bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A tutorial on Bayesian parameter inference for dynamic energy budget models

Philipp H. Boersch-Supan[A,B,C,1] *and Leah R. Johnson*[D,2]

February 3, 2018

[A]*Department of Geography, University of Florida, Gainesville, FL 32611, USA*
[B]*Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA*
[C]*Department of Integrative Biology, University of South Florida, Tampa, FL 33610, USA*
[D]*Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[1]*pboesu@gmail.com - ORCID ID 0000-0001-6723-6833*
[2]*lrjohn@vt.edu - ORCID ID 0000-0002-9922-579X*

Abstract

Mechanistic representations of individual life-history trajectories are powerful tools for the prediction of organismal growth, reproduction and survival under novel environmental conditions. Dynamic energy budget (DEB) theory provides compact models to describe the acquisition and allocation of organisms over their full life cycle of bioenergetics. However, estimating DEB model parameters, and their associated uncertainties and covariances, is not trivial. Bayesian inference provides a coherent way to estimate parameter uncertainty, and propagate it through the model, while also making use of prior information to constrain the parameter space. We outline a Bayesian inference approach for energy budget models and provide two case studies – based on a simplified DEBkiss model, and the standard DEB model – detailing the implementation of such inference procedures using the open-source software package deBInfer. We demonstrate how DEB and DEBkiss parameters can be estimated in a Bayesian framework, but our results also highlight the difficulty of identifying DEB model parameters which serves as a reminder that fitting these models requires statistical caution.

Key words: dynamic energy budget theory, parameter identifiability, informative priors

1 Introduction

Dynamic energy budget (DEB) theory (Kooijman, 2010) provides a powerful and well tested framework to describe the acquisition and use of energy by individual organisms over their entire life cycle. The standard DEB model is built on rules inherent to the process of resource uptake and allocation by all heterotrophs. It is a compact model that is able to describe the full life cycle bioenergetics of a living animal (Kooijman, 2010). DEB models are used as tools to address both fundamental and applied questions in ecology, e.g. about metabolic scaling (Maino et al., 2014), life-history strategies (Kooijman, 2013), in ecotoxicology (Jager et al., 2006; Jager and Zimmer, 2012), or as components of population models (Martin et al., 2012; Smallegange et al., 2017).

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Because of strong correlations between individual parameters, simultaneous parameter inference for DEB models is not trivial. The difficulty of estimation is by no means unique to DEB models, but is commonly encountered in dynamic systems models across scientific disciplines (Aster et al., 2011; Brewer et al., 2008; Johnson and Briggs, 2011). Parameter inference for DEB model parameters themselves is routinely based on a non-linear least squares regression approach, the so-called covariation method (Lika et al., 2011). The covariation method makes use of constraints on parameters that follow from theory (Lika et al., 2014, 2011), as well as by reducing the effective number of parameters by the use of so-called pseudo data: canonical fixed values for certain parameters which enter the estimation procedure with low weights. This approach has been successfully used to parameterize DEB models for over 1000 species. However, one drawback of the method is that uncertainty estimates of parameters are not readily available. Furthermore, while the DEB literature acknowledges the importance of biological variability (e.g. Lika et al., 2014), input data are treated as known without error for the purposes of the parameter estimation. While measurement error for many observable traits used to parameterize DEB models may indeed be small, trait data often exhibits heterogeneity between individuals of a species, which would be expected to reflect individual heterogeneity in the underlying metabolic parameters. Given the potential of DEB theory as a building block for population models, we feel these are important hurdles to overcome, so more value can be added to DEB-based population models by incorporating both better estimates of parameter uncertainty, and a better reflection of individual variability.

In contexts where quantification of uncertainty in parameters is desired, the Bayesian inference framework offers multiple advantages. First, multiple sources of uncertainty can be accounted for. Second, the use of informative priors can constrain the parameter space to biologically sensible outcomes while allowing fairly straightforward assessment of the influence of the prior information. Finally, the implementation of hierarchical models which allow inferences about both individual and population heterogeneity, as well as partial information pooling across individuals, is conceptually straightforward (Gelman et al., 1996).

Bayesian parameter inference for DEB models has been demonstrated by Johnson et al. (2013). However, until recently, general inference for these models in a Bayesian framework has required that the practitioner be able to implement the full inference procedure, from the differential equation model specification through to the sampler used to obtain posterior draws. Here, we present a tutorial for the estimation of model parameters for dynamic energy budget models using the open-source R package `deBInfer` (Boersch-Supan et al., 2017) which makes the approach simpler to implement. We present this tutorial using two case studies. The first is based on a DEBkiss model, a simplified bioenergetic model that builds on DEB theory Jager et al. (2013). We follow this with a case study based on the standard DEB model. In each case study we discuss how the model is implemented and the approach estimate parameters. The R and C code needed to reproduce all of the analyses are available as supplementary materials.

2 Basics of the Bayesian Approach

Bayesian approaches for parameter estimation in complex, nonlinear models are being applied with increasing frequency to a broad range of biological models (e.g. Coelho et al., 2011; Voyles et al., 2012; Johnson et al., 2013; Smith et al., 2015). Here we very briefly explain the rationale behind the approach. Further details on the statistical background and software implementation can be found in the literature, (e.g. Clark, 2007; Gelman et al., 2003; Hobbs and Hooten, 2015; Johnson

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

et al., 2013; Boersch-Supan et al., 2017).

In the Bayesian approach the model, and in particular its parameters, are viewed as random variables. In contrast, the data, once observed, are treated as fixed. This approach to parameter inference is attractive, as it provides a coherent framework that allows the incorporation of uncertainty in the observation process and model parameters, and it relaxes the assumption of normal errors that is inherent in least-squares fitting. It provides us not only with full posterior probability distributions describing the parameters, but also with posterior distributions for any quantity derived from the parameters, including the model trajectories. Further, the Bayesian framework naturally lets us incorporate prior information about the parameter values and examine the sensitivity of our inferences to this incorporated information. This is particularly useful in the context of DEB theory, where there are known biological and theoretical constraints on parameters (Lika et al., 2011, 2014; Johnson et al., 2013). For example, many biological quantities, such as body size, are strictly non-negative, and most DEB parameters have at least one well defined bound, e.g. the allocation fraction κ is bounded on the interval [0,1]. Using informative priors can help constrain the parameter space, aiding with parameter identifiability.

The purpose of our case studies is to show how to estimate the posterior probability distribution of the parameters of an energy budget model $\mathcal{M}$, given an empirical dataset $\mathcal{Y}$, and accounting for the uncertainty in the data. The model takes the general form

$$\mathcal{M} \equiv \frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}_t, t, \boldsymbol{\theta}) \tag{1}$$

where $\boldsymbol{x}$ is a vector of state variables evolving with time; $\boldsymbol{f}$ is a functional operator that takes a time input and a vector of continuous functions $\boldsymbol{x}_t(\theta)$ and generates the vector $\frac{d\boldsymbol{x}}{dt}$ as output; and $\boldsymbol{\theta}$ denotes a set of parameters.

Using Bayes' Theorem (Hobbs and Hooten, 2015) we can calculate the posterior distribution of the model parameters, given the data and the prior information as

$$\Pr(\boldsymbol{\theta}|\mathcal{Y}) = \frac{\Pr(\mathcal{Y}|\boldsymbol{\theta})\Pr(\boldsymbol{\theta})}{\int \Pr(\mathcal{Y}|\boldsymbol{\theta})\Pr(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{2}$$

where Pr() denotes a probability, $\mathcal{Y}$ denotes the data, and $\boldsymbol{\theta}$ denotes the set of model parameters. The product in the numerator is the *joint distribution*, which is made up of the *likelihood* $\Pr(\mathcal{Y}|\boldsymbol{\theta})$ or $\mathcal{L}(\mathcal{Y}|\boldsymbol{\theta})$, which gives the probability of observing $\mathcal{Y}$ given the deterministic model $\mathcal{M}(\boldsymbol{\theta})$, and the *prior distribution* $\Pr(\boldsymbol{\theta})$, which represents the knowledge about $\boldsymbol{\theta}$ before the data were collected.

Closed form solutions for the posterior are practically impossible to obtain for complex non-linear models such as DEB models, but they can be approximated numerically, e.g. by using Markov Chain Monte Carlo (MCMC) integration methods (Gilks et al., 1995).

2.1 Computation using the `deBInfer` package

The `deBInfer` package (Boersch-Supan et al., 2017) provides templates for implementing dynamical models consisting of a deterministic differential equation (DE) model and a stochastic observation model and subsequently for performing inference for these models. To perform inference, R functions or data structures must be specified to represent both the dynamical (here bioenergetic) model and the observation model (i.e., the data likelihood). Further, all model and observation parameters

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

must be declared, including prior distributions for those parameters that are to be estimated or values for those that are fixed. The DE model itself can also be provided as a shared object, e.g. a compiled C function, which can considerably speed up inference (see Appendix S3 in Boersch-Supan et al., 2017). `deBInfer` then samples from the posterior distributions of parameters via MCMC, specifically using Metropolis-Hastings updates nested within a Gibbs sampler (Brooks et al., 2011). As each sample of the posterior distribution is obtained, the differential equation model must be solved numerically within the MCMC procedure.

3 Case Study 1: DEBKiss Model

The standard DEB model is a powerful framework to describe the bioenergetics of an organism across its full life cycle (Kooijman, 2010). However, that power comes at a cost of many complex equations with many parameters needing a great deal of data to parameterize properly. In an effort to develop a simpler model that still exhibits important features of the full DEB Theory Jager et al. (2013) developed the DEBKiss model. It is a model inspired by DEB but "with a strong focus on the KISS principle" (Jager et al., 2013). The main departures from DEB are the absence of a reserve buffer and a maturation state variable. This has the effect of reducing the number of differential equations in the system, as well as reducing the number of parameters. The model is slightly less flexible. For instance, the organism must mature at a fixed length or weight. In contrast, the DEB framework allows maturation to happen once sufficient complexity has been accrued, which can correspond to a different weights or lengths in organisms living in differing food environments.

We use the DEBKiss model as a simplified DEB-like model to show the the basics of the Bayesian approach to fitting models of this sort. We perform inference using the same data set used in the paper introducing DEBKiss: data on growth and reproduction of the pond snail, *Lymnaea stagnalis*. These data come from a series of part of partial life experiments. Juvenile snails that were the same age (113 days from when the egg was laid) and approximately the same size were selected and followed for an additional 140 days (Figure 1, (data from Zimmer et al., 2012)). The reproduction and length data were originally from different experiments, but we treat them together here, following Jager et al. (2013). To keep this example especially simple use data from a single food level treatment, specifically snails that were fed ad libitum over the course of the experiment. Further, we only estimate a subset of the parameters estimated in the original DEBKiss paper (which were, in turn, a subset of all of the parameters) as not all parameters are simultaneously identifiable at a single food level. The model is specified and described in detail in both the main and supplementary text of Jager et al. (2013), so we do not re-produce the equations here. However, the complete implementation of the model in R, including the inference shown here, is included as supplementary materials.

3.1 Bayesian Parameter Estimation

For simplicity, we focus on estimating a subset of parameters from data on snail growth and reproduction at a single food level. The DEBKiss model was implemented as a set of differential equations following (Jager et al., 2013). Similarly to the standard DEB model, the state variables in the DEBKiss framework are not all directly measurable. However, it is possible to specify how measured quantities, such as length and numbers of eggs, are related to the state variables. For

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

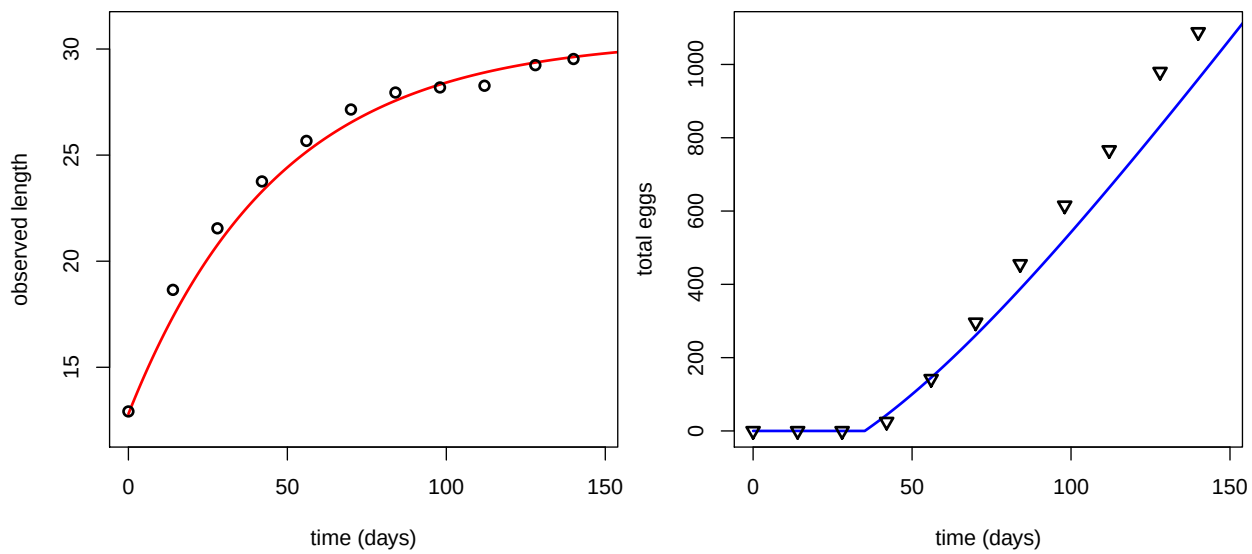

Figure 1: Snail growth and reproduction data (from Zimmer et al., 2012) plotted with predictions from the DEBKiss model (based on parameters from Jager et al., 2013).

this application, we used the formulation of the DEBKiss equations in terms of physical length and cumulative number of eggs produced by the snails.

Likelihood

Once the differential equations have been specified, the likelihood of the data conditional on the parameters and model must next be specified. The `deBInfer` package allows substantial flexibility in the probability distributions that may be used for this purpose. For instance, one could allow Normal errors for lengths and truncated or log-normal, or Poisson for eggs. This allows the user to choose an appropriate distribution that is consistent with the characteristics of the data the user wishes to model. The snail data we use here consists of average lengths (mm) and mean cumulative eggs. Thus both the lengths and eggs are appropriately modeled as continuous distributions. Additionally both are constrained to be positive and have error that increases over time (as small differences between individuals is likely to be magnified as the grow).

To define out likelihood, we must relate our measured quantities to the model outputs. We assume that the observed length, $\tilde{L}$, is the product of two quantities: the predicted physical length from the DEBKiss model, denoted as L_w and a log-normally distributed, multiplicative noise term. Thus:

$$\tilde{L}(t) = L_w(t)\epsilon_L \text{ with } \log \epsilon_L \sim \mathcal{N}(0, \sigma_L^2). \tag{3}$$

Similarly, the resources allocated to reproduction, W_R, are related to the number of eggs. However, the number of eggs created depend on the energy needed per egg, W_{B0}, and the conversion efficiency of assimilated energy to egg, y_{BA}. Again, the noise is assumed to be multiplicative and log-normal, so the cumulative egg production at any given time, F, is given by

$$F(t) = \frac{W_R(t) y_{BA}}{W_{B0}} \epsilon_E \text{ with } \log \epsilon_E \sim \mathcal{N}(0, \sigma_E^2). \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Conditional on the solution to the underlying differential equations we assume that the observed lengths and eggs are independent at each time. Thus the likelihood is constructed by multiplying the (independent) likelihoods of the lengths and fecundity at each time point together.

Priors and sampling details

We chose relatively un-informative priors for the four parameters that we chose to estimate. The choice of prior here was primarily guided by simple constraints on the values that the parameters may take. For example, κ, the proportion of energy directed towards growth processes, must lie between 0 and 1. Thus we used a uniform prior over this full range as the prior. Priors for estimated parameters are specified in Table 1.

Table 1: Prior distributions used for parameter inference in the DEBKiss model.

parameter		prior distribution	parameters	reasoning
κ	`kappa`	Uniform	$a = 0; b = 0$	bounded on [0,1]
$\ln(J_M^v)$	`logJMv`	normal	$\mu = 0; \sigma^2 = 100$	weakly informative prior, constraining the untransformed parameter to be positive.
$\ln \epsilon_L$	`sdlog.L`	lognormal	$\mu = 0; \sigma^2 = 1$	weakly informative prior regularizing towards 0
$\ln \epsilon_E$	`sdlog.E`	lognormal	$\mu = 0; \sigma^2 = 1$	weakly informative prior regularizing towards 0

In addition to a prior distribution, the user must specify a *proposal* distribution for each parameter being sampled (Gilks et al., 1995). This distribution determines how new values of each parameter are chosen as the MCMC algorithm progresses. In the `deBInfer` package one can choose between 3 options: 1) a random walk proposal with a normal proposal distribution centered at the current accepted parameter value; 2) a random walk proposal with a uniform distribution that is asymmetric around the current accepted value; 3) independent draws from the prior distribution. In the example here we chose all random walk proposals. For κ and $\ln(J_M^v)$ we used a normal proposal. For the two standard deviations we used the asymmetric uniform proposals. This latter is especially good for sampling parameters that are bounded on the left with values that may be close to that lower bound.

3.2 DEBKiss Model: Results

MCMC Output Diagnostics

When examining the posterior output from the MCMC produced by `deBInfer`, the first step is always to check for *mixing* and *convergence* of the MCMC chain by plotting traces of the chains (e.g., Figure 2). A "good", well behaved chain (i.e., that mixes adequately and that has converged to the posterior distribution) is sometimes described as resembling a "fuzzy caterpillar". A chain that doesn't look very fuzzy is often called a "sticky" chain – it has high auto-correlation and the effective sample size is low. If the chain has converged a horizontal line should approximately go through the center of the trace and there shouldn't be major patterns, such as a linear trend, visible. The chains for this example appear to be well behaved, and seem to indicate both adequate mixing

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and convergence. The subtleties of assessing mixing and convergence is beyond the scope of this paper, but may be found in textbooks such as Gilks et al. (1995) or Hobbs and Hooten (2015).

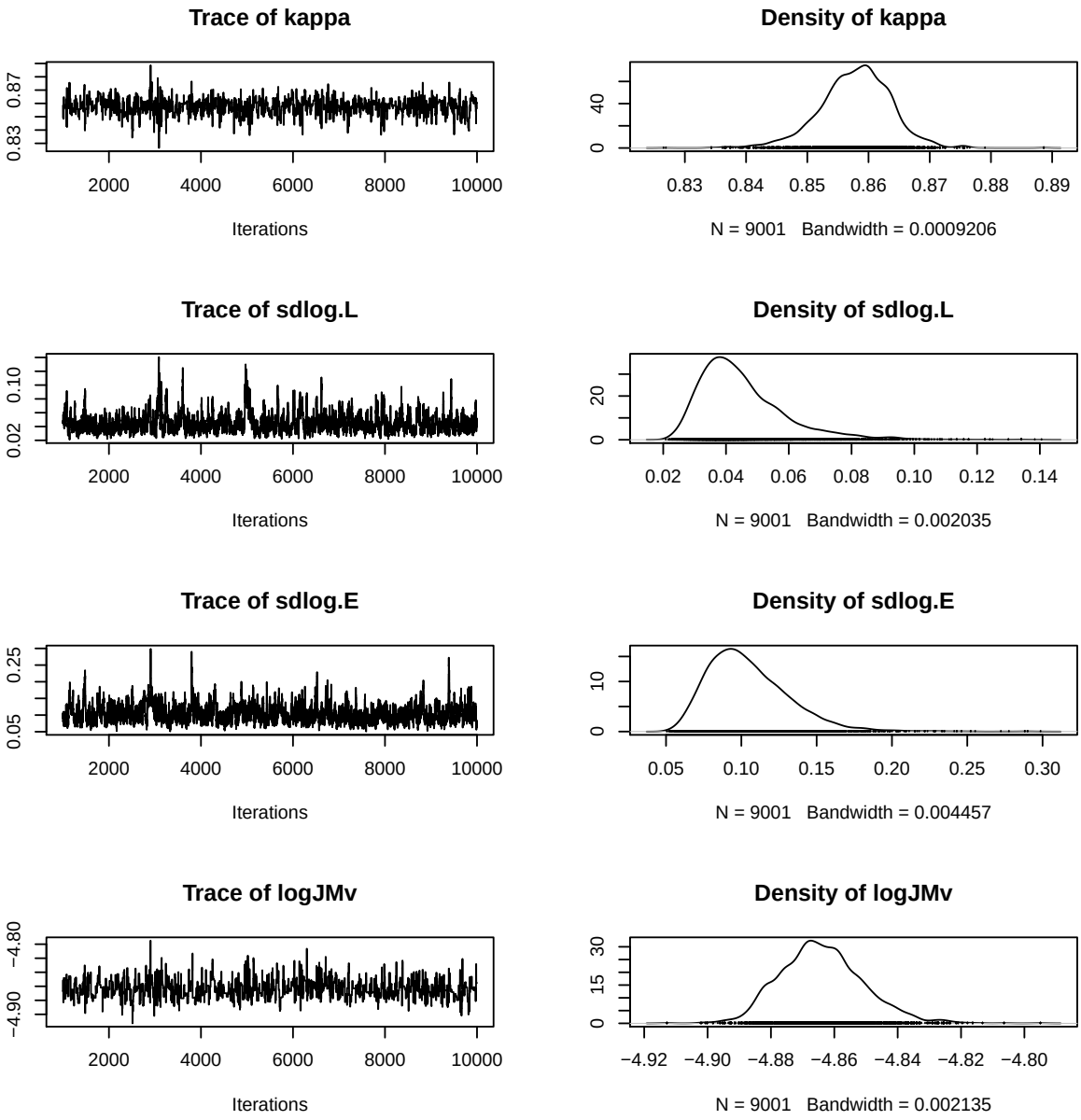

Figure 2: MCMC traces and marginal distribution for 2 observation and 2 primary model parameters estimated for the DEBKiss model.

Once mixing and convergence have been assessed, the next, very important, diagnostic to check is the prior-posterior overlap. Priors in Bayesian analyses can be double edged swords – they allow us to incorporate previous knowledge and constraints into our process in a formal way. However it is possible to inject more prior information than one means to through the prior. If you don't have good information about a parameter value, you ideally want to choose a "vague" prior so that the information in your data can drive the posterior results. Thus it is always a good idea to plot the marginal posterior distribution together with the marginal prior to confirm that your choice of

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

prior is not influencing your posterior more than you intended. In our example, even though we knew the values that (Jager et al., 2013) had previously reported for both parameters, we wanted to incorporate as little additional information as possible in our priors. In Figure 3 we show the marginal posterior distribution for each parameter (in black) overlaid with the prior distribution (in red). Notice that across all 4 panels the red line lies across the very bottom of the panel – the priors have very little mass in the areas corrsponding to the posterior distribution. In all cases the data seem to be informative for the parameters and the posteriors different from the priors.

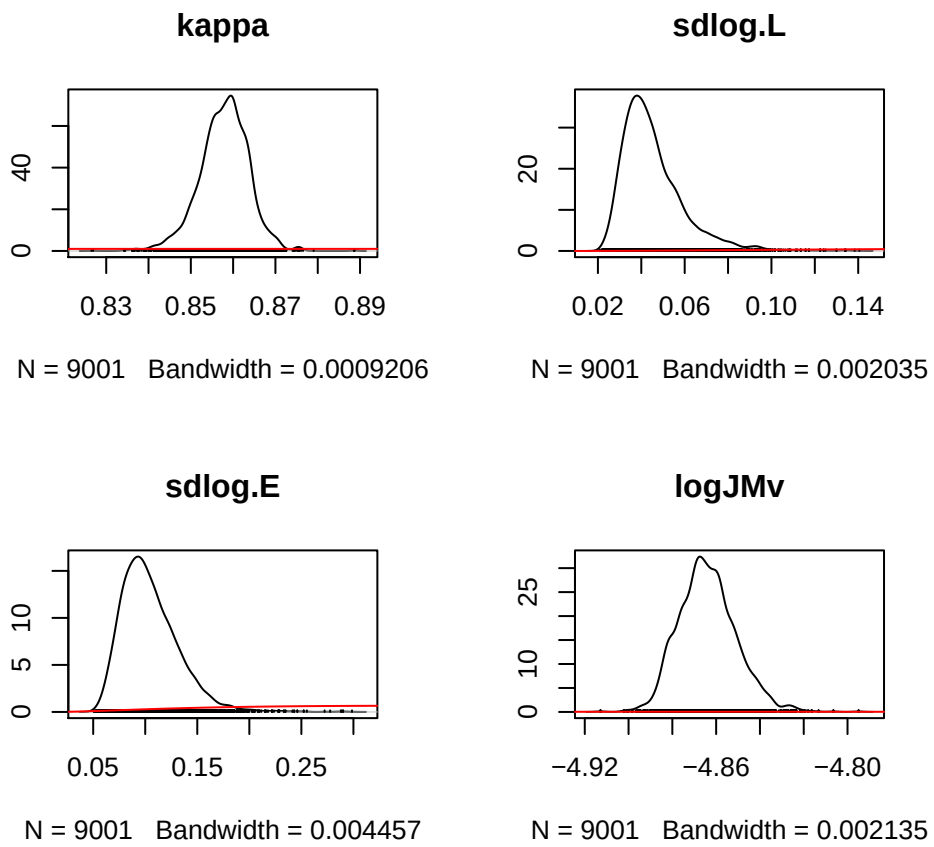

Figure 3: Marginal posterior samples of parameters (black lines) plotted with prior distributions (red). Notice that in all cases the prior is very different from the prior and the data are informative for all parameters.

We also typically examine the full joint posterior distribution by visualizing the pairwise joint distributions (e.g. Figure 4). The pairwise plots can give additional indications about the identifiability of individual parameters, independent of the others. In this example we can see that the correlation between our estimated parameters is overall very low, with the strongest correlation being (not surprisingly) between κ and $\ln(J_M^v)$. This is often the case. For nonlinear systems such as these, often the observation parameters are not highly correlated with the primary parameters, but the model parameters themselves may be. Very strong correlations between parameters can indicate that the data are not fully informative for each parameter individually – for example it may be that the quotient or product is very tightly constrained by the available data, but the marginal uncertainty in the individual parameters is higher. This is not necessarily problematic, per se, but

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

should be kept in mind when using and interpreting posterior samples. Further, some patterns in the posterior, such as extreme nonlinear patterns ("banana" shapes, etc.) can indicate that parameters are not well constrained. For an example of this for DEB models see Johnson et al. (2013).

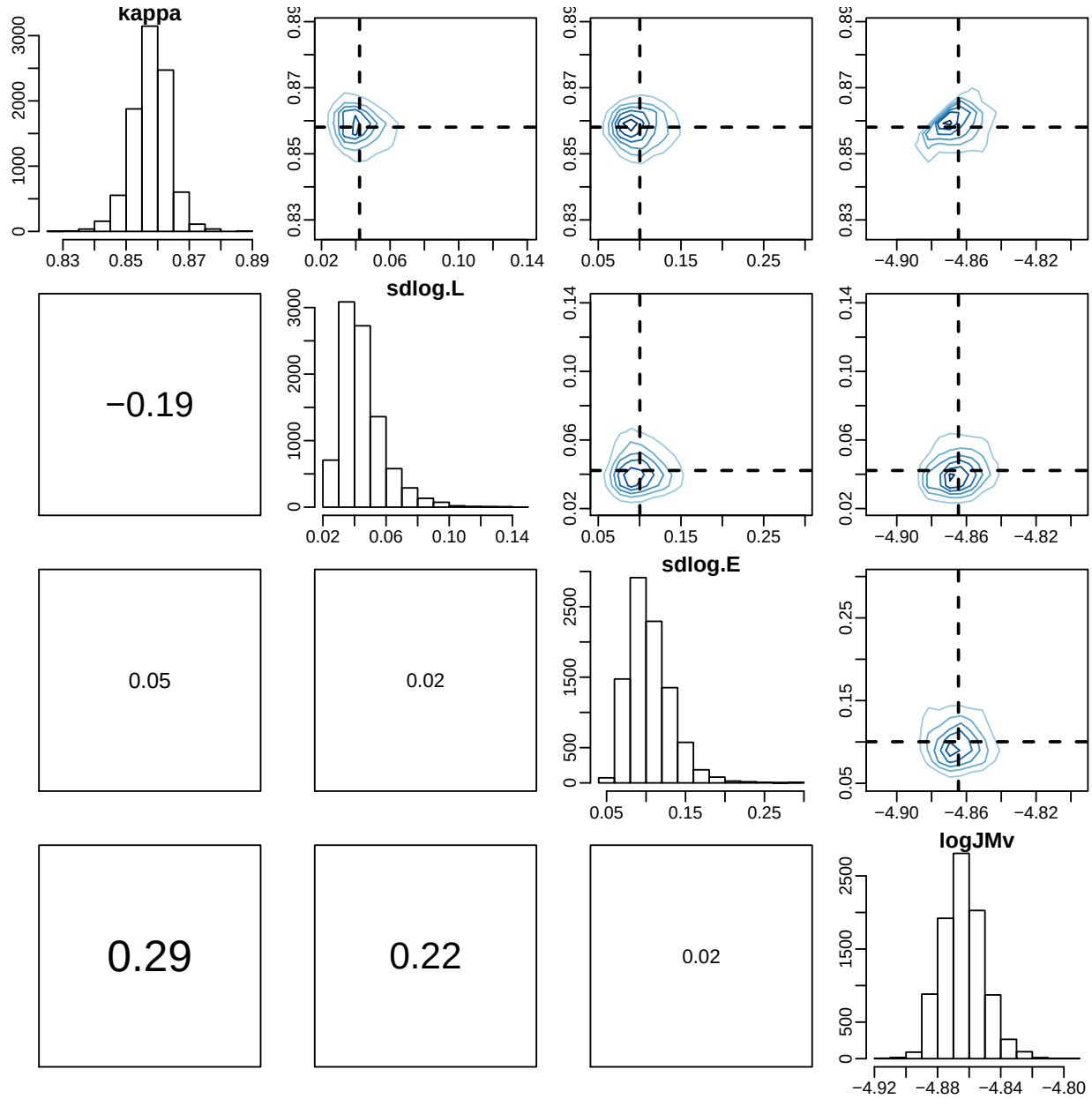

Figure 4: Joint samples from the full posterior of the 2 observation and 2 primary model parameters estimated for the DEBKiss model.

Posterior estimates of parameters

Now that we feel confident in the convergence of the chains and that our parameters are well identified we can interpret and present the inferred parameters, including the posterior uncertainties.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Further, we can use the posterior distribution of parameters to obtain the posterior distributions of other functions of the parameters, such as the trajectories of the system.

In a Bayesian analysis we often report the marginal highest posterior density (HPD) interval or credible interval, which are the Bayesian analogs to confidence intervals. In Figure 5 we show a graphical representation of the HPD intervals for the two primary parameters that we estimated. The shaded region corresponds to the HPD region (i.e., 95% of the posterior weight), and the HPD interval thus corresponds to the locations of the edges of the region. We indicate the posterior mean, often used as the point estimate for the parameter, using a solid line. For comparison we also indicate the "true" parameter (i.e., the estimate obtained by Jager et al. (2013)). Note that Jager et al. (2013) fit the model via maximum likelihood assuming normal errors on data that has been square-root transformed. This is a very different assumption about the underlying data model and is likely part of the reason for the discrepancy. The extra data from other experimental treatment groups would also provide additional information about parameter values.

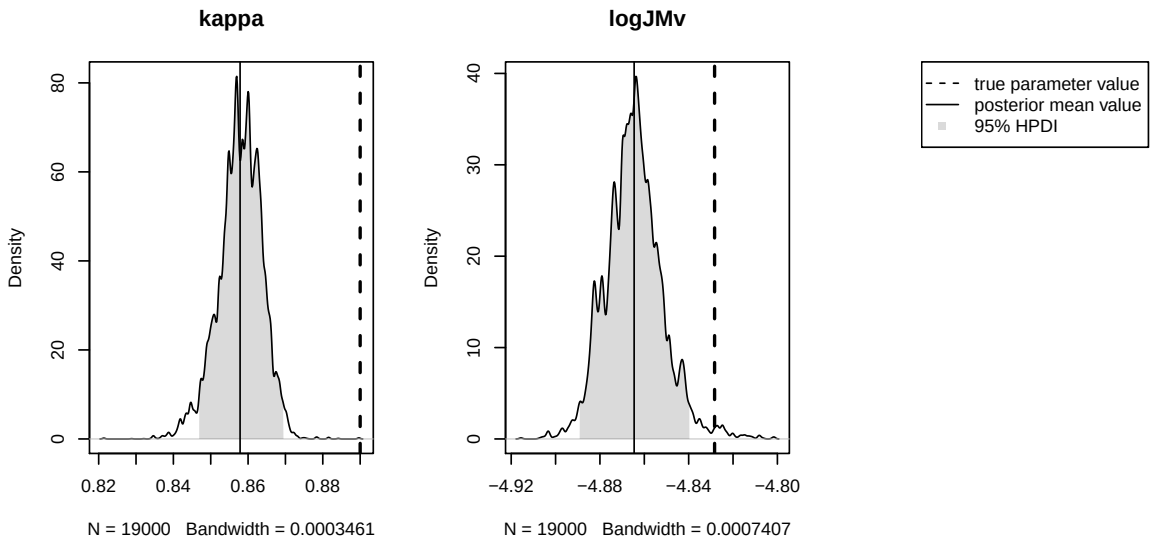

Figure 5: Marginal posterior distributions of two primary parameters κ and $\ln(J_M^v)$. Shaded areas indicate the 95% highest posterior density (HPD) region. The posterior mean is indicated with a solid line, and the "true" parameter (i.e., the estimate obtained by Jager et al. (2013)). Note that Jager et al. (2013) fit the model via maximum likelihood, assuming normal errors on data that has been square-root transformed, compared to our assumption of log-normal errors.

Finally, we can construct and visualize the posterior distributions of any functions of the parameters. For this example, we focus on the posterior distribution of the trajectories of the sets of differential equations. To obtain the posterior distribution of trajectories requires solving the set of differential equations with the parameters set to the estimated values in the posterior sample. For instance, in this example we collected $N = 20000$ samples of the posterior distribution of parameters. We discarded the first 1000 as burn-in (because for part of that the chain had not yet converged), leaving 19000 samples. We then "thinned" these samples (that is sub-sampled them), taking every 10th sample. This leaves 1900 parameter samples. For each of these samples we take the pair or primary parameter estimates together with the fixed parameters and initial conditions and solve the DEs. After repeating this for all 1900 samples we have 1900 trajectories of the system. We can

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

obtain the mean behavior by taking the mean at each time point in the trajectory across the 1900 samples. Similarly we can calculate the credible intervals by obtaining the appropriate values of the quantiles at each time point. The posterior mean and credible intervals of the trajectories for our example are shown in Figure 6. Notice that this fit is slightly different than the fit obtained by Jager et al. (2013). In particular this fit slightly over-estimates the egg production, whereas Jager et al. (2013) slightly under-estimate egg production. Both give similar fits for the length.

Note that, unlike most methods for obtaining uncertainty estimates of parameters (e.g. obtained via maximum likelihood or least squares) we do not need to assume that the parameters are approximately multivariate normal. Although that assumption may not be far off for the simple example here, there are certainly cases where that assumption is a poor representation of the posterior distribution, and where assuming the parameters are jointly normal would result in very different predictions of the system trajectories.

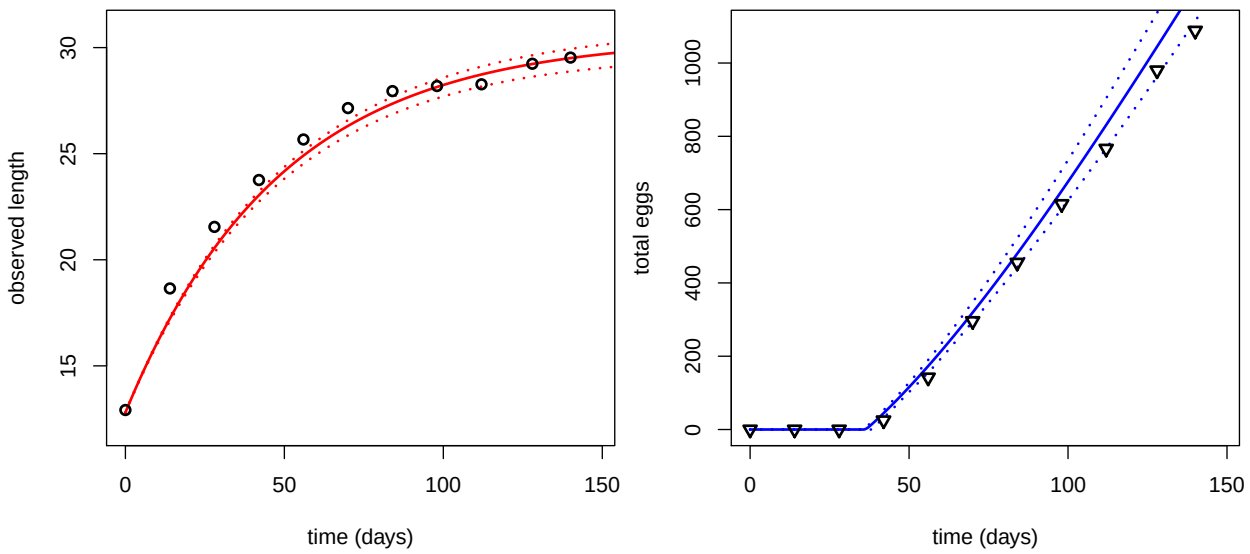

Figure 6: Posterior mean and 95% credible interval (CI) of the predicted growth and reproduction from the DEBKiss model plotted with the pond snail data. Solid lines indicate the posterior mean, and dotted lines the edges of the 95% CI.

4 Case Study 2: The Standard DEB Model

For our second case study we attempt to replicate the DEBtool estimation of growth and reproduction parameters for the standard DEB model for the earthworm *Lumbricus terrestris*. For the sake of simplicity we did not estimate ageing parameters. Data and DEB model parameters estimated with DEBtool_M (https://github.com/add-my-pet/DEBtool_M) in MATLAB R2014b (MathWorks, Inc.) were obtained from the add-my-pet database. To assess parameter identifiability and the precision of the posterior parameter estimates this case study is uses simulated observations based on the AmP parameters, rather than the original earthworm data from Butt (1993). Empirical distributions of DEB parameters for the construction of priors were obtained using AmPtool (https://github.com/add-my-pet/AmPtool).

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

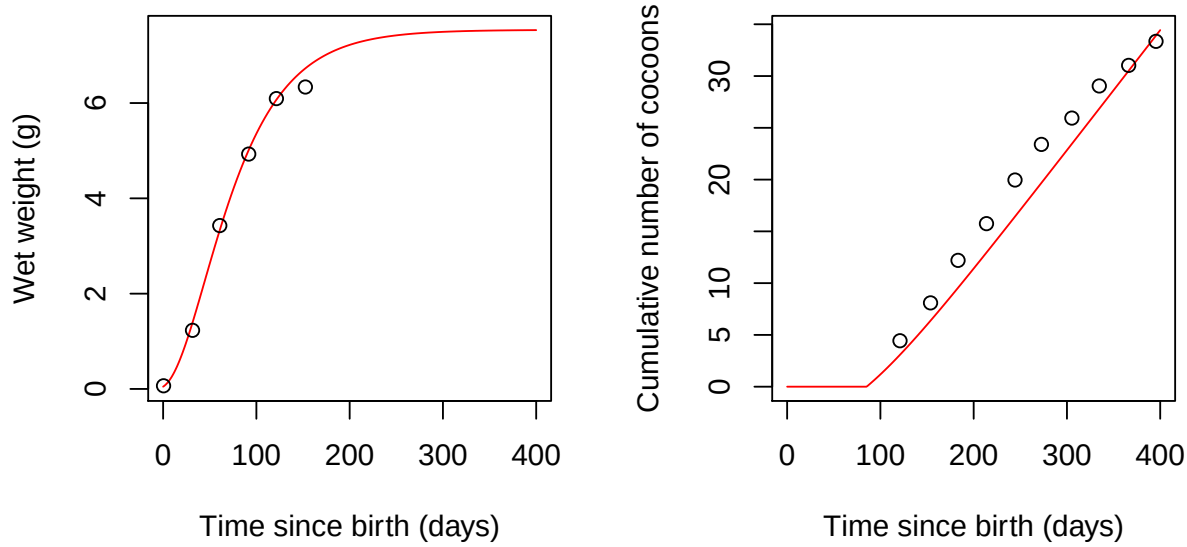

Figure 7: Earthworm growth and reproduction data (Butt, 1993) plotted with predictions from the standard DEB model based on parameters for *Lumbricus terrestris* from the AmP database.

4.1 Bayesian parameter estimation

We implemented the standard DEB model in terms of scaled energy density e, scaled length l, and scaled maturity and reproductive buffer u_H and u_R, respectively, (Kooijman, 2010, Table 2.5) as functions compliant with the ODE solvers provided by the `deSolve` package Soetaert et al. (2010). Further R functions to compute boundary values for the state variables from DEB parameters (Kooijman, 2009) were adapted from DEBtool routines and are available in the R package DEButilities (https://doi.org/10.5281/zenodo.1162331) which we provide in the supplementary materials.

Treating the initial value for the scaled reserve density u_E^0 as parameter dependent (Kooijman, 2009) necessitates a recalculation of two of the initial values, the scaled length l_b and maturity at birth u_H^b, respectively, for the DEB model within the MCMC inference procedure. This computational step is currently only implemented in a development branch of `deBInfer`, which is available on github (https://github.com/pboesu/debinfer/tree/recalc-inits) and provided in the supplementary materials.

Initial inference runs highlighted parameter identifiability issues, in particular the strong, non-linear correlation between L_m, $\{\dot{p}_{Am}\}$ and $\dot{v}$ made it impossible to estimate $[E_G]$. This could partially be resolved by using informative priors based on the empirical distribution of parameters in the AmP database (see below).

The strong parameter correlations further resulted in slow mixing of the MCMC chain, necessitating a large number of samples to explore the posterior distribution. We therefore conducted inference for this model using a compiled ODE model implemented in C to speed up inference.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We ran three separate MCMC chains of 150000 samples each. We discarded the first 10000 samples of each chain and thinned the remaining samples retaining every 10th sample.

4.1.1 Prior distributions

Priors on the parameters were chosen to be weakly informative, based on the principle that priors should contain enough information to rule out unreasonable parameter values but not values that might make sense. Hard bounds were used only where dictated by DEB theory. We further aimed to translate the pseudodata approach of the covariation method (Lika et al., 2011) into our choices of prior distributions and their location and scale parameters. Specific prior choices are detailed in Table 2.

4.1.2 Data model and likelihood

The state variables of the DEB model are abstract quantities that are not directly observable, but can be mapped to observable quantities using auxiliary parameters. We used the following mappings between the so-called zero-variate observable quantities and model states and parameters:

$$Ww_0 = u_E^0 \frac{v^2}{g^2} k_M^3 p_{Am} \frac{w_E}{\mu_E d_E}$$

$$Ww_b = (l_b L_m)^3 (1 + f\omega)$$

$$Ww_p = (l_p L_m)^3 (1 + f\omega)$$

$$a_b = \tau_b / \dot{k}_M$$

$$a_p = \tau_b + \rho_B^{-1} \log \frac{l_\infty - l_b}{l_\infty - lp}$$

where u_E^0, l_*, and τ_b were calculated following Kooijman (2009).

Further, the time series of wet weight $Ww(t)$ and reproductive output $R(t)$ were mapped from the model using the equations

$$Ww(t) = (l(t) L_m)^3 (1 + f * \omega)$$

$$R(t) = \kappa_R * u_R(t) / u_E^0.$$

The full likelihood of error at time t is as follows

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2: Prior distributions used for parameter inference in the standard DEB model.

parameter		prior distribution	parameters	reasoning
L_m	`L_m`	truncated normal	$a = 0; \mu = 1, \sigma^2 = 1$	weakly informative prior in the absence of species specific maximum length or weight data
$\{\dot{p}_{Am}\}$	`p_Am`	normal	$\mu = 900; \sigma^2 = 300$	weakly informative prior based on AmP collection
$\dot{v}$	`v`	truncated normal	$a = 0; \mu = 0.2, \sigma^2 = 0.2$	weakly informative prior based on generalised animal
κ	`kap`	Beta	$\alpha = \beta = 2$	bounded on [0,1], regularizing away from bounds
$[E_G]$	`E_G`	truncated normal	$a = \frac{d_V \mu_V}{w_E}; \mu = 4200; \sigma^2 = 100$	informative prior based on theoretical lower bound and AmP data scaled for $d_V = 0.16$
E_H^b	`E_Hb`	truncated normal	$a = 0; \mu = 0; \sigma^2 = 100$	weakly informative prior regularizing towards 0
E_H^p	`E_Hp`	truncated normal	$a = 0; \mu = 1000; \sigma^2 = 1000$	weakly informative prior regularizing to values ≈ 1000
$\ln \varepsilon_{Ww}$	`sdlog.EWw`	truncated normal	$a = 0; \mu = 0.1; \sigma^2 = 0.1$	weakly informative prior regularizing towards 0
$\ln \varepsilon_R$	`sdlog.R`	truncated normal	$a = 0; \mu = 0.1; \sigma^2 = 0.1$	weakly informative prior regularizing towards 0

Table 3: Fixed parameters used for inference in the standard DEB model. Values are based on the AmP entry for *Lumbricus terrestris*. No uncertainty estimates for ages and weights at stage transitions in *L. terrestris* were available in the literature. We therefore assumed a fixed standard deviation of 1% of the corresponding mean for those values.

parameter		value
$\dot{k}_J$	`k_J`	0.002 cm/d
T_A	`T_A`	5000 K
T_{ref}	`T_ref`	293.15 K
f	`f`	1
w_E	`w_E`	23.9 g/mol
d_V	`d_v`	0.16 g/cm^3
d_E	`d_E`	0.16 g/cm^3
μ_E	`mu_E`	550000 J/mol
μ_V	`mu_v`	500000 J/mol
κ_R	`kap_R`	0.95
σ^2_{a*}		$0.01 \times \mu_{a*}$
σ^2_{Ww*}		$0.01 \times \mu_{Ww*}$

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$
\begin{aligned}
\mathcal{L} = & \Delta_b \, \Delta_p \, \Delta_H \times \\
& \text{truncnorm}(\tilde{W}w_0 | W w_0, \sigma^2_{W w_0}, a = 0, b = \infty) \times \\
& \text{truncnorm}(\tilde{W}w_b | W w_b, \sigma^2_{W w_b}, a = 0, b = \infty) \times \\
& \text{truncnorm}(\tilde{W}w_p | W w_p, \sigma^2_{W w_p}, a = 0, b = \infty) \times \\
& \text{truncnorm}(\tilde{a}_b | a_b, \sigma^2_{a_b}, a = 0, b = \infty) \times \\
& \text{truncnorm}(\tilde{a}_p | a_p, \sigma^2_{a_p}, a = 0, b = \infty) \times \\
& \prod_t \text{lognormal}(\tilde{W}w(t) | \ln W w(t), \ln \varepsilon_{Ww}) \times \\
& \prod_t \text{lognormal}(\tilde{R}(t) | \ln R(t), \ln \varepsilon_R)
\end{aligned}
\tag{5}
$$

where the constraints on the parameter space follow from theoretical considerations detailed in Lika et al. (2014) about the animal reaching the stage transitions at birth and puberty:

$$
\Delta_b = \begin{cases} 0 & \text{for } l_b \geq f \text{ or } k\, v_H^b \geq \frac{f}{g+f}\, l_b^2\, (g + l_b) \\ 1 & \text{otherwise} \end{cases} \tag{6}
$$

$$
\Delta_p = \begin{cases} 0 & \text{for } k\, v_H^p \geq f(f - l_T)^2 \\ 1 & \text{otherwise} \end{cases} \tag{7}
$$

$$
\Delta_H = \begin{cases} 0 & \text{for } H_b \geq H_p \\ 1 & \text{otherwise} \end{cases} \tag{8}
$$

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4.2 Results: Standard DEB Model

Initial inference runs highlighted parameter identifiability issues, in particular the strong, non-linear correlation between L_m, $\{\dot{p}_{Am}\}$ and $\dot{v}$ (see the bana-shaped contours in the joint density plots for the variables in Figure 8) made it impossible to estimate $[E_G]$, even when using informative priors based on the empirical distribution of parameters in the AmP database. We therefore present inferences for a model where $[E_G]$ was fixed at the value of $4150\,\mathrm{J/cm^3}$ (Figure 9).

Despite the strong correlations, the AmP parameter values were recovered within the 95% highest posterior density interval, although the posterior means and modes diverged substantially from the AmP parameter values for L_m, $\{\dot{p}_{Am}\}$, and $\dot{v}$, and to a lesser extent for E_H^b (Figure 10).

Posterior trajectories for the earthworm DEB model (Figure 11) further indicate that the parameter identifiability issues encountered here are a structural property of the model, rather than a result of poor statistical fit. The posterior distribution of model trajectories is relatively narrow on the data scale, which is well constrained by the observations, but wide on the scale of the state variables.

5 Discussion

DEB theory and related bioenergetic models such as the DEBkiss framework have the potential to unify biological theory across levels of organization (Nisbet et al. 2000). However, to realize this potential robust methods are needed to connect the theoretical predictions with the inherently variable and noisy biological data.

We here demonstrate how DEB and DEBkiss parameters can be estimated in a Bayesian framework, a coherent, effective, and well-established approach, using the freely available `deBInfer` package in R. The Bayesian approach is particularly useful for DEB models as it provides a fairly straightforward way to incorporate prior information and at the same time provides a means to quantify uncertainty in parameters and predictions. DEB theory in its very core is focused on the individual animal, and the role of individual trait heterogeneity is increasingly recognized as an important factor underlying ecological dynamics (e.g. Cam et al., 2002; Vindenes et al., 2008; Jenouvrier et al., 2015). The Bayesian approach, in principle, provides a conceptually straightforward avenue for hierarchical inference for bioenergetic models, thereby opening a door to better understanding causes and effects of individual heterogeneity of metabolic traits. However, our results also highlight the difficulty of identifying DEB model parameters which serves as a reminder that fitting these models requires statistical caution.

In particular, we were not able to simultaneously estimate the same number of parameters for the standard DEB model for *Lumbricus terrestris* as are presented in the corresponding AmP entry, even when using priors based on AmP information.

Both the Bayesian approach presented here, and the DEBtool procedure make use of expert opinion to constrain the parameter estimation. The former through the choice of particular prior distributions and likelihoods, the latter by using pseudodata and setting weight coefficients for the least-squares estimation. Furthermore, the weighted least-squares method underlying the DEBtool estimation procedure does in principle provide variances and approximate covariances on parameter estimates, however, these are rarely if ever reported, and are not currently part of the AmP database. To better understand the sensitivities of parameter estimates we would encourage all

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

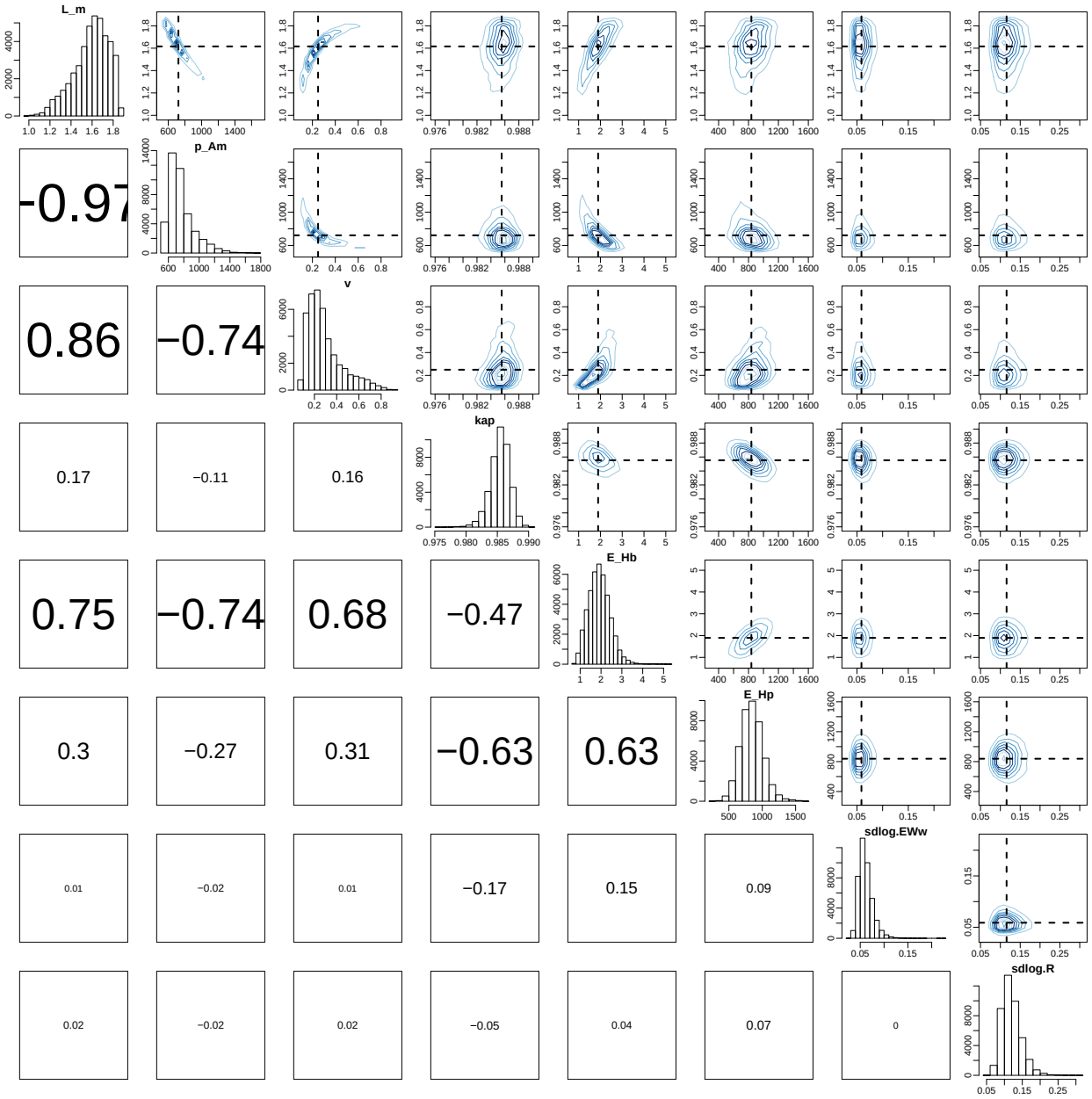

Figure 8: Pairwise correlations of posterior parameter estimates for the standard DEB model. E_G was fixed at $4150\,\mathrm{J/cm^3}$.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

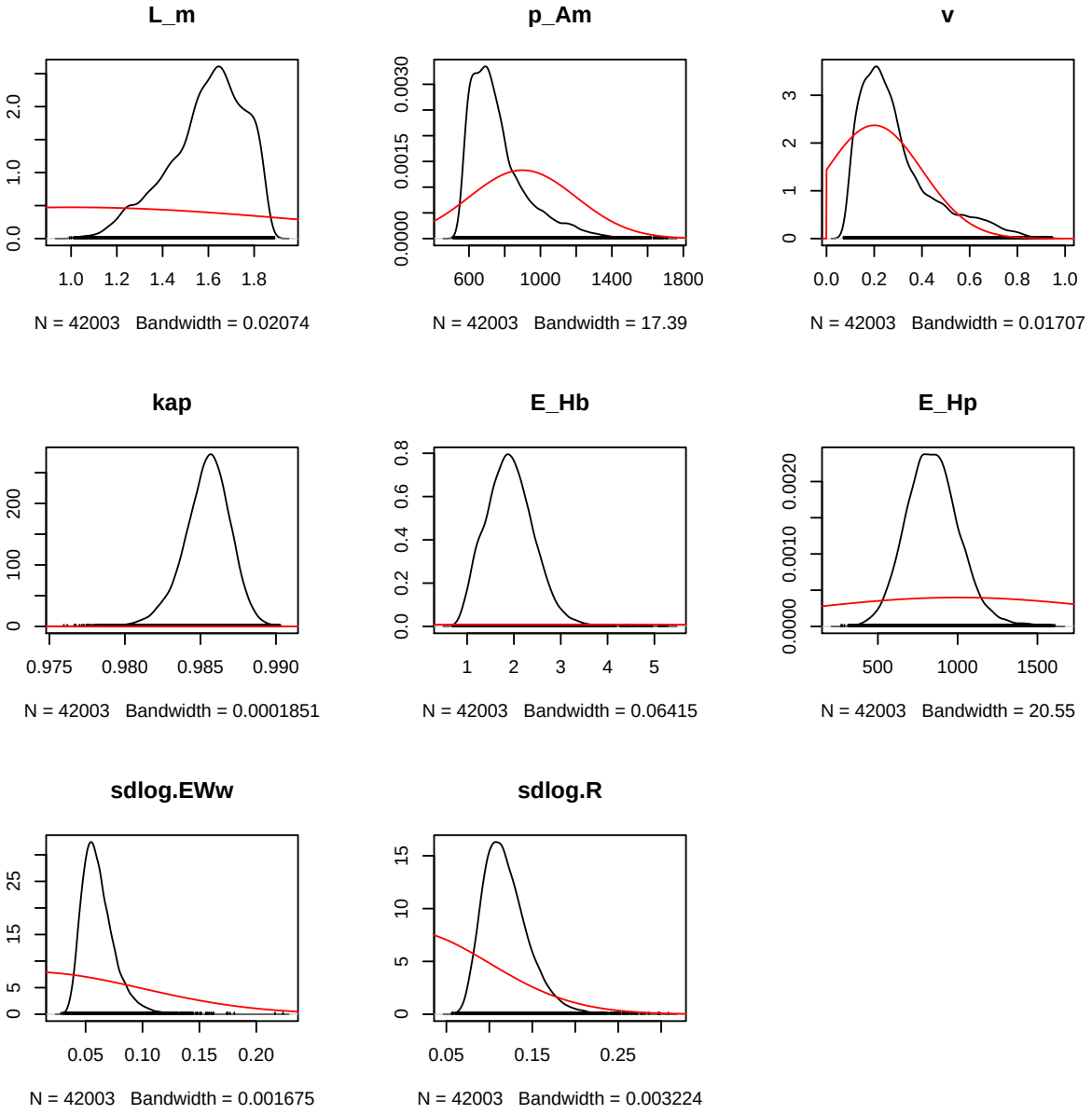

Figure 9: Even with a fixed value of E_G the data likelihood provided little information about the values of $\dot{v}$, L_m, and $\{\dot{p}_{Am}\}$, as is apparent from the substantial overlap between posterior and prior densities .

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

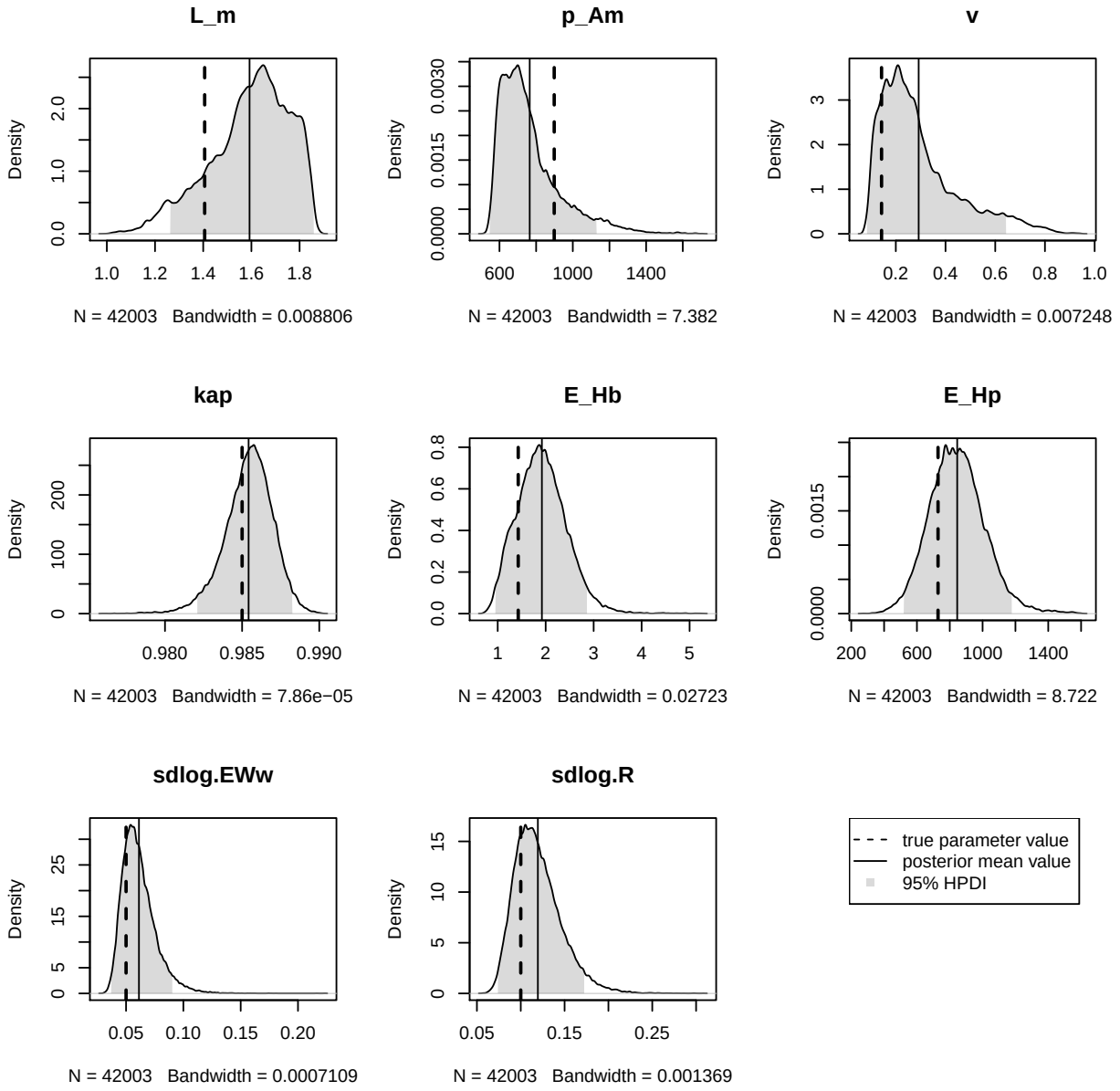

Figure 10: Eventhough several parameters of the standard DEB model were only weakly identified, the true values of all free parameter were recovered within the 95% highest posterior density interval.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

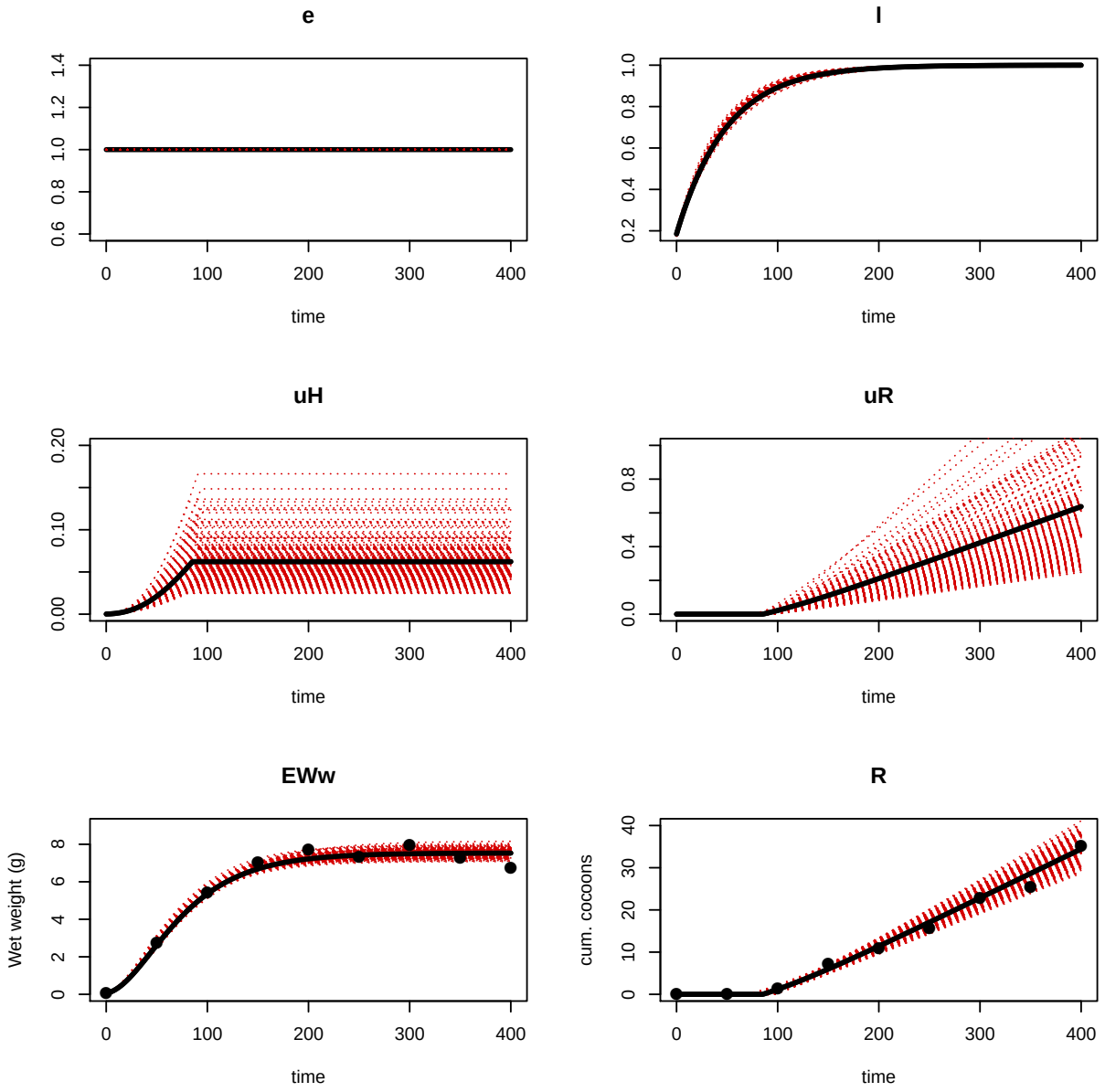

Figure 11: Posterior trajectories for the earthworm DEB model. The posterior distribution of model trajectories is relatively narrow on the data scale, but wide on the scale of the state variables. This indicates that the weak identifiability of several parameters is structural, rather than a consequence of poor statistical fit. The red lines are a random sample of 500 posterior trajectories. The black line represents the "true" trajectories on which the simulated observations are based, black circles show the simulated univariate data used in the inference procedure.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DEB practitioners to report choices made to constrain the parameter estimation, as well as variance and covariance estimates for estimated parameters.

6 Acknowledgements

The authors were supported by the US National Science Foundation (Grant PLR-1341649). We thank Sadie Ryan and the participants of the 5th international symposium on Dynamic Energy Budget theory for comments and discussions on earlier versions of this work.

7 Code and Data availability

The earthworm data and parameters were obtained from the add-my-pet library (AmP; http://bio.vu.nl/thb/deb/deblab/add_my_pet/) entry *Lumbricus terrestris* version 2015/12/07,

Code and derived data sets for this paper are available online.

- DEButilities 0.1.0 is archived on zenodo (https://doi.org/10.5281/zenodo.1162331)
- deBInfer 0.4.1 is available on CRAN (https://CRAN.R-project.org/package=deBInfer)
- deBInfer 0.4.1.9000-recalc-inits is available on github (https://github.com/pboesu/debinfer/tree/recalc-inits)
- Simulation and inference code for this paper are archived on zenodo (https://doi.org/10.5281/zenodo.1162942)

References

R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems*. Academic Press, Amsterdam, 2nd edition, 2011.

P. H. Boersch-Supan, S. J. Ryan, and L. R. Johnson. deBInfer: Bayesian inference for dynamical models of biological systems in R. *Methods in Ecology and Evolution*, 8(4):511–518, 2017.

D. Brewer, M. Barenco, R. Callard, M. Hubank, and J. Stark. Fitting ordinary differential equations to short time course data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 366(1865):519–544, 2008.

S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL, 2011.

K. Butt. Reproduction and growth of three deep-burrowing earthworms (lumbricidae) in laboratory culture in order to assess production for soil restoration. *Biology and Fertility of Soils*, 16(2): 135–138, 1993.

E. Cam, W. A. Link, E. G. Cooch, J. Monnat, and E. Danchin. Individual covariation in life-history traits: Seeing the trees despite the forest. *The American Naturalist*, 159(1):96–105, 2002. doi: 10.1086/324126.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. S. Clark. *Models for Ecological Data: An Introduction*. Princeton University Press, Princeton, New Jersey, USA, 2007.

F. C. Coelho, C. T. Codeço, and M. G. M. Gomes. A Bayesian framework for parameter estimation in dynamical models. *PloS one*, 6(5):e19616, 2011.

A. Gelman, F. Bois, and J. Jiang. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91(436): 1400–1412, 1996.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2003.

W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

N. T. Hobbs and M. B. Hooten. *Bayesian models: a statistical primer for ecologists*. Princeton University Press, Princeton, New Jersey, USA, 2015.

T. Jager and E. I. Zimmer. Simplified dynamic energy budget model for analysing ecotoxicity data. *Ecological Modelling*, 225:74–81, 2012.

T. Jager, E. H. Heugens, and S. A. Kooijman. Making sense of ecotoxicological test results: towards application of process-based models. *Ecotoxicology*, 15(3):305–314, 2006.

T. Jager, B. T. Martin, and E. I. Zimmer. Debkiss or the quest for the simplest generic model of animal life history. *Journal of Theoretical Biology*, 328:9 – 18, 2013.

S. Jenouvrier, C. Péron, and H. Weimerskirch. Extreme climate events and individual heterogeneity shape life-history traits and population dynamics. *Ecological Monographs*, 85(4):605–624, 2015.

L. R. Johnson and C. J. Briggs. Parameter inference for an individual based model of chytridiomycosis in frogs. *Journal of theoretical biology*, 277(1):90–98, 2011.

L. R. Johnson, L. Pecquerie, and R. M. Nisbet. Bayesian inference for bioenergetic models. *Ecology*, 94(4):882–894, 2013.

S. Kooijman. Waste to hurry: dynamic energy budgets explain the need of wasting to fully exploit blooming resources. *Oikos*, 122(3):348–357, 2013.

S. A. Kooijman. What the egg can tell about its hen: embryonic development on the basis of dynamic energy budgets. *Journal of Mathematical Biology*, 58(3):377–394, 2009.

S. A. L. M. Kooijman. *Dynamic energy budget theory for metabolic organisation*. Cambridge University Press, 2010.

K. Lika, M. R. Kearney, V. Freitas, H. W. van der Veer, J. van der Meer, J. W. Wijsman, L. Pecquerie, and S. A. Kooijman. The "covariation method" for estimating the parameters of the standard dynamic energy budget model i: philosophy and approach. *Journal of Sea Research*, 66(4):270–277, 2011.

K. Lika, S. Augustine, L. Pecquerie, and S. A. Kooijman. The bijection from data to parameter space with the standard deb model quantifies the supply–demand spectrum. *Journal of theoretical biology*, 354:35–47, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/259705; this version posted February 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. L. Maino, M. R. Kearney, R. M. Nisbet, and S. A. Kooijman. Reconciling theories for metabolic scaling. *Journal of Animal Ecology*, 83(1):20–29, 2014.

B. T. Martin, E. I. Zimmer, V. Grimm, and T. Jager. Dynamic energy budget theory meets individual-based modelling: a generic and accessible implementation. *Methods in Ecology and Evolution*, 3(2):445–449, 2012.

I. M. Smallegange, H. Caswell, M. E. Toorians, and A. M. Roos. Mechanistic description of population dynamics using dynamic energy budget theory incorporated into integral projection models. *Methods in Ecology and Evolution*, 8(2):146–154, 2017.

M. J. Smith, D. P. Tittensor, V. Lyutsarev, and E. Murphy. Inferred support for disturbance-recovery hypothesis of North Atlantic phytoplankton blooms. *Journal of Geophysical Research: Oceans*, 120(10):7067–7090, 2015.

K. Soetaert, T. Petzoldt, and R. W. Setzer. Solving differential equations in R: package deSolve. *Journal of Statistical Software*, 33(9):1–25, 2010.

Y. Vindenes, S. Engen, and B.-E. Sæther. Individual heterogeneity in vital parameters and demographic stochasticity. *The American Naturalist*, 171(4):455–467, 2008.

J. Voyles, L. R. Johnson, C. J. Briggs, S. D. Cashins, R. A. Alford, L. Berger, L. F. Skerratt, R. Speare, and E. B. Rosenblum. Temperature alters reproductive life history patterns in *Batrachochytrium dendrobatidis*, a lethal pathogen associated with the global loss of amphibians. *Ecology and Evolution*, 2(9):2241–2249, 2012.

E. I. Zimmer, T. Jager, V. Ducrot, L. Lagadic, and S. A. L. M. Kooijman. Juvenile food limitation in standardized tests: a warning to ecotoxicologists. *Ecotoxicology*, 21(8):2195–2204, Nov 2012.