the difference between successive cumulative development times for each sample in which the stage occurs

$$\hat{D}_{s,r} = DC_{s+1,i} - DC_{s,i} \quad (2)$$

For each sample  $i$ , the estimated fraction of copepods in each life stage is

$$f_{s,i} \max[0, \min(1, F_{s+1,i} + F_{s,i})] \quad (3)$$

where  $f_{s,i}$  is the difference in the  $F$  values for successive life stages and the truncation is to ensure no values exceed the bounds  $[0,1]$ , preventing occasional errors due to rounding.

The value for the last stage  $s_{\max}$  (either the last stage counted or the terminal adult stage) was calculated by difference so the  $f_{s,i}$  values would sum exactly to 1:

$$f_{s_{\max},i} = 1 - \sum_{s=1}^{s_{\max}-1} f_{s,i} \quad (4)$$

The frequency distributions  $f$  were fitted to the data as a series of multinomial distributions

$$n_{\bullet,i} \sim \text{multinomial}(f_{\bullet,i}, N_i) \quad (5)$$

where the symbol  $\bullet$  denotes a vector comprising all observed life stages,  $n_{\bullet,i}$  is the number of copepods of each stage observed in sample  $i$ , and  $N_i$  is the total number of copepods counted in sample  $i$ . In cases where development times were determined separately for males and females, we conducted a separate analysis for each sex, decreasing the expected frequencies for sexed stages in Eq. 5 by half (assuming a 50:50 sex ratio in earlier, unsexed stages), then adjusted all frequencies,  $f_{\bullet,i}$ , proportionally to sum to 1.

The parameters  $a_s$  and  $b$  in Eq. 1 introduce some redundancy in that both influence how the durations of the later, longer stages spread in time. To reduce the effect of this redundancy during model fitting, we assumed that the spread parameter  $a_s$  was a linear function of stage:

$$a_s = a_0 + a_1 (s - s_{\text{mid}}) \quad (6)$$

where  $a_0$  and  $a_1$  are parameters to be fit,  $s$  is stage and  $s_{\text{mid}}$  is the middle stage, used to center the regression making the fitting process more efficient.

Prior distributions for the model parameters were generally uninformative except when results of one experiment were available for use in the next. Priors could be set based on previous results from similar species, but there was little difference between results with uninformative and informative priors. Priors were uninformative beyond the assumption that stage durations were unlikely to be more than a few days:

$$D_{s,r} \sim \max[0^+, \text{Normal}(5, 32)] \quad (7)$$

$$\begin{aligned} a_0 &\sim \max[0^+, \text{Normal}(2, 32)] \\ a_1 &\sim \text{Normal}(0, 1) \\ b &\sim \text{beta}(1, 1) \end{aligned} \quad (8)$$

The truncations in Eqs. 7 and 8 are to ensure that values are greater than 0. Note that the priors for stage duration (Eq. 7) are very flat, within 1.5% of the maximum value for durations between 0 and 20 d; a uniform distribution between 0 and 100 d gave the same results in our analyses of simulated data (stage durations within 0.02 d). The prior for  $a_0$  was equally uninformative but with a mean of 2 under the expectation that the true value was likely to be < 10. The prior for  $a_1$  was relatively tight since we were aware that the spread of stage durations did not change much across the life stages we were observing.

The prior for  $b$  was constant between 0<sup>+</sup> and 1<sup>-</sup>, where the requirement that it be < 1 is because the asymmetry is a right skew due to late-developing copepods.

Outputs included the stage duration  $\bar{D}_s$  and the shape parameters  $a_0$ ,  $a_1$ , and  $b$ . In addition, the difference between the mean of all stage durations in each replicate and the grand mean was determined to examine the size of replication errors.

Stage durations  $\bar{D}_s$  were determined using Bayesian hierarchical models (Gelman et al. 2004) fitted using WinBUGS version 1.4.3 (this name is not redundant: BUGS is an acronym for Bayesian inference Using Gibbs Sampling; Lunn et al. 2000). Each model was run with triplicate Markov chains of 10,000 samples following a 1,000-sample burn-in after 10-fold thinning (i.e., retaining only every 10th point, leaving 10,000 points from each chain). Results from the first and second half of the retained samples were nearly identical, indicating that both the burn-in and sample length were adequate. Gelman-Rubin statistics (Gelman et al. 2004) and plots of autocorrelation and time series of simulations were used as a check on model convergence. Results are presented as means with 95% credible intervals, which are the 2.5 and 97.5 percentiles of the sample distribution (i.e., the 3 × 10,000 samples), and are roughly equivalent to confidence intervals. WinBUGS code will be provided on request to the first author.

**Data**—Simulated data were generated by sampling from a gamma distribution with a shape parameter of 3 (Klein Breteler et al. 1994) for relative development times of 100,000 simulated copepods, and calculating the realized development time through 4 stages with nominal durations of 5, 1, 2, and 4 d, and a terminal adult stage. The data were put in a matrix of numbers of each stage on each day, which was sampled randomly in triplicate for each day of the model run. Trials were run with various numbers of copepods per sample to determine the effect of small sample sizes. Additional runs were made: 1) with 50 copepods assuming that the last 2 stages were sexed and had a 50:50 sex ratio (i.e., there were ultimately 25 in the terminal adult stage), and 2) with a gamma shape parameter of 100 indicating much less skew (as found by Souissi and Ban 2001).

Laboratory data were obtained from four experiments in which copepods were grown through multiple life stages. Copepods (the calanoids *Eurytemora affinis* and *Pseudodiaptomus forbesi* and the cyclopoid *Limnithona tetraspina*) were collected from the San Francisco Estuary at salinity of 1–5.

Calanoid copepods were collected by gentle near-surface net tows and brought to the laboratory in insulated 20-L buckets full of surface water from the collection site. Copepods were maintained in culture from several days to a few weeks at 15°C on a 12 h light:12 h dark cycle using mixed cultured phytoplankton as food (*Rhodomonas salina*, *Thalassiosira* sp., and *Nannochloropsis oculata*). Several hundred gravid females of the target species, or all in the sample, were isolated under a dissecting microscope. Copepods were placed in a filter made of a length of PVC pipe with a 150-μm mesh glued to one end,

**Table 1.** Results of analysis of simulated copepod development data, presented as means with 95% credible intervals. The nominal duration (days) was used to establish a population of 100,000 copepods with 5 stages, the final being the terminal stage (adult). The gamma shape parameter was 3 (highly skewed) except in the last sample, where it was 100 (nearly symmetrical). The calculated duration was obtained by fitting Eq. 1 to the cumulative proportions by stage for all data. For each column, three replicate random samples of the number of copepods shown were taken on each day, and the Bayesian algorithm was used to estimate stage durations.

	Stage			
	1	2	3	4
Nominal duration	5	1	2	4
Calculated duration	4.99	0.97	1.99	3.97
N = 10,000 NS*	4.91 ± 0.01	1.03 ± 0.01	1.89 ± 0.01	3.95 ± 0.01
N = 1,000 NS*	4.92 ± 0.02	1.04 ± 0.03	1.88 ± 0.03	3.97 ± 0.04
N = 100 NS*	4.92 ± 0.06	1.04 ± 0.08	1.91 ± 0.10	3.95 ± 0.11
N = 50 NS*	4.92 ± 0.09	1.04 ± 0.12	1.92 ± 0.14	3.94 ± 0.16
N = 50 Sexed	4.84 ± 0.10	1.05 ± 0.12	1.89 ± 0.16	3.98 ± 0.20
N = 15 NS*	4.87 ± 0.16	1.06 ± 0.21	1.93 ± 0.24	3.98 ± 0.27
N = 50 NS*; shape = 100	4.96 ± 0.05	1.00 ± 0.08	1.99 ± 0.10	3.98 ± 0.13

\*Not sexed.

suspended in a 15-L vessel containing water from the collection site. After 24 h, the filter container was lifted from the vessel, leaving eggs and nauplii, and placed in a replicate incubating vessel for another 24-h period. This process was repeated a third time to get triplicate experiments with a 1-d lag between replicates. From the time the adults were removed for every 24 h until copepods reached adulthood, the incubation vessels were stirred and 500 mL samples were taken. Samples were concentrated on a 35- $\mu$ m filter, preserved in 2% buffered formaldehyde, and counted the same day. Nauplii were counted but not identified to stage, and all copepodites were identified to stage. The experiment for *P. forbesi* was begun with a smaller number of adults than that for *E. affinis*, resulting in smaller recoveries of each life stage and wider confidence limits on the results.

The procedure for *Limnoithona tetraspina* differed slightly from that for calanoids. Filter mesh size was 80  $\mu$ m, and adults were obtained by size fractionation rather than by sorting. Experimental temperature was 18°C. Two sets of replicates were prepared during each of 2 successive 12-h periods for a total of 4 replicates. Cultures were maintained with daily additions of a mixture of *Nannochloropsis oculata* and a dilution of Instant Algae brand Shellfish Diet 1800®, which was incubated in estuarine water to develop a natural microzooplankton assemblage to maintain food-saturated conditions. Subsamples of at least 30 individuals were removed every 12 h, stained with Rose Bengal, and preserved in 5% formaldehyde for later counting under an inverted microscope. All stages were identified, and sexes were identified for copepodite stage 5 (C5) and adult.

Two separate experiments were run with *L. tetraspina*. In the first (April 2008), development was slower than expected, resulting in depletion of the cultures so that sample sizes became too small for reliable estimates of stage duration beyond stage ~C2. Therefore, the experiment was repeated in November 2008 except that cultures were not sampled until

late-stage nauplii were predominant. The first experiment was used to determine naupliar stage durations and to provide informative priors for the second experiment on stage duration of copepodites.

In the November experiment, stage durations were determined either with the April experiment as priors, or with uninformative priors to determine the influence of the priors on the posterior distributions. In addition, because the C5 and adult stages were identified to sex, we also calculated stage durations with sexes combined, and separately for females and males. For these calculations, we assumed the C4 stage to have a 50:50 sex ratio. Counts of total C5 and adults revealed that the sex ratio was close to 50:50; although we did not measure mortality, it must have been low because nearly all the copepods collected appeared intact upon examination.

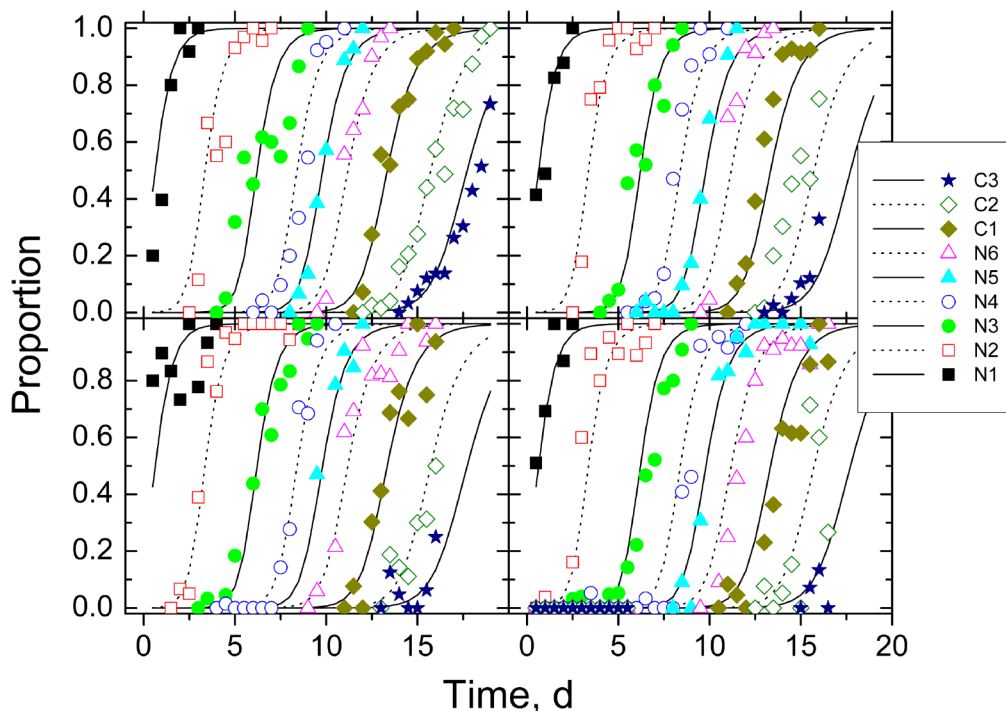
## Assessment

The results from simulations (Table 1) verified the model for estimating stage duration. The model fit to the proportions by life stage from the entire data set closely matched the nominal durations. The mean values with 10,000 copepods per sample were slightly off from the nominal durations, apparently because the temporal distributions of life stages were calculated on the basis of whole days. The values for smaller numbers of copepods per sample included the means for 10,000 copepods within their credible intervals in every case, and credible intervals based on 50 or fewer copepods included the nominal stage durations. Results using 15 copepods had broader credible intervals but were otherwise similar to the results for larger sample sizes. Results for sexed individuals hardly differed from those for unsexed individuals. Results for a lower degree of skewness (shape = 100) were very close to nominal values.

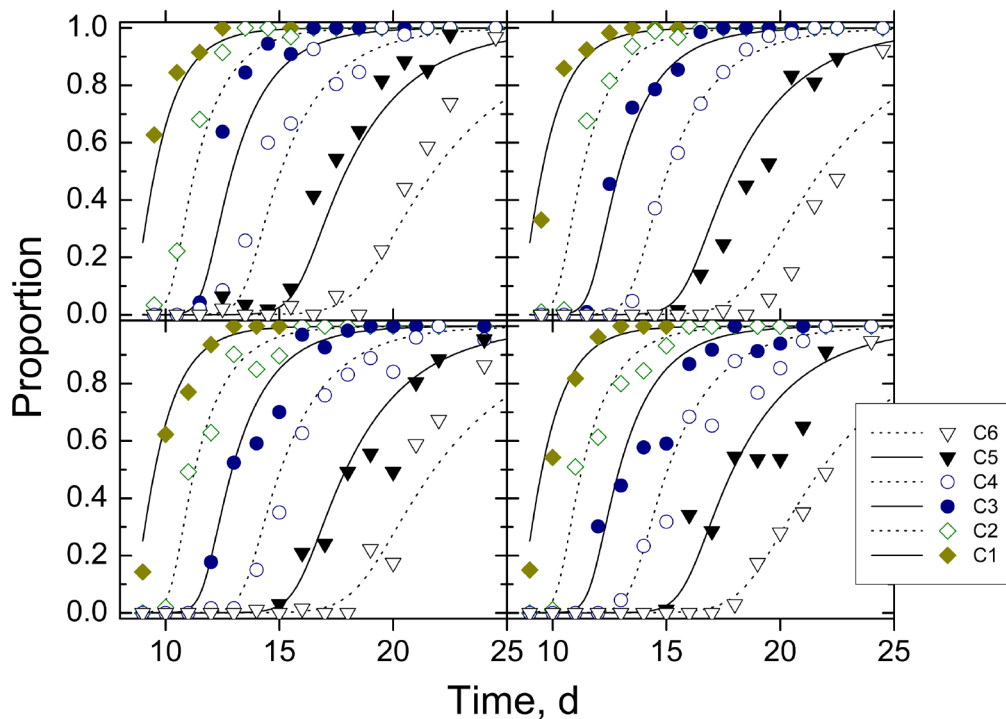
We focus mainly on the results for *Limnoithona tetraspina* since the laboratory data were more complete than those for the other species. Proportions at or beyond each stage show

the usual logistic-like pattern through time as the copepods developed (Figs. 1 and 2). There was no obvious increase in the spread of the stage duration within the nauplii, but the

later copepodites had a broader distribution in time than the nauplii. Bayesian 95% credible intervals (Table 2) were generally around 5% to 10% of the stage durations and higher for



**Fig. 1.** *Limnoithona tetraspina*. April 2008 experiment. Predicted (lines, from parameters of the Bayesian model) and observed (symbols) cumulative fraction in or beyond each stage. Each panel represents a different replicate experiment.



**Fig. 2.** *Limnoithona tetraspina*. As in Fig. 1 for later stages from the November 2008 experiment. Sexes are combined for C5 and adult. Lines are for the Bayesian model with April results (Fig. 1) used as priors.

**Table 2.** Copepod life stage durations and other parameters determined in this study, including stage durations and parameters for spread of the curves ( $a_0$  and  $a_1$ ) and skew ( $b$ , see Eqs. 1 and 6). Each value is the mean with Bayesian 95% credible intervals, which are very close to 1.96 times the standard errors. The first row gives the mean number of copepods per sample for each analysis. Results listed as “Female” and “Male” had only stages C5 and adult sexed, and used April priors.

Life stage/ parameter	<i>Limnoithona tetraspina</i>					<i>Eurytemora</i>	<i>Pseudodiaptomus</i>
	April 2008 Uninformative priors, $\bar{N} = 30$	Uninformative priors, $\bar{N} = 66$	April priors, $\bar{N} = 66$	Female, $\bar{N} = 59$	Male, $\bar{N} = 58$	<i>affinis</i> , $\bar{N} = 37$	<i>forbesi</i> , $\bar{N} = 15$
Egg	0.6 ± 0.10						
N1	2.7 ± 0.14					6.8 ± 0.11	8.4 ± 0.36
N2	2.8 ± 0.15						
N3	2.2 ± 0.16						
N4	1.4 ± 0.15						
N5	1.4 ± 0.15						
N6	2.2 ± 0.16						
C1	2.4 ± 0.19	1.7 ± 0.13	2.2 ± 0.10	2.2 ± 0.08	2.2 ± 0.08	1.3 ± 0.15	1.3 ± 0.39
C2	2.0 ± 0.34	1.6 ± 0.11	1.7 ± 0.11	1.7 ± 0.09	1.7 ± 0.09	1.6 ± 0.17	1.3 ± 0.32
C3		2.1 ± 0.13	2.1 ± 0.15	2.1 ± 0.12	2.1 ± 0.12	1.0 ± 0.16	1.2 ± 0.32
C4		2.9 ± 0.16	2.9 ± 0.18	2.8 ± 0.19	3.0 ± 0.19	1.1 ± 0.17	1.9 ± 0.46
C5		3.9 ± 0.32	3.7 ± 0.23	4.4 ± 0.43	3.3 ± 0.43	2.1 ± 0.22	2.7 ± 0.52
$a_0$	1.85 ± 0.17	3.31 ± 0.50	2.9 ± 0.5	2.9 ± 0.5	3.6 ± 0.7	2.2 ± 0.3	2.7 ± 0.8
$a_1$	-0.18 ± 0.04	-0.68 ± 0.14	-0.61 ± 0.15	-0.63 ± 0.17	-0.83 ± 0.22	-0.28 ± 0.12	-0.43 ± 0.12
$b$	0.58 ± 0.09	0.18 ± 0.03	0.17 ± 0.03	0.17 ± 0.04	0.13 ± 0.03	0.7 ± 0.2	0.3 ± 0.13

later stages. The tighter credible intervals for C1 and C2 in the November experiment compared with the April experiment was partly a result of higher numbers of copepods (especially for C2), but also a result of the refinement in the estimates obtained by using the April results as a prior for the November analysis. This is made plain by comparing the November results with those obtained using uninformative priors (Table 2). The 95% credible interval for C1 but not C2 was smaller with the April results used as priors for the November experiment because of the lower confidence in the value for C2 in the April experiment. The larger credible intervals for the C5 stages with sexes separate likely occurred because we did not increase sample size to allow for sexing of the copepods, so the numbers counted in each sample were about half of those for the earlier, unsexed stages. The stage durations and credible intervals for unsexed stages did not differ whether the later stages were males, females, or both.

Predicted and observed fractions of copepods that had molted into each stage at each time (Fig. 3) from the November experiment were reasonably well tracked by the 1:1 line, indicating a good fit of Eq. 1 to the data. The fit of a generalized additive model with a loess smoother was similar, indicating the straight-line fit was reasonable.

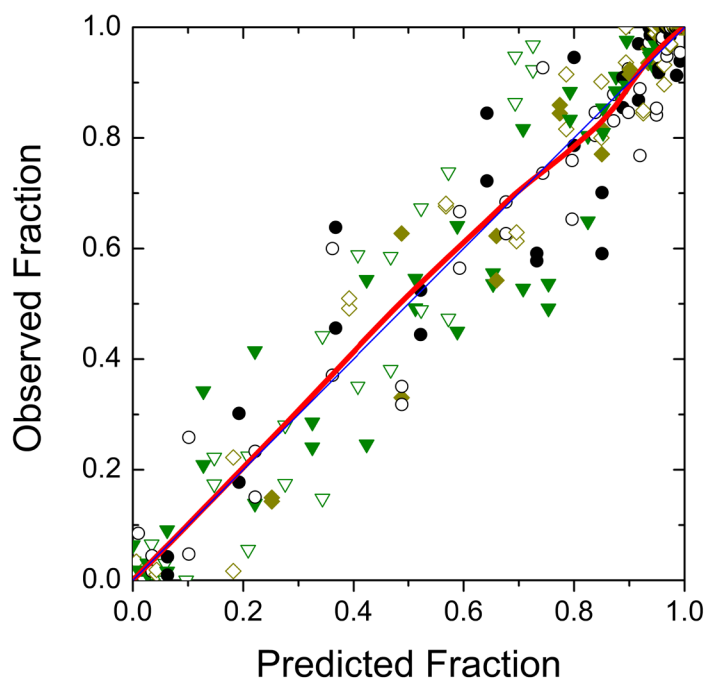
Probability density curves for stage duration from the November experiment (Fig. 4) show a roughly normal distribution with little indication of skewness or other distortion. The jagged appearance near the tops of some of the curves arose from rounding error in the output. Normal probability plots of these results (not shown) indicated that each was very

close to a normal distribution. The spread in time of the later stages is apparent here.

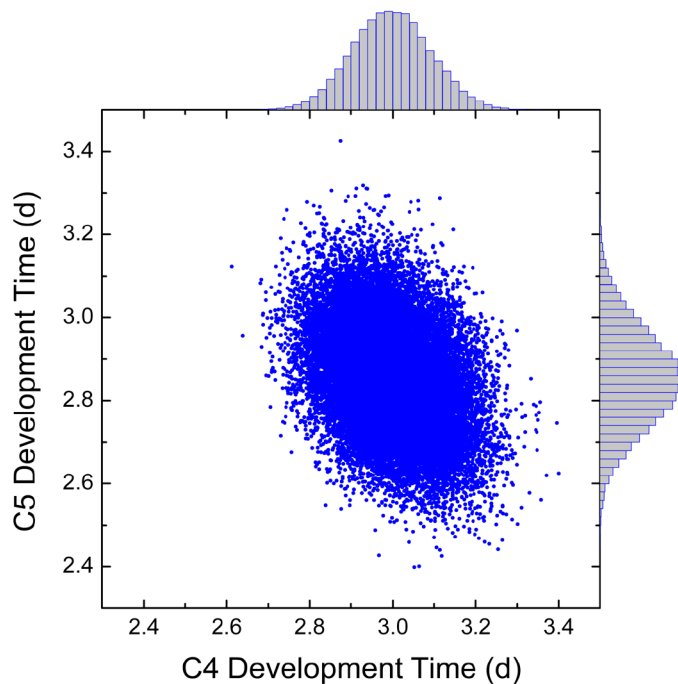
Estimates of duration of successive stages were weakly and negatively correlated ( $r$  between -0.34 and -0.09, Fig. 5). To determine the influence of this correlation on the estimates, we recalculated means of individual stages from those Markov chain samples (of the original 30,000) in which the estimates of duration for the previous and succeeding stage were within 1 standard deviation of the mean to reduce the effect of variation in one stage on the estimate of duration of adjacent stages. The recalculated stage durations had means and standard deviations that differed in the second or third decimal place from the values calculated with the whole data set, indicating that the between-stage correlation had only a minor effect on the estimates.

Stage durations for *Eurytemora affinis* were much shorter, and estimates were somewhat less precise, than those for equivalent stages of *L. tetraspina*. The longer sampling interval in relation to stage durations and the smaller number of copepods per sample contributed to higher credible intervals in relation to estimated stage duration (Table 2). Results for *Pseudodiaptomus forbesi* had still lower precision, as expected from the smaller sample size. Nevertheless, there is a clear difference in stage durations between the two calanoid species, with *P. forbesi* taking ~1.5 d longer to develop through the naupliar stages than *E. affinis* and having longer mean durations for the C4 and C5 stages.

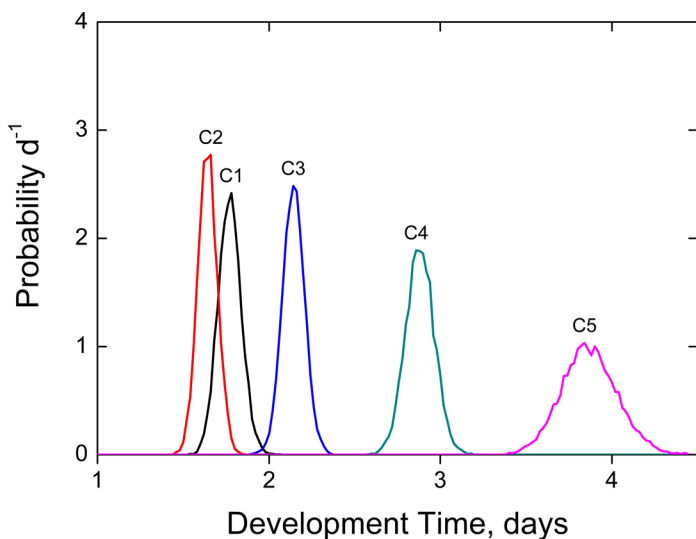
Results differed somewhat among replicate experiments (Table 3). Bayesian credible intervals for the experiment-specific



**Fig. 3.** *Limnoithona tetraspina*. Predicted (lines in Fig. 2 from Bayesian model) and observed (data points in Fig. 2) cumulative fraction molted for all life stages from the November experiment. Symbols as in Fig. 2. Thin line is the 1:1 line. The heavy line is a generalized additive model fit to the data with a loess smoother, span = 0.5 and degree = 1.



**Fig. 5.** *Limnoithona tetraspina*. Scatter plot of stage duration  $\bar{D}_i$  for the two most-correlated ( $r = -0.34$ ) pair of successive stages. Data as in Fig. 4. The histograms give the marginal frequency distributions for each development time ( $n = 30,000$  from 3 independent Markov chains).



**Fig. 4.** *Limnoithona tetraspina*. Probability density of development time for copepodites from the November 2008 experiment with April priors for C1 and C2, sexes combined.

error excluded zero for several replicates in each series, including all those for *E. affinis*. We interpret this as simply normal experimental variability. Quantitatively the replicate errors were small and, since they apply for the entire development period of the copepods, they do not appear to contribute much to the uncertainty in estimates of individual stage duration.

## Discussion

Our simulations demonstrate that the Bayesian approach presented here can recover the actual stage durations (Table 1). These results indicate that even with as few as 15 copepods per sample, reasonable estimates can be obtained with triplicate samples; however, more copepods per sample is preferable, especially if stage durations are to be determined separately for each sex. Our results also demonstrate the influence of prior distributions on the posterior distributions of parameters. The results for *Limnoithona tetraspina* early copepodites were based on the April experiment in which confidence limits became large because of dwindling numbers of copepods, and the November experiment in which numbers were high and confidence limits tight. The April priors had a modest effect on the November results for C1, for which the prior distribution had tight confidence limits, but only a minor influence on C2 for which the prior was less certain (Table 2). This reinforces the general finding in Bayesian analyses that vague priors have little influence on posterior distributions if the data are informative.

An advantage of the Bayesian approach for determining stage duration is that it can generate estimates of stage durations with probability distributions (Fig. 4) from which confidence intervals can be estimated. An alternative way to determine confidence limits on stage duration would be to propagate uncertainties from calculations done using other methods, e.g., by maximum likelihood (Klein Breteler et al. 1994). This requires some assumptions or knowledge about

**Table 3.** Differences of mean stage duration between replicate experiments and grand means. Data are means with 95% credible intervals.

Replicate	<i>Limnoithona tetraspina</i>				<i>Eurytemora affinis</i>	<i>Pseudodiaptomus forbesi</i>
	April 2008	November 2008				
		Female	Male			
1	-0.01 ± 0.05	0.01 ± 0.09	0.03 ± 0.10	-0.13 ± 0.06	-0.09 ± 0.16	
2	-0.11 ± 0.05	0.27 ± 0.10	0.15 ± 0.09	0.22 ± 0.06	-0.23 ± 0.16	
3	-0.02 ± 0.08	-0.16 ± 0.09	-0.06 ± 0.09	-0.09 ± 0.06	0.32 ± 0.13	
4	0.14 ± 0.10	-0.12 ± 0.08	-0.12 ± 0.09	—	—	

the form of the distributions. Although our parameter estimates appeared to have nearly normal distributions (Fig. 4), this may not always be the case. Furthermore, it is not clear how to estimate errors given a limited number of replicate experiments, each of which contains measurable uncertainty arising from the use of count data. In addition, estimates of durations of successive life stages can be negatively correlated, complicating error propagation. We found that this correlation had little influence on the estimates of the stage durations or their standard errors (Fig. 5), but this might not be the case with different data.

McLaren (1995) discussed the distributions of stage durations, in particular their right-skewness, and preferred to use development rate (1/stage duration) to fit data to models using temperature. The Bayesian approach makes it easy (once the model has been set up) to extract development rate or other parameters of interest with their probability distributions.

The use of a multinomial error distribution takes into account the variation inherent in count data from multiple categories. In contrast to many fitting algorithms (e.g., median stage duration, Landry 1983), this takes the greatest advantage of the increase in precision of estimates as the sample size increases.

Although the Bayesian approach can provide the full probability distribution of the stage durations (Fig. 4), this reflects uncertainty in the model parameters for the population, rather than any characteristic of individual variability within the copepod population. This is demonstrated by the lack of skewness in these curves despite obvious skewness in the distributions of individual molting times (Figs. 1 and 2), a consequence of the Central Limit Theorem. The probability distributions of stage durations in Table 2 would be suitable for use in determining population-level characteristics such as mortality estimates using a life-table approach. Individual variability contributes to the stage-specific spread  $a_i$ , although that parameter is also influenced by the duration over which eggs are collected to begin the experiment. In addition, the right-skewness that sometimes appears in stage development data may indicate the presence of “laggards” (Klein Breteler et al. 1994), slow-growing individuals that may have an important effect on emergent population properties. Individual variability would also be necessary input to an individual-based

model. Individual variability could be determined by this approach if the starting time interval were very short, but experiments in which the molting rates of individuals are observed may be more efficient for this purpose (Twombly and Burns 1996).

Other population parameters could be calculated using a Bayesian approach because of its flexibility and the ease of propagating error. For example, estimates of growth rate by the molt rate method (Hirst et al. 2005) require estimates of development time or rate from molt-rate experiments and estimates of weight or carbon per life stage, both of which have substantial measurement error. These estimates also require an underlying model of growth that establishes the relationship between mean weight within a stage (which is measured) and weight upon molting into or out of a stage (which is unobservable; Hirst et al. 2005). Bayesian approaches provide the necessary flexibility in modeling while correctly propagating the measurement errors, and we have applied them to the measurement of growth in *L. tetraspina* (Gould and Kimmerer unpubl. data). The estimation of mortality by a vertical life table method (Kimmerer and McKinnon 1987; Aksnes and Ohman 1996) uses similar mathematics and requires similar assumptions to the estimate of growth, and a Bayesian approach is suitable for that purpose as well.

## References

- Aksnes, D. L., and M. D. Ohman. 1996. A vertical life table approach to zooplankton mortality estimation. *Limnol. Oceanogr.* 41:1461-1469.
- Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8:2-14 [doi:10.1111/j.1461-0248.2004.00702.x].
- Ellison, A. M. 2004. Bayesian inference in ecology. *Ecol. Lett.* 7:509-520 [doi:10.1111/j.1461-0248.2004.00603.x].
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis, 2nd ed. CRC Press.
- Hirst, A. G., W. T. Peterson, and P. Rothery. 2005. Errors in juvenile copepod growth rate estimates are widespread: problems with the Moulting Rate method. *Mar. Ecol. Prog. Ser.* 296:263-279 [doi:10.3354/meps296263].
- Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evo-



lutionary biology. *Science* 294:2310-2314 [doi:10.1126/science.1065889].

Kimmerer, W. J., and A. D. Mckinnon. 1987. Growth, mortality, and secondary production of the copepod *Acartia tranteri* in Westernport Bay, Australia. *Limnol. Oceanogr.* 32:14-28 [doi:10.4319/lo.1987.32.1.0014].

Klein Breteler, W. C. M., N. Schogt, and J. Van Der Meer. 1994. The duration of copepod life stages estimated from stage-frequency data. *J. Plankton Res.* 16:1039-1057 [doi:10.1093/plankt/16.8.1039].

Landry, M. R. 1983. The development of marine calanoid copepods with comment on the isochronal rule. *Limnol. Oceanogr.* 28:614-624 [doi:10.4319/lo.1983.28.4.0614].

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS - A Bayesian modeling framework: Concepts, structure, and extensibility. *Stat. Comput.* 10:325-337 [doi:10.1023/A:1008929526011].

Mclaren, I. A. 1965. Some relationships between temperature and egg size, body size, development rate, and fecundity, of the copepod *Pseudocalanus*. *Limnol. Oceanogr.* 10:528-538 [doi:10.4319/lo.1965.10.4.0528].

———. 1995. Temperature-dependent development in marine copepods: comments on choices of models. *J. Plankton Res.* 16:1385-1390.

Souissi, S., and S. Ban. 2001. The consequences of individual variability in moulting probability and the aggregation of stages for modelling copepod population dynamics. *J. Plankton Res.* 23:1279-1296 [doi:10.1093/plankt/23.11.1279].

Trotta, R. 2008. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* 49:71-104 [doi:10.1080/00107510802066753].

Twombly, S., and C. W. Burns. 1996. Exuvium analysis: A nondestructive method of analyzing copepod growth and development. *Limnol. Oceanogr.* 41:1324-1329.

Van Den Meersche, K., K. Soetaert, and J. J. Middelburg. 2008. A Bayesian Compositional Estimator for microbial taxonomy based on biomarkers. *Limnol. Oceanogr. Methods* 6:190-199.

Vidal, J. 1980. Physioecology of zooplankton. II. Effects of phytoplankton concentration, temperature, and body size on the development and molting rates of *Calanus pacificus* and *Pseudocalanus* sp. *Mar. Biol.* 56:135-146 [doi:10.1007/BF00397130].

Submitted 1 August 2009

Revised 18 December 2009

Accepted 20 January 2010